**Business Intelligence Course Project**

**Dataset: NYC AirBNB**

**Group Members:**

**Ali Khurram (25250), Ali Iqbal (24529)**

1. **Description of Dataset**

The following Airbnb activity is included in this New York dataset:

Listings, including full descriptions and average review score Reviews, including unique id for each reviewer and detailed comments Calendar, including listing id and the price and availability for that day

1. **Data Wrangling Steps:**

**1. Removed Data:**

* Eliminated the **'license'** column due to almost empty entries.
* Removed rows with complete empty entries.
* Removed duplicate rows.
* Eliminated rows with identical latitude, longitude, host name, and price, considering them duplicates. This step significantly reduced the dataset from 102,599 to 69,305 entries.

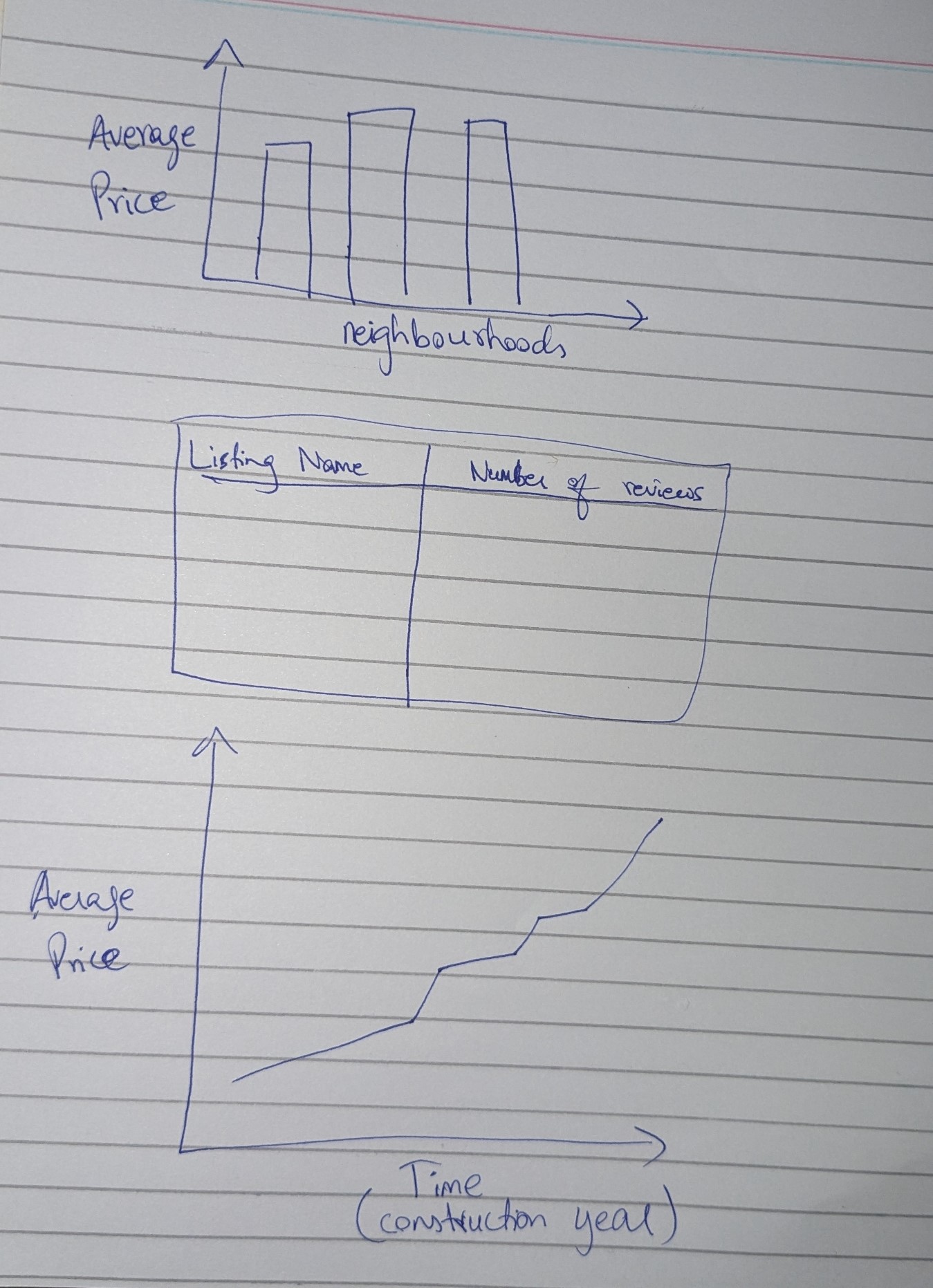
**2. Corrected Inaccuracies:**

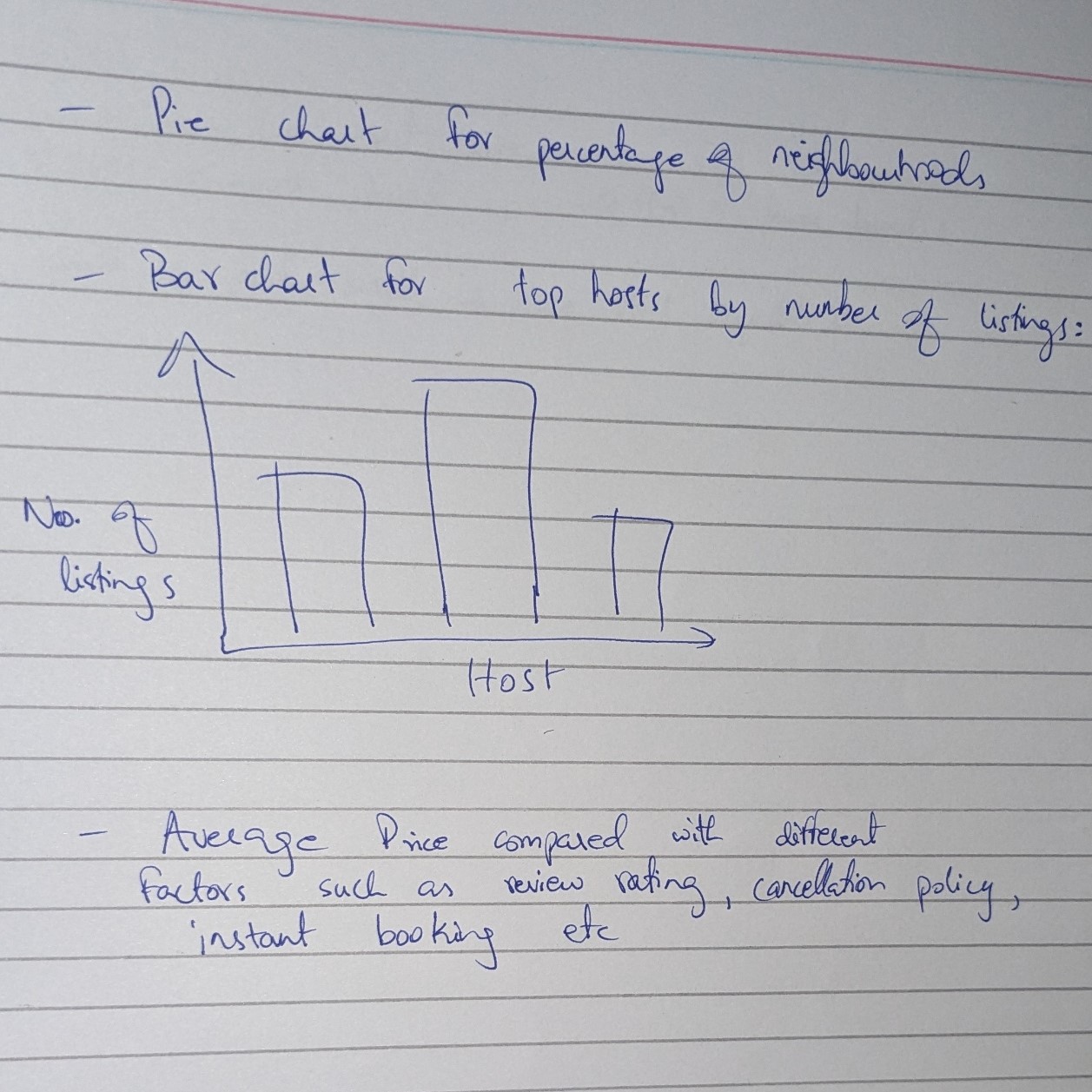
* Adjusted dates in the **'last review'** column greater than the year 2022 to the column's median.
* Changed the data type of **'price'** and **'service fee'** columns from object to float.
* Corrected the spellings of **'Manhattan'** and **'Brooklyn'** in the 'neighbourhood group' column.

**3. Filled Missing Values:**

* Imputed the missing values in the **'house\_rules'**, **'name'**, and **'host name'** columns with the word **'blank'**.
* Imputed missing dates in the **'last review'** column with the earliest date available.
* Used the median to fill missing values in columns such as **'reviews per month', 'availability 365', 'minimum nights', 'review rate number', 'construction year', 'number of reviews', and 'calculated host listings count'.**
* Filled missing values in the **'id'** column with 0.
* Imputed missing values in the **'host\_identity\_verified'** column with **'unconfirmed'**.
* Dropped **'country'** and **'country code**' columns due to zero variability.
* Filled missing values in the **'neighbourhood'** and **'neighbourhood group'** columns using latitudes and longitudes.
* Imputed missing values of **latitude** and **longitude** using existing data for the same neighborhood.
* Filled missing values in the **'price'** and **'service fee'** columns with the mean.
* Imputed missing values in the **'instant bookable'** column with "true".
* Filled missing values in the **'cancellation policy'** column with "moderate".

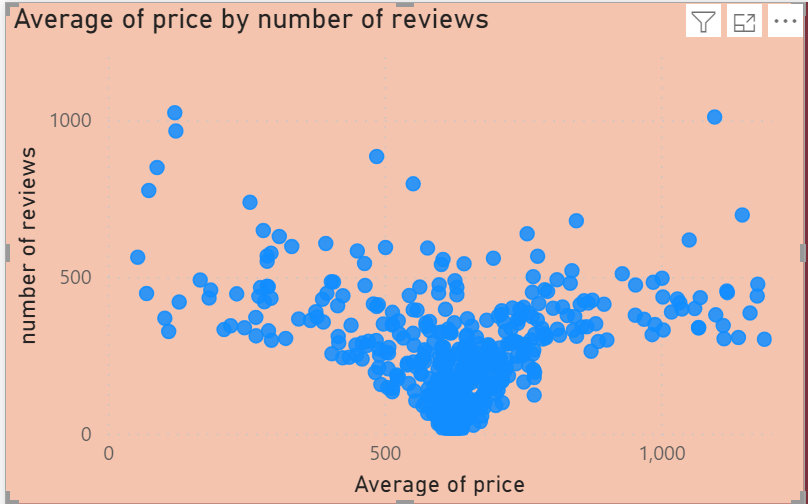
1. **Brainstorming:**

****

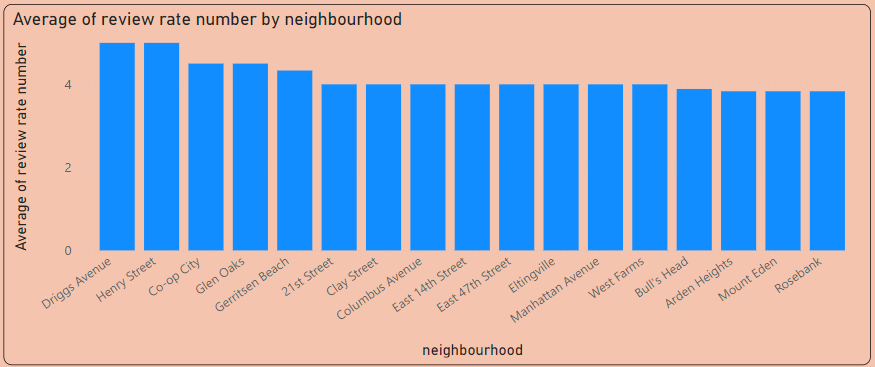
****

1. **Charts Made on PowerBI:**

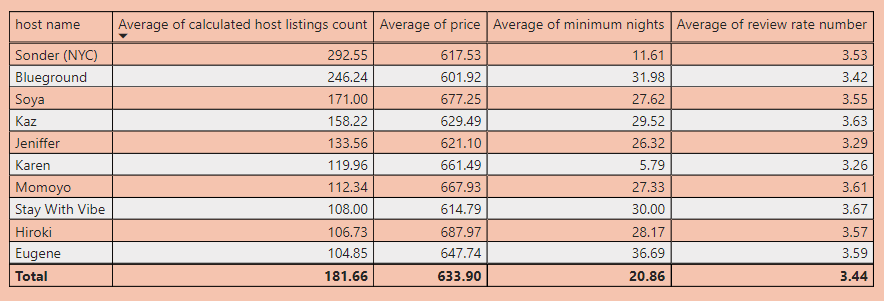
**Analysis of Price vs Number of reviews for a listing**

****

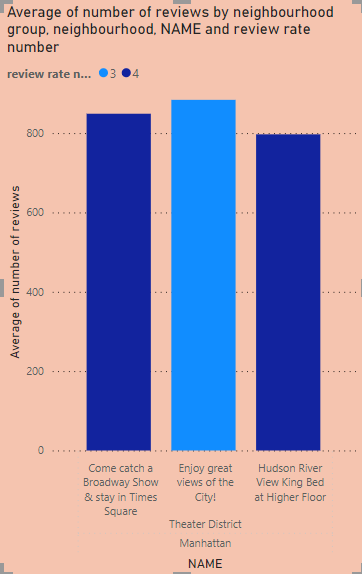
**Neighbourhoods ranked by review ratings:**

****

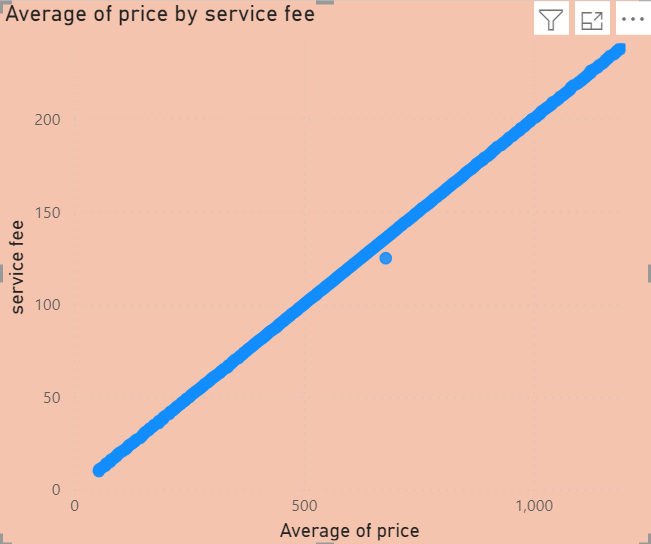
**Top 10 hosts according to number of host listings on AirBNB:**

****

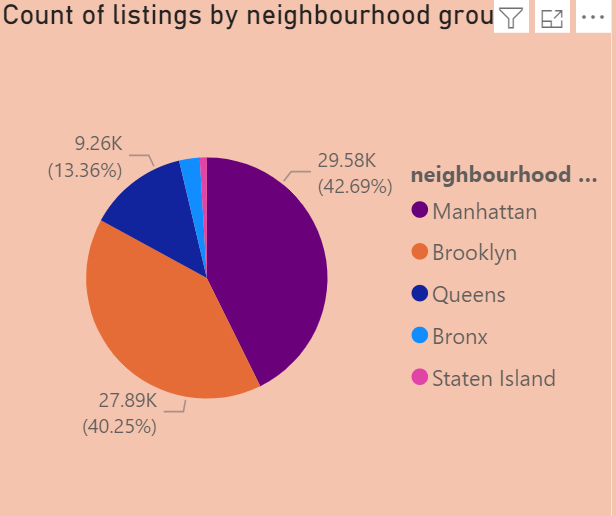
**Listings with highest number of reviews, with their rating indicated:**

****

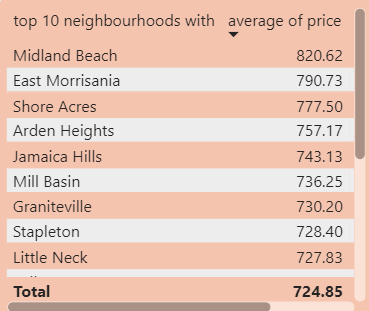
**Average of Price vs Average of Service Fee:**

****

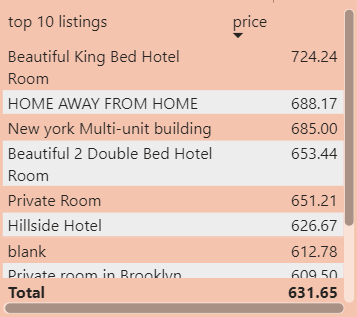
**Count of Listings by neighbourhood Group:**

****

**Top 10 Neighbourhoods by Average Price:**

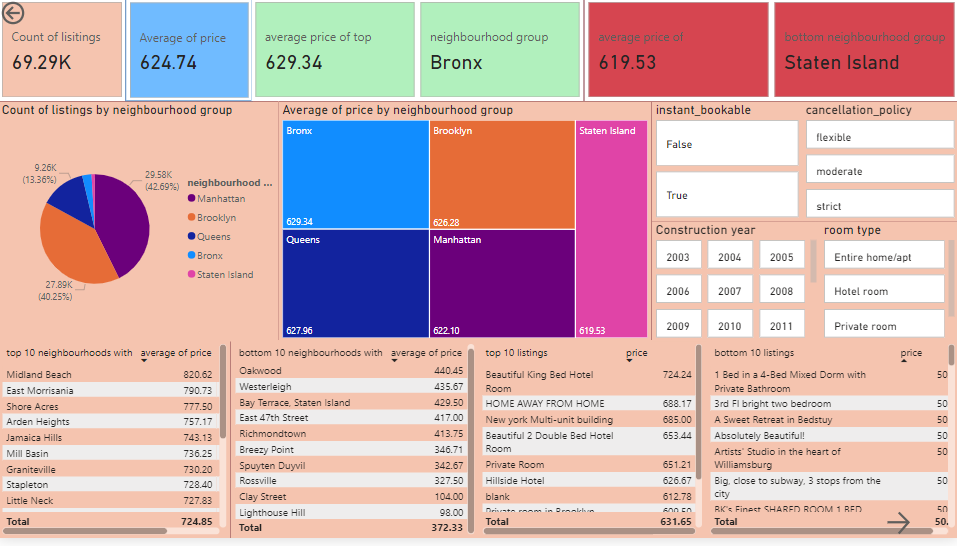
****

**Top 10 Listings by Price:**

****

1. **Dashboards:**

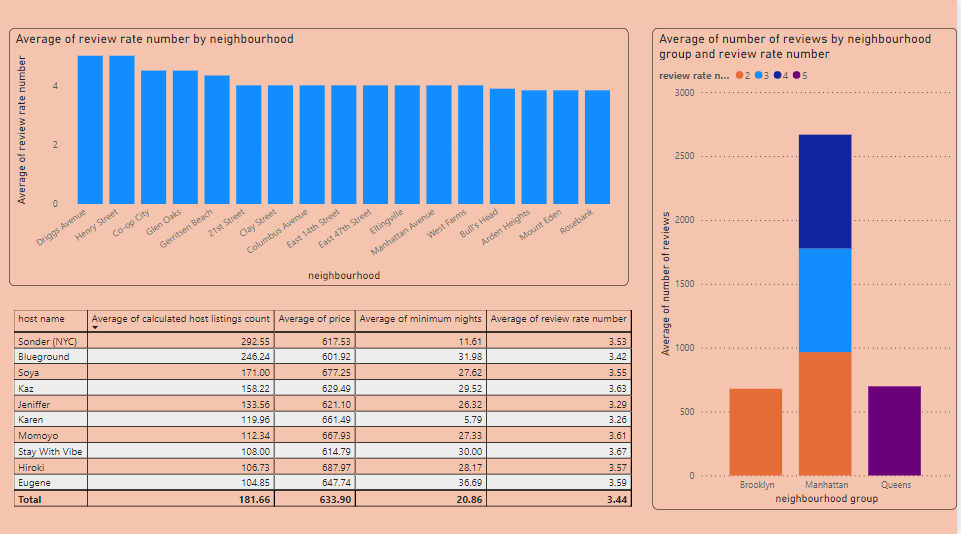
**Price Analysis Dashboard: Overview of Prices**

****

**Second Price Analysis Dashboard: Shows comparison of price with different variables such as number of reviews, service fee etc:**

****

**Third Page with additional charts for analysis of top hosts and listings with highest reviews:**

****

1. **Team Member Contributions:**

Ali Iqbal:

* Made charts on price analysis
* Made the dashboard
* Filled background data knowledge template

Ali Khurram:

* Performed data cleaning
* Made charts on host and reviews analysis
* Made the report