Influence of Population and it's Density on Airbnb Listings and Pricing in US Cities

1 . Introduction

In this modern era of tourism, the conventional hotel culture is shifting towards renting rooms through Airbnb. It was a breakthrough in tourism, as the tourists were no longer tied to the conventional booking methods and rules of hotels. Now, every tourist has the opportunity to book a home or apartment which can be private or shared closest to where they want to visit. However, there are certain demographic features of every individual city which can influence the pricing and type of room listed on Airbnb.

This study focuses on how demographic features of the respective cities of the U.S influence the nature of listings. By doing so, it will give a perspective on how cities with varying demographic characteristics affect the listings.

"Do cities with larger populations have more Airbnb listings? And does higher population density affect higher average Airbnb prices & room type preferences (e.g., shared vs. private rooms)?"

2 . Used Data

2.1 Data Sources

Data source 1: US Airbnb Open Data

- **Metadata URL**: US Airbnb Data
- **Saved CSV Name**: *"AB_US_2023.csv"*
- **Description**: It contains a vast record of Airbnb listings in several cities of the United States.
- **Domain-Specific Value Types**:
    - *AB_US_2023.csv*: The dataset contains both numerical (Price), categorical features (Room type) and textual columns (name, neighbourhood, host name etc). There were more features, but they are not relevant for the scope of this report question.
- **Data Structure and License**: The dataset is structured as a tabular format (CSV file). The licence it uses is CC0 1.0 Universal and since it is public domain, I can work with this dataset without asking permission. Licence details - Licence Details

Data source 2: United States Cities Database

- **Metadata URL**: US Cities Database

- **Saved CSV Name:** *"uscities.csv"*
- **Description:** It contains demographic data of 30,844 cities of the United States of America.
- **Domain-Specific Value Types**:
    - *Uscities.csv*: The dataset contains both numerical (Population, density), categorical features (incorporated) and textual columns (name, city, state name etc). There were more features, but they are not relevant for the scope of this report question.
- **Data Structure and License**: The dataset is structured as a tabular format (CSV file). The licence it uses is Creative Commons Attribution 4.0 and since it is public domain, I can work with this dataset without asking permission. Licence details - Licence Details

2.2 Data Preparation - ETL

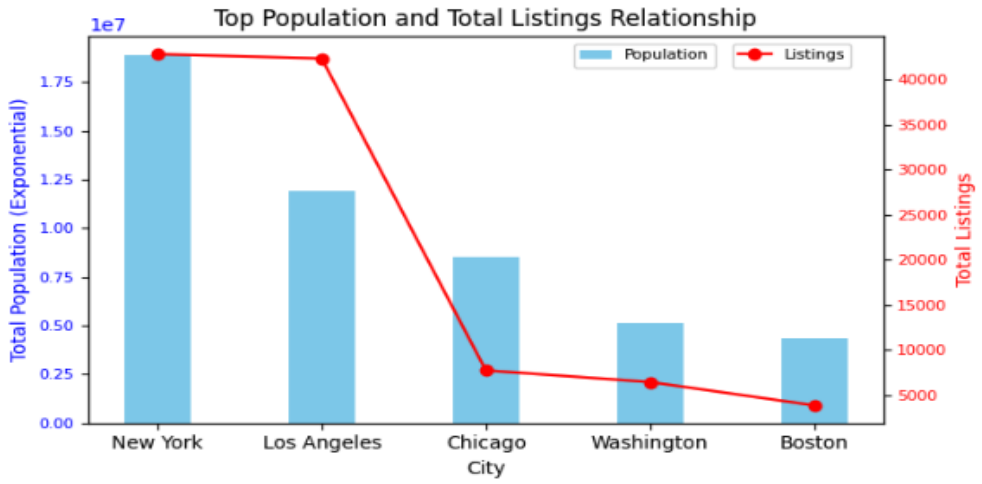
The datasets are first populated through the shell script, keeping a strict check that if the dataset already exists, it does not download the data, so no continuous loading is needed, as the datasets are static. The dataset preparation involves cleaning and organizing data from two sources: Airbnb listings and city demographics. For the Airbnb data, city names are standardized using a replacement dictionary such that city names are similar to the official names in the us_cities database to ensure consistency. Irrelevant columns are removed, and room types are converted into separate columns using one-hot encoding for better analysis. Outliers in airbnb listing prices are identified and removed using statistical methods, using data from 25th and 75th percentile to avoid extreme outliers. Key metrics such as average price and the number of listings by room type (private or shared) are calculated for each city. This ensures that the Airbnb data is accurate, clean, and ready for analysis. The airbnb data finally exists as being grouped by on city level, with for each city, "average price", "sum of private rooms", "sum of shared rooms", "total listings in the city", "total home apartments in the city".

The city demographic data is also cleaned by keeping only relevant details, such as population, density, and city names, while ensuring there are no duplicates. Once both datasets are processed, they are merged based on city names to combine Airbnb metrics with city-level demographic data. The output of the pipeline creates the data named ***"Final_data.csv"*** provides a clear view, making it easier to analyze how population and density influence Airbnb listings and pricing trends.

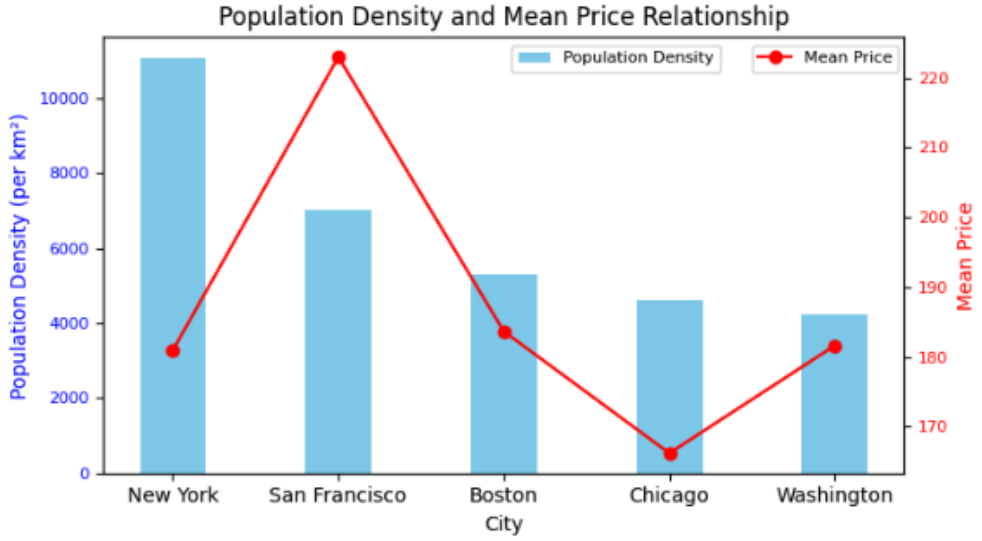
3 . Analysis

3.1 Effect of Population on total listings in a city

- The top five cities by Population came out to be New York, LA, Chicago, Washington & Boston.
- There is a clear correlation of population and total listings in that city. (Table ahead)
- The higher the population of a city, the more airbnb listings that city has, hence answering the first part of my research question.
- **Finding** - A highly populated city in the US has a positive correlation with total Airbnb listings.



3.2 Effect of Population density on average price of airbnb rooms

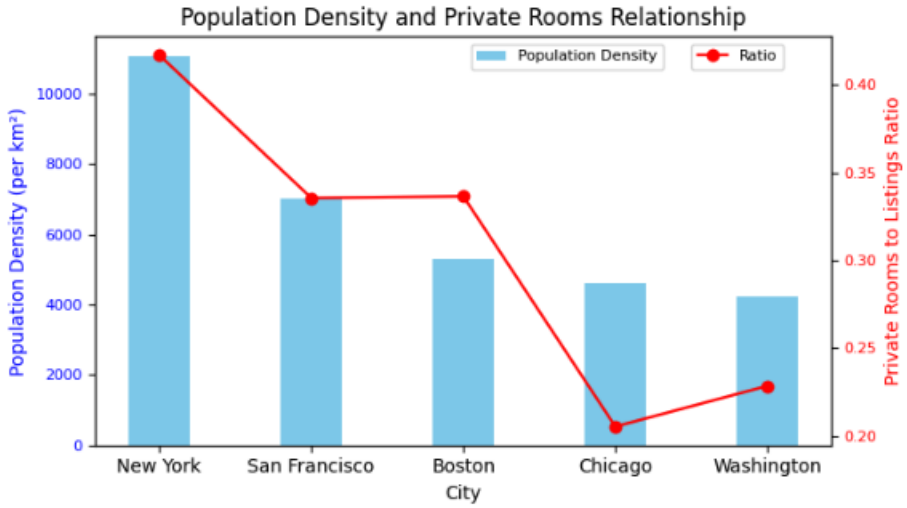


- The top five cities by Population Density per square kilometer came out to be New York, San Francisco, Boston, Chicago & Washington.
- There isn't a clear correlation of population & average price per night of rooms in that city.
- However, for the middle 3 cities, we can see that lower the population density, lower are average prices of the rooms.
- **Finding** - It isn't necessary that a city with high population density has expensive rooms, it might be that there are just a lot of options in a densely populated city (in outskirts), bringing the price down sometimes. Other factors seem to be playing a role, so, no concrete correlation found between population density and average price of AIrbnb rooms in that city.

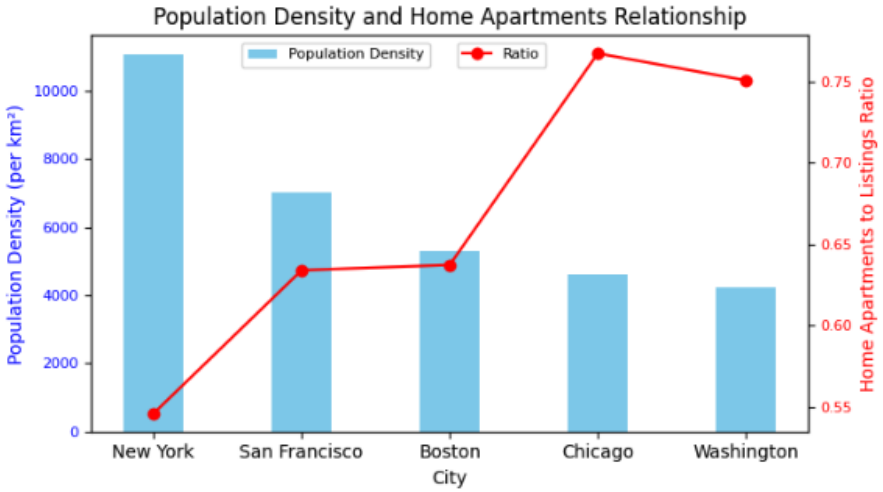
3.3 Effect of Population density on room preferences

- A ratio was first calculated - (total private rooms/total listings), this was done to cater to a variable number of total listings of all cities - keeping things equal just by comparing the ratios.

- It can be seen that (Table ahead), for those top five high population density cities in the previous part, there is a general negative trend.
- For shared rooms also,
- **Finding** - The higher the population density of a city, the more preference customers have for a private room.



3.3 Effect of Population density on Home apartments



- It can be seen that higher the population density of a city, lower is the availability of Home apartments, which makes sense, a city like New york won't have a lot of entire apartments to rent.

3.3 Conclusion

It can be concluded that higher the population of a city, higher are the total listings of Airbnb rooms, giving customers more options. Apart from that, higher the population density of a city, it isn't

necessary that the average price of listings would be expensive, which can be due to a lot of options in the outskirts of the city, bringing price down. Customers prefer a private room more in a city with high population density. Finally, although not in the research question, it was seen that it is difficult to find a Home apartment in a city where there is high population density.