A1: Combined assessment :: Text Mining / NLP

Airbnb in United States and Australia

Victor Isakov

Hult International Business School

Business Analysis with Unstructured Data - DAT-7471 - FMBAN1

Prof. Thomas Kurnicki

April 7th, 2024

Introduction

The following executive summary shows differences and similarities of listings in the United States and Australia. Moreover, this analysis covers the impact of sentiments and various phrases on the listings ratings. It is intended to be used by Airbnb managers to make adjustments to the advertising strategies, communication with homeowners, and how differences in listings by country can be leveraged for the success of both clients and the business.

Insights Report

First idea to drive insights comes from analyzing differences and similarities of descriptions of listings in the United States and Australia. Based on the correlation output and the associated correlogram (Appendix 1), there is a 95% correlation which suggests that listings have mostly similar descriptions and don't vary much overall. At the same time, the US descriptions mention more words like: *vacation, subway, apt, neighborhood, and island.* Whereas in Australia, words like: *centre, opera, hills, castle, and cinema* are more popular. Some assumptions might be based that people in the US would be more interested in using Airbnb services for vacations on an island, whereas in Australia sightseeing would be more popular in a description and catching the eye of the clients. If this analysis will be studied by the professionals who are native or more knowledgeable with each country, they will drive greater insights further.

Visuals 5 and 6 (Appendix 5 and Appendix 6) represent 10 most impactful positive and 10 most impactful negative sentiment words in the United States listings and in Australia respectively. This is insightful because it can help both Airbnb advertising strategy as well as homeowners. For Airbnb marketing strategy, the most impactful positive sentiment words should be leveraged in marketing campaigns whereas the negative ones should be taken out. To dive deeper and explore more word or even phrase options, this analysis can be expanded

which should drive a wider business value. For homeowners, Airbnb could implement sentiment analysis recognition into their platform to help adjust listings descriptions for homeowners to drive a better value for them and the business as well.

At the same time, visuals 7 and 8 (Appendix 7 and Appendix 8) show the effect of "best and worst" sentiment words in US and Australia listings respectively on the rating that the listings have which use these words. And as we see, the ratings with "good" and "bad" words are very similar. It means that the best and worst sentiment words don't have a great effect on the rating customers leave and doesn't change their satisfaction as much, on average.

Visuals 9 and 10 (Appendix 9 and Appendix 10) are the most valuable. They were created through LDA algorithm to divide listings into topics based on words. Based on the LDA visualization of 10 topics for listings in the United States, a few insights are clear:

1. Convenience: words *private, restaurants, walk, located, floor* focus on the convenience of accommodations. This might indicate the proximity to the nearby places.

2. Vacations: words *beach* and *guests* suggest properties for people looking to get a place for vacation on the beach or with friends and family.

3. Enjoyment: words *ocean, beautiful, enjoy* could be focused on properties for individuals who like having great experiences.

LDA visualizations of 10 topics for listings in Australia with a few insights:

1. Lifestyle: words *pool, home, walk, restaurants, private, beach* focus on properties for people interested in a particular lifestyle.

2. Attractions and locations: words *Syndey, station, Bondi, minutes, located* put emphasis on particular locations that clients of homeowners could be interested in.

Comparative analysis of Australia and the United States listings based on LDA visualizations brings the following insights:

- Australia and US listings focus on urban amenities, Australia based properties tend to lead towards a more beach and coastal lifestyle.

- Locations are important for both, but in Australia - focusing more on city names is more relevant than in the US.

Dashboard

The Dashboard shows 5 most impactful visuals with insightful data. It starts with the beginning, where it shows correlation of Australia and US listings. Further, it shows sentiment analysis of both regions to see if there are significant differences, and the final 2 visuals are the most important and impactful. LDA visuals show separation of each region listings into topics (default is 10 and 5 words per each). It is interactive so that stakeholders can adjust it based on their purpose to drive unique insights for various reasons. It shows main topic differences between regions and helps to better understand the markets.

Conclusion

This analysis concludes that Airbnb in Australia should leverage the touristic aspect and put emphasis on cities and renowned locations like Sydney and Bondi beach. While in the US - urban and natural attractions should be dominating. These insights suggest that tourists in Australia seek iconic city experiences and beach vacations, while the US audience appreciate more diverse interest from urban to rural locations.

Marketing strategies for Airbnb and the homeowners should focus on walkable neighborhoods with local access to amenities in Australia, and in the US it is better to emphasize a self-sufficient and comfortable spaces lifestyle. In Australia, advertisements should focus more on emotional and scenic journeys whereas the US should drive personal adventures and travel stories in diverse locations.
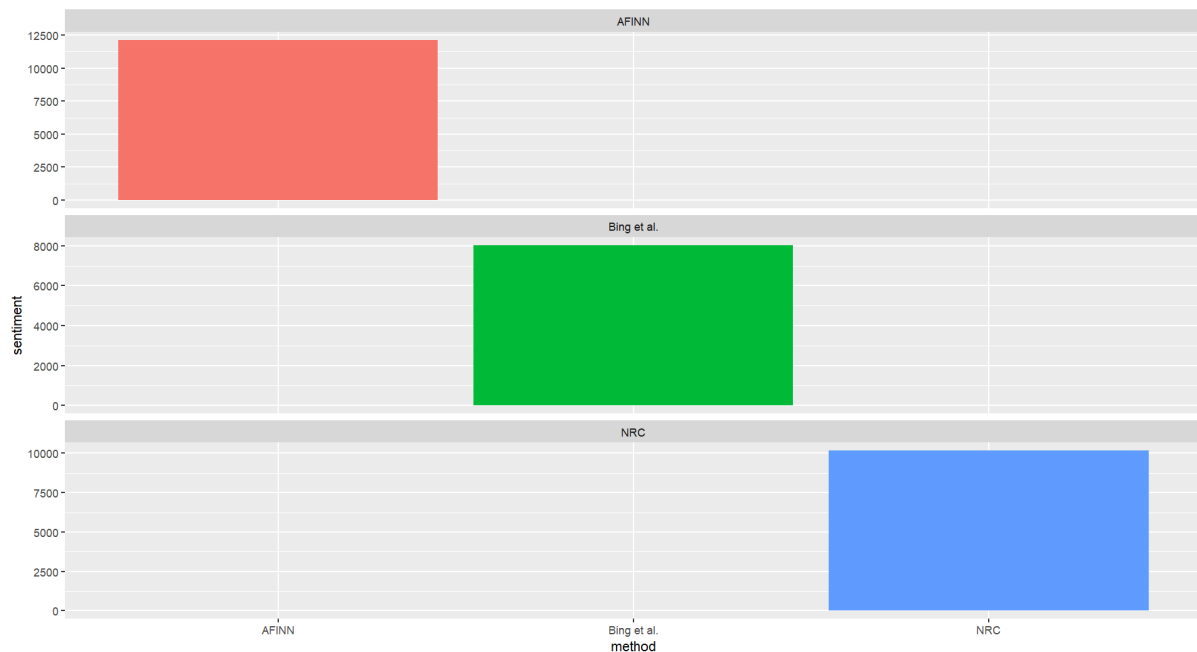
Appendix:
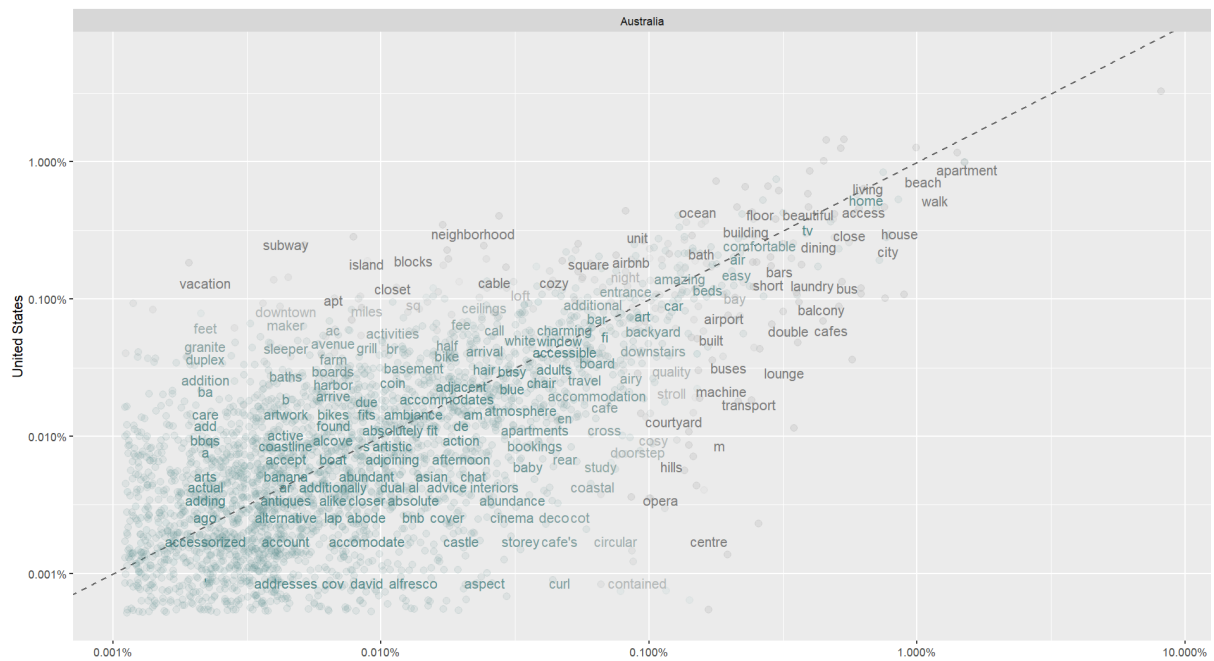
Appendix 1 - correlogram of descriptions of Australia and the United States
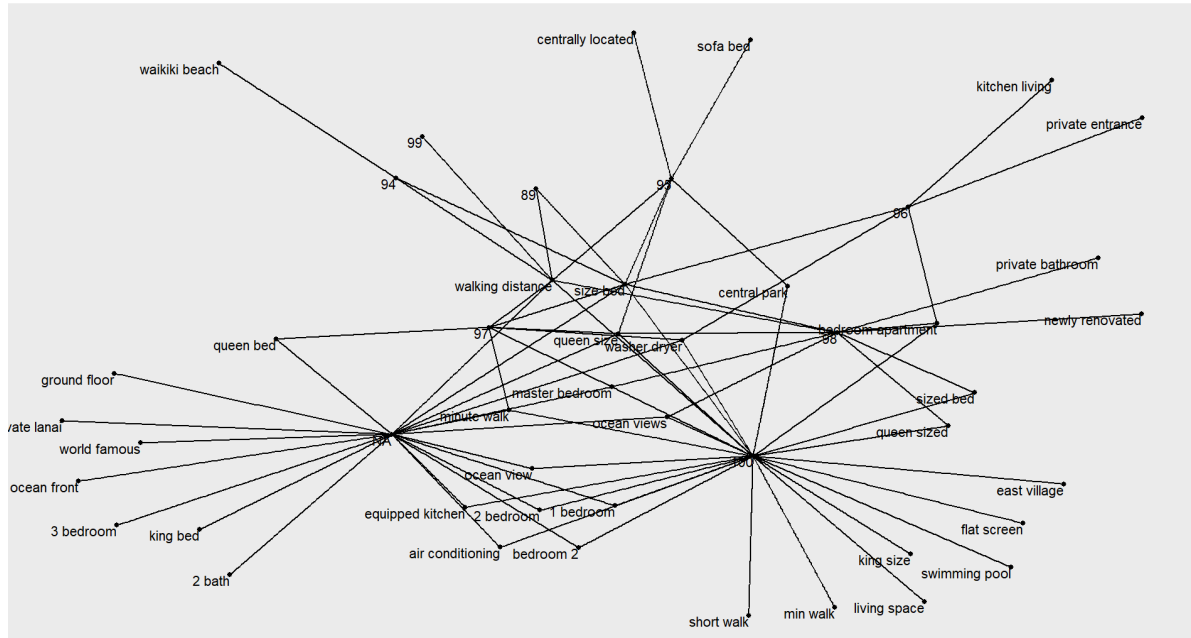
Appendix 2 - bigram graph of US listings' most popular phrases. This visualization

represents the most popular phrases from the descriptions of the listings in the US. (Appendix

2) It shows that the most frequent phrases are related to the amenities of properties. This bigram graph doesn't drive an insight as it is expected to have such phrases in descriptions.


Appendix 3 - Sentiment analysis based on 3 metrics in listings description in US. This Visualization of Sentiment analysis cover 3 main metrics: Afinn, Nrc, Bing. (Appendix 3 and Appendix 4) They might seem confusing because they add up the overall sentiment for each metric. This visual is not super insightful as it would make sense of listings to have a more positive sentiment in descriptions in order to attract customers.
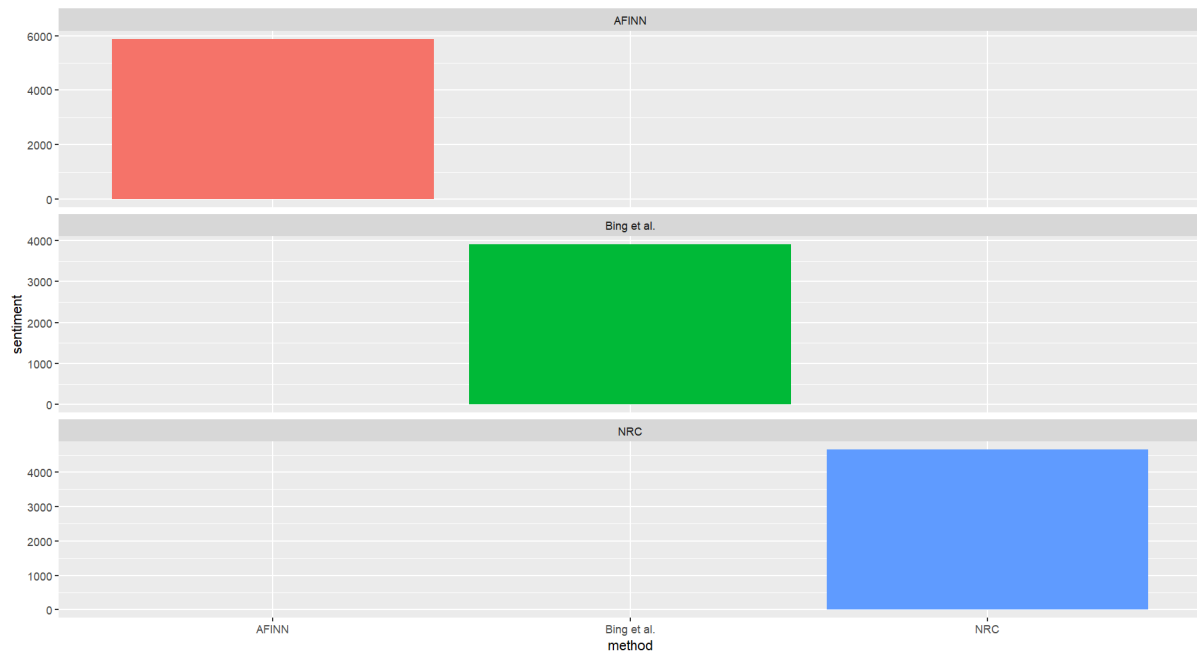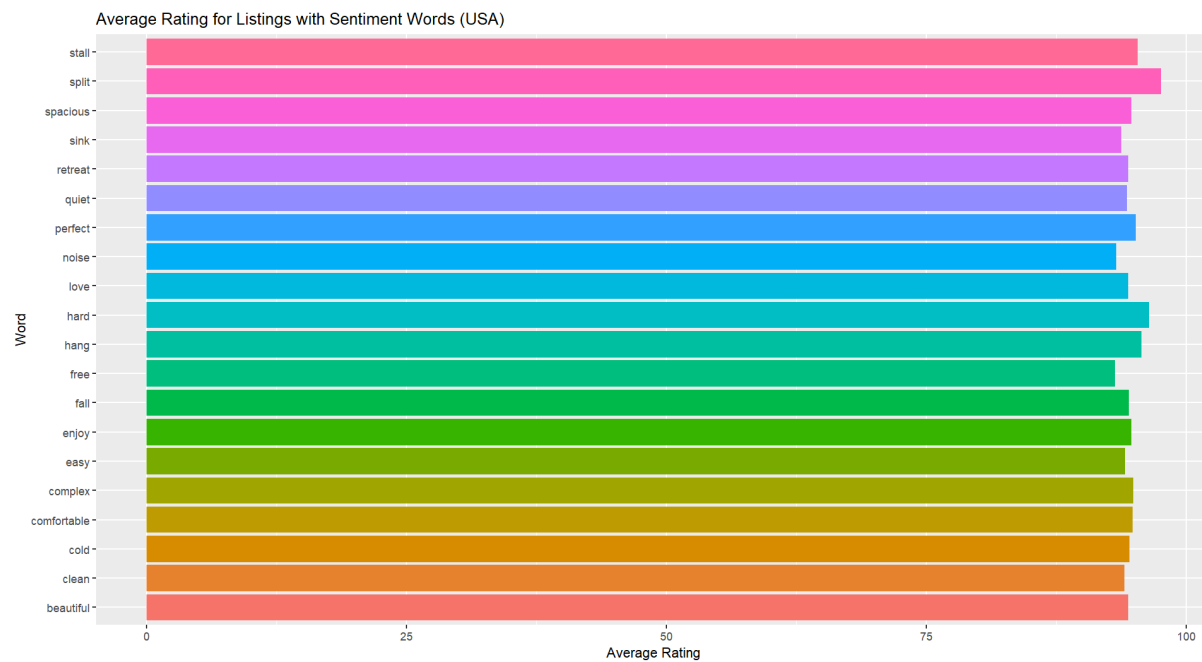
Appendix 4 - Sentiment analysis based on 3 metrics in listings description in Australia
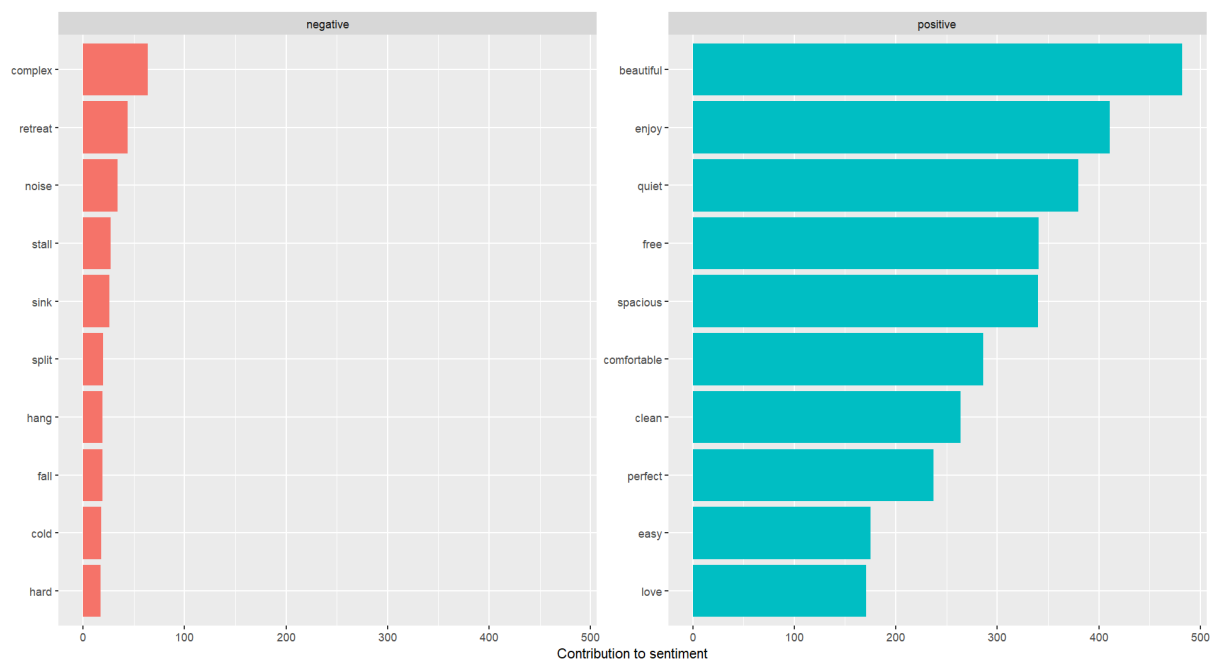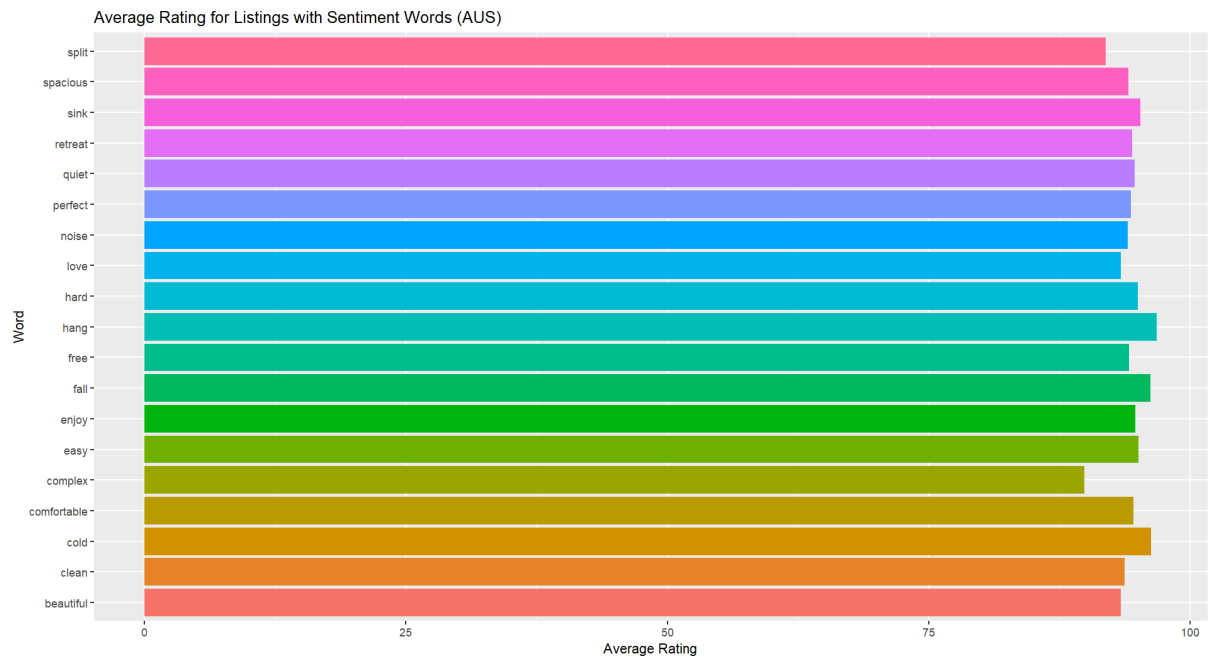

Appendix 5 - USA Sentiment Analysis
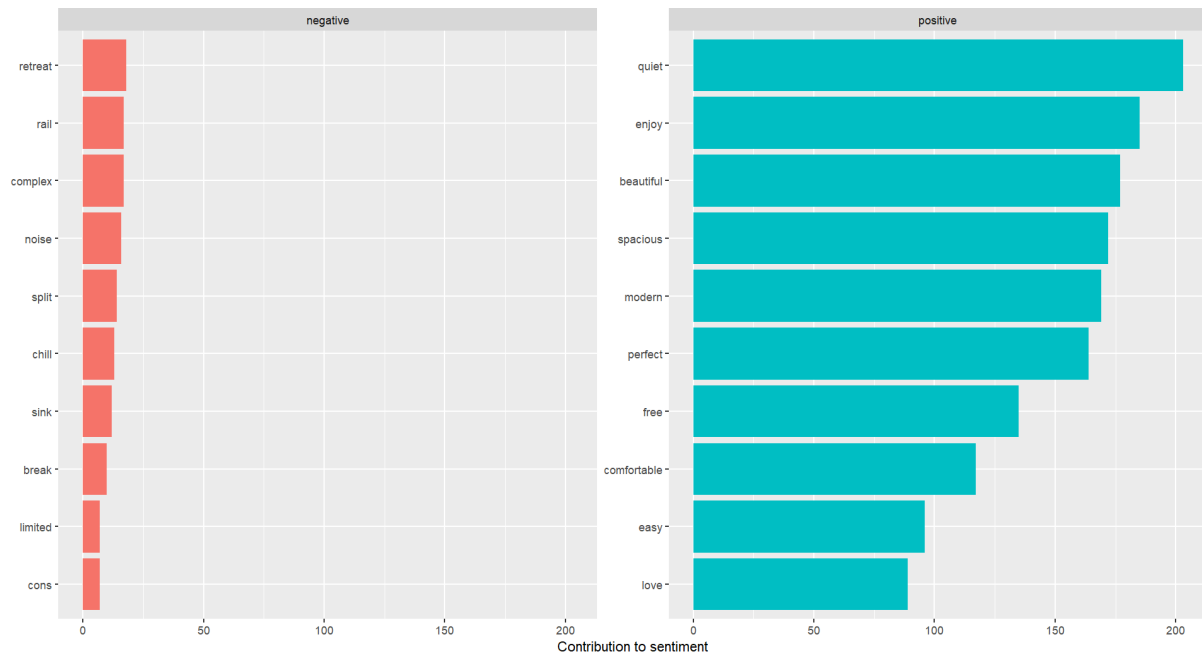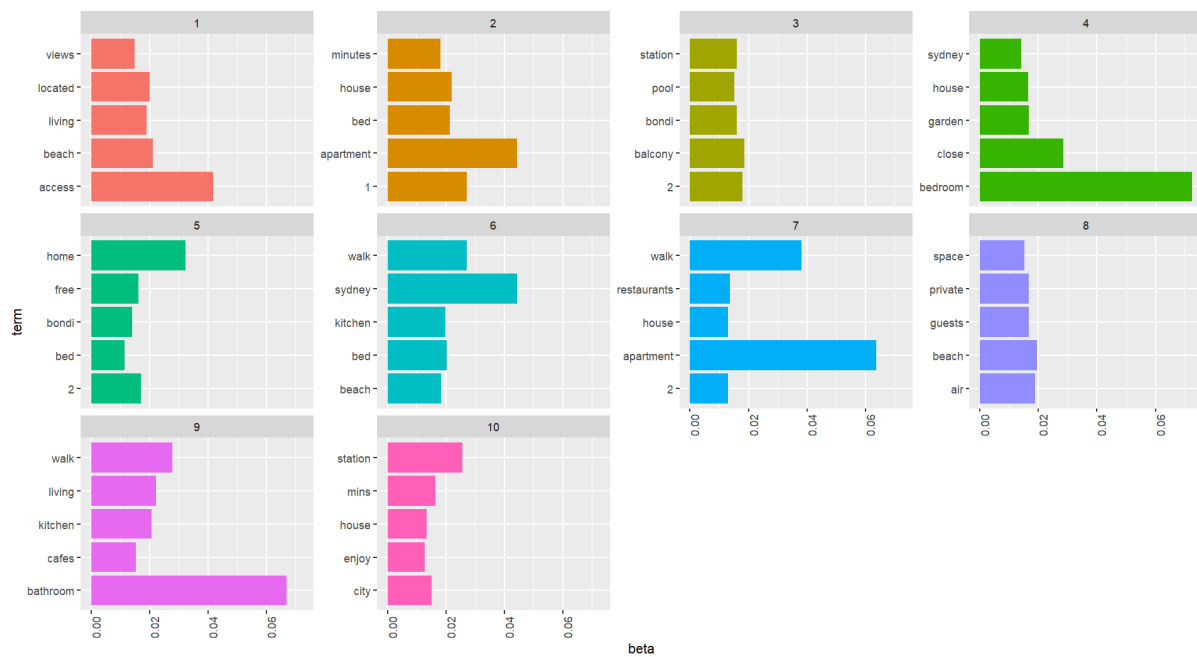
Appendix 6 - Australia Sentiment Analysis

Appendix 7 - average listing rating with best and worst sentiment words in the US

Appendix 8 - average listing rating with best and worst sentiment words in Australia
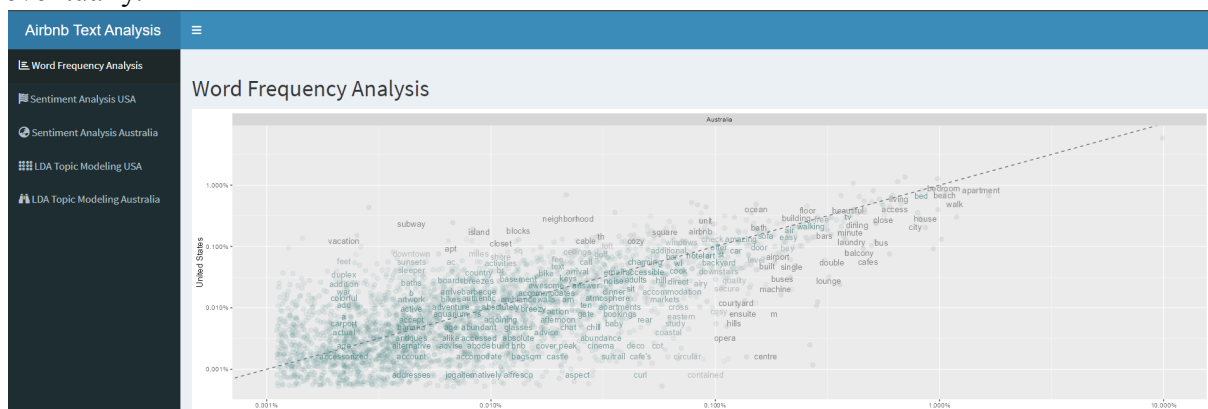

Appendix 9 - USA - defined words per topic per document. Which splits listings into 10 various types (using description).
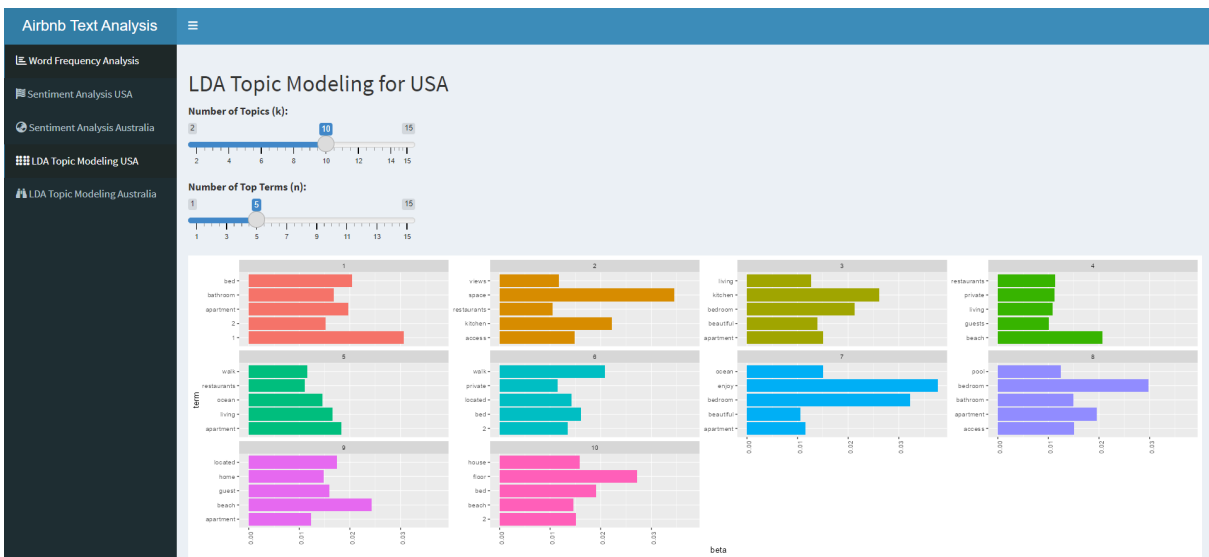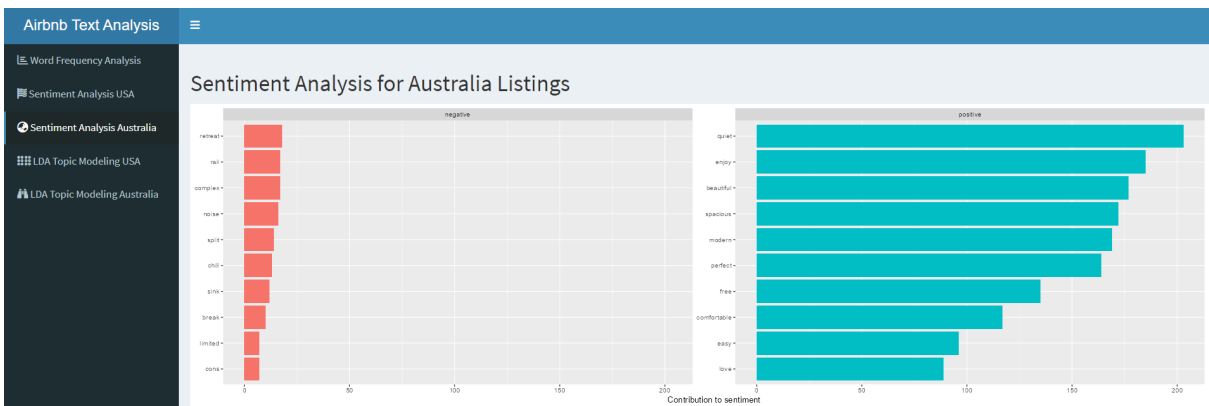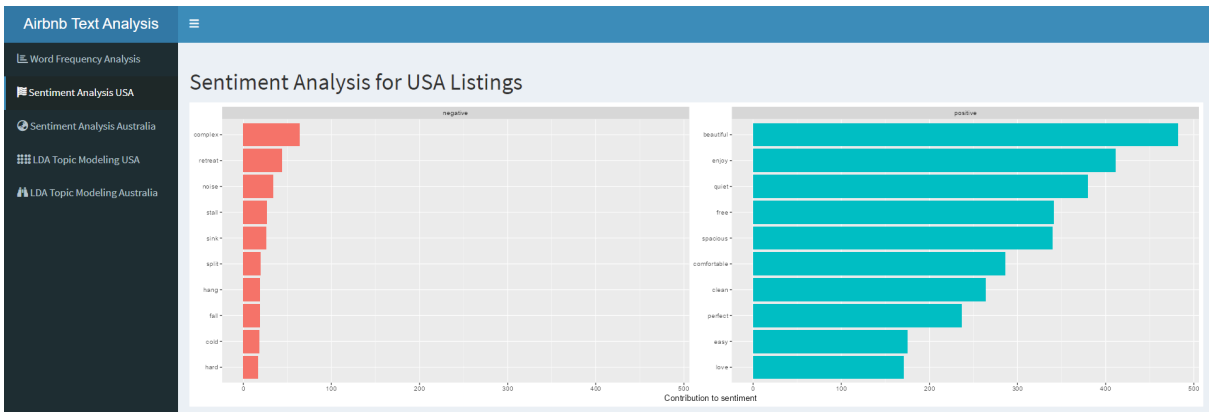
Appendix 10 - Australia - defined 10 topics (types of listings) based on words in description.

Shiny Dashboard:

Disclaimer: To run Shiny App properly and get faster outputs, the main R scripts need to be run prior, in order to create a global environment. Such an approach was chosen to increase the speed of Shiny App adjustments for the LDA parameters. Also, LDA parameters might be running slow if you have a slower computer. But wait until it finishes as it will load eventually.

Sentiment Analysis for USA Listings

Sentiment Analysis for Australia Listings

LDA Topic Modeling for USA

LDA Topic Modeling for Australia

Number of Topics (k):

2 | 10 | 15

2 4 6 8 10 12 14 15

Number of Top Terms (n):

1 | 5 | 15

1 3 5 7 9 11 13 15