INT217

**PROJECT REPORT**

(Project Semester January-April 2025)

Airbnb Dataset Analysis and Dashboard Creation

Submitted by

Aryan Agarwal

Registration No: 12324865

B.Tech (CSE) K23KM

Course Code: INT217

Under the Guidance of

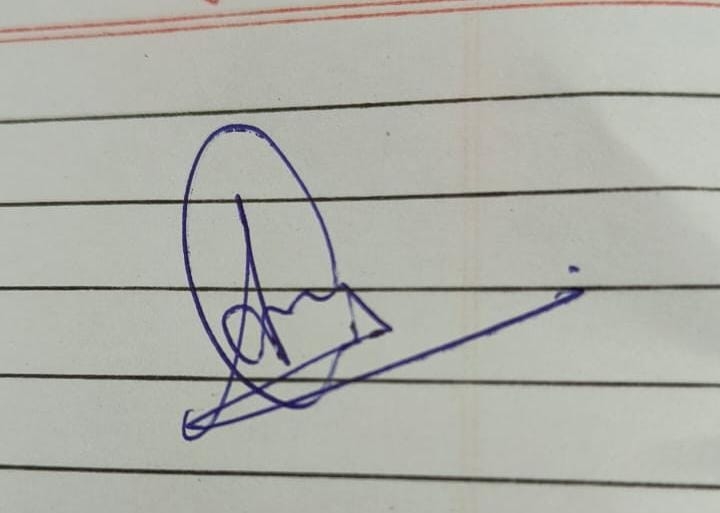
Ms. Nidhi Arora (28373)

Discipline of CSE/IT

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

# DECLARATION

I, Aryan Agarwal, student of B. Tech (CSE) K23KM under CSE/IT Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.  
  
Date: 12/04/2025  
  
Signature:   
  
Registration No. 12324865  
  
Name of the student: Aryan Agarwal

# CERTIFICATE

This is to certify that Aryan Agarwal bearing Registration no. 12324865 has completed INT217 project titled, “Airbnb Dataset Analysis and Dashboard Creation” under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.  
  
Signature and Name of the Supervisor: Ms. Nidhi Arora  
Designation: Faculty  
School of Computer Science and Engineering  
Lovely Professional University  
Phagwara, Punjab.  
Date: 12/04/2025

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my faculty guide, Ms. Nidhi Arora, for her valuable guidance and support throughout the project. Her insights helped me understand the practical application of data analysis and dashboard development in Excel. I also thank Lovely Professional University for providing the platform and resources to complete this project successfully. Lastly, I would like to thank my peers and family for their constant encouragement.

# TABLE OF CONTENTS

1. Introduction

2. Source of Dataset

3. Dataset Preprocessing

4. Analysis on Dataset

i. General Description

ii. Specific Requirements

iii. Analysis Results

iv. Visualization

5. Conclusion

6. Future Scope

7. References

## 1. Introduction

The digital transformation in the travel and tourism industry has opened up numerous opportunities for individuals to monetize their properties via platforms such as Airbnb. Airbnb allows homeowners to list their properties for short-term rental and provides travellers with a variety of accommodation options. With the availability of large volumes of data on Airbnb listings, data analysis can provide valuable insights into pricing strategies, demand patterns, and regional preferences.

This project aims to harness the power of Microsoft Excel to analyse a real-world Airbnb dataset. The primary objective is to clean and process the dataset, identify key trends, and visualize these insights using Excel dashboards. The final dashboard provides an intuitive interface to interact with and derive meaningful conclusions from the data.

## 2. Source of Dataset

The dataset used in this project was sourced from "Inside Airbnb," a platform that provides public access to Airbnb data for several cities. The data is a snapshot of the Airbnb listings at a given point in time and includes a wide range of attributes, such as listing ID, host ID, location, room type, price, number of reviews, and availability.

The specific dataset analysed in this project includes listings for a metropolitan city and contains over 40,000 records. This data was available in CSV format and was imported into Microsoft Excel for pre-processing and analysis.

## 3. Dataset Preprocessing

Data pre-processing is a critical step in the data analysis pipeline. The raw Airbnb dataset contained various missing values, inconsistencies, and irrelevant columns. The following steps were taken during pre-processing:

* **Removal of Null Values**: Columns with significant amounts of missing data were either removed or filled using imputation methods.
* **Column Filtering**: Only relevant columns such as 'name', 'host\_id', 'neighbourhood', 'room\_type', 'price', 'number\_of\_reviews', and 'availability\_365' were retained for analysis.
* **Data Cleaning**: Price fields were converted from text to numeric format. Duplicates and outliers were identified and removed to ensure data integrity.
* **Derived Columns**: New columns such as 'price per day' and 'review rate per month' were calculated to enable further analysis.

These steps helped convert the dataset into a structured and analysable format, suitable for building PivotTables and dashboards.

## 4. Analysis on Dataset

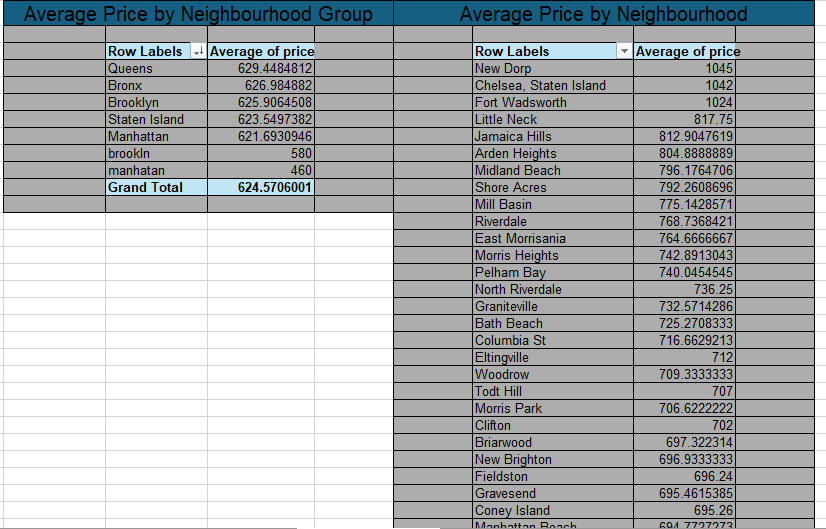
### i. General Description

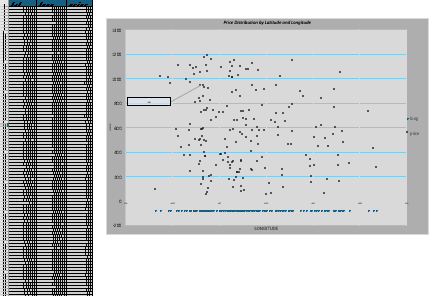
This section gives a high-level overview of the Airbnb dataset:

* The dataset includes over **40,000 listings**, each with attributes like price, location, room type, and availability.
* The **majority of listings** fall under the category **“Entire home/apt”**, indicating users prefer renting full spaces.
* Listings are geographically distributed, with high concentrations in neighbourhoods such as **Manhattan**, **Brooklyn**, and **Queens**.
* The data also highlights host activity, review frequency, and calendar availability.

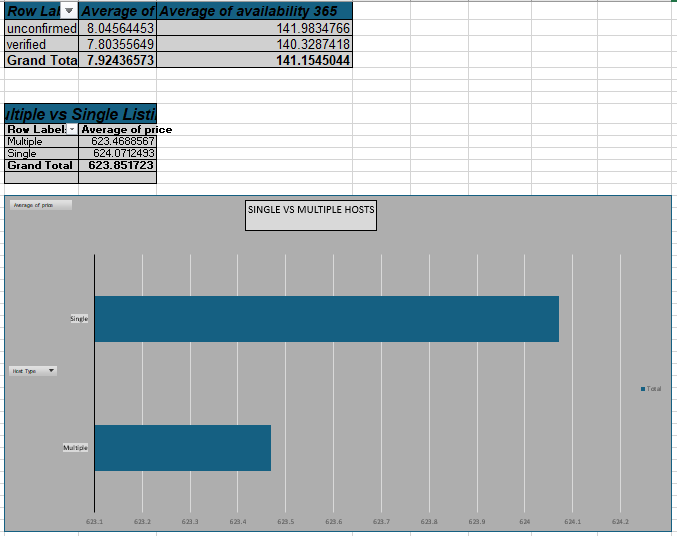
### ii. Specific Requirements

Each dashboard objective is detailed below:

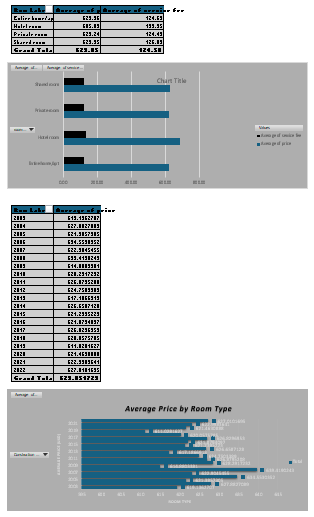
1. **Price Analysis**:
   * The dashboard reveals **average, minimum, and maximum prices** across room types and neighbourhoods.
   * High-end neighbourhoods like **Manhattan** show significantly elevated price points compared to others like **Queens**.
   * 



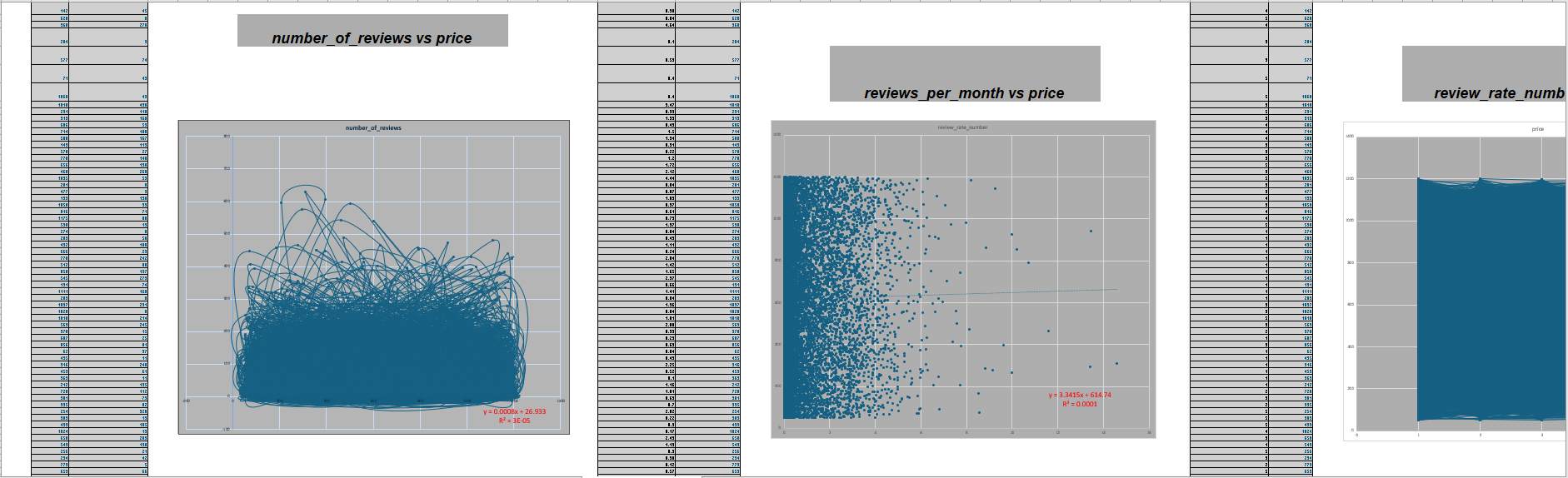
1. **Room Type Analysis**:
   * The breakdown of listings by room type shows that **Entire homes/apartments dominate**, followed by **Private rooms**.
   * Room type has a clear correlation with price, with shared accommodations being the most economical.

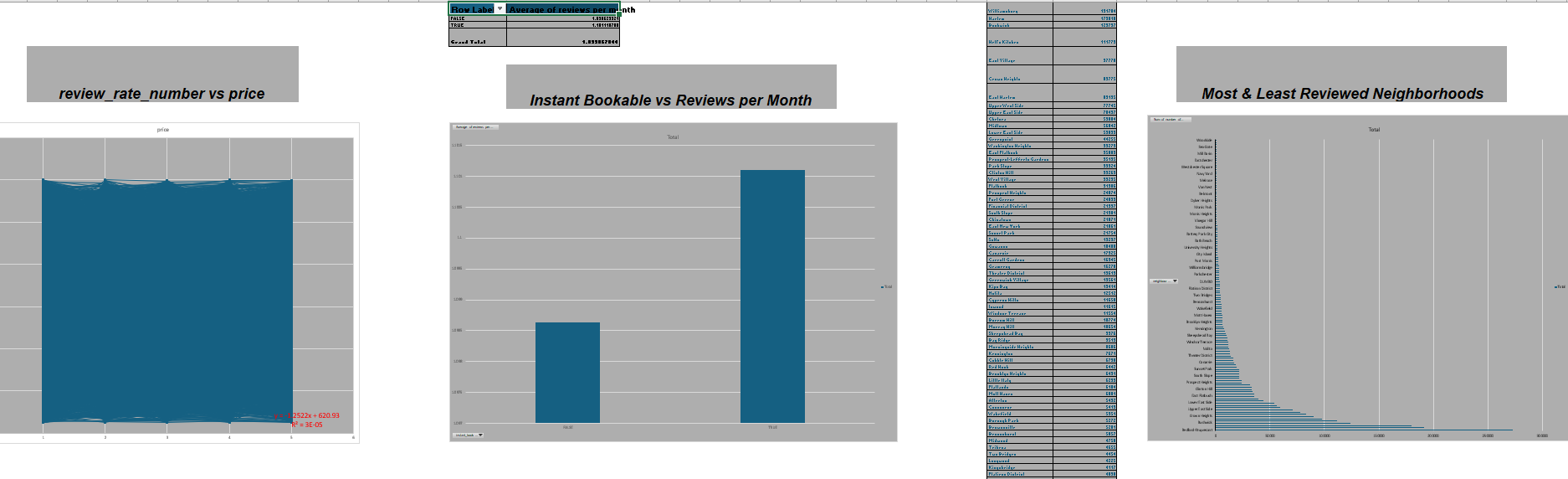


1. **Neighbourhood-Wise Insights**:
   * Analysis of listings by location uncovers **which areas are most saturated**.
   * It also helps identify **budget-friendly vs. premium neighbourhoods** based on average nightly price.

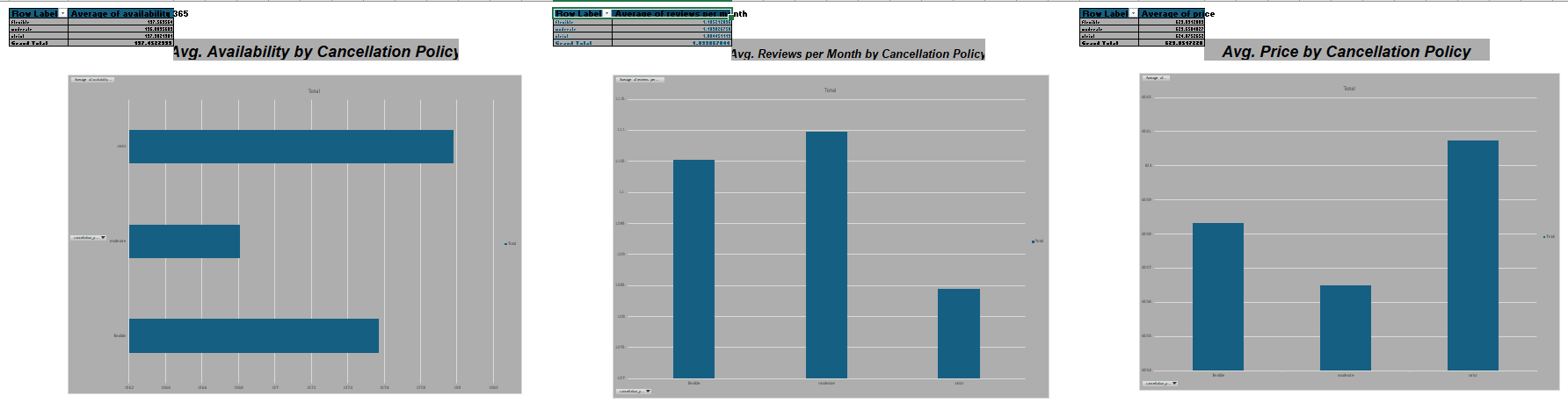


1. **Availability Analysis**:
   * By visualizing availability across 365 days, the dashboard indicates which listings are:
     + Fully available (365 days)
     + Seasonally or partially available
     + Not available (0 days)
   * This supports travel planning and host competitiveness.





1. **Host Listing Frequency**:
   * Reveals top hosts by the number of listings they manage.
   * Some hosts appear to manage 20+ listings, suggesting **professional hosting businesses**.



1. **Review & Popularity Trends**:
   * Listings with **higher review counts** often also have higher availability and competitive pricing.
   * Helps pinpoint popular and high-demand properties.

### iii. Analysis Results

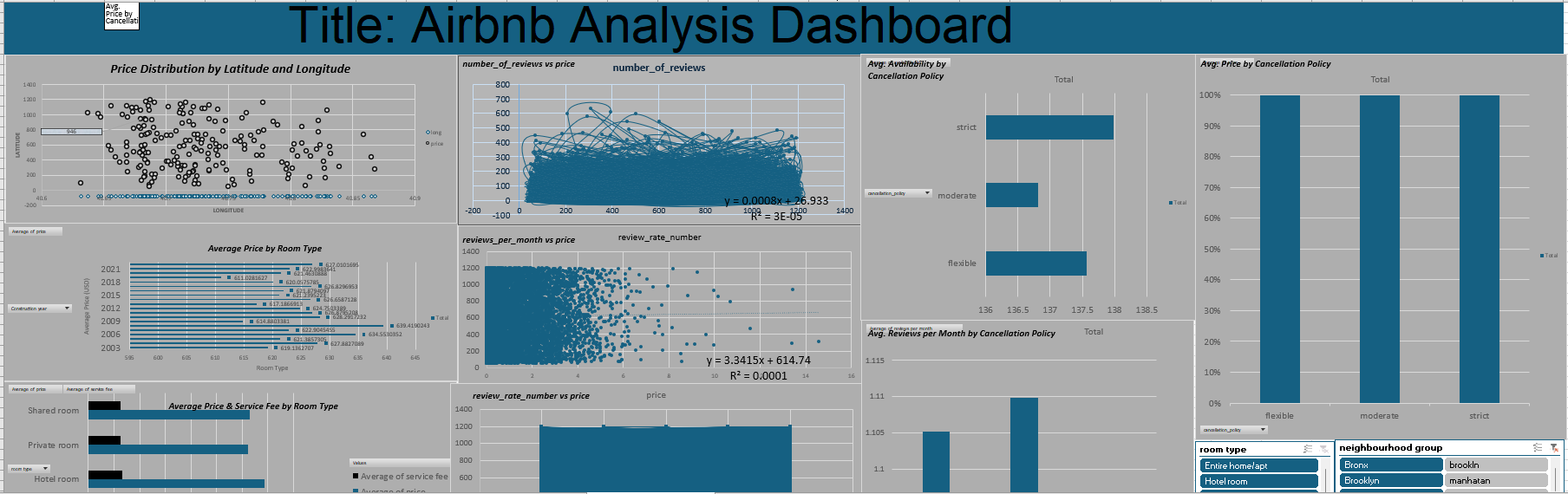
* **Entire home/apt** is the preferred listing type and also the most expensive on average.
* **Manhattan** is the priciest neighbourhood, while **Queens** and **Bronx** are more affordable.
* **High-review listings** are often well-priced and have high year-round availability.
* **Some hosts manage dozens of properties**, implying Airbnb is also used as a commercial platform.
* Listings with **365-day availability** are generally more optimized for frequent bookings.

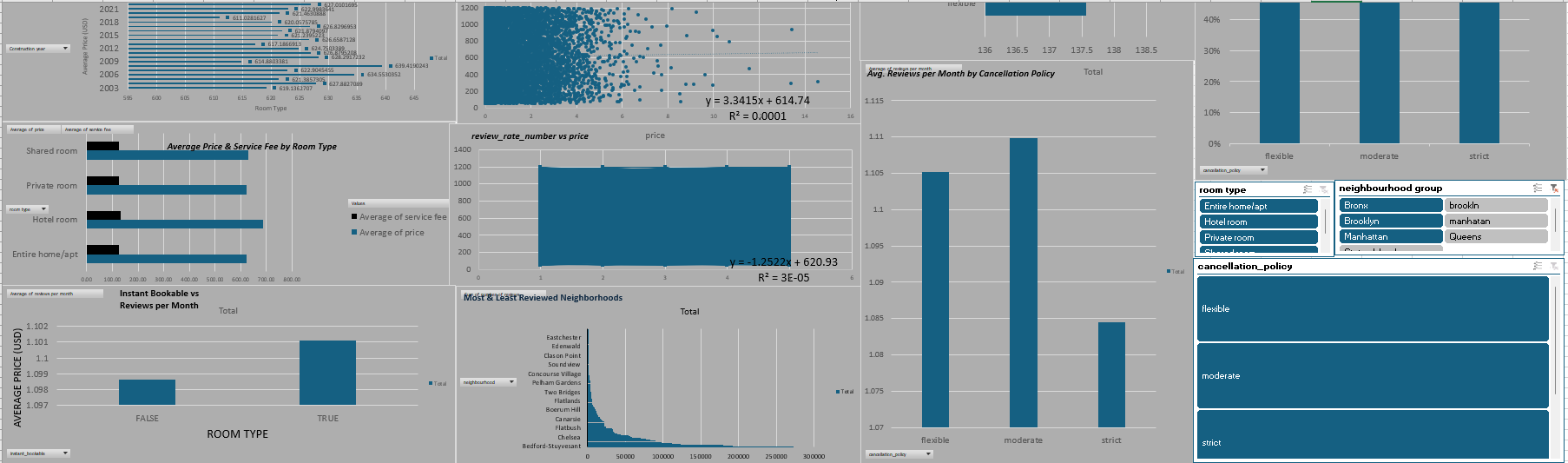
### iv. Visualization

The following dashboard elements visually represent insights:

* **Pivot Tables** summarize data such as:
  + Room type by average price
  + Availability by listing count
  + Reviews by neighbourhood
* **Bar & Pie Charts** show:
  + Room type distribution
  + Listings per neighbourhood
  + Price range comparison
* **Slicers** allow dynamic filtering by:
  + Neighbourhood
  + Room type
  + Availability

**Pivot Charts** were developed to visualize the relationship between price and availability, as well as to show distribution of room types.  
These visuals make the dashboard highly interactive and user-friendly.





## 5. Conclusion

The Airbnb dataset provided a robust basis for learning and applying various Excel-based data analysis techniques. The final dashboard enables a non-technical audience to explore listings based on different dimensions and draw conclusions from trends in pricing, availability, and location.

Through this project, I gained practical knowledge of data cleaning, transformation, visualization, and reporting in Excel. It also provided insight into how data can be used to support strategic decision-making in the hospitality industry.

## 6. Future Scope

While Excel was an effective tool for this level of analysis, the future scope of this project includes:

* **Advanced Data Integration**: Using Power Query to connect live datasets for real-time updates.
* **Enhanced Dashboards**: Creating interactive dashboards using Power BI or Tableau for more advanced visuals and geospatial insights.
* **Predictive Modelling**: Applying machine learning models in Python or R to forecast demand, pricing trends, and customer behavior.
* **Automation**: Automating the reporting pipeline to reduce manual updates.

These improvements would make the analysis more scalable, interactive, and valuable in a business context.

## 7. References

[1] Airbnb. "Inside Airbnb: Adding data to the debate." [Online]. Available: http://insideairbnb.com/get-the-data.html  
[2] Microsoft Support. "Create a PivotTable to analyze worksheet data." [Online]. Available: https://support.microsoft.com  
[3] Lynda.com, "Excel Data Visualization," Online course materials.  
[4] Airbnb Official Blog, "Data Science at Airbnb." [Online]. Available: https://news.airbnb.com  
[5] Statista, "Airbnb average nightly rates in major cities." [Online]. Available: https://www.statista.com