Airbnb Listings Analysis in Beijing & Boston

Team 4, Cohort B, MSBA'20
Questrom School of Business

Dataset Description & Handling Variables / Missing Values

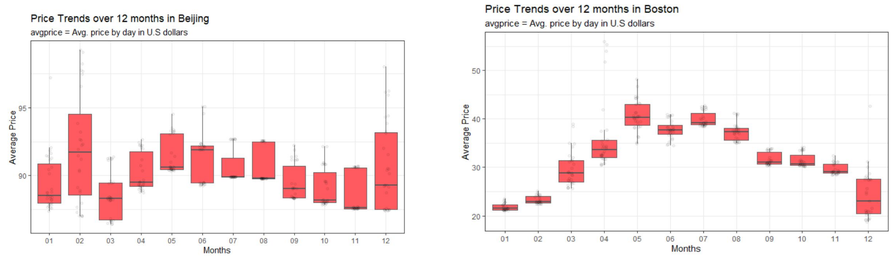
- Three **sub-datasets**
  - listings, reviews and calendars
- Beijing & Boston **Listings** :
  - **18** predictors and **1** output "price"
- Beijing & Boston **Reviews** : **9** variables
- Beijing & Boston **Calendar**: **9** variables
- **Transfer** text and categorical variables into dummies
- Use **Median** and **Mean** to replace missing values
- Perform **Boruta** to select variables according to importance scores

Abstract

- Comparing **demand and supply** between Airbnb homestays in Beijing and Boston
- Establishing supervised machine learning models and use house info to predict the **housing price**
- Implementing text and sentiment analysis to analyze Airbnb homeowners (hosts) and customers (visitors) **preferences and tastes**

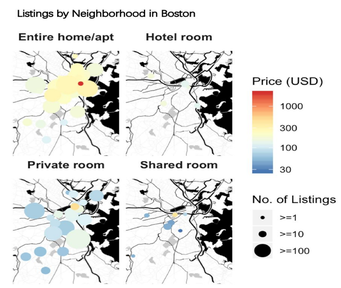
Methodology

- **Exploratory Data Analysis**
  - **Price Trends by Season** – for Beijing (left) & Boston (right)

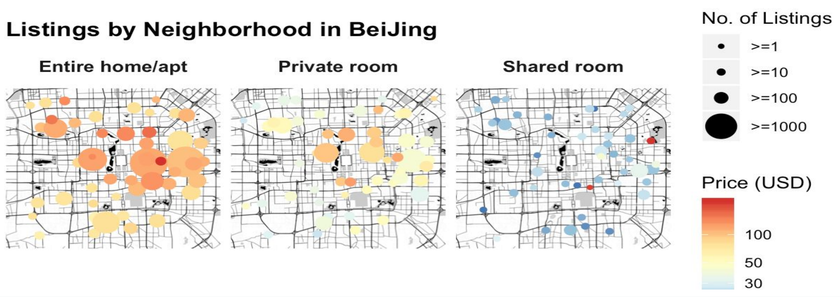


- **Pricing Modeling: Supervised Machine Learning**
- **Text Analysis: Sentiment Analysis for Comments**

- **Exploratory Data Analysis**
  - **Price & Number of Listings in Beijing and Boston by Neighborhoods**



- The price in Beijing is distributed by its intrinsic "Ring-shape" cityscape.
- The price in Boston goes higher as the location gets closer to downtown



Results

- **Linear Regression** – Highest Interpretability
  - **Positively** Related to Prices For Beijing:
    - Bedrooms, Bathrooms, TV
  - **Positively** Related to Prices For Boston:
    - Accommodates, Neighborhoods in Beacon Hill, South Boston & West End
- **Random Forest** – High Predictability
  - **Important Variables For Beijing Listing**
    - Availability of house in 90/365 days
    - The number of listings that host has
    - The number of accommodates
  - **Important Variables For Boston Listing**
    - The number of accommodates
    - The number of listings that host has
    - The number of reviews
- **Gradient Boosting** – High Predictability
  - **Locations** are important in **Boston's Airbnb Marketplace** but not in Beijing's
- **XGBoost** – Highest Predictability with Low Interpretability
  - R-squared For Beijing Listing: **55.71%**
  - R-squared For Boston Listing: **64.80%**
  - Test MSE For Beijing Listing: 6,518
  - Test MSE For Boston Listing: 3,367
- **Text Analysis** – top positive words for Beijing (left) & Boston (right)

Conclusion

- **EDA**: obtain differences in **price trends** and **geographic distributions** of the homestays
- **SML**: predict the **prices** of listings to assist Airbnb hosts in posting suitable prices
- **Sentiment analysis**: gain insights into the different customers' **tastes** in Beijing and Boston
- **Suggestions for Hosts in Beijing**
  - Have friendly and helpful attitudes
  - Set proper renting length and available periods
- **Suggestions for Hosts in Boston**
  - Focus on the quality of the house
  - Have a shorter response time
  - Consider an appropriate housing price
    - natural competitive disadvantages for suburban areas