# Airbnb Geneva Data Integration & Analysis Documentation

## 1. Introduction

This document outlines the process of integrating and analyzing the Airbnb data for Geneva, Switzerland, sourced from Kaggle. The aim was to provide insights into the relationship between property type, average price, and average ratings.

## 2. Data Source

The dataset was downloaded from Kaggle using the following link: (https://www.kaggle.com/datasets/prayankkul/airbnb-geneva-switzerland-29-jun-2021/data)

## 3. Data Integration

### 3.1. Star Schema Selection

A star schema was chosen for this dataset to simplify querying and to promote self-service for analysts. This schema centralizes the main data elements (fact tables) and connects them to descriptive categories (dimension tables), enabling easier and more intuitive querying.

### 3.2. Data Insertion

The dataset was ingested into the database using the SQL COPY INTO command to process data efficiently.

### 3.3. Processing corrupted data with PySpark

PySpark was used for dealing with corrupted files. The trim_whitespace function helps clean the data by removing unnecessary spaces. This method aids in ensuring data quality and uniformity.

# 4. Data cube design

## 4.1. Dimensions:

- **Listing Details Dimension:**

  **Purpose**: This dimension encapsulates the details associated with a specific listing. It serves as a backbone to track how different listings perform in the market.

- **Review Details Dimension:**

  **Purpose**: Reviews play a critical role in a guest's decision-making process. This dimension helps in understanding guest sentiment and feedback..

- **Neighbourhood Information Dimension:**

  **Purpose**: The location of a property plays a pivotal role in its demand and price. This dimension provides insights into how different neighbourhoods fare against each other.

## 4.2. Measures:

- **Average Rating**:

  **Purpose**: Provides an overall sentiment score for listings. It is a crucial metric as listings with higher ratings tend to attract more guests.

  **Calculation**: Total rating points divided by the number of reviews.

- **Average Price:**

  **Purpose**: To understand the price trends. It can provide insights into whether certain listings or neighbourhoods command a premium.

  **Calculation**: Total price from all listings divided by the number of listings.

- **Count of Listings:**

  **Purpose**: Gives an idea about the volume of properties listed. Helps in understanding market supply.

  **Calculation**: Total number of unique listings.

- **Count of Reviews:**

  **Purpose**: Helps in gauging the popularity and feedback volume for listings.

  **Calculation**: Total number of reviews given across all listings.

# 5. Data Processing: Staging

The data loading style chosen is "truncate and insert". This ensures that the staging tables are refreshed with the most recent data, avoiding any duplications.

# 6. Data Warehousing:

**6.1. Dimension and Fact:** The MERGE INTO command was efficiently used for upsert operations – inserting records if they don't exist and updating if they do. The MERGE INTO command helps in maintaining data integrity and ensures the latest data is reflected in the dimension tables. The fact tables are populated with measures aggregated from the staging tables, which can be queried for insights.

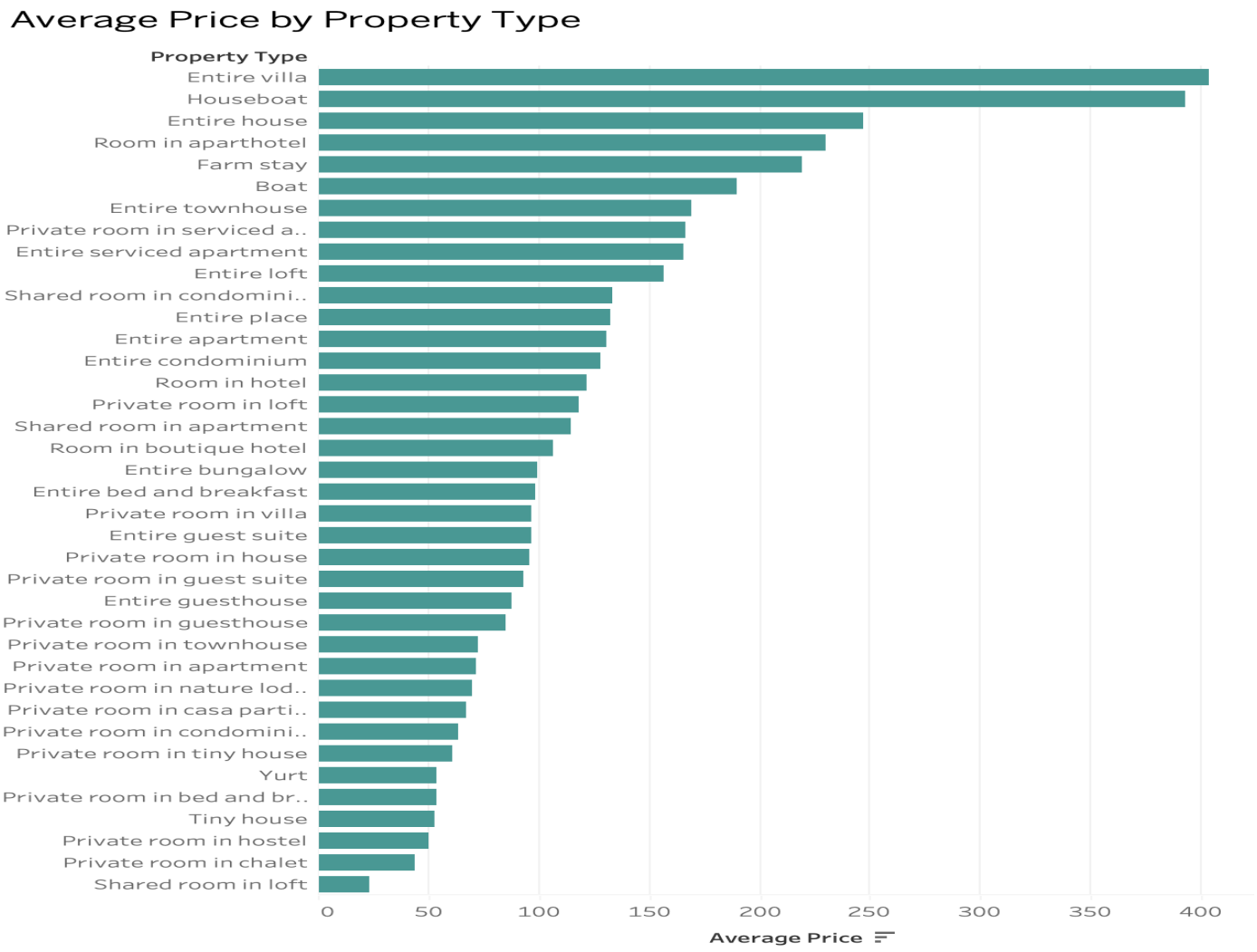
**6.2. Reporting and Visualization**: A SQL view was created to aid in the reporting and visualization of the insights.This view aggregates and ranks the properties, making it easier to derive insights without querying the base tables directly.

# 7. Visualization

By exporting this data and visualizing it using Tableau, we can communicate findings effectively and provide a visual narrative.

## 7.1. Average Price by Property Type:

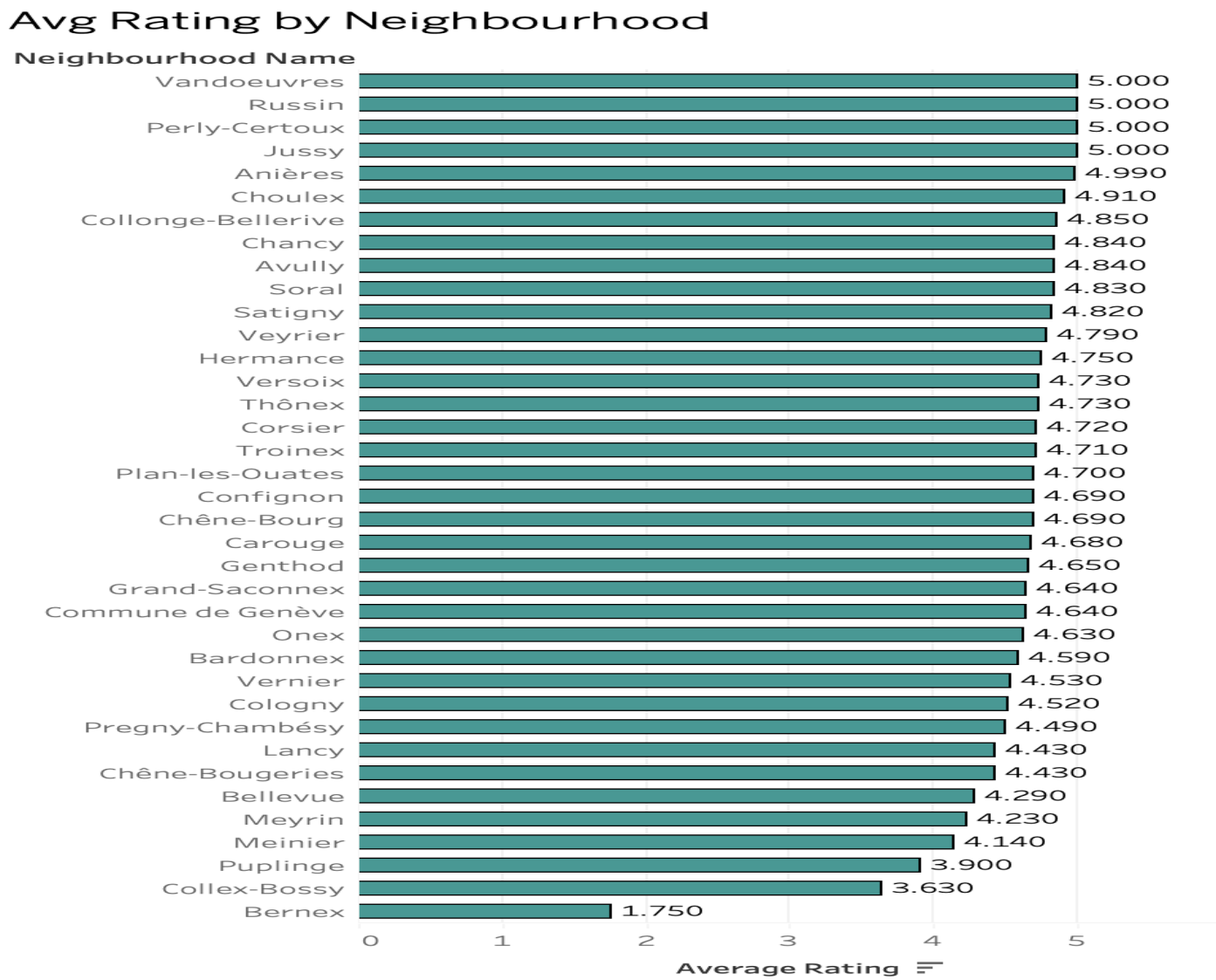
The most expensive property type on average appears to be "Entire villa", followed by "Houseboat" and "Entire house". On the lower end of the spectrum, properties like "Shared room in loft", "Private room in chalet", and "Private room in hostel" have the least average price. Most property types cluster around the price range of 100 to 250.




Average of Avg Price for each Property Type.

## 7.2. Average Rating by Neighbourhood:

The neighbourhoods of Vandoeuvres, Russin, Perly-Certoux, and Jussy boast perfect average ratings of 5.000. In contrast, Bernex has a significantly lower average rating of 1.750. The majority of the neighbourhoods have average ratings between 4.5 and 5.0, indicating generally positive feedback from renters.

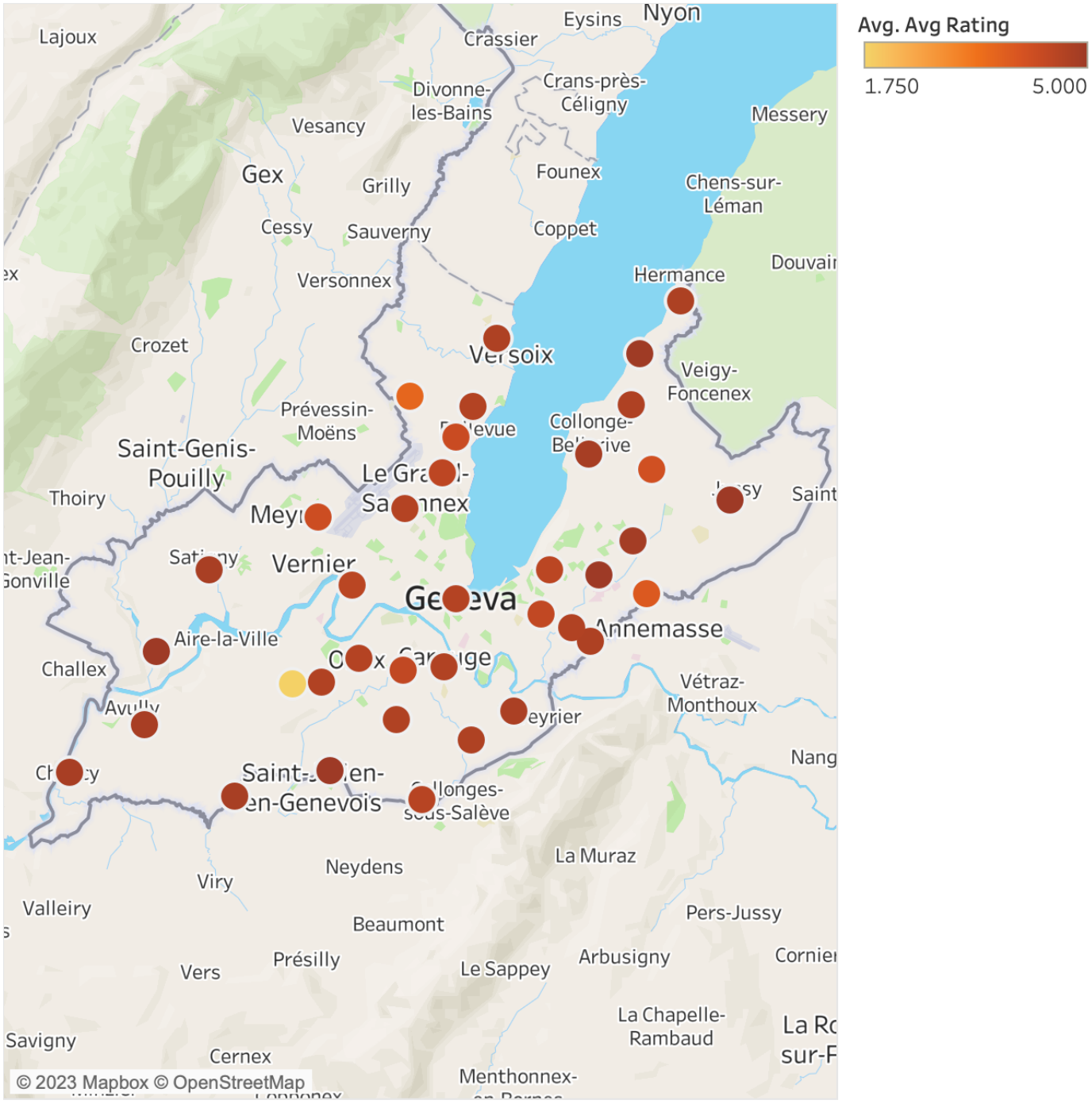



Average of Avg Rating for each Neighbourhood Name. The marks are labeled by average of Avg Rating. The data is filtered on Most Common Property Type, which keeps 15 of 15 members.

## 7.3. Neighbourhood Ratings Map:

This map provides a visual representation of the average ratings of different neighbourhoods around Geneva. Higher ratings are represented by darker shades of red, while lower ratings are in a lighter orange hue. Areas like Vandoeuvres, Russin, and Perly-Certoux (which had perfect average ratings in the bar graph) are easily identifiable as dark red dots. Bernex, with its lower rating, can be identified as a lighter dot.

## Neighbourhood Ratings Map

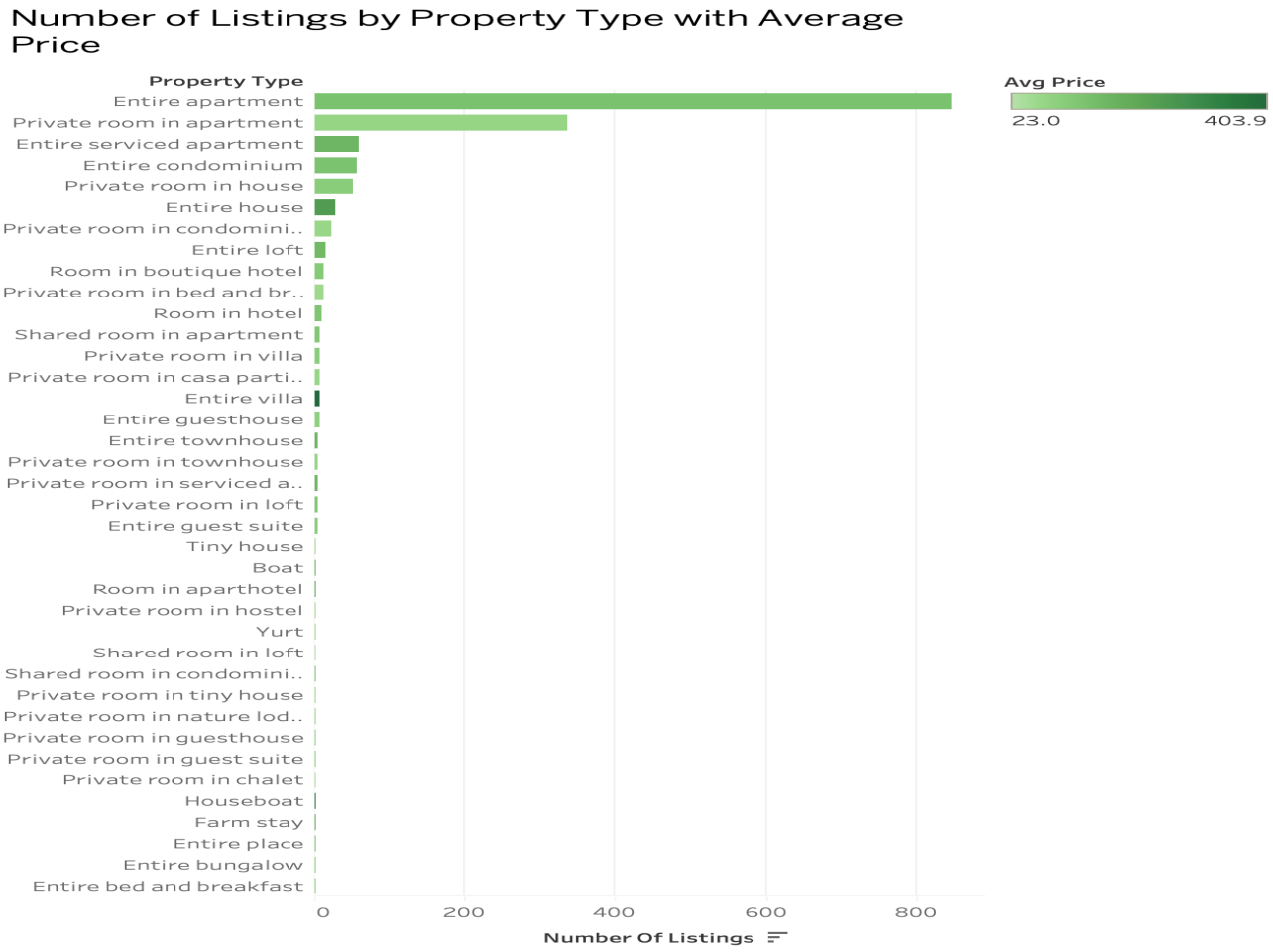



Map based on average of Longitude and average of Latitude. Color shows average of Avg Rating. Details are shown for Neighbourhood Name.

### 7.4. Number of Listings by Property Type with Average Price:

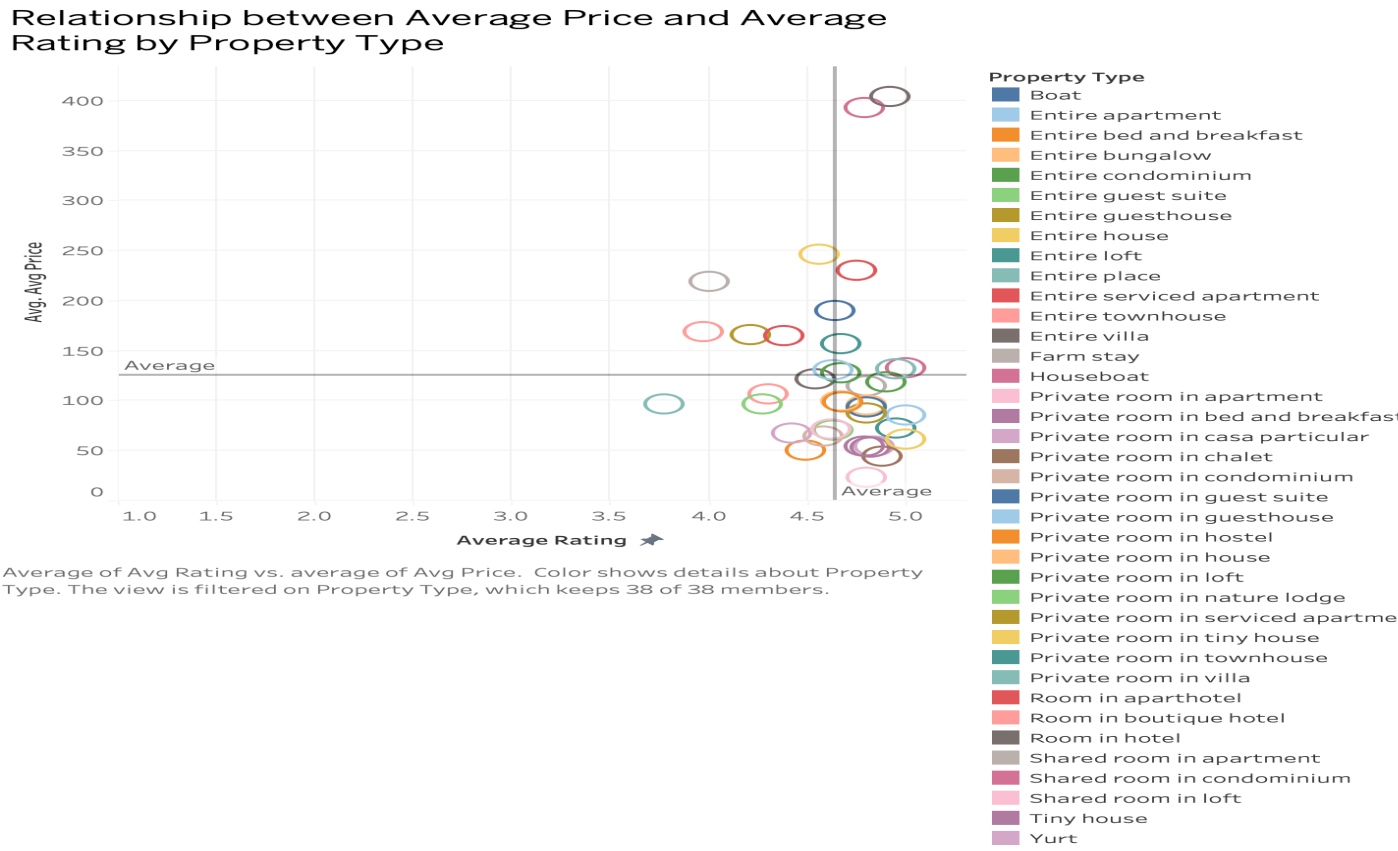
"Entire apartments" have the highest number of listings by a significant margin, suggesting they are the most commonly listed property type. On the contrary, unique listings like "Boat,"

"Tiny house," "Room in aparthotel," and "Private room in a hostel" have fewer listings, indicating their rarity or niche appeal. The color shading (representing average price) shows that property types like "Private room in a villa," "Entire villa," and "Entire guesthouse" have a higher average price compared to other listings.




## 7.5. Relationship between Average Price and Average Rating by Property Type:

It appears that the majority of properties cluster around an average rating between 3.5 to 4.5, irrespective of their average price. The property types "Entire apartment," "Private room in an apartment," and "Entire serviced apartment" seem to be prevalent as they have multiple data points spread across different price ranges. Some unique property types like "Boat" have a higher average price but a narrower range of ratings.




Average of Avg Rating vs. average of Avg Price. Color shows details about Property Type. The view is filtered on Property Type, which keeps 38 of 38 members.

## 8. Conclusion

The analysis provides a comprehensive understanding of the Geneva Airbnb market. The key takeaways are:

- "Entire apartments" are the most popular listings, suggesting that tourists and visitors prefer having a complete space to themselves. Price is not always indicative of guest satisfaction. Many affordable listings have commendable average ratings, emphasizing the importance of quality and service.
- High-end properties like "Entire villa" and "Houseboat" have maintained good average ratings despite their premium pricing. Interestingly, the "Shared room in loft" and "Private room in hostel", despite being on the lower end of the price spectrum, still manage to maintain decent average ratings, indicating that price is not the sole factor determining guest satisfaction.
- Neighborhood plays a significant role in guest satisfaction. Areas like Vandoeuvres, Russin, and Perly-Certoux are leading in terms of average ratings.

- Visualizing the data offers actionable insights for both property hosts and potential guests. Hosts can gauge market demand and adjust their pricing or services accordingly, while guests can make informed decisions based on price, rating, and property type.

## 9. Future Recommendations

To enhance this analysis further:

- Incorporate time-series data to understand seasonal trends and fluctuations in the Geneva Airbnb market.
- Collaborate with other available datasets like city events, hotel pricing, or tourist data to provide a more holistic view of Geneva's hospitality scene.

## 10. Databricks Notebook

Please find the link to the Databricks notebook here.