CIS 500 Data Analysis and Decision Testing

Project Proposal

Project Title: Analysis of Price Discrepancies Among New York City Airbnb Rentals

Introduction:

In recent years, the hospitality industry has undergone a significant transformation with the rise of the sharing economy, exemplified by platforms like Airbnb providing alternative lodging options. Despite its popularity, concerns linger about potential inequities in pricing across diverse demographics and neighborhoods. This project aims to explore and illuminate any price disparities present in NYC Airbnb listings, with the objective of addressing issues related to fairness and equity within the short-term rental market.

Research Questions:

1. Are there significant differences in average listing prices between neighborhoods in NYC?
2. How do prices vary across different types of accommodations (example: entire homes, private rooms, shared rooms)?
3. What factors contribute to pricing variations, such as neighborhood characteristics or property amenities?
4. Can we identify any patterns of price discrimination based on demographic factors like race or ethnicity?
5. Are there regulatory factors impacting pricing, such as licensing requirements or zoning laws?
6. How do seasonal fluctuations affect listing prices?

Data Source:

Dataset for this project will be taken from Kaggle, where it is titled as 'New York City Airbnb 2023, Public Data'. Here the data set consists of attributes such as id, name, host_id, host_name, neighbourhood_group, neighbourhood, latitude, longitude, room_type, price, minimum_nights, number_of_reviews, last_review, reviews_per_month, calculated_host_listings_count, availability_365, number_of_reviews_ltm, license.

Methodology:

The methodology involves thorough data cleaning to remove duplicates and handle missing values, followed by exploratory data analysis to visualize key variables and identify correlations. Statistical analysis will be conducted to examine relationships between variables, while machine learning models will be developed to predict listing performance and prices. Spatial analysis will utilize geographical data to map listings and identify spatial

patterns, while regulatory analysis will investigate the impact of factors like licensing on listing characteristics and performance.

<u>Output:</u>

The outcome of the project will be culminated in a comprehensive report presenting findings from statistical analyses, visualizations, and machine learning models. Key outputs include maps illustrating price distributions, regression coefficients highlighting significant predictors of pricing, and insights into potential disparities or biases in the Airbnb market.

<div align="right">

By Team – 2
Vathasalya Pakalapati
Anurag Pampati
Divya Sai Patnana

</div>