Airbnb Data Analysis

Wireframe Document

Divyansh Verma
Data Analyst Intern
iNeuron AI

Table of Contents

1. Project Overview....3
2. Objectives and Research Questions....4
3. Data Analysis Workflow....5
3.1. Data Import and Preparation....5
3.2. Data Transformation....5
4. Exploratory Data Analysis (EDA)....6
4.1. Host Analysis....6
4.2. Neighbourhood Analysis....6
4.3. Reviews Analysis....7
4.4. Price Analysis....7
5. Visualization with PowerBI....9

1. Project Overview

Problem Statement:

Since 2008, guests and hosts have used AirBNB to expand traveling possibilities and present more unique, personalized ways of experiencing the world. This dataset describes the listing activity and metrics in San Diego, California for 2019.

Content:

This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions, and draw conclusions.

2. Objectives and Research Questions

1. **Regarding the Host:**
    - Identify the top earners.
    - Explore any relationship between monthly earnings and prices.
2. **Regarding the Neighbourhood:**
    - Determine any particular location getting the maximum number of bookings.
    - Analyze the relationship between price and location.
3. **Regarding the Reviews:**
    - Examine the relationship between quality (overall satisfaction) and price.
4. **Regarding Price:**
    - Investigate the relationship between price and amenities.
    - Analyze the relationship between price and location.

3. Data Analysis Workflow

3.1. Data Import and Preparation

Import Required Libraries:

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import statsmodels.api as sm
```

Load Data:

```
airbnb_data =
pd.read_csv("https://raw.githubusercontent.com/divyansh-193/airbnb_ana
lysis/main/airbnb_prices.csv")

airbnb_data.head()
```

Inspect Data:

- Display unique values per column.
- Display DataFrame information and summary statistics.

3.2. Data Transformation

- **Add Derived Columns:**
    - `total_customers = reviews * accommodates`
    - `total_earnings = price * reviews`

4. Exploratory Data Analysis (EDA)

4.1. Host Analysis

Top Earners:

```
top_earners =
airbnb_data.groupby('host_id')['total_earnings'].sum().reset_index()

top_10_earners = top_earners.sort_values(by='total_earnings',
ascending=False).head(10)

top_10_earners
```

Earnings and Prices Relationship:

```
average_price_per_listing =
airbnb_data.groupby('host_id')['price'].mean().reset_index()

host_earnings_price = top_earners.merge(average_price_per_listing,
on='host_id')

correlation =
host_earnings_price['total_earnings'].corr(host_earnings_price['price'
])

sns.regplot(x="price", y="total_earnings", data=host_earnings_price)
```

4.2. Neighbourhood Analysis

Bookings by Neighbourhood:

```
neighborhood_reviews =
airbnb_data.groupby('neighborhood')['reviews'].count().reset_index()

neighborhood_reviews.rename(columns={"reviews": "no_of_bookings"},
inplace=True)
```

Top Earning Neighborhoods:

```
neighborhood_bookings =
airbnb_data.groupby('neighborhood')['reviews'].sum().reset_index()

top_neighborhoods = neighborhood_bookings.sort_values(by='reviews',
ascending=False).head(10)
```

Average Price per Neighborhood:

```
neighborhood_price =
airbnb_data.groupby('neighborhood')['price'].mean().reset_index()
```

4.3. Reviews Analysis

Quality and Price Relationship:

```
correlation_quality_price =
airbnb_data['overall_satisfaction'].corr(airbnb_data['price'])

sns.boxplot(x="overall_satisfaction", y="price", data=airbnb_data)
```

4.4. Price Analysis

Bedrooms Listed on AirBnB:

```
airbnb_data['bedrooms'].value_counts()

X = airbnb_data[['bedrooms', 'bathrooms']].fillna(0)

y = airbnb_data['price']

X = sm.add_constant(X)

model = sm.OLS(y, X).fit()

sns.boxplot(x="bedrooms", y="price", data=airbnb_data)
```

Heatmap of Prices by Latitude and Longitude:

```
latitudes = np.linspace(airbnb_data['latitude'].min(),
airbnb_data['latitude'].max(), 40)

longitudes = np.linspace(airbnb_data['longitude'].min(),
airbnb_data['longitude'].max(), 40)

lat_grid, lon_grid = np.meshgrid(latitudes, longitudes)

heatmap, xedges, yedges = np.histogram2d(airbnb_data['latitude'],
airbnb_data['longitude'], bins=[latitudes, longitudes],
weights=airbnb_data['price'])

sns.heatmap(heatmap.T, xticklabels=xedges.round(2),
yticklabels=yedges.round(2), cmap='viridis')
```

Heatmap of Bookings by Latitude and Longitude:

```
latitudes = np.linspace(airbnb_data['latitude'].min(),
airbnb_data['latitude'].max(), 30)

longitudes = np.linspace(airbnb_data['longitude'].min(),
airbnb_data['longitude'].max(), 30)

lat_grid, lon_grid = np.meshgrid(latitudes, longitudes)

heatmap, xedges, yedges = np.histogram2d(airbnb_data['latitude'],
airbnb_data['longitude'], bins=[latitudes, longitudes],
weights=airbnb_data['total_customers'])

sns.heatmap(heatmap.T, xticklabels=xedges.round(2),
yticklabels=yedges.round(2), cmap='plasma')
```

5. Visualization with PowerBI

- **Interactive Maps:**
    - Displaying prices by geographical location.
    - Displaying the number of bookings by geographical location.
- **Graphs and Charts:**
    - Bar charts showing top earners and average prices by neighborhood.
    - Scatter plots and heatmaps to analyze relationships and distributions.