CS432/532: Final Project Report

Project Title: AirBNB Data Analysis

Team Member(s): Rohith Vardhan Siliveri, Rugvedh Vaidya

I. NOSQL QUERIES

This project addresses some of the complexities of the hosts or owners of AirBNBs, in analyzing the data to improve the quality of AirBNBs to the customers. The nontrivial NoSQL queries that we have implemented in this project are listed below:

Query 1: Variation in Price Elasticity Based on Property Type, Neighborhood.

Attributes used : property_type, neighborhood_cleansed and price

This analysis helps hosts optimize pricing strategies based on demand fluctuations in specific areas. Here we analyzed how the elasticity of prices varies in different neighborhoods and across different property types.

Step-by-step implementation

- We categorized the property_type attribute into property sizes according to their size(VerySmall, Small, Medium, Large and Very large).
- Used the price attribute to find the maximum and minimum prices of the properties in each neighborhood according to their property_size.
- Computed the elasticity of price for the properties using the min price and max price and some computation to get the price_elasticity in each neighborhood by grouping them using the neighboorhood_cleansed attribute. Price Elasticity is the ratio of the percentage change in quantity demanded to the percentage change in price.
- After the above analysis, we have created a program which asks for the neighborhood as an input in the terminal and plots the graph of price_elasticity vs property_size of the neighborhood chosen. The output is shown in Fig 1.

Query 2: Guest Preferences Analysis

Attributes used: property_type, amenities and neighborhood_cleansed

This analysis helps the host in getting the information of guest's preferences in amenities and property type. Here, we analyzed the amenities, property_type and neighbourhood_cleansed attributes and found the most preferred amenities and the property type in a particular neighborhood.

Step-by-step implementation

- We categorized the property_type attribute into property sizes according to their size(VerySmall, Small, Medium, Large and Very large).
- Used the amenities attribute to find the most preferred amenities and given an amenity_score to each listing by matching the most preferred amenities.
- And then we grouped the data according to the neighborhood and came up with a table, which shows the neighborhood, preferred property_type in that neighborhood and amenity_score along with the array of amenities preferred by the customers. The output is shown in Fig 2.

Query 3: Influence of Property Size, and Host Experience on Occupancy Rate

Analyze used: property_type, host_response_time, host_response_rate, host_acceptance_rate, host_listing_count, host_identity_verified and availability_365

This analysis provides insights into factors that contribute to higher occupancy rates for Airbnb listings. Here we analyzed how the size of the property, and the host's experience level affect the occupancy rate.

Step-by-step implementation

- We categorized the property_type attribute into property sizes according to their size(VerySmall, Small, Medium, Large and Very large).
- Used the attributes like host_response_time, host_response_rate, host_acceptance_rate, host_listing_count and host_identity_verified to derive a new attribute called host_experience_score and scored each host(host_id) with a score.
- And we have taken the availability_365 attribute and computed the occupancy rate of each property.
- After the above analysis, we have plotted the graphs of host_experience_score vs occupancy_rate for each type of property_size. The output is shown in Fig 3

II. DATASET AND NoSQL DATABASE

A. Dataset:
- New Orleans Airbnb Listings and Reviews [1]

This dataset has non tabular as well as tabular data.

This data set describes the listing activity of Airbnb homestays in New Orleans, Louisiana, as part of the Inside Airbnb initiative. The data set was compiled on November 7, 2021. There are 49 attributes in the dataset like name, description, neighbourhood_cleansed, host_id, host_response_time, price etc.

B. NoSQL Database:
- We are using one of the popular No-SQL Databases i.e MongoDB[3]
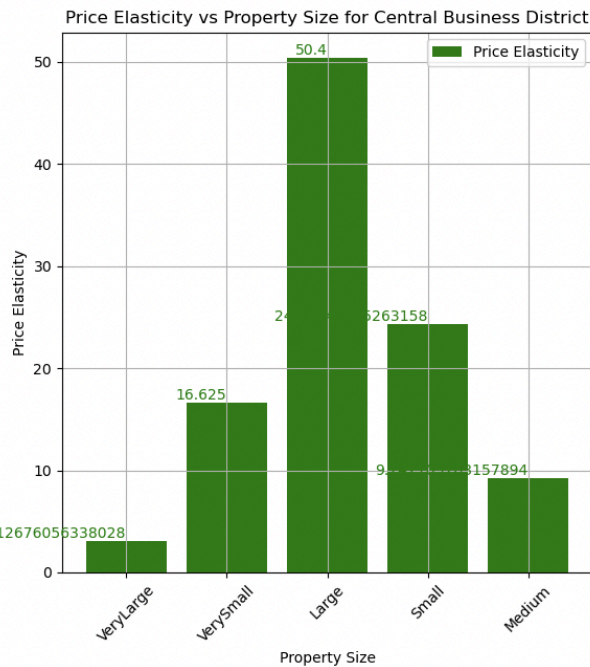
III. PROJECT OUTCOME

Fig. 1. Graph that represent the price_elasticity vs property_size in the Central Business District.

The above figure shows the price_elasticity vs property_size for the Central Business District. This analysis helps hosts optimize pricing strategies based on demand fluctuations in specific areas. This graph shows the price elasticity of different property sizes in the Central Business District.

output2_table

Neighbourhood	Property Size	Amenity Score	Customer Preference
Algiers Point	Small	31.235294117647100	62.2938490464679
Audubon	Large	29.408602150537600	53.13595522168210
B. W. Cooper	Large	26.0	8.21917808219178
Bayou St. John	Large	32.77215189873420	60.416160915554000
Behrman	Large	29.333333333333300	41.3013698630137
Black Pearl	Small	25.85	46.082191780821900
Broadmoor	Large	32.53488372093020	60.18158649251350
Bywater	Medium	30.034013605442200	52.8562109775417
Central Business District	Medium	29.374834437086100	39.076839335934000
Central City	Medium	30.587412587412600	47.62589009164350
City Park	Large	32.84615384615390	57.850368809272900
Desire Area	Large	28.166666666666700	63.37899543379000
Dillard	Small	32.27272727272730	53.499377334993800

Fig. 2. Table that represent the list of neighborhoods along with the preferred property size, amenity score and customer preference.

The above table represents the preferred property size, amenity score, customer preference in each neighborhood. If we take the Bywater neighborhood as an example, the customers prefer to stay in the medium sized property and also they preferred the amenities provided by that property.
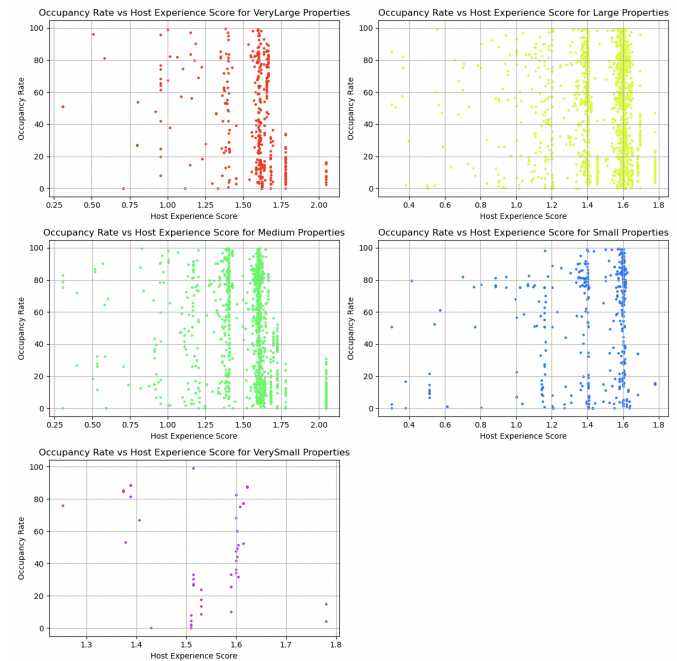

Fig. 3. Series of host experience score vs occupancy rate graphs for each property size.

The graph above shows how the host_experience with the customers affects the occupancy rate of the AirBNB. Each point on the graph represents the relationship between the host experience and the customer occupancy of that AirBNB in the next 365 days. The graph clearly shows that the AirBNBs with high host experience scores have a high occupancy rate.

REFERENCES

[1] Dataset Link: New Orleans Airbnb Listings and Reviews

[2] R. Elmasri and S. B. Navathe, Algorithms, Fundamentals of Database Systems, Seventh Edition, Pearson, 2017.

[3] MongoDB documentation

https://www.mongodb.com/docs/manual/

[4] Michael Ernst, How to Write a Technical Paper,

https://homes.cs.washington.edu/~mernst/advice/write-technical-paper.html