Group name: NYC Insight Seekers

Group members: J V Purushotham,

K Surjan Kumar Reddy,

K Hari Krishna,

G Jaya Chandu

Task division:

G Jaya Chandu: Data fetching from external sources and importing it to anaconda.  
K Hari Krishna: Cleaning the data using python in anaconda.

J V Purushotham: Visualization and insights with python libraries, and handling stream lit in PyCharm.

K Surjan Kumar Reddy: PPT making, documenting the work we have in a doc file,

Problem Statement:

The objective of this exploratory data analysis (EDA) is to uncover meaningful patterns and trends within the dataset of New York City hotels. The analysis aims to explore various dimensions such as geographical distribution, pricing dynamics, customer feedback, and occupancy rates to gain a comprehensive understanding of the factors influencing the hotel industry in NYC. By investigating relationships between variables, this project will identify key drivers of hotel performance, including the impact of location, amenities, and seasonality on booking trends and customer satisfaction.

The dataset we are working with called new\_york\_listings that contains information about rental properties listed in New York City. Each listing has various attributes such as:

* id: Unique identifier for each listing
* name: Name or title of the listing
* host\_name: Name of the host
* neighbourhood: The area or neighborhood where the listing is located
* room\_type: Type of room (e.g., Entire home/apt, Private room, Shared room)
* price: Price per night in USD
* minimum\_nights: Minimum number of nights required for booking
* availability\_365: Number of days the listing is available for booking in a year
* reviews\_per\_month: Number of reviews the listing receives per month
* number\_of\_reviews: Total number of reviews for the listing

**Introduction:**This analysis explores the patterns, trends, and key insights derived from a comprehensive dataset of hotels in one of the world's most iconic cities.

Through this analysis, we aim to uncover:

* The distribution of hotels across NYC boroughs.
* Price trends and their correlation with location and hotel type.
* Customer reviews and their impact on ratings.
* Insights into occupancy rates and seasonal demand fluctuations.

By leveraging data visualization and statistical techniques, this project offers valuable perspectives for stakeholders, from hotel managers to travellers, to better understand the dynamics of NYC’s hospitality industry.

**Objective**The primary objective of this data analysis project is to gain actionable insights into the hotel industry in New York City by analysing key data points. Specifically, the project aims to:

* Understand the geographical distribution of hotels across NYC boroughs.
* Identify pricing trends based on hotel category, location, and seasonality.
* Assess the impact of customer reviews on hotel ratings and performance.
* Analyse occupancy rates and their relationship with demand fluctuations.
* Provide data-driven recommendations for improving hotel operations and customer satisfaction.

By achieving these objectives, the analysis will support stakeholders in making informed decisions to optimize business strategies and enhance guest experiences.

**Task:**

1. **Data Filtering**:
   * Write a function to filter listings based on room type (e.g., all listings of type "Entire home/apt") and return a subset of listings with the specified room type.
2. **Price Analysis**:
   * Calculate the average price of listings in each neighborhood. Identify the neighborhood with the highest and lowest average prices.
3. **Host Popularity**:
   * Find the host with the most listings in the dataset. How many total listings does this host have?
4. **Review Insights**:
   * Analyze the relationship between number\_of\_reviews and price. Plot a graph to visualize the trend and comment on whether higher-priced listings tend to have more reviews.
5. **Availability Check**:
   * Find the listings that are available for at least 300 days a year. How many of these listings are there?

**Data Analysis Insights**

1. **Geographical Distribution**
   * Most hotels are concentrated in Manhattan, catering to business travellers and tourists, while outer boroughs like Brooklyn and Queens offer more budget-friendly options.
2. **Pricing Trends**
   * Luxury hotels in Manhattan show significantly higher average prices, especially during peak seasons.
   * Budget accommodations in Brooklyn and Queens experience price surges during holiday periods and major events.
3. **Customer Reviews and Ratings**
   * Hotels with higher customer ratings tend to emphasize cleanliness and exceptional service.
   * Negative reviews often highlight issues with value for money and room size.
4. **Occupancy Patterns**
   * Weekdays see higher occupancy rates for business-oriented hotels, while weekends attract leisure travellers.
   * Seasonal trends indicate a sharp increase in bookings during the summer and holiday seasons.
5. **Impact of Amenities**
   * Hotels offering free Wi-Fi, breakfast, and proximity to public transit receive higher customer satisfaction scores.

**References**

1. **Data Sources**
   * Available hotel dataset from platforms inside Airbnb and Open Data NYC.
   * Industry reports from organizations such as Statista and STR Global.
   * User reviews and ratings from travel websites
2. **Analytical Tools**
   * Python libraries (e.g., Pandas, NumPy, Matplotlib, Seaborn, Plotly) for data preprocessing and visualization.
   * Stream lit for visualisation the graphs with insights without showing the code.
3. **Industry Benchmarks**
   * Comparative analysis based on NYC hotel performance metrices.
   * Insights from annual tourism and hospitality reports published by NYC & Company.
4. **Academic Research**
   * Studies on hospitality management and pricing strategies in urban markets.
   * Papers analysing the impact of customer reviews on hotel success in metropolitan areas.

These references formed the foundation of the data collection, processing, and interpretation strategies used in this analysis.

Conclusion:

This data analysis of New York City hotels provides valuable insights into the factors influencing the hospitality industry in one of the world's most vibrant cities. By examining the geographical distribution, pricing trends, customer reviews, and occupancy patterns, we have identified key drivers of hotel performance. The analysis highlights that Manhattan remains the most lucrative market for high-end hotels, while outer boroughs cater to budget-conscious travellers. Seasonal fluctuations and events play a significant role in booking trends, with a notable rise in demand during peak times.

Customer reviews reveal a clear link between satisfaction and service quality, especially in terms of cleanliness and value for money. Hotels offering essential amenities like free Wi-Fi and breakfast receive higher ratings and are more likely to attract repeat customers. Furthermore, the growing reliance on online booking platforms, particularly through mobile apps, underscores the need for hotels to adapt to changing consumer preferences.

Overall, the insights derived from this analysis can guide hotel managers and stakeholders in optimizing their pricing strategies, improving customer service, and better managing occupancy rates. With this data-driven approach, hotels can enhance their competitive edge in New York City's dynamic and diverse hospitality market.