Airbnb: Effect of Neighbourhood Factors on Occupancy Rate

Anushital S. Siddiqi asiddiqi30@gatech.edu
Fei Chi fchi6@gatech.edu
Gaurav Chakravarty gchakravarty6@gatech.edu
Mohammed Kashif Uddin Ansari mansari44@gatech.edu
SaiKrishnan Subramanian ssubramanian99@gatech.edu
Sharat Chandra schandra46@gatech.edu

- Airbnb has led to financialization of houses by creating hyper flexible rental market
- There exists demand for decision making tools for Airbnb investments
- ***Motivation*** for this project: answer the difficult questions faced by Airbnb investors e.g. how important neighborhood factors are in determining investment return

Problem Definition

- Explore the impact of neighbourhood factors - restaurants, tourist attractions and median household income on the occupancy rate of Airbnb rentals

Intuition

- Investment decision is based mainly on property, service quality & host controlled factors
- Neighbourhood factors often ignored by the investors but quite important for renters. We hypothesize that neighborhood factors should impact occupancy rate and should be part of the investment decision process

Algorithms & Results:

LASSO

- Least Absolute Shrinkage & Selection Operator
- Adjusted R-Squared value: 0.2167
- Rejected - weak correlation

XGBoost

- Extreme Gradient Boost
- high execution speed & model performance
- Moderate R-Squared value: 0.4614
- Suitable for the study

Innovation

- Using 3 important neighbourhood factors to study their impact on the ***occupancy rate***
- Neighbourhood factors studied : median_household_income_in_1999, restaurant_count, attraction_count
- Data sources used are : InsideAirbnb.com , data.census.gov, Google Places APIs
- Other similar studies :
  - Connection between spatial location of AirBnB properties to hotels
  - Effect of tourism clusters on AirBnB performance
  - Effect of factors in control of the host to AirBnB occupancy rate

Data

- Downloaded
  - Airbnb data from insideairbnb.com : approx. 10k rows, 7.2 MB on disk
  - Median HH income data from census.gov for each zip code : 3698 rows, 182 kb on disk
- Size of Google Places data is l`arge and took hours to collect. It stored about 180K records.
- Called Google Places API to get nearby restaurants and tourist attractions
- Called 2 APIs for each row, each API = approx. 100 KB, total number of API calls = 20,000
- The size on disk of raw input is approximately 2 GB.

Features

- Weak correlation between the predictors and Response variable (Figure 1)
- Occupancy rate = 1 - availability_365 / 365
- Engineered a few columns
  - Filtered out Bedrooms > 9
  - Zipcode down to 4 digits level
  - Amenities reduced from 1000 to 7: chose the most impactful values
- Null Values replaced by :
  - Median HHs Salary -> the median value of the column
  - Num. of Rest and No. of tourist attractions -> zero

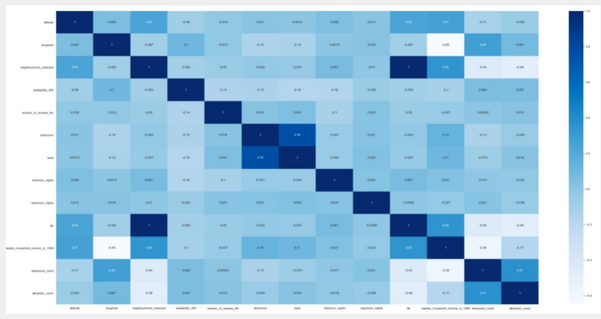

Fig 1: Correlation Heatmap (labels intentionally not visible)

Experiments

- Study the effect of the following on the Occupancy Rate of the Airbnb Property:
  - Number of restaurants in the neighborhood
  - Number of tourists' attraction in the neighborhood
  - Median household income of a neighborhood
  - Combinations of Median household income and number of restaurants, median household income with the number of tourist attractions, number of restaurants with number of tourists' attraction
  - All the three neighbourhood factors combined

UI

- Used Flask to create the GUI (Figure 2)
- User gives some inputs and the GUI returns the occupancy rate

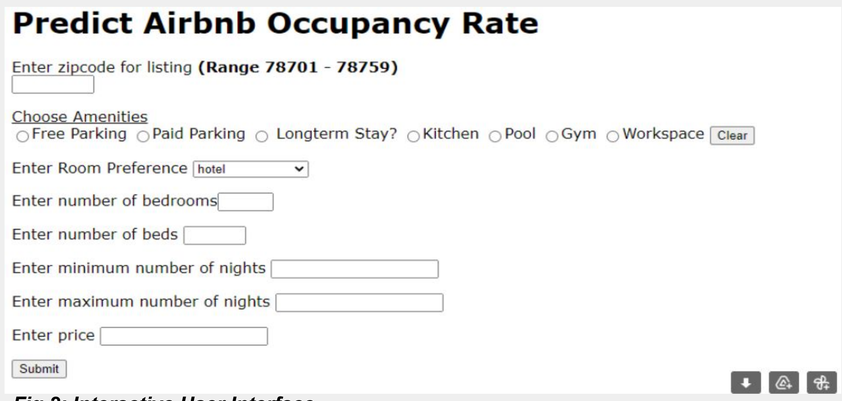

Fig 2: Interactive User Interface

Visualization Chart : Insight

- Occupancy rate for 8 bedrooms jumps significantly when the restaurant count is 18+ (Figure 3)
- Hypothesis: Large properties have better occupancy rate when there are more restaurants around possibly due to being in areas like downtown or posh areas of a city

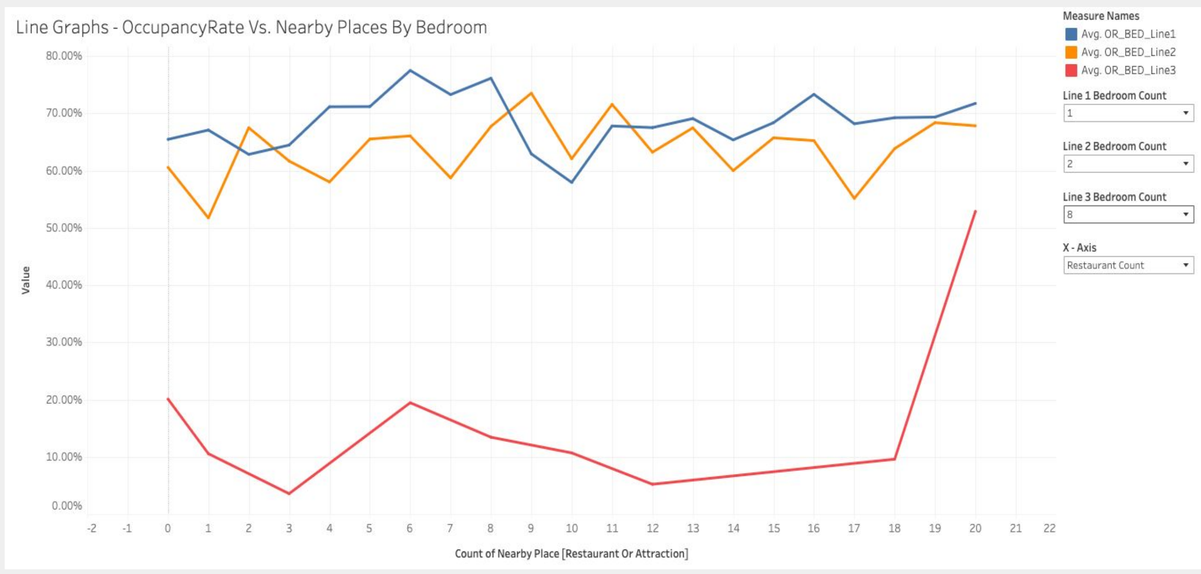

Figure 3: Occupancy rate vs Restaurant Count for 1 bedroom, 2 bedrooms, 8 bedrooms

Conclusion

- Adj R-Squared without neighbourhood factors = 0.46142
- Adj R-Squared with all 3 neighbourhood factors = 0.48565
- Median household income has the most impact on occupancy rate
  - Individually : Adj R-Squared of 0.480232
  - Combined with other factors : Adj R-Squared of 0.480368 (tourist attraction)and 0.482381(restaurants)
- Yes, neighbourhood factors have a positive effect on the occupancy rate
- Richer neighbourhoods have higher occupancy rate
- Visualization insight : Bigger properties have higher occupancy rate in locations with more restaurants possibly downtown
- Investors should take these factors seriously
- Further exploration recommended for other neighbourhood factors like:
  - neighbourhood walkability score, neighbourhood crime rate, neighbourhood gentrification levels, distance from concert venues, distance from sporting venues etc.