INFO-523 FINAL PROJECT
Rental Price Estimation

Team - Group-9

Motivation

- There is a growing need for predictive models that can accurately estimate rental prices.

- Predicting rental price has significant effects: for property Owners, for renters, for real estate platform.

About Data

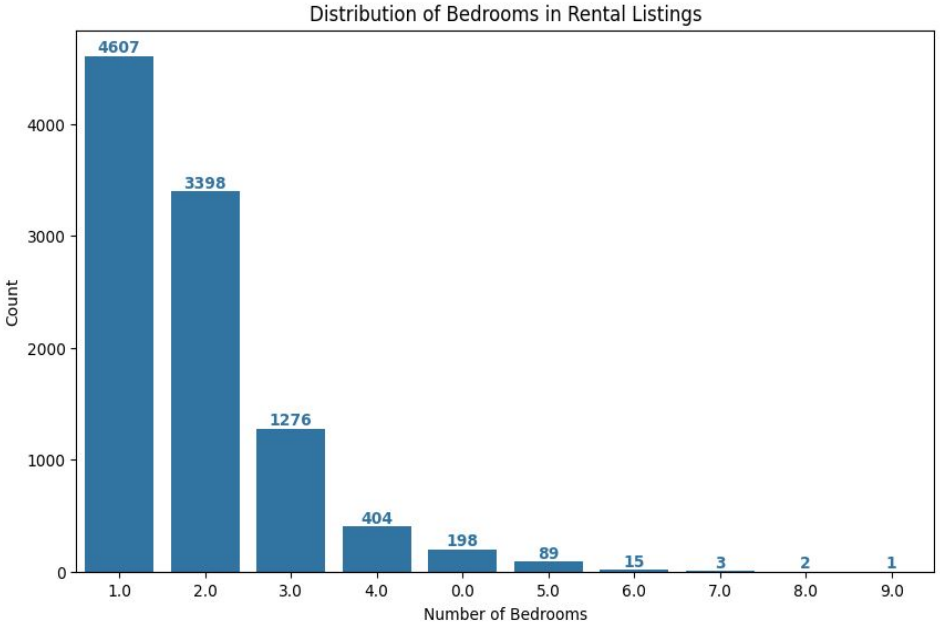
- The dataset for this project is sourced from kaggle which is from the year 2022
- The dataset contains around **10,000** records about the rental prices details
- Following are the parameters:

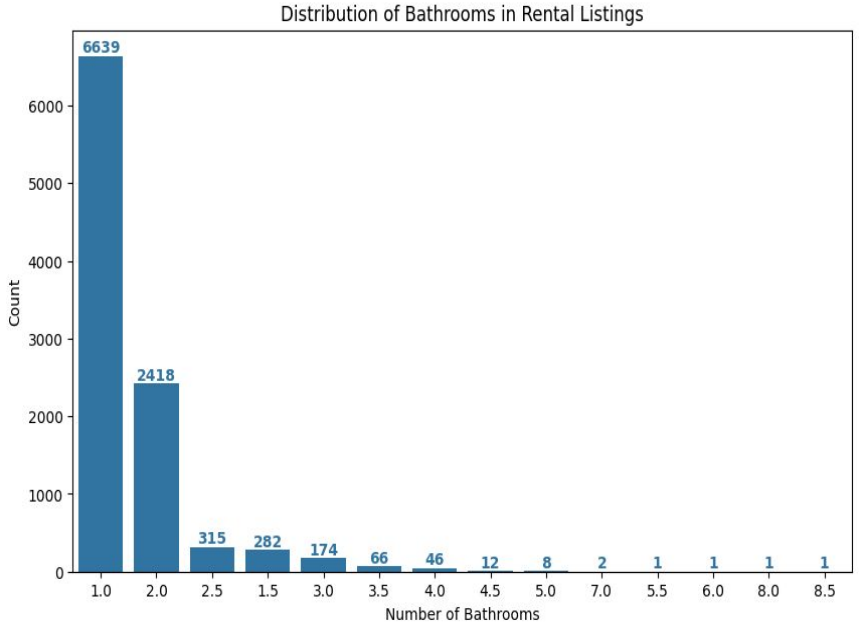
  id', 'category', 'title', 'body', 'amenities', 'bathrooms', 'bedrooms','currency', 'fee', 'has_photo', 'pets_allowed', 'price','price_display', 'price_type', 'square_feet', 'address', 'city name', 'state', 'latitude', 'longitude', 'source', 'time'

Data Pre-Processing

- In the first step handled the null values by replacing the column “amenities” and “Pets_allowed” with specific string values and dropped the columns which are empty.
- Replaced the bedroom and bathrooms columns with median values.
- Rest of the features missing values rows are removed.
- The Title and Body columns do not contribute significantly to predicting rental prices so can be removed from the dataset.

Data Pre-Processing





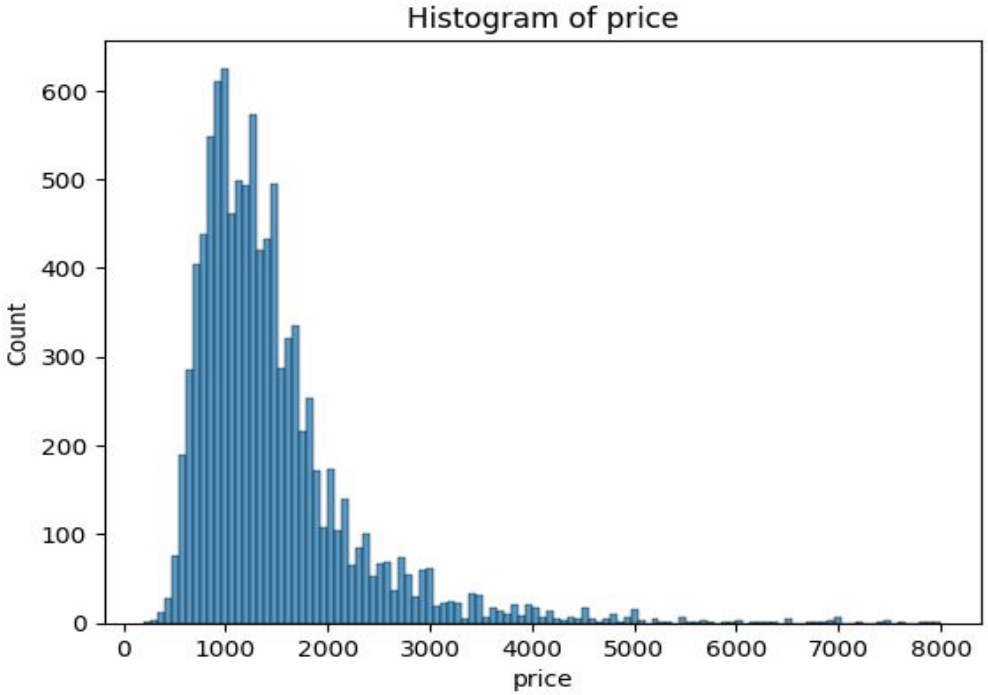
Rows with bedrooms greater than 6 and bathrooms greater than 5 are removed.

Exploratory Data Analysis

Most Common Price: The majority of houses are rented for around $1200 per month.

Price Skew: The distribution of rental prices is skewed to the left. This means there are a few very expensive houses, but most are priced lower.

Outliers: There are some houses with significantly higher rental prices, reaching up to $6000 per month. These are considered outliers as they are much higher than the typical price.



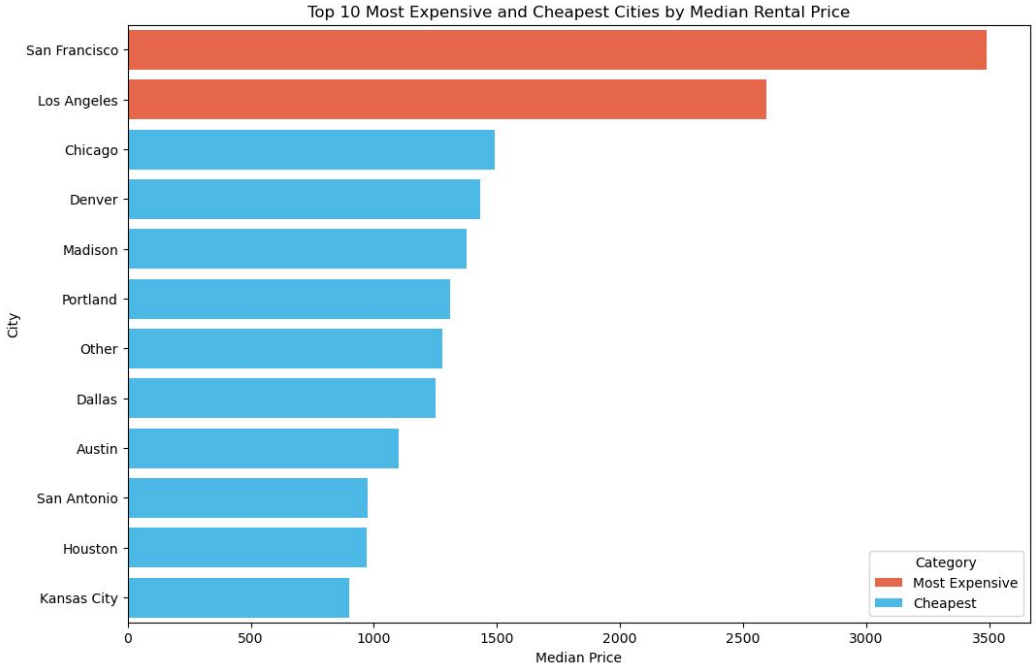
Exploratory data Analysis

Most Expensive Cities:

The cities categorized as the most expensive have significantly higher median rental prices compared to the cheapest cities.

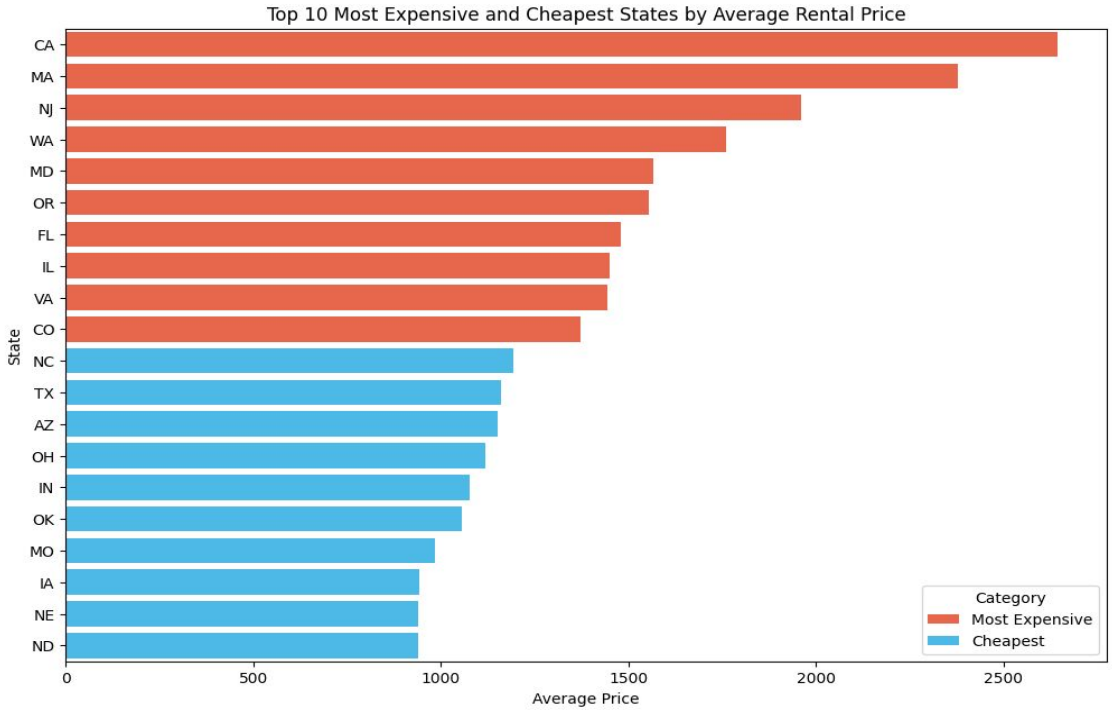
Cheapest Cities:

The cities in the cheapest category have notably lower median rental prices, indicating lower housing demand or ample housing supply.

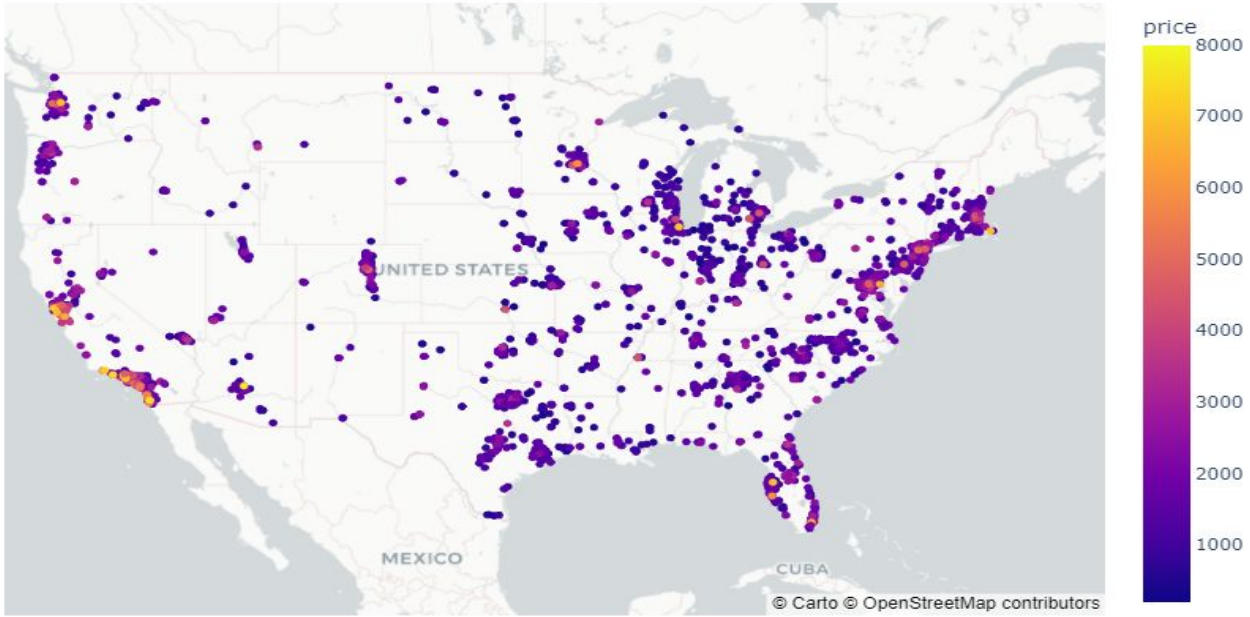


EDA

This visualization helps to identify the cities with the highest and lowest average rental prices, making it easier to compare affordability across different locations.



EDA

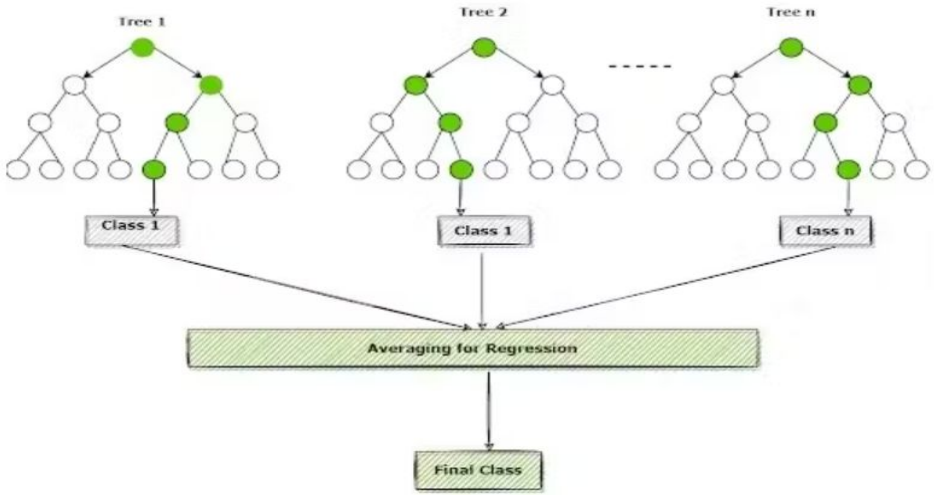


Scatter plot of house rental prices across United states

ALGORITHMS

1. Random Forest algorithm is used for this project to predict the price and then the model is evaluated.

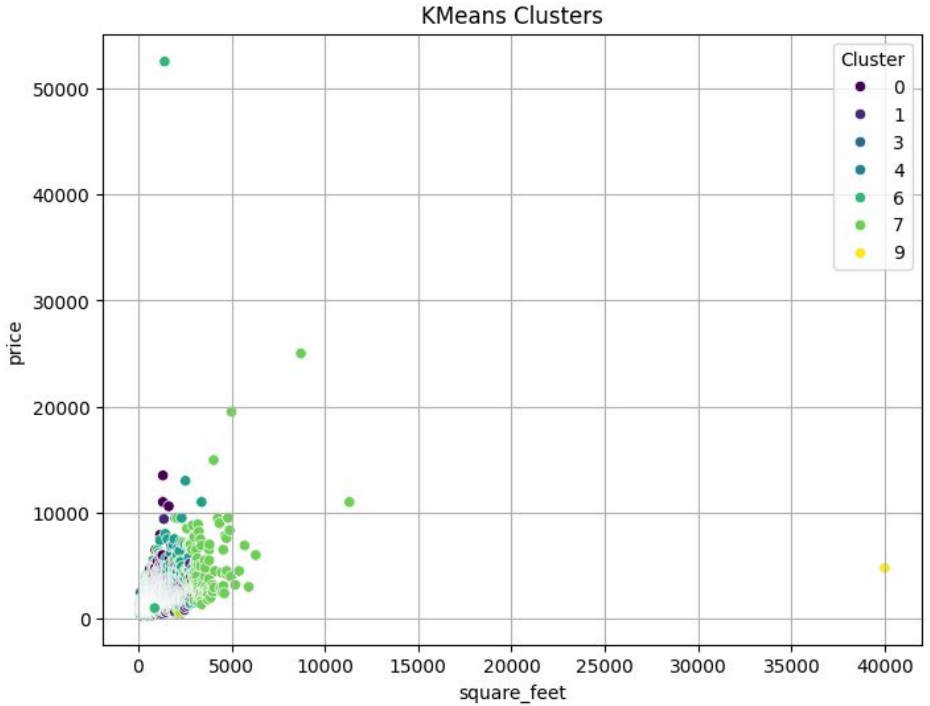
   Results: **Accuracy: 79%**



ALGORITHMS

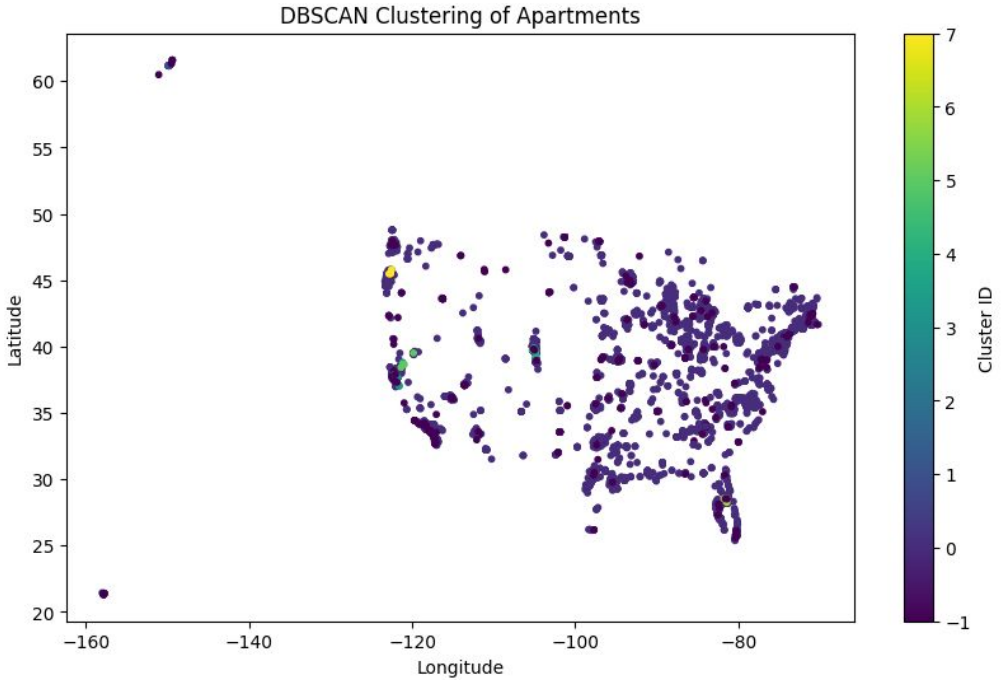
2. Data points are assigned to clusters according to their proximity to a cluster center using the K-Means technique, which finds groups within unlabeled data.

Result: **Silhouette Score - 0.3779**



3. DBSCAN - Density-Based Spatial Clustering of Applications with Noise. A machine learning algorithm that groups data points into clusters based on their density.

Result: **Silhouette score - 0..3479**



CONCLUSION

- House rent prediction system using machine learning has shown great results in accurately predicting the rental prices of properties across US metropolitan cities.

- By using machine learning algorithms such as Random Forest Classifier , K-means , and DBSCAN the system is able to take into account various factors that affect rental prices, such as location, property type, amenities.

- Overall, the house rent prediction system has the potential to change the rental market in US metropolitan cities by providing more transparency and accuracy in rental pricing.

THANK YOU