Master Thesis

“Ethos, Pathos, or Logos? An Empirical Test of Aristotle's Persuasion with Airbnb Listings”

Author: Supervisor:

Konstantinos Tsoumas (599739) Dr. N.M. (Nuno) Almeida Camacho

Co Reader:

Dr. (Oguzhan) O. Vicil

Program:

Master of Science in Data Science and Marketing Analytics

Erasmus School of Economics

2022

Table of Contents

Abstract

Although the demand for digital accommodation platforms is increasing like never before, it is yet not known what the most appropriate promotion content for specific product types is. This research contributes to the previous literature by investigating how hosts, on the Airbnb platform, can persuade the decision of potential guests to book their listings. This was implemented by matching Aristoteles's Rhetoric theory, Logos (Logical Proof), Ethos (Credibility), and Pathos (Emotion) with certain variables, and by manually calibrating the occupancy rate as a performance measurement. From the aforementioned, only the first persuasive mode is matched with utilitarian products, while the remaining were related to hedonic products. Latent Dirichlet Allocation (LDA) is applied to build an understanding of the topics used in the dataset, and then Sentiment Analysis is implemented to capture the valence of hosts when writing describing themselves and their listings. The final model is estimated using Tobit regression. Persuasion through Logos (Logical proof) has the smallest negative effect on the occupancy rate. Likewise, Pathos (Emotion) also has a negative effect on the occupancy in comparison to Ethos (Credibility), which positively affects the occupancy rate suggesting a positive effect also on estimated bookings. I provide several suggestions for further research based on these results.

1 Introduction

The urge for digitalization and the appearance of peer-to-peer platforms accumulatively known as the “sharing economy” has enabled consumers to make use of under-utilized inventory via fee-based sharing (Zervas et al. 2015). Consumers throughout the world have accepted the services offered by new businesses (models) in a sharing economy, creating social and environmental value. In such cases, consumers are valuable “architects” in the creation process (Dellaert, 2019).

Digital sharing economy platforms have reformed the competition and the prospects in industries. Airbnb is such an eminent instance of the new era with sharing economy, which skyrocketed and became the world’s most known global provider of short-term rentals (STRs). Airbnb is a San Francisco-based company, founded in 2007, and it is currently active in over 220 countries and over 100,000 cities with 12.7 million listings as of December 2021, with its main competitors being Booking.com, TripAdvisor Rentals, and Vrbo. A few meaningful insights are that guests are not just traveling on Airbnb; they are living on Airbnb (stays for more than 28nights was over 90% at the end of Q4 2021), and also guests are discovering thousands of small towns and rural communities via Airbnb. In comparison to traditional home rentals and hospitality, Airbnb properties vary enormously in regard to location, amenities, hosts (local/Superhosts), and length of stay. Potential customers then have to choose between these attributes to select their future stay, but all of the information given may not be helpful or might not be given attention. However, by providing extensive and detailed descriptions of properties and hosts, it is found to increase review volume (Sai Liang et al. 2019).

To further understand the theory and intuition of sharing economy, it is determinant to understand the decision-making process from a customer’s point of view. The sharing economy is almost wholly based online, which always contains risks. Therefore, trust plays a primary factor in the sharing economy. On the Airbnb platform, hosts and potential guests can easily communicate and interact with each other online with their fingerprints. That implies that as hosts have publicly available information about the place, they provide but also for themselves (description), potential guests can start their decision-making process a step before not only by the amenities and features of the place but also by the hosts as persons. Various previous studies have focused on the people’s motives for choosing Airbnb (e.g., Ert et al., 2016; Stors and Kagermeier, 2015; Guttentag et al., 2017). (Ert al al., 2016) indicates that although attributes that evoke trust in a host seem to influence potential guests, this is dynamic and subject to change. (Stors and Kagermeier, 2015) however, it identifies financial savings and personal interest (e.g., unique visitor experience, social benefits) as the primary motives for choosing Airbnb.

Regarding persuasion, as defined in one of the masterpieces (Conger, 1998), it’s a continuous learning process in which a persuader directs the audience to a wanted position. Moreover, as mentioned in the same article, numbers do not make an emotional impact. Still, vivid language and stories do, so a persuader should closely observe the audience's emotions to clearly receive the message. It is a process that requires time and preparation in terms of arguments, adjusting positions, understanding what persuasion is and how it works, and matching the aforestated with the specific audience. This study is based on Aristotle's rhetorical theory, which is a general concept with three main persuasion modes: logos, ethos, and pathos.

In brief, logos refers to logical proof (facts and logic), ethos refers to credibility (of the character), and pathos relates to emotion and sympathy (of the audience and character) (Rapp, 2002). In this research, logos regards the host's logical arguments (e.g., star ratings), ethos regards host credibility (e.g,. Superhost badge), and pathos follows to the ability of the host to direct potential customers to make a purchase which is mainly performed by using vivid and emotional language in the description (personal or listing's).

Listings are considered to be either hedonic or utilitarian products. Hedonic products are considered goods which "only pleasure has worth or value,"and they cane be referred to as experience goods. In the Airbnb setting, such goods could be a villa in a centrally located area or an experience place. On the other hand, utilitarian products imply transactional goods (e.g., a place to spend the night). Both types of products are equally crucial in Airbnb as there is a diverse community of potential customers a,nd one place may serve their needs most.

Although the emergence of sharing economy platforms made numerous research available on various research topics, it is yet ambiguous which is the most appropriate content for houses that are more (less) premium and does the type of promotion matter. Considering this literature gap, this research aims to contribute to understanding the persuasion appeal of the description of a listing to different users. Therefore, the following research question and sub-questions have been formulated: RQ:

"**Depending on the product type, how can we optimize contact to sell the right property with the right message?"**

To answer this research question, a sub-question is formulated based on persuasion theory: **"Should a host adapt the selling message whether is it a hedonic or a transactional listing?"**

To answer these questions, Airbnb will be used as a laboratory for the analysis, and a publicly available dataset from the company's website will be utilized. In addition, Aristotle's perspective on the persuasion theory of logos-ethos-pathos will be implemented to investigate which of the three categories a listing can be more appealing to customers and what can be a valuable combination for the listing's description to increase performance. The performance is measured by the occupancy rate, which is a key performance indicator in the hotel industry field (Agarwal et al., 2013). Based on the occupancy rate, one can accurately measure a hedonic (experience) listing on Airbnb compared to a utilitarian (just to spend the night) alternative as this measurement is based on the whole process of choosing a place.

Over and above that, I have used a dataset that contains various information about Airbnb listings in Amsterdam from December 6th, 2021. Unfortunately, as Amsterdam laws are quite strict due to the popularity of its destination, starting as early as 2017, properties listed as "Entire Place" can only be rented up to sixty nights per calendar year which was then changed to thirty days per calendar year in 2021. A license or permit is needed to rent beyond this limit, and a specific form must be filled in. Thus, the occupancy rate has to be calculated by considering specific points. This is further explained in the Data chapter.

In the quest to answer the research questions, I will implement text analysis to formulate a clear picture of the dataset. More specifically, I will apply Latent Dirichlet Allocation (LDA) to investigate the most frequent latent topics used in both the description of each listing and the host's bio. Additionally, I will attribute specific variables to the persuasion modes Aristoteles introduced, namely, Logos, Ethos, and Pathos. More information can be found in the Data chapter. For instance, Logos refers to logical arguments so the number of amenities can hold as an indicator, while. At the same time, Ethos concerns credibility, so an indication can be if the host has a Superhost badge. Lastly, because Pathos affects the emotion of the text to persuade the audience, no single variable can be used. Rather, I apply sentiment analysis based on the National Research Council (NRC) Canada lexicon, which contains a list of 14,182. It measures the emotional tone within a text and sentiment while allowing me to determine to what extent the host appeals to pathos.

Finally, to compare the Airbnb listings, I will need control variables that affect the occupancy rate apart from the core variables attributed to the three appeals mentioned above. To reach the listings, I will use Tobit's regression with the variables above because continuous variables that are bounded in nature are generally addressed using Tobit models, censored regressions, or truncated models. Tobit regressions are suitable for settings in which the dependent variable is bounded at one of the extremes (variable censored in upper or lower bound), presents positive mass of observations at that extreme, and is unbounded otherwise. The values of occupancy rate, the dependent variable, lie between the range of 0 and 1 inclusive, so it cannot take valuesmatters less than zero or more than one. This is because occupancy rate is calculated under certain conditions ,thus, making Tobit's regression suitable for analysis.

By implementing those above, the thesis contributes to the literature in a particular manner: Initially, this research aims to understand how listing descriptions impact listing performance based on different attributes of the listing. There is not much research on this content since most of the literature is based on the listing description to adjust the price in sharing economy or the impact of report to convince guests. However, many studies have investigated the price determination factors in urban areas, Wang and Nicolau (2017) and Porges (2013). In addition, a similar study was conducted depending on Aristotle’s perspective of logos-ethos-pathos to unveil which appeals the most to users (Heejeong et al. 2018), but this is not aimed toward finding the right message with the right property by being dependent on the type of product.

2 Literature Review

The following section will provide an overview of the existing literature in marketing and tourism focused on the sharing economy in general, and on the promotion of listings on Airbnb in particular. Table 1 below, summarizes the key papers reviewed in these literature streams.

2.1 Sharing Economy and Airbnb

The emerging sharing economy, often completely realigned the worldwide tourism sector, referred to as a peer-to-peer economy, platform economy, or gig economy, is undoubtedly a remarkable development (Narasimhan et al., 2017). Although, "to come up with a solid definition for the sharing economy that reflects common usage is nearly impossible," "many companies try to position themselves under the "big tent" of sharing economy because of the positive and symbolic meaning" (Schor, 2014). However, in the academic deliberation, various terminologies have been used by researchers across disciplines such as economics, marketing, sociology (consumer research): access-based consumption (Bardhi & Eckhardt. 2012), and collaborative consumption (Perren & Grauerholz, 2015) and more.

Eckhardt et al., (2019) define sharing economy as "a scalable socio-economic system that employs technology-enabled platforms to provide users with temporary access to tangible and intangible resources that may be crowdsourced" by identifying the characteristics that define sharing economy but also two more features that are typical of many firms involved in sharing economy and traditional market in parallel. The five characteristics above, namely, temporary access, transfer of economic value, platform mediation, expanded consumer role, and crowdsourced supply (Eckhardt et al., 2019), match entirely the example of Airbnb. More specifically, Airbnb provides temporary access to the accommodation without changing ownership, and this lease includes economic transactions that transfer value from one party to another. The platform consolidates supply (from "hosts[1]") and demand (from "guests") by collecting a service fee from each transaction ranging between 13-20%. Additionally, Airbnb relies on its platform as a mediation to attract customers and provide multiple features such as photos, location, videos, text, and previous reviews on its offerings. At the same time, algorithms make the matching with customers accurate, which makes the host trustworthy but also increases the price (Ert et al., 2016). Furthermore, Airbnb also expands consumer roles by making hosts both customers

[1] Hosts may not always be the property owners but people who are in charge for the bookings (e.g., communicating with potential guests, providing keys, resolving issues before and/or during the stay, etc.) Hosts requirements can be found here.

and customers, for example requiring hosts to clean and prepare the property for the following user. Lastly, Airbnb is one of the most successful crowdsourcing companies as they increased the supply of users rapidly beyond their employees (Zimmerman, 2016).

Most often, it includes ratings-based marketplaces and in-payment systems, giving a chance to their employees to work based on their schedules. However, while the implementation of this so-called “sharing economy” is now, the idea behind it is not, dating back in time to rental markets, which have been used for various reasons. The main difference with rental needs is that from a two-way transaction, it’s now a three-way transaction (Narasimhan et al., 2017). The attraction of this model is hidden within the unbelievable speed in that customers can consume, save, and exchange but parallelly feel like being part of a sustainable community (Kathan et al., 2019).

Airbnb is practically synonymous with the sharing economy and its online platform where individuals can rent a “place” (an entire house, a room in a home where the host is present, etc.) mostly primarily as tourist accommodation in various areas of the world from urban to exotic in a range of product types (from spending the night to highly luxurious). From a host’s point of view, the platform’s post includes extended description, photographs, amenities, location, and suggestions for the guests, and it allows hosts to accept/decline reservations (however, Airbnb introduced the “instant booking” feature, which allows will enable guests to book without host’s approval), payments. From a customer point of view, the booking procedure is similar to that of an online travel agency, but it may need to communicate personally with the host at times to arrange needs.

After the stay, both guests and hosts are encouraged to review one another, building a trustworthy platform (they are awarded a “Superhost” badge if the hosts are active and score high in reviews). Airbnb also promotes security by offering additional features such as free property damage protection, free liability insurance, and a refund policy that protects users against issues, for instance, inaccurate listing descriptions. The online peer-to-peer market heavily depends on trust between complete strangers (Ert & Fleischer 2019).

A consumer’s communication with a service provider reflects an emotional attachment that positively affects the development of the relations between these two parties (Bansal, Irving, and Taylor, 2004). Although various users have raised concerns regarding trust issues, Airbnb has long supported meaningful interactions between guests and hosts to build a trusted relationship that may benefit Airbnb significantly, building creating a trusted booking platform.

In addition, Airbnb has filters such as “Self-check-in,” which refers to easy access to the apartment without guidance and “Free cancellation,” which only shows stays that offer free cancellation, and “Airbnb Plus,” which is a service that provides “Thoughtfully designed, Well-equipped, Well-maintained” top-quality properties that have been inspected by Airbnb representatives Airbnb representatives have inspected checked. Airbnb has also cooperated with restaurants and local guides to offer “Experiences,” a feature that includes reservations and tours. Zooming out, Airbnb denies it competes with hotels and instead states that an Airbnb trip changes you (Trenholm, 2015). however, there is also opposition to this statement saying that hotels are worried about Airbnb (Griswold, 2016), but this might have changed in the post-COVID era as the hotel have the edge when it comes to refunding policies, but it’s a draw in terms of hygiene (Glusac, 2020).

Moreover, as the size of the sharing economy has grown, so has the magnitude of its economic impact, and Airbnb has now impacted the traditional, lower-end tourism accommodations as those are the most vulnerable rentals to increased competition, but it indicated that Airbnb supply is differentiated from the hotel supply (Zervas et al., 2017).

Table 1:Existing literature on the sharing economy

Contributors *Journal* [Google Cites]	**Context**	**Research Design**	**Key (Relevant) Findings**	**Cites**
Bansal, Harvir S et al. (2004) *Journal of the Academy of marketing Science*	The role of consumer commitment on consumers' intentions to switch.	Survey	Staying loyal to a single service provider can be desire based, cost based, or obligation based, reflecting differing psychological bases for the relationship customers have with their service providers. Customers do not churn because 'they want to', 'they have to', 'they ought to'.	1510
Bardi & Eckhardt (2012) *Journal of Consumer Research*	Studies the nature of access (access-based consumption) and antithesis ownership and sharing.	Qualitative	Discovered the consumptions dimensions, namely: temporality, anonymity, market mediation, consumer involvement, the type of accessed object, and political consumerism. Internet promote the sharing future.	2748
Dellaert (2019) *Journal of the Academy of Marketing Science*	The need for firms to define new marketing actions to account for consumer co-production.	Literature Review	Proposed a two layered conceptual framework.	134
Ert, Fleischer & Magen (2016) *Tourism Management*	The impact of seller's photos on the consumers decision process.	Experiment/Survey	Trustworthiness of a host is going along price, the more trusted is the host the higher the price of the place. Reputation given by reviews does not effect the listing's price or likelihood.	1340
Guttentag & Smith (2015) *Current Issues in Tourism*	Disruptive innovation of Airbnb.	Survey	Airbnb is expected to substitute the traditional hotel industry by underperforming upscale hotels and outperforming budgets but not completely.	2129
Guttentag et al. (2017) *Journal of Travel Research*	Segmentation and motivation using Airbnb.	Survey	Segmentations: money savers, home seekers, novelty and interactive novelty seekers, collaborative consumers. Motivations: Home benefits and amenities, novelty, interaction, local authenticity and sharing economy ethos.	613

Han et al. (2019) *International Journal of Contemporary Hospitality Management*	Explaining a guest purchase in Airbnb. from Aristotle's appeals point of view.	Experiment/Hypotheses	Logos: positive impact have the price, place, picture, star rating and the occupancy rate has a negative impact on likelihood of purchase. Ethos: positive impact has the Superhost badge, host reviews. Pathos: social words have positive impact.	32
Kathan et al. (2019), *Business Horizons*	Why the sharing economy has the potential to produce a long-term transformation in consumer behavior.	Literature Review	A reconsider process in order to think twice upcoming challenges and the significant amount of possibilities associated with the new consumption practices.	373
Narasimhan et al. (2017), *Customer Needs and Solutions*	The exponential growth of sharing economy creates various questions that needs to be studied.	Literature Review	Sharing economy providers take customers, revenue away from big Hotels (e.g. Marriott) having the advantage of adjust supply quickly to demand. Personalizes approaches reduces friction and increasing matching. Also, the entry of sharing economy could result in customer surplus along with increase in firm profits.	86
Yang et al. (2018), *Journal of Travel & Tourism Marketing*	Users trust in the sharing economy and how platform providers can better operate through Aristotle rhetoric.	Survey	Found that accommodation information is the most influencing factor for the users to trust Airbnb hosts. Aristotle's persuasive modes are positively associated with trust in Airbnb hosts which shifts to an overall trust of the Airbnb brand as a whole.	55
Zervas, Proserpio & Byers (2017), *Journal of Marketing Research*	Estimating the impact of Airbnb on the hotel industry.	Content Analysis	The intro into the Texas market has a noticeable negative impact on local lower-end hotel room revenue.	2606

2.2 Hedonism and Utilitarianism

The terms hedonism and Utilitarianism are pretty frequently used nowadays in the marketing field. The former comes from the Greek word ‘pleasure’. Psychological (or motivational), Ethical (or normative), which supports that pain or displeasure is our only motivation, and Evaluative hedonism, which claims “only pleasure has worth or value and only pain or displeasure has disvalue or the opposite of worth”, are some of the main categories (Moore, 2019). The latter is amongst the most powerful and persuasive approaches in philosophical history, and it follows the perspective that the morally right action is the ultimate action to produce sound but since the 20th Century, the term has faced several refinements (Driver, 2014).

Undoubtedly, products can be categorized in different ways, but the present study follows the classification of products into hedonic and utilitarian types. Hirschman and Holbrook (1982) defined hedonic consumption as “facets of consumer behavior that relate to the multi-sensory, fantasy and emotive aspects of one’s experience with products” while suggesting that hedonic products refer to images, fantasies, and emotional arousal. In comparison, functional performance serves as the primary value factor to consumers, like a tool to perform an assignment. Consumers tend to buy hedonic products based they dream of as reality. Possibly, all the purchase decisions have a sense of joint evaluation as consumers base their buying decisions on previous or future purchases that could be made as a substitute (Okada, 2005). Also, (Okada, 2005) found that people will most probably make hedonic purchases when a stand-alone decision/item does not need any explicit comparisons with other alternatives (Separate Evaluation) and utilitarian investments assets when multiple items are presented simultaneously. There is a clear trade-off between choosing one another (Joint Evaluation).

Shao and Li (2020) recruited 62 undergraduate students randomly assigned to a single factor (utilitarian or hedonic product) within-subjects design to see how these two types of products influence choice preferences. Their findings remarkably showed that consumers are willing to pay more for the ultimate than the approximate best choice regarding a hedonic product. Although customers prefer the ultimate best choice for hedonic products but the approximate best choice for the utilitarian product, undergraduates stated that they are willing to pay a higher price for the approximate than the ultimate best choice concerning practical products with minor differences.

Okada (2005) also researched the purchases of an SLR camera from a local camera store, from which participants responded to a three-question survey voluntarily over six weeks. The study's findings show that the exchange rate of time for hard currency (money) tends to be higher than people suppose for utilitarian hedonic versus products. Customers who consider the good more hedonic were willing to spend more time purchasing the product, while utilitarian view people as more money.

3 Hypotheses

In this chapter, I develop my hypotheses. In the first section, I review Aristotle's rhetoric theory and its three appeals: logos, ethos, and pathos. I also discuss how these three appeals can be applied as a lens through which to evaluate persuasion attempts by hosts in Airbnb listings.

3.1 Aristotle's Rhetoric Theory

Aristotle emphasized rhetoric theory as the "art of persuasion," which is a way to convince people and in good communication. The work of Aristotle, Aristotle's Rhetoric or Art of Rhetoric, consists of three books (Rhetoric I, II, and III); however, most modern authors agree that the first two books illustrate the core of the rhetorical theory. Remarkably, the two themes of Rhetoric III are not mentioned in the original agenda of Rhetoric I and II suggesting that Aristotle at the time referred to the first two books as the complete work. Aristotle identifies three key features of a good persuader, namely, argumentative persuasion that is specific to the respective genre of speech or logical proof (logos), evoking the emotions of the audience (pathos), and the one that depends on the character of the speaker or credibility (ethos) (Rapp, 2002). The table below (Table 2) depicts an example of an employee working in Marketing, trying to pursue his manager about a campaign. The model made these three strategies more tangible for the reader.

Table 2: Example of the three kinds of persuasion

Persuasion appeal	Argument
Logos (logical proof)	"The data is crystal clear; This campaign will make our sales soar."
Ethos (credibility)	"As a Marketing Analyst, I'm qualified to inform you that this
Pathos (emotion)	"Can you imagine the how the sales will change if we fund this

The significance of rhetoric theory has been around for decades in various contexts, but for this research, the marketing and advertising fields are the most relevant. However, certain persuasion modes are the most effective, but these are time-dependent (Brown et al., 2018). Currently, logos appear to be the most notable mean of persuasion since most people have access to valuable information at their fingertips and therefore are looking for well-confirmed statements. (Yang et al., 2018) findings show that each of Aristotle's rhetoric persuasive cues affected customer trust, with logos being the most effective while ethos and pathos come right after.

3.2 Aristotle’s rhetoric on the Airbnb platform:

A relatively limited number of papers have yet to study Aristotle’s rhetoric in the Airbnb setting. In one of the recent works, (Yang et al., 2018) researches the validation of Aristotle’s persuasion theory in creating user trust. Additionally, (Han et al., 2019) studied the explanation of customers booking decisions based on Aristotle’s appeals. However, the goal of this paper is, depending on the type of listing, based on Aristotle’s persuasion theory to optimize contact. Figure 1, which can be found on the next page, summarizes the conceptual framework I tested in this paper. As (Ert et al, 2016) have found, the sharing transaction is constructed based on communications between users.

To rent out a place on Airbnb, the hosts need to have an account on the platform first. Throughout the process of both the listing creation and the account making, individuals are given the opportunity to share more information about their offering and themselves. Having an informative listing description with feeling-home-like pictures actually, formats trust (Yang et al., 2018), while a detailed description not only about a listing but on a personal level improves revenue volume (Liang et al., 2020).

Findings from the latter study also revealed that both the width and depth of the property description but also a comprehensive description of the host are positively associated with the listings review volume, verifying the effectiveness of MGC (textual description constructed by the hosts). The study by (Chen and Xie, 2017) has noted that as there are no available sources to evaluate listings other than host-created information, the given information is ascendant for the customer decision-making process. These attributes are interacted with Aristotle’s persuasion modes and illustrated in Figure1.

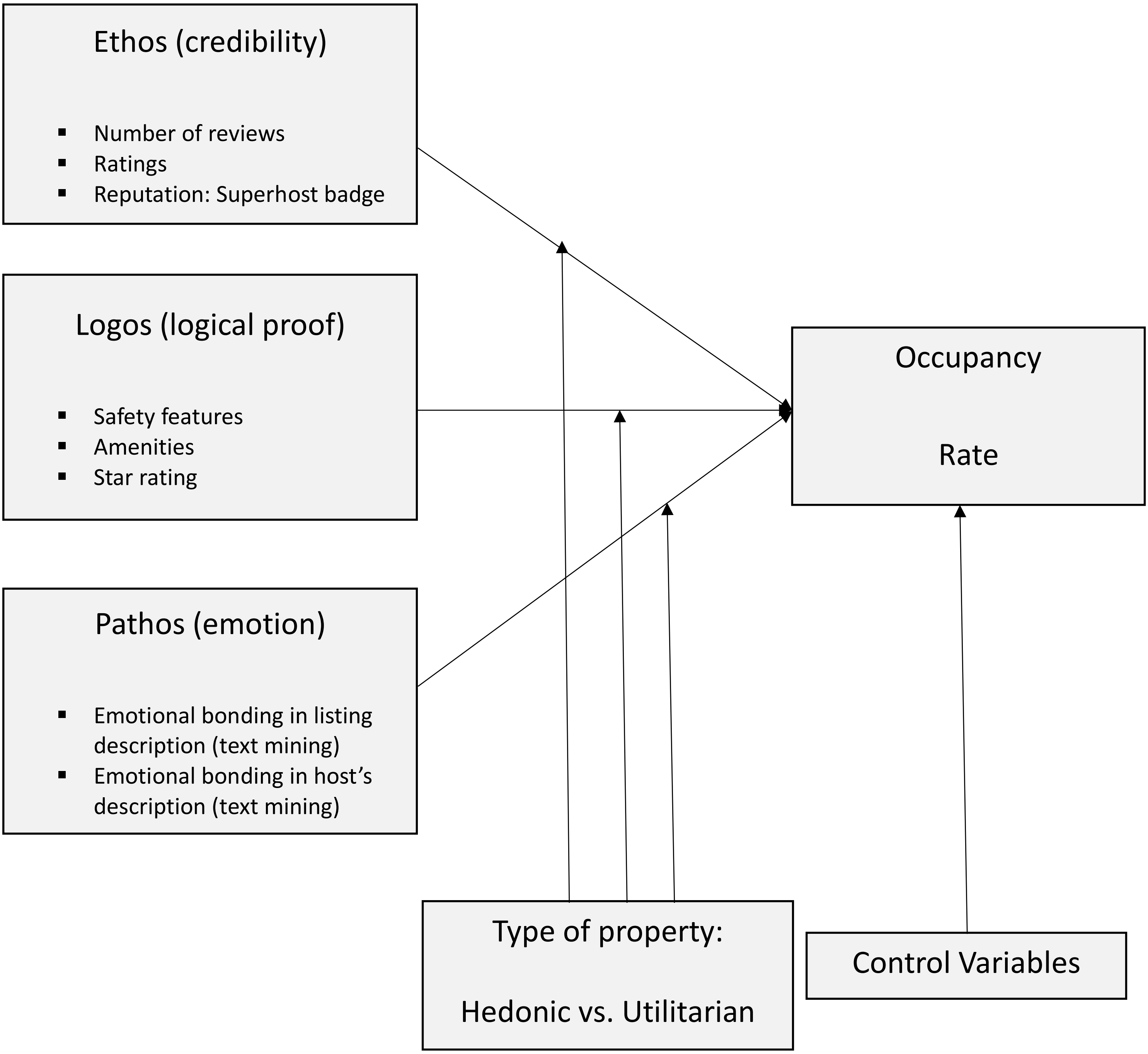


Figure 1: Conceptual Framework

3.2.1 Ethos (credibility) in Airbnb

To pursue customers through Ethos in the Airbnb setting, the strengthening of the host's credibility needs to be addressed. In Aristotle's rhetoric theory, ethos refers to the character (credibility) of the speaker and specifically when the speech is manipulated in a way of presenting the speaker's worth or credence (Rapp, 2002).

It is also mentioned in Rhetoric I, that people are used to following trustworthy speakers quicker, and easier on almost all subjects in which there are no specific criteria to decide but only arbitrary options. In addition, (Jun, 2020) found that brand credibility influences the users the most. Consumers choose credible sellers (or brands) for various different reasons such as safety, word-of-mouth, and experience. (Jill and Swait, 2008) findings suggest that brand credibility also has an impact on customers' satisfaction but also commitment.

Affective-based trust features influence most to users' trust in Airbnb hosts (Yang et al., 2019). Additionally, in the same research, it was found that reputation has the most powerful impact on users' trust in hosts, followed by interactions and familiarity. Based on the studies mentioned before, it could be expected that, all else equal, properties marketed by hosts with higher credibility are a reliable signal of customers purchasing intentions.

H1. All else equal, properties marketed by hosts with higher credibility will have a higher occupancy rate than properties marketed by hosts with lower credibility.

3.2.2 Logos (Logical proof) in Airbnb

Concerning persuasion through logos, can be accomplished by presenting arguments in order for the speaker to make his statement persuasive through logical proof. Persuasion is paramount in the Airbnb environment for the host-guest (stranger-stranger) communication and it serves as a primary element of the relationship between the owner and the potential customer (Yang et al., 2018).

Numerous researchers have tried to address this and suggested that the most appropriate way is to identify different meanings of the word 'pistis' (which means faith in Greek) for the two chapters (Rhetoric I.1, I.2). In the second chapter of Rhetoric, the word may be referred to as 'pisteis', so as to be differentiated.

In the Airbnb setting, extensive information about features such as safety features, price, occupancy, pictures, and ratings impact the decision-making process (Yang et al., 2018). More pictures of the accommodation, higher prices, higher ratings, and lower occupancies increase the attractiveness of a place, according to the same study. However, this study found that the higher the price, the more likely it is for the listing to be booked, which may not always be true. Although most of the time, higher prices indicate higher quality, various guests prefer more economical alternatives. This poses the importance of pictures of the listing as humans cognitively process information based on logical facts and form a belief.

Ert et al. (2016) found that the host's photos which, perceive as trustworthy, increased the price and the probability to book the place.

H2. All else equal, properties that contain logical proof will have a higher occupancy rate than properties that do not.

3.2.3 Pathos (emotion) in Airbnb

According to Aristotle, persuasion through pathos comes when the audience is led by a speech from which people can feel emotion or passion, and this has an impact on the judgment call they are going to make later (Rapp, 2022).

Pathos in the Airbnb setting refers to the features which affect the reader emotionally. The idea behind it is that guests are encouraged to proceed to purchase by reading appealing/effective content in the listing. The more detailed listing descriptions increase review volume and performance for property owners (Liang et al., 2020). Also, the width and depth of the description, for hosts managing multiple places, have a positive impact on review volume (Liang et al., 2020).

There are various factors that contribute to the customer decision-making process, such as persuasive, informative information and images, or emotional text that may drive customer behavior were intended without requiring the consumer's conscious recognition. Cook et al., (2019) found that advertising images may raise several levels of regional brain activity concerning the use of logical persuasion. This implies that people, even if they are unaware of it, may be inclined to choose a property with such characteristics over another.

Tussyadiah and Park (2018) find that, as previously mentioned, a host who is shown as ready to meet new people but also has experienced similar situations with the potential guests, builds trustworthiness which may lead to booking decisions. Such details could be written in the description to make the content more appealing, social, and friendly. In addition, due to the limited number of words used in the descriptions (either the host's or the listing's), the content is easier to invocate emotion.

Gibbs et al. (2017) also, highlighted in their study that sometimes Airbnb hosts act as entrepreneurs or hospitality providers although they thrive in a unique field of tourism, they don't have any specific business or hospitality knowledge. Therefore, the set of the price is driven by emotional considerations and not proper facts.

H3. All else equal, listings that contain emotional words in the description will have a higher occupancy rate than listings that do not contain emotional words (or a very small percentage).

3.3 The Moderating Role of Property Type (Hedonic vs. Utilitarian)

In H1, I argued and hypothesized that properties marketed from people with higher credibility would have a higher occupancy rate, so that ethos (credibility) increases the persuasiveness of an appeal, whereas, in H2, I likewise hypothesized that logos appeals (logical proof/rational arguments) also increase the persuasiveness of an appeal when the listing contains the logical proof. Lastly, in H3, I hypothesized that pathos (emotion), in Airbnb listings containing emotional words, will also increase the persuasiveness of an appeal. However, the strength of these effects depends almost entirely on the type of product in the Airbnb setting. More specifically, some appeals may be more effective for hedonic products compared to utilitarian products, and vice versa. As previously defined, hedonic products can be more described as “experience goods,” such as products that refer to fantasy and images. On the contrary, utilitarian products are best described as valuable to consumers.

The research conclusions from (Fleischer and Ert, 2019) mention that the Superhost badge is more important to host with fewer reputation than the opposite. Furthermore, the study also found that Superhosts seem to receive a price premium owing to their rank between 4% - 6%. This suggests that hedonic products can be well explained by ethos (credibility) as a persuasive mode because hedonic products are referring to images and messages but also experience. On the other hand, in sharing economy, customers searching for transactional (utilitarian) goods are more inclined to pay a higher price for something more relevant than the absolute best choice (Shao and Li, 2020). A mixture of logical proof can be the best combination to pursue new, potential, customers and annotates that logos (proof) match.

Lastly, pathos would probably be best matched with hedonic products as customers are searching for experience goods and not a transactional alternative. Emotion could not only be implemented in writing a listing description but could also be the case of pictures or communication with the host.

H4a: Persuasion through Ethos(credibility) can have a significant outcome to utilitarian products in the Airbnb setting.

H4b: Persuasion through Logos (logical proof/rational arguments) can have a significant outcome to utilitarian products in the Airbnb setting.

H4c: Persuasion through Pathos (emotion) can have a significant outcome to hedonic products in the Airbnb setting.

4 Data

To implement the theoretical framework in a more practical manner, an Airbnb dataset containing listings and reviews was extracted from Inside Airbnb[2]. The website provides publicly available Airbnb data of specific dates throughout the year for various continents and cities. To extract the data, open-source technologies and maps are used for the data to be more visual. For this research, the data used was extracted on the 6th of December 2021. It must be noted that the Airbnb calendar does not individualize unavailable nights and booked nights (also, some hosts might not keep their calendar updated), therefore Inside, Airbnb mark these books as "unavailable".

To be able to measure the performance across listings (hedonic and utilitarian), based on previous literature, the occupancy rate is found to be a valuable metric. This measurement, however, must be calculated manually. Ye et al., (2011), based on previous studies, using the number of reviews as a proxy for sales was found to be a good indicator for hotel sales. Five years later, Dr. Qiang Ye, used the same proxy for hotel sales, which indicates it is a significant estimate of the occupancy rate allowing researchers to measure elasticity between volume dimension and hotel rates.

As it is highly recommended and promoted, although guests are not obliged to, the rate at which consumers leave reviews is important to help property owners/lenders understand how reviews affect the listing performance and to gather additional insights for the customer decision-making process. For instance, properties on Airbnb are booked in advance based on various features but not a physical evaluation of their potential stay. This can be quite expensive, so potential customers invest a significant amount of time going through the reviews (or ratings) to understand" the advantages and disadvantages of the property.

According to Inside Airbnb's "Data Assumptions" (a chapter from the site that provided the data), the website policy is to use an occupancy model to estimate the frequency of a listing occupancy but also the listing's income. The model is a modified version of the "San Francisco Model" which includes three methods to measure occupancy rate (Inside Airbnb, 2022).

Initially, a fifty percent review rate is used to convert reviews to estimated bookings. To elaborate more on this percentage, Alex Marqusee, a housing stability manager, uses a review rate of 72% review rate, to convert reviews to estimated bookings based on Airbnb's CEO Brian Chesky. Even though this might be a legible source, the Budget and Legislative Analyst's Office (San Francisco, May 2015), which also use 72% as a review rate, introduced a more robust model using a 30.5% review rate on public data from New York. Inside Airbnb, after analyzing these percentages, found that a review rate as low as 30.5% is more fact-based but may lack reviews that have been lost due to

[2] Inside Airbnb website, link can be found by clicking here

deleted listings. A 72% is unjustified, but a 50% is acceptable. They have chosen a 50% review rate to be used, and so this research.

The second method used concerns the average length of stay multiplied by the estimated bookings individually for each listing, for a specific period, which gives the occupancy rate as a result. In our case, 3.9 nights is the average, the most recent study (Airbnb (o), 2013) found. Where there are no public statements about the average stay per night, this method uses 3.9 nights per booking or if minimum nights were found to be higher than the average, the minimum nights listing's measure is used instead. Concluding, the occupancy rate is given by multiplying estimated bookings (annually) by the minimum nights, capped at one.

Lastly, the method capped the occupancy rate at the high rate of seventy percent to ensure two main possibilities. During peak season, a host might change their minimum night availability, and this may not be visible in the reviews or in a case where a listing is constantly being reviewed. This is to make sure that the occupancy rate remains conservative at all times.

Although a short-term stay is considered for rentals less or equal to 27 days, in our case, in Amsterdam, there are extra laws that apply. As of 2017, Airbnb limited properties listed as "Entire Place" to 60 per calendar year to make the growth of sharing responsible and sustainable. This changed in 2021 as the city of Amsterdam regulated the maximum nights to stay for "Entire Place" category to half, 30 per calendar year, and still applies to this day. The limit resets on the first day of every year. To rent beyond this limit, a license or permit is needed, and a specific form must be filled in.

To conclude, although occupancy rate can be a valid performance indicator, it must be measured accurately to be of any value. Airbnb cannot automatically distinguish between booked days and unavailable days, so taking a short time into account can lead to ambiguous results. Thus, properties available for more than 30 days were considered for this research. Taking also reviews as a crucial parameter, at least one review is necessary.

4.1 Data Description and Data Cleansing

The data used to perform the analysis contains characteristics and details on listings but also a detailed review of listings in the city of Amsterdam. The final dataset which is used in the analysis is a merged dataset of detailed listings and reviews based on listing ID as given from *InsideAirbnb*.

The dataset represents the specific Airbnb listing by row while the listing's characteristics by column. The initial (merged) dataset included 5.849 rows and 75 columns, however, not all the variables were used in this research, and several columns were deleted. Variables such as URLs, dates, license, hostnames, double counted reviews, but also neighborhood information that is quite like one another (e.g., zip code, city, state, country code, URL, scrape date) were deleted because the information is not usable. Also, some of the columns such as "*bathroom*", "*calendar_updated",* "*neighbourhood_group_cleansed",* were removed as they were completely empty. The Figure below (Figure 2) presents the summary of the dataset right after removing the unusable columns.

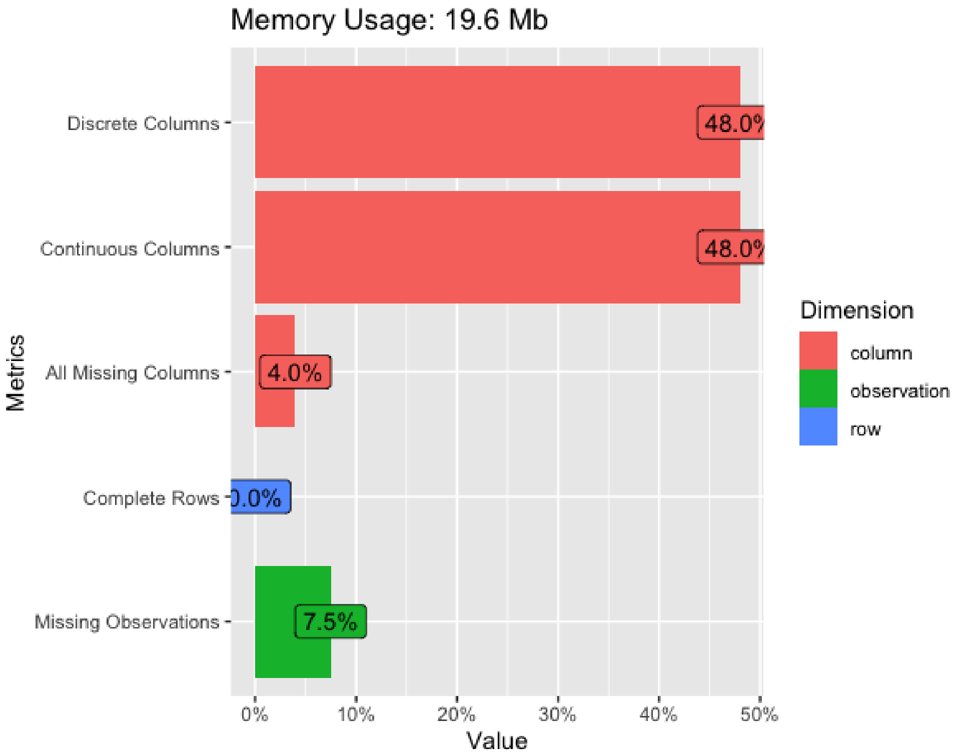


Figure 2: dataset overview

As illustrated from the graph, 7.5% of the dataset are missing values which is not a significant amount. To deal with NA values, I have removed all the rows containing NA values, which in total was 1286 rows. To elaborate more about the cleaning process, the "*host_since"* variable indicates the date when the host created an Airbnb account, and it was converted into days to be more comprehensible. More specifically, I measured the difference of days between (06-02-2022) and the date since the user created an account. In addition, the variables “name” and “description,” which contained the title of the listing, and its description, were merged into one column named “Description” for the topic model to extract words used in both. Hosts bio was not merged with the other two description columns, on purpose, as the listings-host’s descriptions are not constructed similarly and thus, it should separately be examined.

Regarding columns containing textual data, there are instances that had a small number of words (under 5). An example would be in the “*host_about*” column containing text such as “Hi I’m Rosa”. Only “*host_about,*” have 1665 instances containing less than 5 words. A language check was performed on the aforementioned columns, concluding that multiple languages were used in both columns, but only English was used in the analysis. Stopwords and punctuation were also removed from both columns.

Moreover, the amenities of each listing were counted and assigned to a new column named “nr_amenities” (excluding the safety features). Data structure changes were also implemented by converting various columns from character to numeric for usability. Safety features were extracted from the “*amenities”* column, specifically smoke alarm, first aid kit, carbon monoxide alarm, fire extinguisher, and lock on the bedroom door, and assigned to a variable named “*safety_features”* containing a calculation of the safety features for each listing. Additionally, the variable “centre_dis” was created, which calculates the distance from the center of Amsterdam based on the coordinates (using the World Geodetic System(WGS84)). Advancing, the initial dataset contained the review scores (values between 0, lowest, and 5, highest) for each of the six categories (rating, check-in, cleanliness, communication, location, scores value), which were added up and divided by 6 to get the average rating score for each listing and assigned to the “*reviews”* variable. Lastly, an important point to be highlighted is that data scraped Inside Airbnb appear to have an upper bound on the number of words used in the text, which causes an unexpected ending or cut of the sentence. The final dataset contains 2280 observations and 15 columns. The table below presents all variables used in this search.

Table 3: Variables Overview

Variable name	Type	Description
description	Character	Listings description
host_since	Numeric	The date of when the host joined the platform
host_about	Character	Host’s description in summary
host_is_superhost	Boolean	Whether a host has a Superhost badge
neighbourhood_cleansed	Character	Name of the neighborhood of the listing
host_has_profile_pic	Binary	Whether the host has a profile picture
room_type	Factor	The type of the listing (among the 4 options)
host_identity_verified	Boolean	Whether the host has verified his identity or not
number_of_reviews_ltm	Numeric	Number of reviews a listing has (last 12 months)
centre_dis	Numeric	Distance of the listing from the center of Amsterdam(WGS84)
price	Numeric	Listing’s price
description_pol	Character	Whether the host provides an emotional description of the listing
host_about_pol	Character	Whether the host provides an emotional self-bio
minimum_nights	Numeric	Minimum nights a guest can book the listing
comments	Character	User’s comments per listing
number_of_reviews	Numeric	Number of reviews a listing has
instant_bookable	Boolean	Whether the listing can be booked automatically or needs host’s
nr_amenities	Numeric	The number of amenities provided per listing
occupancy_rate	Between 0 and 1	Occupancy rate of each listing
availability_365	Numeric	Listing’s available days per year
safety_features	Numeric	Safety features of each listing
reviews	numeric	Average review stars for each listing (among 6 categories)

4.2 Exploratory Data Analysis

It is pretty important to visualize data before applying any machine learning model. The base maps below in Figure 3 depict how Airbnb listings are spread throughout Amsterdam. The figure presents spatial data according to the common standard World Geodetic System (WGS84). Not surprisingly, it is evident that many listings are gathered around the center of Amsterdam, while a noticeable number of listings can be found on the West side of Amsterdam in comparison to the East side, where not many listings are available.

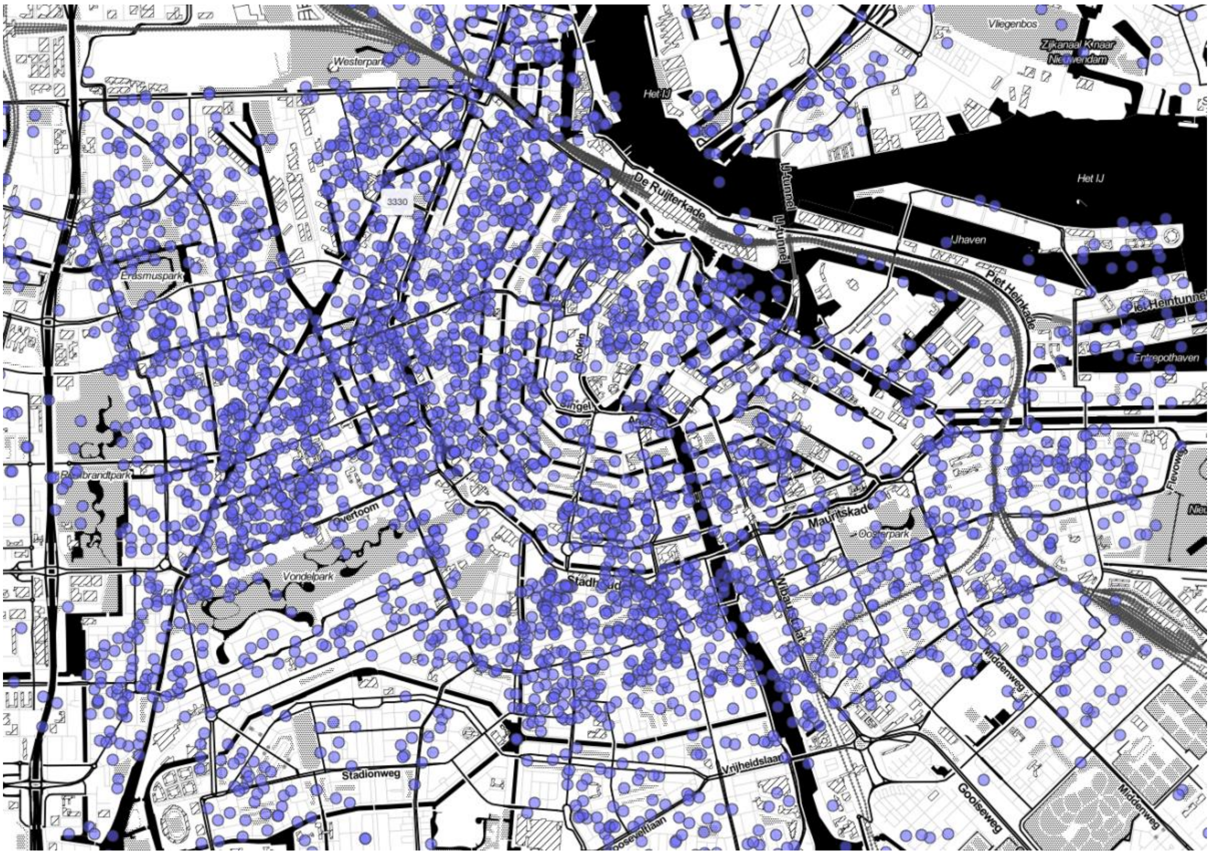


Figure 3: Map of the listings in December 2021

Furthermore, the price distribution across listings in Amsterdam on December 6th, 2021, is illustrated in Figure 4 below. As shown in the figure, the price distribution lacks symmetry; it is positively skewed. Most of the listings are priced between 150-200$ per night, which is indicated to be a frequently used price, perhaps for medium-sized apartments or premium rooms.

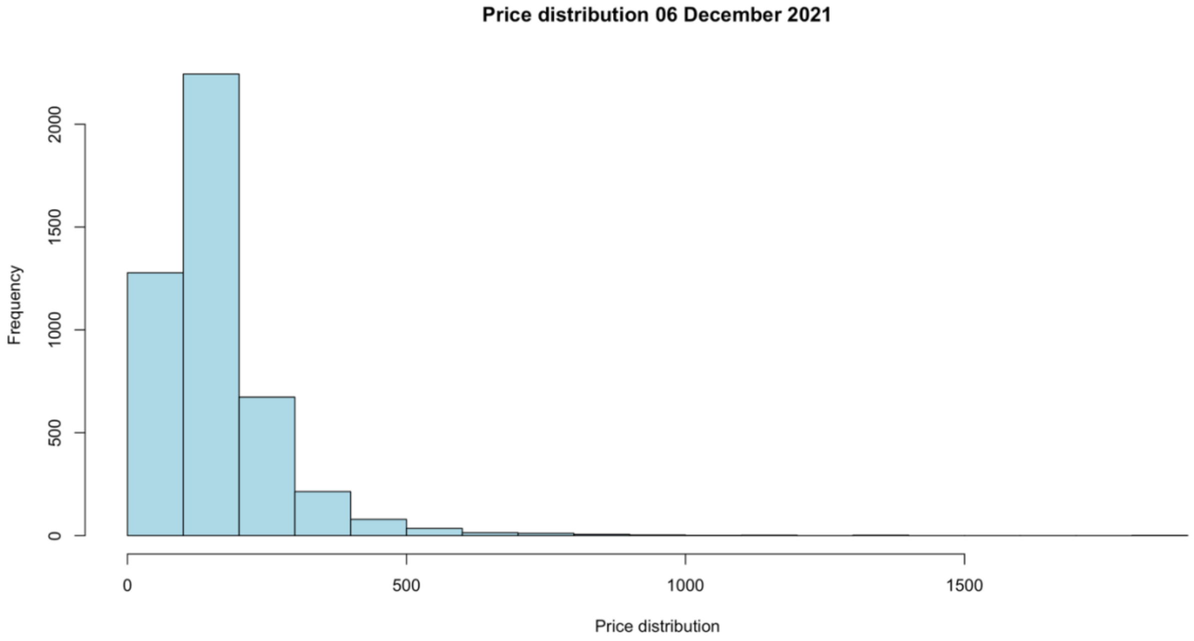


Figure 4: Price distribution December 2021

To explore a bit more in-depth the relation between price and room type, Figure 5 is introduced. Figure 7 illustrates, as the axes and the legend display, the distribution of the prices across the kinds of rooms available in the dataset. It can be inferred from the graph that the most expensive type of accommodation is a private room with an average price of almost 100$ per night. The x-axes are limited between (0,500) for visualization purposes.

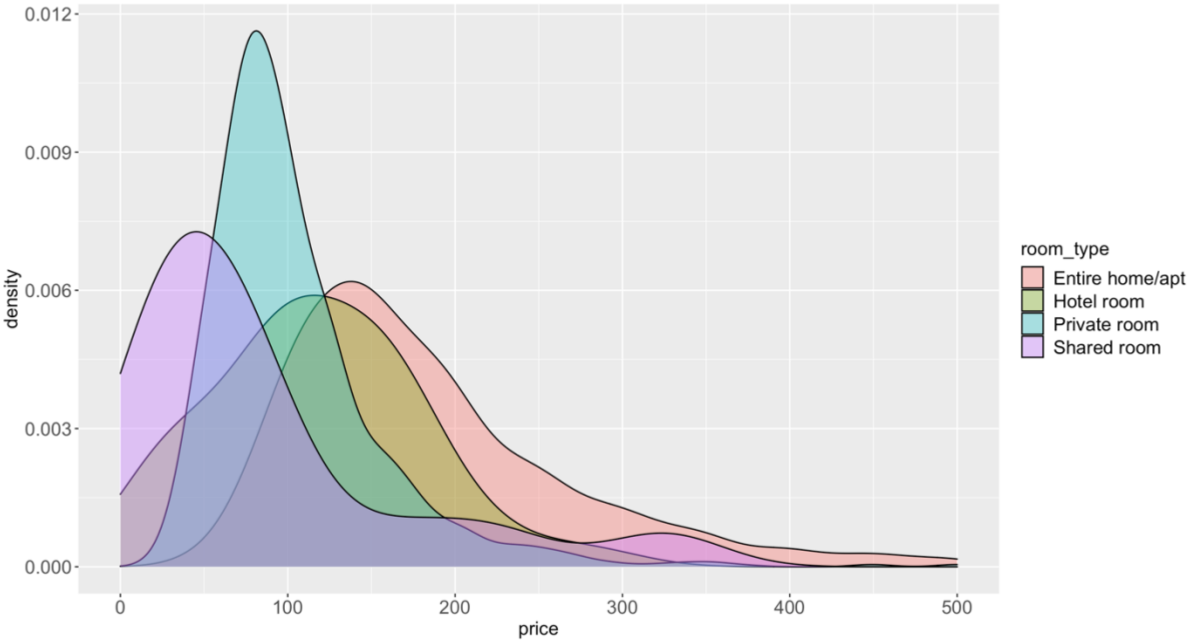


Figure 5: Price distribution across listings

4.3 Variables and measures

Ethos (credibility) in Airbnb

In this study, we are going to use four indicators of the host credibility: a super host badge, the number of reviews & ratings, and communication, ID verification.

Firstly, reviews can only be written by users who have stayed at an accommodation which significantly reduces the number of reviews written, limiting fake, random reviews and therefore can be considered a reliable indicator of the property and host's experience. Thus, the higher number of reviews in a place translates to a sign of popularity. However, both (Mingming and Xin, 2019), (Jun, 2020) studies remarkably note that users tend to base their evaluation review on previous experiences (sometimes past hotel stays). (Tussyadiah and Park, 2018) have found that the host's trustworthiness by potential guests is key to being successful in the sharing economy. It must be highlighted that guests may review the person with whom had been in contact (which is presented as the 'host'), but as mentioned in the literature review, this person might not always be the property owner.

Secondly, communication between strangers is quite common on the Airbnb platform. Host has 24 hours to accept or decline the guests' requests, but differences in time zones can slow down the process. The response rate (which is also needed to become Superhost) is calculated by dividing the responded requests within 24 hours (in the last 30 days) by the total number of inquiries that have been received. To become a Superhost, Airbnb takes the previous 365 days into account to calculate the response rate. Jun (2020) findings highlight that Airbnb customers are familiar with various Airbnb accommodations, and therefore it is highly probable to permit potential performance and physical risks when deciding to purchase. The motives for these risks can be well described by the unique travel motivations of the consumers. Tussyadiah and Park (2018) found that hosts who are represented as willing to meet/connect with new people are considered more trustworthy hosts depicted as world travelers, who are communicative and open and are more happily perceived than hosts that do not reveal personal information. Mingming and Xin (2019) underline that good communication is vital to building initial trust, which could also be accredited to Airbnb's stranger to the stranger transaction system.

Regarding Superhost, the badge is given to a host with outstanding hospitality skills and is considered an example and was first introduced in 2014. The badge is rewarded by an emoji next to the hostname for appearance (Airbnb). To earn a badge, it is essential to meet four specific criteria (Airbnb)

1. Completed at least ten trips or three reservations that total at least 100 nights;

2. Maintain a response rate of 90 percent or higher;

3. Maintain a 1 percent cancellation rate, excluding extenuating circumstances policy;

4. Maintained a 4.8 overall rating (based on the date the guest left a review, not the date they checked out over the past 365 days).

These requirements are evaluated every three months concerning the performance of the previous twelve months for all listings on your account. If the conditions are met, an automatic Superhost badge will be rewarded to the host and it will take up to one week for the badge to be visible (Airbnb). Liang et al. (2020) study verify the positive relationship between the "Superhost" badge and the review volume of the host's properties. The conclusions of the study from Fleischer and Ert (2019) mention that the Superhost badge is more important to host when there are fewer reputation hosts than the most popular. Furthermore, the previous study also found that Superhosts seem to receive a price premium owing to their rank between 4% - 6%. This suggests that hedonic products can be well explained by an ethos as a persuasive mode because hedonic products refer to images and messages but also experience.

Lastly, as far as the host verification is concerned, Airbnb tries to build trust in the community by providing two ways to verify the host's identity (Airbnb).

- A legal name and address;
- A photo of the host's government ID (driver's license, passport, identity card, or Visa);
- Sometimes a photo of the person and/or a profile photo.

To sum up, Ethos persuasive mode has been calculated using the host response rate, if the host has a Superhost badge and if the host identity is verified from Airbnb. The host response rate was initially given as a percentage, which was transformed to a scale from 0-10, and if a NA value was found, it was considered a zero-response rate. Moreover, if the host has a Superhost badge is given by a binary variable "host_is_superhost" as converted from the initial Boolean variable, and the same applies to host verified identity as well. Finally, 'host_identity_verified' contained the values 't' and 'f' which referred to True and False, respectively and were changed accordingly to be recognized as a Boolean variable.

Logos (Logical proof) in Airbnb

As previously mentioned, persuasion through logos refers to presenting logical arguments. In the Airbnb example, various accommodation features could be shown, such as the star rating, safety features, and the number of amenities.

Regarding star rating, guests usually pay attention to the reviews and the stars of the host. As mentioned in Airbnb(k), reviews and stars can be given for the following eight categories.

- Overall experience.
- Cleanliness.
- Accuracy.
- Check-in. It should be easy.
- Communication.
- Location.
- Value.
- Amenities.

(Guttentag, 2019) literature review on Airbnb appears that reviews/star ratings are a crucial feature of Airbnb as they build the trust between the guest and the host.

In addition, when a new place is added to the platform, the host must provide the amenities of the accommodation. Amenities are varied as many people apply filters in the pursuit of finding the right fit. A short sample of the amenities would be a pool, a kitchen, free parking, self-check-in, and so on. Amenities are found to increase the market price because consumers are willing to pay more for specific amenities (Xie, 2017).

Concerning safety features, which can be referred to as one of the most essential parts of the transactions, Airbnb has long been building trust throughout its marketplace. Each host can add six safety features in total. Airbnb promotes hosts who equip their listings with basic safety features such as smoke and carbon monoxide detectors, first aid kits, fire extinguishers, and bedroom locks, given that they also share that information (Han et al, 2019). Gibbs et al., (2017) found that physical characteristics, location, and host characteristics significantly impact price.

Utilitarian products nowadays need logical proof often to justify and pursue consumers. In sharing economy, customers searching for transactional goods are more inclined to pay a higher price for something more relevant than the absolute best choice (Shao and Li, 2020). That said, a mixture of logical proof can be the best combination to pursue new, potential, customers.

To elaborate further on how Logos is calculated in practice, some adjustments need to be made. Initially, the safety features “smoke alarm”, “first aid kit”, “carbon monoxide alarm”, “fire extinguisher” and “lock on bedroom door” were extracted from the amenities column and added up to create the “safety_features” variable. These amenities were also deleted from the initial “amenities” column to avoid duplicates, and the total number of amenities was calculated. Furthermore, the initial dataset contains reviews (on a scale of 1 to 5) for 6 categories (rating, check-in, cleanliness, communication, location, and scores value). These reviews were assigned to a variable named “reviews” with a sum divided by 6 to compute the average review score. Lastly, “*host_has_profile_pic”* contained values ‘t’ if true (host has profile pic) and ‘f’ if false, which were converted to 1 and 0, respectively.

Pathos (emotion) in Airbnb

I cover the methodology I used to calibrate ‘pathos’, i.e., the emotional appeal of a listing, in chapter 5, section 3, on sentiment analysis. In short, I measure the emotional words found in both the “*description”* and *“host_about”* columns to capture the valence of the hosts using emotion when describing their respective listing(s) or when writing about themselves. It has to be noted that both negative and positive sentiments are considered emotional.

5 . Methodology

This chapter displays the clarification of methods applied to answer the research question and the sub-question. I, the researcher, explain why topic modeling and regression analysis are suitable for the thesis. The implemented methods are discussed in more detail, including mathematical explanation, before presenting the results in the next chapter.

5.1 Topic Modeling

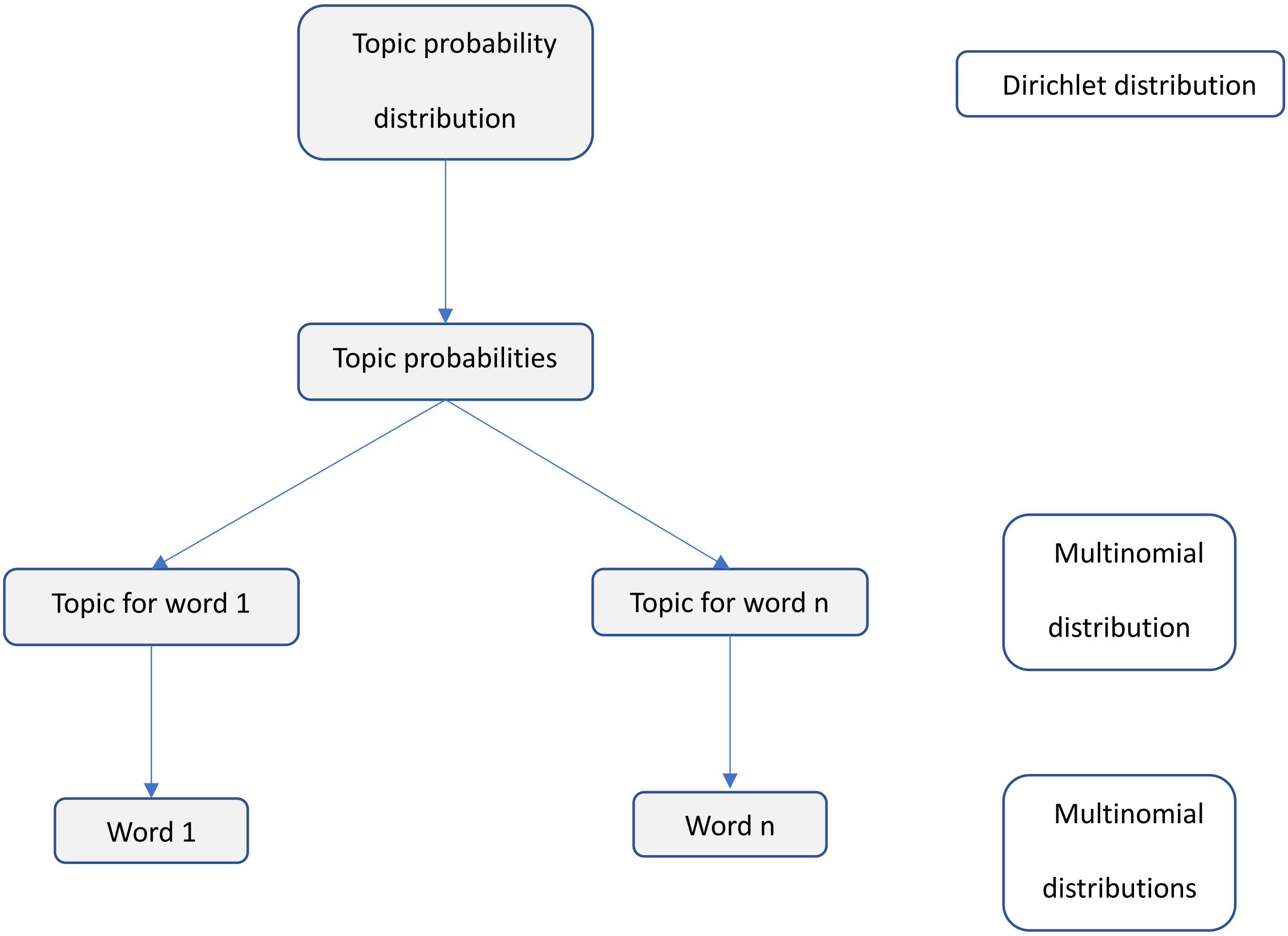
The thesis conducts a topic modeling approach to find the meaningful topics mentioned in the host's description ("About me" section) but also in the listing's description and users' comments to form a holistic picture of each listing. The goal of this analysis is to understand what topics and words are most frequently used by both users and hosts that will contribute to answering not only the main questions but also the sub-question: "Should a host adapt the selling message whether is it a hedonic or a transactional listing?" Topic modeling as mentioned in (Blei, 2012) is a text analysis method that reveals hidden information (themes) across a collection of documents. It is specifically mentioned that "algorithms for discovering the main themes that pervade a large and otherwise unstructured collection of documents" (Blei, 2012, p. 77).

Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) as introduced by (Blei et al., 2003a), (Blei et.al, 2003b) is a mixed membership model for text data, which will be referred to as LDA for abbreviation purposes. It is a generative probabilistic topic model that can be supervised (Mcauliffe & Blei, 2008) or unsupervised (Blei, 2012). It was initially developed for soft-clustering large quantities of textual data to find latent structures. The idea behind soft clustering is that a term can be presented to multiple topics, and a document can be about multiple topics. Compared to classic clustering methods in which membership (vector of continuous non-negative latent variables that add up to 1) is a binary, the schema of soft-clustering methods is that a word may partially belong to all topics, varying in probability. To elaborate further, a short summary of how the algorithm works is that observations (words) are grouped into documents and each document is modeled with a mixture of distributions. The mixture contains topics that are simply multinomial probability distributions over a fixed vocabulary (a recurring pattern of co-occurring words). The document and word order do not matter as a document is presented as a bag of words. As mentioned, the topics are shared across all documents, and the co-occurrence between words is measured. The goal is to discover the hidden topics in a collection of documents that cannot easily be found. A graphical illustration of the LDA process is presented below in Figure 8.

As illustrated from the graph (Figure 8), at the basis of the model that the corpus (C), which is a collection of documents, is made from a mixture of random latent topics. In other words, a corpus is a collection of elements called documents (**D**) and is of a total size M. Each document is a sequence of N words given by $\mathbf{w} = (\mathbf{w}_1, \mathbf{W}_2, \ldots., \mathbf{w}_\mathrm{N})$ where $\boldsymbol{w_N}$ is the $\boldsymbol{nth}$ word of the sequence. The words (as found in the basis of the above figure) are the basic units of the corpus and are defined to be elements from a vocabulary {1, ..., V} of size V. To further elaborate the above terms, a more tangible example is given below that can hold in real world context with the respective terminology.

Figure 6:LDA algorithm Illustration



Shall we want to take a book as the dataset in this case and implement LDA. The name of the book is *After you* (i.e., corpus - C) and contains 30 chapters (i.e., the documents – D) and it is written in English according to the Oxford English Dictionary (i.e., the vocabulary – V). It is a sequel that has been written by *Jojo Moyes*. If we try to implement LDA there is a critical concept that should be taken into consideration. The documents (**D**) are independent in most cases but in the scenario of a book this most likely will not be the case as the chapters are all chapters of the same book (however, books can contain independent chapters but not the one mentioned above). With the LDA algorithm, the researcher can unveil the most frequent topics used in the book but also the most frequent words across the 30 chapters. In addition, the words used in every topic can be found along with their probability of how likely a given word is. If needed, a close examination can be performed to find the most used context in which a certain word appears and to track changes (for example, in a history book, a researcher can investigate how the word 'currency' was used across the ages). As mentioned in (Blei, 2003), the generative process for each document (**D**) in a corpus (C) can summarized by:

- Choose the number of words Ni in the document (**D**i) that follows a Poisson distribution, Ni ~ Poisson (ξ).
- Choose θ_i ~ Dir(α), which is a topic distribution following a Dirichlet distribution. α can be denoted as the topic distribution for each document.
- For each of the Ni words win in the document **D**i:
  (a) Choose a topic z_{in} that follows a multinomial distribution, z_{in} ~ Multinomial (θ).
  (b) Choose a word w_n from $p(wn\,|zn,\beta)$, which is a multinomial probability conditioned on the topic z_n.

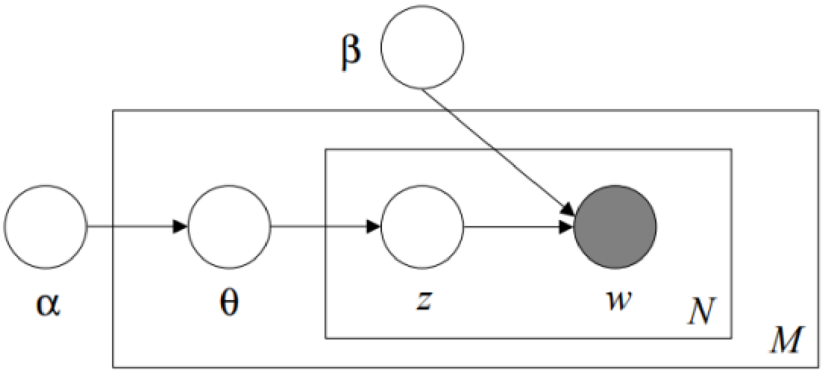


Figure 7: Graphical representation of the LDA model as given from Blei et al. (2003)

As we can see in figure 9 above, LDA involves a three-level model in contrast to the Dirichlet-multinomial clustering model, which consists of only two levels. In LDA the corpus parameters are α and β while θ are document level and z and w are word level variables. It must be noted that α and β parameters are sampled once in the process of corpus generation, θ parameters are sampled once per document while z and w are sampled once for each word in each document. The α and β parameters refer to the sparseness of topic probabilities and word probabilities respectively while low values of these parameters will produce documents that includes a small number of topics (or even a single prevalent) and topics that consists of a small number of words (or even a single prevalent one) Hyperparameters need to be selected prior to the application and are usually symmetric. The above formulas ensures that all words have the same likelihood to appear in a topic and all topics have the same likelihood to appear in a document.

Regarding the calculation of the posterior multivariate distributions there are two main methods most used by researchers namely, Expectation-Maximization (VEM) (Minka & Lafferty, 2002) and collapsed Gibb's sampling (Griffiths & Steyvers, 2004). The main difference is that Gibbs sampling is a Markov Chain (Monte Carlo) technique that uses a stochastic process to calculate and constantly update the parameters (sample from a probability distribution) while in VEM, parameters are computed with maximizing expectation (sample from a joint distribution). For this research, Gibbs sampling is used.

To explore the right set of parameters for the model, the parameters are initially trained using a part of the dataset (80%) and the model performance is evaluated on the remaining part (20%) of the dataset (dataset is spitted into an 80/20 sample as training and validation). To evaluate my findings perplexity measurement was used as a goodness of fit reference (along with human interpretation), albeit the fact that perplexity is found to be more focused more on the model complexity instead of interpretability, as it is still a good measurement for LDA (Cheng et al., 2019). The perplexity metric is calculated using the estimated probabilities of the words that appear in a set of documents. The model will constantly try to find the performance of assigning the high probabilities in the predication of a test set and this will be performed when the certainty is higher, and this is calculated by:

$$Perplexity(M) = P(w1, w2, \dots, w_N)^{-\frac{1}{N}}\ (1),$$

Whereas M is the model that this is applied to, P is the distribution of the total words (w) that occur in a set of documents (defined above as D) and N is the number of words. The probability is calculated by the log probability of total words (that appear in a document D), divided by the total number of words. A low perplexity value indicates that the model is less

complex and vice-versa, respectively and this will be used for the number of topics (defined as K) and α (the per-document topic distribution).

5.2 N-grams and Skip-grams

It goes without saying that humans (also in the Airbnb platform) write in a pretty unorganized format, so machines are challenged to find meaning from raw text. With N-grams plots, we can search for words that often are written together to explore some insights that may not be visible immediately; in other words, it is a commonly used term for a string of words. N-grams consist of at least one (single) word and can be visualized in various ways. Single word grams are called unigrams while two-word grams (two consecutive words) are called bigrams, three-word grams (three consecutive words) trigram, while when the number of words is more than three, the grams are referred to as four grams, five grams, and so on.

Under the hood, this method shares the same assumption with the bag-of-words approach and presents a unique characteristic of text. The most straightforward way to measure n-gram features is to use the most frequently used word or pair of words (N-grams) in the corpus. In most cases, so in this research, unigrams can be misleading or not easily interpretable due to specific sentiment (for example, "Amsterdam" is found to be one of the most frequently used words but does not imply how this word is used). Let me explain what is mentioned above with three examples by taking the sentence “I Love Airbnb”, illustrated in Table 5.

Table 4: N-gram example

Complete Sentence	“I Love Airbnb”
Unigram	[I] [Love] [Airbnb]
Bigram	[I Love] [Love Airbnb]
Trigram	[I Love Airbnb]

5.3 Sentiment Analysis

The next analysis that I apply is sentiment analysis. Although sentiment analysis's primary goal is to extract opinions from textual data while classifying text by their semantic orientation (Nasukawa & Yi, 2003), this research will mainly be focused on capturing the presence of emotions in the listing's description and creating a character variable that takes the value “Emotional” (for both negative and positive sentiments), if an emotion is found, otherwise take “Not Emotional” as a value. This information can be utilized to track brand and product attitudes in the Airbnb marketplace. The idea behind it is that guests are encouraged to proceed to purchase by reading appealing/effective content in the listing. The more detailed listing descriptions increase review volume and performance for property owners (Liang et al., 2020). Also, the width and depth of the description for hosts managing multiple places have a positive impact on review volume (Liang et al., 2020).

To implement this, various researchers have created dictionaries, such as the Linguistic Inquiry and Word Count (LIWC) or the National Research Council (NRC) Canada. For this research, I use the latter, the NRC dictionary as introduced in (Mohammad & Turney, 2010). It is a dictionary capturing emotions as a list of 14,182 English words (however, it has been found that a significant number of affective norms are stable across languages) which are also provided in over one hundred languages by using Google Translate.

Furthermore, NRC provides a score of 1 if a word has negative or positive sentiment or if the word is associated with at least one of the eight emotional categories, namely, anger, anticipation, disgust, fear, joy, sadness, surprise, trust and 0 otherwise. An example for a better explanation would be the word “angry” which results in a negative sentiment but also is associated with anger in comparison to the word “happy” which results in positive sentiment and is associated with joy. Thus, a sentence like “I'm angry at you, but I'm still happy” would score 1 in the anger and joy categories. Another example is presented below (Table 6).

Word	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
alive	0	1	0	0	1	0	1	0

Table 5: NRC Lexicon example

When it comes to pre-processing, it is of the highest importance not to stem words, mainly because the NRC dictionary cannot recognize stemmed words. Even some words that may be understandable from the NRC lexicon, such as “closeness” and “close” may provoke a completely different meaning. Additionally, as semantic, and syntactic processing is important when applying sentiment analysis, it is actually impossible to happen after the words have been stemmed. This leads to similar results, as mentioned before, with a completely different meaning than initially intended. These marks may reveal emotions such as urge or happiness (e.g., an exclamation mark) and are valuable.

5.4 Tobit Regression

The concluding and final step of this research is to measure the effect of the core and control variables on the occupancy rate. To estimate this, I use a Tobit model as introduced by Tobin in 1958, which can also be referred to as the censored regression model or, as Tobin liked to call it, the model of limited dependent variables. The model is best used to apply a regression when continuous variables are bounded (either lower or upper) at one of the extremes and presents a concentration of observations there (e.g., negatively, or positively skewed). It has to be noted that, if the variable is bounded between zero and one, it can only take values between zero and one (but also the extremes). When this is the case, the effect of explanatory variables tends to be non-linear, while variance decreases when the mean gets closer to one of the limits. In this research, the occupancy rate is negatively skewed because it is estimated with the number of reviews (of the last 12 months), and it is slightly different from the exact number of stays. Moreover, I capped the occupancy rate at one since there were instances where the occupancy rate exceeded one and required to be censored.

Tobit regression uses all observations, both those lying at the limits and those above it to estimate a regression. The standard Tobit model, as presented in (McDonald & Moffit, 1980) starts by the below stochastic model equation:

$$y_t = X_t\beta + u_t, \qquad t = 1, 2, .., N, (2)$$

Where N is the number of observations, and y_t is the dependent variable and X_t is a vector of independent variables, and β is a vector of unknown coefficients (coefficient estimates). Furthermore, u_t is a representation of the independently distributed error term and assumes that variance (σ^2) is constant, normal, and has a mean zero. Equation (2) holds when the model $(X_t\beta + u_t) > 0$ thus, the model assumes that there is an underlying stochastic index qualified as an unobserved, latent variable. The expected value of Ey (i.e., all observations) is given by the equation (3) below:

$$Ey = X\beta F(z) + \sigma f(z) \ (3),$$

Where, $z = X\beta/\sigma$, $f(z)$ refers to normal density (normal distribution) and $F(z)$ is the cumulative normal distribution function (excluding individual subscripts). Moreover, the expected value of observations that exceed the limit, $Ey *$, is estimated by $X\beta$ plus the expected value of the truncated normal error term. Equation (4) is represented by:

$$Ey^* = X\beta + \sigma f(z)/F(z) \ (4)$$

The truncated error term ($\sigma f(z)/F(z)$) is calculated in a similar way as the normally distributed error term, with the only difference that the variable is being bounded either below or/and above. As a result, the basic relationship between Ey, $Ey *\ and\ F(z)$ is given by:

$$Ey = F(z)Ey^{*} \ (5)$$

Therefore, Ey is affected by Ey^{*} and by $F(z)$.

To evaluate the models, McFadden's pseudo R^2 is used as introduced in McFadden (1973) and is defined by:

$$\frac{1-LLmod}{LL\mathbf{0}} \ (6)$$

, where $LLmod$ is the log likelihood value for the respective model and $LL0$ is the log likelihood for the null model that only considers the intercept as a predictor in order for every individual to be predicted with the same probability. Values from 0.2 to 0.4 indicates excellent fit as mentioned by McFadden. (1973)

6 Results

6.1 N-grams

In order to identify the latent topics that hosts refer to in their listing's description but also to inspect topics used in each host about section. The model input is a pre-processed corpus of documents (a bag of words). On top of the predefined text processing, a tokenized document was further transformed to an n-gram plot, specifically, bi, tri, and four-gram in order to reveal the most common words used in pairs which may not be highly informational when used separately. However, only bigram and trigram are presented in this analysis as unigram and four gram do not produce meaningful results. For instance, when examining the uni-gram, it was found that Amsterdam is the most frequent word used in the description section, but the way one uses this word could be interpreted in many different ways. The bigram below (Figure 8) displays the 30 most common terms used in a pair of two.

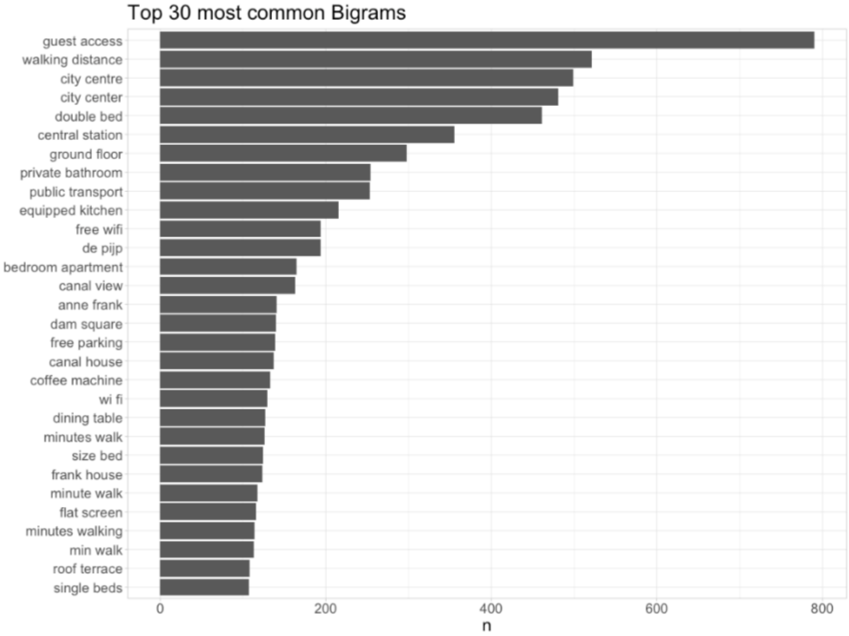


Figure 8: Bigram "description" column

Surprisingly enough, in a pair of two, “guest access” appears to be the most frequent duo, which stresses that potential guests may be interested in ‘access’ to the accommodation. The following pairs are referring to the distance from the center perhaps, or “walking distance” may even refer to a walking distance from a specific place of interest; however people seem to care less about single beds than location. A trigram is provided to enhance further our exploration of the dataset (Figure 9).

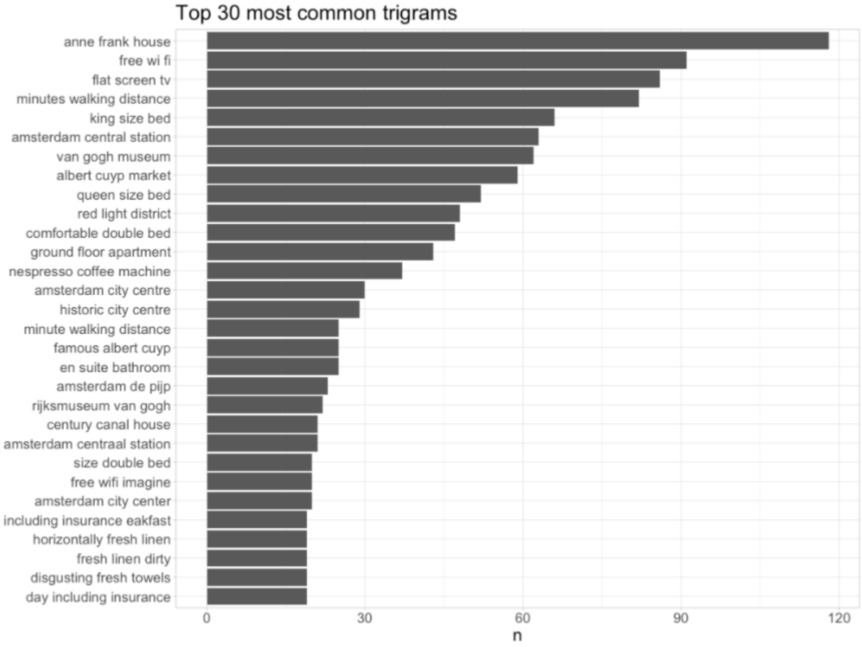


Figure 9: Trigram "description" column

Moreover, "anne frank house" is the most three-word pair used in the listing's description, followed by "free wifi", which is three words because of stemming, plays a vital role for consumers. On the other hand, hosts are not so frequently using words about towels and insurance, which may also imply that people do not put such matters as a top priority. In addition, it can be concluded that museums are more frequently used in the listings description than amenities.

Furthermore, the "*host_about"* bigram and trigram are introduced in the figures below (Figures 10, 11). While the most frequent bigram refers to "bridges houses" probably hints at sentences that refer to the view of Amsterdam, "yays serviced apartments" is the most frequent combination of three words used consecutively. The latter is written in a not formal and happy tone and may concern the services that the host provides to attract guests and build credibility.

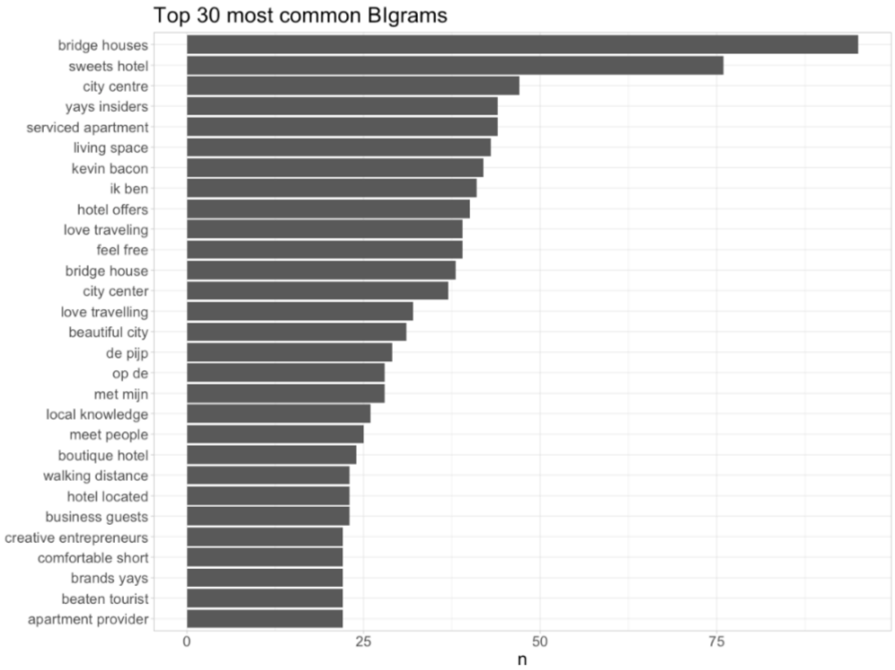


Figure 10: Bigram "host_about" column

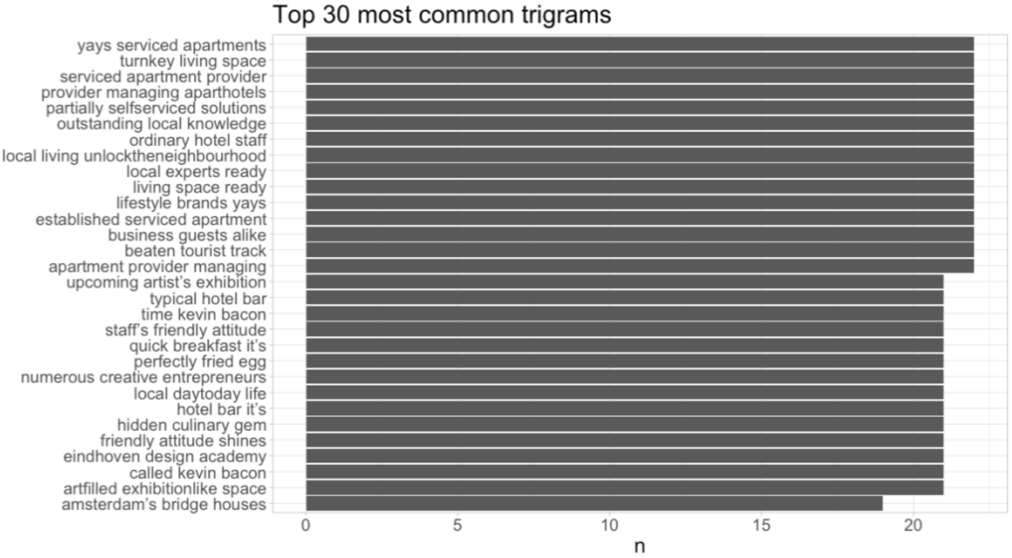


Figure 11: Trigram "host_about" column

To formulate a clearer picture of the dataset, the bigram and trigram of the reviews are also presented to examine what words reviewers use (Figure 12, 13). It must be noted that various reviews are not written in English (mostly French or Spanish); however, no English reviews have been transformed to NA value to be easier to interpret. As illustrated from the graphs, walking distance and museums are also quite frequent in the comments section meaning that guests care about these categories. Although specific words may not be used so frequently, “amazing location”, as expected, reviewers use words such as “highly recommend” and/or comment about the location of the listing, which is quite common when leaving a review in general, especially about accommodation.

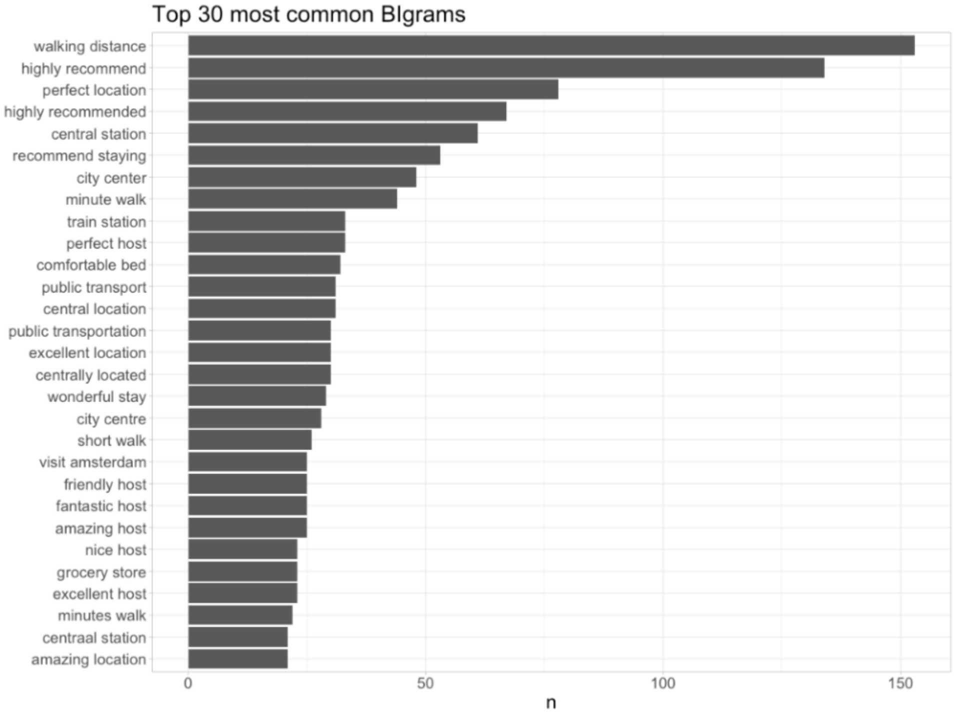


Figure 12: Bigram "comments" column

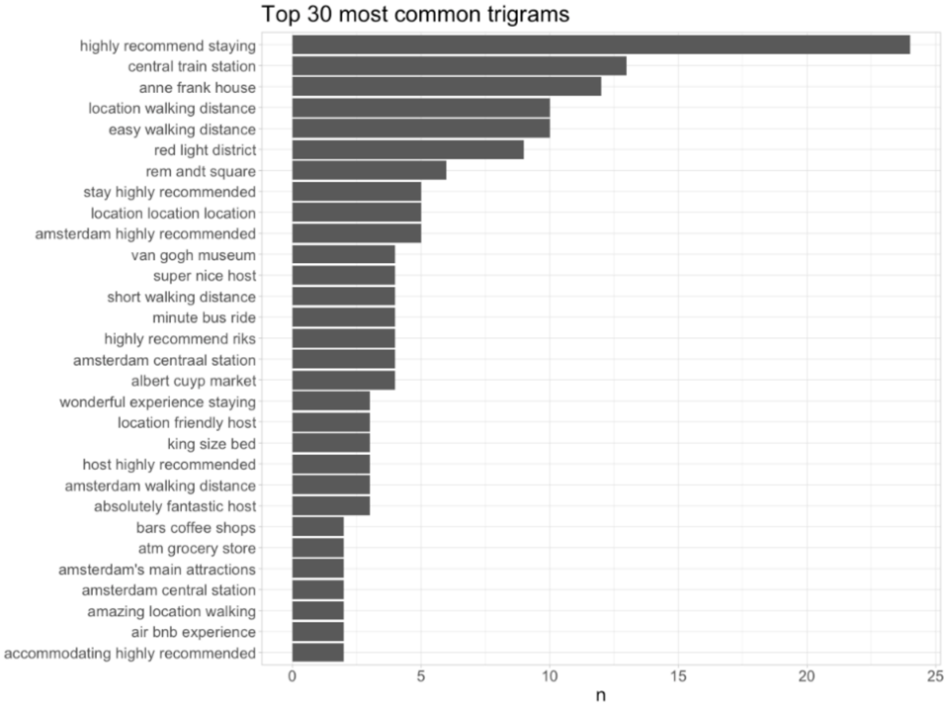


Figure 13: Trigram "comments" column

6.2 Latent Dirichlet Allocation

In this section, after showing the n-grams, I will present the LDA results. As mentioned in the N-grams section, LDA was performed on the "*host_about"* and "description" columns which refer to the host about and the description of each listing, respectively. The results for the "comments" column can be found in the appendix to understand the most frequent topics used by customers.

As far as Latent Dirichlet Allocation (LDA) is concerned, a critical step is to determine the parameters k (number of topics) and alpha (topic density) that will be used for the analysis. As mentioned in the methodology section, the number of topics is determined by the k hyperparameter but needs to be investigated carefully before deciding the final number. If the value of k is small compared to a proper number of topics, then the model will not provide meaningful results as it will be too extemporaneous. On the other hand, if the value of k is larger than an acceptable number of topics, then the model will be very complex and may not be interpretable. The same intuition is used to calibrate the parameter alpha.

There are multiple ways to determine the number of topics, but for this research, the perplexity score will be used as a measure of fit. Although it is generally accepted that a lower perplexity score of the model leads to a better fit, the results must be examined. This also holds for the alpha hyparameter.

The following figure (Figure 14) displays the hyperparameter k to the respective perplexity score when researching the textual data of the description column (description of each listing). The values of k are in a sequence from 10 to 50 by a step of 10. As illustrated from the graph, the perplexity score decreases as k (number of topics) increases which means that the model is less perplexed (complex). Perplexity is the lowest for a k = 30 in both training and validation samples, suggesting accepting 30 topics. Using 30 as the number of topics, the same model was trained on different alpha values (α), which is illustrated in the graph (Figure 25), and alpha 0.5 was selected.

Heading over to the "*host_about*" column, regarding the train set, there are pretty small differences in the perplexity value (Figure 15) when k is between 20 and 30. However, this is best explained by the validation set, in which, for k =30, perplexity is lower. Based on the given results and human interpretation, k = 30 was selected. Additionally, in the plot (Figure 26) found in the appendix, it is found that perplexity scores are proportional to alpha (increasing as alpha increases). The best α (alpha) value is 0.5 as the perplexity is found to be the lowest.

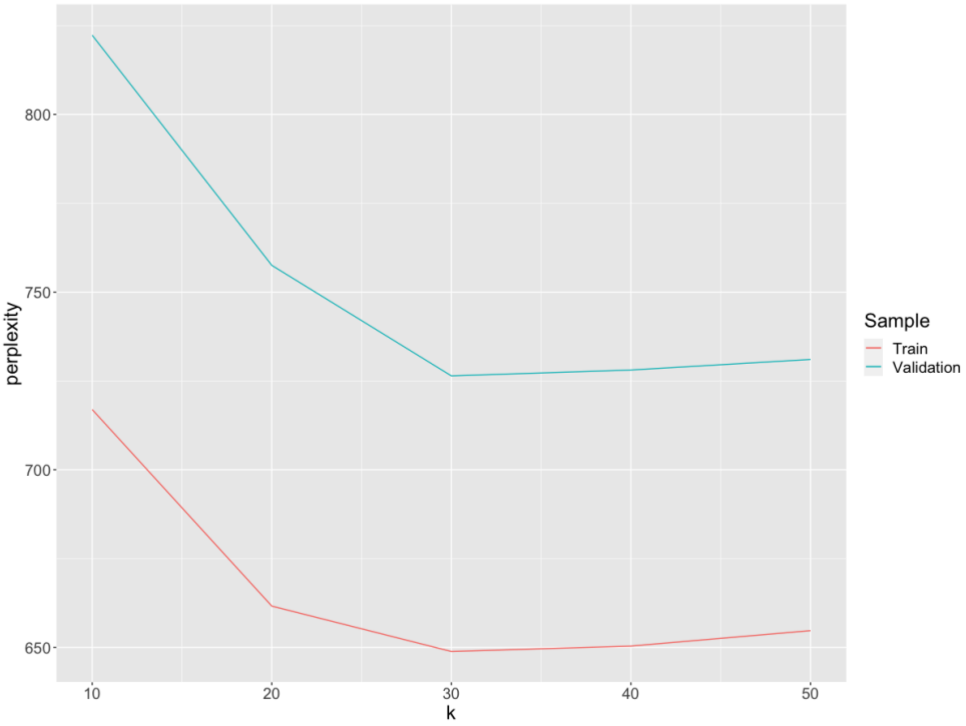


Figure 14: Perplexity plot for the “description” column

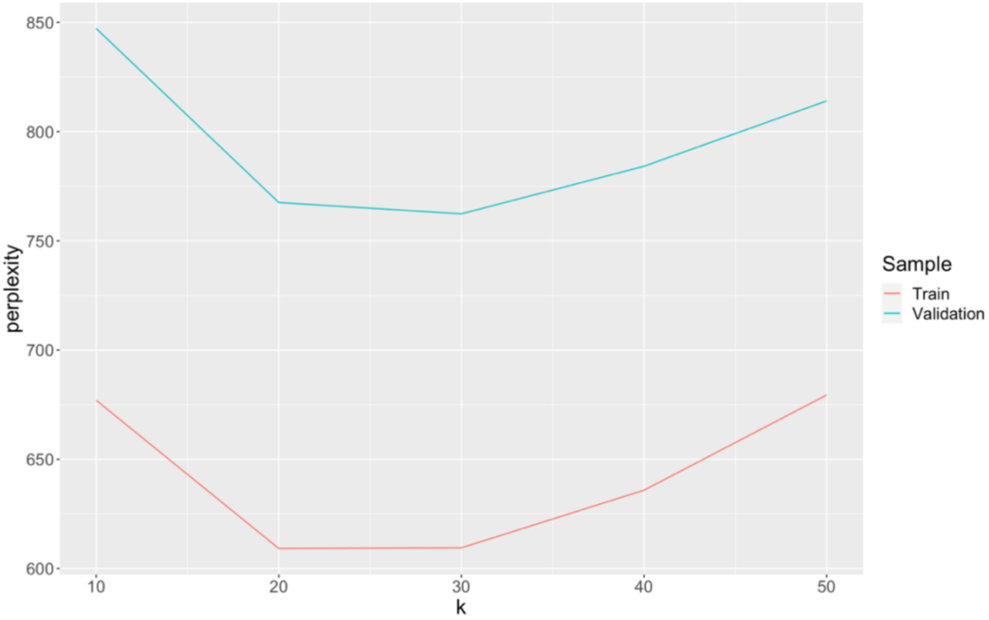


Figure 15: Perplexity plot for the "host_about" column

Concerning the “description” column, the hyperparameters used to apply LDA k = 30 and a = 0.5, had an output of 30 topics. The top 10 topics are presented below (Table 6), and the complete list can be found in the appendix. The figures consist of the top 10 words that contribute to each topic. The topics can contain words that may be misleading at times, so it may be difficult always to rename or accurately rename the topics, so we will only, if applicable, label only the top 5 topics. In our case, Topic 1 can be labeled as “Amenities of the apartment” as it contains various words about it. In contrast, topic 2 can be labeled as “Property type in Amsterdam city center”. In addition, Topic 3 contains words that refer to museums and neighborhoods. Hence, it is more about “Location,” and topic 4 contains terms that may refer to “Access and Location of the property.” Lastly, topic 5 is quite hard to label but implies “Amenities and comfortable living”. As depicted in the table below, terms about amenities of the respective listing are the most used words by hosts in the description section which makes total sense as this is the goal of the description section. It should be noted that, as mentioned earlier, the “description” column is a concatenation of the “name” column that contains the title of each listing and the description. Although we don’t know the topics and the words coming from the text that is considered emotional, we see that naming the amenities and abstractly describing the location of the listing, are on top of the list.

Turning over to the “host_about” section (Table 7), it is evident that topics are not so similar to the ones found in the description section, albeit the fact that words such as “amsterdam” or “room” are present in most of the topics. Topic 1 can be labeled as “House introduction”, while topic 2 contains information about “City”. Topic 3 can be renamed “Offerings around the area” and topic 4 “Travel and work”. Lastly, topic 5 is about “what to visit in Amsterdam”. The words used by hosts to describe themselves contain words to describe the city of Amsterdam or the listing they offer but also themselves. Unlike “*description”*, “*host_about”* contains text which is focused more on the welcoming part and also the attractions that guests may find if they potential rent the host’s listings (e.g., museum or historic). It is essential to highlight that stopwords are entirely erased from this column to avoid words such as “I” for interpretation purposes. These results may be further researched on a deeper level to extract the relation between words and personalities.

Table 6:LDA Results "description" column top 10 topics

Topic 1	**Topic 2**	**Topic 3**	**Topic 4**	**Topic 5**	**Topic 6**	**Topic 7**	**Topic 8**	**Topic 9**	**Topic 10**
"floor"	"amsterdam"	"canal"	"apartment"	"apartment"	"room"	"ship"	"bridge"	"apartment"	"amsterdam"
"room"	"city"	"house"	"amsterdam"	"amsterdam"	"city"	"room"	"house"	"will"	"blicense"
"private"	"room"	"amsterdam"	"room"	"bed"	"spacebbr"	"amsterdam"	"amsterdam"	"stay"	"numberbbr"
"apartment"	"private"	"jordaan"	"bed"	"bedroom"	"things"	"spacebbr"	"hotel"	"living"	"spacebbr"
"terrace"	"house"	"walking"	"double"	"kitchen"	"can"	"sailing"	"architect"	"amsterdam"	"apartment"
"roof"	"centre"	"view"	"spacebbr"	"can"	"accessbbr"	"guests"	"sweets"	"offer"	"close"
"house"	"will"	"distance"	"bathroom"	"comfortable"	"bother"	"space"	"architectural"	"bedroom"	"renovated"
"bathroom"	"suite"	"apartment"	"private"	"modern"	"bguest"	"can"	"spacebbr"	"personal"	"modern"
"kitchen"	"can"	"anne"	"bguest"	"living"	"bed"	"apartment"	"new"	"every"	"stay"
"two"	"stay"	"famous"	"accessbbr"	"spacebbr"	"amsterdam"	"private"	"interior"	"room"	"bathroom"

Table 7: LDA Results “host_about” top 10 topics

Topic 1	**Topic 2**	**Topic 3**	**Topic 4**	**Topic 5**	**Topic 6**	**Topic 7**	**Topic 8**	**Topic 9**	**Topic 10**
"amsterdam"	"love"	"stay"	"amsterdam"	"amsterdam"	"amsterdam"	"amsterdam"	"amsterdam"	"amsterdam"	"amsterdam"
"house"	"people"	"amsterdam"	"enjoy"	"house"	"love"	"love"	"love"	"love"	"love"
"welcome"	"travel"	"apartment"	"love"	"city"	"live"	"enjoy"	"can"	"years"	"living"
"city"	"amsterdam"	"will"	"living"	"famous"	"family"	"live"	"living"	"travel"	"years"
"beautiful"	"city"	"offer"	"years"	"live"	"work"	"years"	"years"	"city"	"travel"
"place"	"world"	"hotel"	"work"	"bars"	"house"	"city"	"city"	"born"	"working"
"centre"	"new"	"personal"	"house"	"historic"	"name"	"stay"	"best"	"apartment"	"places"
"love"	"food"	"best"	"much"	"amsterdams"	"life"	"airbnb"	"people"	"living"	"time"
"live"	"around"	"museum"	"travelling"	"building"	"two"	"can"	"will"	"feel"	"new"
"unique"	"good"	"area"	"live"	"monumental"	"kids"	"born"	"old"	"share"	"old"

6.3 Sentiment Analysis

I researched the emotions of "description", "*host_about*" columns using the NRC lexicon to discover whether the host uses emotional words to construct the "*Ethos*" variable, which is based on the persuasion through emotions as mentioned in the literature review. The same analysis has also been applied to the "*comments"* column for understanding purposes, and the results can be found in Appendix B. Stopwords are removed for the frequency histograms along with numbers but only for the visualization of the frequency histogram of words contributing to both, the real sentiment of each column and for each (positive and negative) sentiment respectively. To highlight this, I have only kept English words for this research. A frequency polygon regarding the "description" column is presented below (Figure 16).

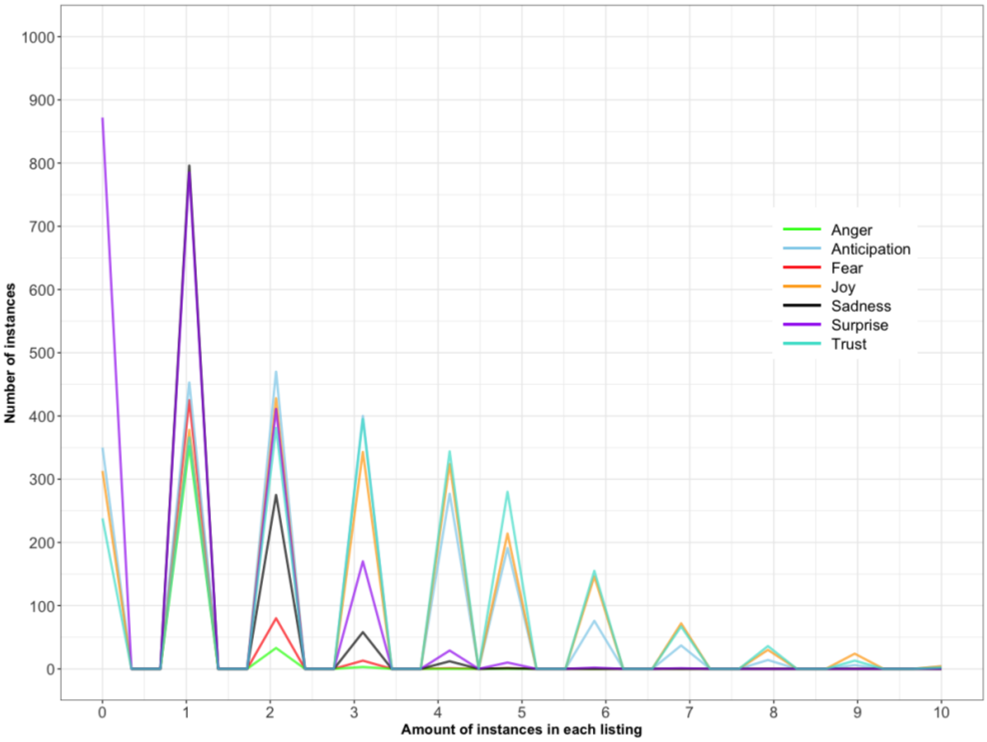


Figure 16:Frequency polygon of emotions "description" column zoomed.

On the y-axes, the count of each emotion is plotted, while in the x-axes, the number of words described like the respective emotions is present. Although "*Surprise*" is prevalent when comparing a single word, "*anticipation*" takes the lead when comparing two words and "*trust*" for three words. This is critical because hosts use words that provoke anticipation for the description of their listings less than trust (when comparing emotions found only once in a listing). Fear, on the other hand, when compared in a set of a single word still, is quite frequent, which implies that many of the words used in the description of the listings are provoking “*anger*”. A descriptive statistics table is presented below (Table 8), including the mean, min, max, and standard deviation (sd).

Variable	Mean	SD	Min	Max
anger	0.1872807	0.4348927	0	3
anticipation	2.455702	1.869307	0	11
disgust	0.352193	0.547117	0	3
fear	0.2776316	0.5649011	0	5
joy	2.872807	2.141251	0	12
sadness	0.6899123	0.8246339	0	5
surprise	1.009649	1.041017	0	7
trust	3.047368	2.053143	0	13

Table 8: Descriptive statistics of emotions "description" column

The emotion analyzed consists of 2.280 observations. The mean score of each emotion is presented in the table, for instance, anger has an average score of 0.1872807, which can be interpreted as found to be more than one time in each listing’s description in comparison to disgust which has an average score of 0.352193 implying that are various descriptions which do not contain words that provoke “*disgust*”. The highest expressed emotion overall is “*trust*”, at a value of 3 words per description on average, which can be backed up by the frequency polygon if the highest counts are added. This is followed by “*joy*” which is present in almost 3 words of each listing’s description. These are quite favorable results because consumers want to read descriptions that provoke “*positive*” emotions rather than “*negative*”.

Additionally, the words that contribute to the total of sentiment of the “*description”* column are presented in the histogram below (Figure 17).

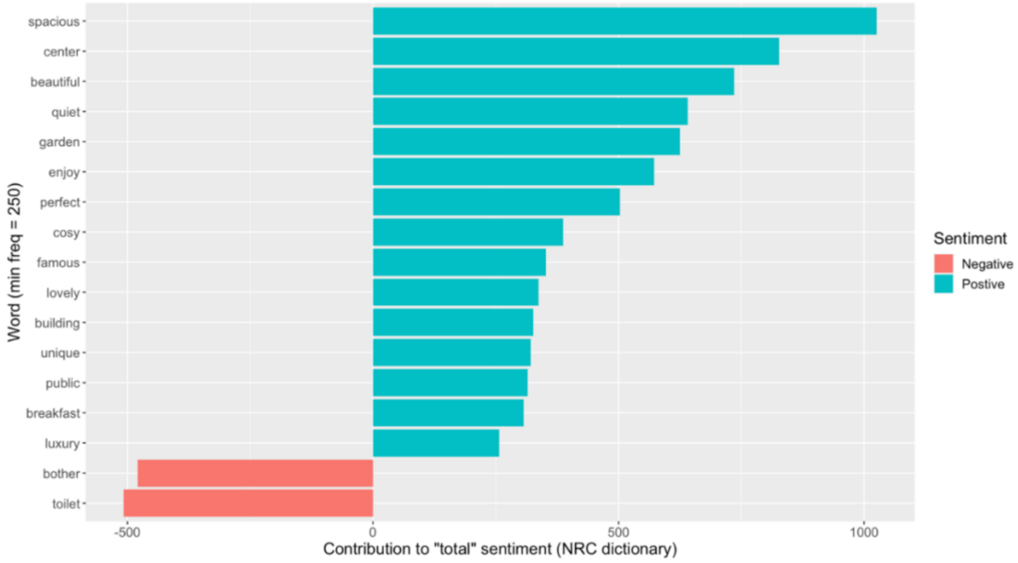


Figure 17: Overall sentiment contribution of words "description" column

Although that there are only two negative words that contribute to the total sentiment the most, there are various positive words with the most frequent being "*spacious*" and "center" implying that hosts may describe their listings to have wider space for the potential guests and are centrally located. Remarkably, the word "*building*" is used positively more than 250 times inferring the characteristics of the building the apart is actually in to pursue guests. Finally, luxury is on the bottom of this list suggesting that hosts do not use the luxury aspect of their listings oftentimes when compared to the location. Furthermore, I have plotted, in histograms, the frequent words for the "*description*" column. To be more precise, I have plotted the difference in words frequencies between positive and negative words because some words (e.g., "Amsterdam") which are really frequent can bias the results as when they are present in a sentence such as, "Although Amsterdam is dirty" but also in "Apartment in Amsterdam center" which are negative and positive sentences, respectively. These non-common words are referred to as 'specific' words and can be found in Figure 18,19.

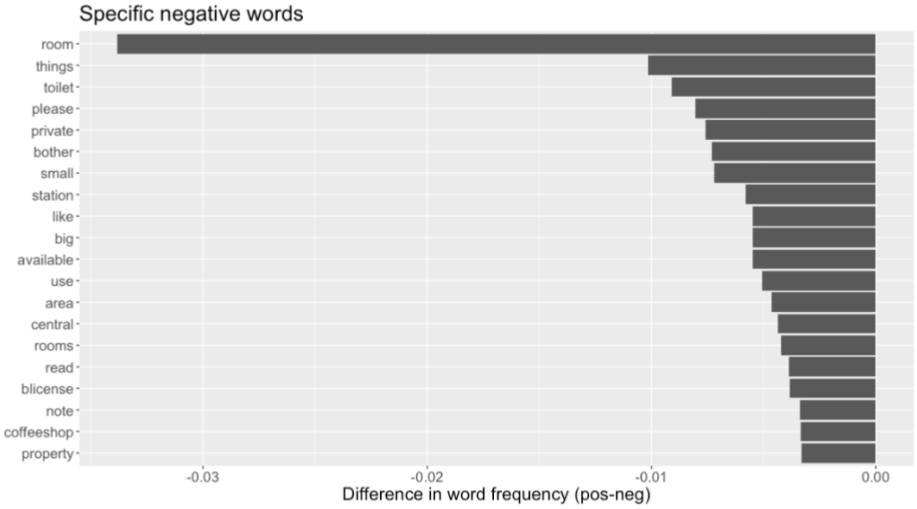


Figure 18: Specific negative words of the "description" column

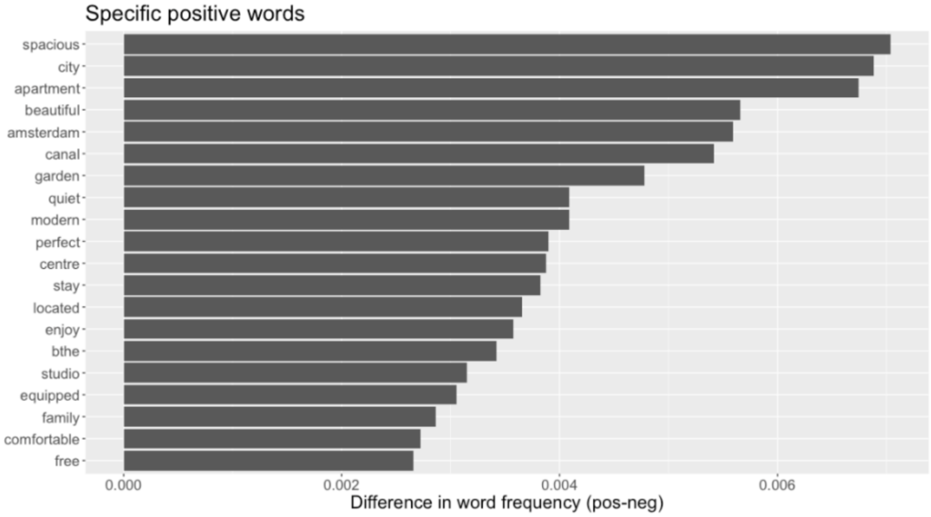


Figure 19: Specific positive words of the "description" column

As far as the "*host_about*" column is concerned, the same investigation was performed to measure the frequency of emotions and the descriptive statistics of emotions for the "*host_about"* column (Figure 20). From the graph, it is visible that "*Surprise*" is actually followed by "*anticipation*" when comparing a single set of words before dropping to the third place, when comparing a set of two words. Unlike the results of the "*description*" column, words expressing "*anger*" are not frequently found the "*host_about*" column as it is found to be the least frequent emotion. This is crucial because we can see a clear difference in the use of "*negative*" emotions between the "*host_about*" and "description" columns, with negative emotions being less frequent in the former. A descriptive statistics table is presented further below (Table 10) to investigate how emotions are used in the "*host_about*" column overall.

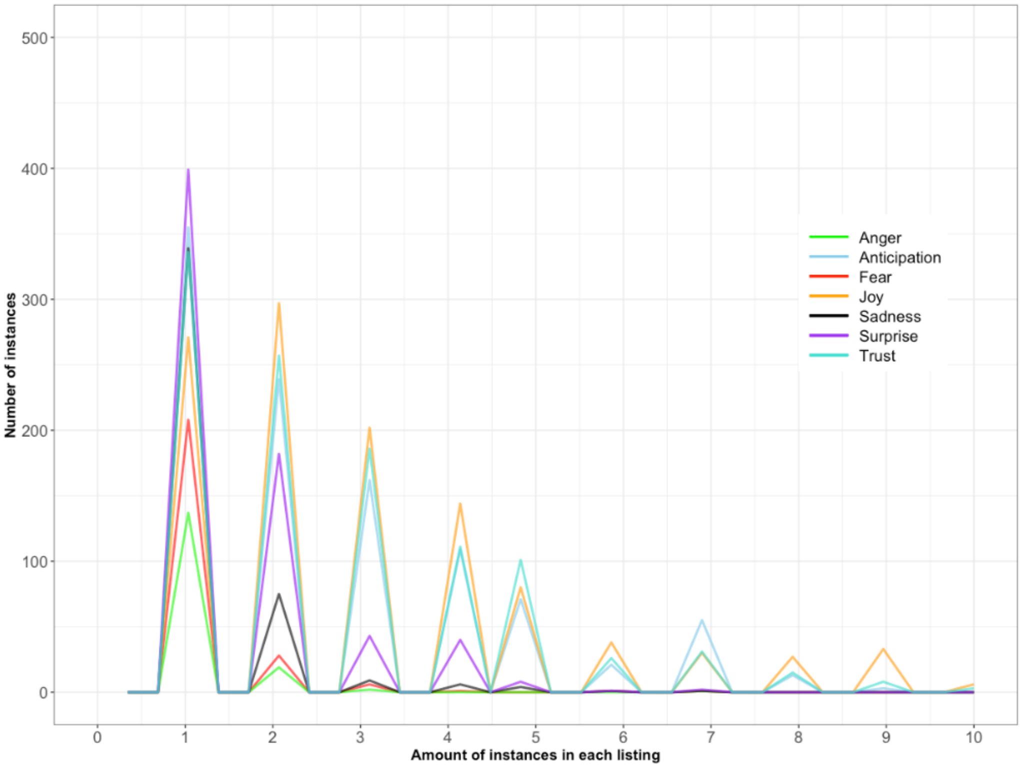


Figure 20:Frequency polygon of emotions "host_about" column zoomed

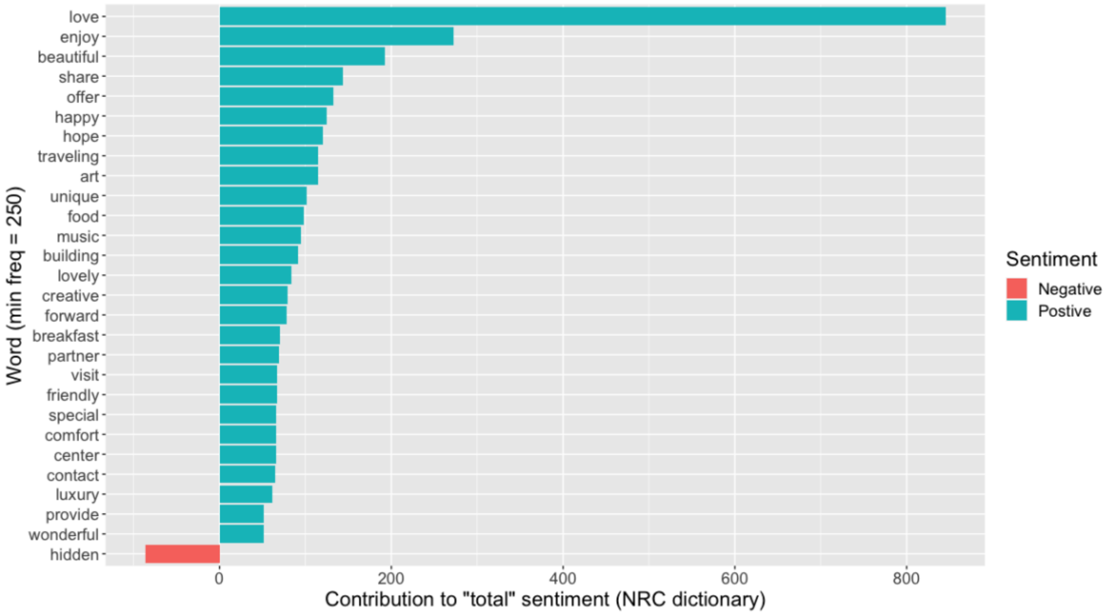
Albeit the fact that values are lower, even half, when compared to the values found in the "*description*" column, "*negative*" emotions such as "anger" or "*disgust*" are found at minimum levels. Surprisingly, "*joy*" is the most prevalent emotion expressed overall, followed by "*trust*" which are precisely the reserved results of the "description" column. This implies that when hosts write about themselves, they strive to use more words that bring "*joy*" to the reader than "surprise" to build a connection between the host and potential customers. For example, "*daughter*" is a word that is marked by the NRC lexicon expressing "*joy*" in comparison to "*credible*" describing "trust. The words that contribute to the total sentiment of the “*host_about*” column are presented in the figure below (Figure 21).

Variable	Mean	SD	Min	Max
anger	0.08245614	0.3406742	0	7
anticipation	1.235088	1.89292	0	15
disgust	0.08464912	0.3394901	0	6
fear	0.1311404	0.4372864	0	7
joy	1.56886	2.219937	0	13
sadness	0.2513158	0.6132213	0	7
surprise	0.4877193	0.9239764	0	7
trust	1.357895	2.015295	0	15

Table 9:Descriptive statistics of emotions "host_about" column

It is visible that most words contributing to the total sentiment are positive, which implies that hosts use way more positive words than negative words to describe themselves, which is no surprise. However, the word that contributes the most to the total sentiment is “*love*” followed by “enjoy” describing a host that expresses love and enjoyment. Over and above that, as shown for the “*description”* column, the specific negative and positive words are given in figures 20, and 21, respectively. As illustrated from the plots, the terms “*possible”* and *“amsterdam”* are the most frequent specific words used in negative and positive sentences, respectively. These words are frequently used in persuasive texts and can be formulated as an example like “Calling is not possible” or “I live in the beautiful city of Amsterdam” respectively. Overall, the words plotted in figures 22, 23 below are not formal and provoke a friendly tone.

Figure 21: Overall sentiment contribution of words "host_about" column



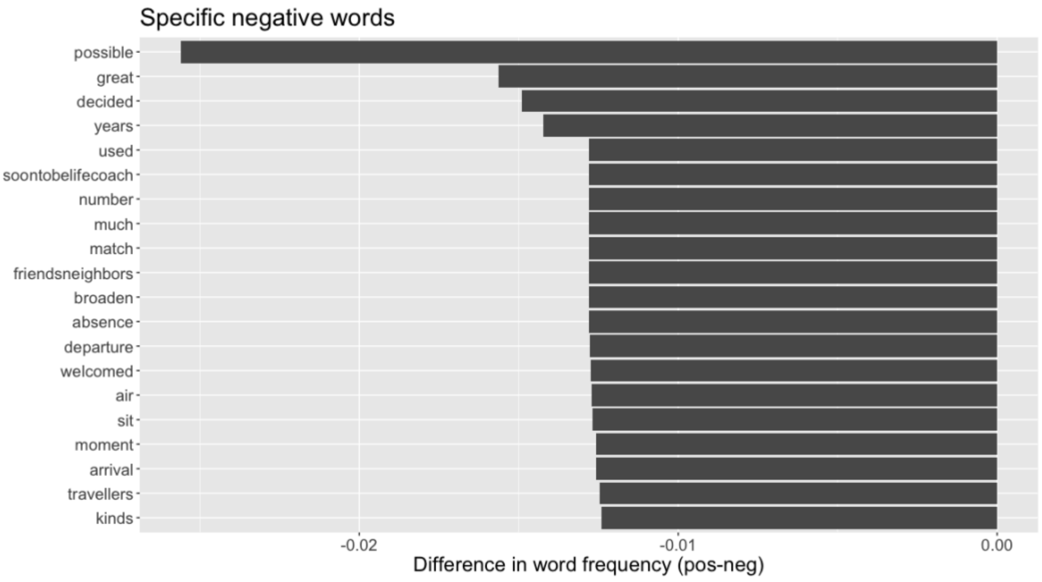


Figure 22: Specific negative words of the "host_about" column

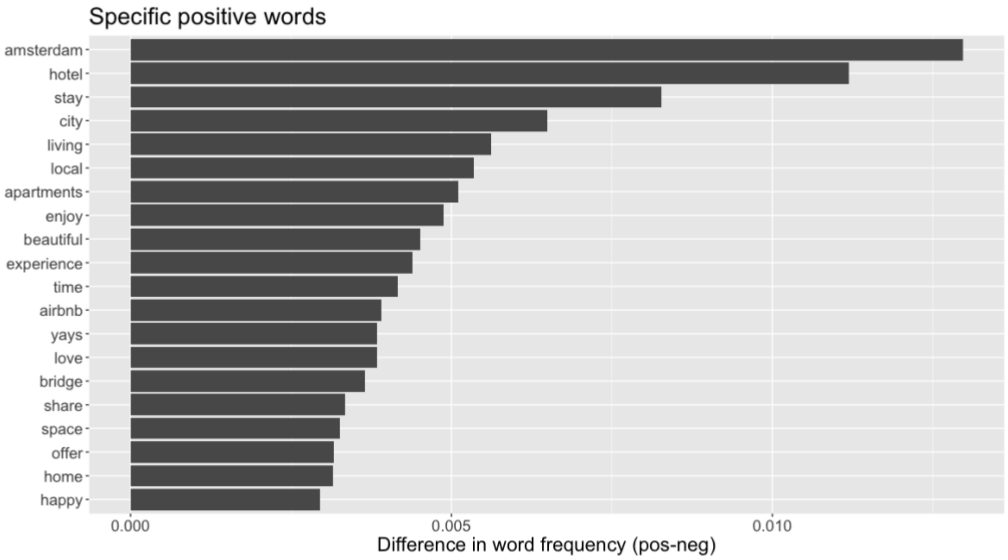


Figure 23:Specific positive words of the "host_about" column

To conclude, the NRC lexicon also provides two additional columns regarding sentiment (negative and positive). Each of these columns contains the amount of either negative or positive words, respectively, which are contained in each listing's description or host about. I have calculated the overall sentiment by subtracting the value of negative sentiment from the positive sentiment for each listing. The distribution of these scores is presented in the next figure (Figure 24). It is clear that the overall positive sentiment is higher when hosts write about themselves rather than when writing about their listing's description. More specifically, the overall positive score for the "*description"* column is 15991, while for the "*host_about"* column is only 6810, almost 2.5 times less. In addition, the overall negative sentiment scores are 2637 for the "*description"* column and only 698 for the "*host_about* column, which means that hosts use a significantly smaller number of negative words when describing themselves. However, such scores are completely logical because the description length is about 500 words in contrast to the host bio's, which are less than 150 words.

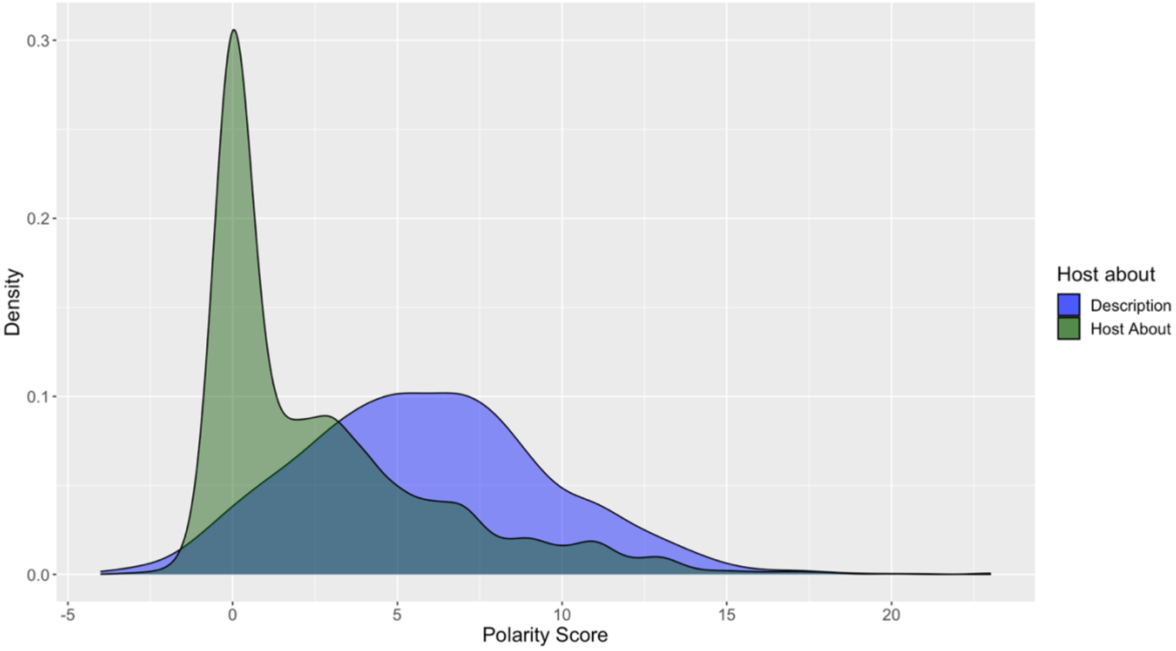


Figure 24: Distribution of overall sentiment scores

6.4 Tobit Regression

In this section, the results of the Tobit regression are presented. Firstly, a descriptive statistics table (Table 10) is introduced, including the descriptive statistics of the core variables used in the regression analysis. Next, Tobit's regression table results (Table 11) are interpreted and explained.

The variables in table 10 are summarized by the three persuasive modes, namely, Logos (logical proof), Pathos (emotion), and Ethos (credibility). To dive into the specifics of the statistics presented, firstly, the average of occupancy rate is 0.181which translates to 66 days per year that a listing is occupied. Additionally, the average number of ratings (out of the 6 categories given) is 4.7 out of 5, which is quite high and could not be a truthful estimate. Regarding Pathos (Emotion), the description column contains 2190 listings that contain emotional words and only 90 listings whose description does not contain emotional words. On the other hand, 1304 hosts about bio listings contain emotional words and 976 are not written in an emotional manner, which implies that the sentiments are varied when hosts write about themselves. However, emotional texts are referred both the negative and positive sentiments of the texts per listing. Lastly, the average response rate is 0.79 out of 1 (or 79% in percentage), showing that hosts are quite responsive to potential guests, and the majority of the hosts have also verified their identity (85%). It is also noticeable that a little over 1 out of three hosts are actually awarded the Superhost badge.

Descriptive Statistics

Statistic	Mean	St. Dev.	Min	Max	Frequency
occupancy_rate	0.181	0.260	0	1	2280
Logos (logical proof)					
safety_features	2.871	1.398	0	5	2280
reviews	4.769	0.244	1	5	2280
host_has_profile_pic	0.999	0.030	0	1	2280
nr_amenities	50.920	27.182	4	170	2280
Pathos (emotion)					
description_pol					
Emotional	0.96	0.960	0	1	2190
Not Emotional	0.0394	0.194	0	1	90
host_about_pol					
Emotional	0.57	0.19	0	1	1304
Not Emotional	0.42	0.49	0	1	976
Ethos (credibility)					
host_response_rate	0.790	0.374	0	1	2280
number_of_reviews	76.136	112.932	1	901	2280
host_identity_verified	0.856	0.351	0	1	2280
host_is_superhost	0.335	0.472	0	1	2280

Table 10: Descriptive statistics of core variables

	Coefficient	St. Error	z value
Constant	0.146	(0.205)	0.716
Core Variables			
Logos (Logical proof)			
safety_features	0.002	(0.004)	0.011
reviews	-0.044 *	(0.022)	-1.939
host_has_profile_pic	0.005	(0.172)	0.035
nr_amenities	0.0004 **	(0.0002)	2.171
Pathos (Emotion)			
description_pol Emotional	0.022	(0.028)	0.833
host_about_pol Emotional	-0.034 ***	(0.011)	-3.139
Ethos (credibility)			
host_response_rate	0.088 ***	(0.015)	5.908
number_of_reviews	0.001 ***	(0.0001)	21.064
host_is_superhost	0.007	(0.012)	0.635
host_identity_verified	0.042 ***	(0.016)	2.678
Control variables			
price	-0.0003 ***	(0.0001)	-5.025
instant_bookable	0.152 ***	(0.012)	13.096
centre_dis	-0.00002 ***	(0.00000)	-7.395
room_typeHotel room	-0.076 **	(0.034)	-2.316
room_typePrivate room	0.046 ***	(0.014)	3.437
room_typeShared room	-0.045	(0.088)	-0.494
logSigma	-1.418 ***	(0.017)	-81.805
Notes: ***p<0.001; **p<0.05; *p<0.01, Log Likelihood = -427.594, Pseudo R2 = 0.5603168, Observations = 2280			

Table 11: Tobit regression results

Lastly, the Tobit model presented above contains all the core and control variables to estimate the overall effect of the three persuasive theories, namely, Logos (logical proof), Pathos (emotion), and Ethos (credibility), on the dependent variable, the occupancy rate. Concerning the Tobit regression, the left limit is 0, and the right limit is 1 for the censored dependent variable (occupancy rate). To evaluate the goodness-of-fit, the log(scale) (also commonly as logSigma) along with the PseudoR2 are taken into consideration. The latter is most commonly known as McFadden's R squared which is calculated as I mentioned in Methodology section. In this model, the PseudoR2 has a value of 0.55, which suggests a good model fit corresponding to the null model. However, to look beyond that value and PseudoR2, the Log Sigma (or Log(scale), the name depends on the software the researcher uses or the package) is considered to measure the goodness-of-fit. Log Sigma is analogous to the logarithmic standard deviation of the residuals, and it is an absolute measure. To interpret the value given in the table (-1.416), the Log Sigma has to be exponentiated and takes the value 0.242, which, in other words, is the difference, on average, between the actual occupancy rate and this mode's estimates and implies that the model performs well.

In regard to the variables and specifically to the three persuasion theories, the results are quite insightful. Initially, in Logos (logical proof), the estimated coefficient of "*nr_amenities*" is positive and statistically significant at significance (using p < 0.05). This implies that if the number of amenities increases by 1 amenity, then the expected occupancy rate increase by 0.0004 and suggests that hosts that have more amenities will have a higher occupancy rate than those who do not. In addition, the "*reviews*" (average star reviews for each listing, among 6 categories, namely, ratings, check-in, cleanliness, communication, location, and value) coefficient is found to be statistically significant and has a negative effect on occupancy rate per star rating (average number of reviews is 4.7, the range is from 1 to 5), specifically 0.044 (p < 0.1), that mean if the number of stars, per listing, increases by 4.7 (which is the average) then occupancy rate will decrease by 0.044. This is valuable because apparently, it follows the logic of "quality over quantity" which translates to quality reviews (and to an extent more truthful) having a better outcome on the occupancy rate than mass highly starred reviews. Surprisingly, the variable "*safety_features*" is not statistically significant. This may be due to the fact that The Netherlands is a very safe country to live in and, to an extent, Dutch people are very friendly, not violent, and feel quite safe overall, with the majority of houses not having even curtains.

Furthermore, as observed in the Data section, most of the listings are centrally located, which can also play a role in this relationship between safety features and occupancy rate as there is less chance of burglary. Moreover, safety features contain alarms and fire extinguishers, but people may not have such attributes at the top of their priority list when renting a listing on Airbnb. Although the literature suggests otherwise that hosts having a profile picture may influence guests, the regression results provide that having a profile picture as a host does not influence the occupancy rate. Additionally, as for the host about bio, writing an emotional text affects the occupancy rate negatively as the coefficient is statistically significant (p <0.01). Specifically, the occupancy rate is 0.034 lower (p < 0.01) for emotional texts written in the host about section.

As far as Pathos (emotion) is concerned, as provided in the regression output above, it surprisingly shows that persuasion through emotion (which is affective content to the reader) in the description, in contrast to the literature, does not affect the occupancy rate. The variable *"description_pol"* contains two values, namely, "Emotional" and "Not Emotional". This is fascinating because, as illustrated in the results of the sentiment analysis, specific frequent words between description, and host about, are quite different from each other apart from the fact that description has a limit of 500 words in contrast to the host's bio which is mostly found to be under 150 words. This implies that potential guests, may are triggered to look deeper than the description of the listing, in the host about section, and/or are looking to read less text to decide.

Moreover, the coefficient of emotional text found in the host about section is statistically significant (p < 0.001) and has a negative effect on the occupancy rate. It should be noted that *"host_about_pol"* contains the same values as the *"description_pol"*. In the host about section, texts that do not contain emotional words (or are not "provoking emotion") are to the advantage of the host since the occupancy rate is 0.034 (p < 0.001) lower for emotional texts. This can be due to the fact that "Emotional" is attributed to both *"host_about_pol", "description_pol"* variables when the text is found to be emotional (either having positive or negative sentiment). In the host about section, 1827 out of 2280 observations were found to have negative sentiment, which is the majority of the listings, and it may influence the results significantly in comparison the description of the listings in which observations with negative sentiment are only about a fifth of the total specifically, 453 that contribute to the total sentiment. These results may imply that people do not want to read bios provoking negative emotions (such as anger or disgust, for instance), which is backed up by literature as potential guests are looking for experience and hosts that can be trusted or vice-versa for the listings description.

Lastly, regarding Ethos (credibility), both "*host_response_rate"* and "*number_of_reviews*" affect the occupancy rate positively and are statistically significant (p < 0.01), suggesting that a host that takes more time to respond and has more reviews in total, in the respective listings, tends to have an occupancy rate that is higher than those who deliver fast response and lack of reviews. Precisely, the occupancy rate tends to be 0.088 higher when the host replies faster, and, the occupancy rate, increases by 0.001 when the number of reviews is higher per listing. This may be due to the fact that people may ask important questions about the listing to the hosts. Thus, potential guests may value hosts that generally reply in less time than hosts who delay replying because oftentimes, people need an immediate answer and are more satisfied when the other party is fast and responsive, which is also backed up by literature. Also, that confirms the findings found in Logos, that occupancy rate is affected negatively by, on average, the higher number of review stars proposing that the occupancy rate of a listing may increase if the listing contains a higher number of reviews but contains a lower number of stars (the average number of stars in this research is 4.7, which is relatively high). Remarkably, the Superhost badge does not affect the occupancy rate in contrast to "*host_identity_verified"* of which, the coefficient is positive and statistically significant (p < 0.01). A verified host has an occupancy rate that is 0.042 (p < 0.01) higher as compared to the hosts who are not verified.

Nevertheless, there are also interesting results provided regarding the control variables. The variables "*price",*" *centre_dis*" are statistically significant (p<0.001) and have a negative effect on the occupancy rate. Both the magnitude (effect) of the coefficients of "*price",*" *centre_dis*" on the occupancy rate are small (0.0003 and 0.00002, respectively), which suggests that if these variables are increased, the occupancy rate will be slightly decreased. The affects effects can be due to the shortage of listings in Amsterdam (and housing in general) but also because the data are scraped from the 6th of December, which is a tourist period. The negatively effect of distance on the occupancy rate means that as we deviate from the center ("*centre_dis*" increases*)*, per listing, its occupancy rate tends to be lower and that could be because a significant amount of guests prefer to live in center. *"room_typeHotel room"* has also a negative effect on the occupancy rate, bigger than the previous variables (0.076), and its coefficient is statistically significant (p < 0.05). That suggests, potential guests are not interested in renting listings that fall under the "Hotel room" category but want to rent a private room as the coefficient of the variable "*room_typePrivate room*" is statistically significant (p < 0.001) and has a smaller but positive effect on the occupancy rate. Lastly, "*instant_bookable*" has the largest positive effect of all variables on the occupancy rate (0.152), and its coefficient is statistically significant (p < 0.001). This can be explained by the rejection of guests from the hosts to their rental requests, due to the enormous amount of demand in Amsterdam (especially in December). Hosts may deny potential guests' rental requests to avoid parties, mass gatherings, mass sleepovers, and other unwanted (to the hosts) activities making the "*instant_bookable*" a special feature in the marketplace.

All in all, the results are quite surprising especially for the three persuasive modes. When Logos (logical proof), Pathos (emotion), Ethos (credibility) are compared, the results are varied. Initially, Logos (logical proof) has the overall bigger negative effect, among the three modes of persuasion, on the occupancy rate as guests are less interested in higher review stars per listing than a higher number of amenities. On the other hand, Ethos (credibility) has the higher, stronger, positive effect on the occupancy rate by significant margin between the three modes. If a host's responses are fast and thus, have a higher response rate, the occupancy rate tends to be much higher than having a higher number of stars, more amenities and an emotional about me bio, combined. Similarly, the verified identity of the host, has a higher effect on the occupancy rate than having an emotional bio. Pathos (emotion) has a negative effect on the occupancy rate expressing that hosts that contain emotional words when describing themselves may face a lower occupancy rate in their listings.

7 Conclusion

The aim of this research was to investigate how hosts can effectively persuade the decision-making process, and, finally, the potential guests, to book the provided accommodation. For this, the following research questions were formulated: “Depending on the type of product, how can we optimize contact to sell the right property with the right message?” along with the following sub-question: “Should a host adapt the selling message whether is it a hedonic or a transactional listing?”. In this paper, I investigated the effects of the three persuasive modes as introduced by Aristoteles in Rhetoric I, II, and III, namely, Logos (Logical proof), Pathos (Emotion), and Ethos (Credibility), by implementing Latent Dirichlet Allocation (LDA), Sentiment analysis and Tobit regression. These modes were carefully matched with specific variables to calibrate the effect of the three persuasive theories on the occupancy rate while accounting for various control variables. The results display that Ethos (Credibility), by a significant margin, has the most effect on the occupancy rate while Logos (Logical proof) has the least.

Existing literature focuses on the effects on price or the impact of Airbnb on the sharing economy in general, rather than a particular performance measurement like the occupancy rate, but also some papers investigate the role of persuasion on the sharing economy. Thus, this research provides meaningful insights into understanding more in-depth the decision-making process, calculated by the occupancy rate, in the sharing economy, while optimizing contact in order to sell the right property to the right audience. It is important to be highlighted that this paper contributes to the existing literature in a unique way as it only analyzes listings that are highly available and have at least one review to conclude in more accurate results.

In chapter 3, I have defined certain hypotheses (H1, H2, H3, H4a, H4b, H4c) regarding each persuasion mode but also concerning the type of the property (hedonic or utilitarian). As I mentioned before, Ethos (Credibility) is shown to be the most effective persuasion theory for the occupancy rate, indicating that persuasion through Ethos (Credibility) can have a significant impact on the number of bookings. Hence, H1 is accepted, suggesting that properties from a credible host have a higher impact on the occupancy rate than listings offered from a less credible host. Furthermore, H2, refers to properties that contain logical proof, (i.e., persuasion through Logos) is rejected because, as presented in the regression results, persuasion through Logos has an overall negative effect on the occupancy rate. Also, quite surprisingly, H3 is rejected as listings that contain emotional words tend to have a negative impact overall on the occupancy rate than listings that do not. However, the strength of these three hypotheses is dependent on the type of listing (hedonic vs utilitarian). Hedonic products are mostly related to Ethos (Credibility) and Pathos (Emotion) rather than Logos (Logical proof), which is matched with utilitarian products, and that is what H4 concerns. As displayed in the regression’s outcome, only Ethos (Credibility) has an overall significant positive effect on the occupancy rate thus, only H4a can be accepted while H4b and H4c are rejected.

Before diving into the specifics of each persuasive mode but also the significant control variables, a partition in the implementation of these persuasion strategies should be applied. Unfortunately, Ethos (credibility) requires more time to be constructed than the other two persuasion modes, and thus, a separation between short-term and long-term strategies should be clarified. I urge hosts to take to their advantage the strategies that can be applied in a short amount of time, namely, Pathos (Emotion) and Logos (Logical proof) by, for instance, writing an emotional listing description to attract potential guests interested in hedonic products or upgrading the amenities that the listing offer for utilitarian listings. However, hosts should also be responsive, identified, and ask customers to review their stay in order to increase their number of reviews and be awarded as Superhosts but also to be benefited from what Ethos (Credibility) has to offer in the long term. To achieve both the short-term and the long-term strategies in less time, short can also lower the prices of their listings (to acquire more reservations, for example), limit the fees, offer more services, and most importantly, make the accommodation instantly bookable. It has to be highlighted, however, that hosts need to focus more on features that yield higher results in a shorter amount of time so even though offering the listing as instantly bookable has a positive effect on the occupancy rate, being responsive to messages of potential guests has a higher positive effect and thus, may require more attention.

As far as the texts are concerned, albeit the fact that containing emotional words in the host about section has a negative effect on the occupancy rate, most of the texts used in this analysis have a higher negative than positive sentiment and thus, suggestions can only be implemented for negative emotional texts. Sentiment analysis provides meaningful results for both the "*host_about"* and "comments" columns on the specific negative words, as it is clear that both figures (Figure 23 & Figure 40) contain identical words. I recommend that hosts, not to be words such as "possible", "decided, "years", or "used" as these words may dissociate readers and lead to a lower occupancy rate. However, using these words in another way (more positively) may provide completely different results and influence the reader. To optimize contact, I urge hosts to clarify the type of listing they are offering (also their audience) and describe the listings and themselves based on the respectively positively significant features, in a way that provokes a positive sentiment (by using some of the 'specific' positive words for example). Undoubtedly, hosts are required to adapt their message and how their description (host about, listing) based on the type of accommodation provided. For instance, a host that offers a utilitarian product can have a bio describing more of the apartments offered, their location in a positive way but also the safety and amenities provided for these accommodations. On the other hand, a bad example would be: "After a great number of years, I decided to leave The Netherlands to broaden my horizon to soon become a life coach, but all kinds of travelers are welcomed".

7.1 Limitations and Future Research

This thesis contains some limitations which can be further searched and improved. Initially, this research is based on a limited amount of information on a specific date of the year (December 6th, 2021) in the city of Amsterdam. Most of the listings are quite centrally located, so variables such as the distance from the center cannot deviate to a big extent constraining the research. Furthermore, as December is a touristic month, especially in Amsterdam, hosts may change their prices and/or their services along with the amenities as Christmas is getting closer and demand increases rapidly. If the occupancy rate is affected by such changes, as the number of listings increases, it can result in misleading results. A further investigation across seasons but also across countries (or even cities) could provide remarkable results and increase the external validity of this research.

Additionally, regarding sentiment analysis, most of the observations in the "*host_about*" section are classified as 'negative' thus, limiting the results, while for the "*description*" column, the total sentiment is varied. On top of that, the processed text analyzed was selected to be only in English, censoring the analysis of Dutch text. The most difficult matter to predict in such algorithms is sarcasm or irony, which can tweak the results. Also, sentiment analysis has been performed based on only lexicon (dictionary) and various words of the dictionary may be classified in another way in different lexicons, so I propose that a comparison could be applied and compared to yield interesting results as further research. Some might argue that making a lexicon is perhaps the best way, but it's quite unrealistic to hold in real-life analysis. *Inside Airbnb* has a limit on the amount of text scraped, which sometimes can influence results, especially in algorithms like sentiment analysis. I recommend scraping textual data from Airbnb and concatenating them based on the listing ID.

Lastly, regarding the regression analysis, as it is widely accepted, control variables are never enough. However, in this research, variables like access to the listing, and parking spaces (and for what vehicle) could be investigated to derive meaningful results. Moreover, other variables that influence the decision-making process could be further investigated to explore the performance in the sharing economy. Lastly, I urge to classify listings as hedonic, or utilitarian based on specific attributes and to deploy choice models to address the heterogeneity between consumers.

References

Airbnb(a), “How to describe your space - resource center”, (accessed 20 March 2022), [available at https://www.airbnb.com/resources/hosting-homes/a/how-to-describe-your-space-308]

Airbnb(b), “How to write an appealing description”, (accessed 20 March 2022), [available at https://www.airbnb.com/resources/hosting-homes/a/how-to-write-an-appealing-description-369]

Airbnb (c), “I rent out my home in Amsterdam. What short-term rental laws apply?”, (accessed 05 June 2022), [available at https://www.airbnb.com/help/article/1624/i-rent-out-my-home-in-amsterdam-what-shortterm-rental-laws-apply]

Airbnb (d), “Night limits in Amsterdam and London: Frequently asked questions”, (accessed 05 June 2022), [available at https://www.airbnb.com/help/article/1628/night-limits-in-amsterdam-and-london-frequently-asked-questions]

Airbnb(e), “It's a good time to update your listing with these tips - resource center”, (accessed 28March, 2022), [available at https://www.airbnb.com/resources/hosting-homes/a/its-a-good-time-to-update-your-listing-with-these-tips-161]

Airbnb(f), "New Study: Airbnb Community Makes Amsterdam Economy Stronger”, (accessed 30 May 2022), [available at https://www.airbnb.nl/press/news/new-study-airbnb-community-makes-amsterdam-economy-stronger]

Airbnb(g), “Rent out my Home in Amsterdam. What Short-Term Rental Laws Apply?”, (accessed 20 March 2022), [available at https://www.airbnb.com/help/article/1624/i-rent-out-my-home-in-amsterdam-what-shortterm-rental-laws-apply]

Airbnb(h), “Sprucing up your listing description”, (accessed 21 March 2022), [available at https://www.airbnb.com/resources/hosting-homes/a/sprucing-up-your-listing-description-13]

Airbnb(i), “What are response rate and response time and how are they calculated?”, (accessed 28 May 2022), [available at https://www.airbnb.com/help/article/430/what-are-response-rate-and-response-time-and-how-are-they-calculated]

Airbnb(j), “What makes Superhosts so “super?”, (accessed 21 March 2022), [available at https://www.airbnb.com/resources/hosting-homes/a/what-makes-superhosts-so-super-56]

Airbnb(k), “About Superhosts”, (accessed 28 May 2022), [available at https://www.airbnb.com/help/article/828/about-superhosts]

Airbnb(l), “How to become a Superhost”, (accessed 28 May 2022), [available at https://www.airbnb.com/help/article/829/how-to-become-a-superhost]

Airbnb(m), “Verifying your identity”, (accessed 28 May 2022), [available at https://www.airbnb.com/help/article/1237/verifying-your-identity]

Airbnb(n), “Star Ratings”, (accessed 29 May 2022), [available at https://www.airbnb.com/help/article/1257/star-ratings]

Airbnb(o), “The amenities guests want”, (accessed 30 May 2022), [available at https://www.airbnb.com/resources/hosting-homes/a/the-amenities-guests-want-25]

Bansal, Harvir S., P. Gregory Irving, and Shirley F. Taylor (2004), “A three-component model of the customer to service providers”, *Journal of the Academy of Marketing Science*, 32, July, 234-250.

Bardhi, Fleura, and Giana M. Eckhardt (2012), “Access-based consumption: The case of car sharing”, *Journal of consumer research*, 39, December, 881-898.

Blei, David M., Andrew Y. Ng and Michael I. Jordan (2003), “Latent Dirichlet Allocation”, *Journal of Machine Learning Research*, 3, January, 993-1022.

Blei, David M., and Michael I. Jordan (2003b), “Modeling annotated data", *In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, July, 127-134.

Brown, Stephen, Chris Hackley, Shelby D. Hunt, Charles Marsh, Nicholas O’Shaughnessy, Barbara J. Phillips, David Tonks, Chris Miles, and Tomas Nilsson (2018), “Marketing (as) Rhetoric: paradigms, provocations, and perspectives”, *Journal of Marketing Management*, 34, October, 1336-1378.

Chang, Jonathan, Sean Gerrish, Chong Wang, Jordan Boyd-Graber, and David Blei, (2009), "Reading tea leaves: How humans interpret topic models" *Advances in neural information processing systems,* 22.

Chen, Yong and Karen Xie (2017), "Consumer valuation of Airbnb listings: A hedonic pricing approach", *International Journal of contemporary hospitality management,* 29, September, 2405-2424.

Cheng, Mingming and Xin Jin (2019), "What do Airbnb users care about? An analysis of online review comments", *International Journal of Hospitality Management,* 76, January, 58-70.

Conger, Jay A. (1998), "The Necessary Art of Persuasion," *Harvard Business Review,* 76, 84-97.

Cook, Ian A., Clay Warren, Sarah K. Pajot, David Schairer, and Andrew F. Leuchter (2011), "Regional brain activation with advertising images", *Journal of Neuroscience Psychology and Economics,* 4, 147.

Dellaert, Benedict GC (2019), "The Consumer Production Journey: Marketing to Consumers as Co-Producers in the Sharing Economy*", Journal of the Academy of Marketing Science,* 47, March, 238–54.

Dogru, Tarik and Osman Pekin (2017), "What do guests value most in Airbnb accommodations? An application of the hedonic pricing approach", *Boston Hospitality Review,* 5, Spring, 1-10.

Driver, Julia (2014), "The History of Utilitarianism", (accessed 16 May, 2022), [available at https://plato.stanford.edu/].

Ert, Eyal, Aliza Fleischer, and Nathan Magen (2016), "Trust and reputation in the sharing economy: The role of personal photos in Airbnb", *Tourism Management*, 55, August, 62-73.

Ert, Eyal, and Aliza Fleischer (2019), "The evolution of trust in Airbnb: A case of home rental", *Annals of Tourism Research*, 75, March, 279-287.

Farrell, Thomas B. (1976), "Knowledge, consensus, and rhetorical theory", *Quarterly Journal of Speech*, 62, February, 1-14.

Gallo, Carmine (2019), "The art of persuasion hasn't changed in 2000 years", *Harvard Business Review,* July, 1-6.

Gibbs, Chris, Daniel Guttentag, Ulrike Gretzel, Jym Morton, and Alasdair Goodwill (2018), "Pricing in the sharing economy: A hedonic pricing model applied to Airbnb listings", *Journal of Travel & Tourism Marketing, 35,* March, 46-56.

———, Stephen Smith, Luke Potwarka, and Mark Havitz (2017), "Why tourists choose Airbnb: A motivation-based segmentation study", *Journal of Travel Research*, 57, April, 342-359.

——— (2015), "Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector", *Current Issues in Tourism,* 18, December, 1192-1217.

Griffiths, Thomas L., and Mark Steyvers, (2004), "Finding scientific topics", *Proceedings of the National academy of Sciences*, 5228-5235.

Han, Heejeong, Seunghun Shin, Namho Chung, and Chulmo Koo (2019), "Which appeals (ethos, pathos, logos) are the most important for Airbnb users to booking?", *International Journal of Contemporary Hospitality Management*, 31, April, 1205-1223.

Inside Airbnb, "Data Assumptions", (accessed 10 June 2022), [available at http://insideairbnb.com/data-assumptions]

Jill, Sweeney, Joffre, Swait (2008), "The effects of brand credibility on customer loyalty", *Journal of Retailing and Consumer Services,* 15, 179-193.

Jun, Soo-Hyun (2020), "The Effects of Perceived Risk, Brand Credibility and Past Experience on Purchase Intention in the Airbnb Context", *Sustainability,* 12, 5212.

Kathan, Wolfgang, Kurt Matzler and Viktoria Veider (2019), "The sharing economy: Your business model's friend or foe?", *Business Horizons,* 59, 663-672.

Liang, Sai, Markus Schuckert, Rob Law, and Chih-Chien Chen (2020), "The importance of marketer-generated content to peer-to-peer property rental platforms: evidence from Airbnb", *International Journal of Hospitality Management,* 84, January, 102329.

Mcauliffe, J. D., & Blei, D. M. (2008). Supervised topic models. In Advances in neural information processing systems (pp. 121–128).

McDonald, John F., and Robert A. Moffitt (1980), "The uses of Tobit analysis", *The review of economics and statistics*, May, 318-321.

McFadden, Daniel (1973), "*Conditional logit analysis of qualitative choice behavior*"

Minka, T., & Lafferty, J. (2002), "Expectation-propagation for the generative aspect model. Dans les actes de Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence", 352–359.

Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining (2021), *Introduction to linear regression analysis*, John Wiley & Sons.

Moore, Andrew (2019), "Hedonism", (accessed 16 May 2022), [available at https://plato.stanford.edu/].

Narasimhan, Chakravarthi, Purushottam Papatla, Baojun Jiang, Praveen K. Kopalle, Paul R. Messinger, Sridhar Moorthy, Davide Proserpio, Upender Subramanian, Chunhua Wu, and Ting Zhu (2018), "Sharing economy: Review of current research and future directions", *Customer needs and solutions*, 5, March, 93-106.

Nasukawa, Tetsuya, and Jeonghee Yi (2003), "Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of the 2nd international conference on Knowledge capture", *Association for Computing Machinery*, October, 70-77.

Okada, Erica Mina (2005), "Justification Effects on Consumer Choice of Hedonic and Utilitarian Goods", *Journal of Marketing Research*, 42, February, 43-53.

Oskam, Jeroen A (2019), *The future of Airbnb and the 'sharing economy'. In The Future of Airbnb and the 'Sharing Economy*, Channel View Publications.

Perren, Rebeca, and Liz Grauerholz (2015), "Collaborative consumption.", *International Encyclopedia of the Social & Behavioral Sciences*, 4, December

Rapp, Christof (2002), "Aristotle's Rhetoric", *The Stanford Encyclopedia of Philosophy*.

Saif Mohammad (2016), "NRC Word-Emotion Association Lexicon", (accessed 14 July 2022), [available at http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm]

Schor, Juliet (2016), "Debating the sharing economy", (accessed 29 March 2022), [available at https://www.greattransition.org/publication/debating-the-sharing-economy]

Shao, Aiping, and Hong Li (2021), "How do utilitarian versus hedonic products influence choice preferences: Mediating effect of social comparison", *Psychology & Marketing*, 38, August, 1250-1261.

Stors, Natalie, & Kagermeier, Andreas (2015), "Motives for using Airbnb in metropolitan tourism: Why do people sleep in the bed of a stranger?", *Regions Magazine*, 299, September, 17–19.

Trenholm, Richard (2015), "Airbnb exec denies competition with hotels, says an Airbnb trip 'changes you' ", (accessed April 28, 2022), [available at https://www.cnet.com/tech/services-and-software/airbnb-exec-denies-competition-with-hotels-says-an-airbnb-trip-changes-you-somehow/].

Tussyadiah, Iis P., and Sangwon Park (2018), "When guests trust hosts for their words: host description and trust in sharing economy", *Tourism Management*, 67, August, pp. 261-272.

Wang, Dan and Juan L. Nicolau (2017), "Price Determinants of Sharing Economy Based Accommodation Rental: A Study of Listings from 33 Cities on Airbnb.com", *International Journal of Hospitality Management*, 62, 120–31.

Xie, Karen, and Zhenxing Mao (2017), "The impacts of quality and quantity attributes of Airbnb hosts on listing performance", *International Journal of Contemporary Hospitality Management*, 29, September, 2240-2260.

Yang, Sung-Byung, Hanna Lee, Kyungmin Lee, and Chulmo Koo (2018), "The application of Aristotle's rhetorical theory to the sharing economy: an empirical study of Airbnb", *Journal of Travel & Tourism Marketing*, 35, May, 938-957.

Yang, Sung-Byung, Kyungmin Lee, Hanna Lee, and Chulmo Koo (2019), "In Airbnb, we trust: Understanding consumers' trust-attachment building mechanisms in the sharing economy", *International Journal of Hospitality Management*, 83, 198-209.

Ye, Qiang, Rob Law, Bin Gu, and Wei Chen (2011), "The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings", *Computers in Human behavior 27*, 2, 634-639.

Zervas, Georgios, Davide Proserpio, and John W. Byers (2017), "The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry", *Journal of marketing research*, 54, October, 687-705.

Zervas, Georgios, Davide Proserpio, and John W. Byers (2021), "The first look at online reputation on Airbnb, where every stay is above average", *Marketing Letters*, 32, November 1-16.

Zimmerman, Kaytie, "How Crowdsourcing Is Transforming The Workplace" (2016), (accessed 28 April 2022), [available at https://www.forbes.com/sites/kaytiezimmerman/2016/07/12/how-crowdsourcing-is-transforming-the-workplace/?sh=6962f69f7030]

Appendix A

LDA Description Results

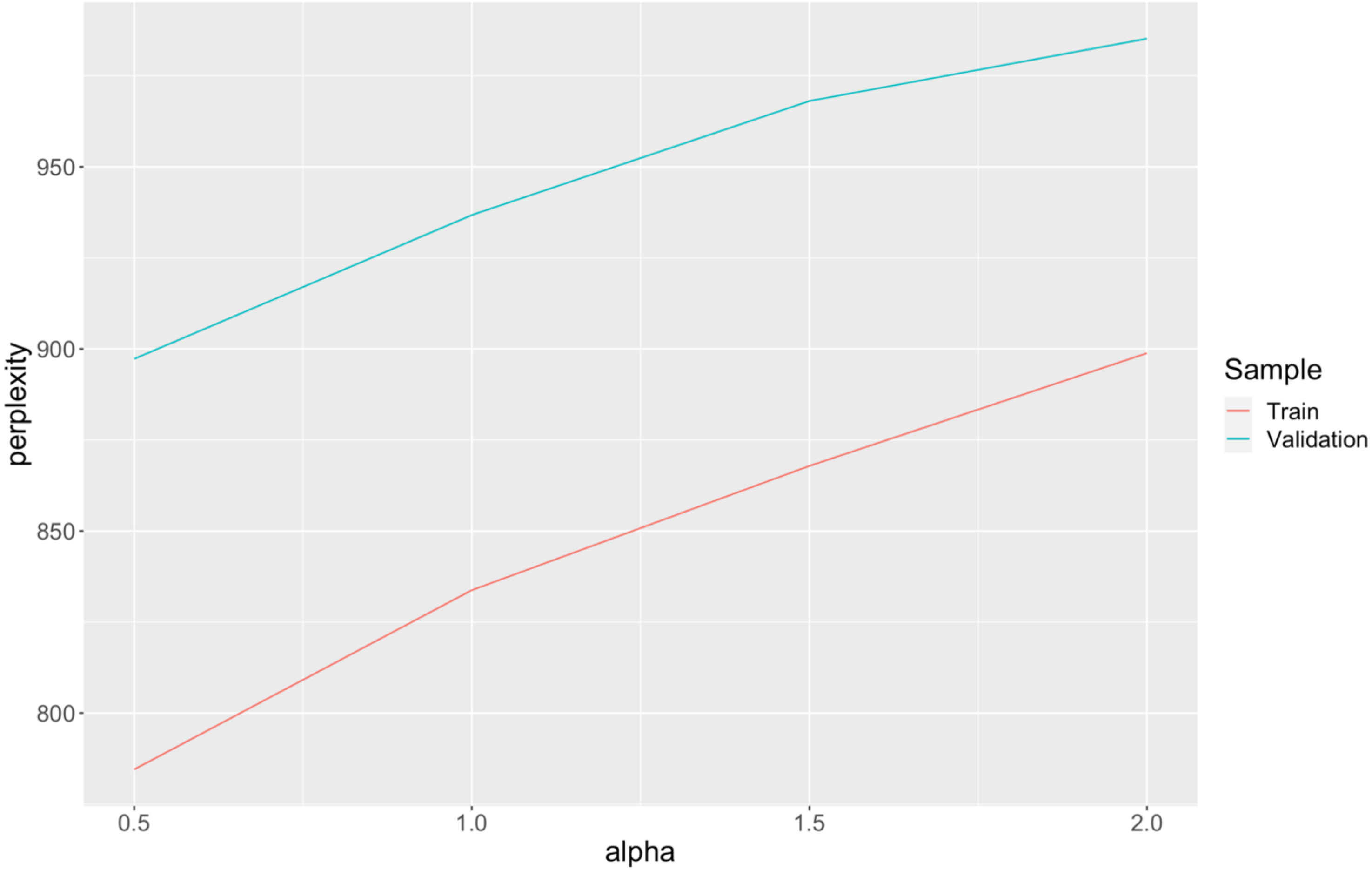


Figure 25: Perplexity – alpha plot "description" column

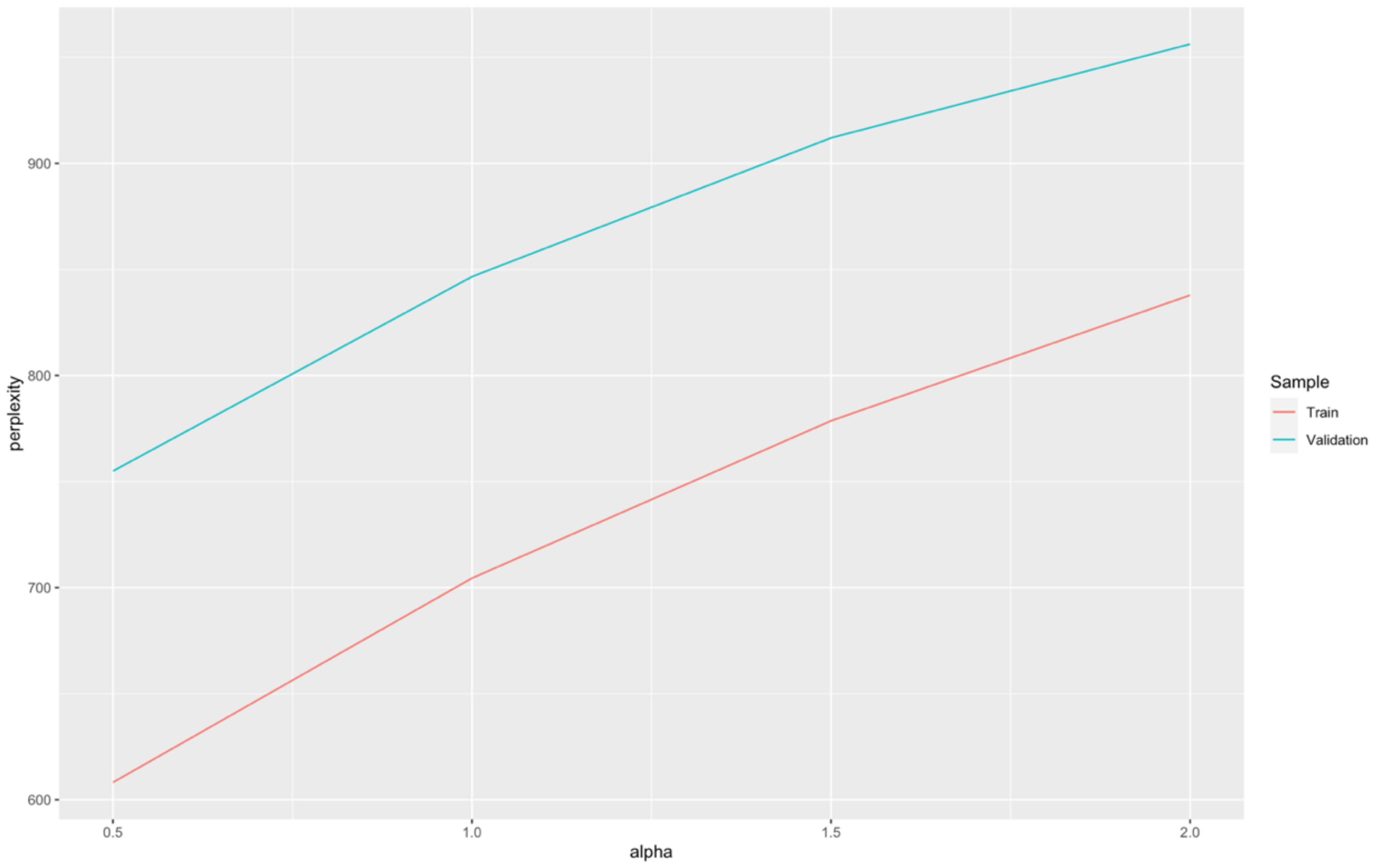


Figure 26: Perplexity - alpha plot "host_about" column

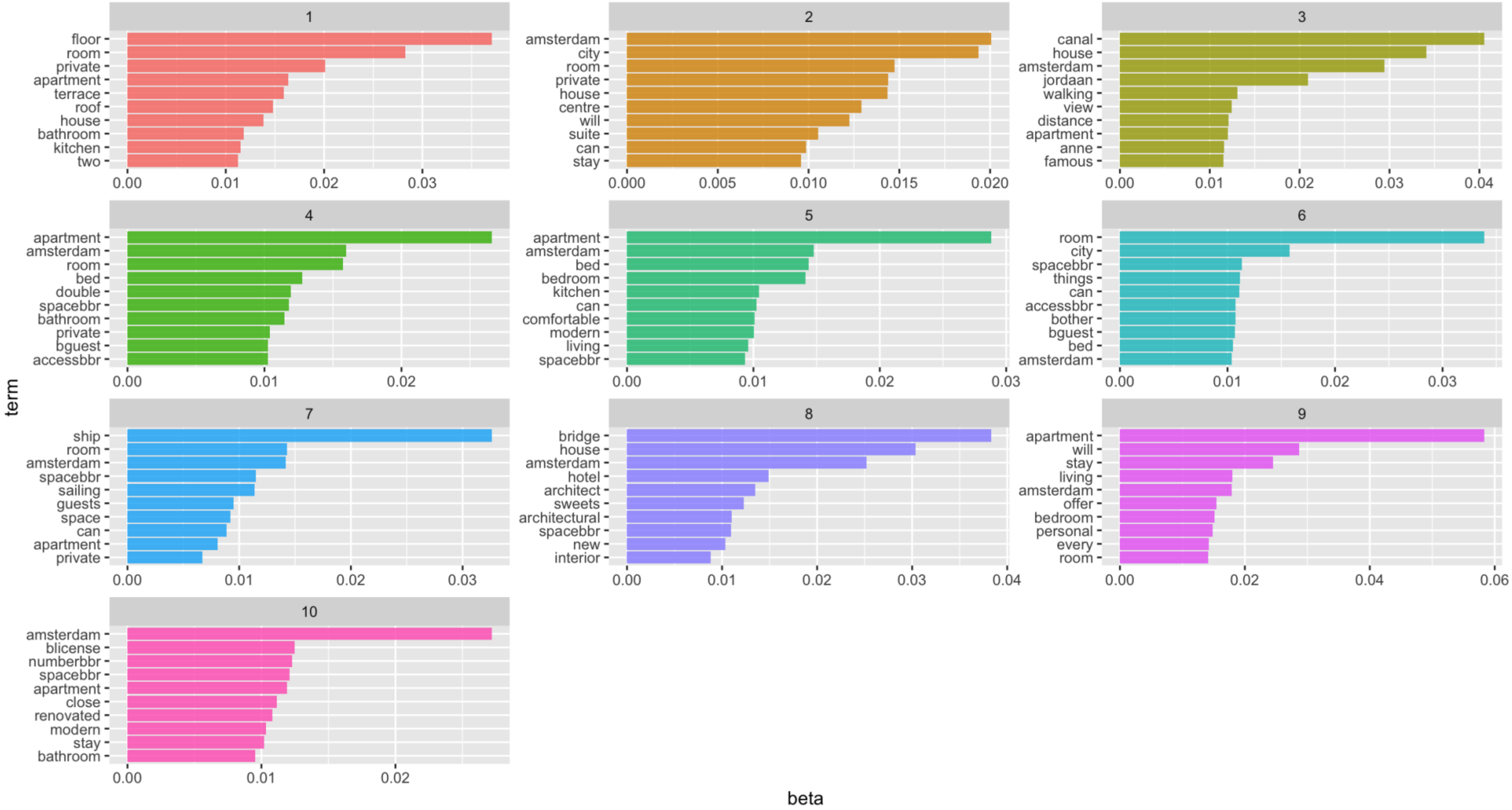


Figure 27:LDA Top 10 Results “description" column

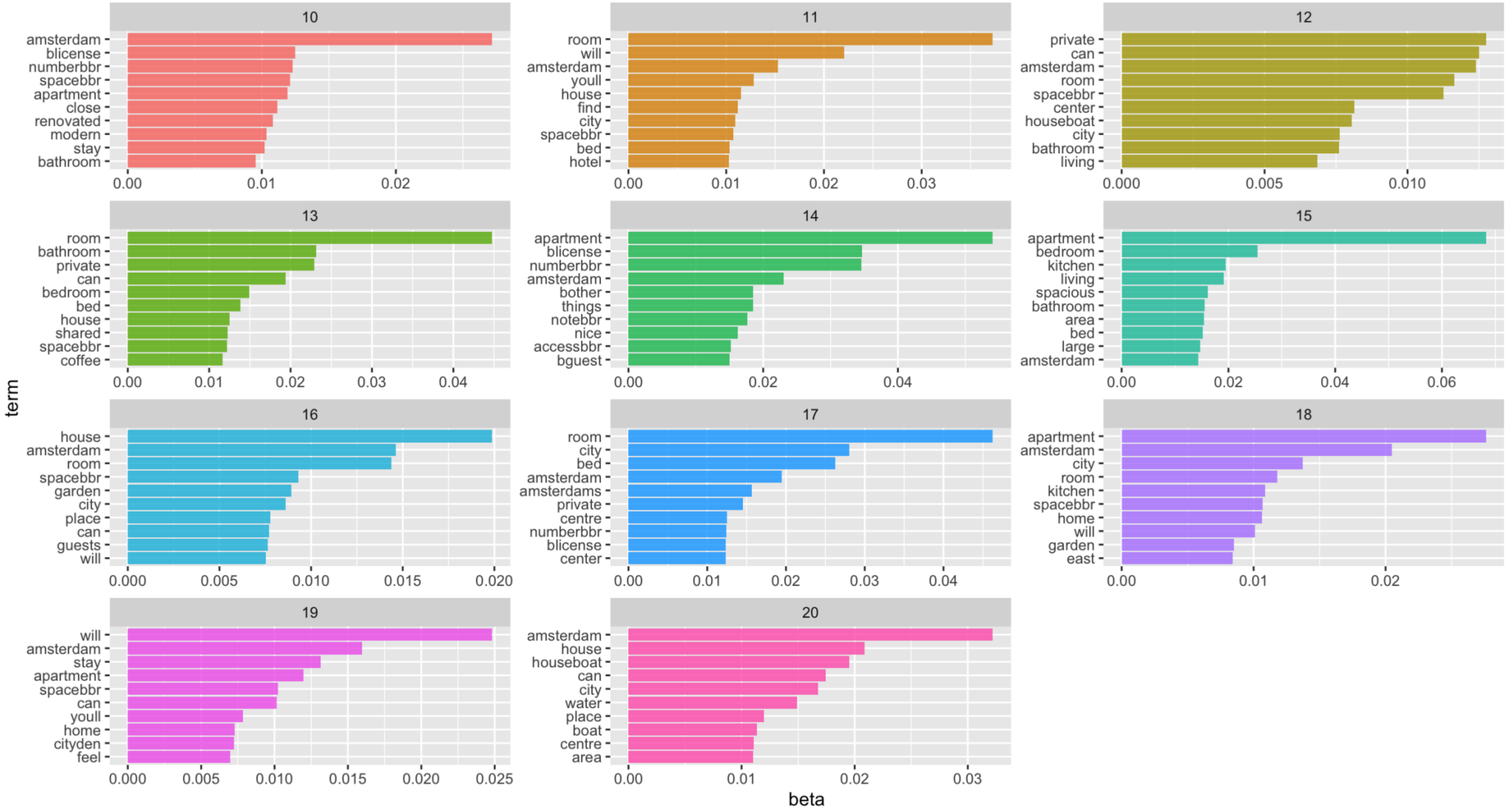


Figure 28:LDA Top 10 - 20 Results "description" column

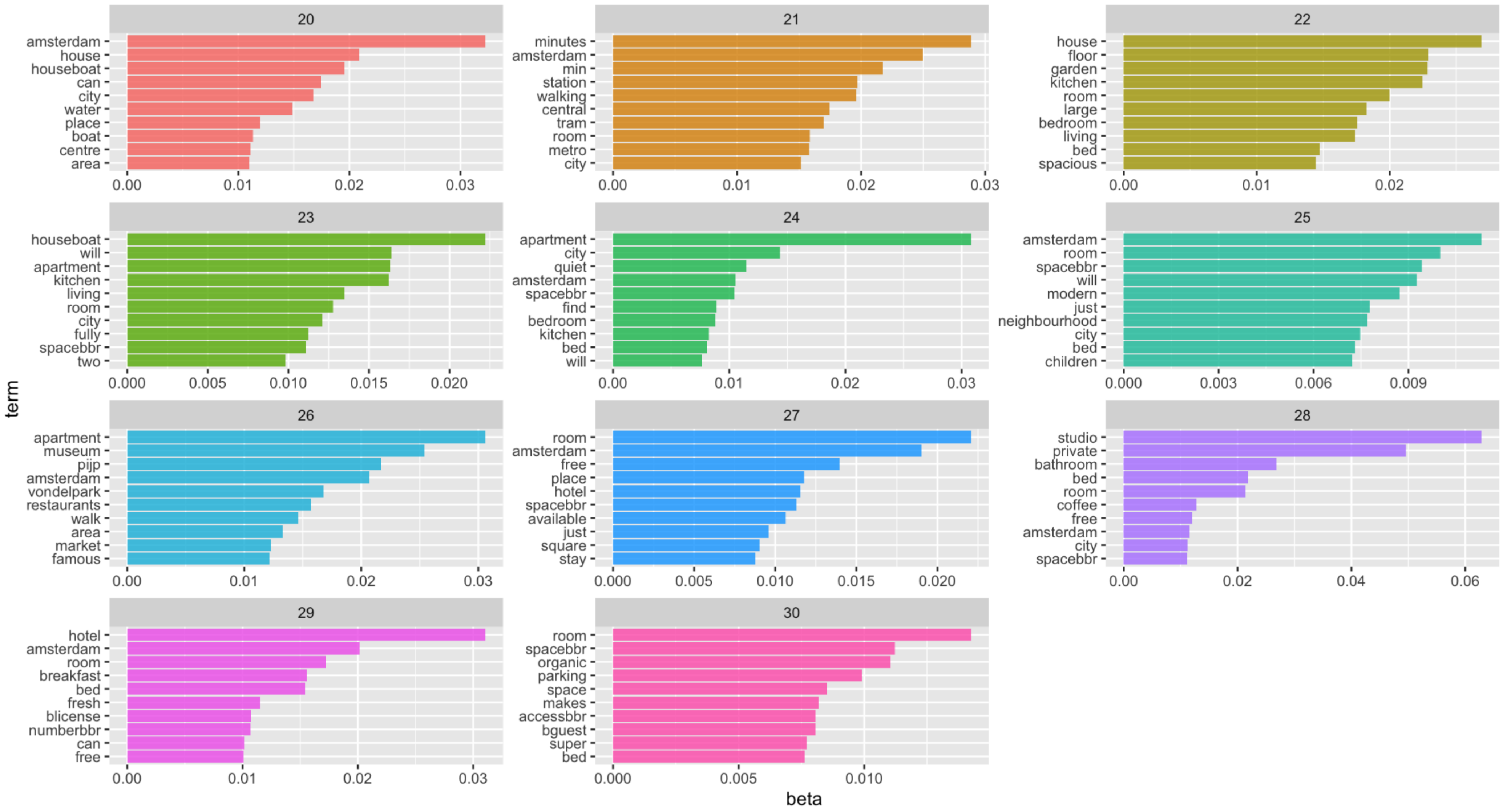


Figure 29:LDA Top 20 - 30 Results "description" column

LDA “host_about” top terms Results

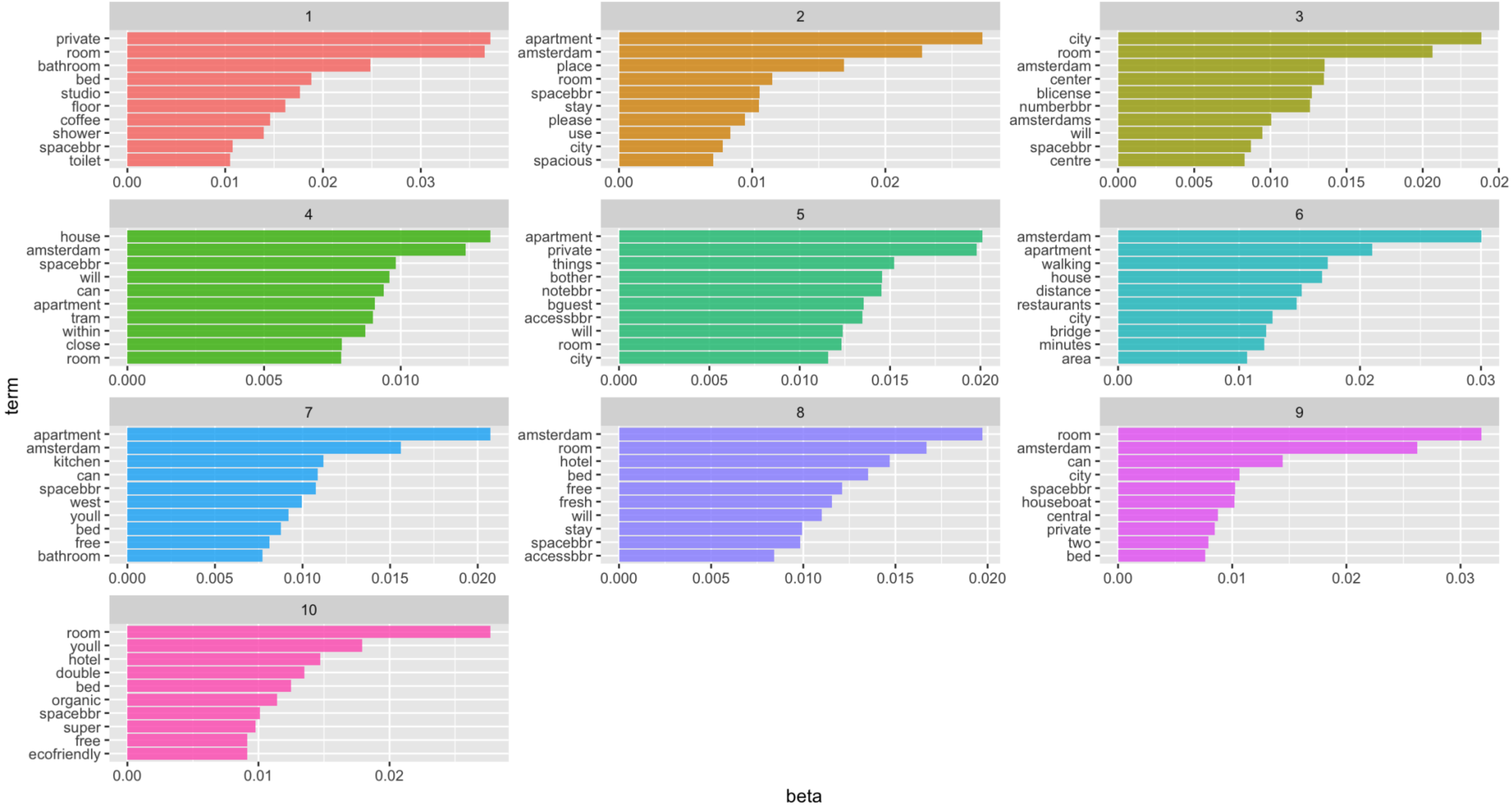


Table 12: LDA Top 10 terms "host_about" column

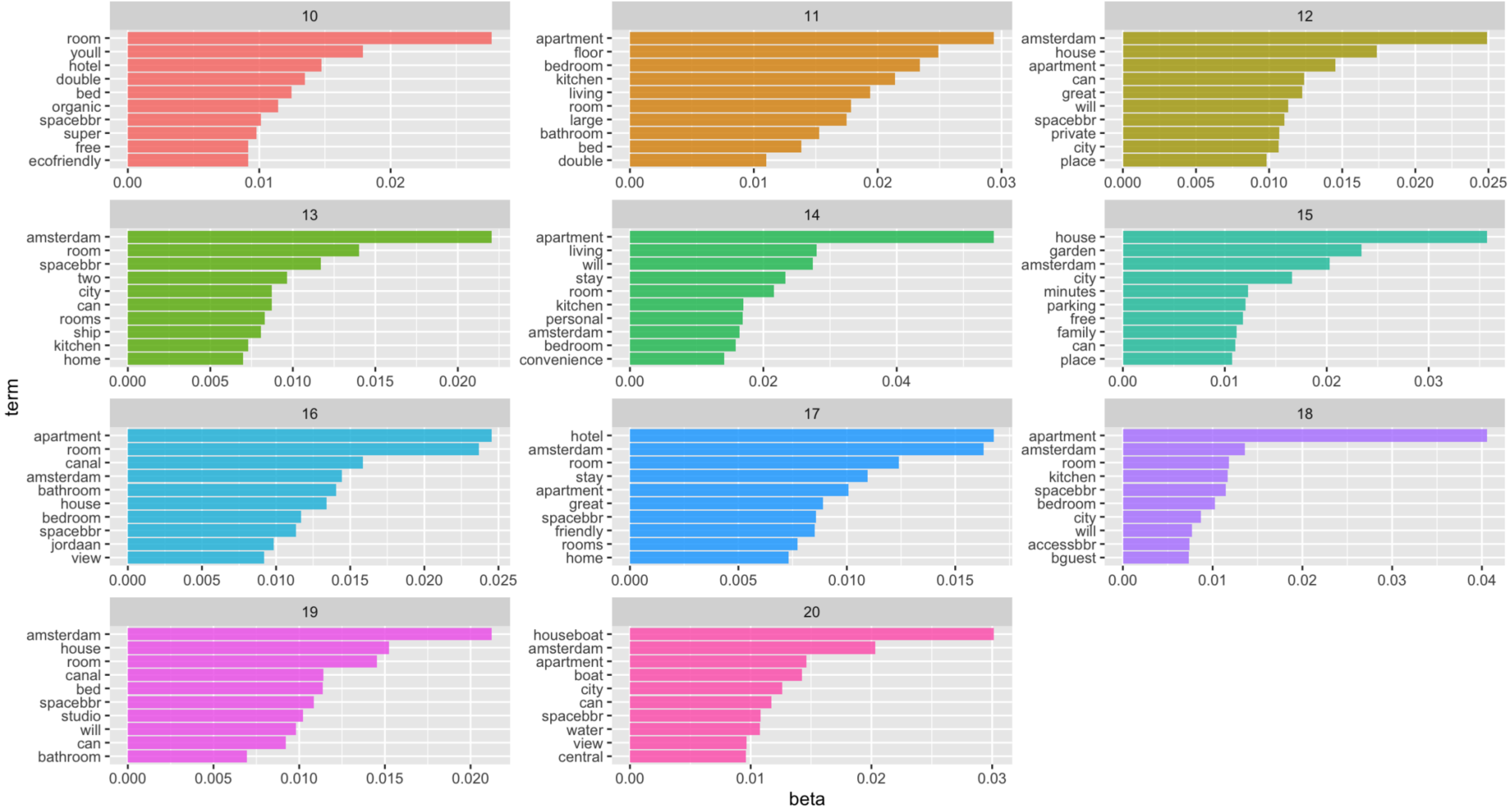


Table 13: LDA Top 10 – 20 terms "host_about" column

LDA “comments” Results

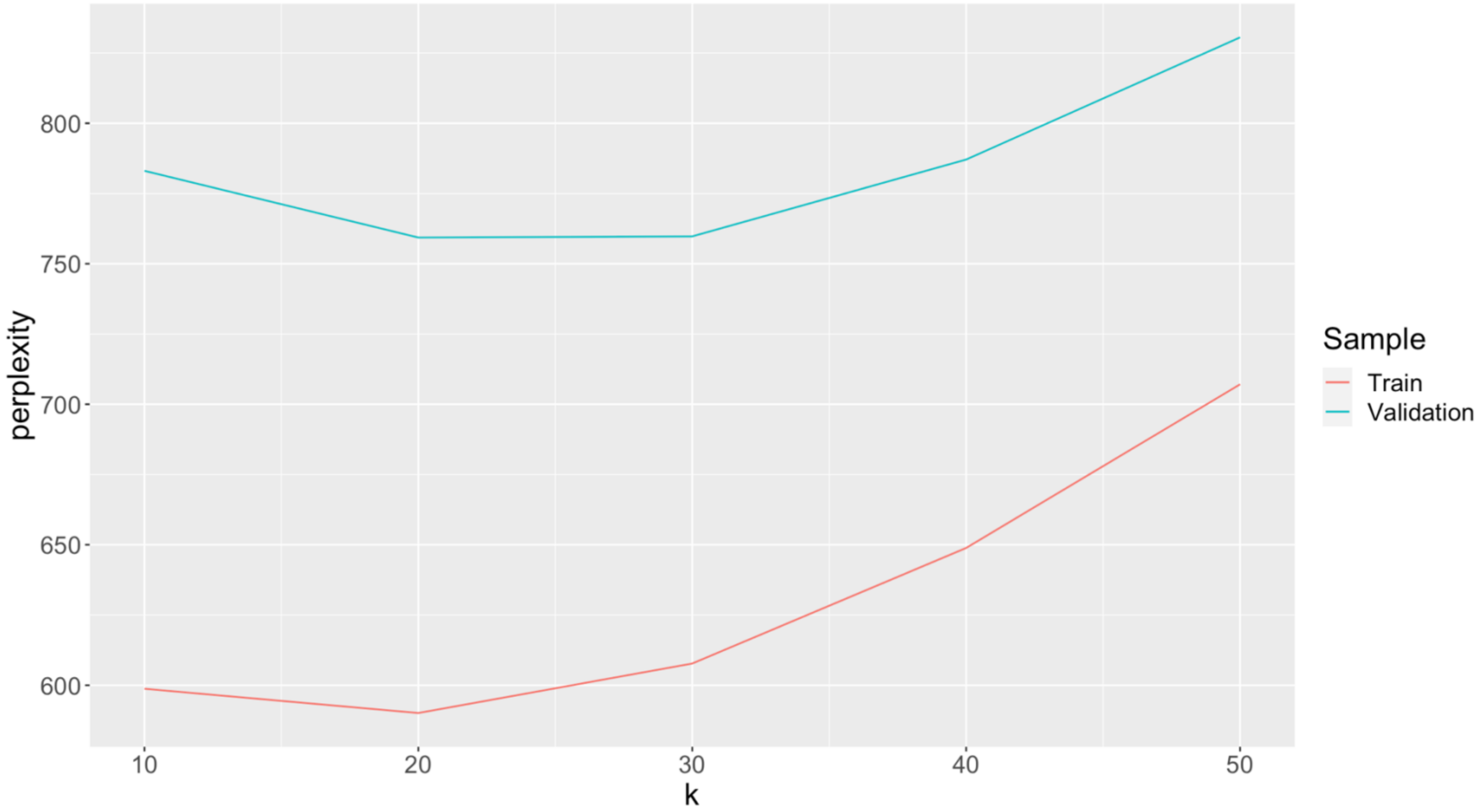


Figure 30: perplexity plot "comments" column

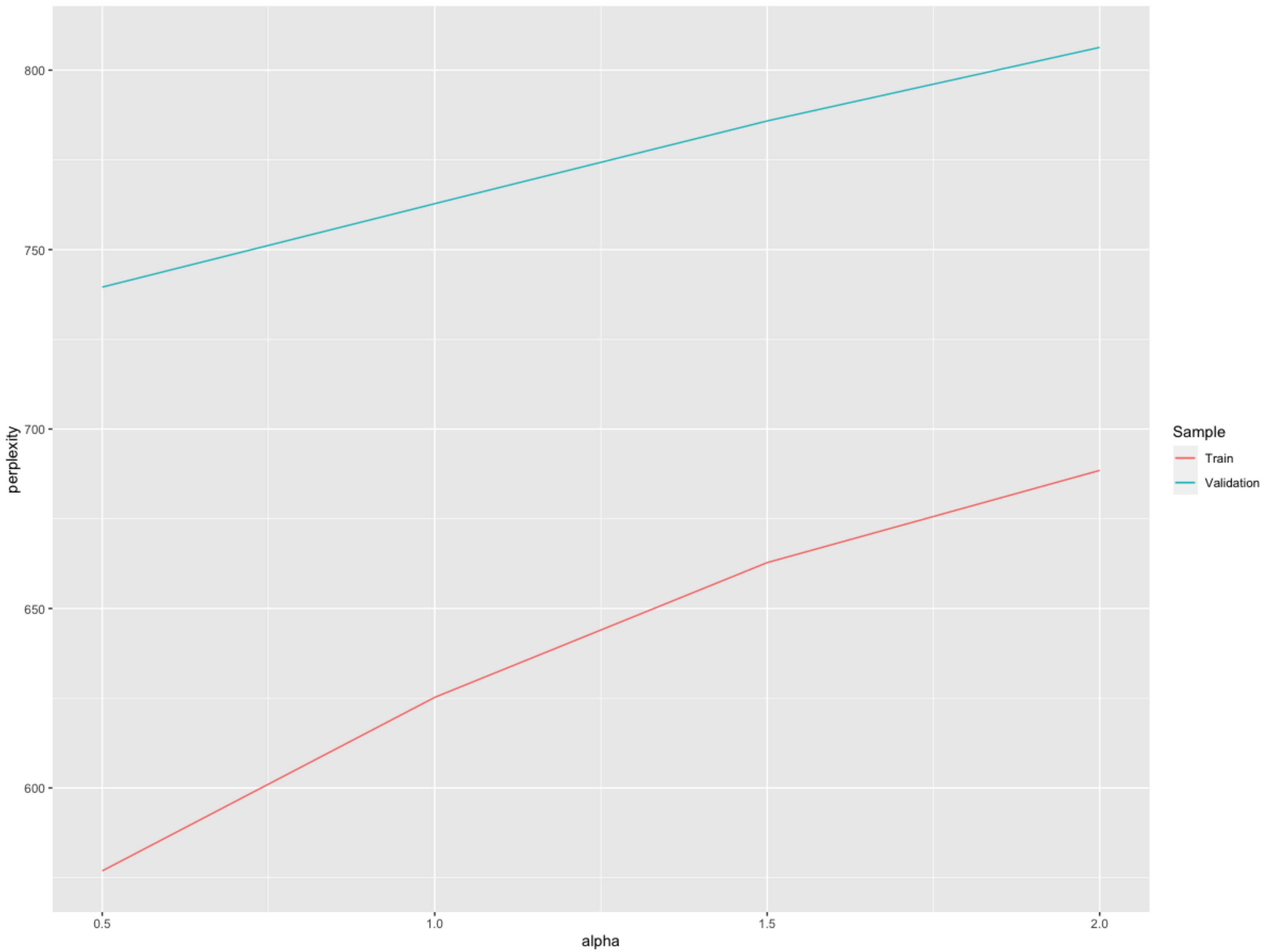


Figure 31: perplexity – alpha plot "comments" column

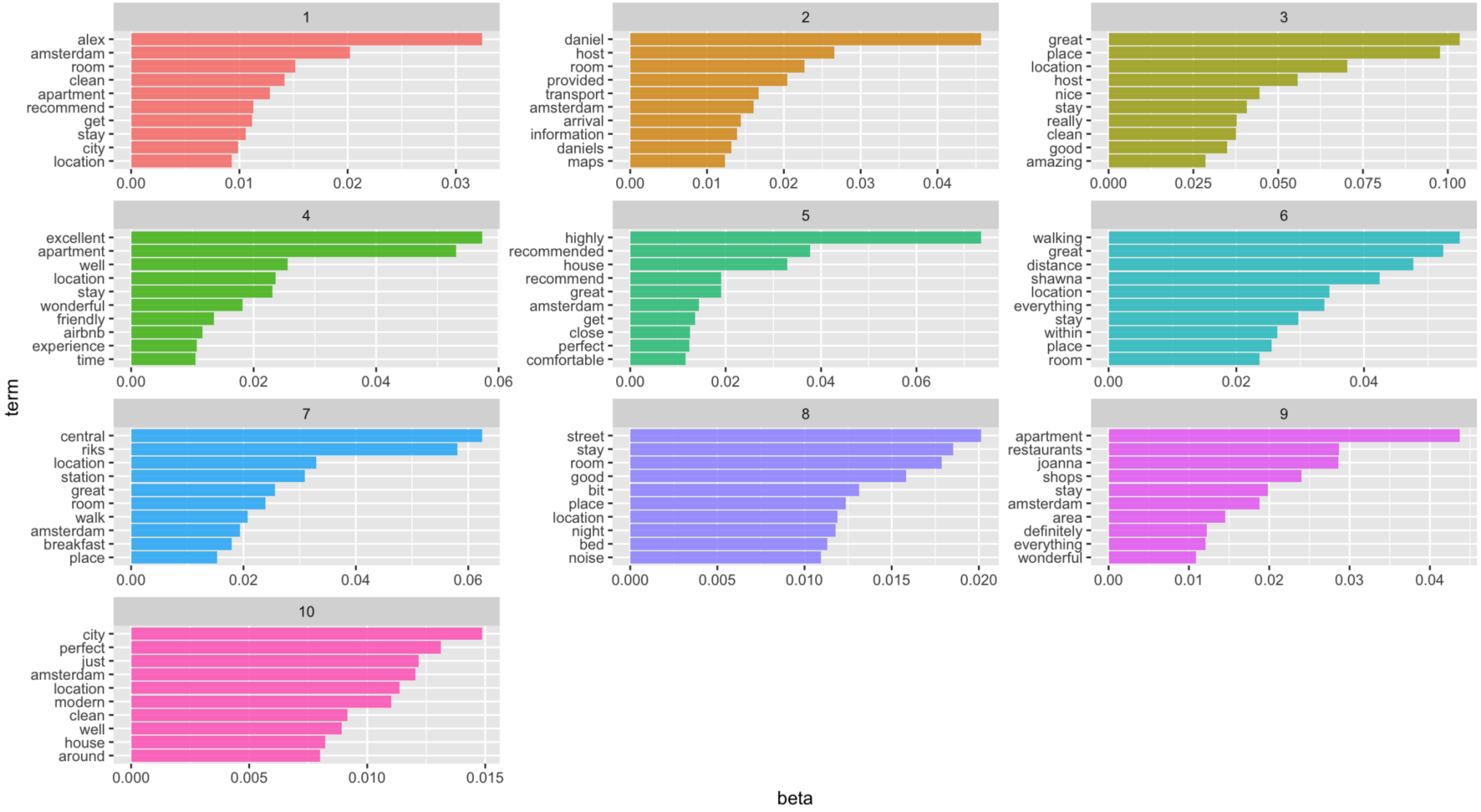


Figure 32: LDA Top 10 terms "comments" column

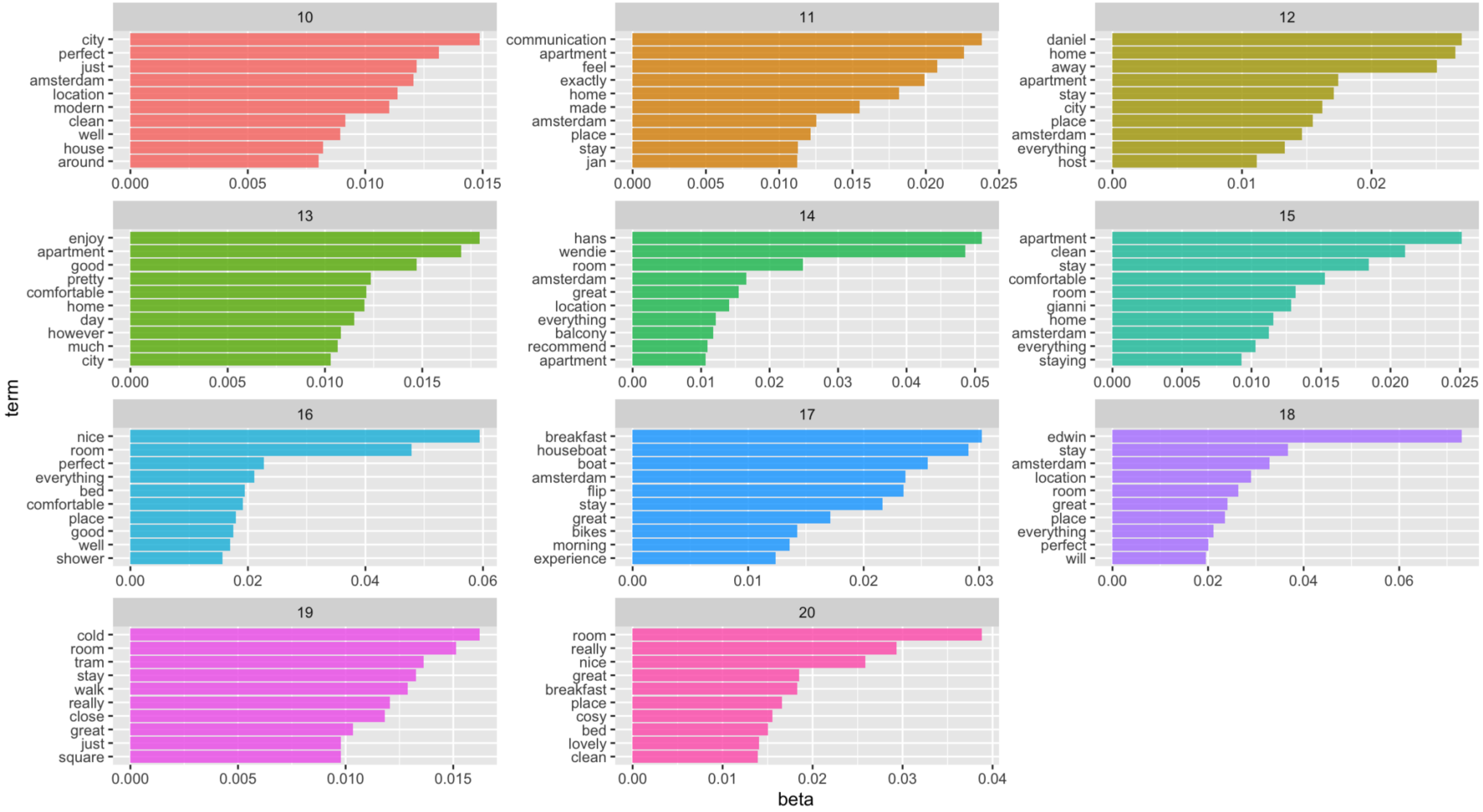


Figure 33: LDA Top 10 – 20 terms "comments " column

Appendix B

Sentiment analysis "*comments*" column

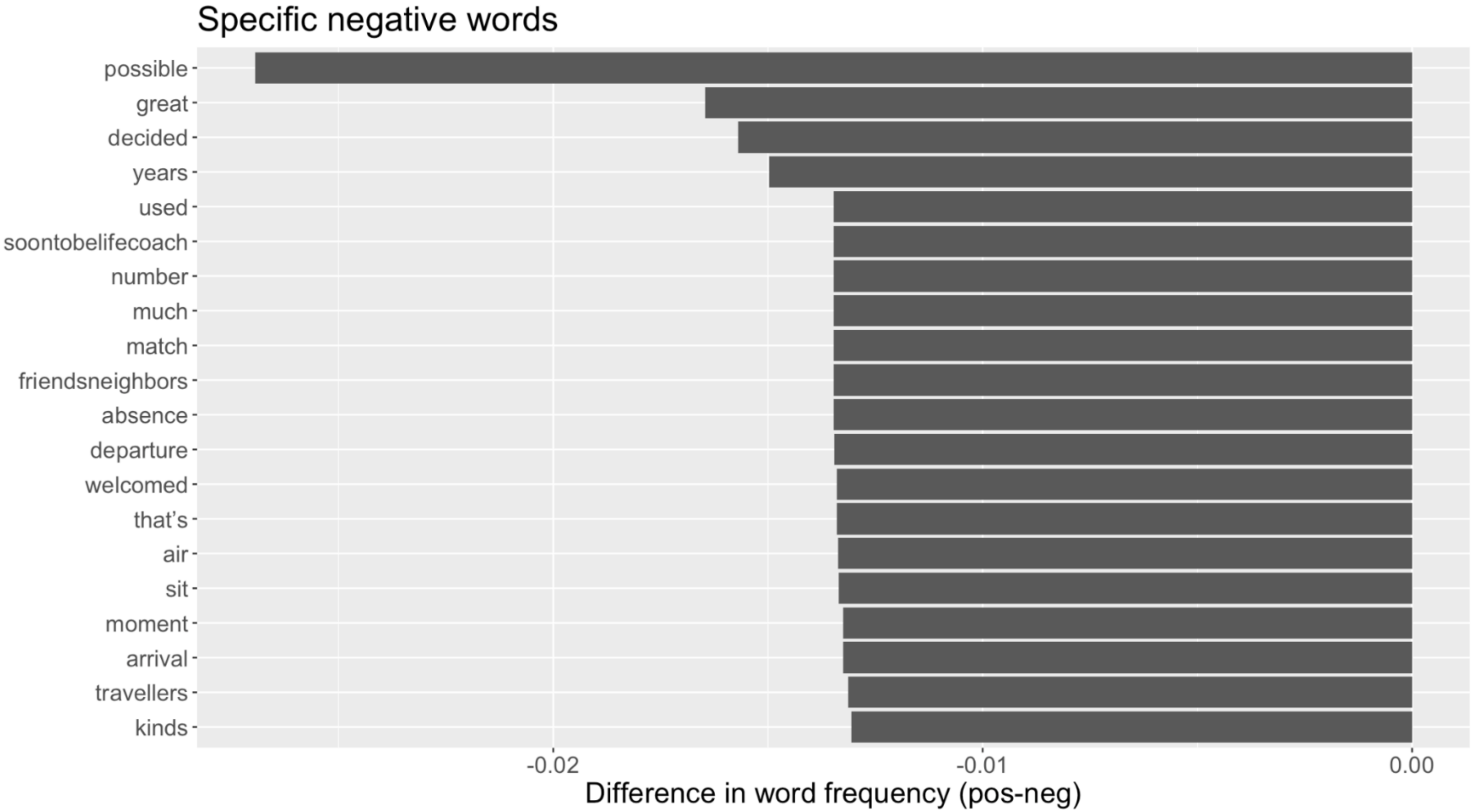


Figure 34:Specific negative words of the "comments" column

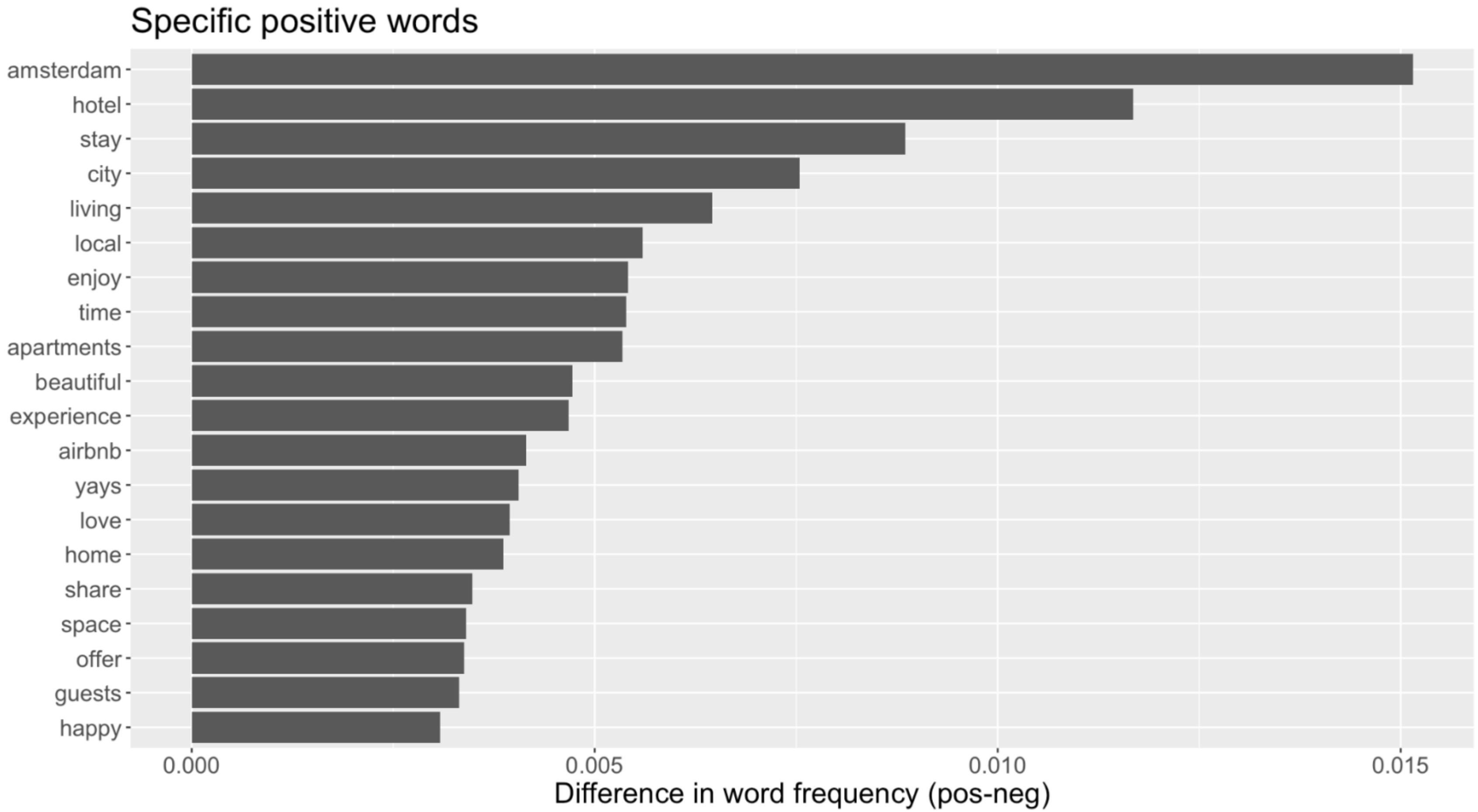


Figure 35:Specific positive words of the "comments" column