REAL ESTATE MARKET IN BAY AREA

Presented By:
Group 7

Navya Areti
Sai Charan Reddy
Harshitha Ananthula
Sahithi Priya
Sai Teja Koya
Srinath Kethavath
Janardhan Reddy
Abhishek Malla

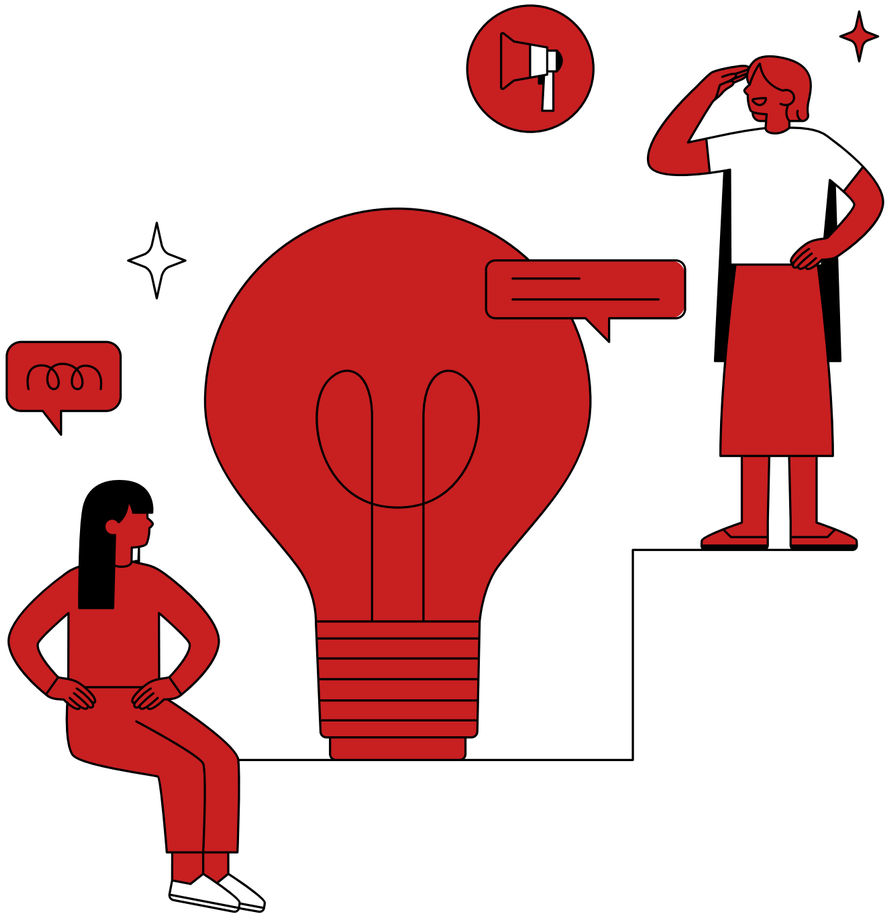

REDFIN

Introduction

-We'll explore how we transformed raw data into valuable insights, performed data cleaning, and applied machine learning techniques to uncover patterns and relationships within the data.

-2,500+ rows of data was scraped from Redfin website for latest dataset analysis.

-The dataset ('redfin_data.csv') was used to explore and analyze real estate property data.



Data Exploration

-Imported libraries include: NumPy, pandas, seaborn, matplotlib and scikit-learn.
-These libraries were used for data manipulation, visualization, and machine learning.
-The dataset was loaded and its shape was examined.
-The dataset contains 562 rows and 30 columns.

Data Cleaning

Columns Cleaning

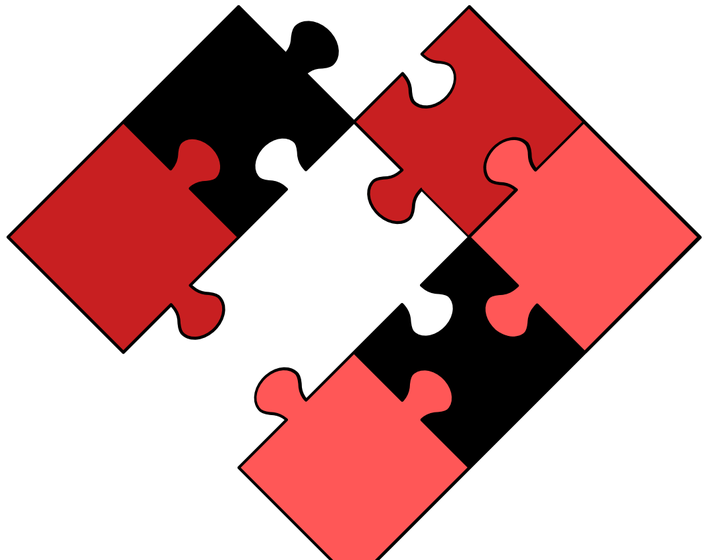
-Few Columns were removed as they were deemed unnecessary for analysis.

-The remaining columns were then renamed for clarity.

	SALE TYPE	PROPERTY TYPE	PRICE	BEDS	BATHS	LOCATION	SQUARE FEET	LOT SIZE	YEAR BUILT	DAYS ON MARKET	WALKSCORE	TRANSITSCORE	BIKESCORE
0	MLS Listing	Single Family Residential	1750000	5.0	3.5	Novato	2934.0	5456.0	2004.0	12	49.0	37.0	56.0
1	MLS Listing	Single Family Residential	1299000	4.0	3.5	Novato	2126.0	4125.0	1998.0	29	39.0	34.0	40.0
2	MLS Listing	Single Family Residential	1199000	3.0	3.0	San Rafael	1977.0	7209.0	1958.0	1	70.0	29.0	46.0
3	MLS Listing	Townhouse	575000	3.0	1.5	San Rafael	1323.0	3598.0	1973.0	1	75.0	42.0	64.0
4	MLS Listing	Vacant Land	165000	NaN	NaN	San Rafael	NaN	8112.0	NaN	1	58.0	50.0	54.0

Addressing Missing Data

- Regular expressions were used to extract and replace non-numeric values in the 'price', 'bedrooms', 'bathrooms', and 'size' columns.
- The 'address' column was converted to a categorical data type.
- Null values were also dropped.
- Numeric columns 'bedrooms' and 'size' were converted to integers for more effective analysis.

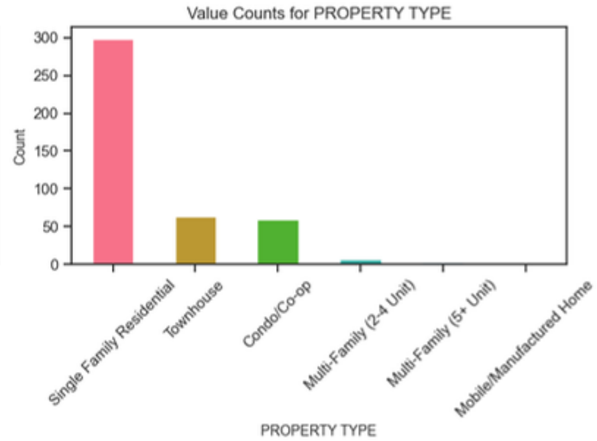


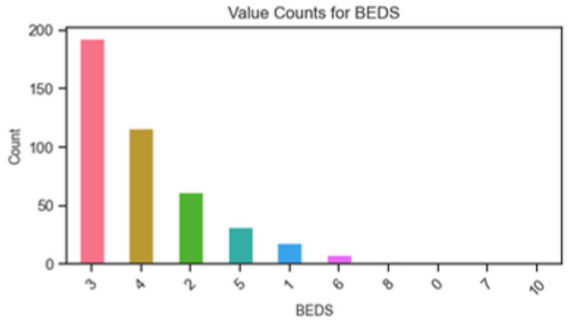
Data Analysis

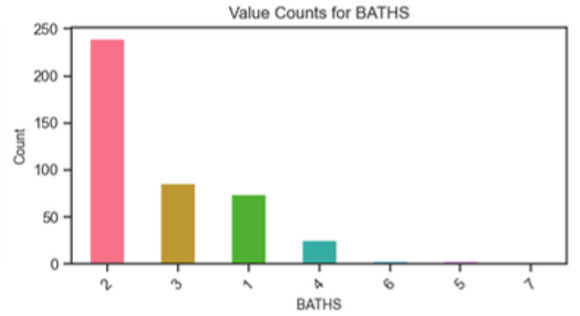
Bedroom and Bathroom Distribution

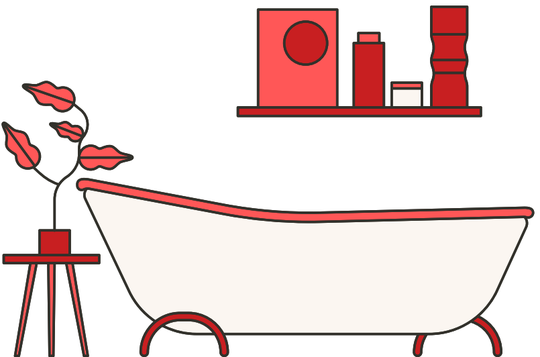
-A bar chart shows the distribution of the number of bedrooms in the dataset.
-Most houses were 3 and 4 bedroomed while the least were 10 and 8 bedroomed

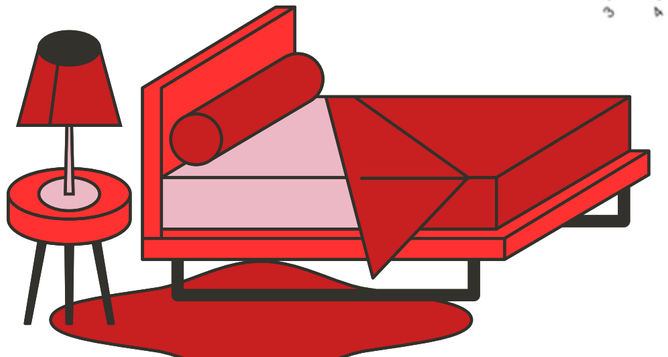
-A bar chart displays the distribution of the number of bathrooms in the dataset.
-Most houses had 2 and 1 bathrooms while the least houses had 4,5 and 6 full bathrooms.





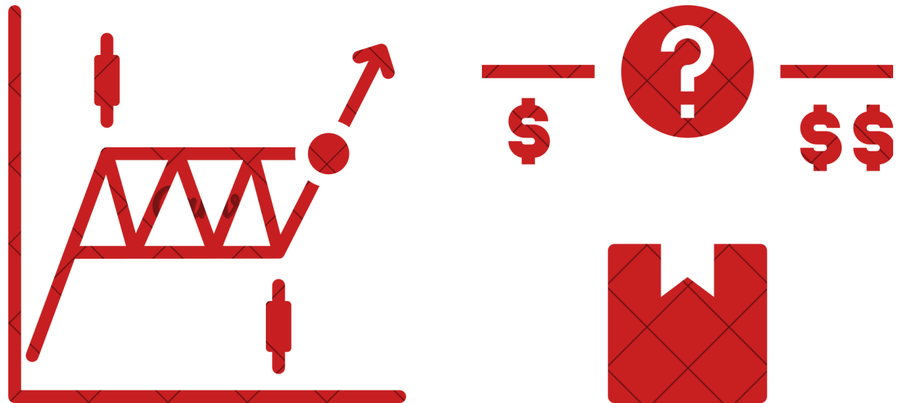


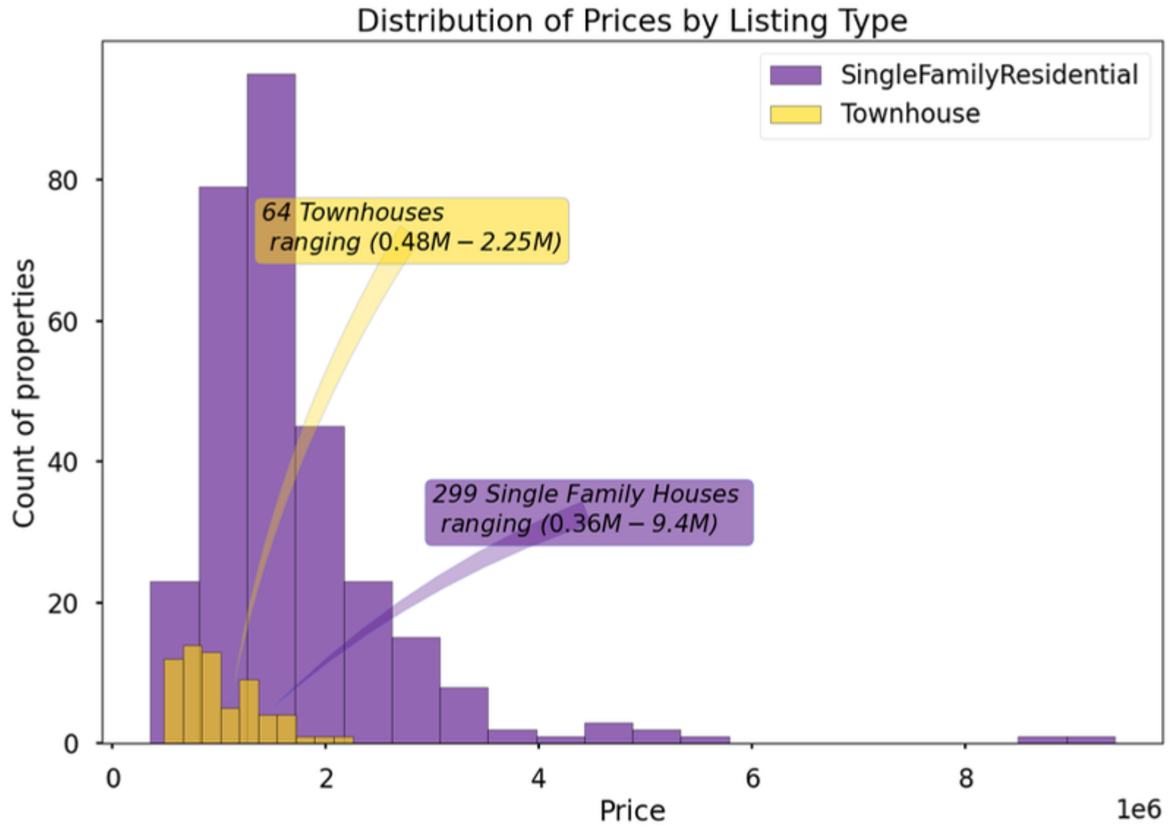




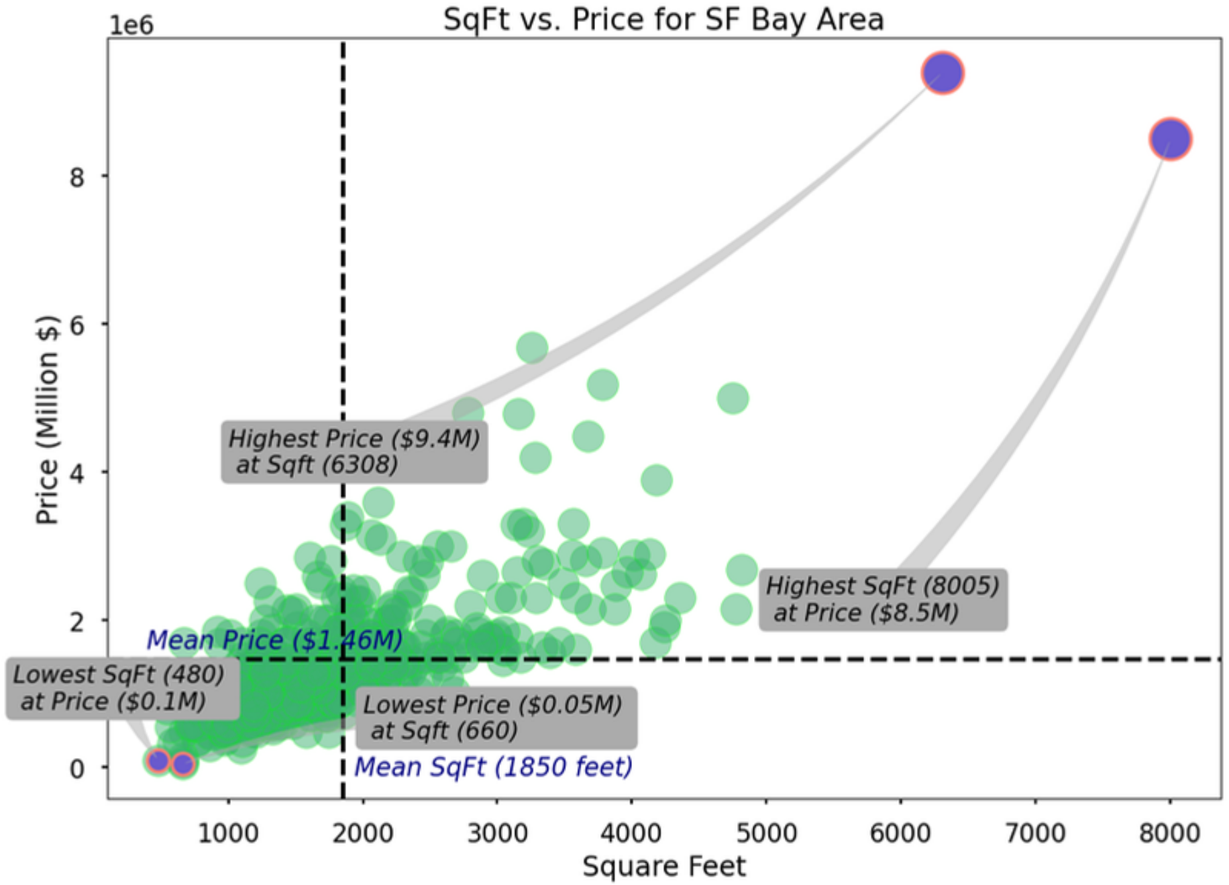
What is the price range of different types of properties?

1.The histogram provides a visual representation of how property prices are distributed for each type of listing.
2.Each type of listing (Single Family, Condo, Townhouse, Multi Family, Manufactured) has its own color and is represented by a separate histogram.





How much space do we get for the lowest price, how much for the higher price and sqft at highest price and sqft at lowest price available.

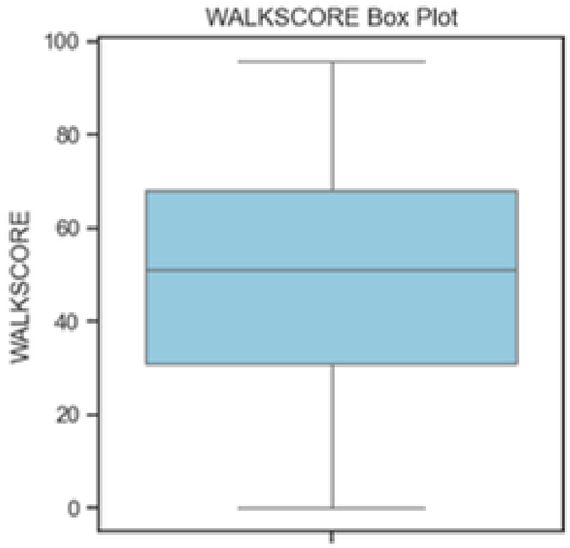


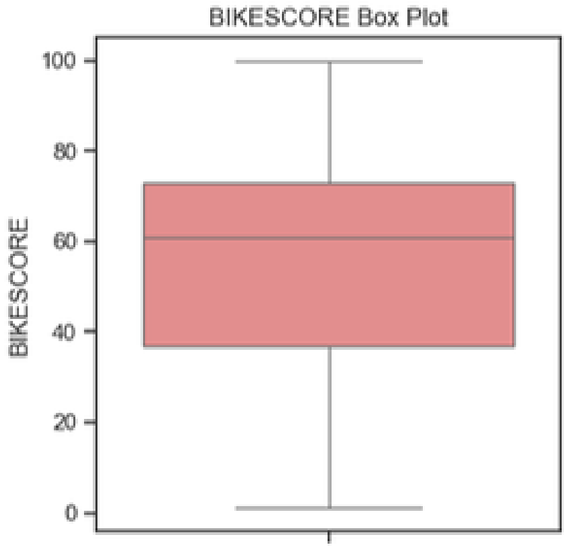
Relationship Between Bedrooms and Price

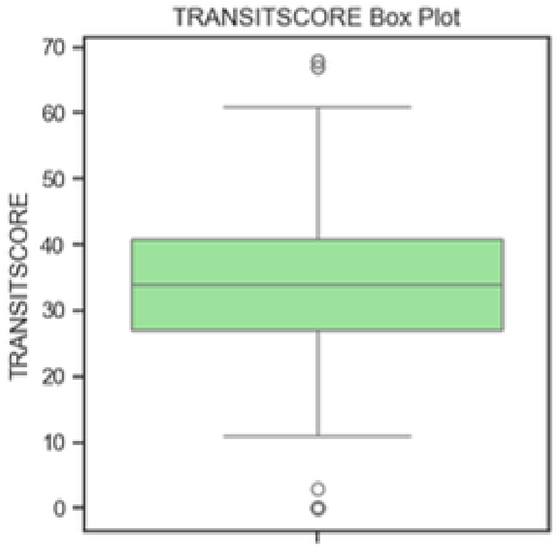
-How does the number of bedrooms influence property prices in the San Francisco Bay Area?

-We explored the pair plot with data points colored by the number of bedrooms.

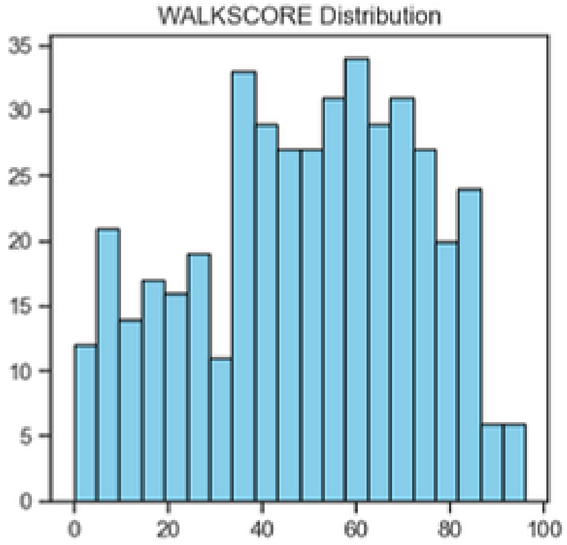
Boxplot for Correlation between Price range and Walk Score,Bike Score and Transitscore

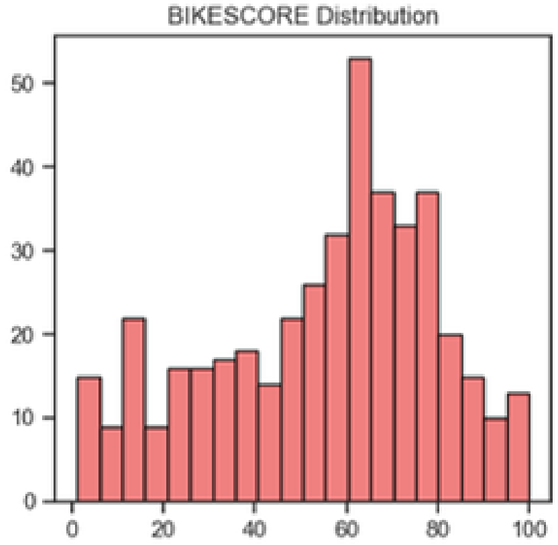


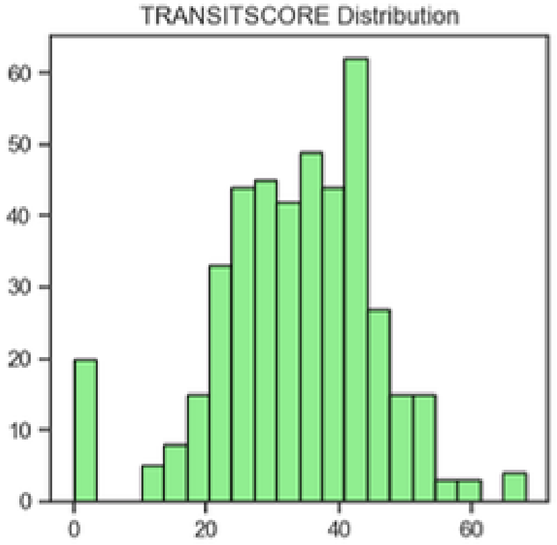




Correlation between Price range and Bike Score rating



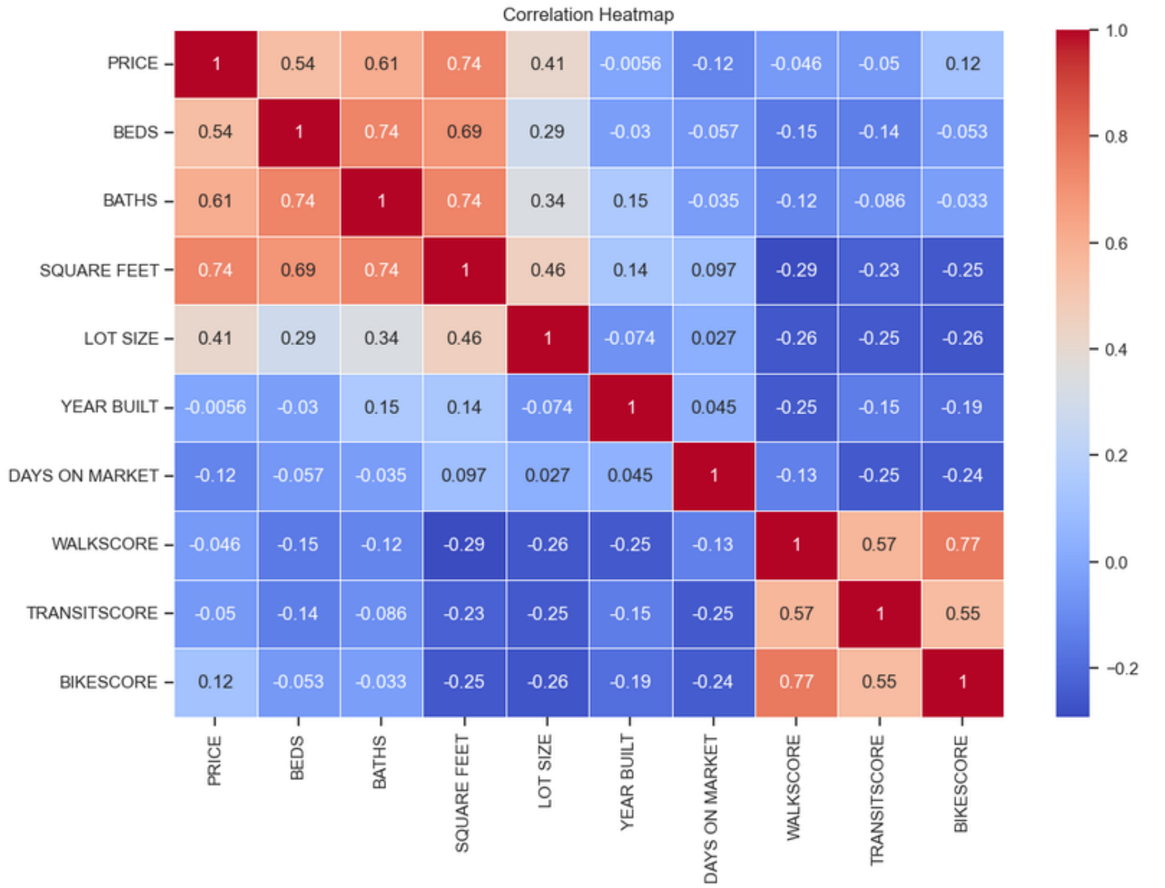




What features exhibit the strongest correlation with property prices in the San Francisco Bay Area?

-We examined the correlation heatmap to identify features with significant correlations to property prices.

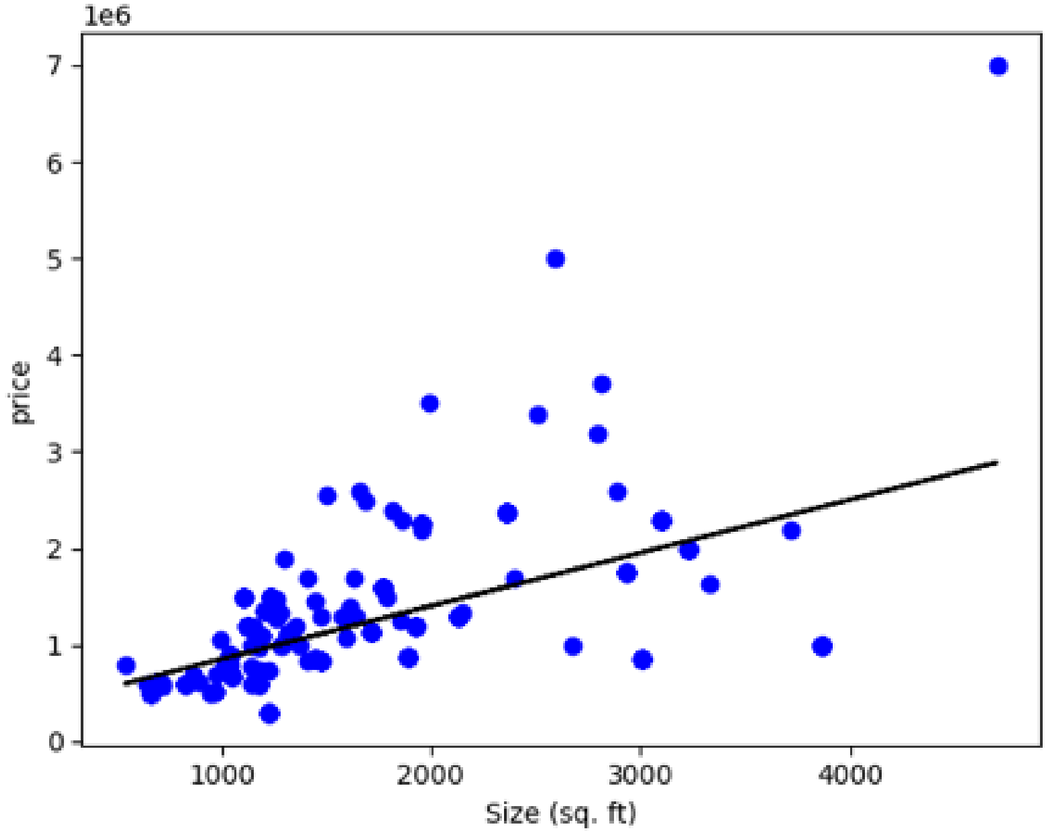
-Price is correlated more with Sqft followed by Bath and Beds



Linear Regression - Size vs. Price

-How does the size (square footage) of a property affect its price in the San Francisco Bay Area?

-We conducted a linear regression analysis to model the relationship between property size and price.



K-Nearest Neighbors (KNN) Classifier

-Can we classify the number of bedrooms in properties based on price and size?

-We employed a KNN to categorize properties by their number of bedrooms.

-The model was able to achieve an accuracy of 63.30% with K value of 22

Accuracy = 63.30%, K = 22

Conclusion

-In conclusion, our extensive analysis of the San Francisco Bay Area real estate market has yielded valuable insights and findings.

-Key Findings:

- The number of bedrooms exhibits a positive correlation with property prices, suggesting that larger properties tend to have higher prices. Property size (square footage) has a statistically significant impact on price, with larger properties generally commanding higher prices.

-Recommendations:

- For homebuyers, understanding the impact of property size and the number of bedrooms on price can inform more strategic purchasing decisions. Future research could explore additional factors affecting real estate prices, such as neighborhood amenities, and historical trends.

Thank you very much!