Project Title: Exploratory Data Analysis (EDA) on Airbnb Listings

Done by:

Fatima Bailoun

Submitted on October 2023

During 1-month internship

With “Coding Samurai”

ACKNOWLEDGMENT

First and foremost, I would like to praise and thank God, the Almighty, who has granted me countless blessings, knowledge, and opportunities.

Second, I send a thank you to Coding Samurai for giving me the chance to take this internship and collect the experience I can take during this month.

ABSTRACT

In this project, I performed basic data analysis on a dataset of Airbnb listings. EDA is a fundamental step in data science that involves exploring and understanding the data before diving into more complex analysis or modeling.

TABLE OF CONTENTS

Exploratory Data Analysis (EDA) on Airbnb Listings

Introduction

Airbnb, Inc. is an American San Francisco-based company operating an online marketplace for short- and long-term homestays and experiences. The company acts as a broker and charges a commission from each booking. The company was founded in 2008 by Brian Chesky, Nathan Blecharczyk, and Joe Gebbia.

Amsterdam has become a prime destination for travelers seeking unique and memorable experiences. In recent years, the sharing economy has played a significant role in shaping the city's tourism landscape, and one platform, in particular, has stood out as a major player in this regard – Airbnb

This report embarks on an exploratory data analysis (EDA) project focused on Airbnb listings in Amsterdam. The primary objective of this study is to gain deeper insights into the dynamics of Airbnb's presence in the city and understand the implications of its growth on various aspects of Amsterdam's socio-economic landscape. By examining and analyzing a rich dataset of Airbnb listings, we aim to uncover valuable patterns and trends that shed light on these accommodations' distribution, pricing, and characteristics.

Data Scraping Details

- **Data Title:** Amsterdam, North Holland, The Netherlands
- **Data Publishing Date:** 03 September, 2023
- **Data Source:** Inside Airbnb
- **Data Source Link:** http://insideairbnb.com/get-the-data/
- **Data File Type:** CSV file

Implementation

1.1 Introduction:

This chapter discusses the steps involved in the implementation and the tools and libraries that were utilized.

1.2 Implementation Tools:

The language used for implementation in this project is: Python3

The libraries in our code are seaborn, pandas, NumPy, matplotlib.pyplot, and folium.

1.3 Implementation steps:

1st, to **scrape** the needed data I chose a data set from the Inside Airbnb website then I downloaded it and read data from my desktop with the pandas.read_csv()

2nd, I started by **data exploration** to understand the dataset I'm working with more using pandas & numpy methods.

3rd, data cleaning was done in which I removed all unneeded columns (attributes), handle missing values, rename columns of clarity, and converted object-numeric data to numeric(integer).

4th, applied **basic statistics** so I re-describe my dataset using df.describe() method to see the updated statistics of my cleaned dataset showing mean, count..etc, then calculated the mean of price concerning List type, correlation analysis between Price and number of beds per room, list the top 10 rooms(id & price) by price in decreasing order, count price distribution, top 5 neighborhoods concerning number of listings. All these topics were done by using methods in matplotlib.pyplot and seaborn libraries.

5th, **visualization** was done so I plotted a bar graph to compare the room type distribution in the Amsterdam neighborhood, implementing a joint plot with scatterplot and marginal histograms to explore the relationship between reviews and how it might be influenced by a room type. Moreover, I counted the distribution of listings among neighborhoods using a bar graph, the distribution of prices per room type using a boxplot graph, and counted true and

false values in the instant bookable attribute in a pie chart. All these visualizations were done by the methods of Seaborn Library.

6th, in the **geospatial analysis** part I plotted the Geographical Distribution of Listings using latitude and longitude columns in a scatterplot graph, and the new-to-me visualization I tried was the heatmap using folium library that shows the frequency of listings distribution among Amsterdam. Then, I saved the heatmap in an html file to be able to present it.

7th, Finally the correlation analysis where I summarized the relation between all attributes using pandas method df.corr()

Data Visualizations and Interpretations

2.1 Introduction:

In this section, I will introduce the visualizations done and interpret it

2.2 The mean of price concerning Listings type:

room_type	price
Entire home/apt	297.268627
Hotel room	220.885714
Private room	164.140110
Shared room	146.333333

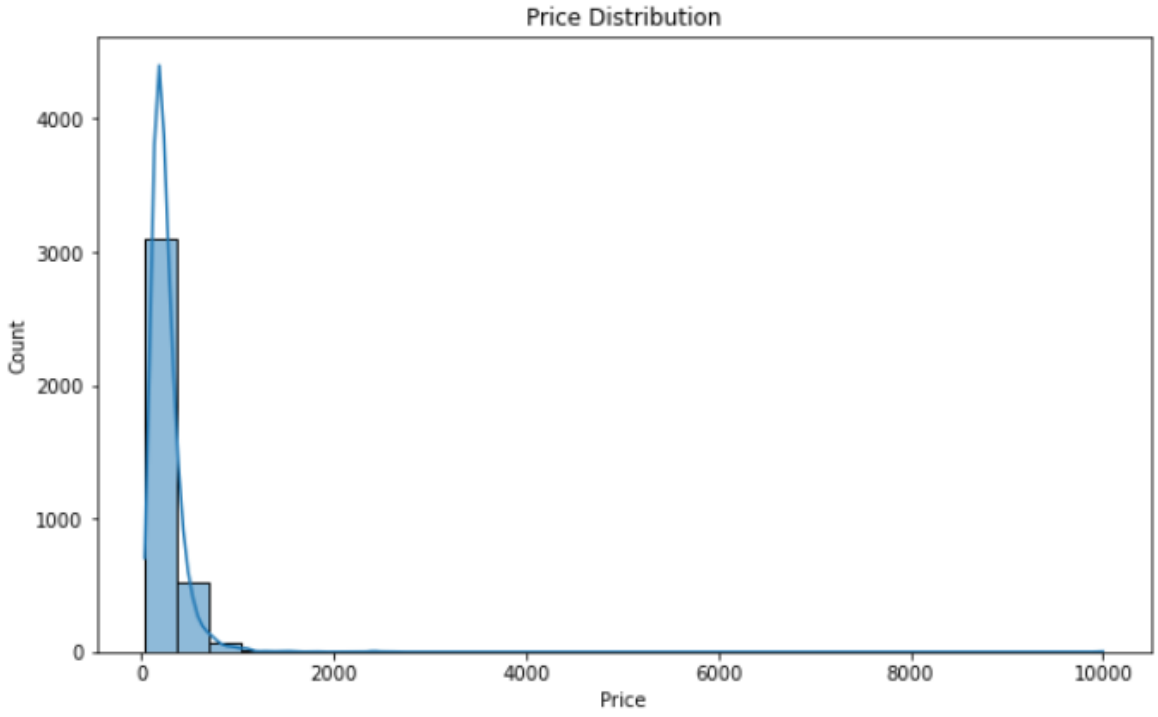
The apartment listings are the most expensive rooms with a range of 297$ whereas the shared rooms are less in the range of 146$.

2.3 Correlation Analysis between Price and number of beds per room

	price	accommodates
price	1.00000	0.33718
accommodates	0.33718	1.00000

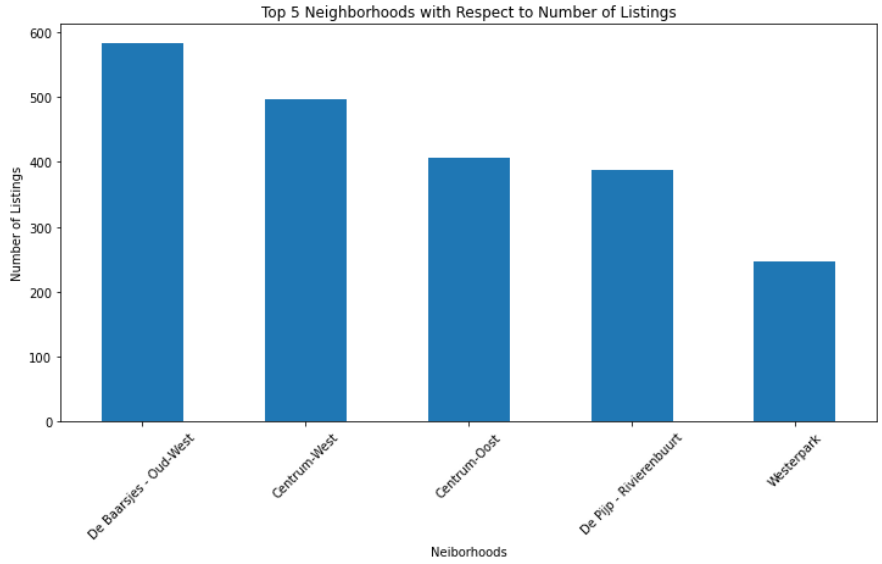
This value (0.33718) suggests a positive relationship between the number of accommodates per listing and the price of it. However, since the value is closer to 0 than 1, it indicates a relatively weak positive correlation. This means that there is some tendency for the two entities to move in the same direction, but it's not a strong or highly reliable relationship. Other factors may also influence their behavior.

2.4 Price Distribution



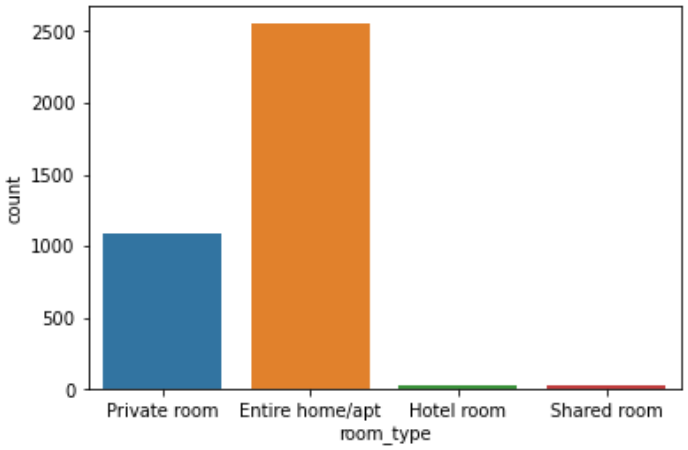
it appears that most listings have prices ranging from $0 to $200, with the highest concentration between $0 and $100. However, there are also some high-priced listings, which result in a long right tail in the distribution.

2.5 Top 5 neighborhoods



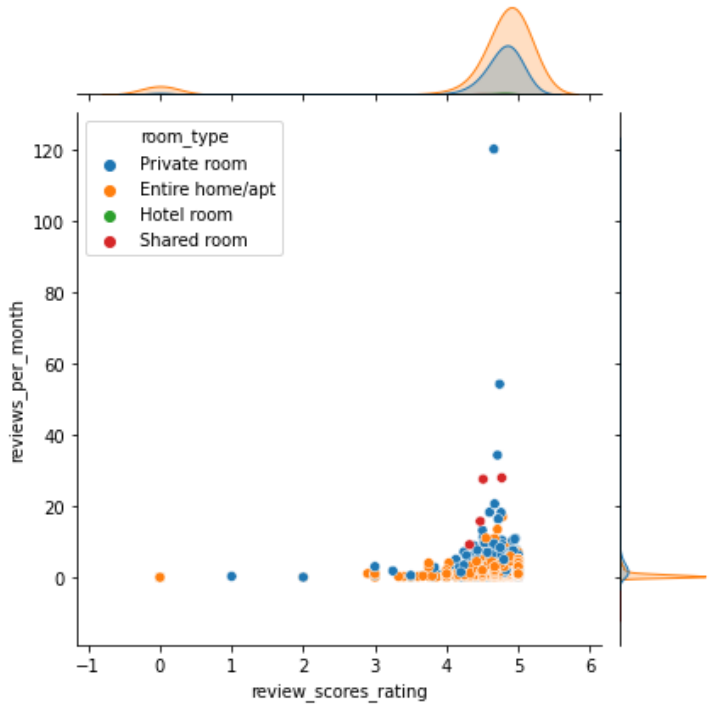
This bar chart allows viewers to quickly identify which neighborhoods have the highest numbers of listings, making it easy to understand the distribution of listings across different areas. Viewers can see that “De baarsjes-Oud-West”, “Centrum-West”, “Centrum-Oost”, “De Pijp-Rivirenbuurt”, and “Wester park” are the most popular or have the most offerings.

2.6 Room Type Distribution in Amsterdam



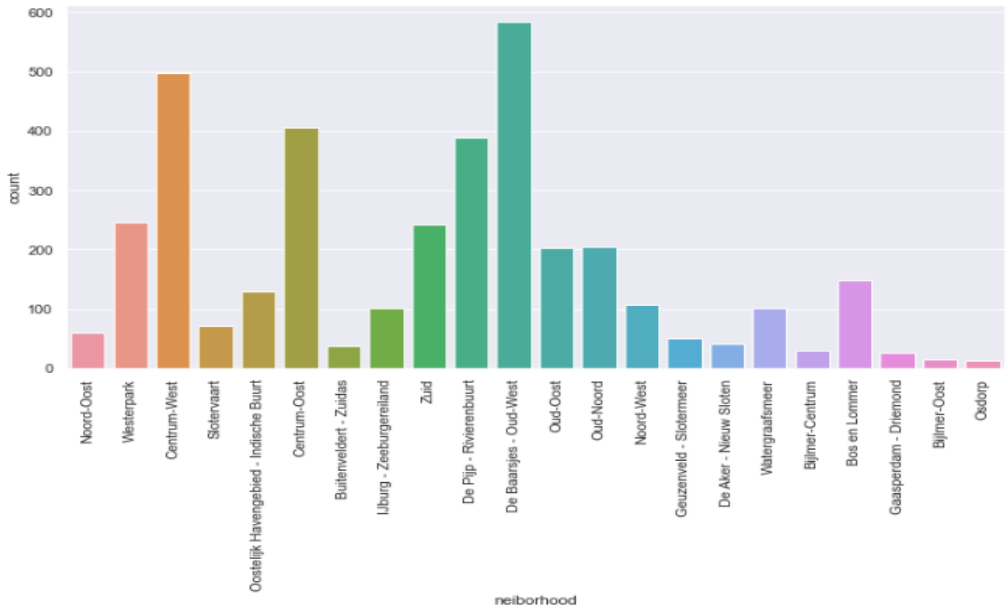
Apartments are the most frequent listings found for rent in Amsterdam, then private rooms. But hotel and shared rooms are the less.

2.7 Reviews with Respect to Listing Type

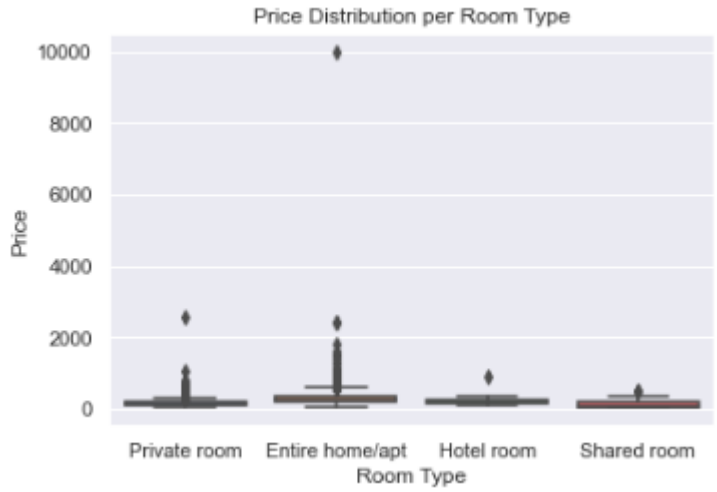


Overall, this visualization shows the relationship between review scores, reviews per month, and room types in the Airbnb listings dataset. Where most reviews are between 3-5/rating and 0-25/monthly. But the most rated rooms are the apartments and private rooms.

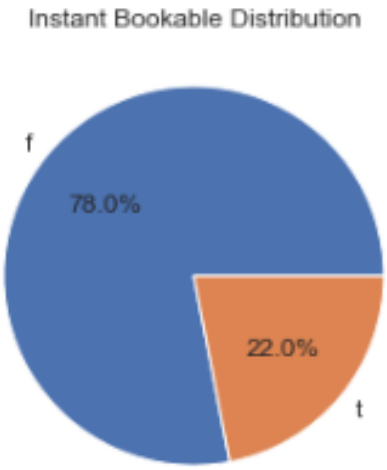
2.8 Distribution of listings among Amsterdam Neighborhoods



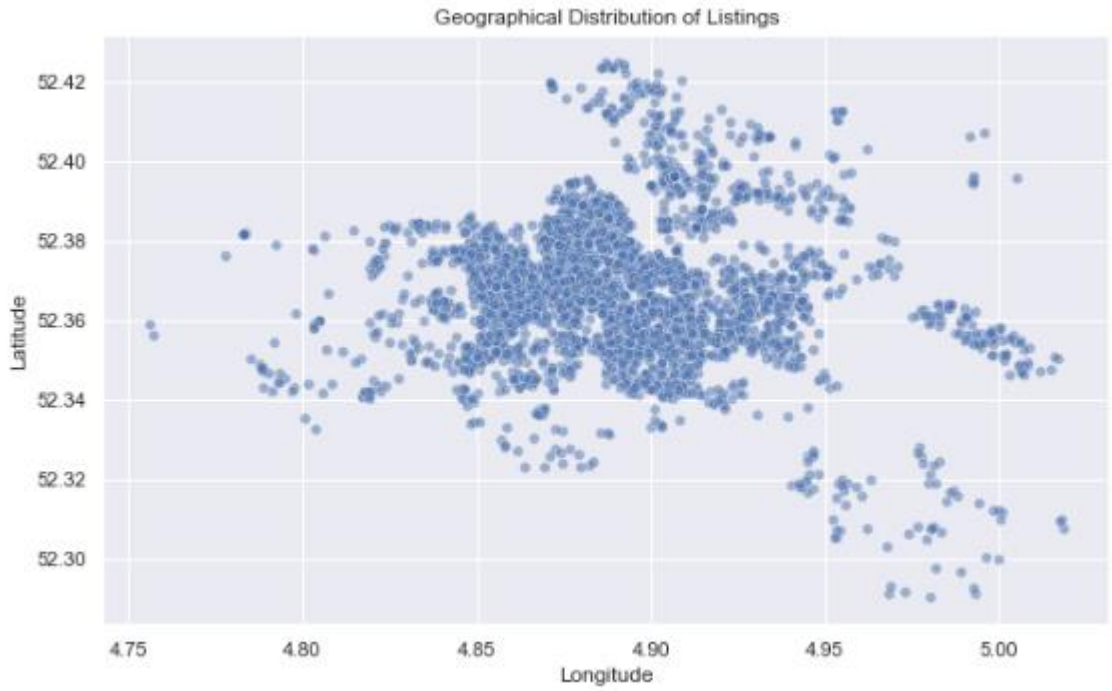
2.9 Distribution of Prices Per Room Type



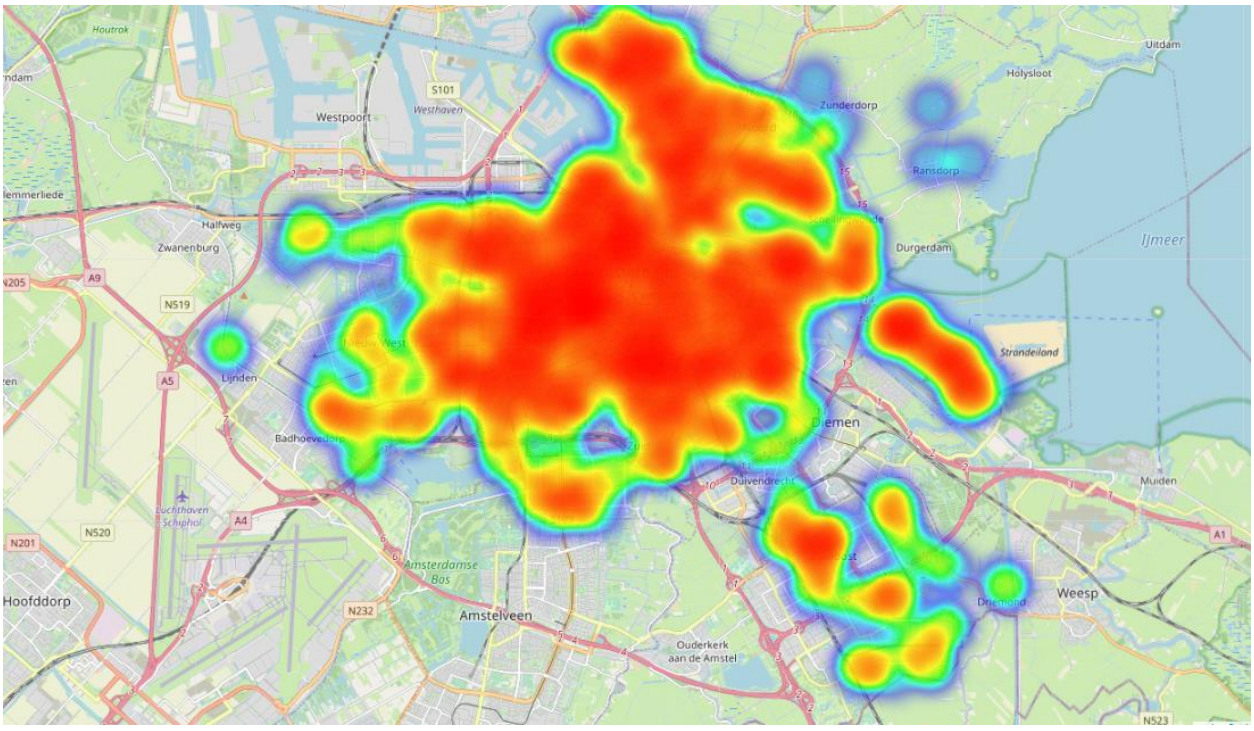
2.10 Count True and False Values in Instant Bookable Attribute



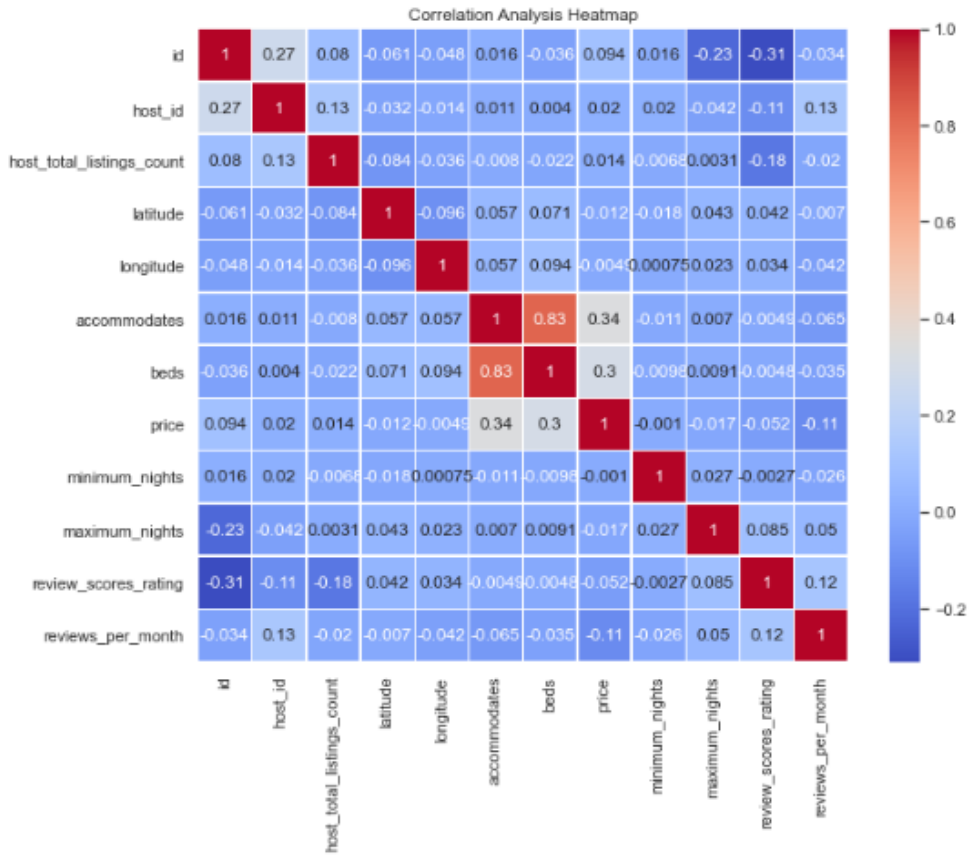
2.11 Geographical Distribution of Listings



2.12 Amsterdam Heatmap



2.13 Correlation Analysis Between Dataset Numerical Attributes



Conclusion

Our findings will be of interest to a wide range of stakeholders, including city planners, policy-makers, residents, prospective Airbnb hosts, and travelers. The outcomes of this analysis may help inform policy decisions, promote sustainable tourism practices, and provide insights into the economic and social aspects of Airbnb's presence in Amsterdam.

This report embarks on an exploratory data analysis (EDA) project focused on Airbnb listings in Amsterdam. The primary objective of this study is to gain deeper insights into the dynamics of Airbnb's presence in the city and understand the implications of its growth on various aspects of Amsterdam's socio-economic landscape. By examining and analyzing a rich dataset of Airbnb listings, we aim to uncover valuable patterns and trends that shed light on the distribution, pricing, and characteristics of these accommodations.