

Dataset explanatory report

DPS960 – Group 1

In this report, we will discuss the dataset that we will use for our course project. The dataset that we have chosen is the "Airbnb Prices in European Cities" dataset, with a focus on Amsterdam city. This dataset is available on Kaggle, and it contains information on Airbnb listings in several European cities, including Amsterdam.

Question 1. What dataset will you use for your course project? (Describe your dataset, and either include a link to where I can find it online or submit it as a spreadsheet along with your report).

The dataset being used for the course project is called "Airbnb prices in European Cities" and can be found on Kaggle using this link: <https://www.kaggle.com/datasets/thedevastator/airbnb-prices-in-european-cities>. This dataset contains information about Airbnb listings across various European cities, including Amsterdam.

Question 2. Describe the dataset. What kind of data does it contain?

The "Airbnb prices in European Cities" dataset contains information on Airbnb listings across various European cities, including Amsterdam. The dataset contains the following columns:

realSum: The total price of the Airbnb listing.

room_type: The type of room being listed, such as private, shared, or entire home/apartment.

host_is_superhost: A boolean value indicating if the host is a superhost or not.

multi: An indicator whether the listing is for multiple rooms or not.

biz: A business indicator.

guest_satisfaction_overall: The overall rating from guests comparing all listings offered by the host.

bedrooms: The number of bedrooms in the Airbnb listing.

dist: The distance from the city center in kilometers.

lng: The longitude coordinates of the Airbnb listing.

lat: The latitude coordinates of the Airbnb listing.

We believe that this dataset will provide valuable insight into how social dynamics and geographical factors affect global markets, and how pricing strategies can be optimized for profitability.

Question 3. Is there anything about your data that you don't understand? (i.e. what a column heading means)? How will you find this out?

There is nothing in the dataset that is unclear or difficult to understand. All of the column headings are clear and concise, and the data is well-organized. However, if any questions or uncertainties arise during the course of the project, we will consult the dataset's documentation or reach out to the Kaggle community for clarification.

In summary, we have chosen the "Airbnb Prices in European Cities" dataset, with a focus on Amsterdam city, for our course project. This dataset provides insights into the Airbnb market in Amsterdam, such as the average price of listings, the most popular neighbourhoods, and the most common room types. We have not found any column headings in the dataset that we don't understand, but we will need to do further research to understand some of the columns better. Overall, we believe that this dataset will be a valuable resource for our course project and will provide us with meaningful insights into the Airbnb market in Amsterdam.