USA Real Estate Analysis

Team Members:

Abhiram Ravipati (A20539084)

Sumanth Kalyan Bandigupthapu (A20544342)

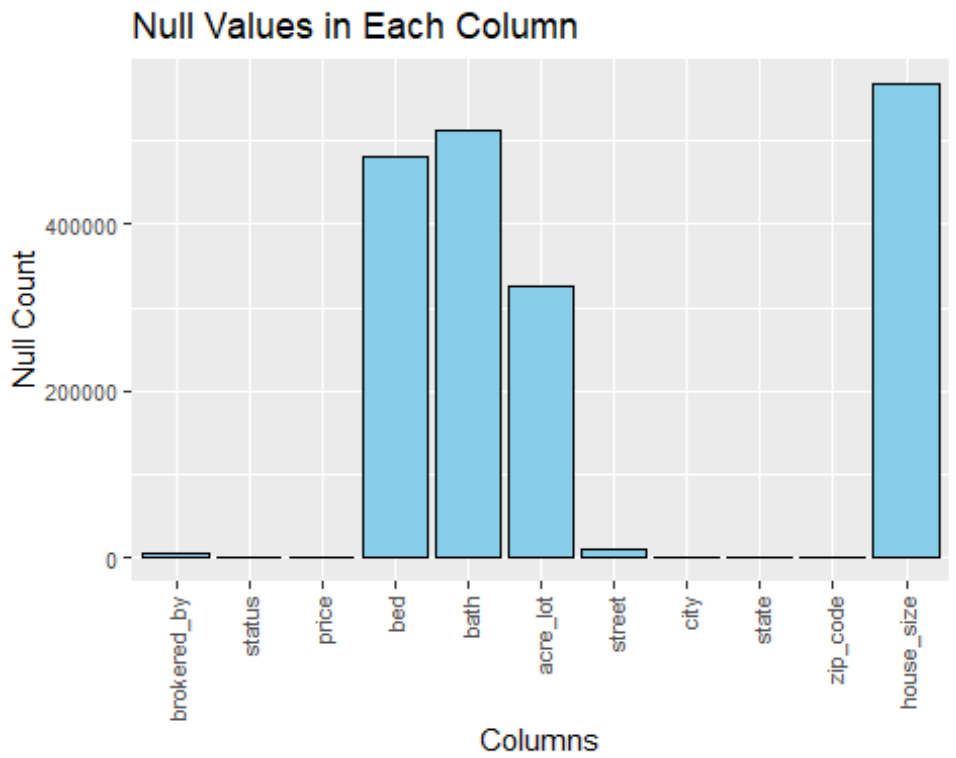
Introduction

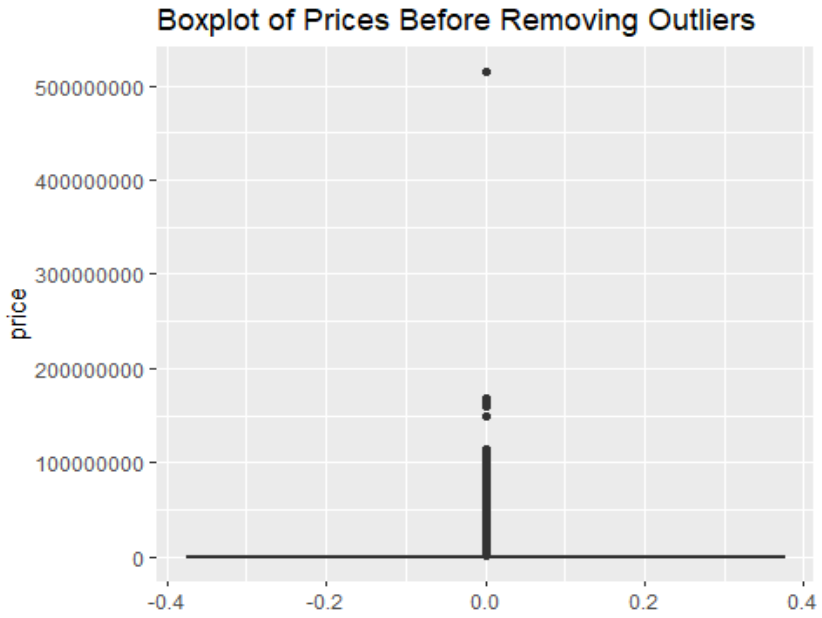
Our project aimed to analyze the real estate market trends and factors affecting property prices using a dataset of real estate listings. The study focused on exploring patterns in property prices across different cities and states, identifying key factors influencing price variations, and developing predictive models to forecast property prices. The research findings provide valuable insights into the dynamics of the real estate market, aiding stakeholders in making informed decisions.

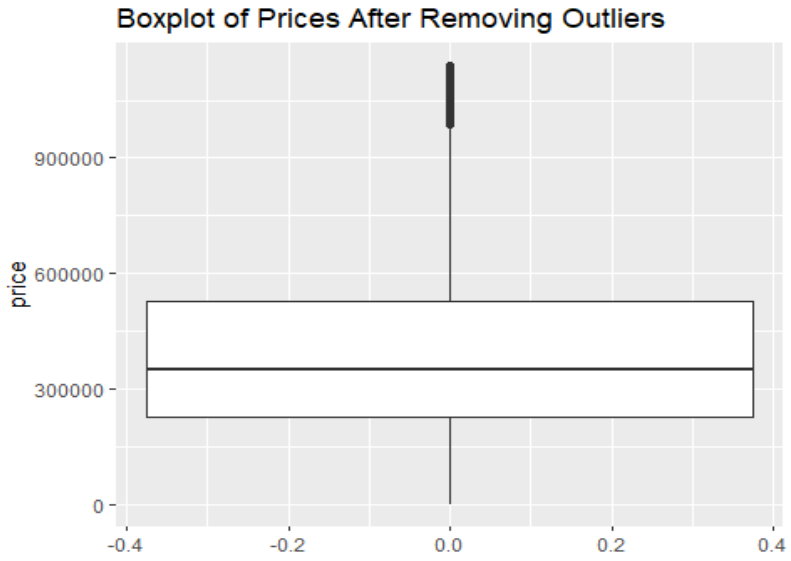
Methodology

- Data Preparation and Cleaning:

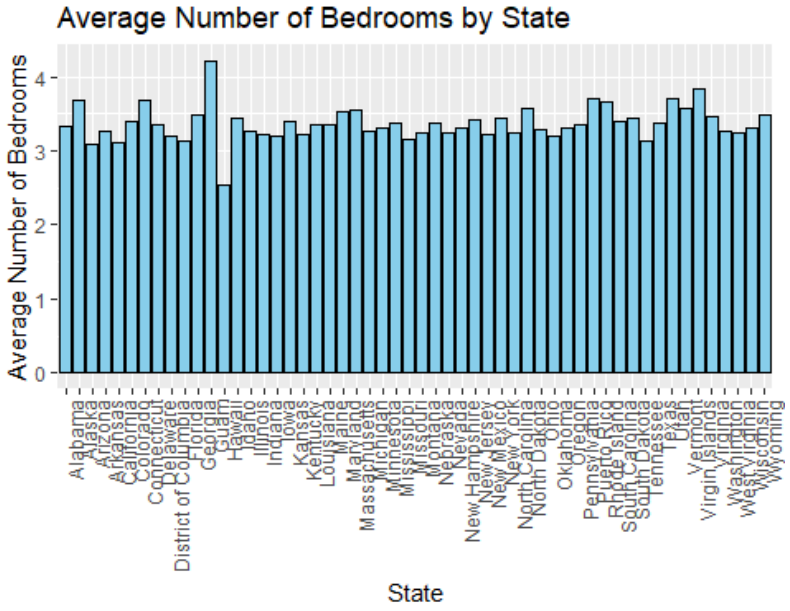
- Dropped prev_sold_date Column
- Removed Null Values
- Handled the Outliers to reduce their
- Negative impact on the Analysis

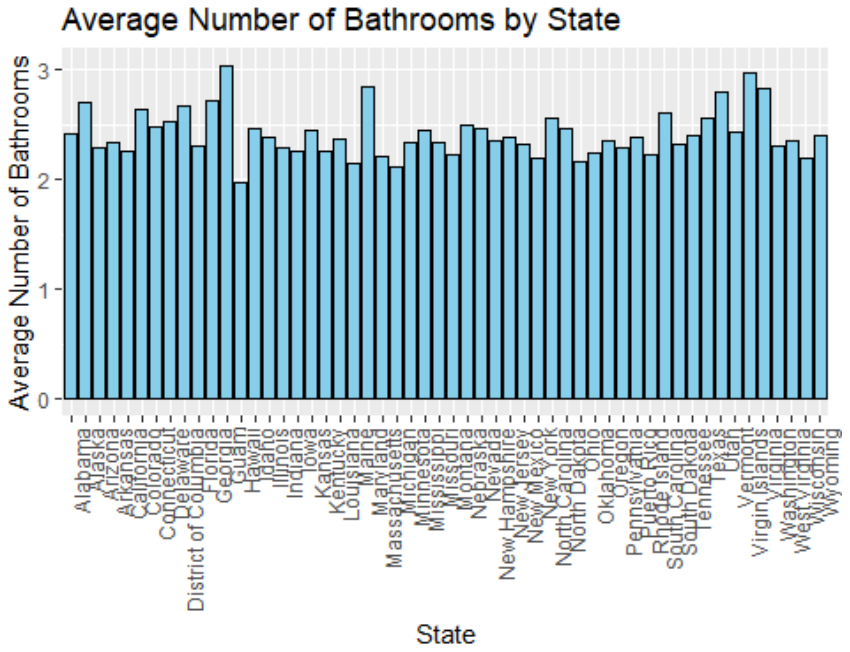


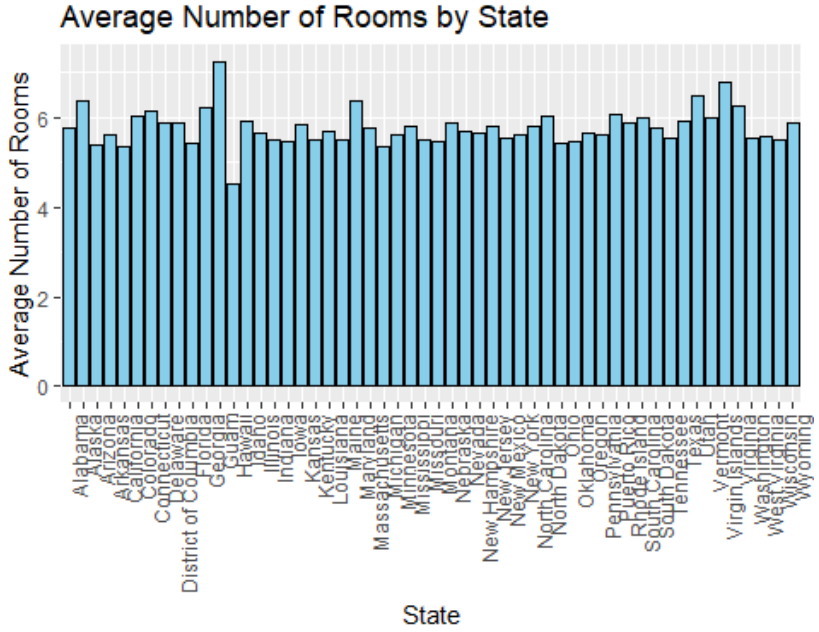

Boxplot of Prices Before Removing Outliers
price
500000000
400000000
300000000
200000000
100000000
0
-0.4
-0.2
0.0
0.2
0.4

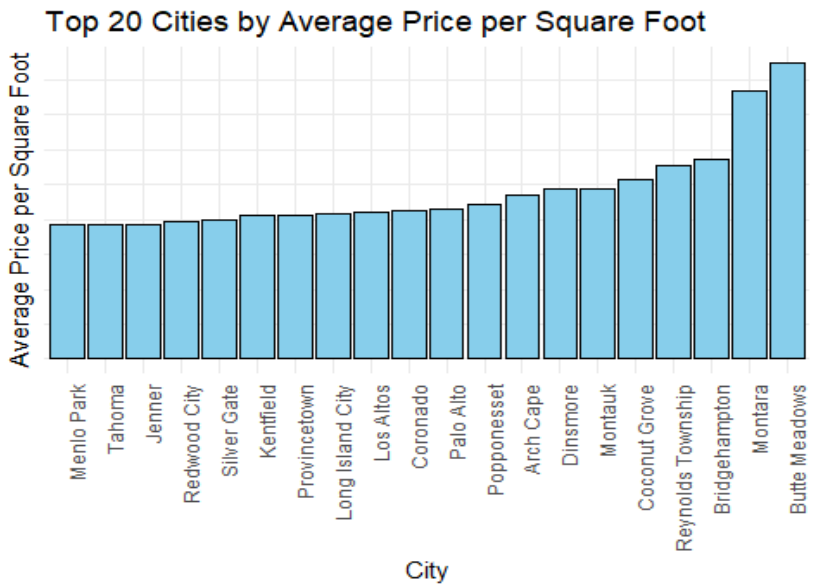

Boxplot of Prices After Removing Outliers
price
900000
600000
300000
0
-0.4
-0.2
0.0
0.2
0.4

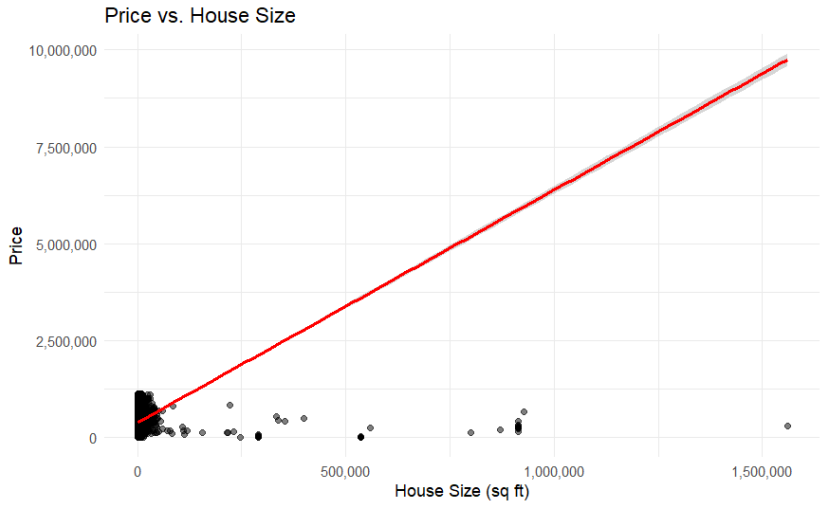
Exploratory Data Analysis

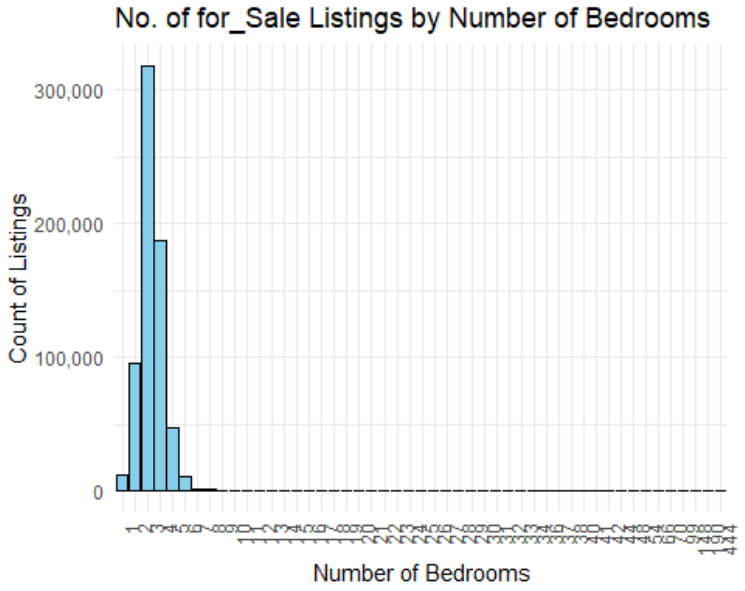

Average Number of Bedrooms by State
Average Number of Bedrooms
0
1
2
3
4
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Guam
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Puerto Rico
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virgin Islands
Virginia
Washington
West Virginia
Wisconsin
Wyoming
State

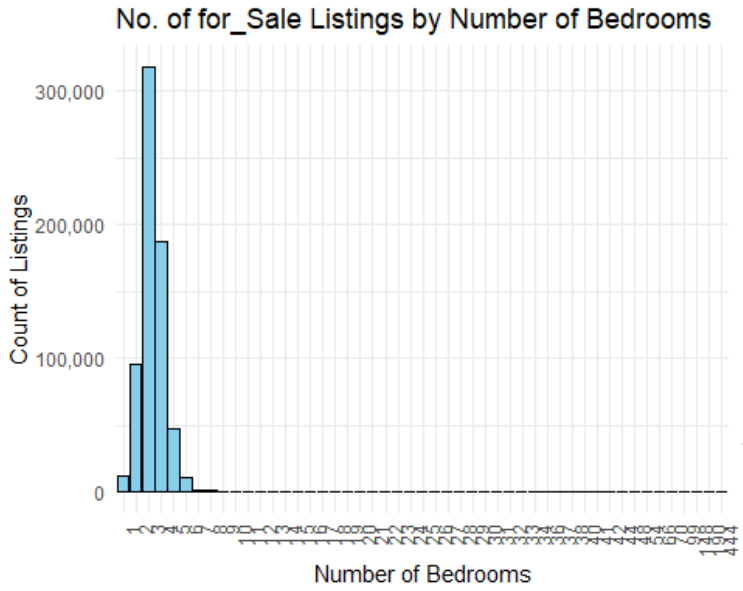

Average Number of Bathrooms by State
Average Number of Bathrooms
0
1
2
3
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Guam
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Puerto Rico
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virgin Islands
Virginia
Washington
West Virginia
Wisconsin
Wyoming
State

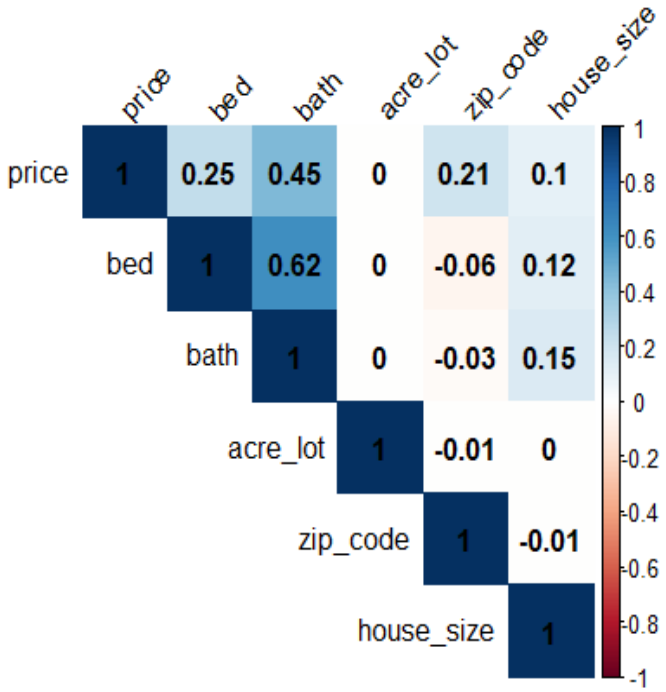

Average Number of Rooms by State
Average Number of Rooms
0
2
4
6
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Guam
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Puerto Rico
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virgin Islands
Virginia
Washington
West Virginia
Wisconsin
Wyoming
State

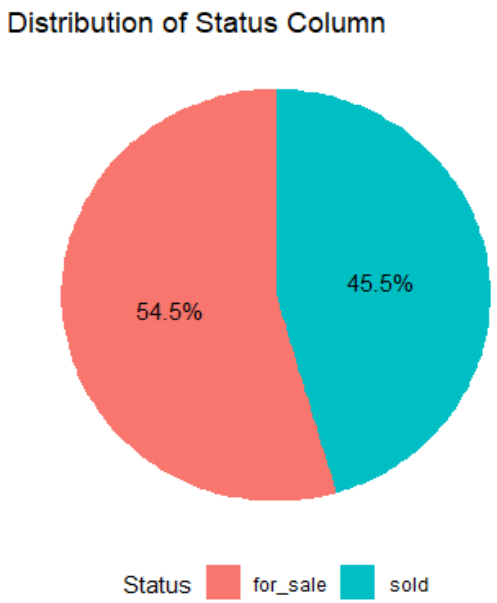

Top 20 Cities by Average Price per Square Foot
Average Price per Square Foot
Menlo Park
Tahoma
Jenner
Redwood City
Silver Gate
Kentfield
Provincetown
Long Island City
Los Altos
Coronado
Palo Alto
Popponesset
Arch Cape
Dinsmore
Montauk
Coconut Grove
Reynolds Township
Bridgehampton
Montara
Butte Meadows
City

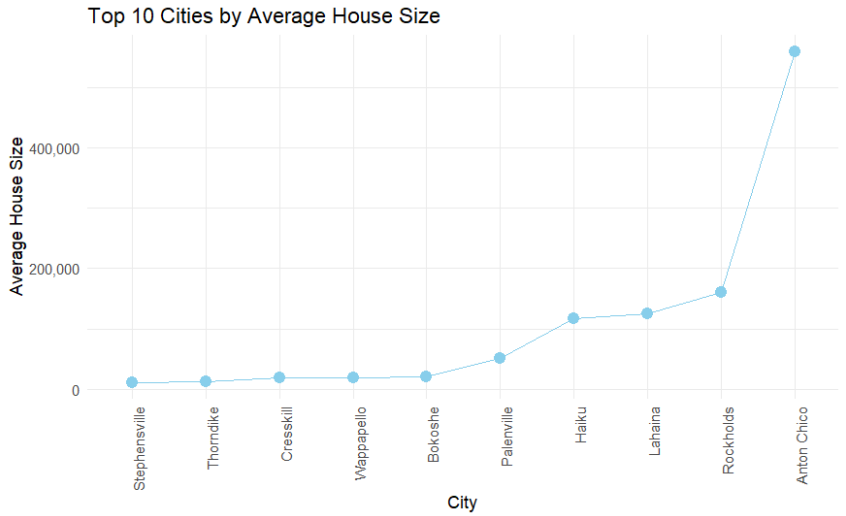

Price vs. House Size
Price
10,000,000
7,500,000
5,000,000
2,500,000
0
0
500,000
1,000,000
1,500,000
House Size (sq ft)

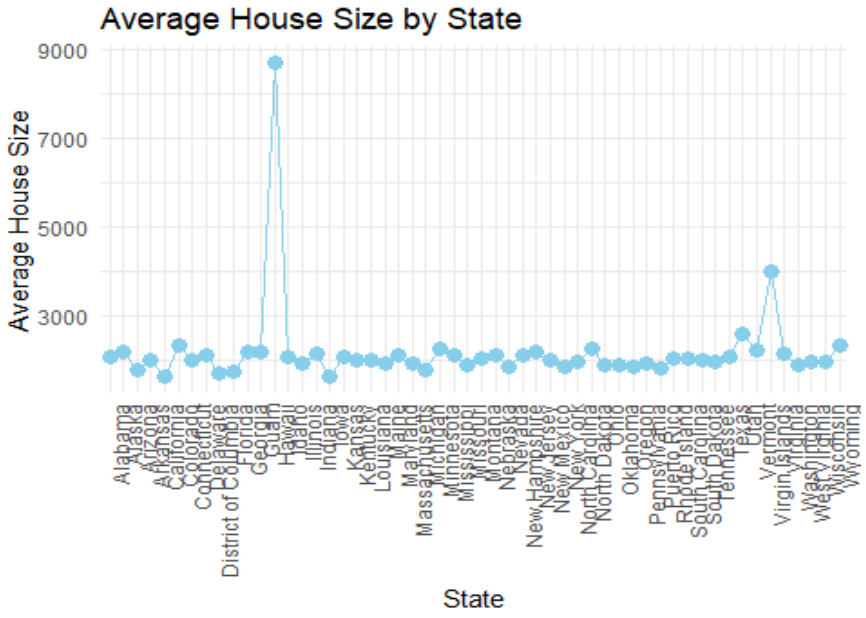

No. of for_Sale Listings by Number of Bedrooms
Count of Listings
300,000
200,000
100,000
0
Number of Bedrooms

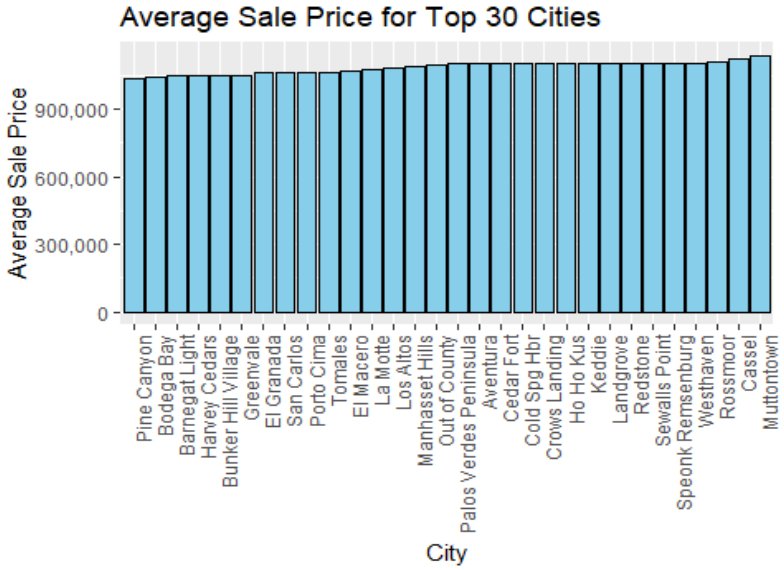

No. of for_Sale Listings by Number of Bedrooms
Count of Listings
300,000
200,000
100,000
0
Number of Bedrooms

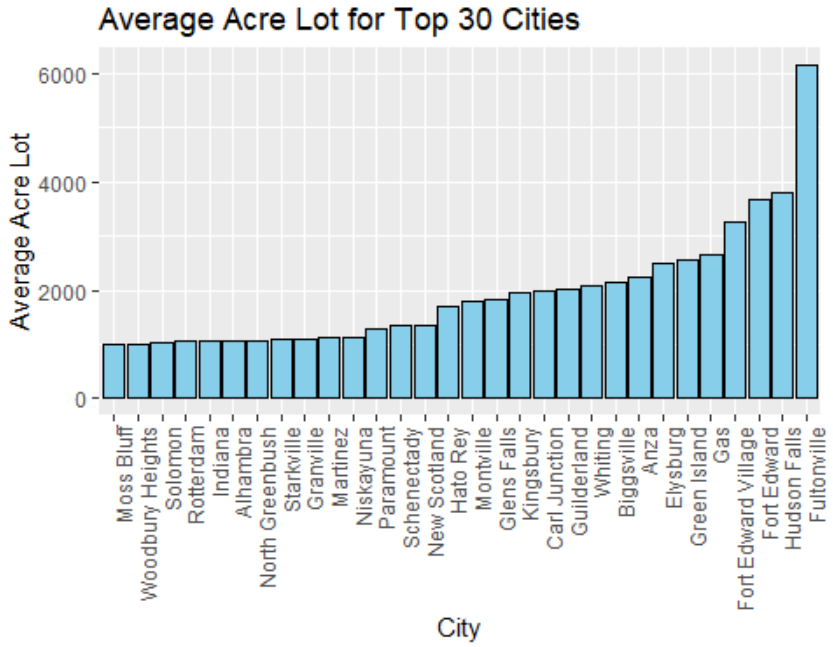

price
bed
bath
acre_lot
zip_code
house_size
price 1 0.25 0.45 0 0.21 0.1
bed 1 0.62 0 -0.06 0.12
bath 1 0 -0.03 0.15
acre_lot 1 -0.01 0
zip_code 1 -0.01
house_size 1
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

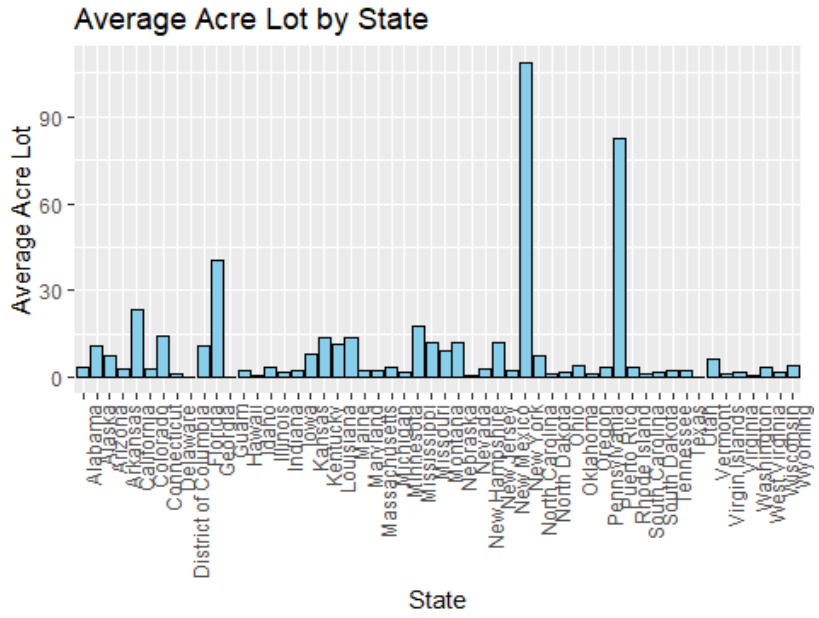

Distribution of Status Column
54.5%
45.5%
Status
for_sale
sold

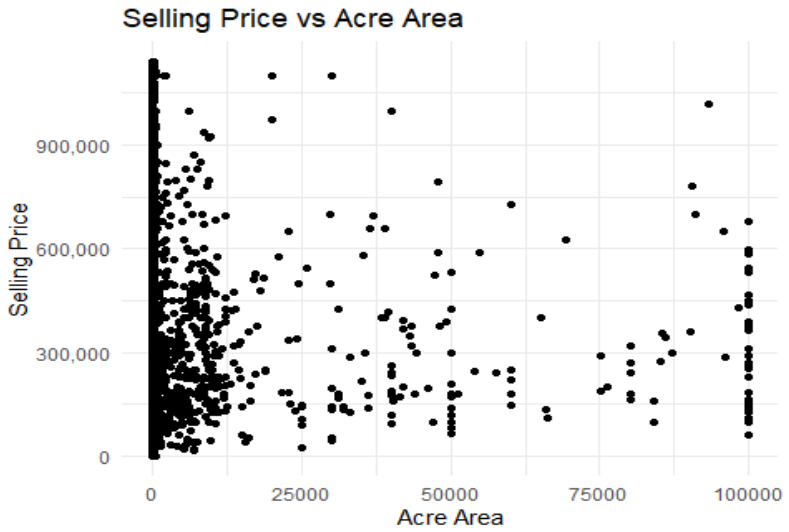

Top 10 Cities by Average House Size
Average House Size
400,000
200,000
0
Stephensville
Thorndike
Cresskill
Wappapello
Bokoshe
Palenville
Haiku
Lahaina
Rockholds
Anton Chico
City

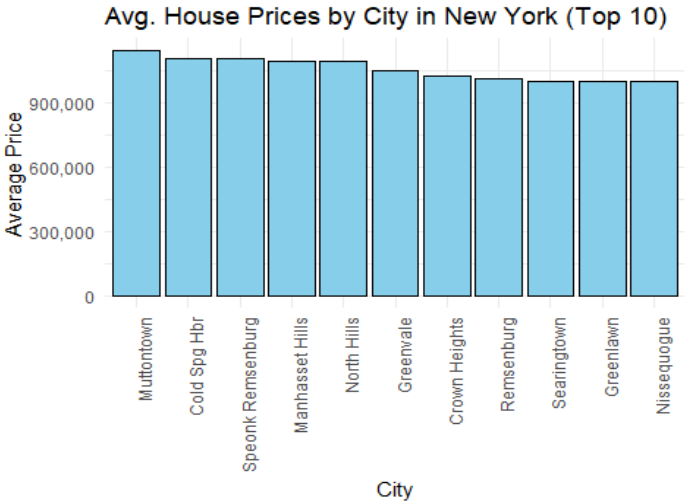

Average House Size by State
Average House Size
9000
7000
5000
3000
State

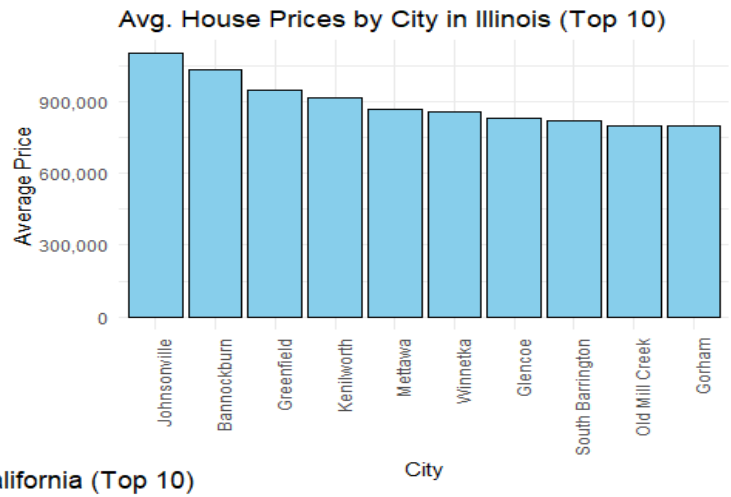

Average Sale Price for Top 30 Cities
Average Sale Price
900,000
600,000
300,000
0
Pine Canyon
Bodega Bay
Barnegat Light
Harvey Cedars
Bunker Hill Village
Greenvale
El Granada
San Carlos
Porto Cima
Tomales
El Macero
La Motte
Los Altos
Manhasset Hills
Out of County
Palos Verdes Peninsula
Aventura
Cedar Fort
Cold Spg Hbr
Crows Landing
Ho Ho Kus
Keddie
Landgrove
Redstone
Sewalls Point
Speonk Remsenburg
Westhaven
Rossmoor
Cassel
Muttontown
City

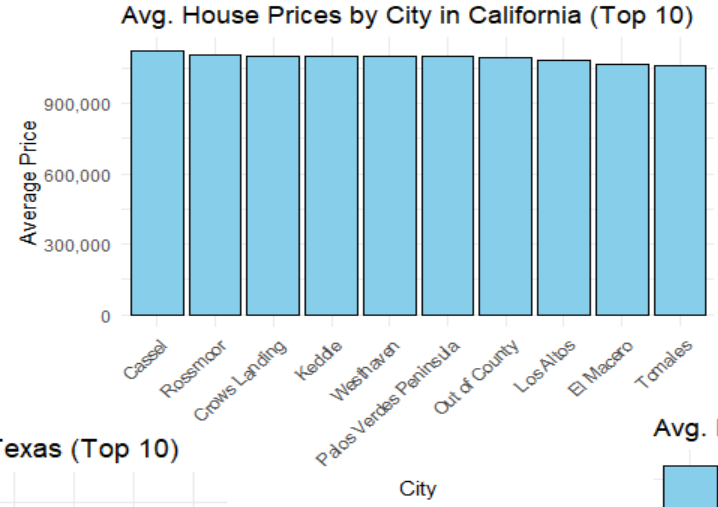

Average Acre Lot for Top 30 Cities
Average Acre Lot
6000
4000
2000
0
Moss Bluff
Woodbury Heights
Solomon
Rotterdam
Indiana
Alhambra
North Greenbush
Starkville
Granville
Martinez
Niskayuna
Paramount
Schenectady
New Scotland
Hato Rey
Montville
Glens Falls
Kingsbury
Carl Junction
Guilderland
Whiting
Biggsville
Anza
Elysburg
Green Island
Gas
Fort Edward Village
Fort Edward
Hudson Falls
Fultonville
City

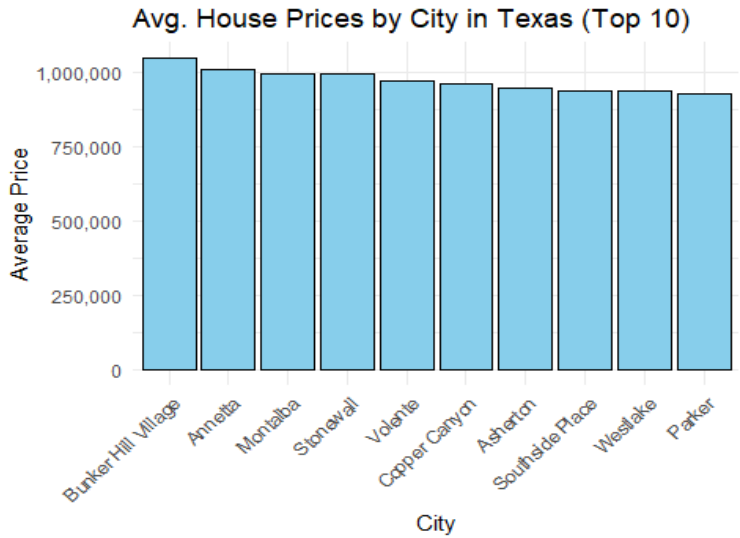

Average Acre Lot by State
Average Acre Lot
90
60
30
0
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Guam
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Puerto Rico
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virgin Islands
Virginia
Washington
West Virginia
Wisconsin
Wyoming
State

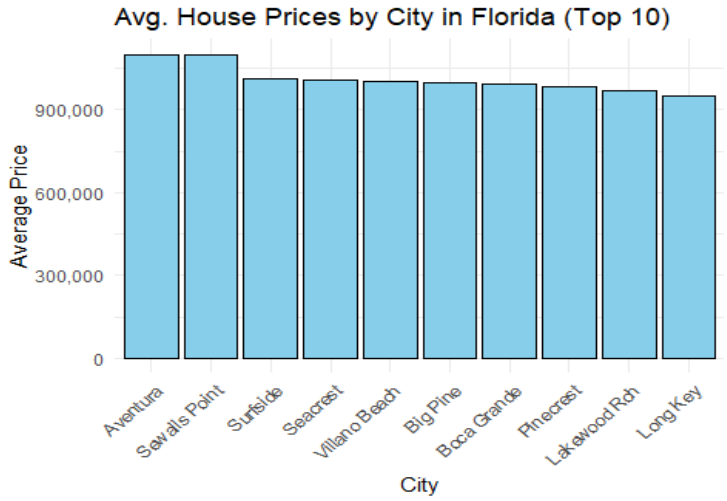

Selling Price vs Acre Area
Selling Price
900,000
600,000
300,000
0
0
25000
50000
75000
100000
Acre Area

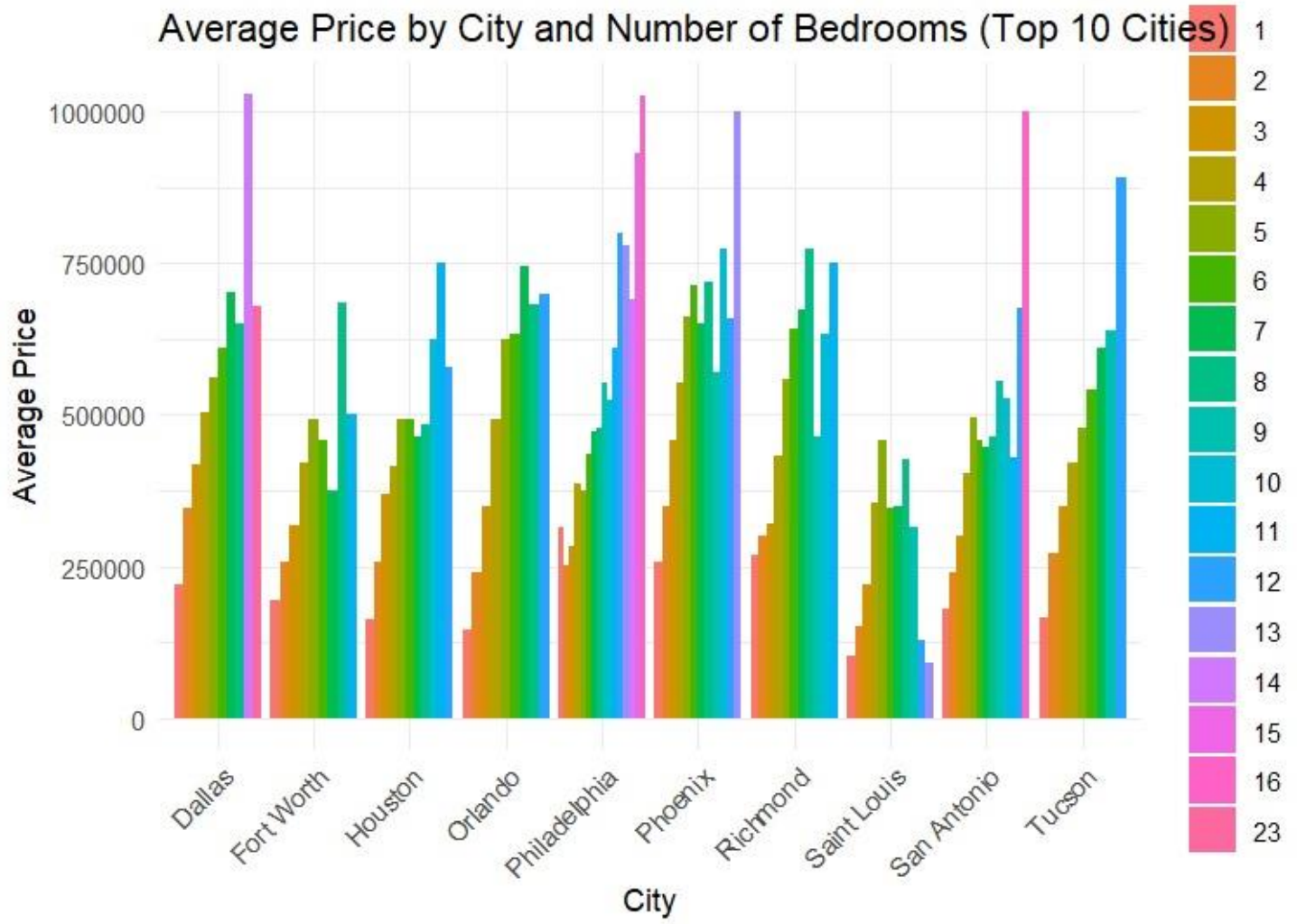

Avg. House Prices by City in New York (Top 10)
Average Price
900,000
600,000
300,000
0
Muttontown
Cold Spg Hbr
Speonk Remsenburg
Manhasset Hills
North Hills
Greenvale
Crown Heights
Remsenburg
Searingtown
Greenlawn
Nissequogue
City


Avg. House Prices by City in Illinois (Top 10)
Average Price
900,000
600,000
300,000
0
Johnsonville
Bannockburn
Greenfield
Kenilworth
Mettawa
Winnetka
Glencoe
South Barrington
Old Mill Creek
Gorham
City


Avg. House Prices by City in California (Top 10)
Average Price
900,000
600,000
300,000
0
Cassel
Rossmoor
Crows Landing
Keddie
Westhaven
Palos Verdes Peninsula
Out of County
Los Altos
El Macero
Tomales
City


Avg. House Prices by City in Texas (Top 10)
Average Price
1,000,000
750,000
500,000
250,000
0
Bunker Hill Village
Annetta
Montalba
Stonewall
Volente
Copper Canyon
Asherton
Southside Place
Westlake
Parker
City


Avg. House Prices by City in Florida (Top 10)
Average Price
900,000
600,000
300,000
0
Aventura
Sewalls Point
Surfside
Seacrest
Villano Beach
Big Pine
Boca Grande
Pinecrest
Lakewood Rch
Long Key
City


Average Price by City and Number of Bedrooms (Top 10 Cities)
Average Price
1000000
750000
500000
250000
0
Dallas
Fort Worth
Houston
Orlando
Philadelphia
Phoenix
Richmond
Saint Louis
San Antonio
Tucson
City
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
23

Modelling

- The linear regression model exhibited a notable accuracy of 78.8%, indicating its proficiency in categorizing property prices. With a minimal Mean Squared Error (MSE) of 42.7 billion and a Mean Absolute Error (MAE) of approximately 158,374, the model demonstrates robust performance. Root Mean Squared Error (RMSE) stands at 206,742, reflecting a relatively low level of prediction error. Despite some limitations, such as the model's sensitivity and specificity, the overall balanced accuracy remains satisfactory. This suggests the model's ability to maintain a reasonable balance between true positive and true negative rates.
- Additionally, the model's Kappa statistic indicates a marginal agreement between observed and predicted classifications. However, further refinement may be required to enhance the model's predictive capabilities. Overall, the model presents a promising tool for analyzing real estate market trends and making informed decisions.

Conclusion, Limitations and Future Work

- The real estate analysis uncovers significant variations in house prices across cities, driven by location and property attributes. Correlations between price, house size, and acre area affirm key determinants of real estate value. Despite achieving 78% accuracy, the model's performance suggests opportunities for enhancement through feature refinement. Market dynamics, as illustrated by the distribution of "for_sale" and "sold" listings, provide valuable context for predicting future trends. Overall, the project offers insights into real estate trends and market dynamics, facilitating informed decision-making in the industry.

- Our current focus on specific features like house size and acre area might be too narrow. We're missing out on other crucial factors like neighborhood quality and nearby amenities, which can significantly impact property prices. To improve our model, we should expand our dataset to include more diverse features and explore different machine learning methods. It's also important to validate our model and fine-tune its parameters to ensure its accuracy and consistency.

Thank You!!!