CPS 5745 – Interactive Information Visualization

Final Project Technical Report

Name: Rayleen Ramos
Course: CPS 5745 – Interactive Information Visualization
Instructor: Dr. Huang
Dataset Source: Inside Airbnb (https://insideairbnb.com/get-the-data/)

1. Project Overview

This project implements an interactive information visualization system using a real-world Airbnb dataset. The webpage allows users to explore pricing trends across New York City Airbnb listings through dynamic charts, filters, and data tables. The system integrates MySQL, PHP, JavaScript (Chart.js & Google Charts), HTML, CSS, and AJAX to demonstrate full-stack data visualization concepts.

The primary goals of this project are:

- To analyze Airbnb pricing patterns by borough, neighborhood, and number of reviews
- To provide meaningful user-controlled interactivity through sliders and dropdowns
- To highlight outliers using quartile (IQR) method
- To maintain user preferences across sessions

2. Demo Login Information

To allow testing and evaluation of the system, the following demo credentials are provided:

- **Login:** tiger
- **Password:** 5920

These credentials grant access to all project functionality, including database loading, interactive visualization, and saving user settings.

3. Dataset Description

The dataset used in this project was obtained from Inside Airbnb, a public, non-commercial data source that provides detailed Airbnb listing information for major cities worldwide. For this project, the New York City, New York, United States dataset was selected. The original dataset contains over 79 attributes and approximately 36,000 records, including both numeric and categorical fields. Approximately 10 columns were excluded to simplify analysis and improve clarity. Figure 1 shows the cleaned dataset after column reduction.

neighbourhood_cleans...	neighbourhood_group_clean...	latitude	longitude	property_type	room_type	accommodat...	bathrooms	bedrooms	beds	price	number_of_revie...
Midtown	Manhattan	40.75356	-73.98559	Entire rental unit	Entire home/apt	1	1.0	0	1	240.00	47
Williamsburg	Brooklyn	40.70935	-73.95342	Entire rental unit	Entire home/apt	3	1.0	2	1	96.00	195
East Harlem	Manhattan	40.80107	-73.94255	Private room in condo	Private room	1	1.0	1	1	59.00	1
East Harlem	Manhattan	40.78778	-73.94759	Private room in rental unit	Private room	1	1.0	2	2	73.00	249

Figure 1.

4. Loading Data & Data Marts

To load the Airbnb dataset into the database, a table named airbnb was first created in MySQL. The cleaned dataset was then inserted into the table using terminal SQL INSERT commands, ensuring each column aligned with its corresponding attribute. Furthermore, to support efficient querying and meaningful visualizations, multiple data marts were derived from the raw airbnb table:

Data Mart 1: Price by Borough

This data mart groups Airbnb listings by borough and calculates the total price, number of listings, and average price for each borough.

borough	total_price	listing_cou...	avg_price
Manhattan	1105227.00	3485	317.14
Brooklyn	575211.00	3195	180.03
Bronx	50198.00	334	150.29
Queens	171199.00	1308	130.89
Staten Island	18402.00	146	126.04

Figure 2.

Data Mart 2: Price by Neighborhood

This data mart aggregates Airbnb listings at the neighborhood level within each borough. It calculates the total price, number of listings, and average price for each neighborhood.

neighborhood	borough	total_price	listing_count	avg_price
Allerton	Bronx	1958.00	13	150.62
Arden Heights	Staten Island	134.00	2	67.00
Arrochar	Staten Island	831.00	6	138.50
Arverne	Queens	9521.00	37	257.32
Astoria	Queens	15565.00	139	111.98

Figure 3.

Data Mart 3: Price by Number of Reviews

This data mart groups listings by their number of reviews and calculates the average price for each review count to analyze the relationship between listing popularity and pricing.

number_of_revie...	total_price	listing_count	avg_price
0	340733.00	780	436.84
1	120116.00	396	303.32
2	54615.00	273	200.05
3	63354.00	215	294.67
4	39716.00	196	202.63

Figure 4.

5. System Design & Layout

The webpage features a clean, single-page layout with a pastel color theme designed for visual appeal and readability.

Layout Structure

- **Header Section:**
  - Left Side: Contains the Login form and a dynamic message area for system feedback.
  - Right Side: Contains action buttons, Load DB Data, Save Settings, and Logout. Above these buttons, there is an accessibility Help button. This split design provides better page organization.
- **Control Panel:**
  - Borough dropdown (categorical filter)
  - Price range slider (numeric filter)
- **Visualization Area:**
  - Chart 1: Average Price by Neighborhood/Borough (Bar Chart)
  - Chart 2: Price vs Number of Reviews (Line Chart)

- **Tabbed Data Tables:**
  - Raw Airbnb data
  - Borough data mart
  - Neighborhood data mart
  - Review data mart

Tabs are used to switch between the raw dataset and aggregated data marts. This approach allows users to explore different views of the data without requiring page reloads.

6. Interactive Features

Slider (Numeric Interaction)

The price range slider provides numeric filtering for the dataset and is directly linked to the average price by neighborhood bar chart visualization. The slider spans the minimum and maximum price values available in the dataset and allows users to adjust the visible data range dynamically.

- Controls the minimum and maximum price values applied to the bar chart.
- Dynamically updates the chart in real time as the slider values change.
- Remains disabled until the user logs in and loads the database.
- Automatically restores previously saved slider values when a user logs in and has saved preferences.

Dropdown (Categorical Interaction)

The borough dropdown is a categorical filter used only for the bar chart visualization. When a borough is selected, the bar chart updates to display only the neighborhoods belonging to that borough.

- Allows filtering neighborhoods by borough for the bar chart.
- Includes an "All" option to display neighborhoods from all boroughs.
- Updates the bar chart immediately upon selection, providing real-time feedback.

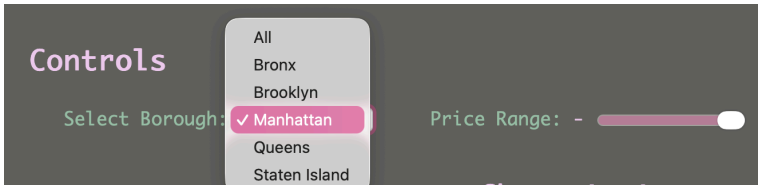

Figure 5.

7. Data Visualization & Analytics

Chart 1: Average Price by Neighborhood

- Displays average Airbnb prices across neighborhoods.
- Helps identify high-cost and low-cost areas.
- Uses aggregated neighborhood data calculated using SQL 'GROUP BY' query.

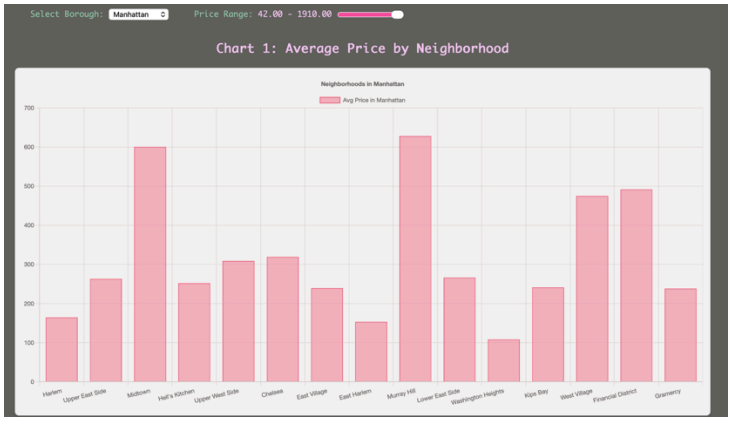

Figure 6.

Chart 2: Price vs Number of Reviews

- Line chart showing correlation between price and review count.
- Reveals pricing trends related to listing popularity.
- Shows whether highly reviewed places tend to be cheaper or more expensive.

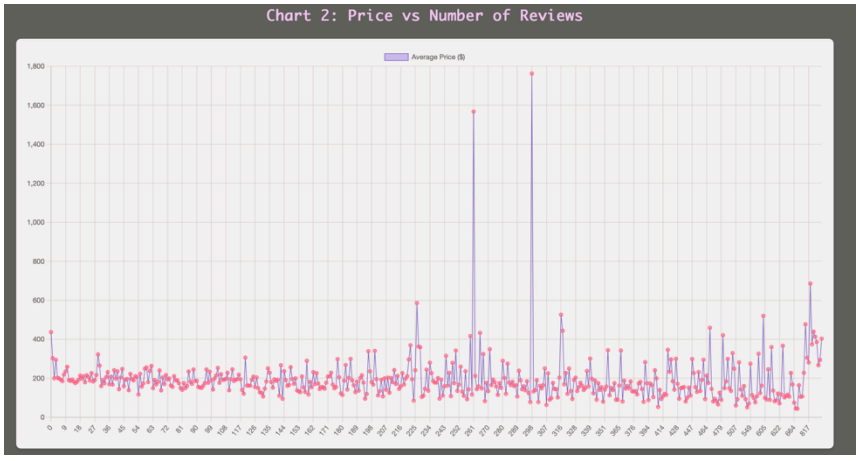

Figure 7.

Outlier Detection

- Uses the quartile (IQR) method.
- Rows containing outlier prices are highlighted in yellow.
- Outlier thresholds are displayed in a message area beneath the datasets for transparency, helping users understand why a row is highlighted.

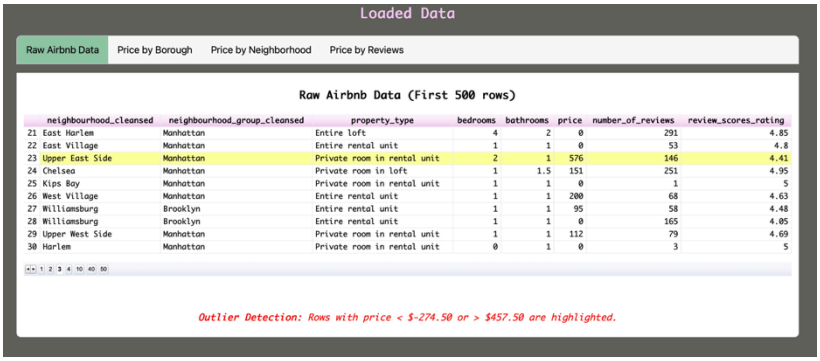

Figure 8.

8. User Settings Functionality

Users can save their preferred price range using the Save Settings button.

Features:

- Settings stored in the *User_Setting* table
- One record per user. Linked by the uid foreign key to the DV_User table in the datamining database.

- save_settings.php:
  - *INSERT* creates a new row for first-time users.
  - *UPDATE* modifies the existing row for returning users.
- app.js:
  - An event listener captures the current slider values and sends them via fetch() POST request.
  - Saved settings automatically load when the user logs in

- *get_settings.php*
  - Retrieves saved user preferences on login to restore the session.

uid	login	slider_low_value	slider_high_val...	datetime
1		42.00	1260.00	2025-12-13 03:29:53

Figure 9.

9. Code Architecture Overview

Each layer communicates cleanly, following a standard three-tier architecture.

Presentation Layer

- ***index.html***- The structure of the webpage. It defines headers, buttons, and chart containers that the user interacts with.
- ***style.css*** - The visual aesthetic of the webpage, such as colors, fonts, and layout. It correlates with HTML by using IDs and classes to apply responsiveness and styling.

Application Layer

- ***app.js***- This file manages all interactive behavior on the webpage. It listens for user actions such as selecting a borough, adjusting the price slider, or clicking buttons. When an action occurs, app.js sends AJAX requests to the server and uses the returned data to update the charts and tables displayed on the page.
- ***login.php*** + ***logout.php***- Handle user authentication and session management.
- ***load_data.php***- Retrieves Airbnb data and aggregated data from the database.

- ***save_setting.php*** - Safely saves user slider preferences to the database.
- ***get_settings.php***- Retrieves saved user preferences on login to restore the session.

Data Layer

- MySQL tables (airbnb, data marts, User_Setting)
- Database connection via ***dbconfig.php***

10. Conclusion

The results of this project show that Airbnb pricing in New York City is mainly driven by neighborhood-level differences rather than being evenly distributed across boroughs. A small number of neighborhoods consistently exhibit higher average prices, while many others remain relatively affordable. Analysis of price versus number of reviews indicates that listing popularity does not strongly determine price, suggesting that location plays a more significant role in pricing.

Overall, this project demonstrates the development of an interactive information visualization application using real-world Airbnb data. By combining dynamic charts, interactive filters, outlier detection, and continuous user preferences, the application enables meaningful exploration of pricing trends across New York City. The project fulfills all functional and technical requirements of CPS 5745 and showcases the practical application of a full-stack web application for data-driven visualization.