COMBINED ASSESSMENT:

DATA VISUALIZATION AND TEXT MINING / NLP

Team 12 - The Out(siders)

Eyra Urrutia
Mohssin Elyounssi
Naila Jan
Panchali Chaudhury
Stefano Sanzo

HULT INTERNATIONAL BUSINESS SCHOOL

TEXT ANALYTICS AND NATURAL LANGUAGE PROCESSING

DATA VISUALIZATION

Thomas Kurnicki, Priyanka Shrivastava

December 14, 2021

Airbnb2b

Air bed and breakfast to breath,

Breath: Experiences around the house, Breath: With activities, Breath: Love of people.

Introduction

We are the '*The Out(siders)*' and were hired by Airbnb to analyze descriptions of postings. By exploring the dataset from MongoDB, we see that the properties listed in Airbnb were from the countries shown in map 1.1 attached to the appendix.

We have selected the top three countries (United States, Australia, Spain) according to the number of reviews listed in the Airbnb dataset.

The main property types that we analyzed for the Airbnb case are mentioned in Figure 1.2 as we can Cleary see the highest property types are the apartments with 3,626 properties around the world according to the dataset.

In addition, with R- Studio we have implemented text mining techniques to analyze and compare the reviews and descriptions. Through our findings, we believe that our recommendation will improve guest experience, increase the business value for the Airbnb hosts, and add value to the Airbnb business model. By adding value to the Airbnb host we add value to the Airbnb business.

Analysis and Findings

By applying text mining with R, we have compared the variables of descriptions, and comments, for the top three countries cited above.

Using the following frameworks: correlograms, bigrams, TF-IDF, sentimental analysis, and LDA (Beta, gamma). We observed a clear pattern from the guest 's comments which is usually talking about their experiences and activities outside the residences. For example, amazing location and central park are the most common bigrams. As you can see in the bigram graph (Figure 2.2)

On the other hand, we observed that the description of the host is mostly focused on the inside of the housing facilities. For example, free Wi-Fi, newly renovated, washer and dryer. (Figure 2.3)

With these observations in mind, we can see a significant difference between the expectations of the guest and the offers of the hosts.

From the textual reviews, we have applied sentiment analysis to measure customer satisfaction to discover the similarities and positive words these countries have in common. Furthermore, sentimental analysis shows us the importance of trust for the guest which is relevant for the business model. (Figure 2.7)

By using tableau, first, we have found that the most expensive country is Australia followed by the United States and Spain. Then we looked for correlation and association between the price per night and the experience of the guest. The finding showed that there is a positive correlation in the United States between prices per night and experiences of the guests while on the other hand for Spain and Australia the correlation is negative in this aspect. As a result, the higher prices don't assure the best experiences. Secondly, we investigate a significant relationship between hosts with verified identity and guests' experience. The data shows that in the United States most of the hosts have a verified identity; on the contrary in Australia and Spain, the highest number of hosts are not verified. Consequently, we found that verified hosts provide a better customer experience according to the rating scores.

Recommendations

Altogether, following our recommendations to Airbnb.

1. The host description should focus more on outside activities to convey a unique experience to the guest.
2. The hosts need to use in the description of their posts words that inspire trust in the guests.

Suggested action: Airbnb should help the hosts by providing them customized templates with keywords for their postings and possible activities close to their property. We believe that this action will increase engagement and trust between host and guest.

Furthermore, Airbnb could partner with international and neighborhood entertainment businesses to expand to a new market and create more business opportunities.

3. The number of hosts with verified identities should be increased in Austria and Spain since they have a lower number of verified hosts.
4. Airbnb should provide a new feature to forecast the quality of the experience.

Suggested action: Airbnb could also provide additional services to verified hosts to increase the number of them and improve the customer experience. Moreover, we recommend creating a new KPI to measure and track the data from the website about the experience related to the price.

Appendix

Figure 1.1 - Overview of the main properties' exposure around the world. (Tableau)

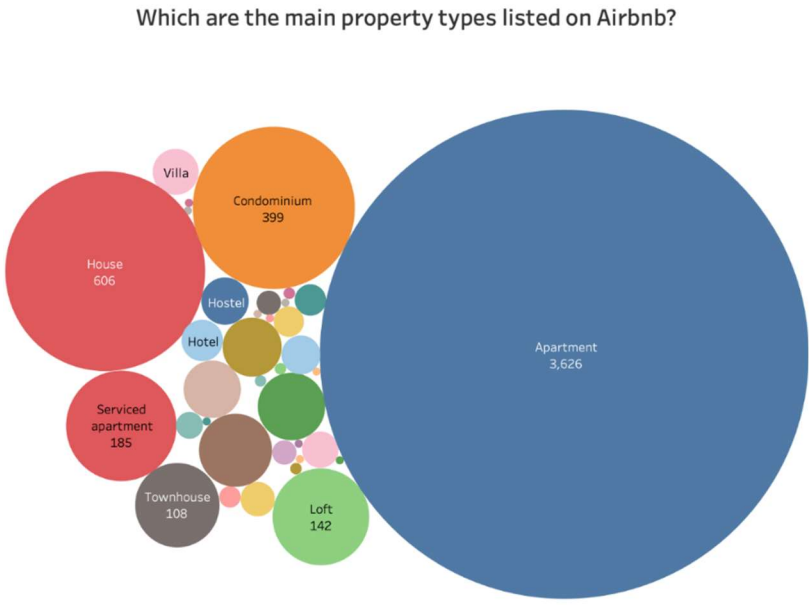

Figure 1.2 - Different type properties for accommodation listed in Airbnb. (Tableau)

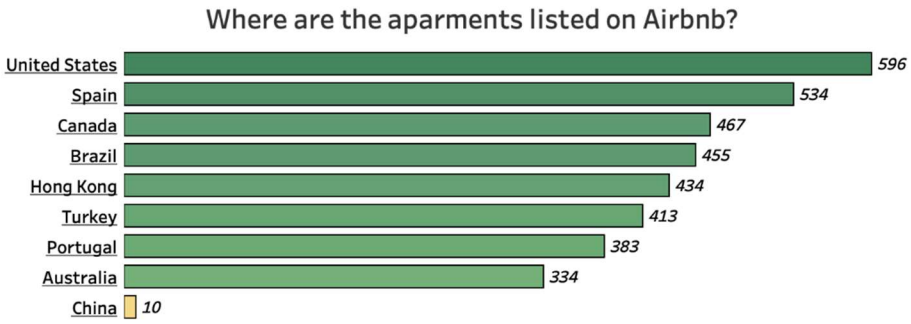

Figure 1.3 - Exposure of the properties around the word. (Tableau)

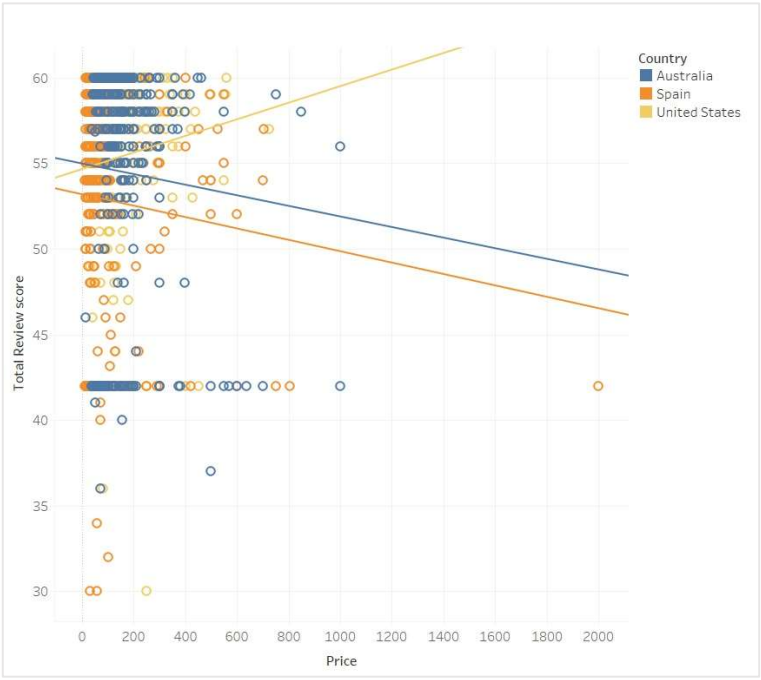

Figure 1.4 - Relation between price and reviews' score. (Tableau)

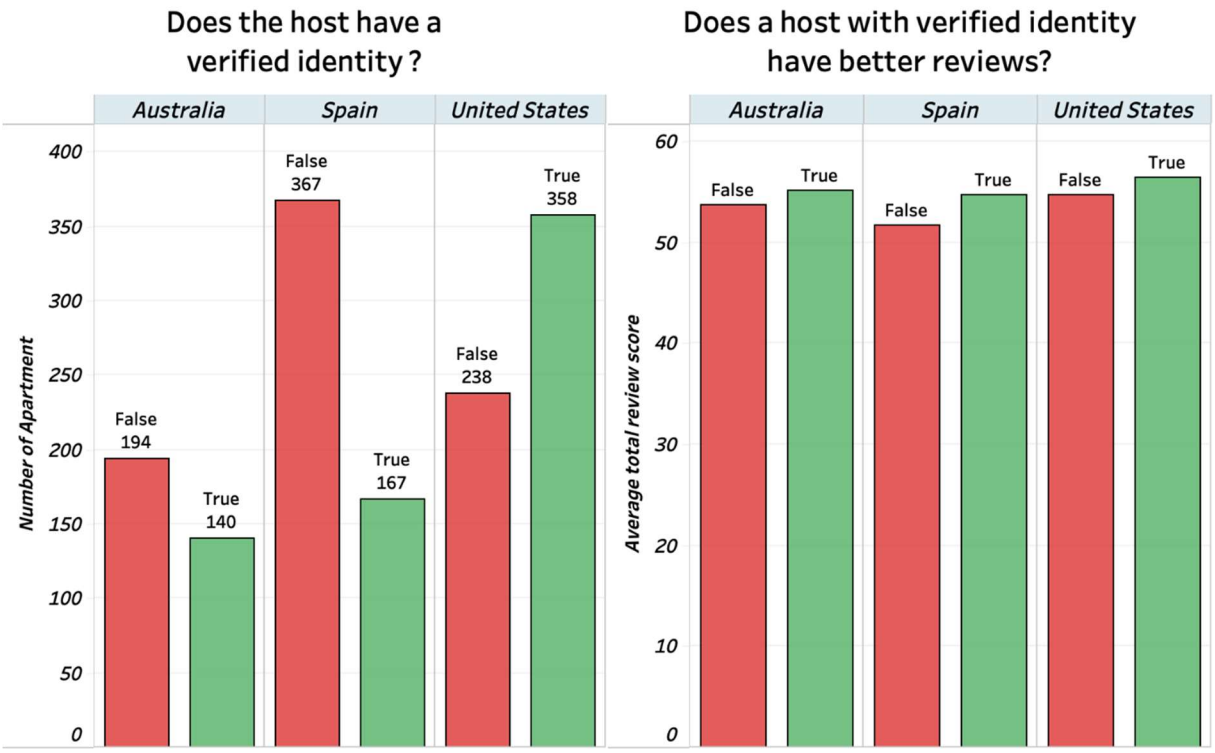

Figure 1.5 - Hosts with a verified identity and their relationship with reviews' score. (Tableau)

Correlograms - R

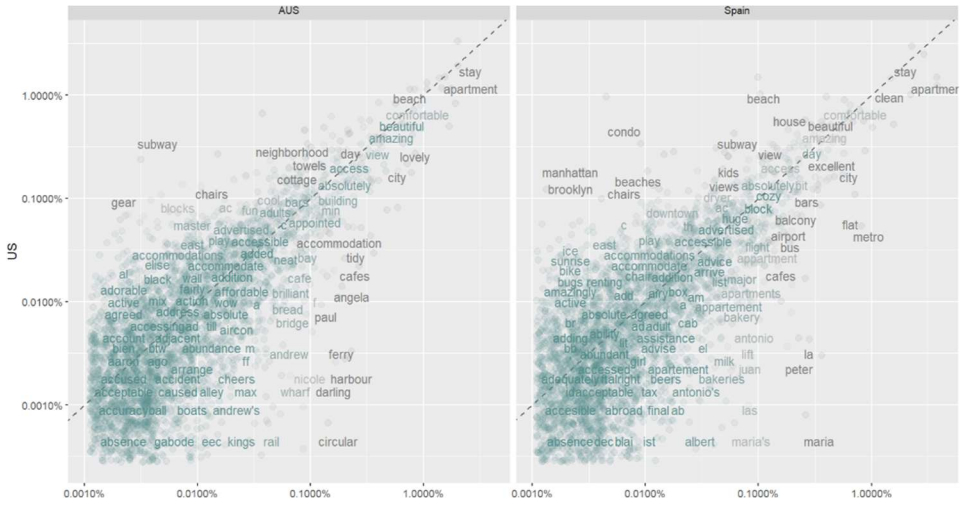

Figure 2.1 - Correlograms of reviews US/AUS and US/Spain - R

Bi-Grams for Comments

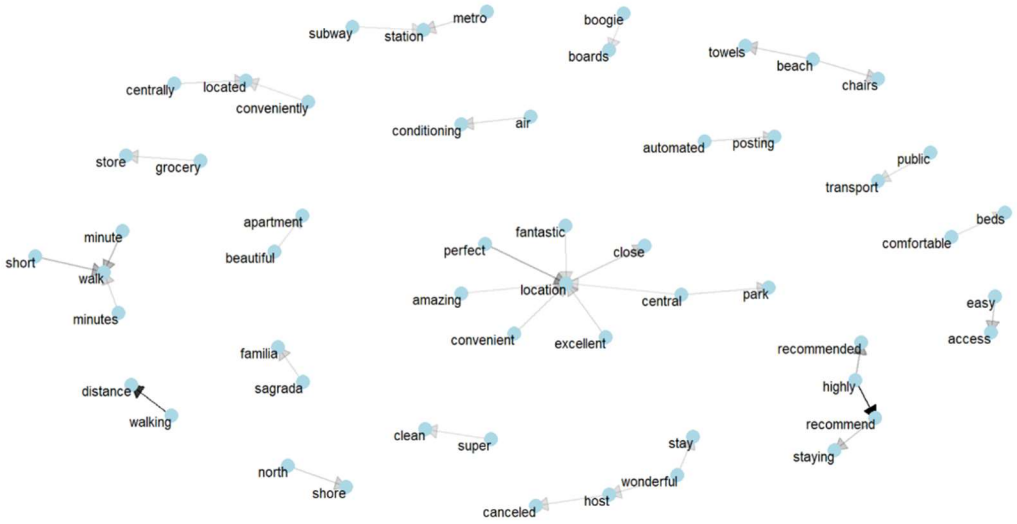

Figure 2.2 - Graph to see the connections between the different bigram words in reviews.

Bi-Grams for descriptions

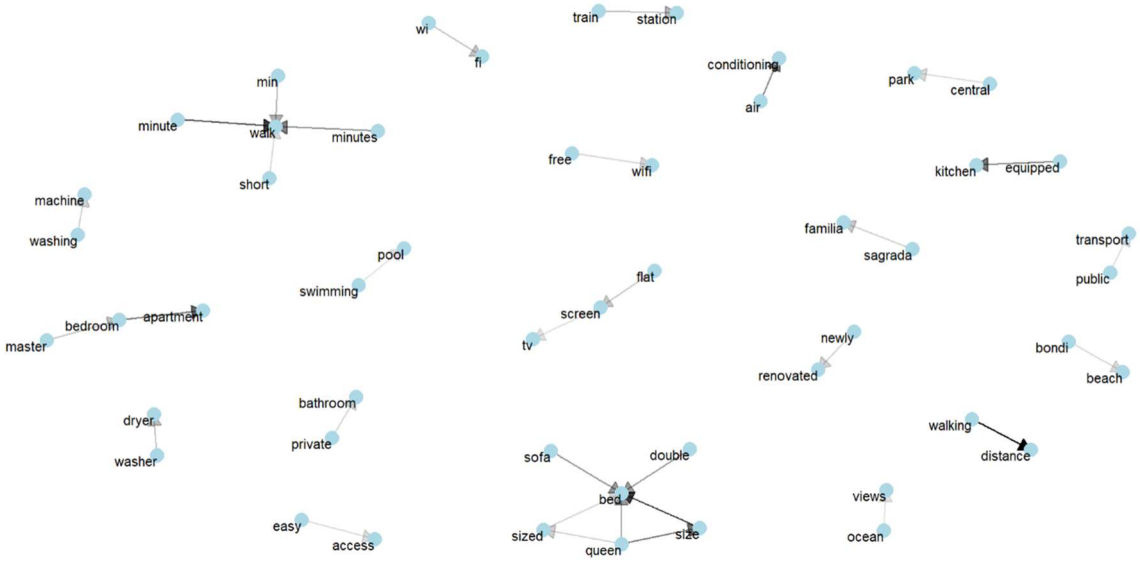

Figure 2.3 - Graph to see the connections between the different bigram words in description

Correlation Bi-Gram cluster of the reviews

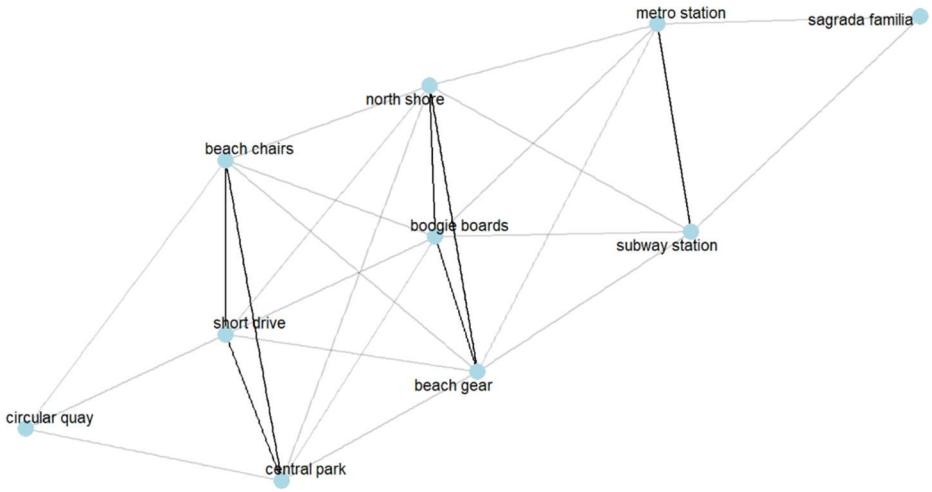

Figure 2.4 - Correlations and clusters of words by the widyr package

TF-IDF for Comments

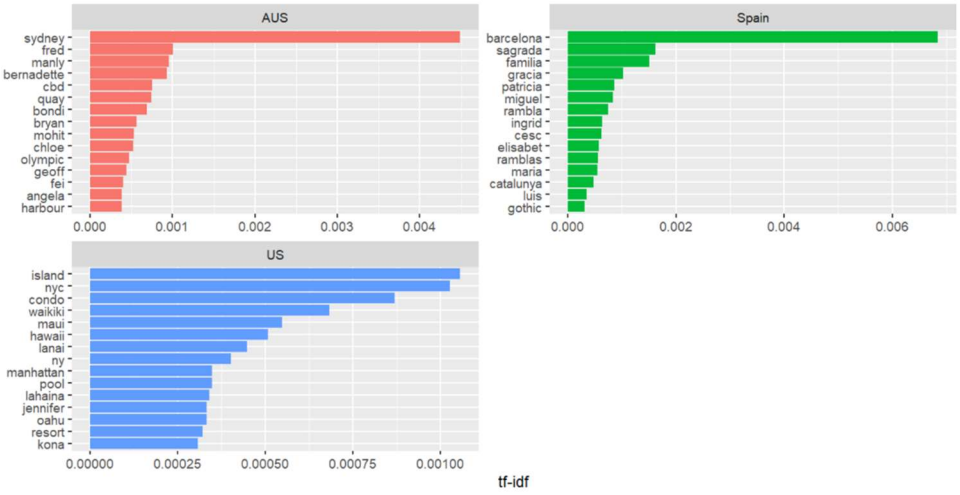

Figure 2.5 - Uniqueness of the word in AUS, Spain and USA by using TF-IDF

Sentiment Analysis with negation words

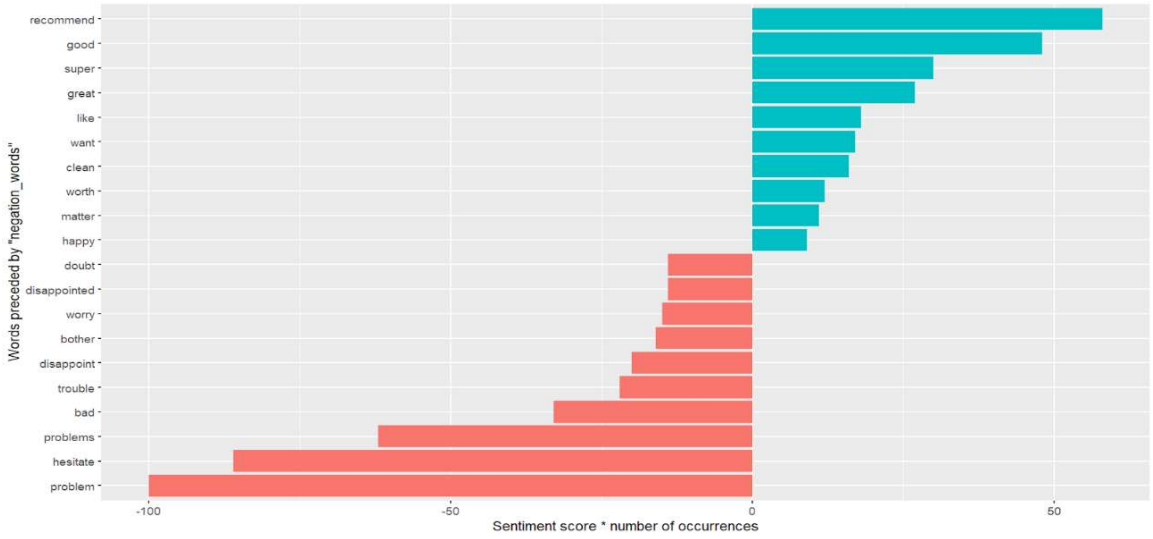

Figure 2.6 - Sentiment analyses with negation words "not", "no", "never", "without"

Sentiment Analysis with NRC

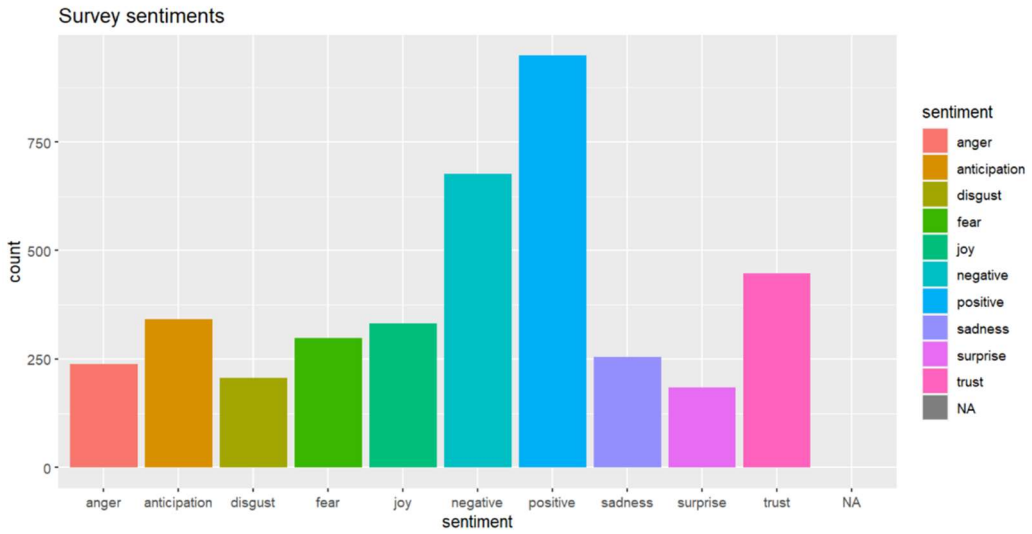

Figure 2.7 - Sentiment Analysis with NRC

LDA – Beta two topics

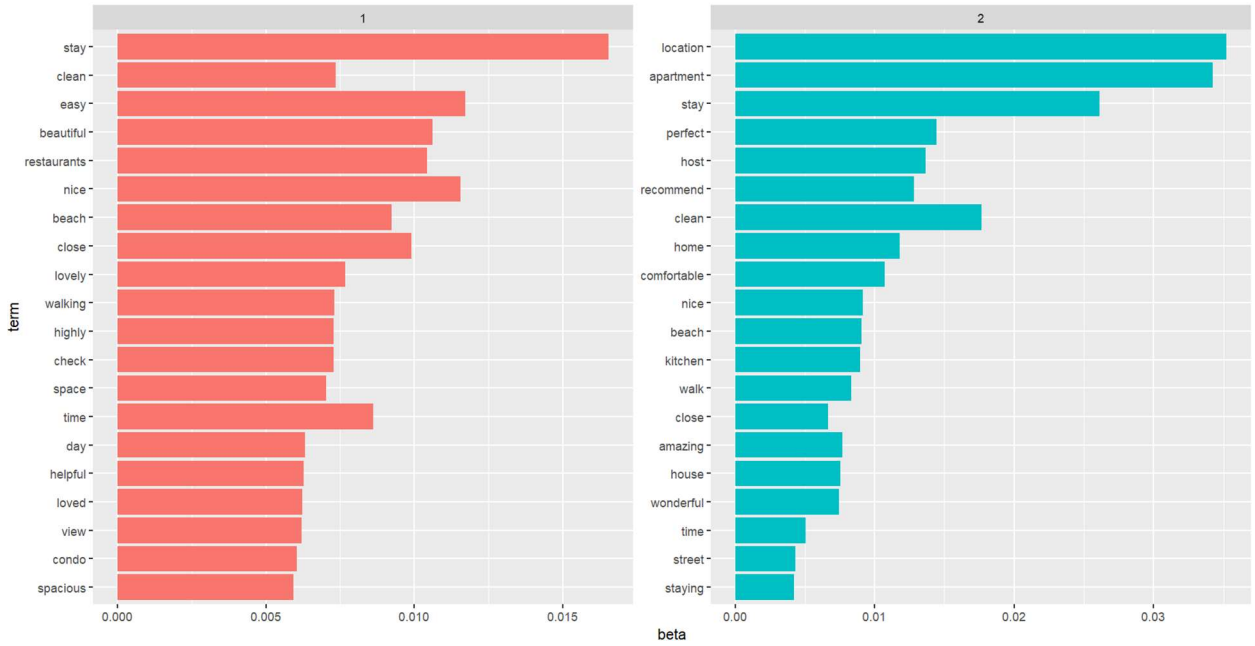

Figure 2.8 - Graph of two topics Latent Dirichlet algorithm (LDA)