Nirav Bariya

Achievement 6

Exercise 6.7

Date: 05/18/2024

## 1. The Story based on my complete analysis

**Airbnb** wants to design a recommendation system that will help hosts with respect to the pricing of the listings to offer better chance of being occupied consequently generating more revenue for the hosts and consequently increasing profit margins for Airbnb.

They want to have a better understanding of the listing types, user types, neighbourhoods, to help better understand the Toronto Airbnb market.

They are also interested in identifying variables that can help predict price of a listing.

**We hypothesized that more bedrooms should result in higher price.**  But linear regression isn’t very useful in this case as it doesn’t completely describe the variation in pricing.

We also hypothesized that if a listing is available for more days in the next 30 days, it is likely to be available in the next 60 days. This hypothesis held true.

Our target is to identify successful listings and recommend a price for the listing. Therefore, we must focus on listings that are occupied frequently and received higher and more number of reviews.

We used K-means clustering algorithm to create clusters so that we can identify successful listings.

We then share the results from the clustering algorithm and what can be done to improve the clustering. This isn’t an ideal situation and the clusters aren’t very clear. None the less, we can glean some important insights from the clusters produced in this case.

* Show results of the analysis.

Finally, give recommendations based on the analysis.

* Include important aspects that I discovered
* Give recommendation for pricing
* Tell how you might improve the clustering process
* Any other recommendations

create a business case by answering the guidance questions (who, what, when, where, and why)

**Who** is this analysis for?

**Who:** The client is Airbnb.

**Why:** Airbnb wants to recommend the fairest prices for their real estate properties.

**What:** The storyboard will provide more information about what determines price of a listing and how pricing should be done.

**When:** It will be used to suggest prices for new and existing hosts based on the information they enter in the data field.

**Where:** Tableau Public.

They are also interested in exploring the host data and want to know the answers to the following questions.

**Questions we will explore:**

Basic Summary of the statistics

What is the distribution of super host?

**How are the listings distributed in Toronto? Are certain neighbourhoods have more listings than others? Which neighbourhoods priced more than others? Use maps to count the listings and price:**

What is the distribution of property type and room type in Toronto? **Property type using Tableau.**

Exploring relationships:

* availability\_30 and availability\_90
* bedrooms and price  
  bedrooms alone cannot help in predicting price the correlation coefficient is not strong.
* **Is there any relationship between property type and price?**

Creating clusters to better understand pricing clusters and how may we recommend the pricing.

Future steps:

NLP What are the common amenities offered in the listings?