Summary Report: Analysis of Airbnb Listings in NYC

Overview:
This project analyzed Airbnb listings in New York City to uncover trends and insights that inform property management and marketing decisions. It focused on understanding how location, room type, and availability impact price through statistical analysis and data visualization.

Dataset Selection:
The dataset, sourced from Kaggle, provides comprehensive information on NYC Airbnb listings, including room types, prices, availability, and neighbourhood details.

Data Cleaning:
The dataset initially included 48,894 records and 16 variables. Columns with missing or irrelevant information (e.g., host_name, name, id, and last_review) were dropped, and missing values in reviews_per_month were filled with 0 to ensure data consistency.

Key Findings:
Entire homes are the most expensive, targeting luxury travellers, while shared rooms are the cheapest, appealing to budget-conscious guests. Private rooms offer a balance of cost and comfort. Manhattan listings command the highest prices, reflecting premium demand, while Queens and the Bronx are budget-friendly alternatives. Brooklyn provides a middle ground between affordability and luxury. Manhattan and Brooklyn have lower availability, indicating high demand. Queens and Staten Island exhibit higher availability, suggesting less competition. Short-term stays (1-2 nights) dominate, emphasizing the preference for brief itineraries.

SQL Database Integration:
The cleaned data was exported to an SQLite database, allowing efficient querying and storage for further analysis. The process involved creating an SQLite database (ab_nyc_data.sqlite), exporting the dataset as a table (ab_nyc_listings), and verifying the successful retrieval of data. This integration facilitates scalability and supports advanced analytics by providing a robust backend for data storage.

Team Contributions:

- **Data Import and Cleaning:** Dhruv Sharma, Krishi Shah, and Devansh Bhavsar
- **Statistical Analysis:** Devarshi Lala and Nisargvan Goswami
- **Visualizations:** Dhruv Sharma, Krishi Shah, and Devansh Bhavsar
- **SQLite:** Dhruv Sharma
- **Reporting and Documentation:** Devarshi Lala and Nisargvan Goswami

Challenges and Research:
Handling missing data, performing advanced statistical analyses, and creating effective visualizations were key challenges. Research leveraged online resources like Kaggle, Stack Overflow, and library documentation for statistical methods and visualization techniques.

Project Expansion:
Future work could include predictive modelling for rental prices, sentiment analysis of reviews, and exploring seasonal trends to optimize pricing strategies.

References: Kaggle API documentation, Pandas and NumPy official guides, Seaborn and Matplotlib tutorials, and Statsmodels