EXPLORING AIRBNB MARKET TRENDS

Lawal’s Project

2024-10-23

Table of contents



NYC Skyline

# 1. Project Overview

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. 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**

You can download the dataset [here](https://github.com/lawaloa/Project_6/blob/main/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**

You can download the dataset [here](https://github.com/lawaloa/Project_6/blob/main/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**

You can download it [here](https://github.com/lawaloa/Project_6/blob/main/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, in YYYY-MM-DD format.

# 3. Exploratory Data Analysis

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. 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.

# 4. Data Analysis

# 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'), airbnb\_last\_review, on='listing\_id', how='inner')  
  
# Converting reviews data to a date format  
price\_room\_review['last\_review'] = pd.to\_datetime(price\_room\_review['last\_review'], errors='coerce')  
  
# 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}")  
  
# 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\_rooms': [pvt\_room], 'avg\_price': [mid\_price]}  
  
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

# 5. Result/Findings

The Analysis results are summarized as follows:

* The earliest review date was 1st January 2019.
* The most recent review date was 9th July 2019.
* The number of private rooms listings are 11356.
* The average listing price is 141.78
* This is the one row DataFrame below: `
* first\_reviewed
* last\_reviewed
* nb\_private\_rooms
* avg\_price
* 0
* 2019-01-01
* 2019-07-09
* 11356
* 141.78
* `{=html}