EXPLORING AIRBNB MARKET TRENDS

Lawal's Project

2024-10-23

Table of contents

Figure 1: NYC Skyline

1 Introduction

Welcome to New York City, one of the most-visited cities in the world. There are many Airbnb listings in New York City to meet the high demand for temporary lodging for travelers, which can be anywhere between a few nights to many months. In this project, we will take a closer look at the New York Airbnb market by combining data from multiple file types like `.csv`, `.tsv`, and `.xlsx`.

2 Project Overview

As a consultant working for a real estate start-up, you have collected Airbnb listing data from various sources to investigate the short-term rental market in New York.

3 Data Source

There are three files in the data folder: `airbnb_price.csv`, `airbnb_room_type.xlsx`, `airbnb_last_review.tsv`.

Recall that **CSV**, **TSV**, and **Excel** files are three common formats for storing data. Three files containing data on 2019 Airbnb listings are available to you:

data/airbnb_price.csv

This is a CSV file containing data on Airbnb listing prices and locations.

- `listing_id`: unique identifier of listing
- `price`: nightly listing price in USD
- `nbhood_full`: name of borough and neighborhood where listing is located

data/airbnb_room_type.xlsx

This is an Excel file containing data on Airbnb listing descriptions and room types.

- `listing_id`: unique identifier of listing
- `description`: listing description
- `room_type`: Airbnb has three types of rooms: shared rooms, private rooms, and entire homes/apartments

data/airbnb_last_review.tsv

This is a TSV file containing data on Airbnb host names and review dates.

- `listing_id`: unique identifier of listing
- `host_name`: name of listing host
- `last_review`: date when the listing was last reviewed

4 Exploratory Data Analysis

You'll analyze this data to provide insights on private rooms to the real estate company.

- What are the dates of the earliest and most recent reviews? Store these values as two separate variables with your preferred names.
- How many of the listings are private rooms? Save this into any variable.
- What is the average listing price? Round to the nearest two decimal places and save into a variable.
- Combine the new variables into one DataFrame called `review_dates` with four columns in the following order: `first_reviewed`, `last_reviewed`, `nb_private_rooms`, and `avg_price`. The DataFrame should only contain one row of values.

5 Data Analysis

```python
# Import necessary packages
import pandas as pd
import numpy as np

# Import the datasets
airbnb_price = pd.read_csv("data/airbnb_price.csv")
airbnb_room = pd.read_excel("data/airbnb_room_type.xlsx")
airbnb_last_review = pd.read_csv("data/airbnb_last_review.tsv", sep='\t')

# Merging the three DataFrames
price_room_review = pd.merge(pd.merge(airbnb_price, airbnb_room, on='listing_id', how='inner'), a

# Converting reviews data to a date format
price_room_review['last_review'] = pd.to_datetime(price_room_review['last_review'], errors='coerc

# Dates of the earliest and most recent reviews
fir_reviewed = price_room_review['last_review'].min()
las_reviewed = price_room_review['last_review'].max()

print(f"Earliest review date: {fir_reviewed}")
print(f"Most recent review date: {las_reviewed}")

# Deal with Value inconsistency in room_type column
price_room_review['room_type'] = price_room_review['room_type'].str.lower()

# Number of listings that are private rooms
pvt_room = price_room_review[price_room_review['room_type'] == 'private room'].shape[0]

print(f"Number of private room listings: {pvt_room}")
```

```python
# Alternative Method: Using value_counts()

# Get the count of each room type
nb_private_room = price_room_review['room_type'].value_counts()

# Display the number of listings that are private rooms
print(f"Number of private room listings: {nb_private_room.get('private room', 0)}")

# The average listing price? Round to the nearest 2 decimal places.
# Firstly, convert to float from strings
price_room_review['price'] = price_room_review['price'].str.strip('dollars')
price_room_review['price'] = price_room_review['price'].astype('float')

# Average listing price
mid_price = round(price_room_review['price'].mean(), 2)

print(f"The average listing pice: {mid_price}")

# Combine the new variables into one DataFrame called review_dates
review_date = {'first_reviewed': [fir_reviewed], 'last_reviewed': [las_reviewed], 'nb_private_roo

review_dates = pd.DataFrame(review_date)

print(review_dates)
```

```
Earliest review date: 2019-01-01 00:00:00
Most recent review date: 2019-07-09 00:00:00
Number of private room listings: 11356
Number of private room listings: 11356
The average listing pice: 141.78
  first_reviewed last_reviewed  nb_private_rooms  avg_price
0     2019-01-01    2019-07-09             11356     141.78
```

6 Result/Findings

The Analysis results are summarize as follows:

- The earliest review date was Timestamp('2019-01-01 00:00:00').
- The most recent review date was Timestamp('2019-07-09 00:00:00').
- The number of private rooms listings are 11356.
- The average listing price is 141.78