Predictive Real Estate Value With Housing Data

-BY
Manoj

Date:09-Sep-2024

Introduction:

- Real estate firms play a crucial role in the property market, acting as intermediaries between buyers and sellers, landlords and tenants, as well as developers and investors.
- Their core function is to help clients navigate the complexities of real estate transactions, ensuring that their interests are well-represented and that they make informed decisions based on market data and trends.

Problem Statement:

- Real estate firms wants to analyze market trends, understand price movements, and identify areas with the highest growth.

Objective:

- Create predictive models to forecast house prices based on key features.
- To Help Real estate firms seeking data-driven insights to shape business strategies, marketing efforts, and inventory management.

OverView Of House DATA Set

- House data contains 16 features and 5000 rows
- Selected relevant features for modeling: zipcodes, lot size, taxes, year built, bedrooms, bathrooms, square footage, and garage availability.
- Split the data into 80% training and 20% testing sets

Model Implementation:

Multiple Linear Regression:

It is a statistical technique that uses several explanatory variables to predict the outcome of a response variable.

Before Tuning

Accuracy:

Train set:80.26%

Test set:80.49%

Mean Absolute Error

Train set:19.74%

Test set:19.51%

After Tuning by Gradient Descent

Accuracy:

Train set:80.84%

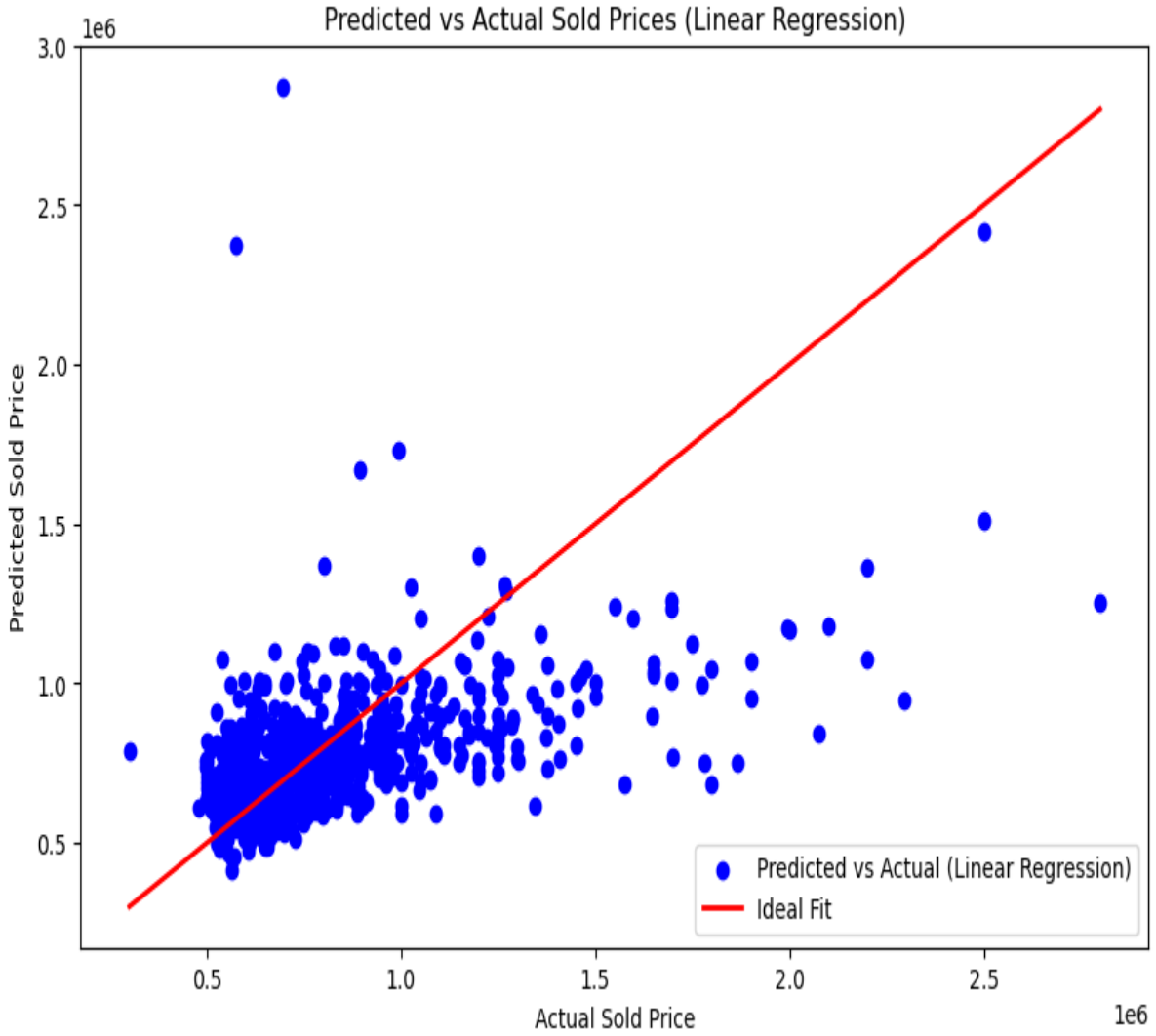
Test set:81.12%

Mean Absolute Error

Train set:19.16%

Test set:18.88%

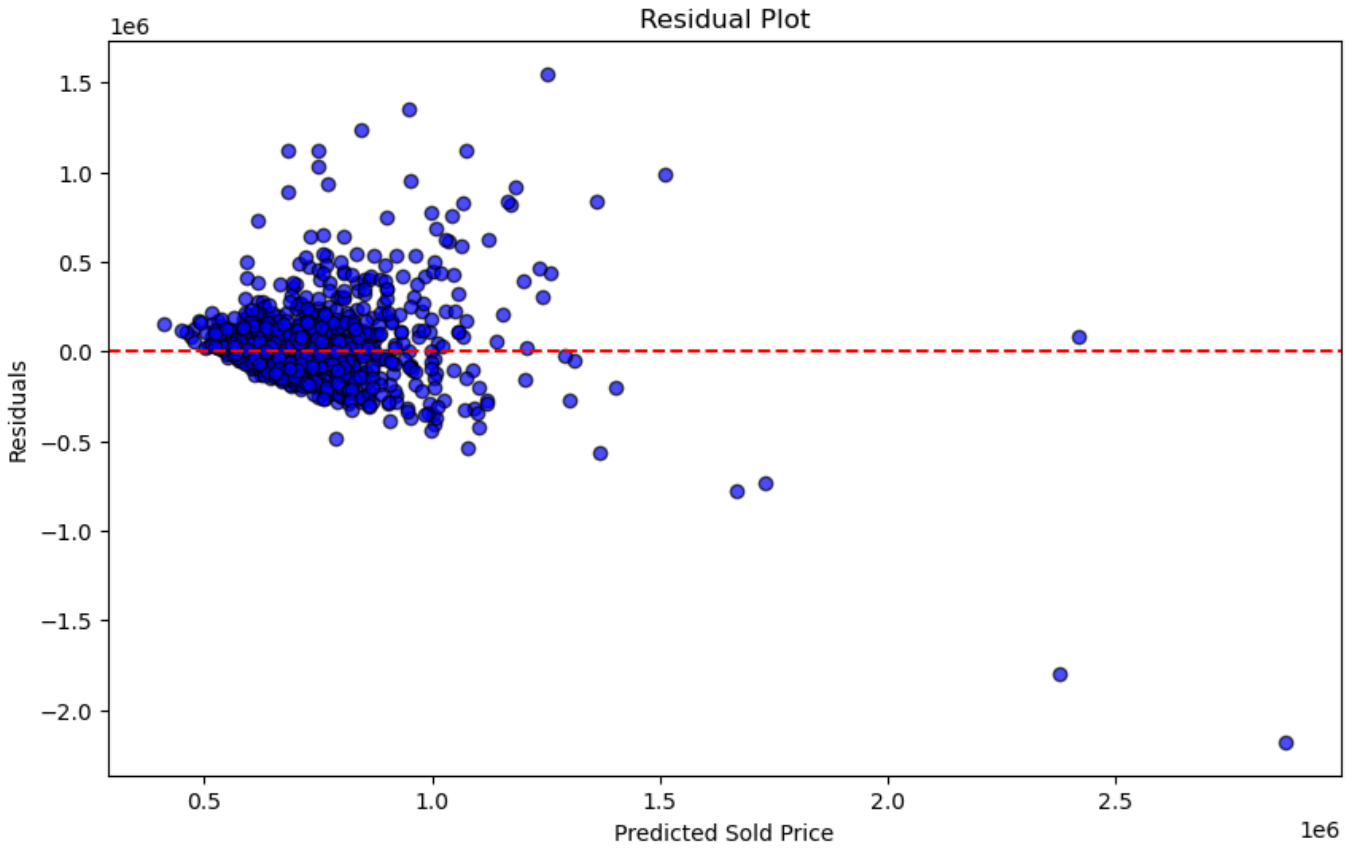
Key Visulas



Insights:

- The **blue points** show the predicted vs. actual sold prices based on the linear regression model.
- The **red line** represents the ideal fit, meaning if the predictions were perfect, all blue points would lie on this line.
- Since the blue points are spread out and deviate from the red line, it suggests that the model's predictions some deviate from the actual sold prices, indicating that the model performing good but with some deviation

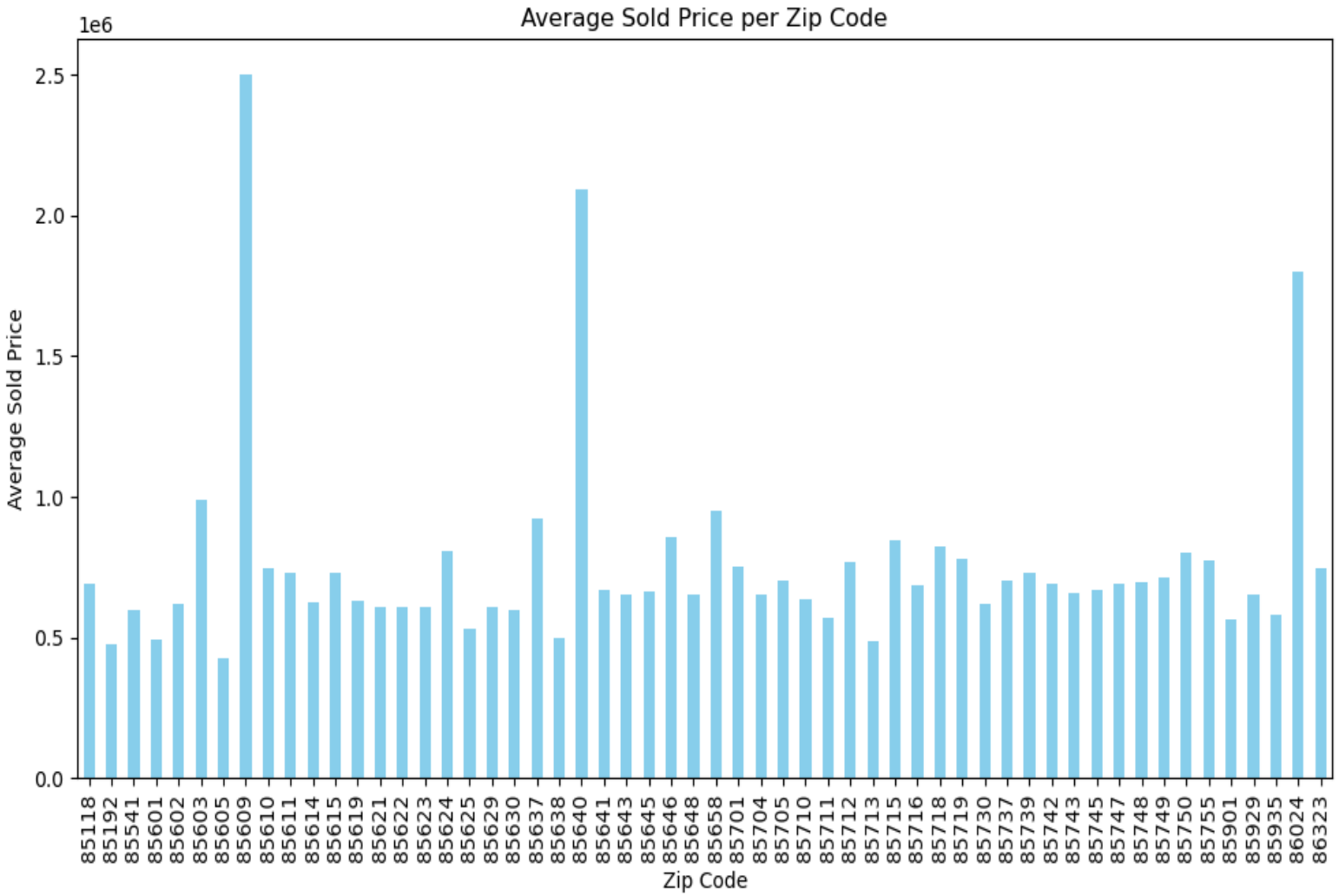
Key Visulas



Insights

The plot shows that while many predictions are close to the actual values, there are Some systematic deviations.

Key Visuals



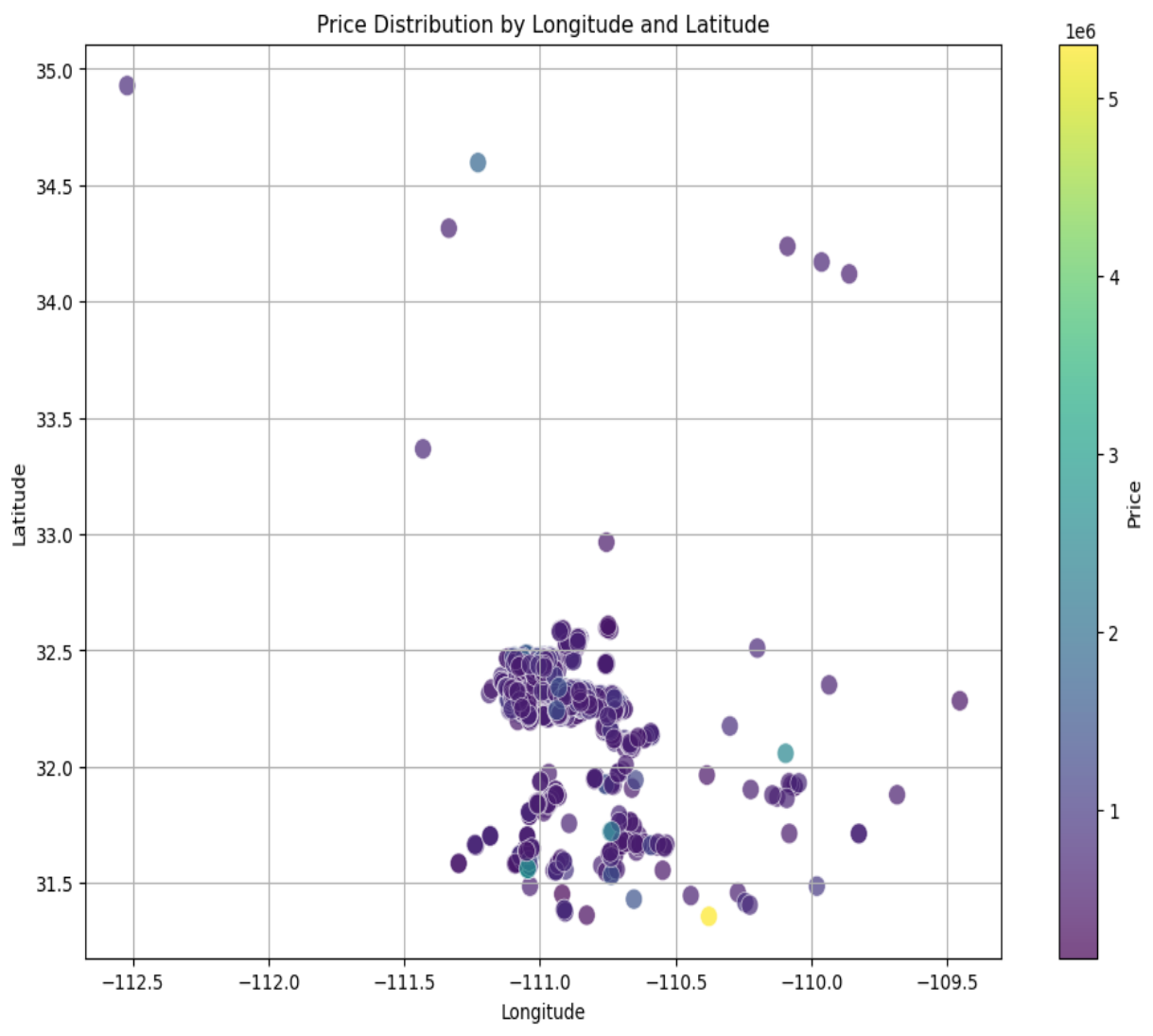
Insights

- Certain zip codes, such as **85602** and **85640**, have significantly higher average sold prices (around 2 million or more), indicating that these areas may be wealthier or have higher property values.

- Most zip codes have average sold prices clustered below 1 million, with a few standing out as outliers.

.

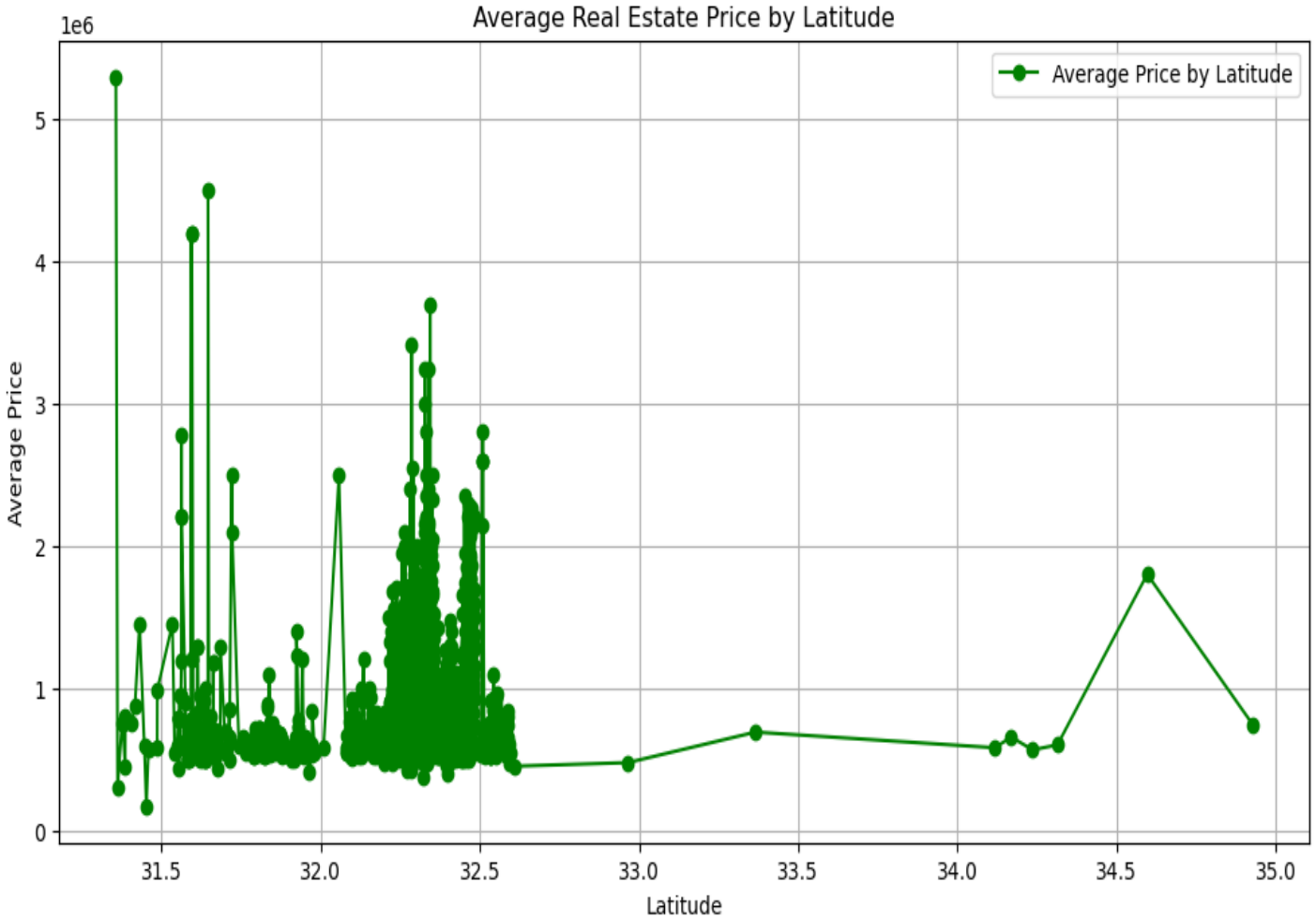
Key Visuals



Insights

- The plot provides insight into how prices vary across different locations, with the clustering suggesting regions of high or low price concentration.
- The most of the locations fall into the lower price ranges (darker colors), with a few locations showing higher prices (lighter colors).

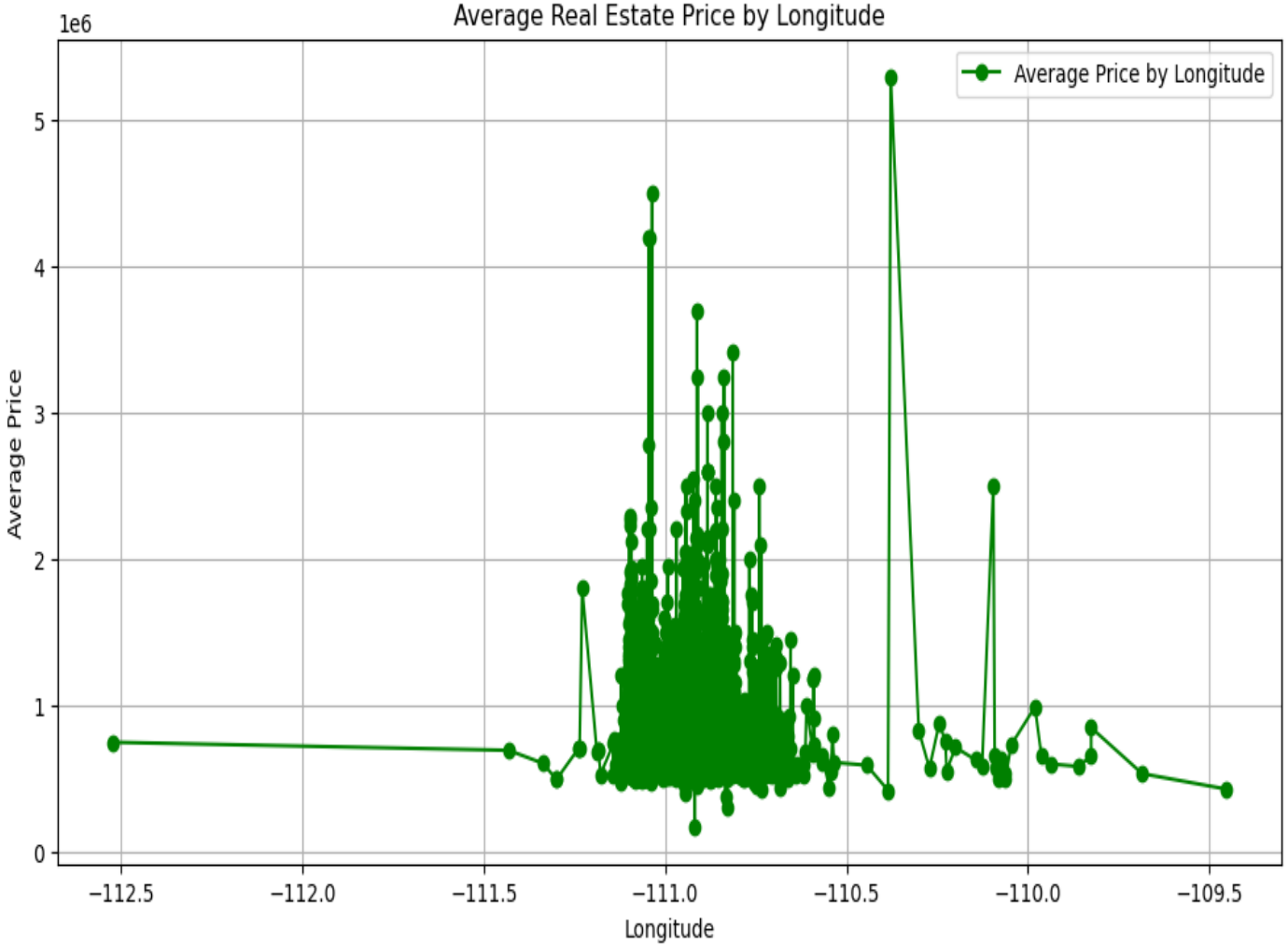
Key Visuals



Insights

- This plot provides insights into how average real estate prices vary with latitude, showing a densely populated lower-latitude region with more price variability, and sparser, more stable pricing at higher latitudes.

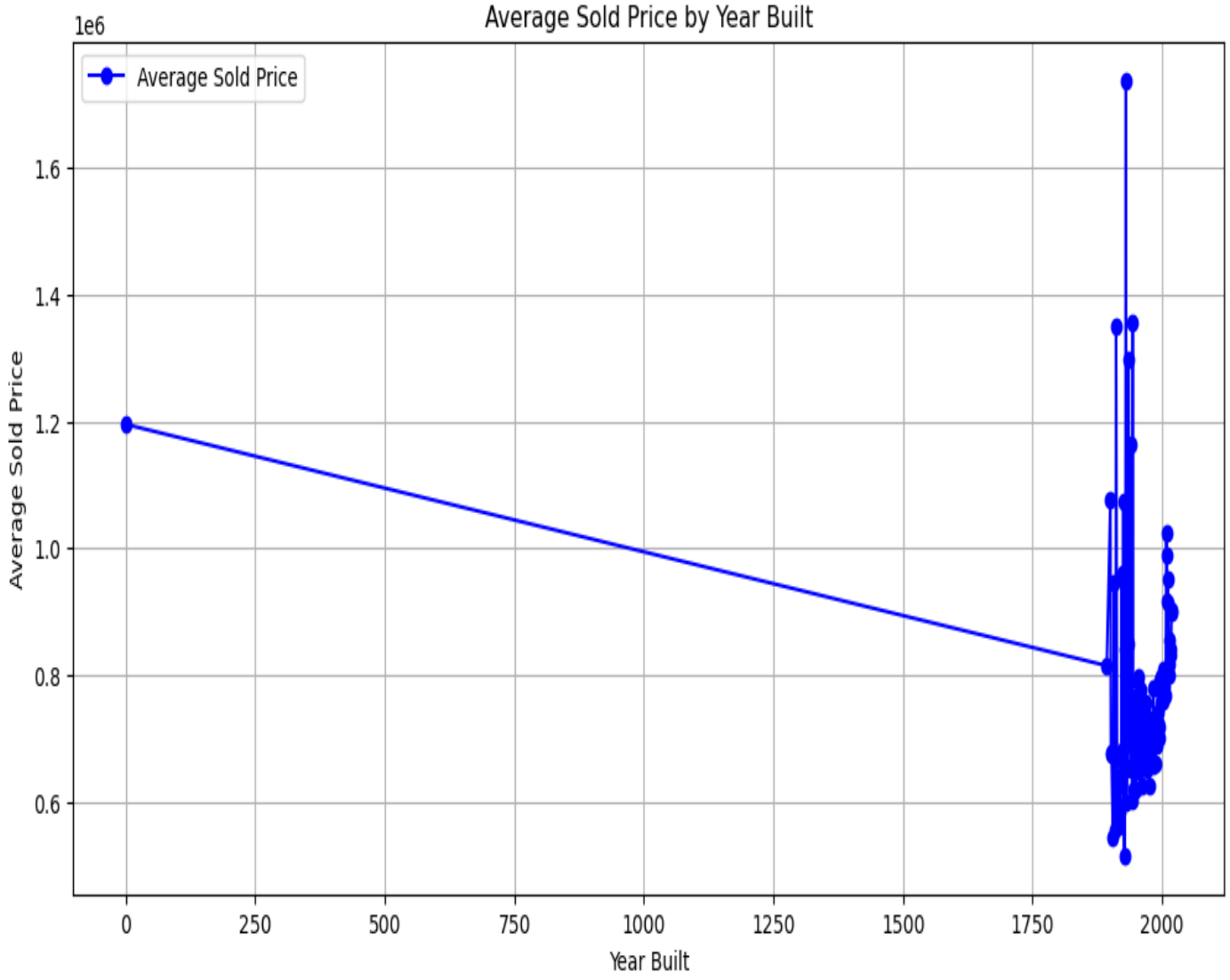
Key Visuals



Insights

- This plot highlights how average real estate prices vary with longitude, showing that most price variability and higher prices occur within a specific longitude range, while areas outside this range tend to have lower, more stable average prices.

Key Visuals



Insights

- There is a sharp decline in average sold prices from year 0 to the year 2000.
- Most data points are concentrated in the years closer to 1900-2000, with prices showing variability.

SUMMARY

- The model gives most accurate predicting results but further tuning or using some other advanced techniques may improve the accuracy of the model.
- Some zip codes have significantly higher average sold prices indicating that these areas may be wealthier or have higher property values.
- The Insights shows with specific regions based on latitude and longitude have higher sold prices.
- Most data points are concentrated in the years closer to 1900-2000, with prices showing variability.
- The fluctuations might be due to market conditions, economic factors, or data inconsistencies. Further analysis is needed to understand the reasons behind these price variations.
- By enhancing the models, real estate firms can use this tool to make better-informed decisions about marketing campaigns, inventory management, and business strategies.
- It helps to real estate firms Identify growth areas to target high-demand regions ,Optimize inventory based on price trends and Improve marketing campaigns by focusing on trending areas.

THANK YOU