

# *Using behavioral data to shed light on consumers' online travel planning*

Master thesis

*Grigor Dimitrov*

A thesis presented for the degree of  
Master of science



Department of Business and Economics  
Erasmus School of Economics  
Rotterdam, The Netherlands  
15 December, 2016"

*Supervisor: Daryna Kolesnyk*

*First-reader: Zhiying Jiang*

## **Abstract**

“This paper investigates the relationship between consumers’ personality traits and their travel-related online behavior. The results prove the positive relationship between general risk and general uncertainty attitudes and online travel-related behavior, which is associated with”unique domains“,”time" and “pageviews”. The results for mobile data, however, show the opposite direction of the effect, associated with uncertainty attitude and no significant relationship between the investigated metrics and risk attitude. The significant difference between desktop and mobile online behavior should be further investigated. The novelty of the paper lays in the methodological contributions, known among practitioners as “data fusion”. It utilizes both behavioral data to gather information about the observable online behavior and survey data to assess consumers’ unobservable personality traits.”

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Literature review</b>	<b>8</b>
2.1	Travel related activities and information search . . . . .	8
2.2	Risk and uncertainty . . . . .	11
2.3	Risk and uncertainty in tourism . . . . .	12
2.4	Big Five Factors Inventory (BFI) . . . . .	14
2.5	Personality traits and tourism information search . . . . .	15
2.6	Hypotheses . . . . .	16
<b>3</b>	<b>Methodology</b>	<b>17</b>
3.1	Introduction . . . . .	17
3.2	Behavioral data & Technology . . . . .	18
3.3	Steps . . . . .	18
3.3.1	Panel and sampling . . . . .	18
3.3.2	Survey . . . . .	19
3.3.3	Data processing . . . . .	21
3.3.4	Categorization . . . . .	22
3.3.5	Analysis . . . . .	23
3.3.6	Model selection . . . . .	23
<b>4</b>	<b>Results</b>	<b>27</b>
4.1	Survey Data . . . . .	27
4.1.1	Risk and uncertainty . . . . .	28
4.1.2	BIG5 . . . . .	28
4.2	Main results . . . . .	29
4.2.1	Desktop Data . . . . .	29
4.2.2	Mobile data . . . . .	31
4.3	Conclusion . . . . .	33
<b>5</b>	<b>Conclusions</b>	<b>35</b>
5.1	Discussion . . . . .	35
5.2	Contribution and Implications . . . . .	37
5.3	Limitations & Recommendations . . . . .	38



# 1 Introduction

In the last few years we created 90% of the world's data, this claim has been made in 2013 and it still consistent today. More precisely it is estimated that every two years in the last three decades the amount of data increases by ten times (Chatfield 2016) . The advantages of such "big data" include possibilities to uncover heterogeneities and subtle population patterns" which are difficult to obtain using self-reported data (Fan, Han, and Liu 2014). However, "big data" has been underutilized due to computational and statistical challenges (Ammu and Irfanuddin 2013, Tole and others (2013), Labrinidis and Jagadish (2012)). Moreover, Nunan and Di Domenico (2013) claim that there are high costs associated with analyzing and storing big data. One of the instances of "big data" is called behavioral data. In the context of this research "behavioral data" refers to recording human behavior, more precisely I am referring to "online behavioral data" which is clickstream data from subject's browser interactions and the Internet.

The survey questionnaires are foremost used when conduction primary market research data collection. Since the emergence of Web 2.0, the online survey made its way as it is a cheap and convenient form or data collection (Evans and Mathur 2005). With the increased penetration of the internet, there were more people online and this availability gave survey market research a huge reach. For example, in 2002 500million dollars were spent on online surveys while this figure doubled in 2004 (Evans and Mathur 2005).

However, 20 years ago when the survey plunged, our economic behavior wasn't focused that much online and the Internet was used mainly for information purpose. Surveys conducted back then explored topics related to our consumer offline behavior replacing standard paper and pencil interviewing popular in the 80's. Now more than ever our economic activity is focused online, our decision-making process is reflected online and it happens that online we can observe and capture information with ease. Consequently, there is an increasing availability of data that reflects this behavior and our online activities are becoming more valuable to researchers. Due to the economic value that online behavior can drive, there are emerging passive data collection technologies. Such technologies allow us to observe and capture all the online consumer behavior.

Behavioral data is purely observational, such data is more accurate in terms of describing actual behavior i.e. information gathering, research of alternatives and transactions which are the ultimate reflection of preference. On the other hand, survey

data is relevant when we have to assess unobservable information such as people's attitudes, motivations, and opinions. Furthermore, it is primary data collected for the specific research goals (Hox and Boeije 2005, Glass (1976)). These two types of data sources have their limitations and researchers are utilizing them separately, however, up to date, there is a limited amount of research that combines both sources of data. With the aid of passive data collection panels, this project provides a rare opportunity to match an actual online behavior with a subject behind it. That is, by combining passive online behavioral data with survey data, in this research I will be able to access information on both consumers' actual online behavior as well access unobservable information about consumers' personality traits.

The context of this research is the online travel industry amongst Dutch consumers. Tourism is highly competitive and fragmented market. It has been disrupted by the Internet at large, the disruption process has started with the emergence of the Internet and continues today. Before the penetration of the Internet the market was dominated by "high street" travel agencies, afterward the market was revolutionized by online travel agencies and direct distribution of travel services as accommodation and transportation.

We are witnessing dynamic segments of consumers emerged because of the technological advancements with constantly changing needs (Xiang, Magnini, and Fesenmaier 2015). This domain has been chosen due to the fact is one of the fastest growing online industry. For example, in the Netherlands for 2 out of 3 trips (69%) consumers book accommodations online according to Eurostat data (Eurostat 2016) Moreover, 40% of the Europeans used the internet for travel related purposes, in the Netherlands this number is higher than the average. Online presence of accommodation businesses in 2015 accounts for 95% of all enterprises compared to 75% for the whole economy. According to Euromonitor data, the sector is one of the fastest growing, rising from 899 million trips in 2009 to 1.4 billion trips in 2019 (Euromonitor 2015). Furthermore, there is an increasing availability of options and all sorts of services bringing the consumers and offers closer together manifesting the ongoing process of disintermediation. Xiang, Magnini, and Fesenmaier (2015) notes that there is growing "bifurcation" or a split among the traditional online travelers to users of traditional travel products and people seeking deeper and authentic experiences. The authors point out that understanding how contemporary travelers use the internet is an important foundation for building successful communication strategies by the business stakeholders. The importance of

the internet has been attributed to three main factors, the extensive amount of travel related information, the development of social networks, travel related social networks and peer-to-peer travel offerings where the user can exchange travel services and experiences and the mobile computing, smartphones in particular (Euromonitor 2015).

The increasing availability of travel related options online and increasing disintermediation lowers the costs of the travel related services and activities. However, this also makes the decision-making process more demanding as consumers themselves are primarily responsible for choosing the best option rather than using an advice from a travel service provider. Thus, the decisions related to travel involve a lot of risk and uncertainties. Arguably one of the risk and uncertainty reducing instruments that consumers use in decision making related to travel is an information search. Urbany, Dickson, and Wilkie (1989) introduces a body of research to support the link between uncertainty and information search. When it comes to perceived risks, in a meta-analysis Gemünden (1985) review one hundred empirical findings to conclude that information search is one of the risk reducing instruments consumers rely on when committing to purchases of complex products.

In an exploratory analysis, Roehl and Fesenmaier (1992) identify seven types of perceived risks related to leisure travel. Namely, equipment risks, financial risks, physical risks, psychological risks, satisfaction risks, social risks and time risks. Lepp and Gibson (2003) look at the general perceptions of risk and uncertainty from the perspective of travel motivations for novelty against familiarity. Their findings suggest that familiarity seekers are more risk averse comparing to less risk averse novelty seekers. With regards to risk attitude and tourism information search, Quintal, Lee, and Soutar (2010) examine the impact of risk and uncertainty avoidance on tourists' information search. The team distinguishes risk from uncertainty and investigate whether the two constructs have a different impact on the information search in terms of a number of sources. Their findings suggest that uncertainty avoidance has a positive significant relationship with the extent of information search while this is not valid for the risk averseness.

The following research further explores the relationship between risk and uncertainty attitudes and different instances of online information search behavior. More precisely, this research explores the relationship between risk and uncertainty attitudes, and information search behavior in tourism context while controlling for trip characteristics

and demographics. The main objectives of this paper are to investigate the relationship between different risk and uncertainty attitudes, captured by means of self-reported data along with activities on travel related domains such as depth, breadth of the information search, travel micromoments and total time spend on travel-related websites captured by means of online passive behavioral data.

One of the main motivations of this paper lays in the methodological and managerial contributions. Namely, demonstrating a full cycle of research, where the research design utilizes observable online behavioral data collected via passive tracking technology combined with subjects' unobservable characteristics collected via online questionnaire, would provide guidelines to managers in how to best process and exploit their tracking data to extract valuable insights. Up to date, there is a limited amount of literature exploring these constructs in relationship to one another and in a relationship with non-self-reported metrics of information sought. Moreover, the research has theoretical contributions as it combines literature streams that have been researched only using self-reported data. Self-reported is prone to biases as opposed to passive data, which is further discussed in the next chapter.

To improve the research quality, I combine both self-reported, recalled information with objective, passively measured data. Using passive metering technology, I can capture the full footprint of consumers across their devices. This data can help understand the time spent planning the travel, the amount of information search with great accuracy and detail. Although prone to biases, survey data is still important as it can capture unobservable information which is still important and valuable. Merging the two datasets can provide an even deeper layer of insights into consumers' travel behavior.

The study findings are relevant to all travel stakeholders such as businesses, travelers, and researchers. The increased availability of data from consumers and its right exploitation can have a huge impact on the design of travel related products, their personalization and cross-platform usage, in-line with the trends reviewed in the industry.



## 2 Literature review

In this chapter, I will review the main literature stream relevant for the hypnotized relationships in this paper. First, I will review online travel related activities and information search. Second, I will focus on risk and uncertainty attitudes and expand to risk and uncertainty attitudes related to travel and tourism. Third, I will briefly explore the big five inventory and its utilization in tourism research. In the final section of this chapter, I will derive the hypotheses of this research.

### 2.1 Travel related activities and information search

Tourism related information search and planning has been a widely researched topic. Internet search behavior in tourism context has been examined primarily from the perspective of demographic variables, motivation and prior knowledge about the destination (Jani, Jang, and Hwang 2014). In a paper exploring longitudinal data for 12 years, Xiang, Magnini, and Fesenmaier (2015) identified as a key trend number one that the internet penetration among consumers using the Internet for travel planning already reached the level of saturation. However, in a research exploring the online behavior from a generation perspective, H. Kim, Xiang, and Fesenmaier (2015) noted that there is no sufficient amount of research with regards to how the newly emerged segments of consumers behave online and use the internet for travel planning.

Xiang, Magnini, and Fesenmaier (2015) turn special attention on travel planning process as a specific type of information search that is an important component of the decision making in tourism. During the planning, consumers obtain information based on which they choose their destinations and form their expectations. There is a substantial amount of research that looks at travel planning from different perspectives, some aim to identify the characteristics of the travelers' demographic characteristics (e.g. D.-Y. Kim, Lehto, and Morrison (2007)), other investigate the way travelers are conducting purchases and navigate in the information stream (e.g. Jun, Vogt, and MacKay (2007)), and recently social media and its influence on travel have been in the focus of researchers as well (e.g. Xiang and Gretzel (2010)).

From a practitioner standpoint, the more consumers are being active online the more prerequisites this creates for the tourism stakeholders marketers to reach them during

their decision-making process. Therefore, identifying the travel planning process i.e. the decision-making journey is a critical step for the brands to intervene and influence the consumers in their direction (Chatfield 2014). Practitioners define that the decision-making process related to the tourism as an array of micro moments which often are not even conscious to the consumers. A micromoment is essentially a user session with the particular goal of obtaining information or committing to a purchase.

For the purpose of this research, it is important to define the main components of the travel planning. In a meta-analysis, Jun, Vogt, and MacKay (2007) point out that there is a consensus among researchers that travel planning cannot be simplified to a single goal-oriented rational action but it is rather viewed as a complex task involving multiple goals and decisions around the different goals and characteristics of the trip. The authors define a conceptual model for travel planning which has three main sequential interrelated components, pre-trip, during trip and post-trip. This research focuses on pre-trip phase. Pre-trip phase itself consists of information search and planning (decision making), furthermore, travel related purchases also occur at the end of this phase as well as during the trip itself.

Jun, Vogt, and MacKay (2007) define travel plan as a complex decision involving an assessment of multiple alternatives organized around the travel goals in mind. The planning process includes setting goals and considering alternatives in order to achieve that goals including an evaluation of different alternatives' outcomes. Planning is dependent on the all information search behavior, utilization of the obtained information, purchase behavior and activities including past experience. Pan and Fesenmaier (2006) define vacation planning over the internet as an interaction between the user and the "online space" related to destinations and tourism. The online space contains content provided by diverse sources and the technology that facilitates the communication. User's "situation, knowledge and skills" combined with the "online space" contribute to the effective search.

It is important to be noted that trip planning is an important and enjoyable part of the vacation experience itself (Stewart and Vogt 1999) and it is likely to be high costs and high involvement decision (Bonn, Furr, and Susskind 1998). Furthermore, travel information search behavior explains travel purchase behavior (Woodside and MacDonald 1994). Quintal, Lee, and Soutar (2010) review numerous aspect of from which information search has been researched including the amount of search, the

number of sources, the search process, involvement, socio-demographic differences, culture etc.

Pan and Fesenmaier (2006) review the consumer vacation planning process from micro level perspective. Their research is motivated by the fact that previous research has been mostly focused on exploring planning and information search on macro level i.e. motivation, need, determinants, and outcomes. Their research focuses on a “snapshot” of travel planning where subjects make choices regarding a hypothetical holiday trip to San Diego. Using such setting it was possible to observe different chapters containing many episodes on how consumers adapt in their online search and come up with a final decision. These episodes or micromoments defined above represent one of the main variables of interest of this research.

Additionally, I will use the metrics defined by Ho and Liu (2005) in “An exploratory investigation of web-based tourist information search behavior”. The researchers explored and characterized the online behavior of 96 subjects in laboratory research by assigning them one tourism related task and recording their desktop activity. Ho and Liu (2005) empirically analyzed the recorded videos and reported on different measures, representing the information search behavior.

Additionally to micromoments, in this paper, I will adapt measures utilized in Ho and Liu (2005). The full list of dependent variables representing the internet search behavior includes:

- Total number of micromoments on travel related websites
- Total number of unique travel related domains (breadth of search)
- Total number of pageviews on travel related domains (depth of search)
- Total time spend on travel related domains
- Total sum of all travel related micromoments (this measure has been introduced due to one particular limitation of the mobile measurement discussed in the limitations section)

Ho and Liu (2005) conducted the research in laboratory environment exploring a single goal-directed session of activity. However, the whole information search planning and purchasing usually happen on many occasions within a longer period. Planning is fragmented on many episodes each characterized with its own purpose reflecting a specific problem (Pan and Fesenmaier 2006). In this research, I will use real-life data

representing the actual subjects' behavior. In this sense, the data is highly accurate and gives a detailed representation of the internet search behavior.

In conclusion, information search behavior is an important phase of the of the overall tourism behavior and more specifically the travel planning. From tourism stakeholders' perspective, it is a crucial phase where the consumer can be influenced by effective communication strategies and communication systems.

## **2.2 Risk and uncertainty**

Risk attitudes are a central part of the economic theory. Classical economic theory of decision under risk states that the risk is related to the probability of the occurrence of the specific outcome. For example, according to expected utility a prospect with probability  $P$  to win  $x$  amount of money opposed to  $1-P$  to win nothing, will be evaluated as follows:  $p.u(x)+(1-p)u(0)$  where  $u$  is the utility function of money. Risk attitudes are defined as follows, risk aversion is an attitude which is manifested by the preference of the sure outcome over a prospect with higher expected value that involves risk. Whereas, risk seeking attitude will occur when the prospect is preferred over the sure amount.

Later economic theory evolves by distinguishing individual level probability weighing and utility weighting by taking into account different psychological variables. (Woodside and MacDonald 1994) propose "Prospect theory" to explain choices among risky prospects that are inconsistent with the standard economic theory. In these recent developments, the risk attitudes evolve. In order to explain decision making under risk scholars explore choices involving a different amount of risk (high risk and low risk) and associated with outcome involving different monetary values (again high and low) as well gains or loses.

In a paper focused on risk measurement in consumer research Mandrik and Bao (2005) summarize that the measurement of risk attitudes typically has been assessed in three ways. The first method involved "choice dilemmas" where subjects are presented with several scenarios and asked for their preference between two courses of action, this results in computing an overall score which is used to determine respondents' risk attitude. The second method involves gambles. Subjects are asked to choose an amount in order to participate in a gamble. Finally, researchers use self-reported

measures. These include the creation of different scales that are measure risk and uncertainty in specific decision situations. The authors validate a novel self-reported scale which measures general risk attitude as a valid psychometrical measure. The construct proposed by Mandrik and Bao (2005) has been utilized in this research as it provides shorter and simple manner of assessing risk attitudes.

### **2.3 Risk and uncertainty in tourism**

As noted above, one of the reasons researchers claims to cause the extensive information search is risk and uncertainty minimization. Stewart and Vogt (1999) attribute uncertainty as an implicit and universal characteristic of every planning process. Furthermore, the authors argue that in order to handle uncertainty the travelers prepare more than one plan for their trips. Sweeney, Soutar, and Johnson (1999) point out that consumers who are more sensitive to risk and uncertainty engage in more extensive search in order to avoid them. Sirakaya and Woodside (2005) claim that because of the intangible nature of tourism products the uncertainty in tourism is higher than comparing to other products or services.

An important remark related to risk and uncertainty is the difference between both constructs. The difference between them lays in the probabilities of their outcomes, while risk is associated outcomes with known probabilities, uncertainty is associated with outcomes with unknown probabilities. Quintal, Lee, and Soutar (2010) investigate the difference between risk and uncertainty on a country level using Hofstede (1980) uncertainty avoidance index (UAI) and risk scale measurement on tourists' information search. The researchers claim that many other papers do not make the distinction and this is especially problematic when researchers are using country UAI scores to explain individual-level behavior as individuals differ with regards to their attitudes of risk and uncertainty.

Quintal, Lee, and Soutar (2010) explain the relationship between uncertainty and risk in tourism and information search to be translated in the following way. In the early stages of their research, consumers search for information extensively and the outcomes are associated with uncertainty because the rate of occurrences of certain threatening events or undesirable outcomes is not known. In a later stage of the decision-making process, when travelers have already selected possible alternatives, the risk attitude is

more likely to have an influence as consumers can assign relative probabilities to a few selected alternatives i.e. alternatives are being compared to one another providing a reference point.

Based on the relationship explained above, Quintal, Lee, and Soutar (2010) provide a hypothesis that the uncertainty attitude has an influence on the extent of the information search while holding risk attitude constant. With regards to the risk attitude, Quintal, Lee, and Soutar (2010) provide a hypothesis that risk attitude doesn't influence the extent of the information search, holding uncertainty attitude constant. This hypothesis, however, has been defended under the notion that all the information search has been performed during the early stage of the trip planning and in the later stage consumers only compare alternatives based on the information they gathered in the early stage. However, in a conceptual model of trip planning presented by Jun, Vogt, and MacKay (2007), the information search is a component of all phases of the trip planning, including pre-trip, during trip and post-trip. Therefore, I find it credible to hypothesize that both risk and uncertainty attitudes influence the extent of the information search.

I further extend my hypotheses, regarding the direction of the effect. My hypotheses are based on Urbany, Dickson, and Wilkie (1989) and Gemünden (1985) claiming that information search is an instrument consumers use to minimize respectively the uncertainties and risks anticipated with regards to future purchases. For example, Gemünden (1985) finds that when the complexity of the decision making is increasing, so does the information search. They attribute this to higher risks involved in that decisions. As travel planning is regarded as a complex task, I can expect the same relationship. Urbany, Dickson, and Wilkie (1989) outlines two types of uncertainty, choice and knowledge uncertainty, choice uncertainty is related to uncertainties choosing a particular alternative amongst many, while knowledge uncertainty is related to familiarity with the product features. Their findings suggest that choice uncertainty has a positive relationship with the extent of the information search.

Based on the literature review above about information search, travel planning and risk and uncertainty hereby I form the following hypotheses:

- H1. Risk seeking attitude decreases the extent of information search, keeping other factors constant.
- H2. Uncertainty seeking attitude decreases the extent of information search,

keeping other factors constant.

## 2.4 Big Five Factors Inventory (BFI)

According to Leung and Law (2010), the usage of personality traits in travel related literature appears to be low even though the academics agreed upon their value in Marketing domain (Baumgartner 2002). Personality is a temperament or person's inherent qualities of mind and strategies according to which one behaves, dispositions and behavioral patterns that are stable across time and can be used to characterize one's behavior. The trait perspective has been frequently utilized in consumer research because their ease of application as a self-reported measure and the measurement outputs can be easily applied in the statistical analysis. (Jani, Jang, and Hwang 2014).

Big five inventory (BFI) is the most well-known and established factor structure to measure personality (Denissen et al. 2008). It has been validated on many occasions and widely utilized by researchers, creating prerequisites to compare findings across different studies. Another important factor for the popularity of BFI is the fact they are freely available to use along with their validated translations in many languages. BFI has been proposed as a fundamental lexical hypothesis by Galton in 1884 (Goldberg 1993), which is a language taxonomy of human temperament based on adjectives describing different personality traits. The theory was put into practice by Goldberg (1993) and it has been greatly developed ever since, leading to the construction of five broad factors. BFI are based on factor analysis where a large group of traits is shown to be correlated and grouped into five universal traits.

BFI consists of Extraversion, Neuroticism, Conscientiousness, Agreeableness and Open to experience. Openness to experience is related to the degree of curiosity, inventiveness, adoption of novelty on the one hand and consistency and cautiousness on the other. That is, persons with high openness tend to be open-minded, adventurous while low openness can describe individuals that are more pragmatic. Conscientiousness reflects on the tendency for one to be organized, non-spontaneous, organized and efficient. Extroversion is related to traits such as outgoing personality, sociability, talkativeness. Personalities exhibiting low extroversion, on the other hand, can be perceived as less open and reserved. Agreeableness is described as a person being more compassionate and cooperative. It measures whether a person can be

trusted or not and if they are well-tempered. High agreeableness personalities are seen as more naïve, while low are seen as more dominative and competitive. Finally, neuroticism explores the emotional stability of individuals. That is a high need for stability results in individuals who are clam and stable, while a low need for stability can describe emotionally unstable individuals.

These five factors are shown to be the main factors that drive human behavior, appear in different cultures, are relatively stable across the lifetime of subjects and have strong predictive validity (Jani, Jang, and Hwang 2014). Researchers using BFI often aim to measure big five personality dimensions using as less as possible items which lead to the development and validation of short scales of BFI [(Denissen et al. 2008).

## **2.5 Personality traits and tourism information search**

In tourism domain, BFI has been examined with regards to general travel behavior, the undertaken activities in the travel destination, adventure tourism and pilgrim tourism (Jani, Jang, and Hwang 2014). There are two articles researching internet search behavior from the perspective of BFI. Jani (2011) proposes a model relating information needs and tourist information behavior from the perspective of BFI and travel personality. It defines tourists' information needs as reasons for collecting information. Information search behavior is defined as breadth and depth of information sources that consumers use to obtain information. Furthermore, Jani, Jang, and Hwang (2014), research the relationship between BFI and internet search behavior. Using self-reported survey data, they observe that travel related information sought online varies with regards to the different personalities. The authors conclude that some of the factors from the BFI can improve the information search behavior predictability.

Jani, Jang, and Hwang (2014) address the research question whether BFI can be used as a predictor of internet search behavior in terms of sources of information and the extent of the information sought. Using self-reported measures for internet information search and channels used, the authors confirm that personality traits can be used as a predictor of information search behavior.

As BFI in tourism has not been researched thoroughly, in this research I will commit to hypotheses concerning the relationship between personality traits and information



search behavior. However, the BFI will be used in the research design and analysis of this paper along with trip characteristics as control variables.

## 2.6 Hypotheses

Based on the literature review, here I define the full list of hypotheses relating each of them to the main variables of interest in regards to travel related activities as listed above. The following relationships between observed behavior and unobservable personality traits are researched:

Hypotheses				
H1	Risk attitude	<i>Decreases</i>	the amount of	<ul style="list-style-type: none"> <li>a. Micromoments</li> <li>b. Domains</li> <li>c. Pageviews</li> <li>d. Time</li> <li>e. Length</li> </ul>
H2	Uncertainty attitude	<i>Decreases</i>	the amount of	<ul style="list-style-type: none"> <li>a. Micromoments</li> <li>b. Domains</li> <li>c. Pageviews</li> <li>d. Time</li> <li>e. Length</li> </ul>

As mentioned above I will run an exploratory analysis of the impact of the big five personality traits on online travel behavior. However, I will not commit to a hypothesis with regards to their effect or direction.

## 3 Methodology

### 3.1 Introduction

The data utilized for this research has been collected from panelists who participate in a large consumer panel in the Netherlands. The data has two main components, behavioral data, and survey data. The methodological section of this paper explains the sampling, data collection and analysis procedures that have been conducted, in the following way.

- Panel and sampling. Sampling based on behavioral data. A sample of respondents has been pulled out of the behavioral dataset based on their online behavior on popular tourism related websites in the Netherlands.
- Survey. An online survey has been administered among the sampled panelists to reveal the “unobservables” from the perspective of the available behavioral data. Namely, these include data points related to trip characteristics associated with respondents’ latest tourist related purchases as well as the question regarding their personality. The code related to this step is available in *Appendix #2 Code. CHUNK #5, Appendix #2 Code. CHUNK #6*
- Data processing. The full behavioral data of the respondents who successfully completed the survey has been sampled out of the full behavioral panel dataset. The code related to this step is available in *Appendix #2 Code. CHUNK #1, Appendix #2 Code. CHUNK #4*
- Categorization. All of the unique websites of the sampled behavioral dataset have been classified into categories. In such a way it was possible to assess whether a certain website was travel related or not. The code related to this step is available in *Appendix #2 Code. CHUNK #2A and Appendix #2 Code. CHUNK #2B, Appendix #2 Code. CHUNK #3A, Appendix #2 Code. CHUNK #3B*
- Analysis. The aggregated information from the behavioral data has been regressed over the survey data to reveal the impact the unobservable personality attitudes and traits on the tourism related online behavior while controlling for the trip characteristics. The code related to this step is available in *Appendix #2 Code. CHUNK #8*

In the following sections, I will first focus on explaining the behavioral data and the

technology behind it. Afterward, the 5 major steps of the methodology are explained in details. Then I proceed with the descriptive results of each of the datasets (survey and behavioral). Finally, I report the analysis results.

## **3.2 Behavioral data & Technology**

The behavioral data, also referred to as observational data, reflects the online behavior of the consumers. It consists of records of the interaction made via consumers' digital devices and the Internet.

The behavioral data has been collected via a technology developed and provided for this research by Wakoopa. The company is a provider of a tracking technology. The technology is utilized primarily for market research purposes. Similarly to the market research consumer panels, where panelists enroll to participate in online surveys for incentives, Wakoopa provides its technology to market research consumer panel companies that are interested in tracking the online behavior of their panelists. After enrolling into the panel and giving their consent to be tracked, panelists install an application on their devices i.e. desktop, mobile, and tablet. The tracking software collects every interaction of the panelists' devices on the Internet which consists of a path or the address the panelists are reaching and the duration of the visit. The software works in a different manner over the different operation systems platforms and devices, but the final result consists of recording raw data containing events. Each event has the address the participant accessed, the duration of the interaction and the client requesting the information i.e. browser, app etc.

## **3.3 Steps**

### **3.3.1 Panel and sampling**

Using Wakoopa's panel it was possible to 'pre-screen' relevant respondents for the purposes of the analysis, that could be invited to participate in the survey I have conducted. The respondents that I was looking for, should have been active on tourism related websites and should have conducted a purchase on such websites in the period of January 2015 to June 2016. Herein, I describe the process that was used to reach

such subjects. Upon starting of the project the panel used for the research had 6682 active panelists and 7103 active devices. The majority of the panelists were being active only on the desktop. See *Appendix #1. Panelists* and *Appendix #2. Devices* for reference.

First, I started looking into tourism related websites in the Netherlands. Initially, I used 300 domains to account for the majority of the tourism related internet traffic in the Netherlands according to the internet analytics company SimilarWeb. Afterward, based on this data I exported the activity of the whole panel over those websites ranging from 01/2015 to 06/2016. The data of this activity was manually analyzed looking for the end pages of the payments, also referred to as “confirmation” pages or “thank you” pages. Confirmation pages are the pages where a customer has been redirected after conducting a purchase at a company website. In general, the analysis consists of looking up keywords within the travel domains URLs and marking down the common patterns with the aid of regular expressions (See Mitkov (2005) for discussion). Identified were the patterns of “confirmation” pages and data for the participants visited from the period of 01/03/2016 to 16/06/2016 was exported. Using this information, it was possible to identify a sample composed of 949 respondents which were to be invited into the online survey as described in the second step below.

Given the estimation of the incidence rate provided by the panel supplier, 20%, and the initially desired sample of 500 respondents more respondents were needed. Therefore, an additional random sample of 123 respondents was selected based on whether they were active on the 300 initial domains but without evidence for their purchases from the data.

### **3.3.2 Survey**

The selected 1039 panelists were invited to participate in an online survey, that aimed to reveal their attitudinal and personal characteristics as well as the trip characteristics of their last travel. The fieldwork was conducted during the last week of June 2016. Out of the 1039 invitations sent 872 started the survey, which resulted in 495 completed interviews. (Note to self, move this to a separate cleaning section). The data was further cleaned by accounting for speeders and flatliners. Speeders were panelists that have finished the survey within less than half of the average length of the

interview. Flatliners refer to respondents who have answered all of the grid questions in a straight line. Moreover, a few respondents attempted to participate in the survey multiple times and thus, they were excluded from further analysis. Thus, the final dataset consists of 426 observations.

The survey consisted of several parts. First, a screening criterion was used. It accounted for the number of travel-related purchases since March 2016. Respondents with no travel related purchases were not allowed to further proceed. Then, subjects who have not purchased neither flight nor accommodation were also excluded from the questionnaire. Business only travelers were not relevant for the analysis, thus they were also screened out, leaving only panelists who have gone on a leisure trip or both on a leisure and business trip. The second part was a demographics section, which consisted of question regarding age, gender, and income. The next part of the questionnaire was in regards to the trip characteristics, which was adapted from Roehl and Fesenmaier (1992) and contained information about moment of booking, destination, planning horizon, information sources used i.e. Internet, advice from friend and relatives, tourist information office, travel agent etc., products purchased online, duration of the trip or number of nights spent away from home, whether the destination was visited before and how many times, number of travel companions, indication whether there was children on the trip and whether subjects visited friends or relatives during their trip.

The next two sections aimed to reveal more about the subjects personality and risk and uncertainty attitude. Items assessing risk and uncertainty originate from Quintal, Lee, and Soutar (2010). Respondents reported their risk attitude on three item scale which the authors adapt from Donthu and Gilliland (1996). Uncertainty attitude has been assessed on four items scale which authors adapt from Yoo and Donthu (2002) and the scale is based on Hofstede (1980) UA items. I have chosen these scales as they are the shortest reliable scales for self-assessing the risk and uncertainty attitudes. Using such scales I can make sure to not tire respondents and keep them engaged. Personality traits related to openness, consciousness, extraversion, agreeableness and neuroticism have been assessed using the short scale from the big five inventory proposed by S. D. Gosling, Rentfrow, and Swann (2003).

### 3.3.3 Data processing

After having collected and cleaned all the data, I further proceed with data processing and analysis. All of the data processing and the analyzation tasks of this paper were done in R Studio. All of the code, including libraries can be seen in the appendix.

The data processing includes merging down the two main behavioral streams of data that originate from desktop and mobile. Additional to that it is needed to classify and derive all the tourism related domains that will be further used for analysis. This is an important step as it will allow to analysis what the relationship between the online travel behavior and the traveler's attitudinal characteristics are. Categorization of the domains was done using machine learning classification algorithms provided by uClassify.com. The domains from the desktop and mobile behavioral data and all of the mobile apps are further classified using keywords.

Next step is, converting the raw data set onto micromoments. They are also further used in the analysis as I examine the relationship between one's micromoment frequency and subject's risk and uncertainty attitude. Micromoments are essentially user sessions where users were active at a given moment of time. For example, if a panelist access a certain domain and spend five minutes on it and then become inactive for more than 3 minutes, this will result in a factor variable grouping all of the observation within those five minutes of activity. If a travel related website has been visited during the micromoment, a dummy variable is assigned to this moment to indicate this. Furthermore, purchases of travel related products were included in the data based on the initial dataset used for pooling the sample out the panel for the survey. The final dataset includes aggregated information about panelists' activity over two main levels. 1. top-level i.e. total activity 2. low-level micromoment level.

This section includes the data processing tasks done on the behavioral data including descriptive statistics and variable derivation, classification and aggregation on the different levels intended for further analysis.

Starting point of the analysis of the behavioral data include procession the data and rendering it in a format suitable for running the analysis. In its initial form the data has been exported in a format containing the following variables: for the desktop data (sample data, see: *Appendix #3 Desktop data*) and for the mobile data (sample data, see: *Appendix #4 Mobile data*). The next step is adding "host" variable to both

datasets. The host variable contains domains and the subdomains that are going to be used for the categorization whether the domains were travel related or not. The next query on the data included identifying all of the micromoments in the data. First, the data has been subsetting on “*panelist\_id*” level, sorted by the the timestamp “*used\_at*” and assigned into a list where each element is the full data for each individual respondent. The sum of “*used\_at*” variables and “*active\_seconds*”, the duration in seconds respondent spend on the page, were compared to “*used\_at*” variable of the next observation. If differences larger than five minutes were found all of the variables prior to this difference were grouped together under a common factor variable.

Table 3.1: Desktop dataset

<i>used_at</i>	host	<i>panelist_id</i>	url	<i>active_seconds</i>	<i>browser_name</i>	<i>mmid</i>	<i>Class_Travel</i>	<i>purchase</i>
----------------	------	--------------------	-----	-----------------------	---------------------	-------------	---------------------	-----------------

Table 3.2: Mobile dataset

<i>panelist_id</i>	<i>device_id</i>	scheme	url	domain	<i>app_id</i>	<i>app_name</i>	<i>used_at</i>	connection	duration	host
--------------------	------------------	--------	-----	--------	---------------	-----------------	----------------	------------	----------	------

### 3.3.4 Categorization

The categorization procedure includes using uClassify machine learning algorithms and also using keywords. The full activity coming from desktop and mobile devices resulted in 194,534 unique domains. The categorization algorithm has been responsible for classifying all domains that have more than ten visits or 47,818 domains in total works as follows. A web-scraper designed for these projects accesses a collection of domains and collects all of the information on the page, then removes the HTML elements along with the punctuation and renders down the information only to a part that is visible to the website visitors. Then, it passes this information to an application programmable interface (API) that returns the probability of this text being into sixteen different categories including travel.

The full list of categories includes:

NULL

Websites with content with the highest probability to be travel were assigned a value of a dummy variable 1 or 0 otherwise.

Due to technical and time constraints the API service wasn't able to classify all of the domains. Therefore, the full dataset of unique domain names has also been scanned for keywords. Keywords include travel, tourism, accommodation, hotels, flight etc. (full list of the keywords can be found in the code in the appendix). If any of the keywords appear within the domain name, the respective domain has also been assigned to the list of travel domains. The same keyword approach has been used for classification of the application on mobile devices, where the apps have been classified using their names.

### **3.3.5 Analysis**

This paper utilizes regression analysis, factor analysis, and stepwise regression. The functional form has been selected based on regression analysis. The underlying factors regard to risk and uncertainty attitudes and BGI have been analyzed by exploratory factor analysis. Finally, the main model has been compiled using stepwise regression. The analysis techniques utilized by this research please refer to the appendix. The "Analysis" chapter contains brief elaboration on regression model "OLS", factor analysis and all of the tests concerning the underlying assumptions of the "OLS" assuring the best model performance.

### **3.3.6 Model selection**

#### **3.3.6.1 Restricted Model**

There are two distinct sources of data resulting in three data sets to be investigated.

1. Dataset from desktop
2. Dataset from mobile
3. Combined dataset from desktop and mobile

Furthermore, there are five dependent variables that are intended to be examined during this research. Namely,

1. Number of travel micro-moments
2. Number of unique travel domains visited
3. Number of travel pageviews



4. Total time in seconds spent on travel domains
5. Total length in seconds of travel micro-moments

The difference between the last two (4 and 5) is as follows. Total time in seconds (4.) has been measured once a panelist arrives at a certain website that has been classified as travel, whereas length (5.) is the total length (i.e. the difference between the start and end) of micro-moments amongst which a participant visited a travel related domain.

The restricted model for each of the three datasets takes the following form:

Table 3.3: Restricted model(s)

Model	Dependent variables	Independent variables	Control *	Control2
# 1	Number of travel micro-moments	Risk + Uncertainty +	Total micro-moments	Total purchases + Days active in the panel
# 2	Number of unique travel domains visited		Total domains	
# 3	Number of travel pageviews		Total pageviews	
# 4	Total time in seconds spent on travel domains		Total time	
# 5	Total length in seconds of travel micro-moments		Total micro-moments length	

\* As there was only one purchase detected on mobile devices this term wasn't present in the mobile model

Running Ramsey-Reset test per each of the basic models showed evidence that the functional form of the models is not well specified, thus the models were rejected.

Consequently, all of the dependent variables along with their corresponding controls were transformed into logs accounting for additional five models per each dataset. The results from the consequent running of the Ramsey-Reset test were still not satisfying, as the test showed significant results indicating that the functional form of the model is not well specified. Consequently, the corresponding control variables, accounting for total activity and purchases, were transformed into binary variables with values 1-indicating a participant belongs to a highly active groups of participants (having values above the mean of the sample) and 0 vice versa. The latter transformation accounted for investigating ten more models per each dataset reaching the total number of thirty models to be reviewed. The functional form of the model has been selected based on the dataset containing desktop data as 1) this dataset is the most complete in terms of a number of observation and 2) the participants were selected based on their desktop behavior.

Among all tested models, based on the results from the Ramsey-Reset test, I select the following functional form of the model for Desktop, Mobile, and Combined datasets. Shapiro-Wilk test for normality also performs best for this functional form, yet the sample is large enough so we can relax the normality assumption.

Table 3.4: Restricted model(s) functional form

Model	Dependent variables	Independent variables	Control *	Control2 **
# 1	log(#) of travel micro-moments	Risk + Uncertainty + Interaction	(D) Total micro- moments	(D) Total purchases + Days active in the panel
# 2	log(#) of unique travel domains visited		(D) Total domains	
# 3	log(#) of travel pageviews		(D) Total pageviews	
# 4	log(#) time in seconds spent on travel domains		(D) Total time	

Model	Dependent variables	Independent	
		variables	Control *      Control2 **
# 5	log(#) length in seconds of travel micro-moments		(D) Total micro- moments length

\* Note: as there was only one purchase detected on mobile devices this term wasn't present in the mobile model.

\*\* Dummy variables take value 1 if the respondents' activity is above the average activity of the sample and 0 otherwise.

A table with the results from the restricted model along with the model performance metrics can be found in the appendix:

*Table 4.1: Desktop data, restricted model*    *Table 4.2: Mobile data, restricted model*

*Table 4.3: Combined data, restricted model*    *Table 4.4: Restricted models tests*

Due to poor model performance is shown in *Table 4.4: Restricted models tests* row **11 to 15**, the model shown in *Table 4.3: Combined data, restricted model* has been rejected. See Appendix *Table 4.4: Restricted models tests* for reference.

### 3.3.6.2 Final Model

Based on the selected functional form of the model(s) I proceed with stepwise selection in order to select the final model across desktop and mobile. First, the models are presented then results of the OLS performance tests are shown in the appendix under *Table 7: Full model tests, Desktop* and *Table 9: Full model tests, Mobile*. The tests have shown evidence for heteroscedasticity. Therefore, in all of the tables of the results, the standard errors have been corrected using white standard errors. The procedure is explained in the appendix.

## 4 Results

### 4.1 Survey Data

The original data consisted of 495 observations. Before proceeding with analysis, the data was cleaned based on several considerations. Namely, I looked at whether 1) respondents were speeding throughout the survey, 2) subjects were flatliners, e.g. if they have answered the grid question with the same answer, and 3) if they had low activity less than 1000 observation in the behavioral data. 4) if they attempted to participate the survey more than once. The final dataset thus consists of 426 observations.

Looking at the demographic characteristics, we see that there is an equal split between males and females. Seeing the distribution among age, the population above 50 years old is overrepresented, accounting for 59% of the sample. In terms of income, the distribution is normal, most of the subject indicated yearly income between 33.000 €“ 66.000 euro.

All subjects were actively involved in the trip planning as the main decision maker or contributed to the decision, subject who didn't participate in the decision making were screened out. The business trip only travelers were screened out as well, those left in the sample are either leisure or leisure and travelers. Finally, we see that the majority of respondents indicated they have booked both flight and accommodation.

The next part of the survey explored the trip characteristics with regards to their general planning, online behavior, and destination. These data points were used as control variables for the analysis. On average it took respondents 9 weeks to plan the whole trip, for about 40% it took between 1 and 4 weeks to plan everything, while 15% indicated they needed more than 20 weeks of planning.

The majority (80%) visited countries within Europe, and 57% have not visited the country before. On average, people who have visited the country before went there for the 4th time. Typically together were traveling 2 people, the largest travel group consisted of 14 people and they have spent approximately 9 nights away from home. About 34% of the respondents indicated they have spent more than 10 days. Only 20% of the subjects have stayed with relatives.

The majority of subjects researched for the trip using the Internet, but many have also

asked for advice from friends or relatives or approached a travel agent. Moreover, the internet is widely used for booking accommodation and not so much for traveling or booking “entertainment” online.

#### **4.1.1 Risk and uncertainty**

For understanding risk and uncertainty attitudes, the seven-item scale was analyzed using exploratory factor analysis. According to the analysis results, I cannot accept the hypothesis that two factors are sufficient to explain the seven-item scale. However, exploring further, the results show that the factor loadings of the different variables have the highest loadings on the assumed two underlying factors i.e. risk and uncertainty. Furthermore, running factor analysis when assuming 3 factors results in separating item #4 as a distinct factor on its own with relatively low factor loading. According to Yong and Pearce (2013) choosing too few factors may lead to leaving out important variance out of the analysis. However, I accept the high factor loadings onto the two assumed factors as a sufficient reason to continue the analysis with two factors. Namely, items 1 to 3 accounting for underlying risk attitude and items 4 to 7 accounting for attitude towards uncertainty. Factor analysis results on risk and uncertainty items are available in the *Appendix 1: Factor analysis results*

#### **4.1.2 BIG5**

Running exploratory factor analysis on BIG5 items resulted in confirming the initially assumed personality traits. Furthermore, based on the results we can accept the hypothesis that 5 factors are sufficient to explain the variance of the 11 item scale. Factor analysis results on BIG5 items are available in the *Appendix 1: BIG5*

Next, with the aid of factor analysis respondents were grouped into 2x2 groups based on the mean score of the items per each factor: 1) risk “averse” and “seeking” and 2) uncertainty “averse” and “seeking”. With regards to “big five”, the respondents have also being allocated into two groups based on the mean score of items accounting per each factor i.e. “high” and “low”. The procedure has been adopted based on the approach from Jani, Jang, and Hwang (2014).

It is interesting to observe some of the descriptive statistics which can already provide

some insights on what the relationships between people’s risk and uncertainty attitude and their travel behavior based on the survey data. First of all, the majority of subjects are risk averse and uncertainty averse. The subject being risk seeking account for 30% of the population, while the uncertainty seeking are 40%.

Table 4.1: Risk attitude

averse	seeking
324	170

Table 4.2: Uncertainty attitude

averse	seeking
299	195

Interestingly enough, given the data, it is observed that risk averse and uncertainty avoidant subjects actually prefer to go to a destination they haven’t visited before, while risk and uncertainty seeking respondents have a slightly higher preference to go to a known destination. Further looking at the data, I find that on average risk averse people tend to spend more time planning the trip than risk seeking people (10 days vs 7 days). This effect is significant at 5% significance level. However, uncertainty averse and seeking people exhibit similar behavior (9 versus 10 days), the mean difference is not significant at 5% significance level, and thus there is no difference in the planning horizon for uncertainty averse and seeking travelers.

## 4.2 Main results

### 4.2.1 Desktop Data

Looking at a table *Full model: Desktop, robust*, I can say that most of the hypotheses of this paper have been confirmed. The table with the main results is available below under *Table 5.3: Full model: Desktop, robust* while the table with the full results is available in the Appendix #1 *Table 5: Full model: Desktop, robust*

First, looking at the effect of risk seeking attitude towards travel behavior, I see that risk seeking attitude decreases individuals’ activities related to travel behavior such as

travel domains ( $\log(\text{TD})$ ), travel pageviews ( $\log(\text{TDPV})$ ), travel time ( $\log(\text{TT})$ ) and total travel-related micromoment length ( $\log(\text{TL})$ ). For example, risk seeking individuals tend to use 23% fewer travel domains on average compared to risk averse individuals, *ceteris paribus*. Moreover, risk seeking individuals spend 46% less time and have 37% fewer pageviews on travel related domains on average compared to risk averse individuals, *ceteris paribus*. Finally, when it comes to the two variables accounting for micromoments, we see that risk seeking attitude doesn't have a significant effect over the total number of travel related micromoments, while when it comes to overall micromoment time (containing travel related activity) it does have a significant effect. On average, risk seeking attitude decreases the overall time of micromoments by 26% compared to risk averse attitude, *ceteris paribus*.

Second, looking at the effect of uncertainty seeking attitude towards travel behavior, I see that uncertainty seeking attitude decreases individuals' activities related to travel behavior such as travel domains ( $\log(\text{TD})$ ), travel pageviews ( $\log(\text{TDPV})$ ) and travel time ( $\log(\text{TT})$ ). Uncertainty seeking individuals tend to use 25% fewer travel domains on average compared to uncertainty averse individuals, *ceteris paribus*. Moreover, uncertainty seeking individuals spend 37% less time and have 34% fewer pageviews on travel related domains on average compared to uncertainty averse individuals, *ceteris paribus*. Finally, when it comes to the two variables accounting for micromoments, we see that uncertainty seeking attitude doesn't have a significant effect over both the total number of travel related micromoments and overall micromoment time (containing travel related activity).

Third, the interaction term between risk seeking and uncertainty seeking attitudes has a significant effect only on time spent on travel-related websites. Interestingly, being both risk seeking and uncertainty seeking increases the time spent on travel related domains by 76%, *ceteris paribus*.

Exploring the effect of personality traits on online travel behavior the following effects are revealed. Scoring low on personality trait "Openness to experience" decreases the number of pageviews on travel related websites by 32% and the total time spent on travel-related websites by 33%, *ceteris paribus*. Scoring low on personality trait "Consciousness" decreases the total time spent on travel-related websites by 20%, *ceteris paribus*.

Table 4.3: Full model: Desktop, robust

	<i>Dependent variable:</i>				
	D log(MM) (1)	D log(TD) (2)	D log(TDPV) (3)	D log(TT) (4)	D log(TL) (5)
Risk Seeking	-0.11 p = 0.36	-0.23* p = 0.08	-0.37*** p = 0.005	-0.46*** p = 0.0002	-0.26** p = 0.03
Uncertainty Seeking	-0.13 p = 0.26	-0.25** p = 0.04	-0.34*** p = 0.004	-0.37*** p = 0.002	-0.17 p = 0.12
Risk(x)Uncertainty: Seeking	0.09 p = 0.73	0.10 p = 0.69	0.39 p = 0.16	0.76** p = 0.02	0.30 p = 0.22
Days active	0.01*** p = 0.001	0.01*** p = 0.002	0.01** p = 0.02	0.01** p = 0.04	0.005*** p = 0.01
D Purchase	0.53*** p = 0.0000	0.63*** p = 0.0000	0.84*** p = 0.00	0.88*** p = 0.00	0.50*** p = 0.0000
D Micromoments	1.13*** p = 0.00				
D Domains		1.03*** p = 0.00			
D PageViews			0.83*** p = 0.00		
D Time				0.80*** p = 0.00	
D MM Length					1.36*** p = 0.00

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## 4.2.2 Mobile data

Looking at the table *Full model: Mobile, robust*, I see a big difference in respondents behavior compared to the desktop behavior. Most of the hypotheses with regards to the mobile behavior of this paper have not been confirmed. The table with the main results is available below under *Table 5.4: Full model: Mobile, robust* while the table with the full results is available in the Appendix #1 *Table 6: Full model: Mobile, robust*

First, looking at the effect of risk seeking attitude I do not observe any significant relationship with the dependent variables. Second, looking at the effect of uncertainty seeking attitude towards travel behavior reveals some surprising results. Uncertainty seeking attitude increases individuals' activities related to travel behavior such as travel micromoments (log(MM)), travel domains (log(TD)), travel time (log((TT)) and overall micromoment time (log(TL)).

Third, on the other hand, the interaction term between risk seeking and uncertainty seeking attitudes decreases greatly individuals' activities related to travel behavior such as travel micromoments (log(MM)), travel domains (log(TD)), travel time (log((TT))



and overall micromoment time ( $\log(TL)$ ).

Forth, scoring low on personality trait “Consciousness” increases the number of micromoments by 60% and travel domains by 53%%, *ceteris paribus*. Scoring low on “Extraversion” increases the pageviews on travel related domains by 88%, *ceteris paribus*. Scoring low on “Agreeableness” significantly decreases travel micromoments ( $\log(MM)$ ), travel domains ( $\log(TD)$ ), travel pageviews ( $\log(TP)$ ) and travel time ( $\log(TT)$ ). Scoring low on “Neuroticism” increases travel domains ( $\log(TD)$ ), travel pageviews ( $\log(TP)$ ), travel time ( $\log(TT)$ ) and overall micromoment time ( $\log(TL)$ ).

Table 4.4: Full model: Mobile, robust

	<i>Dependent variable:</i>				
	D log(MM) (1)	D log(TD) (2)	D log(TDPV) (3)	D log(TT) (4)	D log(TL) (5)
Risk Seeking	0.16 p = 0.66	0.13 p = 0.61	0.53 p = 0.22	1.32 p = 0.17	1.09 p = 0.21
Uncertainty Seeking	0.72** p = 0.04	0.73** p = 0.02	0.57 p = 0.11	1.98* p = 0.06	2.05** p = 0.02
Risk(x)Uncertainty: Seeking	-1.76*** p = 0.003	-1.80*** p = 0.0000	-0.75 p = 0.34	-7.89*** p = 0.00	-7.37*** p = 0.0000
Days active	0.01*** p = 0.001	0.01*** p = 0.0001	0.004 p = 0.12	0.02*** p = 0.0000	0.02*** p = 0.0001
D Micromoments	1.03*** p = 0.004				
D Domains		1.77*** p = 0.0000			
D PageViews			2.44*** p = 0.0000		
D Time				2.06*** p = 0.01	
D MM Length					2.20*** p = 0.003

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

### 4.3 Conclusion

Based on the results I conclude that risk and uncertainty attitudes have an impact of online travel related behavior. I can confirm that the following hypothesized relationships exist.

Based on the analysis the following hypotheses are confirmed/rejected with regards to consumers' **Desktop** activity:

Hypotheses				
H1	Risk attitude	<b>Decreases</b> the amount of	a. Micromoments	<i>Rejected</i>
			b. Domains	<b>Confirmed</b>
			c. Pageviews	<b>Confirmed</b>
			d. Time	<b>Confirmed</b>
			e. Length	<b>Confirmed</b>
H2	Uncertainty attitude	<b>Decreases</b> the amount of	a. Micromoments	<i>Rejected</i>
			b. Domains	<b>Confirmed</b>
			c. Pageviews	<b>Confirmed</b>
			d. Time	<b>Confirmed</b>

Hypotheses		
		e. Length <i>Rejected</i>

Based on the analysis the following hypotheses are confirmed/rejected with regards to consumers' **Mobile** activity:

Hypotheses		
H1	Risk attitude	<i>Decreases</i> the amount of
		a. Micromoments <i>Rejected</i>
		b. Domains <i>Rejected</i>
		c. Pageviews <i>Rejected</i>
		d. Time <i>Rejected</i>
		e. Length <i>Rejected</i>
H2	Uncertainty attitude	<i>Decreases</i> the amount of
		a. Micromoments <b>Opposite effect</b>
		b. Domains <b>Opposite effect</b>
		c. Pageviews <i>Rejected</i>
		d. Time <b>Opposite effect</b>
		e. Length <b>Opposite effect</b>

In the next chapter, the results are discussed thoroughly.

## 5 Conclusions

### 5.1 Discussion

The present study aimed at exploring the personality traits such as risk and uncertainty attitudes and their relationship with online travel behavior. The online travel behavior was viewed both from desktop and mobile perspective, in line with recommendations from Chatfield (2014) and Xiang, Magnini, and Fesenmaier (2015). It is also observed that mobile online behavior is one of the major trends happening in the online travel industry, therefore, nowadays it is essential to review the consumer journey on every channel. The results suggest that the same personality traits have a different impact on desktop and mobile online travel behavior.

The finding with regards to desktop online travel behavior is in line with the main hypotheses of the study. It has been found that risk seeking attitude (H1) decreases the aspect of online travel related behavior related to a number of travel related domains, pageviews on travel related websites, time spent on travel-related websites, overall length of travel micromoments, while the effect is not significant with regards to a number of travel micromoments. The uncertainty attitude has been found to influence a number of travel related domains, pageviews on travel related websites, time spent on travel related websites, while it does not have a significant relationship with micromoments and overall length of travel micromoments. These findings are in line with previous finds in the literature that information search is uncertainty reducing instrument (Urbany, Dickson, and Wilkie 1989). The findings confirm the relationship between uncertainty and the extend of information search studied by Quintal, Lee, and Soutar (2010). However, the analysis also reveals that there is a relationship between risk attitude and the extend of information search in travel context, which contradicts to Quintal, Lee, and Soutar (2010). This finding is in line with the literature that information search is risk reducing instrument (Gemünden 1985). The contradiction may come from the fact that Quintal, Lee, and Soutar (2010) uses self-reported data on different information sources used when making travel decisions. In a design, metrics on the extend of information searches such as time or pageviews cannot be accurately accessed.

When it comes to mobile behavior the findings are surprising. The risk attitude hasn't

been found to have a relationship with the online travel related behavior (in line with Quintal, Lee, and Soutar (2010)) while the uncertainty attitude has been found to have an opposite direction of the effect i.e. uncertainty seeking behavior increases the extend of online travel information search. Furthermore, the interaction term between risk and uncertainty seeking attitude has a huge impact (e.g. more than 700% of time). Such a big fluctuations and unusual results can be explained by the fact that the sample size on mobile is really small 109 respondents and there is no indication what kind of mobile device is the source of data e.g. smartphone or tablet which leaves a big part of the variance in the error term. Furthermore, there is a measurement error when measuring time on mobile (explained in the limitations). Nevertheless, this difference is interesting because of the rising role of mobile in the online travel industry and its worth further investigation.

With regards to big five personality traits, only a few effects found in previous research have been recognized. The desktop behavioral, suggest that the effect of scoring low on the factor “openness to experience” decreases the number of pageviews on travel related websites and the total time spent on them. These findings support the notion that “closed-mindedness” as opposite to “openness” is related to more reserved views towards new experiences. Being low in “consciousness” i.e. “carelessness” decreases the total time spent on travel related websites. Again, the intuition suggests that careless individuals would not spend more efforts obtaining information to decrease probabilities of undesired future outcomes compared to conscious ones.

The findings from the mobile behavior are different. Low in personality trait “consciousness” increases the number of micromoments and travel domains. Low in “extraversion” increases the pageviews on travel related domains. Low in “agreeableness” decreases travel micromoments, travel domains, travel pageviews and time. This contradicts the notions that “antagonism” opposed to “agreeableness” is related to more skeptical and analytical minded individuals which would be expected to spend more time obtaining new information. Low in “Neuroticism” increases travel domains, travel pageviews, travel time and overall micromoment time. Again, these findings are on the opposite side of the intuition which suggest that more “emotionally stable” or “confident” individuals opposed to “neurotic” would be less likely to spend time gathering information and researching alternatives online.

The results from the previous literature (Jani, Jang, and Hwang 2014) suggests that

travel information search varies with the different factors except from “extraversion”, and the online sources utilized for travel information search vary with the different factors except for “conscientiousness”. One of the most obvious reason for the difference in the research methodology, as Jani, Jang, and Hwang (2014) utilized survey data for reporting online activities which are shown to be not particularly accurate (Wakoopa 2016). However, as noted above, some of the findings are contra intuitive, therefore further research is needed in order to explore these relationships in details.

## 5.2 Contribution and Implications

Behavioral data gives us an amazing capacity to understand the online behavior of the consumers as never before. Although the level of understanding of the consumers reached by the academia and practitioners is still great before the age of behavioral data, the capability of zooming into details of the real consumer behavior and the robustness of the insights are truly fascinating. Practically behavioral data empowers researchers to answer every question concerning the observable online behavior that comes to mind in a way beyond the capacity even of the consumer himself. Due to the complexity and fragmentation, the consumer is unable to recall the information in such rich details as when it is recorded.

Due to this fact, the relationship under investigation has not been investigated before in such a detailed manner. This research confirms the finding from previous research in the area that the attitude towards uncertainty, but not the risk is responsible for choosing a number of sources used for travel decision making. Our findings contribute to previous research through passive metering by investigating the time people spend on travel websites, as well as the efforts in terms of pageviews. The results from the analysis strongly confirm the hypotheses that both risk and uncertainty attitudes account for changes in time and pageviews spend desktop devices.

This study contributes to 4 areas, it has been used as promotional material, methodologically to the area of behavioral research, to the stream of the literature of decision making under risk and uncertainty, to tourism research and makes managerial implications for better online travel related marketing activities.

The findings serve as a tangible example of behavioral data employed in practice. The work has been used as promotional marketing material to promote the business of

Wakoopa. The company offers technology for behavioral data collection. This work has been released as part of blogpost series: “How behavioral data sheds light on real consumers’ travel planning behavior”:

- **Part #1**
- **Part #2**
- **Part #3**

The main contribution of the research is adding methodologically to the emerging stream of research in the behavioral observational research. The research explores unobservable attitudes of one’s personality assessed via questionnaires in relation to a real worlds observational data. Even though it uses an established and well-known analysis techniques, the novelty in data collection techniques gives significant methodological contribution for practitioners and the academia.

The research contributes to the area of behavioral economics more precisely in decision under risk and uncertainty. It expands the area in relation to of online travel related behavior. It confirms the notion ones that attitude toward risk and uncertainty play a significant role in decision making concerning travel related information search and that individuals with plausible attitude toward both constructs spend less time in planning whereas those more concerned with the risk and uncertainty, in general, are more careful planners.

### **5.3 Limitations & Recommendations**

1. Risk and uncertainty measurement. The risk and uncertainty attitudes have been measured using the shorted possible way for measurement assessing only the general attitude of these personality traits. Arguably there could be discrepancies between the general attitude and domain specific attitude. Therefore, the experiment can be reproduced using a more sophisticated measurement of risk and uncertainty attitudes such as gambles and lotteries as well as assessing perceived risks in travel related domains.
2. Classification. The classification used in the research has been used as it is provided, however, it is possible to employ better classification algorithms providing better fragmentation of the travel related behavior. Therefore, for

future research, I suggest the use of better classification. Micro-moment, the length of a micro-moment has been set to be 5 minutes without any reference or estimation whatsoever. There is a research that suggests how the best way cut-off point in order to measure user sessions can be estimated. Furthermore, we use the same cut-off point of 5 minutes across desktop and mobile device. As the activity of both type of devices largely differs it could be necessary to use different cut-off points per desktop and mobile. This would also mean that the comparison between desktop and mobile is harder as it is generally different.

3. Multiple users per device. It is known that some desktop devices are shared in the households, therefore we are recording data streams from different users and assessing these data as coming from a single user. Thus, the analysis of this data can be biased as we 1) cannot distinguish the different users and 2) cannot connect properly the survey data with the behavioral data.
4. Time is inferred. The measurement cut-off point is 2 seconds anything shorter than two seconds is not recorded. Crosstab browsing on desktop devices cannot be accessed. There is a measurement error on mobile devices which makes it possible to have events with time zero.
5. Selection bias. One should not overlook the possibility of selection bias as the people who agree to be tracked can be arguably different than the ones who are more privacy conscious and would not agree to share such information even for research purposes. Selection bias states that the selected sample who agrees to participate in the research may yield different characteristics than the sample of interest drawn from the general population.
6. Programming. The program written to process and analyze the data contains several thousand lines of code. Developing such a code in an absence of a quality assurance team and/or a proper quality assurance procedures is prompt to errors.
7. Survey measurement. There are two issues related to the survey methodology. First, the income scale was related to individuals' income, while the questions have been asked with regards to household income. This prevented me from using this variable in the research design. Second, the fourth item from the BFI has been wrongly included as it contains two items instead of one. Yet, they are both from the same traits, thus the impact of this error is potentially small.



## References

- Ammu, Nrusimham, and Mohd Irfanuddin. 2013. "Big Data Challenges." *International Journal of Advanced Trends in Computer Science and Engineering* 2 (1). Citeseer: 613–15.
- Baumgartner, Hans. 2002. "Toward a Personology of the Consumer." *Journal of Consumer Research* 29 (2). The Oxford University Press: 286–92.
- Bonn, M. a., H. L. Furr, and a. M. Susskind. 1998. "Using the Internet as a Pleasure Travel Planning Tool: an Examination of the Sociodemographic and Behavioral Characteristics Among Internet Users and Nonusers." doi:10.1177/109634809802200307.
- Chatfield, T. 2014. "Google Travel Survey 2014." *Google Travel Research*. Google/Ipsos MediaCT. [https://storage.googleapis.com/think/docs/2014-travelers-road-to-decision\\_research\\_studies.pdf](https://storage.googleapis.com/think/docs/2014-travelers-road-to-decision_research_studies.pdf).
- . 2016. "The Trouble of Big Data Its Called Recency Bias." *BBC*. BBC. <http://www.bbc.com/future/story/20160605-the-trouble-with-big-data-its-called-the-recency-bias>.
- Denissen, Jaap JA, Rinie Geenen, Marcel AG Van Aken, Samuel D Gosling, and Jeff Potter. 2008. "Development and Validation of a Dutch Translation of the Big Five Inventory (Bfi)." *Journal of Personality Assessment* 90 (2). Taylor & Francis: 152–57.
- Donthu, Naveen, and David Gilliland. 1996. "The Infomercial Shopper." *Journal of Advertising Research* 36 (2). World Advertising Research Center Ltd.: 69–77.
- Euromonitor. 2015. "Travel Industry and Online Travel Global Overview." *Euromonitor*. Euromonitor.
- Eurostat. 2016. "Statistics on Ict Use in Tourism." *Eurostat*. Eurostat. [http://ec.europa.eu/eurostat/statistics-explained/index.php/Statistics\\_on\\_ICT\\_use\\_in\\_tourism](http://ec.europa.eu/eurostat/statistics-explained/index.php/Statistics_on_ICT_use_in_tourism).
- Evans, Joel R, and Anil Mathur. 2005. "The Value of Online Surveys." *Internet Research* 15 (2). Emerald Group Publishing Limited: 195–219.
- Fan, Jianqing, Fang Han, and Han Liu. 2014. "Challenges of Big Data Analysis."

- National Science Review* 1 (2). Oxford University Press: 293–314.
- Gemünden, Hans Georg. 1985. “Perceived Risk and Information Search. a Systematic Meta-Analysis of the Empirical Evidence.” *International Journal of Research in Marketing* 2 (2). Elsevier: 79–100.
- Glass, Gene V. 1976. “Primary, Secondary, and Meta-Analysis of Research.” *Educational Researcher* 5 (10). JSTOR: 3–8.
- Goldberg, Lewis R. 1993. “The Structure of Phenotypic Personality Traits.” *American Psychologist* 48 (1). American Psychological Association: 26.
- Gosling, Samuel D., Peter J. Rentfrow, and William B. Swann. 2003. “A very brief measure of the Big-Five personality domains.” *Journal of Research in Personality* 37 (6): 504–28. doi:10.1016/S0092-6566(03)00046-1.
- Ho, Chaang-ian, and Yung-ping Liu. 2005. “An exploratory investigation of web-based tourist information search behavior.” *Asia Pacific Journal of Tourism Research* 10 (4): 351–60. doi:10.1080/10941660500363645.
- Hofstede, Geert. 1980. “Motivation, Leadership, and Organization: Do American Theories Apply Abroad?” *Organizational Dynamics* 9 (1). Elsevier: 42–63.
- Hox, Joop J, and Hennie R Boeije. 2005. “Data Collection, Primary Vs. Secondary.” *Encyclopedia of Social Measurement* 1: 593–99.
- Jani, Dev. 2011. “The influence of personality on tourist information behaviour.” *E-Review of Tourism Research* 9 (3): 88–95.
- Jani, Dev, Jun-Ho Jang, and Yeong-Hyeon Hwang. 2014. “Big Five Factors of Personality and Tourists’ Internet Search Behavior.” *Asia Pacific Journal of Tourism Research* 19 (5): 600–615. doi:10.1080/10941665.2013.773922.
- Jun, S H, C A Vogt, and K J MacKay. 2007. “Relationships between Travel Information Search and Travel Product Purchase in Pretrip Contexts.” *Journal of Travel Research* 45 (April): 266–74. doi:10.1177/0047287506295945.
- Kim, Dae-Young, Xinran Y Lehto, and Alastair M Morrison. 2007. “Gender Differences in Online Travel Information Search: Implications for Marketing Communications on the Internet.” *Tourism Management* 28 (2). Elsevier: 423–33.
- Kim, Heejun, Zheng Xiang, and Daniel R Fesenmaier. 2015. “Use of The Internet for

Trip Planning: A Generational Analysis.” *Journal of Travel & Tourism Marketing* 32 (3): 276–89. doi:10.1080/10548408.2014.896765.

Labrinidis, Alexandros, and Hosagrahar V Jagadish. 2012. “Challenges and Opportunities with Big Data.” *Proceedings of the VLDB Endowment* 5 (12). VLDB Endowment: 2032–3.

Lepp, Andrew, and Heather Gibson. 2003. “Tourist roles, perceived risk and international tourism.” *Annals of Tourism Research* 30 (3): 606–24. doi:10.1016/S0160-7383(03)00024-0.

Leung, Rosanna, and Rob Law. 2010. “A Review of Personality Research in the Tourism and Hospitality Context.” *Journal of Travel & Tourism Marketing* 27 (5). Taylor & Francis: 439–59.

Mandrik, Carter A, and Yeqing Bao. 2005. “Exploring the Concept and Measurement of General Risk Aversion.” *Advances in Consumer Research* 32 (32): 531–39.

Mitkov, Ruslan. 2005. *The Oxford Handbook of Computational Linguistics*. Oxford University Press.

Nunan, Daniel, and MariaLaura Di Domenico. 2013. “Market Research and the Ethics of Big Data.” *International Journal of Market Research* 55 (4). The Market Research Society: 2–13.

Pan, Bing, and Daniel R. Fesenmaier. 2006. “Online Information Search. Vacation Planning Process.” *Annals of Tourism Research* 33 (3): 809–32. doi:10.1016/j.annals.2006.03.006.

Quintal, Vanessa Ann, Julie Anne Lee, and Geoffrey N. Soutar. 2010. “Tourists’ information search: The differential impact of risk and uncertainty avoidance.” *International Journal of Tourism Research* 12 (4): 321–33. doi:10.1002/jtr.753.

Roehl, W. S., and D. R. Fesenmaier. 1992. “Risk perceptions and pleasure travel: an exploratory analysis.” *Journal of Travel Research* 30 (4): 17–26. doi:10.1177/004728759203000403.

Sirakaya, Ercan, and Arch G Woodside. 2005. “Building and Testing Theories of Decision Making by Travellers.” *Tourism Management* 26 (6). Elsevier: 815–32.

Stewart, Susan I, and Christine A Vogt. 1999. “A Case-Based Approach to

- Understanding Vacation Planning.” doi:10.1080/014904099273165.
- Sweeney, Julian C, Geoffrey N Soutar, and Lester W Johnson. 1999. “The Role of Perceived Risk in the Quality-Value Relationship: A Study in a Retail Environment.” *Journal of Retailing* 75 (1). Elsevier: 77–105.
- Tole, Alexandru Adrian, and others. 2013. “Big Data Challenges.” *Database Syst J* 4 (3): 31–40.
- Urbany, Joel E, Peter R Dickson, and William L Wilkie. 1989. “Buyer Uncertainty and Information Search.” *Journal of Consumer Research* 16 (2). The Oxford University Press: 208–15.
- Wakoopa. 2016. “Mind the Gap.” *Wakoopa*. Wakoopa. <https://blog.wakoopa.com/mind-the-gap-a-case-study-about-reported-and-actual-behavior-on-mobile-devices-2/>.
- Woodside, Arch G, and Roberta MacDonald. 1994. “General System Framework of Customer Choice Processes of Tourism Services.” *Spoilt for Choice*. Kulturverl: Thaur, Germany, 30–59.
- Xiang, Zheng, and Ulrike Gretzel. 2010. “Role of Social Media in Online Travel Information Search.” *Tourism Management* 31 (2). Elsevier: 179–88.
- Xiang, Zheng, Vincent P. Magnini, and Daniel R. Fesenmaier. 2015. “Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet.” *Journal of Retailing and Consumer Services* 22. Elsevier: 244–49. doi:10.1016/j.jretconser.2014.08.005.
- Yong, An Gie, and Sean Pearce. 2013. “A Beginner’s Guide to Factor Analysis: Focusing on Exploratory Factor Analysis.” *Tutorials in Quantitative Methods for Psychology* 9 (2): 79–94.
- Yoo, Boonghee, and Naveen Donthu. 2002. “Testing Cross-Cultural Invariance of the Brand Equity Creation Process.” *Journal of Product & Brand Management* 11 (6). MCB UP Ltd: 380–98.

# Appendix 1

Grigor Dimitrov

September 27, 2016

## Panel

### How many panelists are active?

---

Total panelists

**6,682**

-4.8% compared with last month

---

New panelists

-93.1% compared with last month

**18**

---

Reactivated panelists

-29.3% compared with last month

**246**

---

Panelists that were already active

+0.1% compared with last month

**6,418**

---

### What combination of devices did they use?

**+** **+** **+** Only desktop

-252 compared with last month

5,910 **88.4%**

---

**+** **+** **+** Only smartphone

-39 compared with last month

340 **5.1%**

---

**+** **+** **+** Only tablet

-6 compared with last month

41 **0.6%**

---

**+** **+** **+** Desktop + Smartphone

-31 compared with last month

269 **4.0%**

---

**+** **+** **+** Desktop + Tablet

-6 compared with last month

87 **1.3%**

---

**+** **+** **+** Smartphone + Tablet

+0 compared with last month

11 **0.2%**

---

**+** **+** **+** Desktop + Smartphone + Tablet

-1 compared with last month

24 **0.4%**

---

Figure 1: Panelists

## How many devices were active?

### Total active devices

**7,103**

-5.0% compared with last month

### 🖥️ Total active desktops

**6,294** 88.6%

-4.4% compared with last month

#### New

-93.1% compared with last month

**19**

#### Reactivated

-24.4% compared with last month

**233**

#### Already active

+0.6% compared with last month

**6,042**

### 📱 Total active tablets

**165** 2.3%

-7.3% compared with last month

#### New

-

**0**

#### Reactivated

-50.0% compared with last month

**6**

#### Already active

-3.0% compared with last month

**159**

### 📱 Total active smartphones

**644** 9.1%

-9.9% compared with last month

#### New

-88.9% compared with last month

**2**

#### Reactivated

-29.7% compared with last month

**45**

#### Already active

-5.7% compared with last month

**597**

Figure 2: Devices

Data and variables exaple

##		used_at	host	panelist_id	url
## 1	2016-01-01	00:47:06	telegraaf.nl	2112	telegraaf.nl
## 2	2016-01-01	00:48:30	telegraaf.nl	2112	telegraaf.nl
## 3	2016-01-01	00:51:28	telegraaf.nl	2112	telegraaf.nl
## 4	2016-01-01	00:53:29	telegraaf.nl	2112	telegraaf.nl
## 5	2016-01-01	01:28:52	telegraaf.nl	2112	telegraaf.nl
## 6	2016-01-01	02:01:00	telegraaf.nl	2112	telegraaf.nl
## 7	2016-01-01	02:01:12	twitter.com	2112	twitter.com
## 8	2016-01-01	02:04:07	twitter.com	2112	twitter.com
## 9	2016-01-01	02:05:07	twitter.com	2112	twitter.com
## 10	2016-01-01	02:27:31	twitter.com	2112	twitter.com
##	active_seconds	browser_name	mmid	Class_Travel	purchase
## 1	72	safari	1	0	NA
## 2	152	safari	1	0	NA
## 3	0	safari	1	0	NA
## 4	0	safari	1	0	NA
## 5	71	safari	2	0	NA
## 6	11	safari	3	0	NA
## 7	168	safari	3	0	NA
## 8	60	safari	3	0	NA
## 9	0	safari	3	0	NA
## 10	15	safari	4	0	NA

Figure 3: Desktop data

```

##          app_name host panelist_id device_id scheme url domain
## 1 ABN AMRO Mobiel Bankieren <NA> 1008505 10093 <NA> <NA> <NA>
## 2 ABN AMRO Mobiel Bankieren <NA> 1008505 10093 <NA> <NA> <NA>
## 3 ABN AMRO Mobiel Bankieren <NA> 1008505 10093 <NA> <NA> <NA>
## 4 ABN AMRO Mobiel Bankieren <NA> 1008505 10093 <NA> <NA> <NA>
## 5 ABN AMRO Mobiel Bankieren <NA> 1008505 10093 <NA> <NA> <NA>
## 6 ABN AMRO Mobiel Bankieren <NA> 1008505 10093 <NA> <NA> <NA>
## 7 ABN AMRO Mobiel Bankieren <NA> 1008505 10093 <NA> <NA> <NA>
## 8 ABN AMRO Mobiel Bankieren <NA> 1008505 10093 <NA> <NA> <NA>
## 9 ABN AMRO Mobiel Bankieren <NA> 1008505 10093 <NA> <NA> <NA>
## 10 ABN AMRO Mobiel Bankieren <NA> 1008505 10093 <NA> <NA> <NA>
##          app_id          used_at connection
## 1 cb46bcba-7258-4b47-8557-de3ff607b456 2016-02-22 13:43:15      wifi
## 2 cb46bcba-7258-4b47-8557-de3ff607b456 2016-02-03 18:03:34    cellular
## 3 cb46bcba-7258-4b47-8557-de3ff607b456 2016-01-23 15:38:49      wifi
## 4 cb46bcba-7258-4b47-8557-de3ff607b456 2016-01-16 15:11:21      wifi
## 5 cb46bcba-7258-4b47-8557-de3ff607b456 2016-02-03 18:43:44      wifi
## 6 cb46bcba-7258-4b47-8557-de3ff607b456 2016-01-08 14:15:29      wifi
## 7 cb46bcba-7258-4b47-8557-de3ff607b456 2016-01-24 10:20:45      wifi
## 8 cb46bcba-7258-4b47-8557-de3ff607b456 2016-02-03 17:00:56    cellular
## 9 cb46bcba-7258-4b47-8557-de3ff607b456 2016-01-22 21:34:28    cellular
## 10 cb46bcba-7258-4b47-8557-de3ff607b456 2016-01-06 07:24:50      wifi
## duration mmid Class_Travel TravelApp
## 1      21  799          0          0
## 2      19  551          0          0
## 3      34  392          0          0
## 4      11  283          0          0
## 5      21  552          0          0
## 6      23  139          0          0
## 7      25  396          0          0
## 8       9  549          0          0
## 9      45  381          0          0
## 10     15   98          0          0

```

Figure 4: Mobile data

## Analysis

The following section contains an explanation of the main techniques performed during the analysis along with their assumptions, followed by the results of the ordinary least squares diagnostic tests of the restricted model. Once the functional form of the restricted model has been selected I proceed with variable selection in order to come up with the final model. Finally, I ran the diagnostics of tests over the final model again.

## Factor Analysis

Factor analysis is a widely used technique used for explaining the variance in several variables by a smaller set of latent variables. As in the current case, it is often used to consolidate several survey variables onto their “underlying” factors in order to reduce the dimensionality of the data. Factor analysis groups variables together, that is, using a lot of variables one can potentially reduce them to certain factors representing the latent underlying factors representing them by accounting the similar patterns in the variables. The intuition behind the analysis is as follows. The analysis groups together observed, correlated variables into smaller groups of unobserved (latent) variables (Yong and Pearce 2013).

In this case, I use factor analysis to to reduce the seven survey items regarding the risk and uncertainty attitude down to two constructs namely risk and uncertainty. Also to reduce the eleven-item scale of BIG5 to 5 factors representing



each of on the of the five personality traits.

## Regression Analysis

For testing the hypotheses of this paper, regression analysis will be utilized. The regression model or ordinary least squares (hereafter OLS) is the “cornerstone of econometrics” (Verbeek 2008). It aims at explaining a variable,  $y$ , in terms of another variable,  $x$ . In other words, using OLS researchers are able to find how will  $y$  vary as  $x$  changes, the ultimate goal being to infer the causal effect  $x$  has on  $y$ . Using such models allows to find relationships between various variables, present the effect the independent variables,  $x_i$  has on the dependent variable,  $y$  in order to be able to make predictions.

The general linear regression models is represented as follows:

$$y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \dots + X_k\beta_k + \varepsilon$$

Where:  $y$  is the dependent variable

$X_1$  to  $X_k$  are the independent variables, which explain  $y$

$\beta_0$  is the intercept, indicating the expected value of  $y$  when all the independent variables are equal to 0

$\beta_1$  to  $\beta_k$  are the coefficients which determine the effect  $x$  has on  $y$

$\varepsilon$  is the error term

## Goodness of fit and model selection

The standard measures of fit include the R-squared and the adjusted R-squared, which measures the variance that is explained in the model for the independent variable by the dependent variables. The measure can be interpreted directly. For example, if the R-squared is equal to 0.45, it means that the variables included in the model explain 45% of the variation of the independent variable,  $y$ . The higher the value, the higher its predictive power. However, it should be noted that adjusted R-squared penalizes for the additional number of parameters. Thus, applying additional variables to the model, I should test if they are jointly significant in order to assess whether they are relevant or not in the model. This is typically applied by using the F-test (Wald test). Using both the R-squared, the overall F-test and applying the F-test to certain variables I can compare best which model fits the data best.

## Akaike information criterion (AIC)

Model selection has been done over Akaike information criterion (AIC) introduced by Akaike (1974). AIC is a metric traditionally used for model selection. It compares the goodness of fit for a number of explanatory variables and penalizes for each additional explanatory variable.

## BLUE Assumptions

There are several assumptions that need to be met when applying OLS explained in the section below. Namely Gauss-Markov assumptions for full ideal conditions for OLS. The model needs to be best linear unbiased estimator (“BLUE”) (Verbeek 2008). It is crucial for the assumptions to be met as to compute unbiased and consistent estimates that explain the variation in the dependent variable. Now, I will go through each assumption: Linear in parameters This implies that the model should have linear parameters,  $\beta$ , however, there can be nonlinearities in the variables,  $x$ . This assumption is met as my specified model does not include non-linearities in the parameters.

## Normality

The error terms should follow a normal distribution. In large datasets, however, even if the error term does not follow a normal distribution the regression estimators are ‘asymptotically normally distributed’, meaning that following non-normal distribution is not crucial as the estimates will still be consistent and unbiased. The Shapiro-Wilk test can

is adopted here and results presented below. The test works under null hypothesis: “the sample comes from normally distributed population” Shapiro and Wilk (1965)

### **Random sample**

The data collection should be done randomly, meaning that the each subject should have the same probability of being selected. In this research, both in the behavioral and survey data collection parts, I can say that subjects were randomly selected for further analysis.

### **Multicollinearity**

Multicollinearity implies that there is no perfect linear relationship between the independent (explanatory) variables as this can lead to ‘unreliable regression estimated’ (Verbeek 2008). For example, adding both male and female in the analysis would lead to perfect collinearity (as  $\text{male} + \text{female} = 1$ ) and the estimations would not work. In this example, removing one of the variables would solve the problem, however, there can be other variables that are highly correlated. Having multicollinearity would not lead to biased estimates, but to inaccurate estimates. In such a case, excluding variables from the model should be considered. There are no tests that specifically look for multicollinearity, however, there are certain indications. For instance, having two variables that are jointly significant (have big F-statistics), but independently are not significant can be a sign of multicollinearity

### **Homoscedasticity**

Homoscedasticity implies that the variance of the error term should be the same for all values of the independent variables. If this does not hold, there is a problem with heteroscedasticity meaning that the estimates of the regression are inconsistent due to an inaccuracy of their standard errors, meaning that the t-statistics and thus the significance level of the estimates are not valid anymore. To test for homoscedasticity, I perform the Breusch-Pagan test, which hypothesizes that there is a constant variance of the error terms.

### **Endogeneity**

The last assumption is crucial to be met as otherwise, the regression estimates are biased and inconsistent. Endogeneity implies that there is a correlation between an independent variable and the error term. There are several reasons why this assumption does not hold:

1. The model is misspecified. That is, nonlinearities are missing from the model or interaction effects are not accounted for. To account for that I perform the Ramsey-Reset test. The tests add fitted values on power and re-estimates the model. The intuition behind it is that if a nonlinear combination of independent variables can explain the dependent variable there are evidence the model is misspecified. The Ramsey-Reset test work under null hypothesis that the model has no important omitted non-linearities (Ramsey 1974)
2. Endogeneity, meaning that we are either missing important variables that explain the variance in the independent variables or we have reverse causality, that means that there can be a loop of causality between the independent and dependent variable.

### **Stepwise regression**

The idea of stepwise regression has been introduced by Hastie and Pregibon (1992) and further improved by Ripley (2002). It is an iterative function ran over a restricted model and a set of candidate models. Each candidate model consists of a different set of explanatory variables. The function computes iteratively Akaike information criterion (AIC) values for the models comparing them to the best performing models from the previous iteration and based on the performance chooses whether to continue the loop with the new model or remain with the old one. The final output is the best performing model.

# Results

## Factor analysis

```
##  
## Call:  
## factanal(x = RAUdata, factors = 2, rotation = "varimax")  
##  
## Uniquenesses:  
## Q16_1 Q16_2 Q16_3 Q16_4 Q16_5 Q16_6 Q16_7  
## 0.35 0.20 0.64 0.63 0.33 0.25 0.60  
##  
## Loadings:  
##      Factor1 Factor2  
## Q16_1      0.71  
## Q16_2      0.88  
## Q16_3      0.53  
## Q16_4 0.49  
## Q16_5 0.78  
## Q16_6 0.84  
## Q16_7 0.55  
##  
##      Factor1 Factor2  
## SS loadings      2.11  1.90  
## Proportion Var   0.30  0.27  
## Cumulative Var   0.30  0.57  
##  
## Test of the hypothesis that 2 factors are sufficient.  
## The chi square statistic is 29.15 on 8 degrees of freedom.  
## The p-value is 0.000299
```

Figure 5: Factor analysis results, risk and uncertainty

```

##
## Call:
## factanal(x = BIG5data, factors = 5, rotation = "varimax")
##
## Uniquenesses:
##  Q15_1  Q15_2  Q15_3  Q15_4  Q15_5  Q15_6  Q15_7  Q15_8  Q15_9  Q15_10
##  0.70   0.49   0.68   0.52   0.60   0.61   0.70   0.21   0.63   0.80
## Q15_11
##  0.48
##
## Loadings:
##      Factor1 Factor2 Factor3 Factor4 Factor5
## Q15_1                -0.52
## Q15_2                0.69
## Q15_3                    0.49
## Q15_4                -0.63
## Q15_5                0.61
## Q15_6                0.60
## Q15_7 -0.40
## Q15_8  0.89
## Q15_9                0.60
## Q15_10
## Q15_11                0.69
##
##      Factor1 Factor2 Factor3 Factor4 Factor5
## SS loadings    0.99   0.97   0.97   0.91   0.74
## Proportion Var  0.09   0.09   0.09   0.08   0.07
## Cumulative Var  0.09   0.18   0.27   0.35   0.42
##
## Test of the hypothesis that 5 factors are sufficient.
## The chi square statistic is 4.15 on 10 degrees of freedom.
## The p-value is 0.94

```

Figure 6: Factor analysis results, BIG5

## OLS

Table 1: Desktop data, restricted model

	<i>Dependent variable:</i>				
	log(MM)	log(Domains)	log(PV)	log(Time)	log(Length)
	(1)	(2)	(3)	(4)	(5)
Risk:seek	-0.16 (0.12)	-0.25** (0.12)	-0.42*** (0.13)	-0.47*** (0.13)	-0.28** (0.12)
Uncertainty:seek	-0.15 (0.11)	-0.26** (0.11)	-0.31*** (0.12)	-0.32*** (0.12)	-0.21* (0.11)
Days	0.01*** (0.001)	0.01*** (0.001)	0.004*** (0.002)	0.004*** (0.002)	0.01*** (0.001)
D MM	1.17*** (0.10)				
D Domains		1.08*** (0.10)			
D PV			0.88*** (0.11)		
D Time				0.86*** (0.11)	
D Length					1.34*** (0.10)
D Purchase	0.50*** (0.11)	0.62*** (0.11)	0.86*** (0.11)	0.87*** (0.11)	0.48*** (0.10)
Risk x Uncertainty	0.27 (0.30)	0.32 (0.31)	0.66** (0.32)	0.83** (0.32)	0.48 (0.29)
Constant	3.12*** (0.27)	3.59*** (0.27)	5.66*** (0.29)	8.94*** (0.29)	10.99*** (0.26)
Observations	426	426	426	426	426
R <sup>2</sup>	0.36	0.35	0.30	0.31	0.41
Adjusted R <sup>2</sup>	0.35	0.34	0.29	0.30	0.41
Residual Std. Error (df = 419)	0.97	0.98	1.02	1.03	0.93
F Statistic (df = 6; 419)	39.89***	37.21***	29.69***	31.07***	49.37***

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 2: Mobile data, restricted model

	<i>Dependent variable:</i>				
	log(MM)	log(Domains)	log(PV)	log(Time)	log(Length)
	(1)	(2)	(3)	(4)	(5)
Risk:seek	0.36 (0.39)	0.29 (0.31)	0.42 (0.42)	0.96 (0.91)	1.26 (0.92)
Uncertainty:seek	0.84** (0.38)	0.66** (0.30)	0.54 (0.41)	1.49* (0.89)	2.26** (0.89)
Days	0.01** (0.003)	0.01*** (0.002)	0.004 (0.003)	0.02*** (0.01)	0.02*** (0.01)
D MM	1.50*** (0.36)				
D Domains		2.28*** (0.29)			
D PV			2.55*** (0.40)		
D Time				3.09*** (0.82)	
D Length					3.13*** (0.82)
Risk x Uncertainty	-2.50 (1.58)	-2.07 (1.25)	-2.07 (1.72)	-7.86** (3.71)	-8.74** (3.75)
Constant	0.10 (0.46)	0.03 (0.36)	0.37 (0.50)	1.60 (1.08)	1.58 (1.08)
Observations	101	101	101	101	101
R <sup>2</sup>	0.36	0.55	0.42	0.37	0.36
Adjusted R <sup>2</sup>	0.33	0.52	0.38	0.33	0.33
Residual Std. Error (df = 95)	1.50	1.19	1.64	3.52	3.56
F Statistic (df = 5; 95)	10.84***	22.94***	13.49***	10.99***	10.72***

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 3: Combined data, restricted model

	<i>Dependent variable:</i>				
	log(MM)	log(Domains)	log(PV)	log(Time)	log(Length)
	(1)	(2)	(3)	(4)	(5)
Risk:seek	-0.20 (0.13)	-0.20 (0.13)	-0.35*** (0.13)	-0.38*** (0.14)	-0.19 (0.15)
Uncertainty:seek	-0.13 (0.12)	-0.15 (0.12)	-0.23* (0.13)	-0.19 (0.13)	-0.17 (0.14)
Days	0.01*** (0.001)	0.01*** (0.001)	0.01*** (0.002)	0.01*** (0.002)	0.01*** (0.002)
D MM	0.91*** (0.11)				
D Domains		1.01*** (0.10)			
D PV			0.82*** (0.11)		
D Time				0.77*** (0.12)	
D Length					0.98*** (0.12)
D Purchase	0.59*** (0.11)	0.62*** (0.11)	0.90*** (0.12)	0.85*** (0.12)	0.51*** (0.13)
Risk x Uncertainty	0.25 (0.32)	0.27 (0.31)	0.58* (0.34)	0.58 (0.35)	0.48 (0.37)
Constant	2.67*** (0.29)	3.16*** (0.28)	5.08*** (0.30)	8.10*** (0.31)	10.01*** (0.32)
Observations	429	429	429	429	429
R <sup>2</sup>	0.30	0.34	0.30	0.29	0.29
Adjusted R <sup>2</sup>	0.29	0.33	0.29	0.28	0.28
Residual Std. Error (df = 422)	1.03	1.00	1.07	1.13	1.16
F Statistic (df = 6; 422)	30.57***	36.53***	29.52***	29.22***	29.27***

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 4: OLS Assumptions - Restricted model

data_name	Residual_Standard_Err	F_Stat	NumDF	FDenDF	R_Sq	Adj_R_Sq	Shapiro_Wilk_Stat	Shapiro_Wilk_P_val	Reset_Stat	Reset_P_val	BP_Stat	BP_P_val
DESKTOP:REST:log(MM)	0.97	39.89	6	419	0.36	0.35	1.00	0.94	2.03	0.13	4.65	0.59
DESKTOP:REST:log(TD)	0.98	37.21	6	419	0.35	0.34	1.00	0.24	1.20	0.30	10.48	0.11
DESKTOP:REST:log(TDPV)	1.02	29.69	6	419	0.30	0.29	0.99	0.16	0.79	0.45	6.82	0.34
DESKTOP:REST:log(TT)	1.03	31.07	6	419	0.31	0.30	0.98	0.0001	1.31	0.27	10.88	0.09
DESKTOP:REST:log(TL)	0.93	49.37	6	419	0.41	0.41	0.99	0.01	1.97	0.14	7.77	0.26
MOBILE:REST:log(MM)	1.50	10.84	5	95	0.36	0.33	0.96	0.002	0.89	0.42	2.90	0.72
MOBILE:REST:log(TD)	1.19	22.94	5	95	0.55	0.52	0.97	0.01	2.40	0.10	3.37	0.64
MOBILE:REST:log(TDPV)	1.64	13.49	5	95	0.42	0.38	0.91	0.0000	2.04	0.14	1.84	0.87
MOBILE:REST:log(TT)	3.52	10.99	5	95	0.37	0.33	0.98	0.06	0.03	0.97	16.62	0.01
MOBILE:REST:log(TL)	3.56	10.72	5	95	0.36	0.33	0.97	0.02	0.45	0.64	14.80	0.01
COMBINED:REST:log(MM)	1.03	30.57	6	422	0.30	0.29	0.99	0.02	10.19	0.0000	11.52	0.07
COMBINED:REST:log(TD)	1.00	36.53	6	422	0.34	0.33	0.99	0.001	8.68	0.0002	11.67	0.07
COMBINED:REST:log(TDPV)	1.07	29.52	6	422	0.30	0.29	0.98	0.0000	9.92	0.0001	19.72	0.003
COMBINED:REST:log(TT)	1.13	29.22	6	422	0.29	0.28	0.93	0	12.93	0.0000	26.66	0.0002
COMBINED:REST:log(TL)	1.16	29.27	6	422	0.29	0.28	0.85	0	3.94	0.02	30.10	0.0000

Note:

Based on the results of Reset-test, all of the models containing COMBINED:REST have been rejected



Table 5: Full model: Desktop, robust

	<i>Dependent variable:</i>				
	D log(MM) (1)	D log(TD) (2)	D log(TDPV) (3)	D log(TT) (4)	D log(TL) (5)
Risk Seeking	-0.11 p = 0.36	-0.23* p = 0.08	-0.37*** p = 0.005	-0.46*** p = 0.0002	-0.26** p = 0.03
Uncertainty Seeking	-0.13 p = 0.26	-0.25** p = 0.04	-0.34*** p = 0.004	-0.37*** p = 0.002	-0.17 p = 0.12
Risk(x)Uncertainty: Seeking	0.09 p = 0.73	0.10 p = 0.69	0.39 p = 0.16	0.76** p = 0.02	0.30 p = 0.22
Days active	0.01*** p = 0.001	0.01*** p = 0.002	0.01** p = 0.02	0.01** p = 0.04	0.005*** p = 0.01
D Purchase	0.53*** p = 0.0000	0.63*** p = 0.0000	0.84*** p = 0.00	0.88*** p = 0.00	0.50*** p = 0.0000
D Micromoments	1.13*** p = 0.00				
D Domains		1.03*** p = 0.00			
D PageViews			0.83*** p = 0.00		
D Time				0.80*** p = 0.00	
D MM Length					1.36*** p = 0.00
S1. x2 travel related purchase		0.21* p = 0.06	0.21* p = 0.08	0.29** p = 0.02	
S1. x3+ travel related purchase		0.15 p = 0.21	0.20 p = 0.12	0.28** p = 0.02	
S3. Not the only decision maker	-0.20* p = 0.06	-0.28*** p = 0.01	-0.33*** p = 0.002	-0.26** p = 0.02	-0.26*** p = 0.005
D1. Age	1.01 p = 0.33	2.58** p = 0.02	-0.52 p = 0.63		
D1. Age Sq	1.87* p = 0.06	2.14** p = 0.02	2.01** p = 0.04		
Q1. Country: Asia (Base Europe)		-0.51 p = 0.20	-0.30 p = 0.45	0.08 p = 0.84	0.33* p = 0.08
Q1. Country: North America (Base Europe)		0.65 p = 0.34	0.65 p = 0.37	0.57 p = 0.44	0.38** p = 0.02
Q1. Country: South America (Base Europe)		-0.01 p = 0.98	-0.82 p = 0.18	-1.13* p = 0.07	0.25 p = 0.29
Q1. Country: Australia (Base Europe)		-0.64 p = 0.21	-0.65 p = 0.25	-0.36 p = 0.60	-0.63 p = 0.14
Q1. Country: Africa (Base Europe)		-0.05 p = 0.83	0.12 p = 0.59	0.35 p = 0.19	-0.14 p = 0.63
Q4. Planning horizon (Weeks)			0.01 p = 0.18		
Q5.3. Used tourist information office			0.43* p = 0.07		0.39 p = 0.16
Q5.5. Used travel agent	-0.28** p = 0.04	-0.38** p = 0.02	-0.37** p = 0.02	-0.31** p = 0.05	
Q6.1. Purchased online: Transport	-0.21** p = 0.05	-0.24** p = 0.02	-0.29*** p = 0.005	-0.37*** p = 0.0003	-0.17* p = 0.09
Q7. Trip longer than 3 nights	0.18** p = 0.05	0.20** p = 0.04	0.14 p = 0.17	0.20** p = 0.05	0.19** p = 0.05
Q8. Visited before	-0.16 p = 0.14	-0.17* p = 0.10			
Q11.1. Children (N)	-0.27* p = 0.07	-0.25* p = 0.09	-0.34** p = 0.03	-0.23 p = 0.11	
Q11.3. Stayed at Friends/Relatives (N)	0.27* p = 0.10	0.36** p = 0.02	0.37** p = 0.03		0.38** p = 0.02
Q11.5. Group Trip (N)	-0.43* p = 0.07				-0.41** p = 0.04
Openness to experience (low)			-0.32* p = 0.06	-0.33** p = 0.04	
Consciousness (low)				-0.20* p = 0.06	
Agreeableness (low)			-0.40 p = 0.22	-0.36 p = 0.21	
Asia x Travel Agent		1.16*** p = 0.01	0.79* p = 0.08	0.54 p = 0.23	
North America x Travel Agent		-0.31 p = 0.65	-0.46 p = 0.53	-0.37 p = 0.62	
South America x Travel Agent		0.32 p = 0.61	1.12* p = 0.09	1.74*** p = 0.01	
Australia x Travel Agent		0.68 p = 0.20	1.09* p = 0.09	0.55 p = 0.45	
Constant	3.91*** p = 0.00	3.89*** p = 0.00	5.36*** p = 0.00	9.29*** p = 0.00	10.69*** p = 0.00

Table 6: Full model: Mobile, robust

	<i>Dependent variable:</i>				
	M log(MM) (1)	M log(TD) (2)	M log(TDPV) (3)	M log(TT) (4)	M log(TL) (5)
Risk Seeking	0.16 p = 0.66	0.13 p = 0.61	0.53 p = 0.22	1.32 p = 0.17	1.09 p = 0.21
Uncertainty Seeking	0.72** p = 0.04	0.73** p = 0.02	0.57 p = 0.11	1.98* p = 0.06	2.05** p = 0.02
Risk(x)Uncertainty: Seeking	-1.76*** p = 0.003	-1.80*** p = 0.0000	-0.75 p = 0.34	-7.89*** p = 0.00	-7.37*** p = 0.0000
Days active	0.01*** p = 0.001	0.01*** p = 0.0001	0.004 p = 0.12	0.02*** p = 0.0000	0.02*** p = 0.0001
D Micromoments	1.03*** p = 0.004				
D Domains		1.77*** p = 0.0000			
D PageViews			2.44*** p = 0.0000		
D Time				2.06*** p = 0.01	
D MM Length					2.20*** p = 0.003
S2. Primary purchase: Flight	-1.09*** p = 0.0002	-0.55** p = 0.03	-1.02*** p = 0.01	-1.22* p = 0.07	-1.06 p = 0.12
S3. Not the only decision maker	-0.70** p = 0.02	-0.40* p = 0.10	-1.03*** p = 0.0005	-2.19*** p = 0.001	-2.51*** p = 0.0003
D1. Age			-3.51** p = 0.02		
D1. Age Sq			-0.97 p = 0.55		
Q1. Country: Asia (Base Europe)			1.99*** p = 0.01		
Q1. Country: North America (Base Europe)			0.33 p = 0.50		
Q1. Country: South America (Base Europe)			0.40 p = 0.52		
Q1. Country: Australia (Base Europe)			0.90 p = 0.30		
Q1. Country: Africa (Base Europe)	0.04** p = 0.02	0.03** p = 0.04	0.04* p = 0.06	0.06** p = 0.05	0.05 p = 0.14
Q4. Planning horizon (Weeks)			0.80** p = 0.03		
Q5.2. Used advice of friends or relatives				-2.86** p = 0.04	
Q5.3. Used tourist information office	-0.49 p = 0.19	-0.39 p = 0.18	-0.95** p = 0.02	-1.29 p = 0.19	-1.60* p = 0.10
Q5.4. Used travel magazines			1.37* p = 0.09		
Q5.5. Used travel agent		0.48* p = 0.08		2.47*** p = 0.001	2.05*** p = 0.004
Q6.1. Purchased online: Transport	0.62 p = 0.11		0.90** p = 0.04		
Q6.3. Purchased online: Entertainment	-0.89*** p = 0.001	-0.62*** p = 0.01	-0.53 p = 0.13	-1.87*** p = 0.004	-1.73*** p = 0.01
Q7. Trip longer than 3 nights	0.59** p = 0.02		0.53* p = 0.07	1.45** p = 0.02	1.77*** p = 0.004
Q8. Visited before					0.58*** p = 0.01
Q11.2. Visited Friends/Relatives (N)			1.23*** p = 0.01		
Q11.3. Stayed at Friends/Relatives (N)	1.08*** p = 0.004	0.93*** p = 0.005		3.49*** p = 0.0002	3.11*** p = 0.002
Q11.4. Stayed at Hotel/Motel/AirBNB (N)	0.91* p = 0.07	0.61 p = 0.12	0.88 p = 0.12	2.21** p = 0.05	2.01** p = 0.05
Consciousness (low)	0.60** p = 0.04	0.53** p = 0.03			
Extraversion (low)			0.88** p = 0.02		
Agreeableness (low)	-1.17* p = 0.08	-1.26*** p = 0.005	-1.86*** p = 0.01	-3.56* p = 0.06	-3.08 p = 0.13
Neuroticism (low)		0.67*** p = 0.002	0.74** p = 0.03	2.88*** p = 0.004	3.27*** p = 0.0001
North America x Travel Agent			-2.20** p = 0.03		
Constant	-0.13 p = 0.86	-0.98 p = 0.13	-2.64* p = 0.06	-0.98 p = 0.66	-4.42** p = 0.03

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 7: OLS Assumption - Full model tests, Desktop

	Residual_Standard_Err	F_Stat	NumDF	FDenDF	R_Sq	Adj_R_Sq	Shapiro_Wilk_Stat	Shapiro_Wilk_P_val	Reset_Stat	Reset_P_val	BP_Stat	BP_P_val
log(MM)	0.942	17.762	16	409	0.410	0.387	0.997	0.655	0.789	0.455	12.209	0.729
log(TD)	0.943	11.378	26	399	0.426	0.388	0.993	0.056	2.085	0.126	23.757	0.644
log(TD(PV))	0.979	12.899	19	406	0.376	0.347	0.991	0.008	0.044	0.957	29.519	0.058
log(TT)	0.982	10.797	24	401	0.393	0.356	0.977	0.00000	3.194	0.042	36.150	0.069
log(TL)	0.903	20.830	17	408	0.465	0.442	0.993	0.068	0.575	0.563	19.391	0.307

Table 8: OLS Assumption - Full model, Mobile

	Residual_Standard_Err	F_Stat	NumDF	FDenDF	R_Sq	Adj_R_Sq	Shapiro_Wilk_Stat	Shapiro_Wilk_P_val	Reset_Stat	Reset_P_val	BP_Stat	BP_P_val
log(MM)	1.288	8.393	14	86	0.577	0.509	0.989	0.543	3.061	0.052	17.158	0.248
log(TD)	1.037	11.351	17	83	0.699	0.638	0.982	0.188	1.682	0.192	10.712	0.871
log(TDPV)	1.373	5.848	27	73	0.684	0.567	0.991	0.761	1.816	0.170	29.234	0.505
log(TT)	2.855	8.979	16	84	0.631	0.561	0.987	0.452	3.920	0.024	21.297	0.167
log(TL)	2.909	8.691	16	84	0.623	0.552	0.988	0.520	4.248	0.018	21.391	0.164

## Reference

- Akaike, Hirotugu. 1974. "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control* 19 (6). Ieee: 716–23.
- Hastie, TJ, and D Pregibon. 1992. "Statistical Models in S, Chapter Generalized Linear Models." *Wadsworth & Brooks/Cole* 51.
- Ramsey, James B. 1974. "Classical Model Selection Through Specification Error Tests." *Frontiers in Econometrics*. Academic Press New York, 13–47.
- Ripley, BD. 2002. "Modern Applied Statistics with S." Springer-Verlag, New York.
- Shapiro, Samuel Sanford, and Martin B Wilk. 1965. "An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika* 52 (3/4). JSTOR: 591–611.
- Verbeek, Marno. 2008. *A Guide to Modern Econometrics*. John Wiley & Sons.
- Yong, An Gie, and Sean Pearce. 2013. "A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis." *Tutorials in Quantitative Methods for Psychology* 9 (2): 79–94.

# Appendix 2: CODE

## Contents

CODE CHUNK #1. INITIAL DATA PROCESSING	1
CODE CHUNK #2A. TRAVEL WEBSITES KEYWORDS CATEGORIZATION	4
CODE CHUNK #2B. TRAVEL WEBSITES WEB SCRAPER + UCLASSIFY.COM API	6
CODE CHUNK #3A. ADD PURCHASE DATA	8
CODE CHUNK #3B. ADD TRAVEL CATEGORIZATION	9
CODE CHUNK #4. AGGREGATE DATA	12
CODE CHUNK #5. SURVEY DATA PROCESSING	17
CODE CHUNK #6. MERGE SURVEY DATA AND PROCESSED BEHAVIORAL DATA	18
CODE CHUNK #8. FINAL ANALYSIS AND OLS TESTS	20

## CODE CHUNK #1. INITIAL DATA PROCESSING

```
##### LOAD #####
load(file = "DesktopPV.Rdata")
load(file = "MobilePV.Rdata")
##### DESKTOP #####
library(urltools)
# GET THE HOST NAME (DOMAIN + SUBDOMAIN)
testD$host <- suffix_extract(domain(testD$url))[1]$host
visits <- testD
# GET USER IDS
uids <- sort(unique(visits$panelist_id))
# CREATE EMPTY LIST
chunks <- list()
# FILL IN THE LIST WITH DATA PER EACH ID AND CREATE
# MICROMOMENTS
for (ivar in 1:length(uids)) {
  workarr <- subset(visits, visits$panelist_id == uids[ivar])
  print(c("Start", ivar, nrow(workarr), as.character(Sys.time())))
  workarr <- workarr[order(workarr$used_at), ]
  mmid <- 1
  for (ivar2 in 1:(nrow(workarr) - 1)) {
    if (as.POSIXct(workarr$used_at[ivar2]) + workarr$active_second[ivar2] +
        300 < as.POSIXct(workarr$used_at[ivar2 + 1])) {
      workarr$mmid[ivar2] <- mmid
      mmid <- mmid + 1
    } else {
      workarr$mmid[ivar2] <- mmid
    }
  }
}
if (as.POSIXct(workarr$used_at[nrow(workarr) - 1]) + workarr$active_second[nrow(workarr) -
1] + 300 < as.POSIXct(workarr$used_at[nrow(workarr)])) {
```

```

    workarr$mmid[nrow(workarr)] <- mmid
  } else {
    workarr$mmid[nrow(workarr)] <- mmid - 1
  }
  chunks[[ivar]] <- workarr
  print(c("End", ivar, nrow(workarr), as.character(Sys.time())))
  rm(workarr)
}
# FIX OBS. 184
save(chunks, file = "chunks.Rdata")
##### Aggregate DESKTOP #####
# CREATE EMPTY LIST FOR THE AGGREGATED DATA ON MICROMOMENT
# LEVEL
aggchunks <- list()
# FILL IN THE LIST FOR THE AGGREGATED DATA ON MICROMOMENT
# LEVEL
for (forvar in 1:length(chunks)) {
  testdata <- NULL
  aggtest <- NULL
  testdata <- chunks[[forvar]]
  print(c("Start", nrow(testdata), forvar, as.character(Sys.time())))
  # NUMBER OF DOMAINS VISITS
  aggtest <- aggregate(testdata$host ~ testdata$mmid, data = testdata,
    FUN = function(x) length(unique(x)))
  colnames(aggtest) <- c("mmid", "domains")
  # NUMBER OF PAGEVIEWS VISITS
  aggtest$pageviews <- unlist(aggregate(testdata$url ~ testdata$mmid,
    data = testdata, FUN = length)[2])
  # SUM OF ACTIVE SECONDS
  aggtest$active_seconds <- unlist(aggregate(testdata$active_seconds ~
    testdata$mmid, data = testdata, FUN = sum, na.rm = TRUE)[2])
  # START OF THE MOMENT (MIN VALUE OF USED AT)
  aggtest$start <- unlist(aggregate(testdata$used_at ~ testdata$mmid,
    data = testdata, function(x) min(x))[2])
  # END OF THE MOMENT (MAX VALUE OF SUM OF USED_AT AND ACTIVE
  # SECONDS OF THE LAST ACTIVITY BEFORE IDLE)
  aggtest$end <- unlist(aggregate(testdata$used_at ~ testdata$mmid,
    data = testdata, function(x) as.character(as.POSIXct(max(x) +
    testdata$active_seconds[length(x)]))[2])
  # CONVERT TO DATE AND GET DIFFERENCE BETWEEN END AND START OF
  # THE MOMENT
  aggtest$start <- strptime(aggtest$start, format = "%Y-%m-%d %T")
  aggtest$end <- strptime(aggtest$end, format = "%Y-%m-%d %T")
  aggtest$length <- difftime(strptime(aggtest$end, format = "%Y-%m-%d %T"),
    strptime(aggtest$start, format = "%Y-%m-%d %T"), units = c("secs"))
  aggtest$app <- NA
  aggtest$device <- "desktop"
  aggchunks[[forvar]] <- aggtest
  print(c("End", nrow(testdata), forvar, as.character(Sys.time())))
  rm(testdata, aggtest)
}
save(aggchunks, file = "aggchunks.Rdata")
##### MOBILE #####
mobilevisits1 <- testM
# REPLACE EMPTY WITH NA
mobilevisits1[mobilevisits1 == ""] = NA
# REMOVE IDS OCCURRING LESS THAN 5 TIMES IN THE DATA

```

```

mobilevisits1 <- mobilevisits1[mobilevisits1$panelist_id %in%
  names(table(mobilevisits1$panelist_id))[table(mobilevisits1$panelist_id) >=
    5], ]
# GET THE HOST NAME (DOMAIN + SUBDOMAIN)
mobilevisits1$host <- suffix_extract(domain(mobilevisits1$url))[1]$host
# CONVERT DURATUION TO NUMERIC AND REMOVE '/N'
mobilevisits1$duration <- as.numeric(mobilevisits1$duration)
mobilevisits1$used_at <- as.character(mobilevisits1$used_at)
mobilevisits1$duration <- ifelse(is.na(mobilevisits1$duration),
  0, mobilevisits1$duration)
mobilevisits1$used_at <- strptime(mobilevisits1$used_at, format = "%Y-%m-%d %T")
# GET USER IDS
uidsm <- unique(mobilevisits1$panelist_id)
# CREATE EMPTY LIST
chunksm <- list()
# FILL IN THE LIST WITH DATA PER EACH ID AND CREATE
# MICROMOMENTS
for (ivar in 1:length(uidsm)) {
  workarr <- subset(mobilevisits1, mobilevisits1$panelist_id ==
    uidsm[ivar])
  print(c("Start", nrow(workarr), ivar, as.character(Sys.time())))
  if (ivar == 83) {
    workarr <- head(workarr, 75000)
  }
  workarr <- workarr[order(workarr$used_at), ]
  mmid <- 1
  for (ivar2 in 1:(nrow(workarr) - 1)) {
    if (as.POSIXct(workarr$used_at[ivar2]) + workarr$duration[ivar2] +
      300 < as.POSIXct(workarr$used_at[ivar2 + 1])) {
      workarr$mmid[ivar2] <- mmid
      mmid <- mmid + 1
    } else {
      workarr$mmid[ivar2] <- mmid
    }
  }
  if (as.POSIXct(workarr$used_at[nrow(workarr) - 1]) + workarr$duration[nrow(workarr) -
    1] + 300 < as.POSIXct(workarr$used_at[nrow(workarr)])) {
    workarr$mmid[nrow(workarr)] <- mmid
  } else {
    workarr$mmid[nrow(workarr)] <- mmid - 1
  }
  chunksm[[ivar]] <- workarr
  rm(workarr)
  print(c("End", ivar, as.character(Sys.time())))
}
save(chunksm, file = "chunksm.Rdata")
##### Aggregate MOBILE #####
aggchunksm <- list()
for (forvar in 1:length(chunksm)) {
  print(c("Start", forvar, as.character(Sys.time())))
  testdata <- NULL
  aggttest <- NULL
  testdata <- chunksm[[forvar]]
  testdata$used_at <- as.character(testdata$used_at)
  aggttest <- aggregate(testdata$panelist_id ~ testdata$mmid,
    FUN = unique)
  colnames(aggttest) <- c("mmid", "panelist_id")
}

```



```

aggtest <- aggtest[, c(2, 1)]
aggtest$host <- unlist(aggregate(host ~ panelist_id + mmid,
  data = testdata, FUN = function(x) if (any(!is.na(x))) {
    length(unique(x[which(!is.na(x)]))
  } else {
    x = NA
  }, na.action = na.pass)[3])
# NUMBER OF PAGEVIEWS VISITS
aggtest$pageviews <- unlist(aggregate(url ~ panelist_id +
  mmid, data = testdata, FUN = function(x) if (any(!is.na(x))) {
    length(x[which(!is.na(x)]))
  } else {
    x = NA
  }, na.action = na.pass)[3])
# SUM OF ACTIVE SECONDS
aggtest$active_seconds <- unlist(aggregate(duration ~ panelist_id +
  mmid, data = testdata, FUN = sum, na.rm = TRUE)[3])
# START OF THE MOMENT (MIN VALUE OF USED AT)
aggtest$start <- unlist(aggregate(used_at ~ panelist_id +
  mmid, data = testdata, function(x) min(x))[3])
# END OF THE MOMENT (MAX VALUE OF USED_TILL I.E. SUM OF
# USED_AT AND ACTIVE SECONDS OF THE LAST ACTIVITY BEFORE
# IDLE)
aggtest$end <- unlist(aggregate(used_at ~ panelist_id + mmid,
  data = testdata, FUN = function(x) as.character(as.POSIXct(max(x) +
  testdata$duration[length(x)]))[3])
# CONVERT TO DATE AND GET DIFFERENCE BETWEEN END AND START OF
# THE MOMENT
aggtest$start <- strptime(aggtest$start, format = "%Y-%m-%d %T")
aggtest$end <- strptime(aggtest$end, format = "%Y-%m-%d %T")
aggtest$length <- difftime(strptime(aggtest$end, format = "%Y-%m-%d %T"),
  strptime(aggtest$start, format = "%Y-%m-%d %T"), units = c("secs"))
aggtest$appids <- unlist(aggregate(app_id ~ panelist_id + mmid,
  data = testdata, FUN = function(x) if (any(!is.na(x))) {
    length(unique(x[which(!is.na(x)]))
  } else {
    x = NA
  }, na.action = na.pass)[3])
aggtest$device <- "mobile"
aggchunksm[[forvar]] <- aggtest
rm(testdata, aggtest)
print(c("End", forvar, as.character(Sys.time())))
}
save(aggchunksm, file = "aggchunksm.Rdata")
# PRODUCT OF THE SCRIPT THE SCRIPT IS FED WITH THE RAW
# FULLDATASET THE RESULT OF THE SCRIPT IS A LIST WITH
# ELEMENTS CONTAINING THE INDIVIDUAL DATA PER EACH ID
# INCLUDING THE MICROMOMENTS

```

---

## CODE CHUNK #2A. TRAVEL WEBSITES KEYWORDS CATEGORIZATION

---

```

### APPLY DOMAIN VS. PATH
library(plyr)
##### DOMAINS FROM DEKSTOP AND MOBILE #####
# GET UNIQUE DOMAINS AND FREQ
domains_D <- count(df = testD, vars = "host")
domains_M <- count(df = testM, vars = "host")
# MERGE AND SUM FREQ
domains_full <- merge(domains_D, domains_M, by = "host", all = TRUE)
domains_full$freq <- rowSums(domains_full[c("freq.x", "freq.y")],
  na.rm = TRUE)
# FIND KEYWORDS AND ASSING CLASSIFY
# KEYWORDS
travelwebskeywords <- c("airbnb", "hotel", "hotels", "flight",
  "flights", "air", "booking", "expedia", "trivago", "travel",
  "tourism", "trip", "vacation", "vakant", "reis", "orbitz",
  "kayak", "weekendje", "vliegtickets", "vliegen", "villa")
for (web1 in travelwebskeywords) {
  domains_full[web1] <- ifelse(grepl(pattern = web1, x = domains_full$host,
    ignore.case = TRUE), 1, 0)
}
domains_full$book <- ifelse(grepl(pattern = "book\\.\"", x = domains_full$host,
  ignore.case = TRUE), 1, 0)
domains_full[grepl(pattern = "facebook", x = domains_full$host,
  ignore.case = TRUE), c("book")] <- 0
domains_full[grepl(pattern = "solitaire", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "hair", x = domains_full$host, ignore.case = TRUE),
  c("air")] <- 0
domains_full[grepl(pattern = "fair", x = domains_full$host, ignore.case = TRUE),
  c("air")] <- 0
domains_full[grepl(pattern = "millionaire", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "billionaire", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "questionnaire", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "airfry", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "repair", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "clair", x = domains_full$host, ignore.case = TRUE),
  c("air")] <- 0
domains_full[grepl(pattern = "nuclair", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "airmax", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "primaire", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "flair", x = domains_full$host, ignore.case = TRUE),
  c("air")] <- 0
domains_full[grepl(pattern = "miljonair", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "ipad", x = domains_full$host, ignore.case = TRUE),
  c("air")] <- 0
domains_full[grepl(pattern = "airbag", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "airco", x = domains_full$host, ignore.case = TRUE),

```

```

c("air")] <- 0
domains_full[grepl(pattern = "airfilter", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "puzzel", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "polair", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "aupair", x = domains_full$host,
  ignore.case = TRUE), c("air")] <- 0
domains_full[grepl(pattern = "kayako", x = domains_full$host,
  ignore.case = TRUE), c("kayak")] <- 0
# CREATE IS TRAVEL
domains_full$IsTravel <- apply(domains_full[5:ncol(domains_full)],
  1, FUN = function(x) if (any(x == 1)) {
    1
  } else {
    0
  })
##### APPS #####
Appused <- count(testM, "app_name")
for (app1 in travelwebskeywords) {
  Appused[app1] <- ifelse(grepl(pattern = app1, x = Appused$app_name,
    ignore.case = TRUE), 1, 0)
}
Appused[grepl(pattern = "solitaire", x = Appused$app_name, ignore.case = TRUE),
  c("air")] <- 0
Appused$TravelApp <- apply(Appused[travelwebskeywords], 1,
  FUN = function(x) if (any(x == 1)) {
    1
  } else {
    0
  })
# PRODUCT OF THE SCRIPT THE SCRIPT TAKES THE UNIQUE DOMAINS
# VISITED VIA DESKTOP AND MOBILE ALSO THE APPS THE RESULT OF
# THE SCRIPT IS A DATAFRAME WITH CATEGORIES PER EACH DOMAIN
# BASED ON KEYWORDS

```

---

## CODE CHUNK #2B. TRAVEL WEBSITES WEB SCRAPER + UCLASSIFY.COM API

---

```

### ABOVE 10 LABRARIES
library(httr)
library(jsonlite)
library(stringr)
library(Hmisc)
# AGG OVER HOST
testclassifytest1 <- aggregate(fullwebvisits$host, by = list(fullwebvisits$host),
  length)
# RENAME
colnames(testclassifytest1) <- c("url", "freq")
# testclassifytest2 <- subset(testclassifytest1, freq>10)
testclassify <- subset(testclassifytest1, freq > 10)

```

```

# GET COLNAMES IN VECTOR
cnames <- c("Arts and Entertainment", "Autos", "Businss Finance",
  "Celebrity", "College", "Cooking", "Dating and Romance",
  "Exercise", "Fashion and Beauty", "Games", "Health", "Home Improvement",
  "News", "Parents and Family", "Technology", "Travel")
# CREATE COLNAME FOR EACH NAME OF COLNAME VECTOR
for (i in cnames) {
  testclassify[, i] <- NA
}
for (item in 1:nrow(testclassify)) {
  # for (item in 1:50) {
  # TRY TO OPEN CONNECTION
  try(fullhtml <- GET(paste0("http://", testclassify$url[item])))
  if (exists("fullhtml")) {
    # SCRAPE THE TEXT
    contenthtml <- content(fullhtml, "text")
    try(onlytest <- htmlParse(contenthtml, asText = TRUE))
    # try ( plaintext <- xpathSApply(onlytest, '//p', xmlValue) )
    try(plaintext <- xpathSApply(onlytest, "//text()[not(ancestor::script)][not(ancestor::style)][not(ancestor::pre)]", xmlValue))
    text <- paste(plaintext, collapse = "")
    text <- str_replace_all(text, "[\r\n\t]", " ")
    text <- str_replace_all(text, "[[:punct:]]", "")
    text <- str_replace_all(text, " +", "+")
    # IF THERE IS TEXT
    if (text != "") {
      # ENCODE AS URL
      encodedurlstring <- URLEncode(text, reserved = FALSE,
        repeated = FALSE)
      encodedurlstring <- str_replace_all(encodedurlstring,
        "%2", "+")
      apiurlstring <- "https://api.uclassify.com/v1/ephraimbaro/Categories/classify/?readKey=70rn6c"
      fullurlstring <- paste0(apiurlstring, encodedurlstring,
        collapse = "")
      # SEND IT TO THE UCLASSIFY API
      try(test1 <- fromJSON(fullurlstring))
      # FILL THE DATA IN DATAFRAME
      testclassify[item, cnames] <- data.frame(matrix(unlist(test1),
        nrow = 1, byrow = T))
      print(item)
    }
    rm(fullhtml)
  }
}
# GET THE HIGHEST CATEGORY
testclassify$class <- unlist(apply(testclassify[5:20], 1, function(x) if (!is.na(x[1]) &&
  x[1] != 0.0625) {
  which(x == max(x))
} else {
  17
}))
cnames2 <- c(cnames, "NO category")
# GET CATEGORY NAME
testclassify$class1 <- unlist(lapply(testclassify$class, function(x) if (is.numeric(x)) {
  cnames2[as.numeric(x)]
}))
# REMOVE SCIENTIFIC NOTATION options(scipen=999)

```

```

# save(testclassify, file =
# 'testclassify_2907_until_48k.Rdata')
# PERCENTAGE VECTOR NAMES
percentagecols <- paste0("P", cnames)
# PERCENTAGE CREATE EMPTY COLUMNS
for (i in percentagecols) testclassify[, i] <- NA
# CREATE COLUMNS IN PERCENTAGES
for (ivar in 1:16) {
  shift = sum(ivar, 4)
  testclassify[percentagecols[ivar]] <- round(testclassify[shift]/rowSums(testclassify[5:20]),
  3)
}
describe(testclassify$class1)
class48k <- head(testclassify, 47818)
save(class48k, file = "class48k.Rdata")
# PRODUCT OF THE SCRIPT THE SCRIPT TAKES THE UNIQUE DOMAINS
# VISITED VIA DESKTOP AND MOBILE ALSO THE APPS IT VISITS THE
# DOMAIN AND SCRAPES DOWN THE INFORMATION STORES IT LOCALLY
# AND SENDS IT OUT TO UCLASSIFY.COM THE RESULT OF THE SCRIPT
# IS A DATAFRAME WITH CATEGORIES PER EACH DOMAIN BASED

```

---

## CODE CHUNK #3A. ADD PURCHASE DATA

```

#### MERGE PURCHASE INTO DESKTOP CHUNKS
# LOAD PURCHASE
library(lubridate)
purchase <- read.csv("gdselect-pageviews-20160101-20160731.csv",
  stringsAsFactors = FALSE)
for (ivar1 in 1:length(uids)) {
  cid <- uids[ivar1]
  workarr <- chunks[[ivar1]]
  workarr$purchase <- NULL
  purchworkarr <- subset(purchase, purchase$ID == cid)
  print(c("start", nrow(workarr), nrow(purchworkarr), ivar1,
  as.character(Sys.time())))
  if (nrow(purchworkarr) >= 1) {
    purchworkarr <- purchworkarr[c("url", "used_at")]
    purchworkarr$used_at <- as.numeric(ymd_hms(purchworkarr$used_at))
    purchworkarr <- dply(purchworkarr, "url", numcolwise(min))
    purchworkarr$url <- NULL
    purchworkarr$purchase <- 1
    purchworkarr$used_at <- as.character(as.POSIXct(purchworkarr$used_at,
    origin = "1970-01-01 00:00.00"))
    workarr <- merge(workarr, purchworkarr, by.x = "used_at",
    by.y = "used_at", all.x = TRUE)
    chunks[[ivar1]] <- workarr
  }
  print(c("end", nrow(workarr), nrow(purchworkarr), ivar1,
  as.character(Sys.time())))
  rm(workarr)
}
save(chunks, file = "chunks.Rdata")

```

```

#### MERGE PURCHASE INTO MOBILE CHUNKS
# LOAD PURCHASE
purchasem <- read.csv("gdselect-mobile-pageviews-20160101-20160731.csv",
  stringsAsFactors = FALSE)
for (ivar1 in 1:length(uidsm)) {
  cid <- uidsm[ivar1]
  workarr <- chunksm[[ivar1]]
  workarr$purchase <- NULL
  purchworkarr <- subset(purchasem, purchasem$ID == cid)
  print(c("start", nrow(workarr), nrow(purchworkarr), ivar1,
    as.character(Sys.time())))
  if (nrow(purchworkarr) >= 1) {
    purchworkarr <- purchworkarr[c("url", "used_at")]
    purchworkarr$used_at <- as.numeric(ymd_hms(purchworkarr$used_at))
    purchworkarr <- ddply(purchworkarr, "url", numcolwise(min))
    purchworkarr$url <- NULL
    purchworkarr$purchase <- 1
    purchworkarr$used_at <- as.character(as.POSIXct(purchworkarr$used_at,
      origin = "1970-01-01 00:00.00"))
    workarr$used_at <- as.character(workarr$used_at)
    workarr <- merge(workarr, purchworkarr, by.x = "used_at",
      by.y = "used_at", all.x = TRUE)
    chunksm[[ivar1]] <- workarr
  }
  print(c("end", nrow(workarr), nrow(purchworkarr), ivar1,
    as.character(Sys.time())))
  rm(workarr)
}

```

---

## CODE CHUNK #3B. ADD TRAVEL CATEGORIZATION

```

#### MERGE DOMAINS FROM SCRAPER AND KEYWORDS
domains_full_classified <- merge(domains_full, class48k, by.x = "host",
  by.y = "url", all = TRUE)
domains_full_classified$class_full <- ifelse(is.na(domains_full_classified$class),
  0, domains_full_classified$class)
domains_full_classified$Class_Travel <- ifelse(domains_full_classified$class_full ==
  16, 1, 0)
domains_full_classified$Class_Travel <- ifelse(domains_full_classified$IsTravel ==
  1, 1, domains_full_classified$Class_Travel)
domain_travel <- domains_full_classified[c("host", "freq", "Class_Travel")]
# FINAL LIST OF DOMAINS
save(domain_travel, file = "final_domain_list.Rdata")
#### MERGE TRAVEL INTO DESKTOP CHUNKS
# MERGING TRAVEL DOMAINS WITH THE CHUNKS
for (ivar in 1:length(chunks)) {
  # GET FROM THE LIST
  workarr <- chunks[[ivar]]
  print(c("start", nrow(workarr), ivar, as.character(Sys.time())))
  # MERGE
  workarr <- merge(workarr, domain_travel[c("host", "Class_Travel")],
    by.x = "host", by.y = "host", all.x = TRUE)
}

```

```

# BACK TO THE LIST
chunks[[ivar]] <- workarr
print(c("end", nrow(workarr), ivar, as.character(Sys.time())))
rm(workarr)
}
save(chunks, file = "chunks.Rdata")
for (ivar in 1:length(chunks)) {
  # GET FROM THE LIST
  workarr <- chunks[[ivar]]
  aggarr <- aggchunks[[ivar]]
  print(c("start", nrow(workarr), nrow(aggarr), ivar, as.character(Sys.time())))
  # WAS TRAVEL MOMENT
  aggarr$travel <- unlist(aggregate(Class_Travel ~ mmid, data = workarr,
    FUN = function(x) if (sum(x, na.rm = TRUE) > 0) {
      1
    } else {
      0
    }, na.action = na.pass)[2])
  # Travel Domains
  aggarr$traveldomain <- unlist(merge(aggarr, subset(aggregate(host ~
    Class_Travel + mmid, data = workarr, FUN = function(x) length(unique(x))),
    Class_Travel == 1)[c("mmid", "host")], by = "mmid", all.x = TRUE)[c("host")])
  # NUMBER OF TRAVEL PAGEVIEWS
  aggarr$travelPV <- unlist(aggregate(Class_Travel ~ mmid,
    data = workarr, FUN = function(x) if (any(!is.na(x))) {
      length(which(x == 1))
    } else {
      0
    }, na.action = na.pass)[2])
  # BACK TO THE LIST
  aggchunks[[ivar]] <- aggarr
  print(c("end", nrow(workarr), nrow(aggarr), ivar, as.character(Sys.time())))
  rm(workarr)
  rm(aggarr)
}
save(aggchunks, file = "aggchunks.Rdata")
#### MERGE TRAVEL INTO MOBILE CHUNKS
# MERGING TRAVEL DOMAINS AND APPS WITH THE CHUNKS
for (ivar in 1:length(chunksm)) {
  # GET FROM THE LIST
  workarr <- chunksm[[ivar]]
  print(c("start", nrow(workarr), ivar, as.character(Sys.time())))
  # MERGE
  workarr <- merge(workarr, domain_travel[c("host", "Class_Travel")],
    by.x = "host", by.y = "host", all.x = TRUE)
  workarr <- merge(workarr, Appused[c("app_name", "TravelApp")],
    by = "app_name", all.x = TRUE)
  # BACK TO THE LIST
  chunksm[[ivar]] <- workarr
  print(c("end", nrow(workarr), ivar, as.character(Sys.time())))
  rm(workarr)
}
save(chunksm, file = "chunksm.Rdata")
#
for (ivar in 1:length(chunksm)) {
  print(c("start", ivar, as.character(Sys.time())))
  # GET FROM THE LIST

```



```

workarr <- chunksm[[ivar]]
aggarr <- aggchunksm[[ivar]]
# WAS TRAVEL MOMENT aggarr$travel <- unlist( aggregate(travel
# ~ mmid, data=workarr, FUN = function(x) if(sum(x, na.rm =
# TRUE)>0){1}else{0}, na.action = na.pass)[2] ) Travel
# Domains aggarr$traveldomain <- unlist(merge(aggarr,
# subset(aggregate( host ~ travel + mmid, data=workarr, FUN =
# function(x) length(unique(x)) ),
# travel==1)[c('mmid','host')], by = 'mmid', all.x =
# TRUE)[c('host')]) NUMBER OF TRAVEL PAGEVIEWS
# aggarr$travelPV <- unlist( aggregate(travel ~ mmid,
# data=workarr, FUN =
# function(x)if(any(!is.na(x))){length(which(x==1))}else{0},
# na.action = na.pass)[2] )
# WAS TRAVEL WEB MOMENT
aggarr$travel_mm_w <- unlist(aggregate(Class_Travel ~ mmid,
  data = workarr, FUN = function(x) if (any(x[which(!is.na(x))] >
    0)) {
    1
  } else {
    0
  }, na.action = na.pass)[2])
aggarr$travel_mm_a <- unlist(aggregate(TravelApp ~ mmid,
  data = workarr, FUN = function(x) if (any(x[which(!is.na(x))] >
    0)) {
    1
  } else {
    0
  }, na.action = na.pass)[2])
aggarr$travel <- ifelse(rowSums(aggarr[c("travel_mm_w", "travel_mm_a")]) >
  0, 1, 0)
# NUMBER OF TRAVEL PAGEVIEWS aggarr$travelPV <- unlist(
# aggregate(Class_Travel ~ mmid, data=workarr, FUN =
# function(x)if(any(!is.na(x))){length(which(x==1))}else{0},
# na.action = na.pass)[2] ) WAS TRAVEL APP MOMENT
# aggarr$travelapp <- unlist( aggregate(TravelApp ~ mmid,
# data=workarr, FUN = function(x)if(sum(x, na.rm =
# TRUE)>0){1}else{0}, na.action = na.pass)[2] ) Travel
# Domains DIRTY PATCH BUT FUCK THOSE PPL
if (ivar != 52 & ivar != 75 & ivar != 76 & ivar != 80 & ivar !=
  94) {
  aggarr$traveldomain <- unlist(merge(aggarr, subset(aggregate(host ~
    Class_Travel + mmid, data = workarr, FUN = function(x) length(unique(x))),
    Class_Travel == 1)[c("mmid", "host")], by = "mmid",
    all.x = TRUE)[c("host.y")])
} else {
  aggarr$traveldomain <- NA
}
# NUMBER OF TRAVEL PAGEVIEWS
aggarr$travelPV <- unlist(aggregate(Class_Travel ~ mmid,
  data = workarr, FUN = function(x) if (any(!is.na(x))) {
    length(which(x == 1))
  } else {
    0
  }, na.action = na.pass)[2])
# NUMBER OF TRAVEL APPS
aggarr$travel_app <- unlist(merge(aggarr, subset(aggregate(app_name ~

```



```

    TravelApp + mmid, data = workarr, FUN = function(x) length(unique(x))),
    TravelApp == 1)[c("mmid", "app_name")], by = "mmid",
    all.x = TRUE)[c("app_name")]
  # BACK TO THE LIST
  aggchunksm[[ivar]] <- aggarr
  print(c("end", ivar, as.character(Sys.time())))
  rm(aggarr, workarr)
}
save(aggchunksm, file = "aggchunksm.Rdata")

```

---

## CODE CHUNK #4. AGGREGATE DATA

---

```

##### AGGREGATE DESKTOP #####
# FAUTY ONES
aggchunks[[431]]$end[which(is.na(aggchunks[[431]]$end))] <- as.POSIXct("2016-03-08 23:59:20")
aggchunks[[325]]$end[which(is.na(aggchunks[[325]]$end))] <- as.POSIXct("2016-01-13 23:58:14")
aggchunks[[224]]$end[which(is.na(aggchunks[[224]]$end))] <- as.POSIXct("2016-02-18 23:56:08")
aggchunks[[431]]$length[which(is.na(aggchunks[[431]]$length))] <- difftime(strptime(aggchunks[[431]]$end[520],
  format = "%Y-%m-%d %T"), strptime(aggchunks[[431]]$start[526],
  format = "%Y-%m-%d %T"), units = c("secs"))
aggchunks[[325]]$length[which(is.na(aggchunks[[325]]$length))] <- difftime(strptime(aggchunks[[325]]$end[910],
  format = "%Y-%m-%d %T"), strptime(aggchunks[[325]]$start[91],
  format = "%Y-%m-%d %T"), units = c("secs"))
aggchunks[[224]]$length[which(is.na(aggchunks[[224]]$length))] <- difftime(strptime(aggchunks[[224]]$end[510],
  format = "%Y-%m-%d %T"), strptime(aggchunks[[224]]$start[516],
  format = "%Y-%m-%d %T"), units = c("secs"))
aggall <- data.frame()
for (ivar in 1:length(uids)) {
  uid <- uids[ivar]
  aggarr <- aggchunks[[ivar]]
  aggarr$id <- uid
  aggall <- rbind(aggall, aggarr)
}
fullagginfo <- aggregate(url ~ panelist_id, data = testD, FUN = length)
colnames(fullagginfo) <- c("ID", "obs")
# DOMAINS
fullagginfo$domains <- unlist(aggregate(host ~ panelist_id, data = testD,
  FUN = function(x) length(unique(x)))[2])
# PAGEVIEWS fullagginfo$domains_PV <- unlist(aggregate(url ~
# panelist_id, data = testD, FUN = function(x)
# length(unique(x)))[2])
fullagginfo$domains_PV <- unlist(aggregate(url ~ panelist_id,
  data = testD, FUN = function(x) length(x))[2])
# MICROMOMENTS
fullagginfo$mm <- unlist(aggregate(mmid ~ id, data = aggall,
  FUN = function(x) length(x))[2])
# TRAVEL DOMAINS
fullagginfo$travel_domains <- unlist(aggregate(traveldomain ~
  id, data = aggall, FUN = function(x) if (any(!is.na(x))) {
  sum(x[which(!is.na(x))])
} else {
  0

```

```

}, na.action = na.pass)[2])
# TRAVEL PAGEVIEWS
fullagginfo$travel_domains_PV <- unlist(aggregate(travelPV ~
  id, data = aggall, FUN = function(x) if (any(!is.na(x))) {
    sum(x[which(!is.na(x))])
  } else {
    0
  }, na.action = na.pass)[2])
# TRAVEL MICORMOMENTS
fullagginfo$travel_mm <- unlist(aggregate(travel ~ id, data = aggall,
  FUN = function(x) sum(x))[2])
for (ivar in 1:length(uids)) {
  print(c(ivar, as.character(Sys.time())))
  fullagginfo$total_time[ivar] <- unlist(aggregate(Class_Travel ~
    panelist_id, data = chunks[[ivar]], FUN = function(x) sum(chunks[[ivar]]$active_seconds))[2])
  fullagginfo$total_time_nottravel[ivar] <- unlist(aggregate(Class_Travel ~
    panelist_id, data = chunks[[ivar]], FUN = function(x) sum(chunks[[ivar]]$active_seconds[which(x ==
    0)]))[2])
  fullagginfo$total_time_travel[ivar] <- unlist(aggregate(Class_Travel ~
    panelist_id, data = chunks[[ivar]], FUN = function(x) sum(chunks[[ivar]]$active_seconds[which(x !=
    0)]))[2])
  fullagginfo$total_time2[[ivar]] <- as.numeric(sum(aggchunks[[ivar]]$length))
  fullagginfo$total_time_nottravel2[ivar] <- as.numeric(sum(aggchunks[[ivar]]$length[which(aggchunks[[ivar]]
    0)]))
  fullagginfo$total_time_travel2[ivar] <- as.numeric(sum(aggchunks[[ivar]]$length[which(aggchunks[[ivar]]
    1)]))
}
for (ivar in 1:length(uids)) {
  print(c(ivar, as.character(Sys.time())))
  fullagginfo$f_date[ivar] <- as.character(min(aggchunks[[ivar]]$start))
  fullagginfo$l_date[ivar] <- as.character(max(aggchunks[[ivar]]$end))
  fullagginfo$purchase[ivar] <- sum(chunks[[ivar]]$purchase,
    na.rm = TRUE)
}
fullagginfo$days_act <- round(difftime(fullagginfo$l_date, fullagginfo$f_date,
  units = "d"))
fullagginfo$log_travel_mm <- log(fullagginfo$travel_mm)
fullagginfo$log_travel_domains <- log(fullagginfo$travel_domains)
fullagginfo$log_travel_domains_PV <- log(fullagginfo$travel_domains_PV)
fullagginfo$log_total_time_travel <- log(fullagginfo$total_time_travel)
fullagginfo$log_purchase <- log(fullagginfo$purchase)
fullagginfo$d_purchase <- ifelse(fullagginfo$purchase < mean(fullagginfo$purchase),
  0, 1)
fullagginfo$log_mm <- log(fullagginfo$mm)
fullagginfo$log_domains <- log(fullagginfo$domains)
fullagginfo$log_domains_PV <- log(fullagginfo$domains_PV)
fullagginfo$log_time <- log(fullagginfo$total_time)
fullagginfo$log_time2 <- log(fullagginfo$total_time2)
fullagginfo$log_total_time_travel2 <- log(fullagginfo$total_time_travel2)
fullagginfo$log_purchase <- ifelse(fullagginfo$log_purchase ==
  "-Inf", 0, fullagginfo$log_purchase)
fullagginfo$device <- 1
fullagginfo$act_mm <- ifelse(fullagginfo$mm < mean(fullagginfo$mm),
  0, 1)
fullagginfo$act_total_time <- ifelse(fullagginfo$total_time <
  mean(fullagginfo$total_time), 0, 1)
fullagginfo$act_total_time2 <- ifelse(fullagginfo$total_time2 <

```

```

    mean(fullagginfo$total_time2, na.rm = TRUE), 0, 1)
fullagginfo$act_domains <- ifelse(fullagginfo$domains < mean(fullagginfo$domains),
    0, 1)
fullagginfo$act_domains_PV <- ifelse(fullagginfo$domains_PV <
    mean(fullagginfo$domains_PV), 0, 1)
fullagginfo$share_mm <- (fullagginfo$travel_mm/as.numeric(fullagginfo$days_act))/(fullagginfo$mm/as.numeric
fullagginfo$share_total_time <- (fullagginfo$total_time_travel/as.numeric(fullagginfo$days_act))/(fullagginfo
fullagginfo$share_total_time2 <- (fullagginfo$total_time_travel2/as.numeric(fullagginfo$days_act))/(fullagginfo
fullagginfo$share_domains <- (fullagginfo$travel_domains/as.numeric(fullagginfo$days_act))/(fullagginfo$doma
fullagginfo$share_domains_PV <- (fullagginfo$travel_domains_PV/as.numeric(fullagginfo$days_act))/(fullagginfo
save(fullagginfo, file = "fullagginfo.Rdata")
##### AGGREGATE MOBILE #####
aggallm <- data.frame()
for (ivar in 1:length(uidsm)) {
  uid <- uidsm[ivar]
  aggarr <- aggchunksm[[ivar]]
  aggarr$id <- uid
  aggallm <- rbind(aggallm, aggarr)
}
fullagginfom <- aggregate(url ~ panelist_id, data = mobilevisits1,
  FUN = length, na.action = na.pass)
colnames(fullagginfom) <- c("ID", "obs")
# DOMAINS
fullagginfom$domains <- unlist(aggregate(host ~ panelist_id,
  data = mobilevisits1, FUN = function(x) length(unique(x)),
  na.action = na.pass)[2])
# PAGEVIEWS fullagginfom$domains_PV <- unlist(aggregate(url
# ~ panelist_id, data = mobilevisits1, FUN = function(x)
# length(unique(x)) , na.action = na.pass)[2])
fullagginfom$domains_PV <- unlist(aggregate(url ~ panelist_id,
  data = mobilevisits1, FUN = function(x) length(x), na.action = na.pass)[2])
# MICROMOMENTS
fullagginfom$mm <- unlist(aggregate(mmid ~ id, data = aggallm,
  FUN = function(x) length(x), na.action = na.pass)[2])
# TRAVEL DOMAINS
fullagginfom$travel_domains <- unlist(aggregate(travel ~ id,
  data = aggallm, FUN = function(x) if (any(!is.na(x))) {
    sum(x[which(!is.na(x))])
  } else {
    0
  }, na.action = na.pass)[2])
# TRAVEL PAGEVIEWS
fullagginfom$travel_domains_PV <- unlist(aggregate(travelPV ~
  id, data = aggallm, FUN = function(x) if (any(!is.na(x))) {
    sum(x[which(!is.na(x))])
  } else {
    0
  }, na.action = na.pass)[2])
# TRAVEL MICROMOMENTS
fullagginfom$travel_mm_web <- unlist(aggregate(travel_mm_w ~
  id, data = aggallm, FUN = function(x) if (any(!is.na(x))) {
    sum(x[which(!is.na(x))])
  } else {
    0
  }, na.action = na.pass)[2])
fullagginfom$travel_mm_app <- unlist(aggregate(travel_mm_a ~
  id, data = aggallm, FUN = function(x) if (any(!is.na(x))) {

```

```

    sum(x[which(!is.na(x))])
} else {
  0
}, na.action = na.pass)[2])
fullagginfom$travel_mm <- rowSums(fullagginfom[c("travel_mm_web",
"travel_mm_app")], na.rm = TRUE)
fullagginfom <- fullagginfom[match(uidsm, fullagginfom$ID), ]
for (ivar in 1:length(uidsm)) {
  print(c(ivar, as.character(Sys.time())))
  fullagginfom$total_time[[ivar]] <- as.numeric(sum(aggchunksm[[ivar]]$active_seconds))
  fullagginfom$total_time_nottravel[ivar] <- as.numeric(sum(aggchunksm[[ivar]]$active_seconds[which(aggchunksm[[ivar]]$active_seconds != 0)]))
  fullagginfom$total_time_travel[ivar] <- as.numeric(sum(aggchunksm[[ivar]]$active_seconds[which(aggchunksm[[ivar]]$active_seconds == 1)]))
  fullagginfom$total_time2[[ivar]] <- as.numeric(sum(aggchunksm[[ivar]]$length))
  fullagginfom$total_time_nottravel2[ivar] <- as.numeric(sum(aggchunksm[[ivar]]$length[which(aggchunksm[[ivar]]$length != 0)]))
  fullagginfom$total_time_travel2[ivar] <- as.numeric(sum(aggchunksm[[ivar]]$length[which(aggchunksm[[ivar]]$length == 1)]))
}
# aggregate(travel ~ panelist_id, data = aggchunksm[[1]], FUN
# = function(x) sum(
# as.numeric(aggchunksm[[1]]$active_seconds) ) )
# aggregate(travel ~ panelist_id, data = aggchunksm[[1]], FUN
# = function(x) sum( as.numeric(aggchunksm[[1]]$length) ) )
# fullagginfom$total_time_travel <- rowSums(
# fullagginfom[c('total_time_travel_web',
# 'total_time_travel_app')], na.rm = TRUE)
fullagginfom$app <- unlist(aggregate(app_name ~ panelist_id,
data = mobilevisits1, FUN = function(x) length(unique(x)),
na.action = na.pass)[2])
for (ivar in 1:length(uidsm)) {
  print(c(ivar, as.character(Sys.time())))
  fullagginfom$f_date[ivar] <- as.character(min(aggchunksm[[ivar]]$start))
  fullagginfom$l_date[ivar] <- as.character(max(aggchunksm[[ivar]]$end))
  fullagginfom$purchase[ivar] <- sum(chunksm[[ivar]]$purchase,
na.rm = TRUE)
}
fullagginfom$days_act <- round(difftime(fullagginfom$l_date,
fullagginfom$f_date, units = "d"))
fullagginfom$log_travel_mm <- log(fullagginfom$travel_mm)
fullagginfom$log_travel_domains <- log(fullagginfom$travel_domains)
fullagginfom$log_travel_domains_PV <- log(fullagginfom$travel_domains_PV)
fullagginfom$log_total_time_travel <- log(fullagginfom$total_time_travel)
fullagginfom$log_travel_mm <- ifelse(fullagginfom$log_travel_mm ==
"-Inf", 0, fullagginfom$log_travel_mm)
fullagginfom$log_travel_domains <- ifelse(fullagginfom$log_travel_domains ==
"-Inf", 0, fullagginfom$log_travel_domains)
fullagginfom$log_travel_domains_PV <- ifelse(fullagginfom$log_travel_domains_PV ==
"-Inf", 0, fullagginfom$log_travel_domains_PV)
fullagginfom$log_total_time_travel <- ifelse(fullagginfom$log_total_time_travel ==
"-Inf", 0, fullagginfom$log_total_time_travel)
fullagginfom$log_mm <- log(fullagginfom$m)
fullagginfom$log_domains <- log(fullagginfom$domains)
fullagginfom$log_domains_PV <- log(fullagginfom$domains_PV)
fullagginfom$log_time <- log(fullagginfom$total_time)
fullagginfom$log_time2 <- log(fullagginfom$total_time2)

```

```

fullagginfom$log_total_time_travel2 <- log(fullagginfom$total_time_travel2)
fullagginfom$log_mm <- ifelse(fullagginfom$log_mm == "-Inf",
  0, fullagginfom$log_mm)
fullagginfom$log_domains <- ifelse(fullagginfom$log_domains ==
  "-Inf", 0, fullagginfom$log_domains)
fullagginfom$log_domains_PV <- ifelse(fullagginfom$log_domains_PV ==
  "-Inf", 0, fullagginfom$log_domains_PV)
fullagginfom$log_time <- ifelse(fullagginfom$log_time == "-Inf",
  0, fullagginfom$log_time)
fullagginfom$log_total_time_travel2 <- ifelse(fullagginfom$log_total_time_travel2 ==
  "-Inf", 0, fullagginfom$log_total_time_travel2)
fullagginfom$device <- 2
fullagginfom$act_mm <- ifelse(fullagginfom$mm < mean(fullagginfom$mm),
  0, 1)
fullagginfom$act_total_time <- ifelse(fullagginfom$total_time <
  mean(fullagginfom$total_time), 0, 1)
fullagginfom$act_total_time2 <- ifelse(fullagginfom$total_time2 <
  mean(fullagginfom$total_time2, na.rm = TRUE), 0, 1)
fullagginfom$act_domains <- ifelse(fullagginfom$domains < mean(fullagginfom$domains),
  0, 1)
fullagginfom$act_domains_PV <- ifelse(fullagginfom$domains_PV <
  mean(fullagginfom$domains_PV), 0, 1)
save(fullagginfom, file = "fullagginfom.Rdata")
##### AGGREGATE DESKTOP AND MOBILE ####
commonvarnames <- colnames(fullagginfo)[c(1:9, 11:12, 14, 17:18,
  30)]
library(plyr)
# fullagginfo_all <- rbind(fullagginfo[c(
# colnames(fullagginfo)[c(1:8,11,19,20,9,12,14)] )],
# fullagginfom[c(
# colnames(fullagginfom)[c(1:7,10,16,26,20,11,26,28)] )] )
fullagginfo_all <- rbind(fullagginfo[commonvarnames], fullagginfom[commonvarnames])
fullagginfo_all <- ddply(fullagginfo_all, "ID", numcolwise(sum))
fullagginfo_all_datearr <- merge(aggregate(f_date ~ ID, data = rbind(fullagginfo[c(colnames(fullagginfo)[c(
  15:16)])), fullagginfom[c(colnames(fullagginfom)[c(1, 18:19)])]),
  min), aggregate(l_date ~ ID, data = rbind(fullagginfo[c(colnames(fullagginfo)[c(1,
  15:16)])), fullagginfom[c(colnames(fullagginfom)[c(1, 18:19)])]),
  max), by = "ID", all = TRUE)
fullagginfo_all <- merge(fullagginfo_all, fullagginfo_all_datearr,
  by = "ID", all = TRUE)
fullagginfo_all$days_act <- round(difftime(fullagginfo_all$l_date,
  fullagginfo_all$f_date, units = "d"))
fullagginfo_all$log_travel_mm <- log(fullagginfo_all$travel_mm)
fullagginfo_all$log_travel_domains <- log(fullagginfo_all$travel_domains)
fullagginfo_all$log_travel_domains_PV <- log(fullagginfo_all$travel_domains_PV)
fullagginfo_all$log_total_time_travel <- log(fullagginfo_all$total_time_travel)
fullagginfo_all$log_total_time_travel2 <- log(fullagginfo_all$total_time_travel2)
fullagginfo_all$log_purchase <- log(fullagginfo_all$purchase)
fullagginfo_all$d_purchase <- ifelse(fullagginfo_all$purchase <
  mean(fullagginfo_all$purchase), 0, 1)
fullagginfo_all$log_mm <- log(fullagginfo_all$mm)
fullagginfo_all$log_domains <- log(fullagginfo_all$domains)
fullagginfo_all$log_domains_PV <- log(fullagginfo_all$domains_PV)
fullagginfo_all$log_time <- log(fullagginfo_all$total_time)
fullagginfo_all$log_time2 <- log(fullagginfo_all$total_time2)
fullagginfo_all$log_travel_mm <- ifelse(fullagginfo_all$log_travel_mm ==
  "-Inf", 0, fullagginfo_all$log_travel_mm)

```

```

fullagginfo_all$log_travel_domains <- ifelse(fullagginfo_all$log_travel_domains ==
  "-Inf", 0, fullagginfo_all$log_travel_domains)
fullagginfo_all$log_travel_domains_PV <- ifelse(fullagginfo_all$log_travel_domains_PV ==
  "-Inf", 0, fullagginfo_all$log_travel_domains_PV)
fullagginfo_all$log_total_time_travel <- ifelse(fullagginfo_all$log_total_time_travel ==
  "-Inf", 0, fullagginfo_all$log_total_time_travel)
fullagginfo_all$log_total_time_travel2 <- ifelse(fullagginfo_all$log_total_time_travel2 ==
  "-Inf", 0, fullagginfo_all$log_total_time_travel2)
fullagginfo_all$log_purchase <- ifelse(fullagginfo_all$log_purchase ==
  "-Inf", 0, fullagginfo_all$log_purchase)
fullagginfo_all$log_mm <- ifelse(fullagginfo_all$log_mm == "-Inf",
  0, fullagginfo_all$log_mm)
fullagginfo_all$log_domains <- ifelse(fullagginfo_all$log_domains ==
  "-Inf", 0, fullagginfo_all$log_domains)
fullagginfo_all$log_domains_PV <- ifelse(fullagginfo_all$log_domains_PV ==
  "-Inf", 0, fullagginfo_all$log_domains_PV)
fullagginfo_all$log_time <- ifelse(fullagginfo_all$log_time ==
  "-Inf", 0, fullagginfo_all$log_time)
fullagginfo_all$log_time2 <- ifelse(fullagginfo_all$log_time ==
  "-Inf", 0, fullagginfo_all$log_time2)
fullagginfo_all$act_mm <- ifelse(fullagginfo_all$mm < mean(fullagginfo_all$mm),
  0, 1)
fullagginfo_all$act_total_time <- ifelse(fullagginfo_all$total_time <
  mean(fullagginfo_all$total_time), 0, 1)
fullagginfo_all$act_total_time2 <- ifelse(fullagginfo_all$total_time2 <
  mean(fullagginfo_all$total_time2, na.rm = TRUE), 0, 1)
fullagginfo_all$act_domains <- ifelse(fullagginfo_all$domains <
  mean(fullagginfo_all$domains), 0, 1)
fullagginfo_all$act_domains_PV <- ifelse(fullagginfo_all$domains_PV <
  mean(fullagginfo_all$domains_PV), 0, 1)

```

---

## CODE CHUNK #5. SURVEY DATA PROCESSING

---

```

rawsurveycompletes1$Q7new <- ifelse(rawsurveycompletes1$Q7 >
  3, 1, 0)
rawsurveycompletes1$Q9new <- ifelse(rawsurveycompletes1$Q8 ==
  2, 0, rawsurveycompletes1$Q9)
rawsurveycompletes1$dRA <- ifelse(rawsurveycompletes1$RA == "seeking",
  1, 0)
rawsurveycompletes1$dUA <- ifelse(rawsurveycompletes1$UA == "seeking",
  1, 0)
which(rawsurveycompletes1$dRA == 1) == which(rawsurveycompletes1$RA ==
  "averse")
which(rawsurveycompletes1$dRA == 0) == which(rawsurveycompletes1$RA ==
  "seeking")
which(rawsurveycompletes1$dUA == 1) == which(rawsurveycompletes1$UA ==
  "averse")
which(rawsurveycompletes1$dUA == 0) == which(rawsurveycompletes1$UA ==
  "seeking")
rawsurveycompletes1$RAUA <- ifelse(rawsurveycompletes1$RA ==
  "averse" & rawsurveycompletes1$UA == "averse", "RAUA", ifelse(rawsurveycompletes1$RA ==
  "averse" & rawsurveycompletes1$UA == "seeking", "RAUS", ifelse(rawsurveycompletes1$RA ==

```



```

"seeking" & rawsurveycompletes1$UA == "averse", "RSUA", ifelse(rawsurveycompletes1$RA ==
"seeking" & rawsurveycompletes1$UA == "seeking", "RSUS",
"NA"))))
rawsurveycompletes1$D3new <- ifelse(as.numeric(rawsurveycompletes1$D3) <
5, 0, 1)
rawsurveycompletes1$Q18new <- ifelse(rawsurveycompletes1$Q18.1.value !=
"{null}", rawsurveycompletes1$Q18.1.value, round(mean(as.numeric(rawsurveycompletes1$Q18.1.value),
na.rm = TRUE), digits = 0))
rawsurveycompletes1$Q18new[242] <- 400
rawsurveycompletes1$Q18new[396] <- 600
rawsurveycompletes1$Q18new[402] <- 400
which(is.numeric(as.numeric(rawsurveycompletes1$Q18.1.value)))
summary(as.factor(rawsurveycompletes1$Q18Anew2))
rawsurveycompletes1$Q18Anew2 <- ifelse(as.numeric(rawsurveycompletes1$Q18new) >
mean(as.numeric(rawsurveycompletes1$Q18new)), 1, 0)
##### FACTOR analysis #####
RAUAdata <- rawsurveycompletes1[, 58:64]
RAUAdata <- as.data.frame(lapply(RAUAdata, as.numeric))
RAUAFactors <- factanal(RAUAdata, factors = 2, rotation = "varimax")
print(RAUAFactors, digits = 2, cutoff = 0.4, sort = FALSE)
BIG5data <- fullagginfo_sdata[c(paste0("Q15_", 1:11))]
BIG5data <- as.data.frame(lapply(BIG5data, as.numeric))
BIG5factors <- factanal(BIG5data, factors = 5, rotation = "varimax")
print(BIG5factors, digits = 2, cutoff = 0.4, sort = FALSE)

```

---

## CODE CHUNK #6. MERGE SURVEY DATA AND PROCESSED BEHAVIORAL DATA

---

```

library(stargazer)
library(lmtest)
##### LOAD SURVEY DATA #####
load(file = "~/trailspot-channelexport/final_files/final_survey_completes.Rdata")
##### DESKTOP #####
fullagginfo_sdata <- merge(fullagginfo, rawsurveycompletes1,
by.x = "ID", by.y = "ID")
fullagginfo_sdata <- subset(fullagginfo_sdata, obs > 1000 & QC ==
0)
fullagginfo_sdata <- fullagginfo_sdata[-which(duplicated(fullagginfo_sdata$ID)),
]
finalRA <- aggregate(fullagginfo_sdata[2], by = list(RA = fullagginfo_sdata$RA),
data = fullagginfo_sdata, FUN = mean)
finalUA <- aggregate(fullagginfo_sdata[2], by = list(UA = fullagginfo_sdata$UA),
data = fullagginfo_sdata, FUN = mean)
finalRAUA <- aggregate(fullagginfo_sdata[2], by = list(RA = fullagginfo_sdata$RA,
UA = fullagginfo_sdata$UA), data = fullagginfo_sdata, FUN = mean)
for (ivar in 2:11) {
finalRA[ivar] <- aggregate(fullagginfo_sdata[ivar], by = list(RA = fullagginfo_sdata$RA),
data = fullagginfo_sdata, FUN = mean, na.rm = TRUE)[2]
finalUA[ivar] <- aggregate(fullagginfo_sdata[ivar], by = list(UA = fullagginfo_sdata$UA),
data = fullagginfo_sdata, FUN = mean, na.rm = TRUE)[2]
finalRAUA[ivar + 1] <- aggregate(fullagginfo_sdata[ivar],
by = list(RA = fullagginfo_sdata$RA, UA = fullagginfo_sdata$UA),

```

```

    data = fullagginfo_sdata, FUN = mean, na.rm = TRUE)[3]
}
mean(fullagginfo_sdata$days_act)
fullagginfo_sdata$l_date[209]
fullagginfo_sdata$f_date[209]
fullagginfo_sdata$days_act[209]
##### MOBILE #####
fullagginfom_sdata <- merge(fullagginfom, rawsurveycompletes1,
  by.x = "ID", by.y = "ID")
fullagginfom_sdata <- subset(fullagginfom_sdata, QC == 0)
fullagginfom_sdata <- fullagginfom_sdata[-which(duplicated(fullagginfom_sdata$ID)),
]
finalRA_m <- aggregate(fullagginfom_sdata[2], by = list(RA = fullagginfom_sdata$RA),
  data = fullagginfom_sdata, FUN = mean)
finalUA_m <- aggregate(fullagginfom_sdata[2], by = list(UA = fullagginfom_sdata$UA),
  data = fullagginfom_sdata, FUN = mean)
finalRAUA_m <- aggregate(fullagginfom_sdata[2], by = list(RA = fullagginfom_sdata$RA,
  UA = fullagginfom_sdata$UA), data = fullagginfom_sdata, FUN = mean)
for (ivar in 2:17) {
  finalRA_m[ivar] <- aggregate(fullagginfom_sdata[ivar], by = list(RA = fullagginfom_sdata$RA),
    data = fullagginfom_sdata, FUN = mean, na.rm = TRUE)[2]
  finalUA_m[ivar] <- aggregate(fullagginfom_sdata[ivar], by = list(UA = fullagginfom_sdata$UA),
    data = fullagginfom_sdata, FUN = mean, na.rm = TRUE)[2]
  finalRAUA_m[ivar + 1] <- aggregate(fullagginfom_sdata[ivar],
    by = list(RA = fullagginfom_sdata$RA, UA = fullagginfom_sdata$UA),
    data = fullagginfom_sdata, FUN = mean, na.rm = TRUE)[3]
}
ftable(fullagginfom_sdata$RA, fullagginfom_sdata$UA)
##### DESKTOP AND MOBILE #####
fullagginfo_all_sdata <- merge(fullagginfo_all, rawsurveycompletes1,
  by.x = "ID", by.y = "ID")
fullagginfo_all_sdata <- subset(fullagginfo_all_sdata, QC ==
  0 & device != 2)
fullagginfo_all_sdata <- fullagginfo_all_sdata[-which(duplicated(fullagginfo_all_sdata$ID)),
]
finalRA_all <- aggregate(fullagginfo_all_sdata[2], by = list(RA = fullagginfo_all_sdata$RA),
  data = fullagginfo_all_sdata, FUN = mean)
finalUA_all <- aggregate(fullagginfo_all_sdata[2], by = list(UA = fullagginfo_all_sdata$UA),
  data = fullagginfo_all_sdata, FUN = mean)
finalRAUA_all <- aggregate(fullagginfo_all_sdata[2], by = list(RA = fullagginfo_all_sdata$RA,
  UA = fullagginfo_all_sdata$UA), data = fullagginfo_all_sdata,
  FUN = mean)
for (ivar in 2:13) {
  finalRA_all[ivar] <- aggregate(fullagginfo_all_sdata[ivar],
    by = list(RA = fullagginfo_all_sdata$RA), data = fullagginfo_all_sdata,
    FUN = mean, na.rm = TRUE)[2]
  finalUA_all[ivar] <- aggregate(fullagginfo_all_sdata[ivar],
    by = list(UA = fullagginfo_all_sdata$UA), data = fullagginfo_all_sdata,
    FUN = mean, na.rm = TRUE)[2]
  finalRAUA_all[ivar + 1] <- aggregate(fullagginfo_all_sdata[ivar],
    by = list(RA = fullagginfo_all_sdata$RA, UA = fullagginfo_all_sdata$UA),
    data = fullagginfo_all_sdata, FUN = mean, na.rm = TRUE)[3]
}

```



## CODE CHUNK #8. FINAL ANALYSIS AND OLS TESTS

```
library(stargazer)
library(sandwich)
library(lmtest)

#####
##### RESTRICTED MODEL #####
#####

##### INDEPENDENT VARIABLE LIST #####
#####

listvars <- c( "I(RA == 'seeking')", "I(UA == 'seeking')", "I(RA == 'seeking')*I(UA == 'seeking')", "days_act

core_ind_var_set <- paste0( listvars , collapse = " + " )

#RESTRICTED
ind_var_set_r <- paste0( c(core_ind_var_set), collapse = " + " )

#####
##### DESKTOP #####
#####

ols <- list()
ols[["restricted_desktop_1"]] <- lm( paste( "log_travel_mm" , "~", ind_var_set_r, "+", "log_mm"
ols[["restricted_desktop_2"]] <- lm( paste( "log_travel_domains" , "~", ind_var_set_r, "+", "log_dom
ols[["restricted_desktop_3"]] <- lm( paste( "log_travel_domains_PV" , "~", ind_var_set_r, "+", "log_dom
ols[["restricted_desktop_4"]] <- lm( paste( "log_total_time_travel" , "~", ind_var_set_r, "+", "log_tin
ols[["restricted_desktop_5"]] <- lm( paste( "log_total_time_travel2" , "~", ind_var_set_r, "+", "log_tin
ols[["restricted_desktop_6"]] <- lm( paste( "travel_mm" , "~", ind_var_set_r, "+", "mm"
ols[["restricted_desktop_7"]] <- lm( paste( "travel_domains_PV" , "~", ind_var_set_r, "+", "domains
ols[["restricted_desktop_8"]] <- lm( paste( "travel_domains" , "~", ind_var_set_r, "+", "domains
ols[["restricted_desktop_9"]] <- lm( paste( "total_time_travel" , "~", ind_var_set_r, "+", "total_t
ols[["restricted_desktop_10"]] <- lm( paste( "total_time_travel2" , "~", ind_var_set_r, "+", "total_t
ols[["restricted_desktop_11"]] <- lm( paste( "log_travel_mm" , "~", ind_var_set_r, "+", "act_mm"
ols[["restricted_desktop_12"]] <- lm( paste( "log_travel_domains" , "~", ind_var_set_r, "+", "act_dom
ols[["restricted_desktop_13"]] <- lm( paste( "log_travel_domains_PV" , "~", ind_var_set_r, "+", "act_dom
ols[["restricted_desktop_14"]] <- lm( paste( "log_total_time_travel" , "~", ind_var_set_r, "+", "act_tot
ols[["restricted_desktop_15"]] <- lm( paste( "log_total_time_travel2" , "~", ind_var_set_r, "+", "act_tot
ols[["restricted_desktop_16"]] <- lm( paste( "travel_mm" , "~", ind_var_set_r, "+", "act_mm"
ols[["restricted_desktop_17"]] <- lm( paste( "travel_domains_PV" , "~", ind_var_set_r, "+", "act_dom
ols[["restricted_desktop_18"]] <- lm( paste( "travel_domains" , "~", ind_var_set_r, "+", "act_dom
ols[["restricted_desktop_19"]] <- lm( paste( "total_time_travel" , "~", ind_var_set_r, "+", "act_tot
ols[["restricted_desktop_20"]] <- lm( paste( "total_time_travel2" , "~", ind_var_set_r, "+", "act_tot

table_tests <- list()
table_tests[["restricted_desktop"]] <- as.data.frame( paste0( "REST:", c("log(MM)", "log(TD)", "log(TDPV)",
colnames(table_tests[["restricted_desktop"]]) <- 'data_name'
```

```

for (ivar in 1:nrow(table_tests[["restricted_desktop"]])) {
  table_tests[["restricted_desktop"]]$rname [ivar] <- ivar
  table_tests[["restricted_desktop"]]$Residual_Standard_Err[ivar] <- summary( ols[[paste0("restricted
  table_tests[["restricted_desktop"]]$F.Stat [ivar] <- summary( ols[[paste0("restricted
  table_tests[["restricted_desktop"]]$NumDF [ivar] <- summary( ols[[paste0("restricted
  table_tests[["restricted_desktop"]]$FDenDF [ivar] <- summary( ols[[paste0("restricted
  table_tests[["restricted_desktop"]]$R.Sq [ivar] <- summary( ols[[paste0("restricted
  table_tests[["restricted_desktop"]]$Adj.R.Sq [ivar] <- summary( ols[[paste0("restricted
  table_tests[["restricted_desktop"]]$Shapiro_Wilk_Stat [ivar] <- shapiro.test( ols[[paste0("restricted
  table_tests[["restricted_desktop"]]$Shapiro_Wilk_P.val [ivar] <- shapiro.test( ols[[paste0("restricted
  table_tests[["restricted_desktop"]]$Reset_Stat [ivar] <- resettest( ols[[paste0("restricted
  table_tests[["restricted_desktop"]]$Reset_P.val [ivar] <- resettest( ols[[paste0("restricted
  table_tests[["restricted_desktop"]]$BP_Stat [ivar] <- bptest( ols[[paste0("restricted
  table_tests[["restricted_desktop"]]$BP_P.val [ivar] <- bptest( ols[[paste0("restricted
  table_tests[["restricted_desktop"]]$Wald.F.Stat [ivar] <- waldtest( ols[[paste0("restricted
  table_tests[["restricted_desktop"]]$Wald.P.Stat [ivar] <- waldtest( ols[[paste0("restricted
}

```

```

stargazer( ols[["restricted_desktop_1" ]],
  ols[["restricted_desktop_2" ]],
  ols[["restricted_desktop_3" ]],
  ols[["restricted_desktop_4" ]],
  ols[["restricted_desktop_5" ]],
  ols[["restricted_desktop_6" ]],
  ols[["restricted_desktop_7" ]],
  ols[["restricted_desktop_8" ]],
  ols[["restricted_desktop_9" ]],
  ols[["restricted_desktop_10"]],
  ols[["restricted_desktop_11"]],
  ols[["restricted_desktop_12"]],
  ols[["restricted_desktop_13"]],
  ols[["restricted_desktop_14"]],
  ols[["restricted_desktop_15"]],
  ols[["restricted_desktop_16"]],
  ols[["restricted_desktop_17"]],
  ols[["restricted_desktop_18"]],
  ols[["restricted_desktop_19"]],
  ols[["restricted_desktop_20"]],
  title="Results Desktop Restricted",
  align=TRUE,
  type = "text",
  #covariate.labels = c("Risk Seeking", "Uncertainty Seeking", "days in panel", "RS * US"),
  dep.var.labels = c("log(MM)", "log(TD)", "log(TDPV)", "log(TT)", "log(TL)", "MM", "TD", "TDPV", "
  report=('vc*p'),
  no.space = TRUE
)

```

```

stargazer(table_tests[["restricted_desktop"]], type = "text", summary=FALSE)

```

```

#####
##### MOBILE #####

```

```
#####
```

```
#FULL
```

```
#REMOVE Q11_5 bc all the same values
```

```
surv_ind_var_set_m <- paste0( fulllist[-which(fulllist=="Q11_5")] , collapse = " + " )  
ind_var_set_m <- paste0( c(core_ind_var_set, surv_ind_var_set_m), collapse = " + " )
```

```
ols[["restricted_mobile_1" ]] <- lm( paste( "log_travel_mm" , "~", ind_var_set_r, "+", "log_mm"  
ols[["restricted_mobile_2" ]] <- lm( paste( "log_travel_domains" , "~", ind_var_set_r, "+", "log_doma  
ols[["restricted_mobile_3" ]] <- lm( paste( "log_travel_domains_PV" , "~", ind_var_set_r, "+", "log_doma  
ols[["restricted_mobile_4" ]] <- lm( paste( "log_total_time_travel" , "~", ind_var_set_r, "+", "log_time  
ols[["restricted_mobile_5" ]] <- lm( paste( "log_total_time_travel2" , "~", ind_var_set_r, "+", "log_time  
ols[["restricted_mobile_6" ]] <- lm( paste( "travel_mm" , "~", ind_var_set_r, "+", "mm"  
ols[["restricted_mobile_7" ]] <- lm( paste( "travel_domains_PV" , "~", ind_var_set_r, "+", "domains  
ols[["restricted_mobile_8" ]] <- lm( paste( "travel_domains" , "~", ind_var_set_r, "+", "domains  
ols[["restricted_mobile_9" ]] <- lm( paste( "total_time_travel" , "~", ind_var_set_r, "+", "total_t  
ols[["restricted_mobile_10" ]] <- lm( paste( "total_time_travel2" , "~", ind_var_set_r, "+", "total_t  
ols[["restricted_mobile_11" ]] <- lm( paste( "log_travel_mm" , "~", ind_var_set_r, "+", "act_mm"  
ols[["restricted_mobile_12" ]] <- lm( paste( "log_travel_domains" , "~", ind_var_set_r, "+", "act_doma  
ols[["restricted_mobile_13" ]] <- lm( paste( "log_travel_domains_PV" , "~", ind_var_set_r, "+", "act_doma  
ols[["restricted_mobile_14" ]] <- lm( paste( "log_total_time_travel" , "~", ind_var_set_r, "+", "act_tota  
ols[["restricted_mobile_15" ]] <- lm( paste( "log_total_time_travel2" , "~", ind_var_set_r, "+", "act_tota  
ols[["restricted_mobile_16" ]] <- lm( paste( "travel_mm" , "~", ind_var_set_r, "+", "act_mm"  
ols[["restricted_mobile_17" ]] <- lm( paste( "travel_domains_PV" , "~", ind_var_set_r, "+", "act_doma  
ols[["restricted_mobile_18" ]] <- lm( paste( "travel_domains" , "~", ind_var_set_r, "+", "act_doma  
ols[["restricted_mobile_19" ]] <- lm( paste( "total_time_travel" , "~", ind_var_set_r, "+", "act_tota  
ols[["restricted_mobile_20" ]] <- lm( paste( "total_time_travel2" , "~", ind_var_set_r, "+", "act_tota
```

```
table_tests[["restricted_mobile"]] <- as.data.frame( paste0( "REST:", c("log(MM)", "log(TD)", "log(TDPV)", "  
colnames(table_tests[["restricted_mobile"]]) <- 'data_name'
```

```
for (ivar in 1:nrow(table_tests[["restricted_mobile"]])) {  
  table_tests[["restricted_mobile"]]$name [ivar] <- ivar  
  table_tests[["restricted_mobile"]]$Residual_Standard_Err[ivar] <- summary( ols[[paste0("restricted_r  
  table_tests[["restricted_mobile"]]$F.Stat [ivar] <- summary( ols[[paste0("restricted_r  
  table_tests[["restricted_mobile"]]$NumDF [ivar] <- summary( ols[[paste0("restricted_r  
  table_tests[["restricted_mobile"]]$FDenDF [ivar] <- summary( ols[[paste0("restricted_r  
  table_tests[["restricted_mobile"]]$R.Sq [ivar] <- summary( ols[[paste0("restricted_r  
  table_tests[["restricted_mobile"]]$Adj.R.Sq [ivar] <- summary( ols[[paste0("restricted_r  
  table_tests[["restricted_mobile"]]$Shapiro_Wilk_Stat [ivar] <- shapiro.test( ols[[paste0("restricted_r  
  table_tests[["restricted_mobile"]]$Shapiro_Wilk_P.val [ivar] <- shapiro.test( ols[[paste0("restricted_r  
  table_tests[["restricted_mobile"]]$Reset_Stat [ivar] <- resettest( ols[[paste0("restricted_r  
  table_tests[["restricted_mobile"]]$Reset_P.val [ivar] <- resettest( ols[[paste0("restricted_r  
  table_tests[["restricted_mobile"]]$BP_Stat [ivar] <- bptest( ols[[paste0("restricted_r  
  table_tests[["restricted_mobile"]]$BP_P.val [ivar] <- bptest( ols[[paste0("restricted_r  
  table_tests[["restricted_mobile"]]$Wald.F.Stat [ivar] <- waldtest( ols[[paste0("restricted_r  
  table_tests[["restricted_mobile"]]$Wald.P.Stat [ivar] <- waldtest( ols[[paste0("restricted_r  
}
```

```
stargazer( ols[["restricted_mobile_1" ]],
```

```

ols[["restricted_mobile_2" ]],
ols[["restricted_mobile_3" ]],
ols[["restricted_mobile_4" ]],
ols[["restricted_mobile_5" ]],
ols[["restricted_mobile_6" ]],
ols[["restricted_mobile_7" ]],
ols[["restricted_mobile_8" ]],
ols[["restricted_mobile_9" ]],
ols[["restricted_mobile_10" ]],
ols[["restricted_mobile_11" ]],
ols[["restricted_mobile_12" ]],
ols[["restricted_mobile_13" ]],
ols[["restricted_mobile_14" ]],
ols[["restricted_mobile_15" ]],
ols[["restricted_mobile_16" ]],
ols[["restricted_mobile_17" ]],
ols[["restricted_mobile_18" ]],
ols[["restricted_mobile_19" ]],
ols[["restricted_mobile_20" ]],
title="Results Mobile Restricted",
align=TRUE,
type = "text",
#covariate.labels = c("Risk Seeking", "Uncertainty Seeking", "days in panel", "RS * US"),
dep.var.labels = c("log(MM)", "log(TD)", "log(TDPV)", "log(TT)", "log(TL)", "MM", "TD", "TDPV", "TT", "TL"),
report=('vc*p'),
no.space = TRUE
)

```

```
stargazer(table_tests[["restricted_mobile"]], type = "text", summary=FALSE)
```

```

#####
#####  DESKTOP AND MOBILE COMBINED      #####
#####

```

```

ols[["restricted_combined_1" ]] <- lm( paste( "log_travel_mm" , "~", ind_var_set_r, "+", "log_mm"
ols[["restricted_combined_2" ]] <- lm( paste( "log_travel_domains" , "~", ind_var_set_r, "+", "log_de
ols[["restricted_combined_3" ]] <- lm( paste( "log_travel_domains_PV" , "~", ind_var_set_r, "+", "log_de
ols[["restricted_combined_4" ]] <- lm( paste( "log_total_time_travel" , "~", ind_var_set_r, "+", "log_t
ols[["restricted_combined_5" ]] <- lm( paste( "log_total_time_travel2" , "~", ind_var_set_r, "+", "log_t
ols[["restricted_combined_6" ]] <- lm( paste( "travel_mm" , "~", ind_var_set_r, "+", "mm"
ols[["restricted_combined_7" ]] <- lm( paste( "travel_domains_PV" , "~", ind_var_set_r, "+", "domain
ols[["restricted_combined_8" ]] <- lm( paste( "travel_domains" , "~", ind_var_set_r, "+", "domain
ols[["restricted_combined_9" ]] <- lm( paste( "total_time_travel" , "~", ind_var_set_r, "+", "total
ols[["restricted_combined_10" ]] <- lm( paste( "total_time_travel2" , "~", ind_var_set_r, "+", "total
ols[["restricted_combined_11" ]] <- lm( paste( "log_travel_mm" , "~", ind_var_set_r, "+", "act_mr
ols[["restricted_combined_12" ]] <- lm( paste( "log_travel_domains" , "~", ind_var_set_r, "+", "act_de
ols[["restricted_combined_13" ]] <- lm( paste( "log_travel_domains_PV" , "~", ind_var_set_r, "+", "act_de
ols[["restricted_combined_14" ]] <- lm( paste( "log_total_time_travel" , "~", ind_var_set_r, "+", "act_to
ols[["restricted_combined_15" ]] <- lm( paste( "log_total_time_travel2" , "~", ind_var_set_r, "+", "act_to
ols[["restricted_combined_16" ]] <- lm( paste( "travel_mm" , "~", ind_var_set_r, "+", "act_mr
ols[["restricted_combined_17" ]] <- lm( paste( "travel_domains_PV" , "~", ind_var_set_r, "+", "act_de
ols[["restricted_combined_18" ]] <- lm( paste( "travel_domains" , "~", ind_var_set_r, "+", "act_de

```

```

ols[["restricted_combined_19"]] <- lm( paste( "total_time_travel" , "~", ind_var_set_r, "+", "act_t
ols[["restricted_combined_20"]] <- lm( paste( "total_time_travel2" , "~", ind_var_set_r, "+", "act_t

table_tests[["restricted_combined"]] <- as.data.frame( paste0( "REST:", c("log(MM)", "log(TD)", "log(TDPV)"
colnames(table_tests[["restricted_combined"]]) <- 'data_name'

for (ivar in 1:nrow(table_tests[["restricted_combined"]])) {

  table_tests[["restricted_combined"]]$rname [ivar] <- ivar
  table_tests[["restricted_combined"]]$Residual_Standard_Err[ivar] <- summary( ols[[paste0("restricted
  table_tests[["restricted_combined"]]$F.Stat [ivar] <- summary( ols[[paste0("restricted
  table_tests[["restricted_combined"]]$NumDF [ivar] <- summary( ols[[paste0("restricted
  table_tests[["restricted_combined"]]$FDenDF [ivar] <- summary( ols[[paste0("restricted
  table_tests[["restricted_combined"]]$R.Sq [ivar] <- summary( ols[[paste0("restricted
  table_tests[["restricted_combined"]]$Adj.R.Sq [ivar] <- summary( ols[[paste0("restricted
  table_tests[["restricted_combined"]]$Shapiro_Wilk_Stat [ivar] <- shapiro.test( ols[[paste0("restricted
  table_tests[["restricted_combined"]]$Shapiro_Wilk_P.val [ivar] <- shapiro.test( ols[[paste0("restricted
  table_tests[["restricted_combined"]]$Reset_Stat [ivar] <- resettest( ols[[paste0("restricted
  table_tests[["restricted_combined"]]$Reset_P.val [ivar] <- resettest( ols[[paste0("restricted
  table_tests[["restricted_combined"]]$BP_Stat [ivar] <- bptest( ols[[paste0("restricted
  table_tests[["restricted_combined"]]$BP_P.val [ivar] <- bptest( ols[[paste0("restricted
  table_tests[["restricted_combined"]]$Wald.F.Stat [ivar] <- waldtest( ols[[paste0("restricted
  table_tests[["restricted_combined"]]$Wald.P.Stat [ivar] <- waldtest( ols[[paste0("restricted

}

stargazer( ols[["restricted_combined_1" ]],
  ols[["restricted_combined_2" ]],
  ols[["restricted_combined_3" ]],
  ols[["restricted_combined_4" ]],
  ols[["restricted_combined_5" ]],
  ols[["restricted_combined_6" ]],
  ols[["restricted_combined_7" ]],
  ols[["restricted_combined_8" ]],
  ols[["restricted_combined_9" ]],
  ols[["restricted_combined_10"]],
  ols[["restricted_combined_11"]],
  ols[["restricted_combined_12"]],
  ols[["restricted_combined_13"]],
  ols[["restricted_combined_14"]],
  ols[["restricted_combined_15"]],
  ols[["restricted_combined_16"]],
  ols[["restricted_combined_17"]],
  ols[["restricted_combined_18"]],
  ols[["restricted_combined_19"]],
  ols[["restricted_combined_20"]],
  title="Results Combined Restricted",
  align=TRUE,
  type = "text",
  #covariate.labels = c("Risk Seeking", "Uncertainty Seeking", "days in panel", "RS * US"),
  dep.var.labels = c("log(MM)", "log(TD)", "log(TDPV)", "log(TT)", "log(TL)", "MM", "TD", "TDPV", "
  report=('vc* p'),

```

```

        no.space = TRUE
    )

stargazer(table_tests[["restricted_combined"]], type = "text", summary=FALSE)

#####
##### Summary ALL RESTRICTED MODELS #####
#####

stargazer(ols[["restricted_desktop_11"]],
  ols[["restricted_desktop_12"]],
  ols[["restricted_desktop_13"]],
  ols[["restricted_desktop_14"]],
  ols[["restricted_desktop_15"]],
  ols[["restricted_mobile_11"]],
  ols[["restricted_mobile_12"]],
  ols[["restricted_mobile_13"]],
  ols[["restricted_mobile_14"]],
  ols[["restricted_mobile_15"]],
  ols[["restricted_combined_11"]],
  ols[["restricted_combined_12"]],
  ols[["restricted_combined_13"]],
  ols[["restricted_combined_14"]],
  ols[["restricted_combined_15"]],
  title="Results Restricted Desktop, Restricted Mobile, Restricted Desktop AND Mobile",
  align=TRUE,
  type = "text",
  #covariate.labels = c("Risk Seeking", "Uncertainty Seeking", "days in panel", "RS * US"),
  dep.var.labels = c( paste("DESK", c("log(MM)", "log(TD)", "log(TDPV)", "log(TT)", "log(TL)")), paste0("log(", c("MM", "TD", "TDPV", "TT", "TL"), ")")),
  report=('vc*p'),
  omit.stat = "all"
)

table_tests[["final_restricted"]] <- rbind(table_tests[["restricted_desktop"]] ,
  table_tests[["restricted_mobile"]] ,
  table_tests[["restricted_combined"]] )

table_tests[["final_restricted"]]$data_name <- c( paste0( "DESKTOP:", table_tests[["restricted_desktop"]]$data_name,
  paste0( "MOBILE:", table_tests[["restricted_mobile"]]$data_name,
  paste0( "COMBINED:", table_tests[["restricted_combined"]]$data_name,

#table_tests[["final_restricted"]] <- table_tests[["final_restricted"]][order(table_tests[["final_restricted"]][1:12], type = "text", summary=FALSE, rownames = FALSE)
#stargazer(table_tests[["final_restricted"]][1:12], type = "text", summary=FALSE, rownames = FALSE)
stargazer(table_tests[["final_restricted"]][c(11:15,31:35,51:55),], type = "text", summary=FALSE, rownames = FALSE)
#apply(table_tests[["final_restricted"]][1:12],2, FUN = function(x) paste0(x , sep = ",")

#####
##### FULL MODEL #####
#####

#####
##### INDEPENDENT VARIABLE LIST #####
#####

```



```

fulllist <- c(
  #SCREEN
  #REPORTED NUMBER OF TRIPS
  'factor( S1 )',
  #REPORTED PRIMARY BOOKING
  'factor( S2 )',
  #ROLE IN DECISION MAKING
  'factor( S3 )',
  # BUSINESS and LEASURE vs. LEASURE (NOT ENOUGH)
  #'factor( S4 )',
  #DEMO,
  #AGE
  #'as.numeric(D1)',
  #AGE SQ
  #'D2' ,
  'poly( as.numeric(D1),degree = 2 )',
  #INCOME
  #'factor(D3)' ,
  #INCOME DUMMY
  #'factor(D3new)' ,
  #TRIPCHAR,
  'factor(Q1)',
  'factor(Q1)*factor(Q5_5)',
  #'as.numeric( Q2)',
  'as.numeric(Q4)',
  #SOURCE,
  'factor(Q5_1)',
  'factor(Q5_2)',
  'factor(Q5_3)',
  'factor(Q5_4)',
  'factor(Q5_5)',
  'factor(Q5_6)',
  #'factor(Q5Anew)',
  #BOUGHT WHAT,
  'factor( Q6.1 )',
  'factor( Q6.2 )',
  'factor( Q6.3 )',
  #AWAY FROM HOME ,
  #as.numeric(Q7) ,
  'factor( Q7new )',
  #VISITED BEFORE,
  'factor( Q8 )',
  #as.numeric(Q9),
  #'poly( as.numeric(Q9new),degree = 2 )',
  'as.numeric(Q10)',
  #KIDS, V relatives, Stayed at hotel, airbnb, grouptrip,
  'factor( Q11_1 )' ,
  'factor( Q11_2 )' ,
  'factor( Q11_3 )' ,
  'factor( Q11_4 )' ,
  'factor( Q11_5 )' ,
  # as.numeric(Q18.1.value),
  #poly( as.numeric(Q18new),2),
  #as.numeric( Q18new),
  #'as.numeric( Q18Anew2 )',
  'factor( 0 )',

```

```

'factor( C )',
'factor( E )',
'factor( A )',
'factor( N )'
)

listvars <- c( "I(RA == 'seeking')", "I(UA == 'seeking')", "I(RA == 'seeking')*I(UA == 'seeking')" , "days_act

core_ind_var_set <- paste0( listvars , collapse = " + " )
surv_ind_var_set <- paste0( fulllist , collapse = " + " )

#RESTRICTED
ind_var_set_r <- paste0( c(core_ind_var_set), collapse = " + " )
#FULL
ind_var_set_f <- paste0( c(core_ind_var_set, surv_ind_var_set), collapse = " + " )

stepwise_select <- list()

#####
#####          DESKTOP          #####
#####
stepwise_select[["final_desktop_1"]] <- step( lm( paste( "log_t
scope = list(lower = as.formula(paste( "log_t
upper = as.formula(paste( "log_t

stepwise_select[["final_desktop_2"]] <- step( lm( paste( "log_t
scope = list(lower = as.formula(paste( "log_t
upper = as.formula(paste( "log_t

stepwise_select[["final_desktop_3"]] <- step( lm( paste( "log_t
scope = list(lower = as.formula(paste( "log_t
upper = as.formula(paste( "log_t

stepwise_select[["final_desktop_4"]] <- step( lm( paste( "log_t
scope = list(lower = as.formula(paste( "log_t
upper = as.formula(paste( "log_t

stepwise_select[["final_desktop_5"]] <- step( lm( paste( "log_t
scope = list(lower = as.formula(paste( "log_t
upper = as.formula(paste( "log_t

#####
#####          MOBILE          #####
#####
surv_ind_var_set_m <- paste0( fulllist[-which(fulllist=="factor( Q11_5 )")] , collapse = " + " )
ind_var_set_m <- paste0( c(core_ind_var_set, surv_ind_var_set_m), collapse = " + " )

stepwise_select[["final_mobile_1"]] <- step( lm( paste( "log_t
scope = list(lower = as.formula(paste( "log_t
upper = as.formula(paste( "log_t

stepwise_select[["final_mobile_2"]] <- step( lm( paste( "log_t
scope = list(lower = as.formula(paste( "log_t
upper = as.formula(paste( "log_t

stepwise_select[["final_mobile_3"]] <- step( lm( paste( "log_t
scope = list(lower = as.formula(paste( "log_t
upper = as.formula(paste( "log_t

stepwise_select[["final_mobile_4"]] <- step( lm( paste( "log_t
scope = list(lower = as.formula(paste( "log_t

```



```

stepwise_select[["final_mobile_5"]] <-
#####
#####          COMBINED          #####
#####

stepwise_select[["final_combined_1"]] <-
step( lm(
  scope = list(lower = as.formula(paste( "log_tr
  upper = as.formula(paste( "log_tr

stepwise_select[["final_combined_2"]] <-
step( lm(
  scope = list(lower = as.formula(paste( "log_tr
  upper = as.formula(paste( "log_tr

stepwise_select[["final_combined_3"]] <-
step( lm(
  scope = list(lower = as.formula(paste( "log_tr
  upper = as.formula(paste( "log_tr

stepwise_select[["final_combined_4"]] <-
step( lm(
  scope = list(lower = as.formula(paste( "log_to
  upper = as.formula(paste( "log_to

stepwise_select[["final_combined_5"]] <-
step( lm(
  scope = list(lower = as.formula(paste( "log_to
  upper = as.formula(paste( "log_to

stargazer(lm(stepwise_select[["final_desktop_1"]]),
  lm(stepwise_select[["final_desktop_2"]]),
  lm(stepwise_select[["final_desktop_3"]]),
  lm(stepwise_select[["final_desktop_4"]]),
  lm(stepwise_select[["final_desktop_5"]]),
  lm(stepwise_select[["final_mobile_1"]]),
  lm(stepwise_select[["final_mobile_2"]]),
  lm(stepwise_select[["final_mobile_3"]]),
  lm(stepwise_select[["final_mobile_4"]]),
  lm(stepwise_select[["final_mobile_5"]]),
  lm(stepwise_select[["final_combined_1"]]),
  lm(stepwise_select[["final_combined_2"]]),
  lm(stepwise_select[["final_combined_3"]]),
  lm(stepwise_select[["final_combined_4"]]),
  lm(stepwise_select[["final_combined_5"]]),
  column.labels = c("Desktop", "Mobile", "Combined"),
  column.separate = c(5, 5, 5),

  #dep.var.labels = c( paste("D", c("log(MM)", "log(TD)", "log(TDPV)", "log(TT)", "log(TL)")),paste
  dep.var.labels = c( paste("D. Model#", 1:5),paste("M. Model#", 1:5),paste("C. Model#", 1:5) ),
  multicolumn = FALSE,
  omit.stat = "all",
  covariate.labels = c("Risk Seeking", "Uncertainty Seeking", "Risk(x)Uncertainty: Seeking",
    "Days active", "D Purchase", "D Micromoments", "D Domains", "D PageViews", "D Time
    "S1. x1 travel related purchase", "S1. x2 travel related purchase", "S1. x3+ tr
    "Q1. Country: Asia (Base Europe)", "Q1. Country: North America (Base Europe)",
    "Q4. Planning horizon (Weeks)",
    "Q5.2. Used advice of friends or relatives", "Q5.3. Used tourist information o
    "Q6.1. Purchased online: Transport", "Q6.2. Purchased online: Accommodation",
    "Q7. Trip longer than 3 nights",
    "Q8. Visited before", "Q10. Numer of people",

```

```

        "Q11.1. Children (N)", "Q11.2. Visited Friends/Relatives (N)", "Q11.3. Stayed at home (N)",
        "Openness to experience (low)", "Extraversion (low)", "Neuroticism (low)", "Agreeableness (low)",
        "Asia x Travel Agent", "North America x Travel Agent", "South America x Travel Agent"
    ),
    order = c(1,2,42,3,4,5,6,12,7,8,
              9:11,13:21,23,
              27,24,28,22,
              25,29,26,
              30:32,
              33,37,34,36,35
    ),
    type="text")

final_results <- list()

for(ivar1 in c("final_desktop_", "final_mobile_", "final_combined_")){
  for(ivar in 1:5){

    print(paste0(ivar1,ivar))
    final_results[[ paste0(ivar1,ivar) ]]$dataname <- paste0(ivar1,ivar)

    final_results[[ paste0(ivar1,ivar) ]]$Residual_Standard_Err[ paste0(ivar1,ivar) ] <- summary(
    final_results[[ paste0(ivar1,ivar) ]]$F.Stat [ paste0(ivar1,ivar) ] <- summary(
    final_results[[ paste0(ivar1,ivar) ]]$NumDF [ paste0(ivar1,ivar) ] <- summary(
    final_results[[ paste0(ivar1,ivar) ]]$FDenDF [ paste0(ivar1,ivar) ] <- summary(
    final_results[[ paste0(ivar1,ivar) ]]$R.Sq [ paste0(ivar1,ivar) ] <- summary(
    final_results[[ paste0(ivar1,ivar) ]]$Adj.R.Sq [ paste0(ivar1,ivar) ] <- summary(
    final_results[[ paste0(ivar1,ivar) ]]$Shapiro_Wilk_Stat [ paste0(ivar1,ivar) ] <- shapiro.test(
    final_results[[ paste0(ivar1,ivar) ]]$Shapiro_Wilk_P.val [ paste0(ivar1,ivar) ] <- shapiro.test(
    final_results[[ paste0(ivar1,ivar) ]]$Reset_Stat [ paste0(ivar1,ivar) ] <- resettest(
    final_results[[ paste0(ivar1,ivar) ]]$Reset_P.val [ paste0(ivar1,ivar) ] <- resettest(
    final_results[[ paste0(ivar1,ivar) ]]$BP_Stat [ paste0(ivar1,ivar) ] <- bptest(
    final_results[[ paste0(ivar1,ivar) ]]$BP_P.val [ paste0(ivar1,ivar) ] <- bptest(
    try(final_results[[ paste0(ivar1,ivar) ]]$Wald.F.Stat [ paste0(ivar1,ivar) ] <- waldtest(
    try(final_results[[ paste0(ivar1,ivar) ]]$Wald.P.Stat [ paste0(ivar1,ivar) ] <- waldtest(

  }
}

final_results[[ "all" ]] <- rbind(final_results[["final_desktop_1"]],
                                final_results[["final_desktop_2"]],
                                final_results[["final_desktop_3"]],
                                final_results[["final_desktop_4"]],
                                final_results[["final_desktop_5"]],
                                final_results[["final_mobile_1"]],
                                final_results[["final_mobile_2"]],
                                final_results[["final_mobile_3"]],
                                final_results[["final_mobile_4"]],
                                final_results[["final_mobile_5"]],

```

```
final_results[["final_combined_1"]],  
final_results[["final_combined_2"]],  
final_results[["final_combined_3"]],  
final_results[["final_combined_4"]],  
final_results[["final_combined_5"]])
```

```
stargazer(final_results[["all"]], type="text" )
```

---

## Appendix #3. Survey Questions

	Hoe veel reis gerelateerde aankopen (transport of onderkomen) heb je gemaakt sinds Maart 2016?	How many travel related purchases (either transportation or accomodations) have you made since March 2016?
S1.1	0	0
S1.2	1	1
S1.3	2	2
S1.4	3+	3+
	Kun je aangeven wat je hebt aangeschaft?	Please specify what have you purchased? Flight/Hotel
S2.1	Vlucht	Flight
S2.2	Hotel/apartment/villa	Hotel/Apartment/Villa
S2.4	Geen van de bovengenoemde	None of the above
	Wat was je rol in het besluitproces rondom deze reis?	What was your role in the decisions around this trip during the trip planning?
S3.1	Ik was de belangrijkste beslisser	Main planner
S3.2	Ik heb deelgenomen aan de beslissing	Took part of the decision
S3.3	Ik had geen invloed op de reis	No part of the decision making
	Wat is het doel van je reis?	What was the purpose of your trip?
S4.1	ontspanning	Leasure
S4.2	zaken	Business
S4.3	zaken en ontspanning	Leasure and business
	<p>Geachte deelnemer            Welkom bij onze enquete.            Het onderwerp van deze enquete is Reizen. We willen graag meer leren over jouw en je beslissingen over aankopen gerelateerd aan reizen. Deze enquete neemt niet meer dan 10 minuten in beslag.            Bedankt!</p>	
	Wat is je leeftijd?	Age
D1	jaar	years
	Wat is je geslacht?	Gender
D2.1	Man	Man

D2.2 Vrouw Woman

Wat is het totale bruto jaarinkomen van het huishouden? (Bruto jaarinkomen is de som van het fiscale inkomen + winst van alle betaalde banen, uitkeringen en pensioenen die personen in het huishouden hadden in het jaar.) Income

- D3.1 Minimum (minder dan € 12.500)
- D3.2 Beneden modaal (€ 12.500 - < € 26.500)
- D3.3 Bijna modaal (€ 26.500 - < € 33.000)
- D3.4 Modaal (€ 33.000 - < € 39.500)
- Tussen 1 en 2 keer modaal (€ 39.500 - < € 66.000)
- D3.5 66.000)
- D3.6 Twee keer modaal (€ 66.000 - < € 78.500)
- D3.7 Meer dan twee keer modaal (€ 78.500 of meer)
- D3.8 Onbekend

Kun je je meest recente, reis gerelateerde, aankopen herinneren? (vlucht of onderkomen) Continent visited

Wat was de bestemming?

- Q1.1 Europa Europe
- Q1.2 Azië Asia
- Q1.3 Noord Amerika North America
- Q1.4 Zuid Amerika South America
- Q1.5 Australië Australia
- Q1.6 Afrika Africa

Welk land in (Q1) heb je bezocht? Country

In het geval van meerdere landen noem je het eerste land

Q1b.65

Wanneer heb je, bij benadering, geboekt? Approximately when did you made this booking?

- Q2.1 vóór week 9  
Week 09, van: February 29, 2016, naar: March 6, 2016
- Q2.2 6, 2016

- Q2.3 Week 10, van: March 7, 2016, naar: March 13, 2016
- Q2.4 Week 11, van: March 14, 2016, naar: March 20, 2016
- Q2.5 Week 12, van: March 21, 2016, naar: March 27, 2016
- Q2.6 Week 13, van: March 28, 2016, naar: April 3, 2016
- Q2.7 Week 14, van: April 4, 2016, naar: April 10, 2016
- Q2.8 Week 15, van: April 11, 2016, naar: April 17, 2016
- Q2.9 Week 16, van: April 18, 2016, naar: April 24, 2016
- Q2.10 Week 17, van: April 25, 2016, naar: May 1, 2016
- Q2.11 Week 18, van: May 2, 2016, naar: May 8, 2016
- Q2.12 Week 19, van: May 9, 2016, naar: May 15, 2016
- Q2.13 Week 20, van: May 16, 2016, naar: May 22, 2016
- Q2.14 Week 21, van: May 23, 2016, naar: May 29, 2016
- Q2.15 Week 22, van: May 30, 2016, naar: June 5, 2016
- Q2.16 Week 23, van: June 6, 2016, naar: June 12, 2016
- Q2.17 Week 24, van: June 13, 2016, naar: June 19, 2016
- Q2.18 na Week 24

Hoe lang voorafgaand aan de reis, ben je begonnen met de planning voor de reis?

Approximately how long prior to (Q2) you've started planning this trip?

Houdt er rekening mee dat de tijdindicatie betrekking heeft op de periode waarin je besloot om op reis te gaan, dit heeft niet persé betrekking op de uiteindelijke bestemming

Please note that the "planning timeframe" refers to the time when you decided to take this trip, which is not necessary related to the destination choice

Q4 Weken

Welke bronnen heb je gebruikt bij het plannen van deze reis?

Please indicate what resources have you used for planning this trip?

- 1 Ja
- 2 Nee

Q5_1	Online onderzoek	Researched alternatives online
Q5_2	Advies van vrienden en kennissen	Used advice of friends or relatives
Q5_3	De VVV	Used tourist information office
Q5_4	Reismagazines	Used travel magazines
Q5_5	Een reisbureau	Used travel agent
Q5_6	Anders	Other
	Wat was de belangrijkste informatiebron?	What was the main information source?
Q5A.1	Online onderzoek	Researched alternatives online
Q5A.2	Advies van vrienden en kennissen	Used advice of friends or relatives
Q5A.3	De VVV	Used tourist information office
Q5A.4	Reismagazines	Used travel magazines
Q5A.5	Een reisbureau	Used travel agent
Q5A.6	Anders	Other
	Welke, van de volgende reis gerelateerde, producten en/of services heb je online aangeschaft voor deze reis?	Which of the following travel related product/service have you purchased online, for this trip?
Q6.1	Reis producten (trein of bus kaartjes)	Travel (plane tickets, train tickets, bus tickets, car renting)
Q6.2	Accommodatie (hotel, hostel, AirBnB etc.)	Accommodation (hotel, hostel, AirBnB, apartment, camping)
Q6.3	Entertainment (museumkaartjes, kaartjes voor bezienswaardigheden, pretpark kaartjes etc.)	Entertainment (Museum admission tickets, sightseeing tickets, park tickets etc.)
	Hoeveel nachten duurde de reis?	How many nights away from home are you planning for this trip?
Q7	nachten	
	Had je deze reisbestemming al eens eerder bezocht?	Have you ever visited this particular destination before?
Q8.1	Ja	
Q8.2	Nee	
	Hoe vaak heb je deze reisbestemming bezocht voor het bezoek?	How many times have you been at this destination prior to the visit ?

Q9	keer	
	Voor hoeveel mensen heeft u geboekt (inclusief jezelf)?	People with you on the trip
	In het geval van een wisselende samenstelling graag het gemiddelde geven	In case the number vary during the period please fill in the average
Q10	mensen	
	Beantwoord de volgende opties	Please indicate the following options
	1 Ja	
	2 Nee	
Q11_1	Was je tijdens deze reis verantwoordelijk voor kinderen jonger dan 18?	On this trip, were there any children age 18 or less that you were responsible for?
Q11_2	Heb je vrienden of familie bezocht tijdens de reis?	Have you visited friends or relatives during your trip?
Q11_3	Heb je bij vrienden of familie thuis overnacht?	Have you stayed at home of friends or relatives during your trip?
Q11_4	Heb je in een hotel/motel of appartement overnacht?	Have you stayed in hotel/motel/airbnb?
Q11_5	Is het een georganiseerde groepsreis?	Was it a group trip?
	Wat is de geschatte budget voor deze reis (per persoon in EURO)?	What is the approximate budget for this trip (per person in EURO)?
Q18	Euro	
Q18.2	Zeg ik liever niet	
	Gelieve aan te geven in welke mate je het eens of oneens bent met de volgende uitspraken die uw persoonlijkheid omschrijven	Please indicate to what extend you agree or disagree that the following statements describe your personality?
	Ik zie mezelf als:	
	1 Zeer oneens	
	2 Oneens	
	3 Niet eens en niet oneens	
	4 Eens	
	5 Zeer eens	
Q15_1	Ontspannen is, goed met stress kan omgaan	Extraverted, enthusiastic.



Q15_2	Somber is	Critical, quarrelsome.
Q15_3	Spraakzaam is	Dependable, self-disciplined.
Q15_4	Terughoudend is Soms verlegen, geremd is	Anxious, easily upset.
Q15_5	Origineel is, met nieuwe ideeën komt	Open to new experiences, complex.
Q15_6	Benieuwd is naar veel verschillende dingen	Reserved, quiet.
Q15_7	Grondig te werk gaat	Sympathetic, warm.
Q15_8	Een beetje nonchalant kan zijn Behulpzaam en onzelfzuchtig ten opzichte van	Disorganized, careless.
Q15_9	anderen is	Calm, emotionally stable.
Q15_10	Geneigd is kritiek te hebben op anderen	Conventional, uncreative.
Q15_11	Attent en aardig is voor bijna iedereen	

Geef aan in welke mate de volgende punten je algemene houding beschrijven

- 1 Zeer oneens
- 2 Oneens
- 3 Niet eens en niet oneens
- 4 Eens
- 5 Zeer eens

Q16_1	Ik neem het zekere voor het onzekere	I would rather be safe than sorry.
Q16_2	Ik wil zekerheid voor ik iets aanschaf	I want to be sure before I purchase anything.
Q16_3	Ik mijd risicovolle zaken	I avoid risky things.
Q16_4	Ik vind het belangrijk om duidelijke instructies te krijgen, zodat ik altijd weet wat er van mij verwacht wordt	It is important to have instructions spelled out in detail so I always know what I am expected to do.
Q16_5	Het is belangrijk om instructies en procedures strak op te volgen	It is important to closely follow instructions and procedures.
Q16_6	Regelgeving is belangrijk omdat het mij verteld wat er van me verwacht wordt	Rules and regulations are important because they tell me what is expected of me.
Q16_7	Gestandaardiseerde werkprocedures zijn nuttig	Standardised work procedures are helpful.
Q16_8	Ik kies niet voor risicovolle alternatieven	I do not choose risky alternatives
Q16_9	Ik maak alleen een keuze wanneer ik denk dat ik de uitkomst kan voorspellen	I only make a decision when I think I can predict the outcomes
Q16_10	Ik zoek naar zekerheden in risicovolle aankopen	I look for guarantees in risky purchases
Q16_11	Operationele instructies zijn belangrijk	Instructions for operations are important