US Rental Market Analysis

Shilpa Shivarudraiah [016079973]

Department of Applied Data Science, San Jose State University

DATA 230-27 Data Visualization

Andrew H. Bond

December 06, 2023

ABSTRACT

In the United States, rental prices have displayed significant fluctuations, triggering investigations into their root causes, particularly in the context of mounting concerns regarding housing affordability. The dynamic nature of rental rates is influenced by a complex interplay of factors, including regional economic conditions, demographic shifts, urbanization trends, and changes in tenant preferences. To gain a comprehensive understanding of these fluctuations and their implications, it is imperative to conduct an exhaustive examination of rental rate disparities across diverse regions of the country. The study harnesses a comprehensive Kaggle dataset, encompassing rental properties across the entirety of the United States. This dataset comprises a wide spectrum of attributes, spanning from property dimensions and locations to amenities, with a particular focus on the dynamic aspect of rental prices. Data visualization serves as an indispensable tool, facilitating the interpretation and extraction of insights from datasets, primarily by making trends readily visible. It also aids in enhancing data comprehension. The project is centered on the scrutiny of a dataset centered on house rentals. To unlock the insights, research employs Tableau Desktop 2023.2, enabling the creation of an array of visual representations, including bar charts, line graphs, scatter plots, geospatial maps, and tables. The outcome of this analysis holds profound significance for a diverse array of stakeholders. Real estate professionals will gain invaluable insights to guide their investment decisions. Prospective renters will find empowerment in making informed location choices, and policymakers will access essential information to formulate strategies aimed at addressing the pressing issue of housing affordability.

Acknowledgements

I would like to thank Professor Andrew H. Bond for his constant guidance and assistance throughout the course work and for the successful development of my project.

Table of Contents

1 Introduction 5

1.1 Project Goals and Objective 5

1.2 Project Deliverables 5

2 Dataset Description 5

2.1 Attributes Description 7

3 Data Pre-Processing 7

3.1 Handling Null Values 9

3.2 Handling Redundant Data 10

3.3 Handling Data Type 10

4 Tableau Visualization Dashboard Overview 13

4.1 Area Information (Real Estate Statistics) 13

4.2 Price Information (Rental Statistics) 14

4.3 Price Analysis based on Amenities 16

4.4 Amenities Information (Amenities Distribution) 17

4.5 Regional Information (Real Estate & Rental Statistics) 19

5 Summary 20

1 Introduction

1.1 Project Goals and Objective

The primary objective of this project is to conduct a comprehensive analysis to identify specific regions within the United States characterized by higher rental prices and to discern the key amenities that contribute significantly to the escalation of rental prices for houses. The ultimate goal is to gather valuable insights that will aid stakeholders in making strategic decisions.

To comprehend fluctuations and their implications, a thorough examination of rental rate disparities across diverse U.S. regions is crucial. The study utilizes a comprehensive Kaggle dataset covering rental properties nationwide, including varied attributes such as property dimensions, locations, and amenities, with a focus on dynamic rental prices. Data visualization, facilitated by Tableau, plays a key role in interpreting and extracting insights from the dataset, making trends visible and enhancing data comprehension. The project centers on scrutinizing a dataset focused on house rentals, employing Tableau to create visual representations like bar charts, line graphs, scatter plots, geospatial maps, and tables for unlocking valuable insights.

1.2 Project Deliverables

Deliverables	**Scheduled Date**
Project Topic / Abstract	23-Aug-2023
Data Collection	13-Sep-2023
Data Cleaning	27-Sep-2023
EDA	11-Oct-2023
Data Visualization Design	25-Oct-2023
Visualization Development	8-Nov-2023
Dashboard & Analysis	22-Nov-2023
Reporting & Documentation	01-Dec-2023
Presentation	06-Dec-2023

Table 1. Project Deliverables

2 Dataset Description

The dataset employed in this project originates from Kaggle and was initially curated by a user named Rekib Ahmed. This dataset is in the public domain and was originally gathered by

Austin Reese on January 7, 2020, using Craiglist.org as the source. It includes a comprehensive set of 265,190 records related to rental houses, providing information on rental prices and various attributes such as the number of bedrooms, bathrooms, parking availability, and laundry facilities. These attributes play a crucial role in evaluating the overall expenses associated with a rental property. The dataset, in its entirety, comprises 22 columns. Analyzing these factors and identifying relevant trends is pivotal for making strategic decisions that empower stakeholders to make informed choices.https://www.kaggle.com/datasets/rkb0023/houserentpredictiondataset

Figure 1 below shows a sample of records that are present in the housing rental dataset.

Figure 1

Rental Dataset

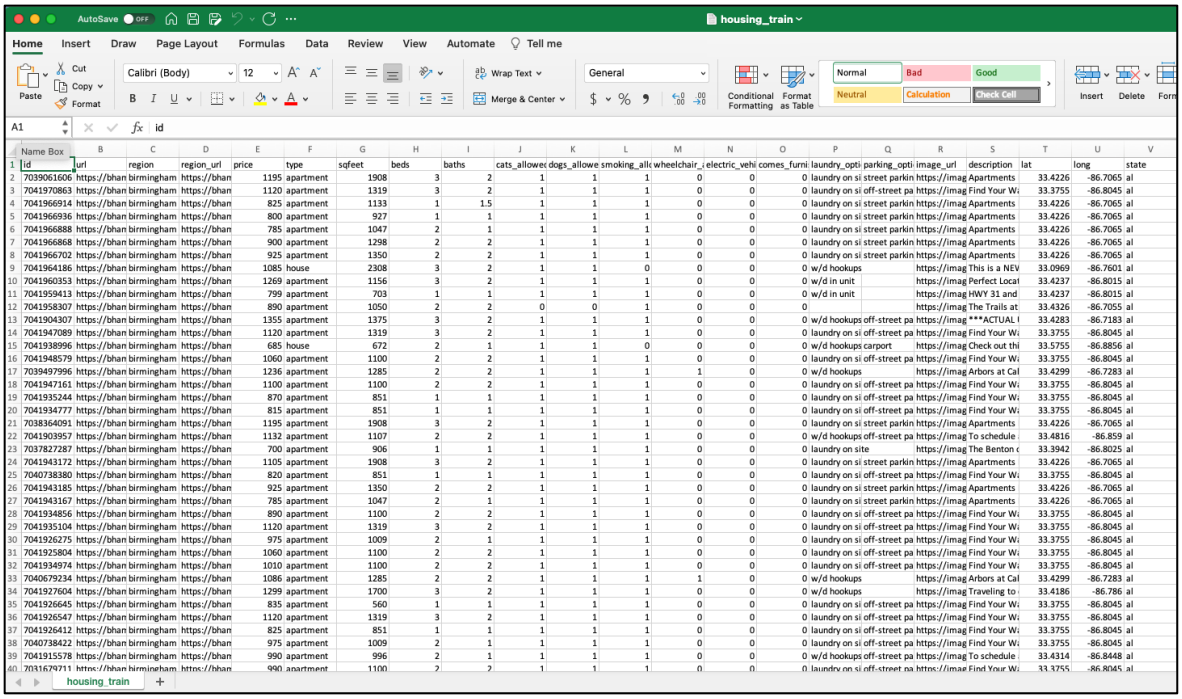


Figure 2 below shows the list of attributes present in the dataset.

Figure 2

Attributes Present in the Dataset

```
rental_csv.columns
```

```
Index(['id', 'url', 'region', 'region_url', 'price', 'type', 'sqfeet', 'beds',
       'baths', 'cats_allowed', 'dogs_allowed', 'smoking_allowed',
       'wheelchair_access', 'electric_vehicle_charge', 'comes_furnished',
       'laundry_options', 'parking_options', 'image_url', 'description', 'lat',
       'long', 'state'],
      dtype='object')
```

2.1 Attributes Description

1. Id: Listing id
2. url: Listing URL
3. region: Craigslist region
4. region_url: Region URL
5. price: Rent per month
6. type: Housing type
7. sqfeet: Total square footage
8. beds: Number of bedrooms
9. baths: Number of bathrooms
10. cats_allowed: Cats allowed Boolean (1 = Yes, 0 = No)
11. dogs_allowed: Dogs allowed Boolean (1 = Yes, 0 = No)
12. smoking_allowed: Smoking allowed Boolean (1 = Yes, 0 = No)
13. wheelchair_access: Has wheelchair access Boolean (1 = Yes, 0 = No)
14. electric_vehicle_charge: Has electric vehicle charger Boolean (1 = Yes, 0 = No)
15. comes_furnished: Comes with furniture Boolean (1 = Yes, 0 = No)
16. laundry_options: Laundry options available
17. parking_options: Parking options available
18. image_url: Image URL
19. description: Description by the poster
20. lat: Latitude
21. long: Longitude
22. state: State of listing

3 Data Pre-Processing

This segment will outline the procedures employed in data cleaning, essential for deriving insights and information applicable to the organization. As the primary phase in scrutinizing the raw data, an examination of the dataset information is conducted to assess data types and non-null counts, enhancing comprehension. Figure 3 illustrates details concerning the data type and count for each column.

Figure 3

Housing Dataset Information

```
rental_csv.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 265190 entries, 0 to 265189
Data columns (total 22 columns):
 #   Column                  Non-Null Count   Dtype  
---  ------                  --------------   -----  
 0   id                      265190 non-null  int64  
 1   url                     265190 non-null  object 
 2   region                  265190 non-null  object 
 3   region_url              265190 non-null  object 
 4   price                   265190 non-null  int64  
 5   type                    265190 non-null  object 
 6   sqfeet                  265190 non-null  int64  
 7   beds                    265190 non-null  int64  
 8   baths                   265190 non-null  float64
 9   cats_allowed            265190 non-null  int64  
 10  dogs_allowed            265190 non-null  int64  
 11  smoking_allowed         265190 non-null  int64  
 12  wheelchair_access       265190 non-null  int64  
 13  electric_vehicle_charge 265190 non-null  int64  
 14  comes_furnished         265190 non-null  int64  
 15  laundry_options         210879 non-null  object 
 16  parking_options         170055 non-null  object 
 17  image_url               265190 non-null  object 
 18  description             265188 non-null  object 
 19  lat                     263771 non-null  float64
 20  long                    263771 non-null  float64
 21  state                   265189 non-null  object 
dtypes: float64(3), int64(10), object(9)
memory usage: 44.5+ MB
```

Upon the preliminary examination, certain aspects requiring attention in the pre-processing phase included addressing null values, verifying data types of attributes, and identifying redundant data. Subsequent sections will elaborate on the approaches taken to address these issues in alignment with the dataset. Prior to initiating pre-processing, it is essential to examine the overall count of records. Figure 4 illustrates the total number of records and columns present in the rental dataset before pre-processing.

Figure 4

Total Number of records before Data Pre-Processing

```
print('Total number of records in the dataset:')
rental_csv.shape
```

```
Total number of records in the dataset:

(265190, 22)
```

3.1 Handling Null Values

This section addresses the management of null values within the dataset. Upon the initial dataset analysis, certain attributes were observed to have a low value count. Upon closer examination, it is evident that the parking options feature has the highest number of null values, followed by laundry options. Given that these features are categorical, filling them with the highest mode is not a suitable option due to the dataset's large and rich nature; therefore, removing them is more conducive to obtaining meaningful insights.

Another set of features, including description, latitude, and longitude, cannot be appropriately filled with values such as mean or median. Consequently, these records will also be removed. Figure 5 illustrates the count of null records present in the rental dataset.

Figure 5

Count of Null Records in Rental Dataset

As stated earlier, these records have been eliminated from the dataset. Figure 6 below illustrates the resolution of null values within the dataset.

```
print("The number of null records in the dataset:\n")
print(result)
```

```
The number of null records in the dataset:

                         count_nulls
id                                 0
url                                0
region                             0
region_url                         0
price                              0
type                               0
sqfeet                             0
beds                               0
baths                              0
cats_allowed                       0
dogs_allowed                       0
smoking_allowed                    0
wheelchair_access                  0
electric_vehicle_charge            0
comes_furnished                    0
laundry_options                54311
parking_options                95135
image_url                          0
description                        2
lat                             1419
long                            1419
state                              1
```

Figure 6

Count of Null Records after Data Pre-Processing

```
print(result)
print('\n All records with null are removed successfully!')
```

```
                        count_nulls
id                                0
url                               0
region                            0
region_url                        0
price                             0
type                              0
sqfeet                            0
beds                              0
baths                             0
cats_allowed                      0
dogs_allowed                      0
smoking_allowed                   0
wheelchair_access                 0
electric_vehicle_charge           0
comes_furnished                   0
laundry_options                   0
parking_options                   0
image_url                         0
description                       0
lat                               0
long                              0
state                             0

 All records with null are removed successfully!
```

3.2 Handling Redundant Data

The presence of data redundancy in the dataset can compromise data integrity, as having the same data multiple times can introduce inconsistencies in the obtained results. Duplicates in the data may arise from insertion, deletion, or updating anomalies caused by human error and therefore require attention. Figure 7 illustrates that there are no duplicates present in the dataset.

Figure 7

Number of Duplicate Records in the Dataset

```
print('Total Duplicates :',rental.duplicated().sum())
```

```
Total Duplicates : 0
```

3.3 Handling Data Type

The dataset comprises 22 attributes, and the data type is crucial for effective data analysis. Any issues related to data type must be addressed as it dictates how the computer interprets these values. It was imperative to handle these issues because they influence how the computer should

interpret these values. For instance, the "price" attribute, initially assigned as an integer, was updated to float. Features indicating amenities, such as "cats and dogs allowed", "electric charging available", "furnished status", and "smoking permission", originally had values of 0 or 1 represented as integers; therefore, these features were converted to Boolean.

Figures 8 and 9 illustrate the data types of attributes in the dataset before and after handling. Monitoring and adjusting the data types contribute to evaluating records and optimizing features for gaining deeper insights into patterns and trends.

Figure 8

Datatype of Attributes in the Rental Dataset

```
print("The Datatypes of the attributes: ")
rental_csv.dtypes
```

```
The Datatypes of the attributes:

id                          int64
url                        object
region                     object
region_url                 object
price                       int64
type                       object
sqfeet                      int64
beds                        int64
baths                     float64
cats_allowed                int64
dogs_allowed                int64
smoking_allowed             int64
wheelchair_access           int64
electric_vehicle_charge     int64
comes_furnished             int64
laundry_options            object
parking_options            object
image_url                  object
description                object
lat                       float64
long                      float64
state                      object
dtype: object
```

Figure 9

Datatype of Attributes in the Rental Dataset after Data Pre-Processing

```
print("New datatypes of the attributes : ")
rental.dtypes
```

```
New datatypes of the attributes :

id                           int64
url                         object
region                      object
region_url                  object
price                      float64
type                        object
sqfeet                       int64
beds                         int64
baths                      float64
cats_allowed                  bool
dogs_allowed                  bool
smoking_allowed               bool
wheelchair_access             bool
electric_vehicle_charge       bool
comes_furnished               bool
laundry_options             object
parking_options             object
image_url                   object
description                 object
lat                        float64
long                       float64
state                       object
dtype: object
```

The count of records following the completion of all pre-processing steps is presented in Figure 10 below. The data, now cleaned and prepared, is ready for further analysis and insights through the creation of graphs and charts.

Figure 10

Final Number of records after Data Pre-Processing

```
print('Total records in the dataset after pre-processing:')
rental.shape
```

```
Total records in the dataset after pre-processing:

(164283, 22)
```

4 Tableau Visualization Dashboard Overview

This segment includes various analyses conducted on the refined data post-preprocessing. The visualizations were created using Tableau 2023.2, aiming to effectively communicate information to users and stakeholders. Visualization plays a crucial role in uncovering trends and patterns within extensive datasets.

A total of 14 charts have been created to visualize and analyze trends. These charts are categorized into different dashboard categories based on the analysis. The categories are Area Information (Real Estate Statistics), Price Information (Rental Statistics), and Amenities Information (Amenities Distribution). Each of these category dashboards, along with their respective charts, is discussed in the following sections.

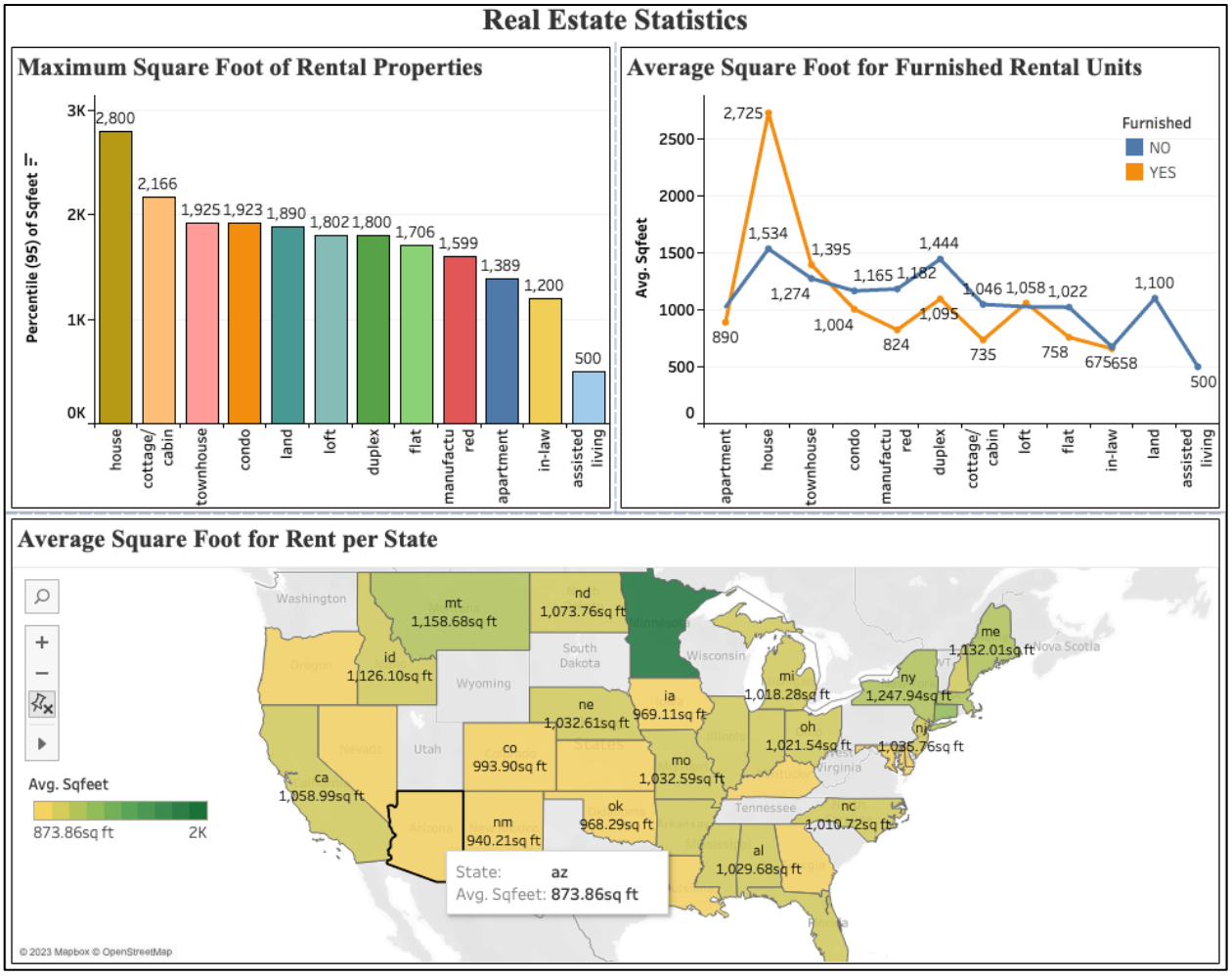
4.1 Area Information (Real Estate Statistics)

This dashboard focuses on the square footage analysis of various types of housing properties in the United States. The properties included in this dataset encompass houses, townhouses, land, condos, lots, cottages, flats, duplexes, assisted living, and apartments. Figure 11 displays all the charts within the dashboard. The existing charts include:

- **Average Square Foot for Rent Per State:** A geospatial chart illustrating the average square footage analysis of all property types. As seen in the chart we can see the average square foot available for rent of any property type in each of the regions respectively.
- **Maximum Square Foot of Rental Properties**: A bar graph displaying the maximum square footage across all available rental properties. This maximum square footage is for the 95th percentile of the property types. The 95th percentile indicates the value at which 95% of the property have a smaller area and only 5% have a larger area. As shown in the figure 11, 95% of houses in the dataset have square footage below 2800.
- **Average Square Foot for Furnished Rental Units:** A line chart comparing the average square footage of different property types based on their furnished status. As shown in the figure 11 we can see that furnished ‘house’ has higher average square foot available for rent compared to other furnished properties. Also, unfurnished ‘house’ and ‘duplex’ have approximately same average square foot available for rent.

Figure 11

Dashboard for Coverage Area Information



4.2 Price Information (Rental Statistics)

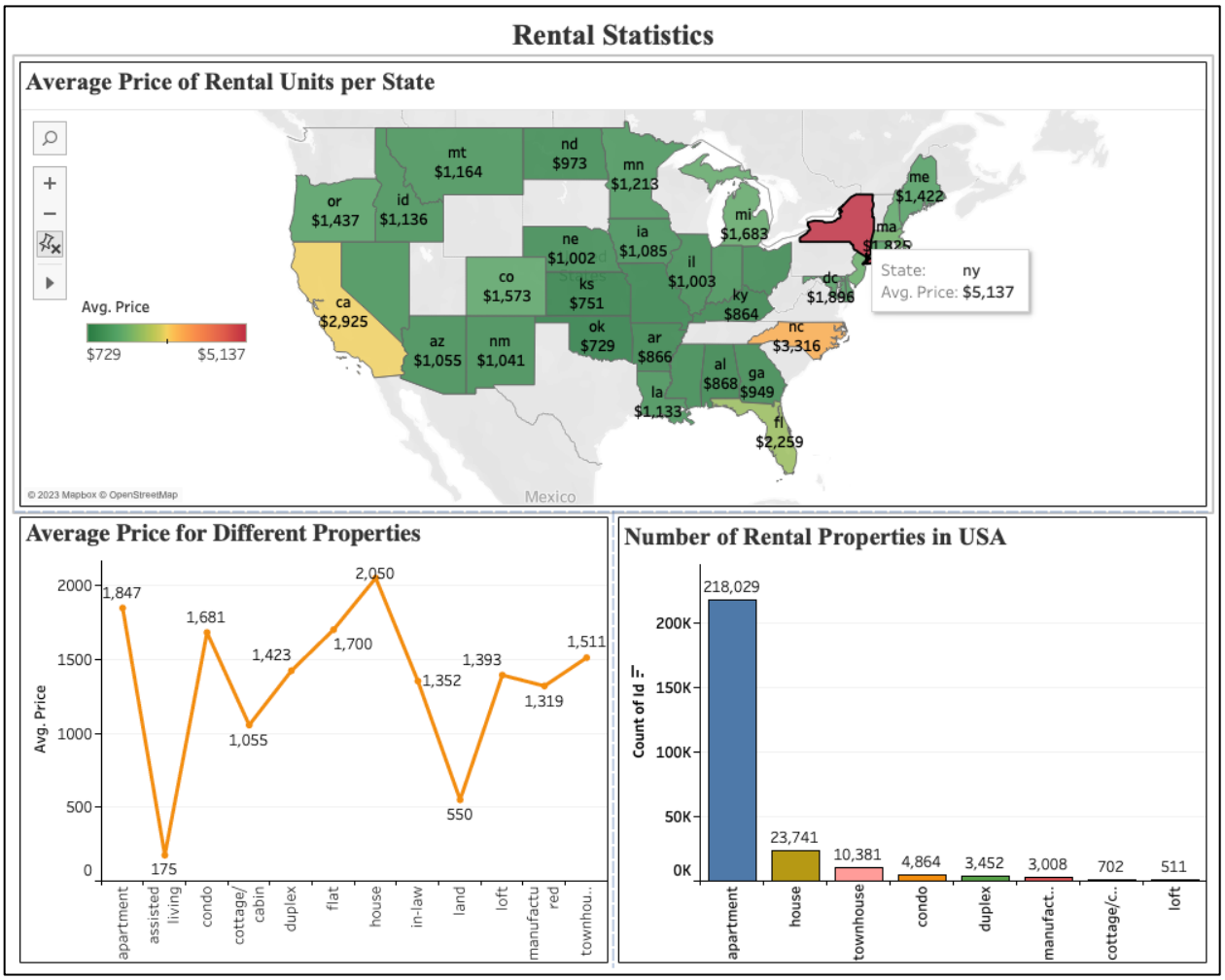
Following the previous dashboard, this visual display focuses on the price analysis of properties across the entire USA. For users, this provides an understanding of the budget associated with renting specific property types, potentially revealing additional avenues for investigation. These charts serve to convey information about the price ranges associated with various property types, aiding users or stakeholders in gaining valuable insights. Figure 12 illustrates the dashboard presenting rental statistics for the dataset. The featured charts include:

- **Average Price for Different Properties:** A line chart depicting the average price of different properties. From the chart we can depict that 'house' in general has higher average rental price compared to other property types.

- **Number of Rental Properties in USA:** A bar chart displaying the top 8 rental properties along with the corresponding number of available accommodations. ‘Apartment’ type of property has more accommodations available in comparison with other property types.
- **Average Price of Rental Units per State:** A geospatial chart illustrating the average rental price across the USA. As per the chart shown in the figure, we see that ‘New York’ state has very high average rental price compared to other regions.

Figure 12

Dashboard for Price Information



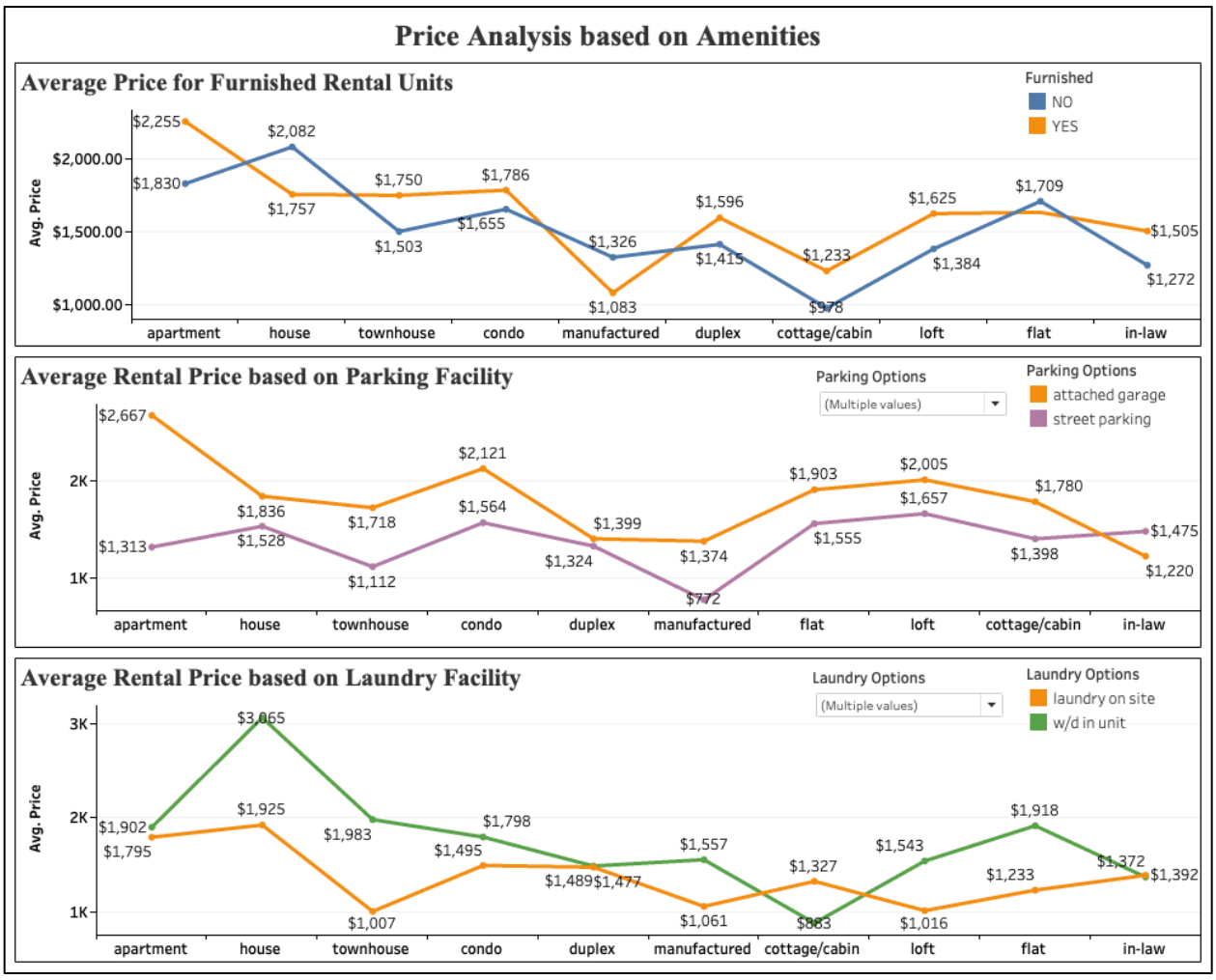
4.3 Price Analysis based on Amenities

Following the previous dashboard, this visual display focuses on the price analysis of properties across the entire USA based on Amenities. Figure 13 illustrates the dashboard presenting pricing analysis for properties based on amenities associated. The featured charts include:

- **Average Price for Furnished Rental Units**: A line chart depicting the average price of different properties based on furnished and unfurnished status of the rental units. From the chart we can depict that 'apartment' in general has higher average rental price compared to other property types followed by 'Condo' property type. Interestingly furnished 'apartment' and unfurnished 'house' properties have closer average rental price.
- **Average Rental Price based on Parking Facility:** A line chart depicting the average price of different properties based on parking facility of the rental units. An 'apartment' with attached garage has higher average rental price when compared to other property types with attached garage. Also 'apartment' with attached garage has higher average rental price and the price drastically drops to half of it when it is associated with street parking facility.
- **Average Rental Price based on Laundry Facility:** A line chart depicting the average price of different properties based on laundry facility of the rental units. 'house' property type has higher average rental price if it has in unit laundry when compared to other property types.

Figure 13

Dashboard for Price Analysis based on Amenities



4.4 Amenities Information (Amenities Distribution)

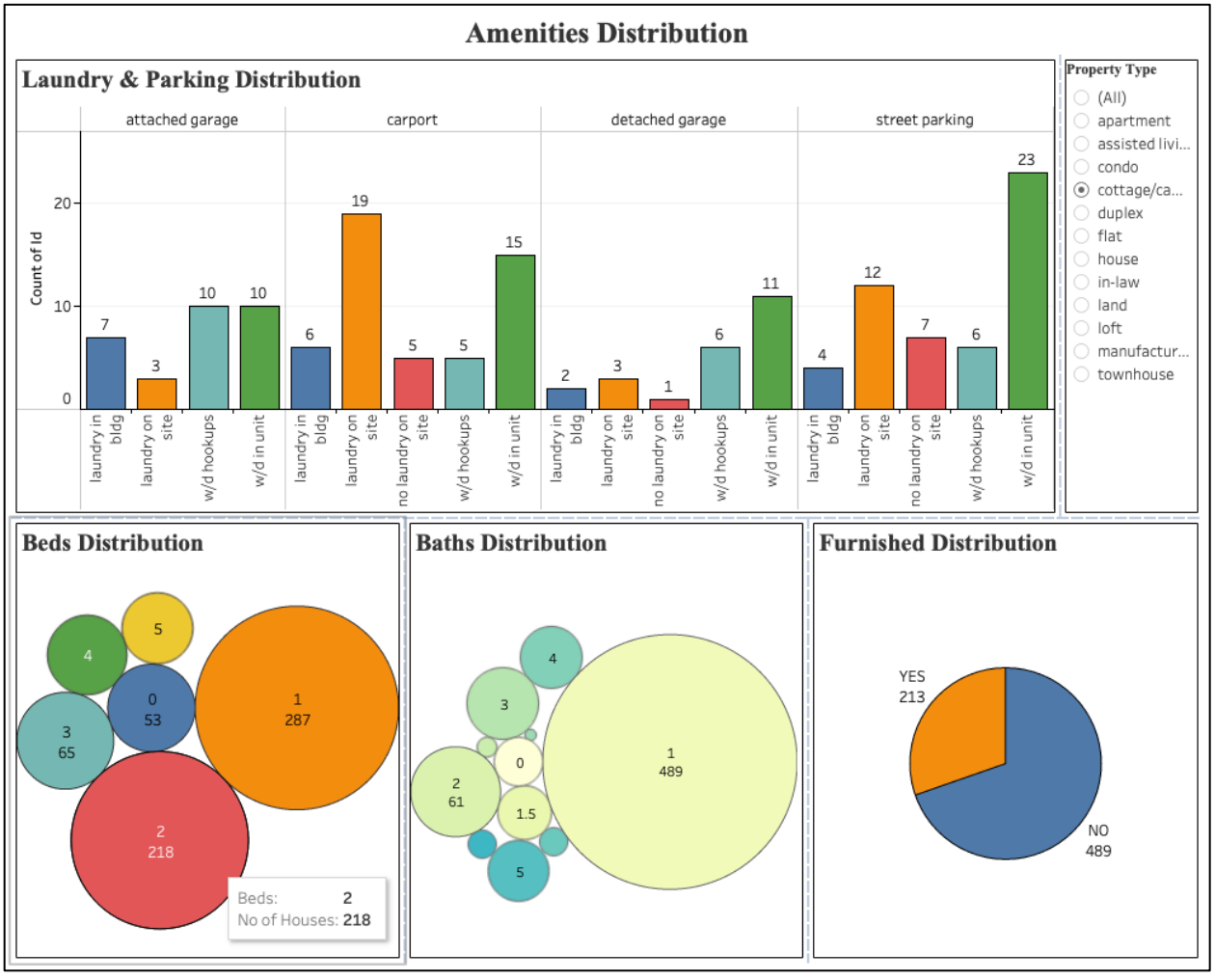
This dashboard addresses the various amenities found in the majority of housing properties. Through this visual display, information about commonly available amenities across all properties becomes accessible, aiding in understanding prevalent features. It also assists in narrowing down searches to specific areas based on desired features. Figure 14, presented below, showcases the amenities dashboard developed as part of the housing rental analysis. The featured charts include:

- **Baths Distribution:** Bubble charts displaying the distribution of bathrooms among the properties. From the figure below we can see that there are 489 properties with 1 bathroom, 61 units with 2 bathrooms and so forth based on the property type filtered for ‘Cottage’.

- **Beds Distribution:** Bubble charts displaying the distribution of beds among the properties. From the figure below we can see that there are 287properties with 1 bedroom, 218 units with 2 bedrooms and so forth based on the property type filtered for ‘Cottage’.
- **Laundry & Parking Distribution:** A bar chart illustrating laundry and parking information for the property type ‘cottage’
- **Furnished Distribution:** A pie chart depicting the number of furnished properties and unfurnished properties based on property type ‘Cottage’. 213 cottage units are furnished whereas 489 cottage units are unfurnished.

Figure 14

Dashboard for Amenities Information

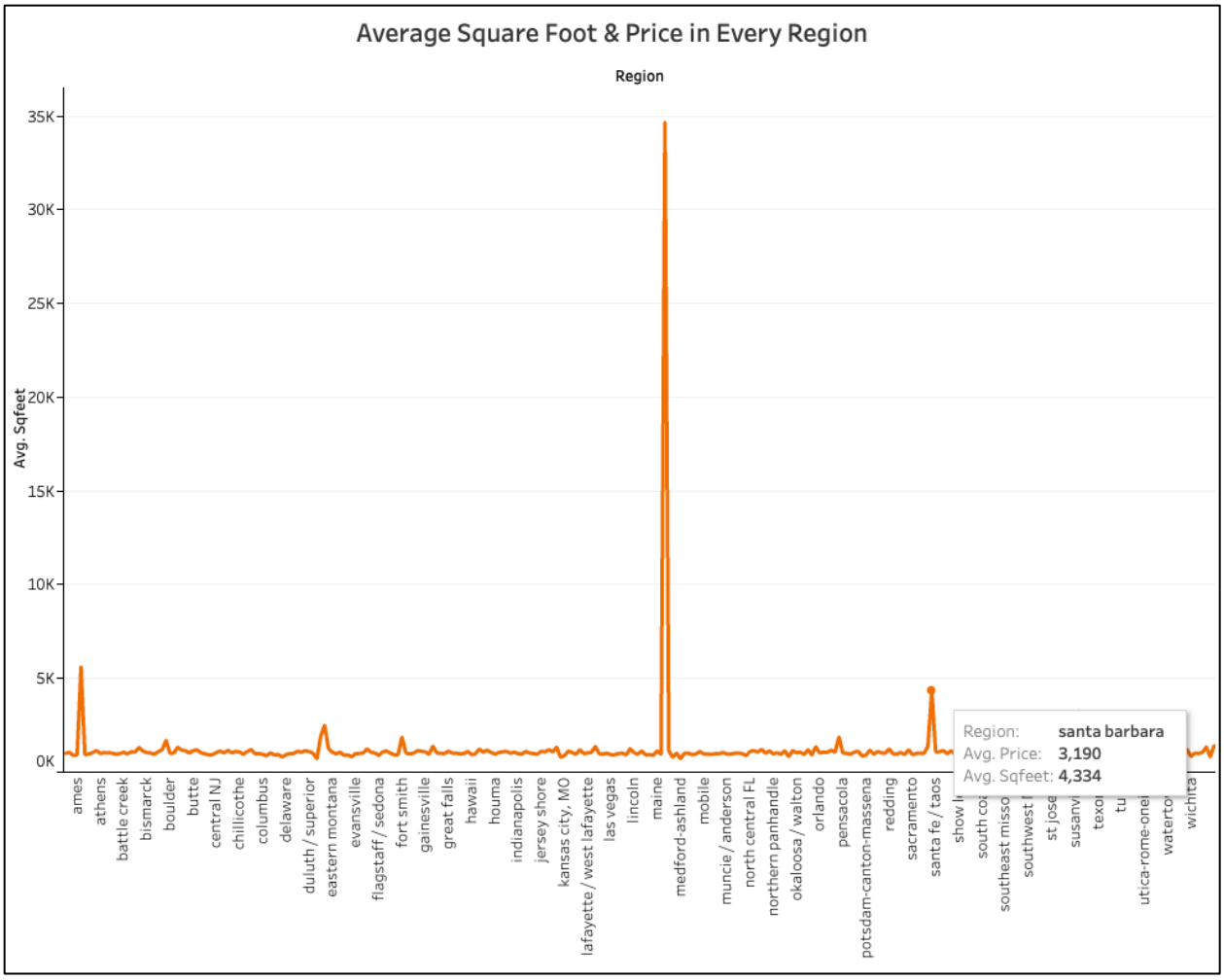


4.5 Regional Information (Real Estate & Rental Statistics)

Figure 15 displays the average square footage available and the average price of properties in each region across the USA. This chart offers a dual analysis encompassing both real estate and rental statistics. As seen in below figure 'Santa Barbara' has an average rental price of $3,190 and average square foot available for rent is 4,334 sq ft.

Figure 15

Average Square Foot and Price of Property in Every Region



5 Summary

The project aimed to empower users and stakeholders by providing comprehensive visualizations for obtaining detailed information on housing prices and other property characteristics. Users can leverage these visualizations to acquire insights into the current rental prices for specific amenities they are seeking. Meanwhile, stakeholders can use this information to identify strategic investment opportunities for optimal returns.

Key insights derived from these visualizations include the prevalence of apartments, which are both numerous and relatively expensive in terms of rent. Additionally, homes in most states tend to have larger square footage compared to other property types. The presence of amenities varies across regions, indicating regional preferences. Overall, the bedroom-to-bathroom ratio in all listed properties is generally favorable.

Four dashboards were generated, accompanied by a total of 14 charts, to visually represent the data and facilitate the extraction of information related to housing rentals. The project successfully facilitated user exploration, making it easy to derive valuable observations. The incorporation of interactive features further enabled effortless discoveries with minimal effort.

GitHub Link:

https://github.com/ShilpaShivarudraiah/DATA230_Final_Project