Modeling Price and Popularity of AirBnB listings in New-York

Melody Jiang and **Raphael Morsomme** and **Ezinne Nwankwo**

Abstract

This is abstract.

1. Introduction

This is introduction

2. Methods

2.1. Data

Here we describe data.

2.2. Feature Engineering

Here we describe feature engineering.

2.3. Ordinal Logistic Regression Model

Here is our model.

3. Results

3.1. EDA

Figure 1 shows the distribution of required minimum number of nights to book a listing. We observe that minimum number of nights is concetnrated at below 14 days and around 30 days. Figure 2 displays the distribution of days that listings are available for booking in a year. We see that there are data concentrated at 0, which means these listings are not open for booking. Such observation would inform our data cleaning.

To address the quesstion of whether the type of listing (shared room, private room, entire home) vary across neighbourhoods, we performed a chi-square test and plotted the results in Figure 3. The p-value of the chi-square test was less than 10^{-16}, indicating that the type of listing does vary across neighbourhoods. In Figure 3, the size of dots represents the absolute standardized residuals. The color represents the value of standardized residuals. We see that difference in room types is most pronounced in Manhattan and Queens. Manhattan has more entire home than expected and Queens has more private room than expected.

Figure 4 shows a spatial map of of listings, metro stations, and attractions. Black dots represent metro stations, gree dots attractions, blue dots listings priced at bottom 80%, and red dots are listings priced at top 20%. We observe that listings priced at top 20% distribute close to metro stations. This observation motivates us to include spatial in formation of metro stations as explanatory variables.

3.2. Main Findings

Main findings.

3.3. Sensitivity Analysis

Sensitivity analysis.

4. Conclusions and further discussion

Conclusion and further discussions.

References

Appendix A. Figures and Tables

Figure 1: Distribution of minimum number of nights.

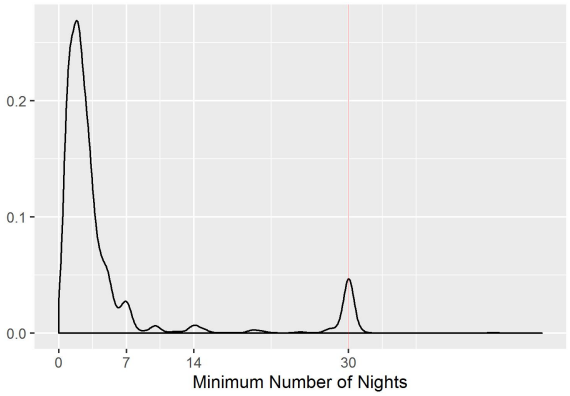


Figure 2: Distribution of number of days available for booking.

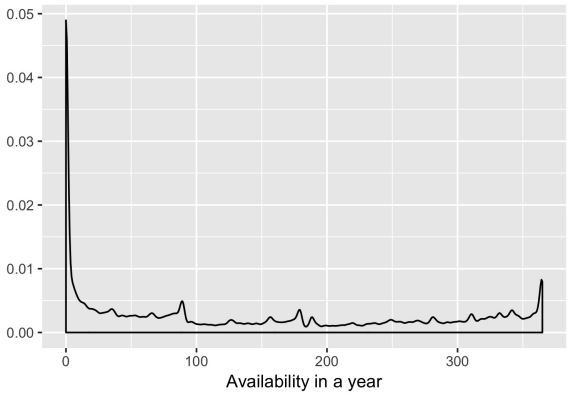


Figure 3: Output from chi-squared test.

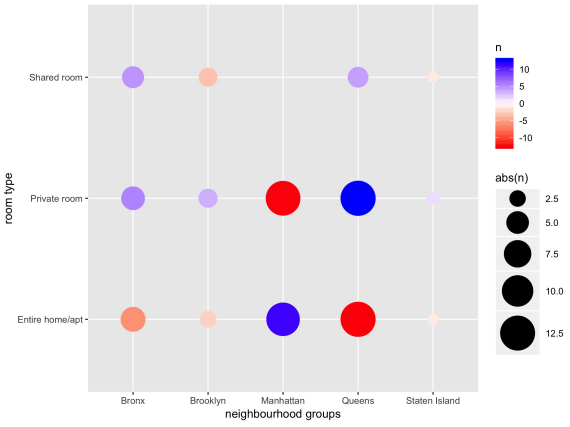
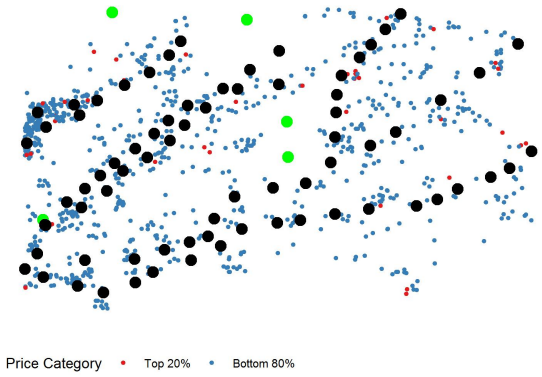


Figure 4: Map of listings, metro stations, and attractions. Black dots are metro stations, gree dots are attractions, blue dots are listings priced at bottom 80%, and red dots are listings priced at top 20%



Figures here.
Tables here.

Appendix B. Model Checking

Model Checking here.

Appendix C. Full Model Output

Full model output here.