

## **Executive Summary**

This report analyzes how Airbnb hosts in the U.S., Canada, and Australia use language in their accommodation descriptions and how word choice correlates with price, trust, and guest reviews. Using TF-IDF analysis and visualizations, key differences in vocabulary across countries and rating levels are identified. Well-structured and relevant phrases are consistently associated with higher scores. The presented findings offer opportunities to improve content quality, user experience and ad performance.

## **Methodology**

The analytical findings were also presented in an interactive dashboard created on the Shiny platform. Its structure and capabilities are described below.

The analysis covers textual and numerical data from Airbnb listings. Comparative cross-country analysis, trigram frequency analysis and segmented TF-IDF (by price and rating) were used. Visualizations were built in the R language (code provided in the appendix). The goal is to identify textual elements that contribute to placement success.

## **Airbnb Text Insights dashboard description**

To visualize the results of the analysis, an interactive dashboard was developed on the Shiny (R) platform. It contains tabs in the following areas: word frequency (by country), bigrams, trigrams and 4-grams, word association with price and rating, and country comparison charts. Navigation is done through the top panel. Each tab presents graphs built with ggplot2. Visualizations are updated automatically when the corresponding section is selected. The design theme used is Flatly. Dashboard allows you to explore textual patterns in Airbnb descriptions and compare the behavioral characteristics of hosts by region.

## Analysis and results

### 1. Differences and commonalities in word usage by country (see Appendix Figure 1):

Corellogram reflects the emphases in hosts' texts. Australian descriptions emphasize transportation accessibility (subway, airport access). Canadian ads often contain French words (e.g. cuisine, centre) and mention specific neighborhoods (Griffintown, Mont Royal), signaling proximity to downtown and transportation. In the US, universal terms focusing on housing characteristics (apartment, access) predominate. Despite the differences, the desire for comfort and predictability is evident in all countries.

2. The most popular **three-word (3-grams see Appendix Figure 2)** phrases such as “flat screen TV”, “king size bed”, “2 bedroom apartment” are especially common in the USA. In Canada and Australia, expressions related to climate and logistics (e.g. “reverse cycle air”, “transportation en commun”) are common. Such standardized language helps to set expectations and signals reliability. They can be used as a basis for text recommendations.

3. **TF-IDF by price segments (see Appendix Figure 3):** In addition to differences by price category, guest reviews remain a key factor, based on which rating groups were identified.

Low-price listings often contain irrelevant or unstructured words (privée, salon). In high-budget listings, terms indicating exclusivity and brands (miele, fireworks, flatiron). This reflects the desire to emphasize features that are perceived as valuable. TF-IDF can be used to automatically assess the quality of descriptions.

4. **TF-IDF by guest ratings (see Appendix Figure 4):** To analyze the effect of text on housing perception, ads were divided by rating into three categories: Low (<85), Medium (85-90), High (>90). The low-rated ones contain random or unrelated words (coffeehouses, kamakiki). High-

rated ones have clear, trustworthy language (read, community, https). Mentioning links such as https can indicate openness and a willingness to provide additional information, which reinforces trust and structure.

## **Strategic Insights**

### **Airbnb can use the findings to develop platform tools:**

- Offer template phrases when creating listings based on region and accommodation type
- Implement automatic text quality assessment based on TF-IDF.
- Support localization through language prompts
- Provide hosts with recommendations to improve structure and content

### **For hosts:**

- Use specific, tested phrases that match guest expectations
- Avoid generic and content-less descriptions
- Emphasize the strengths of the accommodation and take into account the local context

## **Conclusion**

Language is not just a description, but a conversion tool. The right language plays the role of brand signals, shaping expectations and reinforcing perceptions of value. In a highly competitive environment, structured text becomes a growth lever. With these findings in mind, you can create a scalable content strategy that benefits both hosts and the platform.

## Appendix

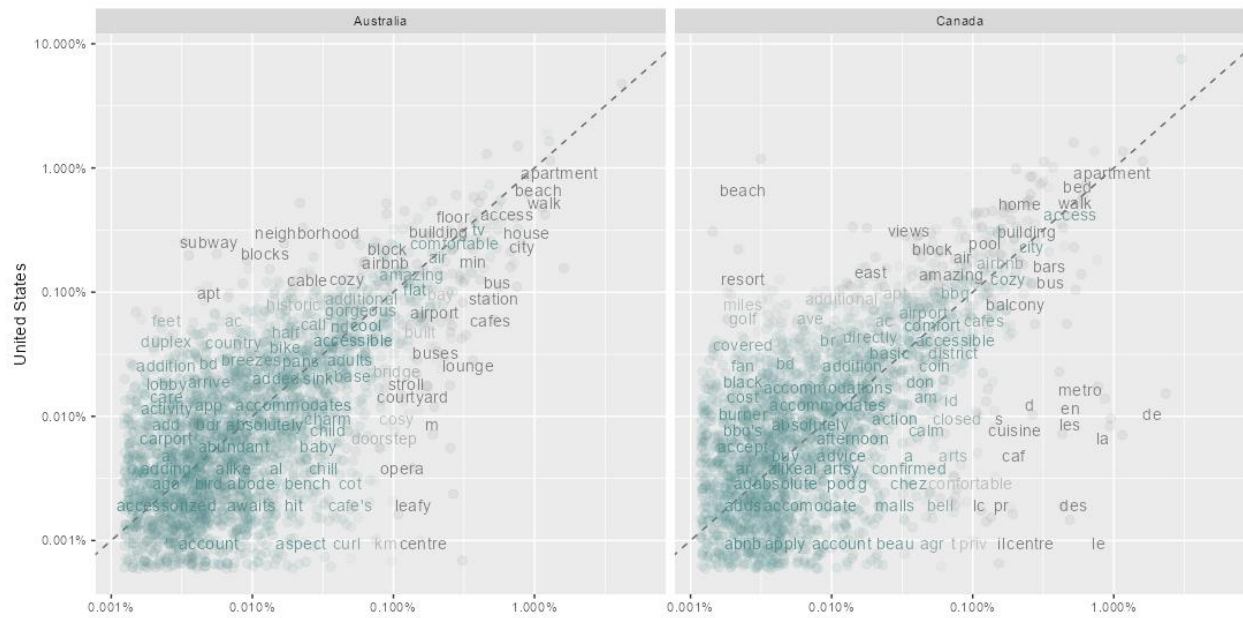


Figure 1 Corelogram - Word Frequency Comparison (US vs Australia / Canada)

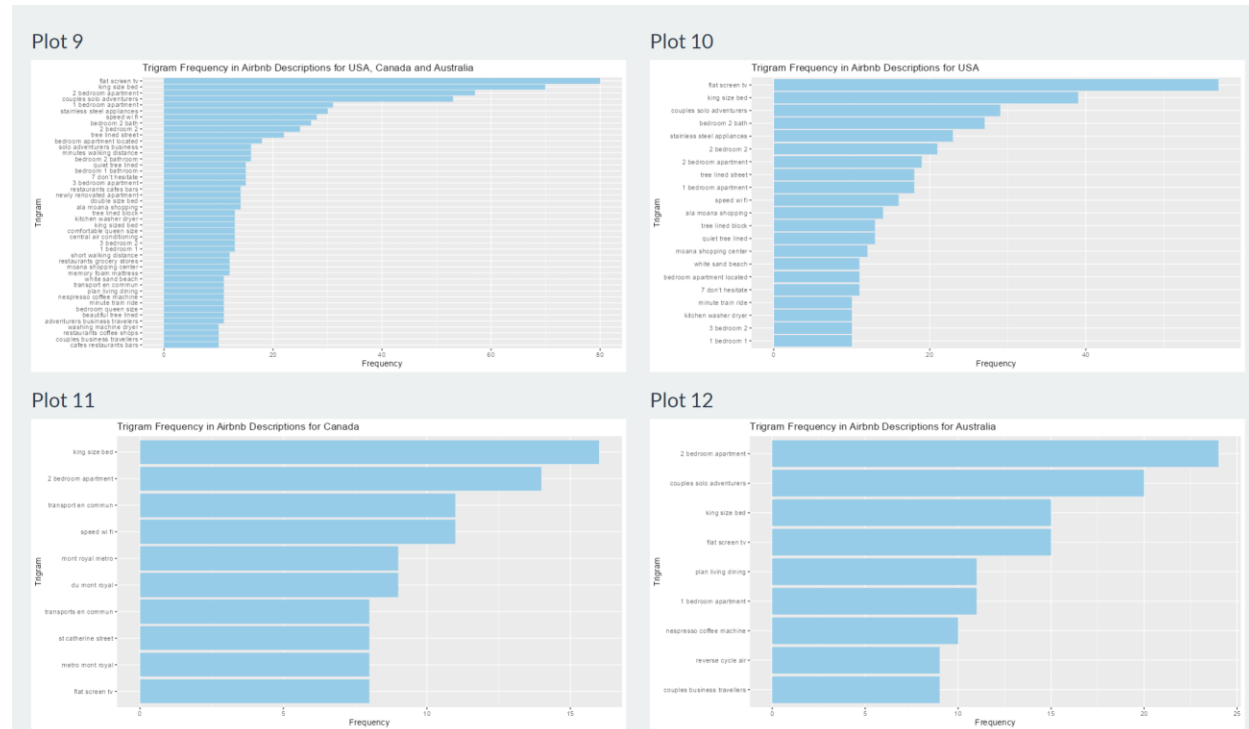


Figure 2 . Frequent Trigrams by country "The most common three-word phrases in the United States, Canada, and Australia"

A1: Airbnb Text Analysis

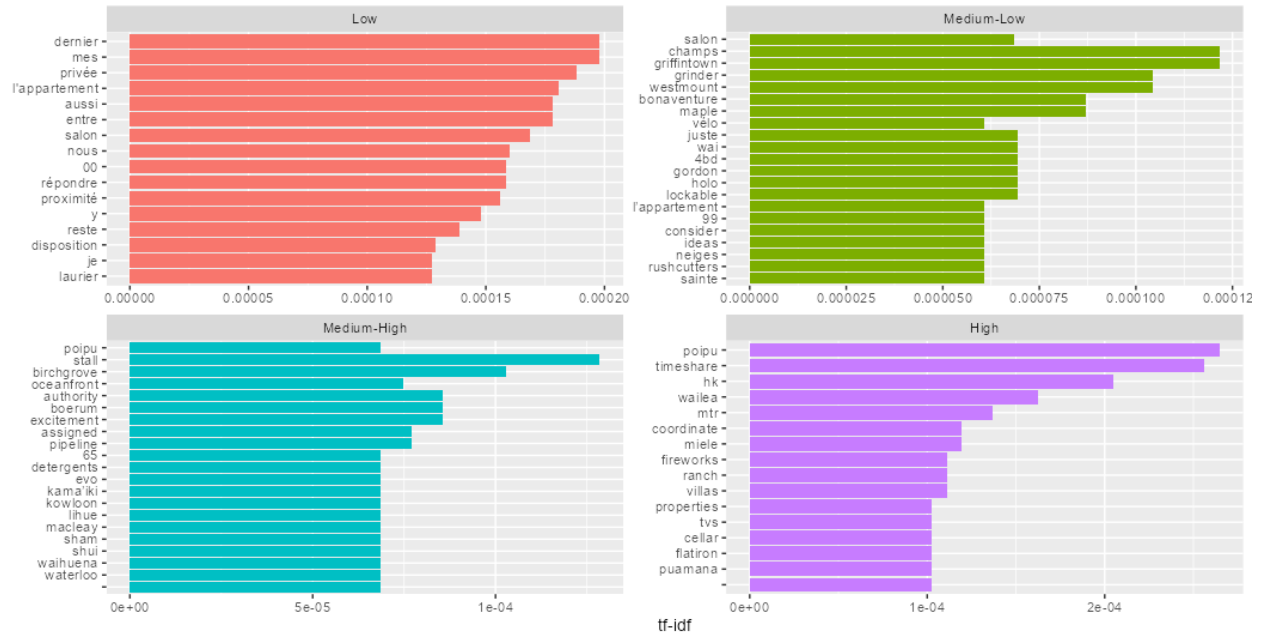


Figure 3 TF-IDF by price segments "Characteristic words for different price groups"

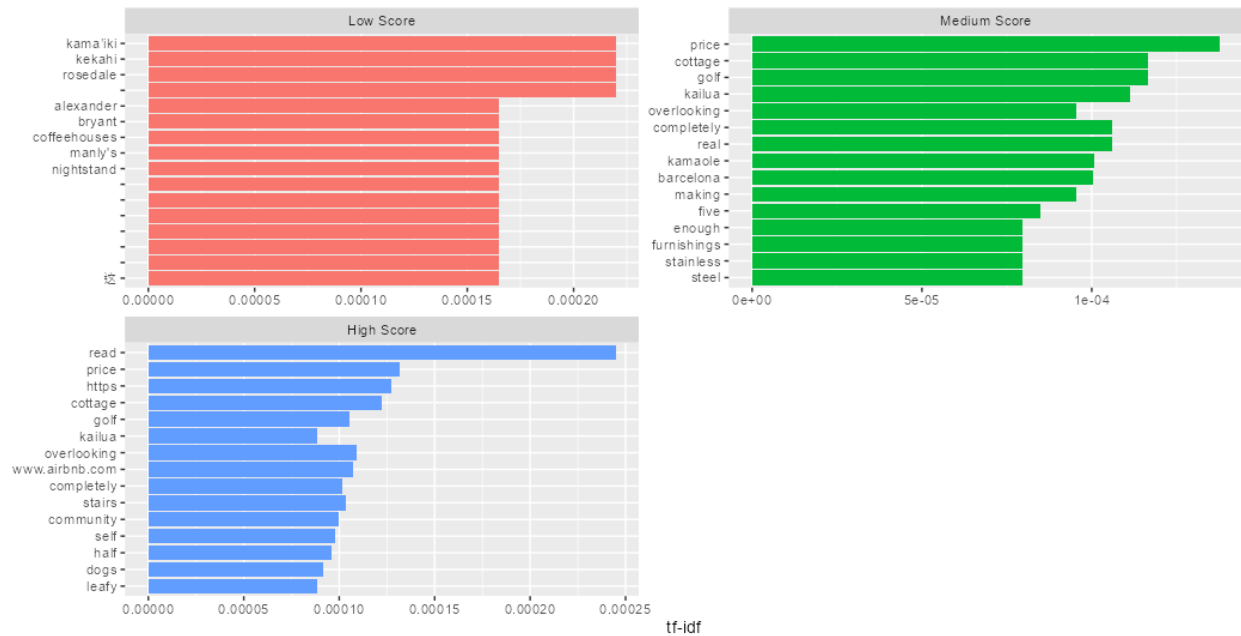


Figure 4 TF-IDF by guest ratings "Keywords by rating level"