

## **Assignment: ASSIGNMENT 12.3 Final Project Step 3**

**Name: Ghanta Madhavi**

**Date: 2023-05-30**

**Project: Analysis of how AirBnB rentals prices affects the nearby housing rental prices in Chicago**

### **Introduction**

Airbnb, Inc operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. The platform is accessible via website and mobile app. Airbnb does not own any of the listed properties; instead, it profits by receiving commission from each booking.

The company has been criticized for a direct correlation between increases in the number of its listings and increases in nearby rent prices and creating nuisances for those living near leased properties.

The problem here I am addressing is how the the prices of Chicago AirBnB rentals affect the prices of the nearby neighborhood rent prices.

Data science algorithm will help here to predict the prices of Chicago AirBnB rentals and also help to understand the correlation between the prices of Chicago AirBnB rentals and neighborhood rent prices.

### **Problem statement**

The company has been criticized for a direct correlation between increases in the number of its listings and increases in nearby rent prices and creating nuisances for those living near leased properties.

### **Research questions**

- 1) What are the Airbnb rental prices for different areas in Chicago?
- 2) What is the correlation between the Airbnb rental prices and Chicago neighborhood rent prices?
- 3) What are the average rent prices by the neighborhood?
- 4) What are the average rent prices for Airbnb by the neighborhood?
- 5) What type of houses are most rented on Airbnb?
- 6) What is the monthly rent from the Airbnb properties?
- 7) What are the rental property options by neighborhood? 8) How much profit does Airbnb made monthly?

### **How you addressed this problem statement**

- 1) **Data research and collection** – I have used inside airbnb website to gather the data for AibBnb listing data, Affordable rental housing data and Average rent Chicago neighborhood data. Each row of the data represents property lister on Airbnb, their prices, neighborhood, zip codes, average rents in that neighborhood. The data is focused on 'Chicago' city.

- 2) **Data preparation and cleansing** – Identified that there were some missing data in the datasets. I have removed those records which have NA. I have also dropped some of the fields that are not use for the analysis. I have merged the 3 datasets in one final datasets for the analysis.

### **Analysis**

**EDA (Exploratory data analysis)** – I did the correlation analysis between variables and noted the strength and weaknesses of relationships.

- a) I found that zip codes, neighborhood, price, average rent, property type are good predictors for the analysis. Once the predictors are decided then I looked into the R2 , Adjusted R2 statistics, p-value.
- b) Visualize different aspects of the data to gain more knowledge.
- c) I then calculated the betas for the predictors in the regression model. It show me how the 1 standard deviation change in predictor will impact dependent (response) variable.
- d) I then calculated confidence intervals which indicate that the estimates of how the model are likely to be representative of the true population values.
- e) I then performed an analysis of variance on all models to compare the performance of different models.
- f) I then calculated standardized residuals, the leverage, cooks' distance, and covariance rations
- g) At last, I checked if the regression model was unbiased.

### **Implications to consumer**

The implication of the research is that the prices of the housing rental have direct impact on AirBnB listing in that same neighborhood. The recommendation from research is that there should be federal rule on how much the housing prices should increase year by year.

Also AirBnB should consider the neighborhood housing prices when deciding prices for their listed property.

### **Limitations of the analysis**

- 1) The research is limited to Chicago city only.
- 2) The research needs more sample size for accurate analysis.
- 3) The research datasets are gathered from only one source.
- 4) The research is limited based on neighborhood. There are other factors needs to be considered for more analysis.
- 5) The research is done using linear regression. There is a scope of improvement by fitting other ML algorithms.

### **Concluding Remarks**

The research helped in applying the concepts and knowledge of the statistics gain in the course.

The research project provides hand-on experience in the real life case study.

The research methods defined in the courses were very helpful

The visualization technique learned through this case study can be used in other research too.