RentWise

Predicting Rental Property Prices in US

Made by: Kartik

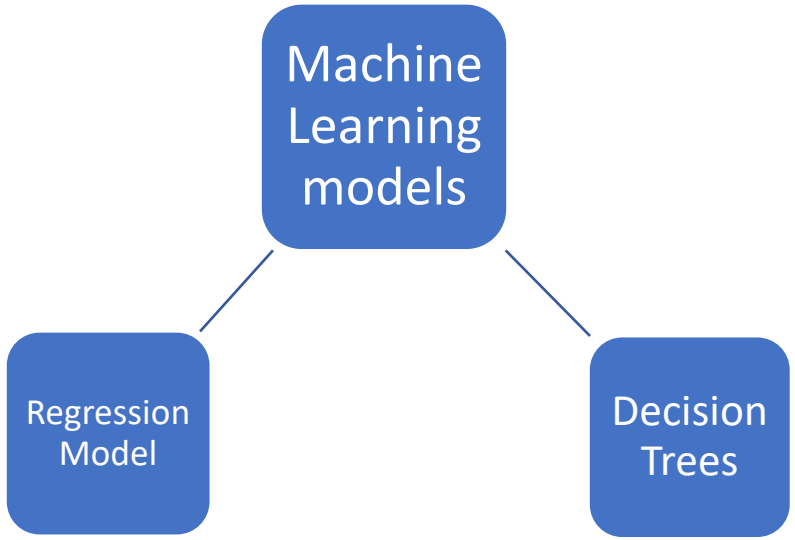
Real Estate Industry

- Massive market.
- Prices have sky-rocketed in the last decade.
- Renting over Buying.
- But the problem is, Rents have also gone up.
- People are struggling to find affordable rentals out there.
- Like Me!!

My solution: RentWise

I hope to use my knowledge in Data Science and Machine learning to develop a robust predictive model that can accurately estimate these rental prices, offering some insights for:

- renters,
- landlords or
- the ever-evolving real estate industry.



Dataset Introduction

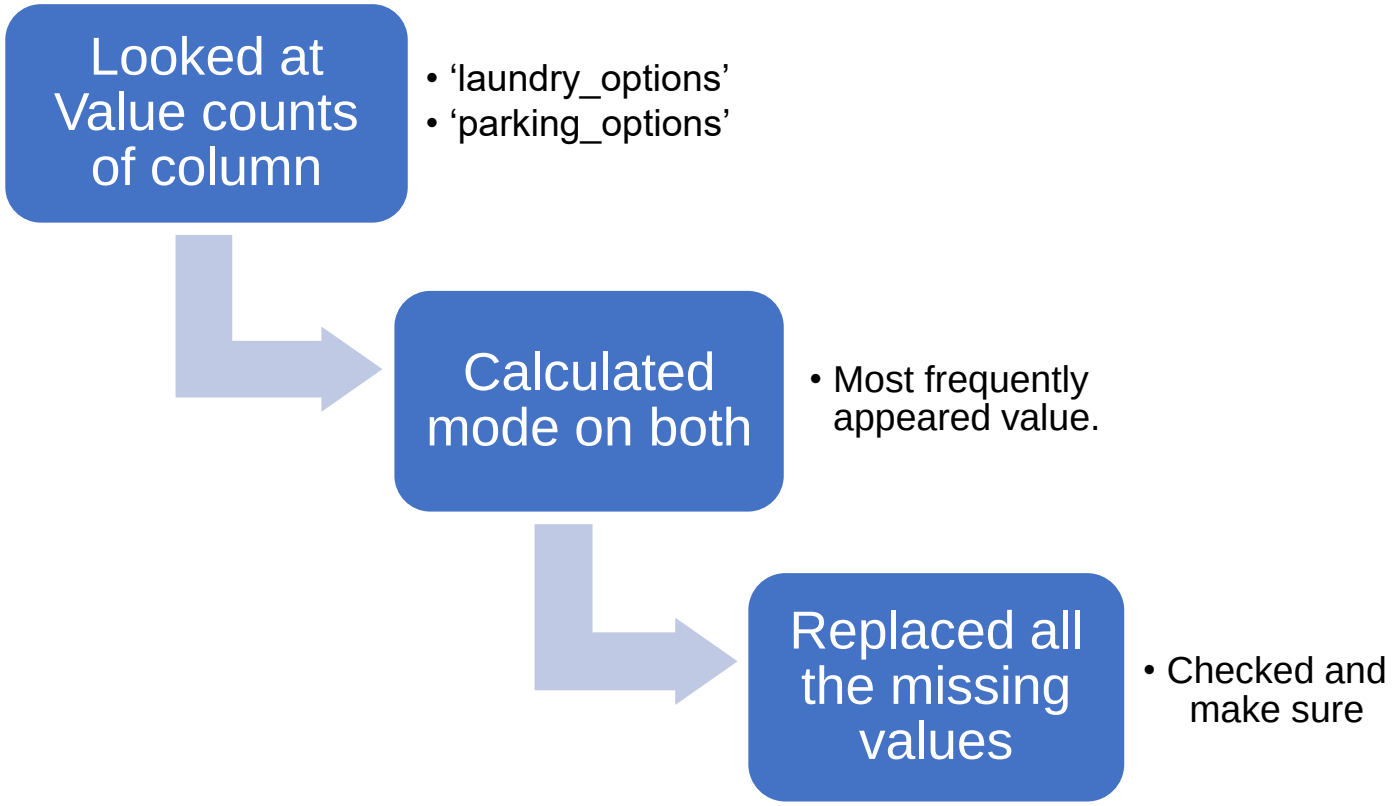
- **Source:** Kaggle (www.kaggle.com)
- **No. of rows:** 384,977
- Each row depicting a **unique data entry** (rental property).
- **No Duplicated** rows
- **No. of columns:** 22
- **Different types of datatypes**: int64, float64 and objects.
- Had a lot of missing values in:
    1. **'laundry_options':** More than 20% (79,026)
    2. **'parking_options':** More than 36.5% (140,687)
    3. **'lat':** 0.49% (1,918)
    4. **'long':** 0.49% (1,918)

Attributes:

- **Id**: Listing id
- **url**: Listing URL
- **region**: Craigslist region
- **region_url**: Craigslist region URL
- **price**: Rent per month (Target Column)
- **type**: Housing type
- **sqfeet**: Total square footage
- **beds**:Number of Beds
- **baths**:Number of Bathrooms
- **cats_allowed**: Cats allowed boolean (1 = yes, 0 = no)
- **dogs_allowed**: Dogs allowed boolean (1 = yes, 0 = no)
- **smoking_allowed**: Smoking allowed boolean (1 = yes, 0 = no)
- **wheelchair_access**: Has wheelchair access boolean (1 = yes, 0 = no)
- **electric_vehicle_charge**: Has electric vehicle charger boolean (1 = yes, 0 = no)
- **comes_furnished**: Comes with furniture boolean (1 = yes, 0 = no)
- **laundry_options**: Laundry options available
- **parking_options**: Parking options available
- **image_url**: URL of the image
- **description**: Description by poster
- **lat**: Latitude
- **long**: Longitude
- **state**: State of listing

Data Cleaning Steps

Dealing with Missing values



Dropping columns

1. **'lat' and 'long':** Because they have such low percentage of missing values, we could safely drop them.

2. **'id':** was a unique identifier for each data point, providing no real value.

3. **'url', 'region_url' and 'image_url':** They all included image links, but a lot of them were similar, so we dropped them,

Exploratory Data Analysis (EDA)

With Outliers

Distribution of ‘price’

min value: 0

Max value: 2,768,307,249.00

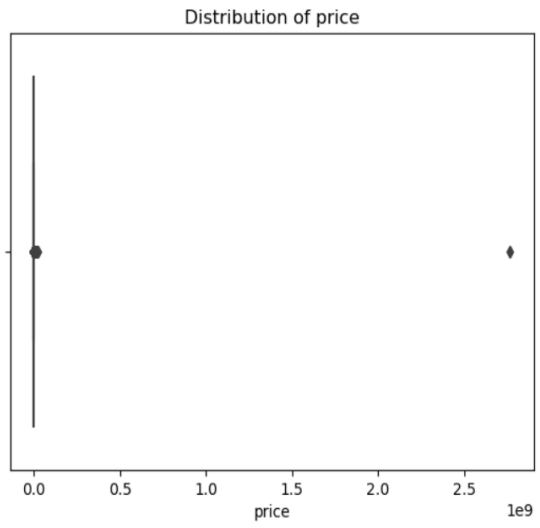
Mean: 8,897.79

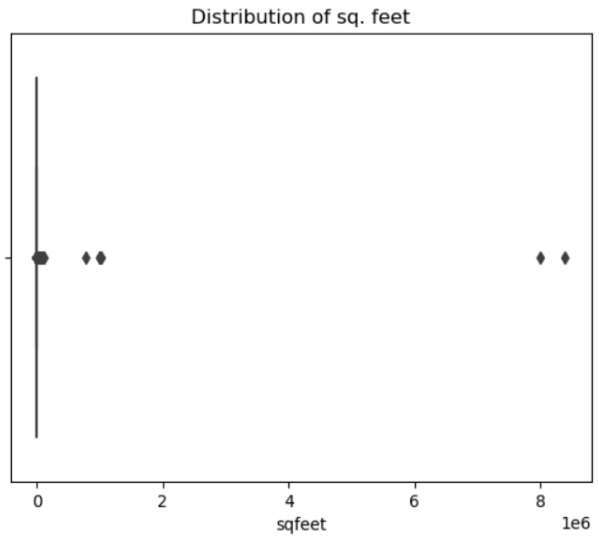
Distribution of ‘sqfeet’

Min value: 0

Max value: 8,388,607.00

Mean: 1,062.35





Without Outliers

Redistribution of ‘price’

Min value: 100

Max value: 5000

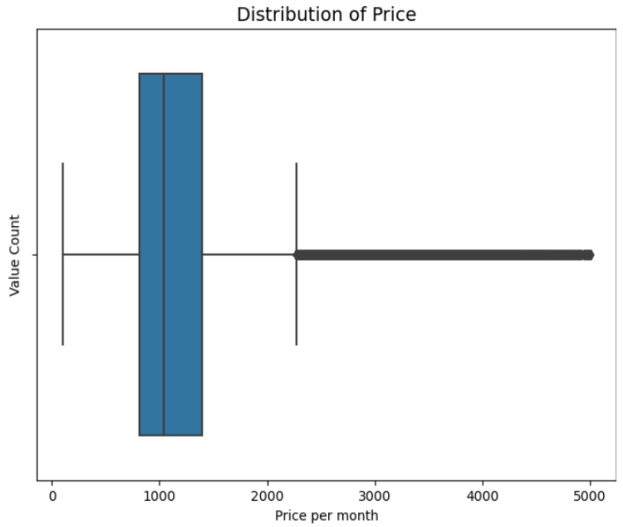
Mean: 1177.46

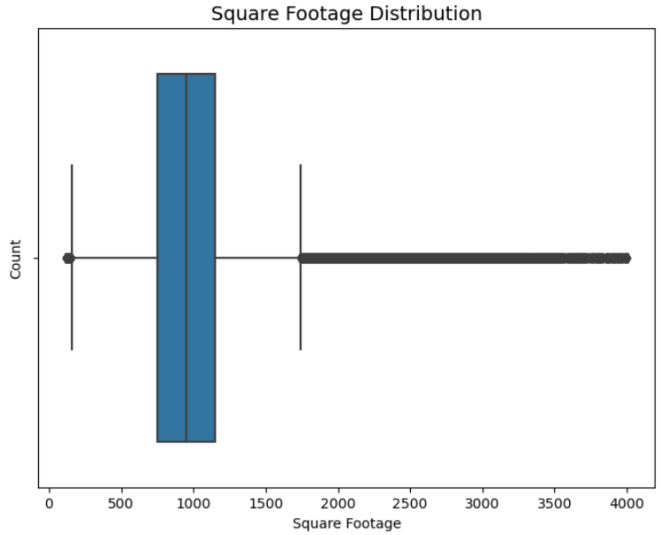
Redistribution of ‘sqfeet’

Min value: 120

Max value: 4000

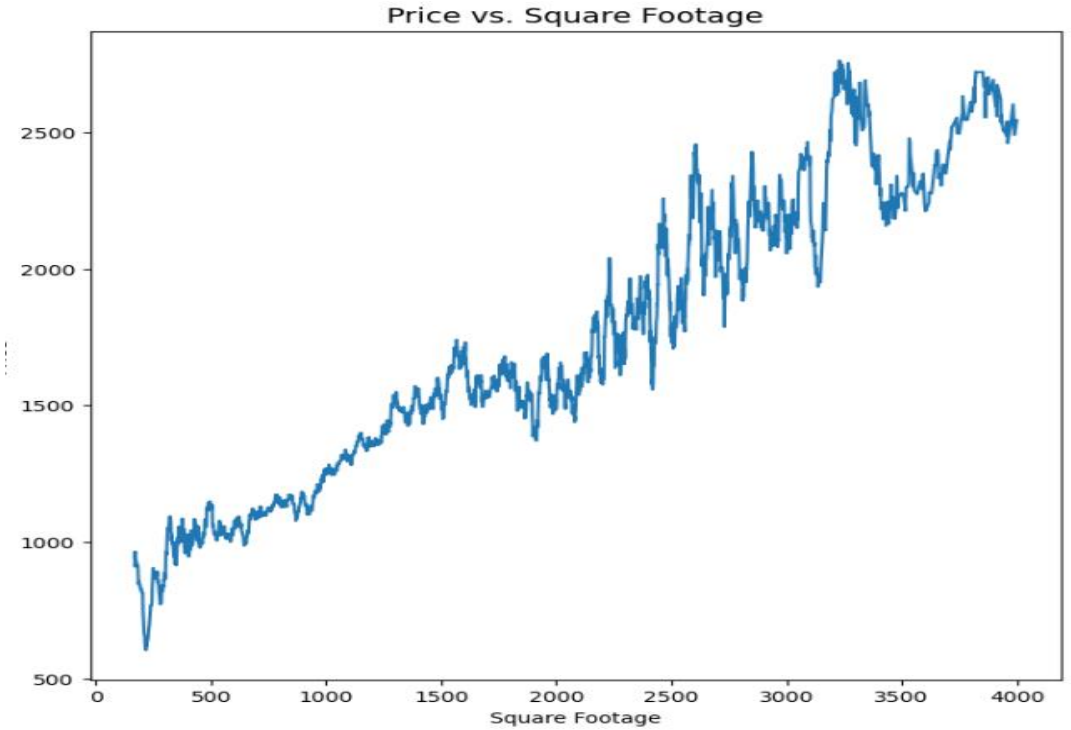
Mean: 989.75



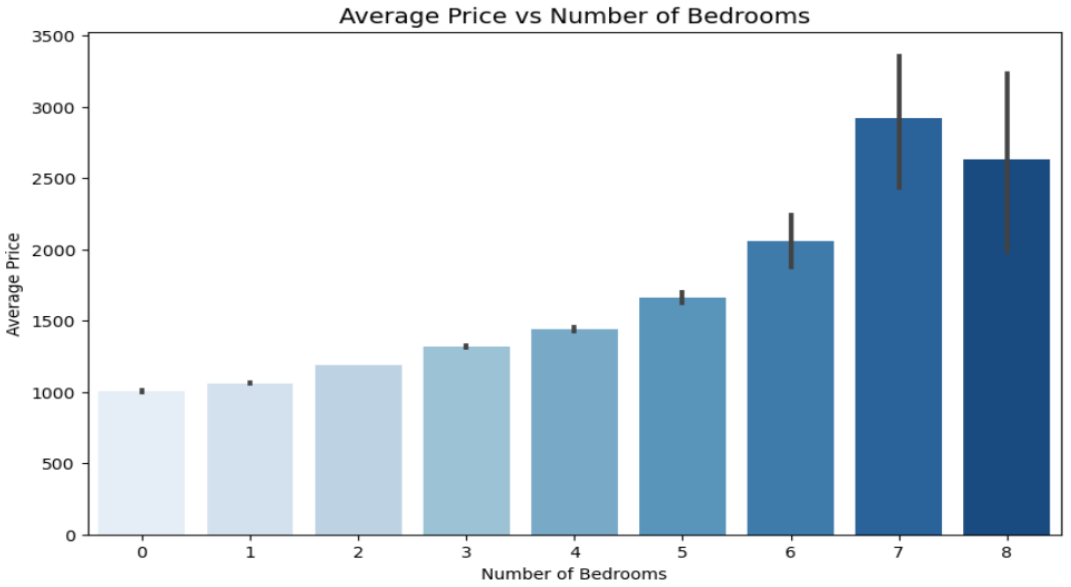


Preliminary Findings

- **Price** and **sqfeet** are positively correlated. Bigger the place, more expensive it is.



- **Price** is also positively correlated with the **number of bedrooms**. Higher number of rooms means place would be more expensive.



Here, we can see the last one, despite having more bedrooms was less expensive than the seven bedrooms, this could be because of outside factors like state, region, condition of the building.

Next steps

- **Data Processing:** My next steps involve data transformation and processing to prepare the dataset for machine learning.
- **Feature Engineering:** I will be refining some features, ensuring they are meaningful and suitable for machine learning.
- **Baseline Modeling:** My project roadmap includes selecting, training, and evaluating machine learning models. I plan to test various models, including Logistic Regression and Decision Trees.

Thank you !!!