CSE-564 VISUALIZATION

LAB REPORT – I

Seattle Airbnb Open Data

By: Mayuresh Pingale

SBU ID: 114910589

Aim: The project aims to visualize a dataset by using various charts developed using D3.JS library. The dataset should contain a good mixture of Numerical and Categorical variables and should have at least 15 variables. The dashboard should provide bar graph, histogram and scatter plot

Dataset:

The Seattle Airbnb open dataset was taken from the Kaggle publicly made available by the Airbnb. This dataset describes the listing activity of homestays in Seattle, Washington. It includes full descriptions, average score, reviews, availability, etc. The dataset contains Seattle Airbnb listings, ratings, and related data. The dataset is a good mix of numerical and categorical variables with 3818 properties and 92 attributes describing them.

Link to Dataset: https://www.kaggle.com/datasets/airbnb/seattle

Dataset Pre-processing:

The dataset was reduced horizontally as well as sampled vertically using Excel to extract meaningful data and reduce the number of samples. A sample with incomplete information was cleaned, and then around 500 samples were taken randomly. The final features selected contained an equal amount of categorical and numerical variables.

Feature Selection:

The following are the definitions of the attributes in the sampled data.

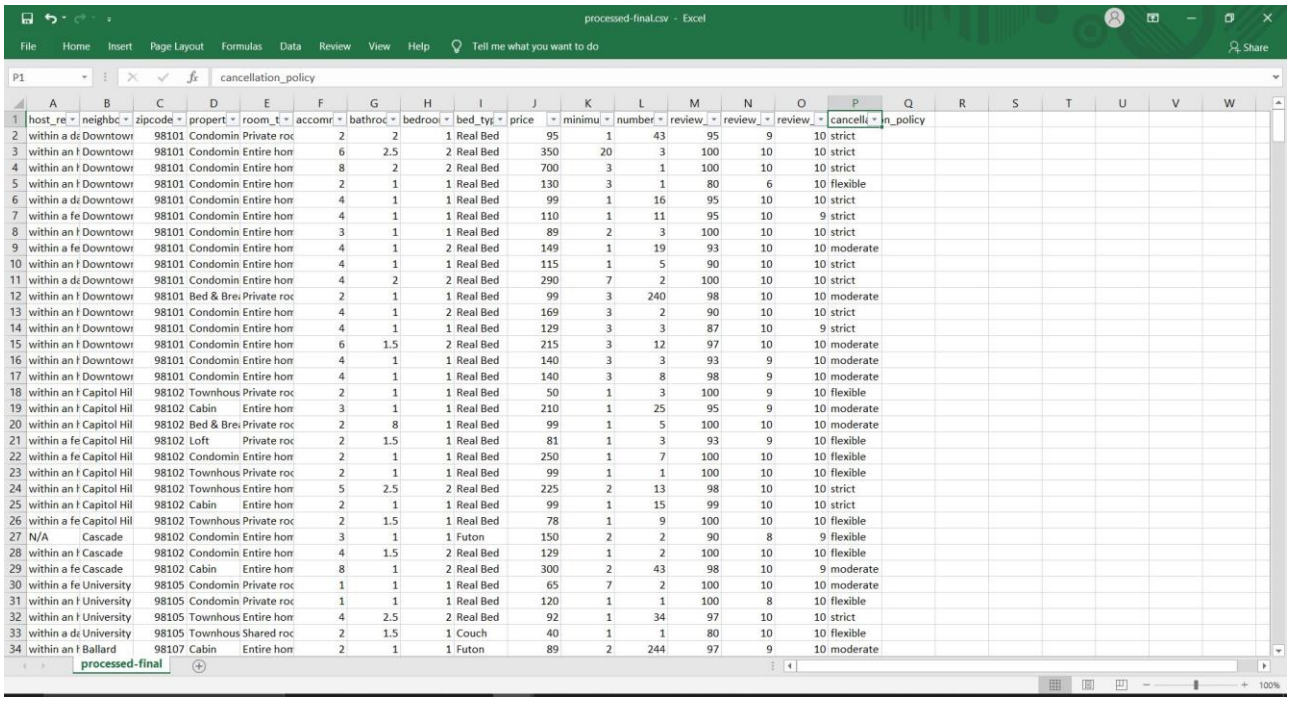
Numerical Variables

1. Accommodates - The number of people/guests that can stay on the property
2. Bathrooms - The number of restrooms available in the house

3. Bedrooms - The number of bedrooms in the house
4. Price - Price of the property in dollars per day
5. Minimum Nights - Minimum number of days the property needs to be booked.
6. Number of Reviews - Number of reviews available on the property
7. Review Score Ratings - The ratings of the property given by the previous customers.
8. Review Score Cleanliness- The cleanliness rating of the property
9. Review Score Location - Ratings of the location of the property

Categorical Variables

1. Neighbourhood - The neighbourhood of the property is located in the Seatle
2. Zip code - The zip code of the property
3. Property type - The type of property. It includes values such as apartments, condominiums, bungalows, Cabins, etc.
4. Room Type - The type of room in the property. It includes values like Shared Rooms, Private rooms, and the entire house.
5. Host Response Time - The time is taken by the house to reply. The categorical values are within an hour, day, etc.
6. Cancellation Policy - It consists of values such as strict, moderate, flexible
7. Bed Type - It describes the type of bed. The value consists of Couch, Futon, Real Bed, and Pull Bed.



Processed CSV

Importance/ Value of Dataset:

This dataset was selected because it gave the real-world picture of renting and listings properties on Airbnb in Seattle. The dataset had a good mixture of categorical and numerical variables that was useful in the visualization of different charts. The data is helpful to find the vibe of each Seattle neighbourhood. Also, it can be used to analyse what factor affects the pricing of the property.

Deployment:

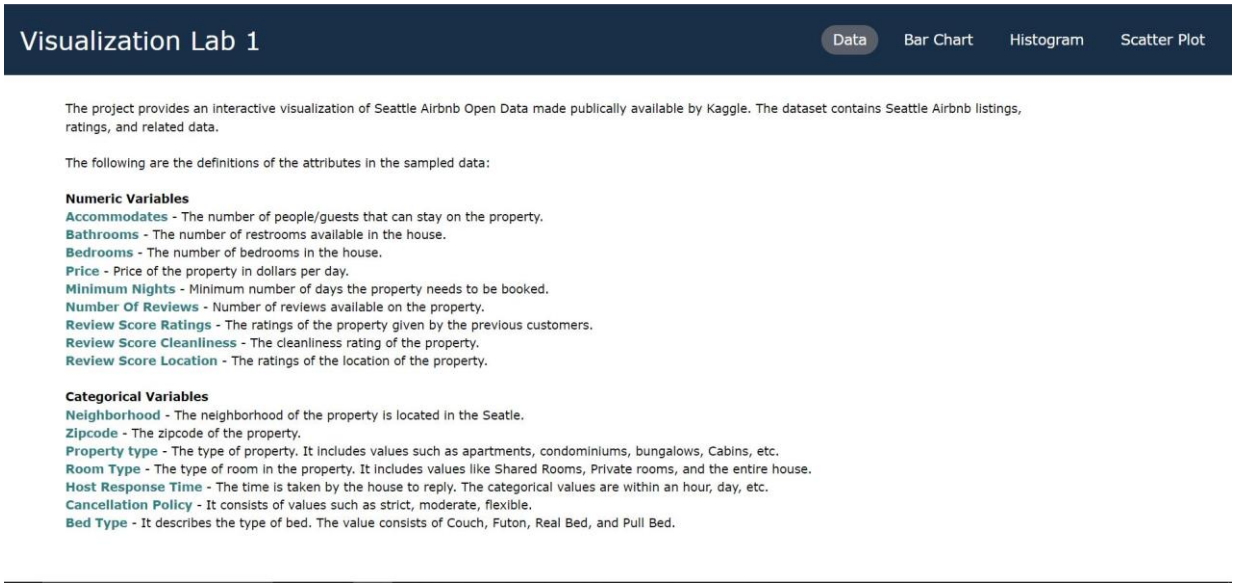
To run this project, just install any web server.

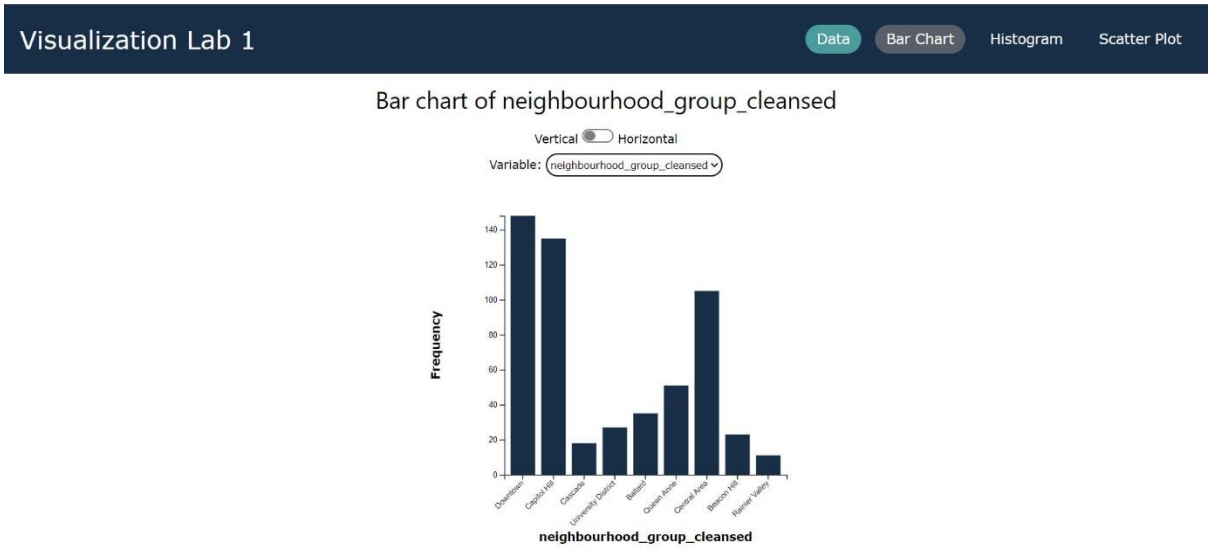
Run: python -m http.server 8080

Open Web Brower and type: localhost:8080

Features of Application:

The application opens with the data page that provides information about the dataset. The navigation bar provided ease to switch between tabs. The user can select the type of graph, and the corresponding page will be displayed. The project has 4 HTML files, each corresponding to a different tab. According to the requirements, categorical variables are plotted on the bar chart, and numerical variables are on the histogram.





For bar charts and histograms, the user can select one column, and a chart of that particular variable is displayed while the other axis is frequency. Users have an option of the toggle to switch between horizontal and vertical graphs. For example, above image shows a Bar chart of the categorical variable neighbourhood. Similarly, a histogram is plotted for a numerical variable.

For plotting a scatter plot, click on the scatter plot tab and provide two variables. A single dropdown provides the option to choose variables while the radio button selects the axis to be updated. If both variables are categorical, points are jittered to make them denser.

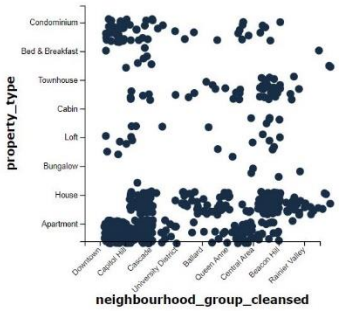
Visualization Lab 1

Data | Bar chart | Histogram | Scatter plot

property_type vs neighbourhood_group_cleansed

X-axis ◉ Y-axis ○

Variable: neighbourhood_group_cleansed



Type here to search

6°C Clear ENG 21:12 14-02-2023