PROJECT 2

Milestone 1- Rojan Khatri

Topic:

“Unveiling Dallas’ Top Airbnb: A Data-Driven Guide.” This project aims to identify and rank the top 10 Airbnb properties in Dallas, TX, based on a unique combination of average ratings and the number of guest ratings.

Business Problem:

Even though Dallas was ranked 22nd among the top 25 cities with the lowest average occupancy rates for vacation rental properties in 2023, according to mashvisor.com, the city still boasts a significant number of Airbnb listings. Top This project seeks to uncover the top-performing properties in a relatively underutilized market, aiding property owners in understanding key success factors and potential guests in making informed accommodation choices. The research will explore which properties stand out in terms of guest satisfaction and popularity, providing insights into what makes them successful despite Dallas's overall low occupancy rate.

Datasets:

Data for this project will be sourced from three main avenues:

Flat File: A comprehensive list of Airbnb listings in Dallas, TX, obtained in CSV format from Inside Airbnb.

API: Recent reviews and ratings data for selected properties will be fetched using an API provided by Stevesie.com to ensure the analysis incorporates the most current guest feedback.

Methods:

The project will employ a mix of data cleaning, exploratory data analysis (EDA), and visualization techniques to analyze the data. Initial analysis will use ratings and reviews data from the flat file to identify the top 15 properties. This list will then be refined to the top 10 after updating with the latest reviews and ratings via the API. Visualization tools like Plotly and

Seaborn will be used to present findings, including geographic distribution and property features analysis. The project will also leverage SQL for data manipulation and integration across different datasets.

Ethical Considerations:

Handling publicly available data on property listings and reviews carries ethical considerations regarding privacy and data use. It is crucial to ensure that no personal information of hosts or guests is disclosed or misused. Additionally, the analysis should avoid inadvertently disadvantaging properties or areas not appearing in the top listings, recognizing the limitations and scope of the data used.

Challenges/Issues:

Anticipated challenges include the integration of datasets from diverse sources with varying formats and levels of completeness. Data collection may be hindered by access restrictions or costs associated with API usage. Ensuring the recency and relevance of data, especially from external APIs, could also pose a challenge. Methodological issues must be carefully managed, such as bias in guest ratings and the potential for newer properties to have fewer reviews.

References:

Mashvisor. (2024, January 8). What Airbnb Occupancy Rate Can You Expect in the US Market? https://www.mashvisor.com/blog/what-airbnb-occupancy-rate-can-you-expect/"

Stevesie.com for API access to the latest Airbnb reviews and ratings (https://stevesie.com/cloud/apis/airbnb/reviews).