MORINGA
Discover · Grow · Transform

KING COUNTY HOUSING PROJECT.

Moringa School

Phase 2 Project

MEET OUR TEAM

- Yussuf Hersi
- Bravin Mugangasia
- Pauline Wambui
- Kibet Kemboi
- Brian Kisilu
- Ronald Nyagaka

BACKGROUND

Business Problem

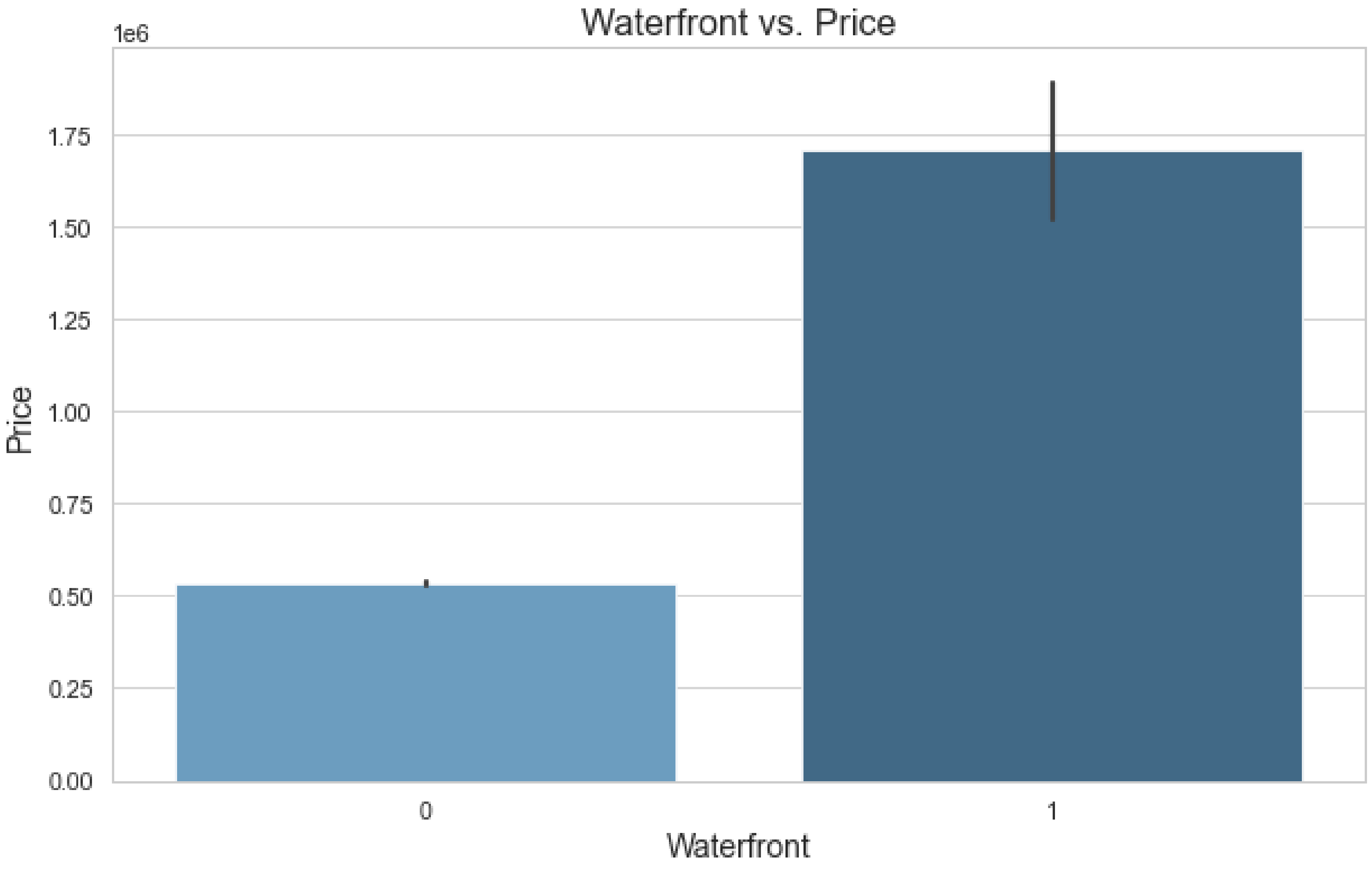
- The client wants to create a successful platform for buying and selling houses in King County.
- Accurate estimates of house prices are crucial for both buyers and sellers.
- The platform needs a model that can infer the most important features that determine house prices in King County.
- The model needs to be trained on data that accurately represents the real estate market in King County.
- The model can be integrated into the platform to provide buyers and sellers with accurate estimates of house prices.

OBJECTIVES

- To identify the key features that significantly influence house prices in King County.
- To develop a model that accurately estimates house prices based on the identified features.
- To evaluate the performance of the developed model in estimating house prices in King County.

DATA EXPLORATION

Exploratory Data Analysis: Waterfront Feature



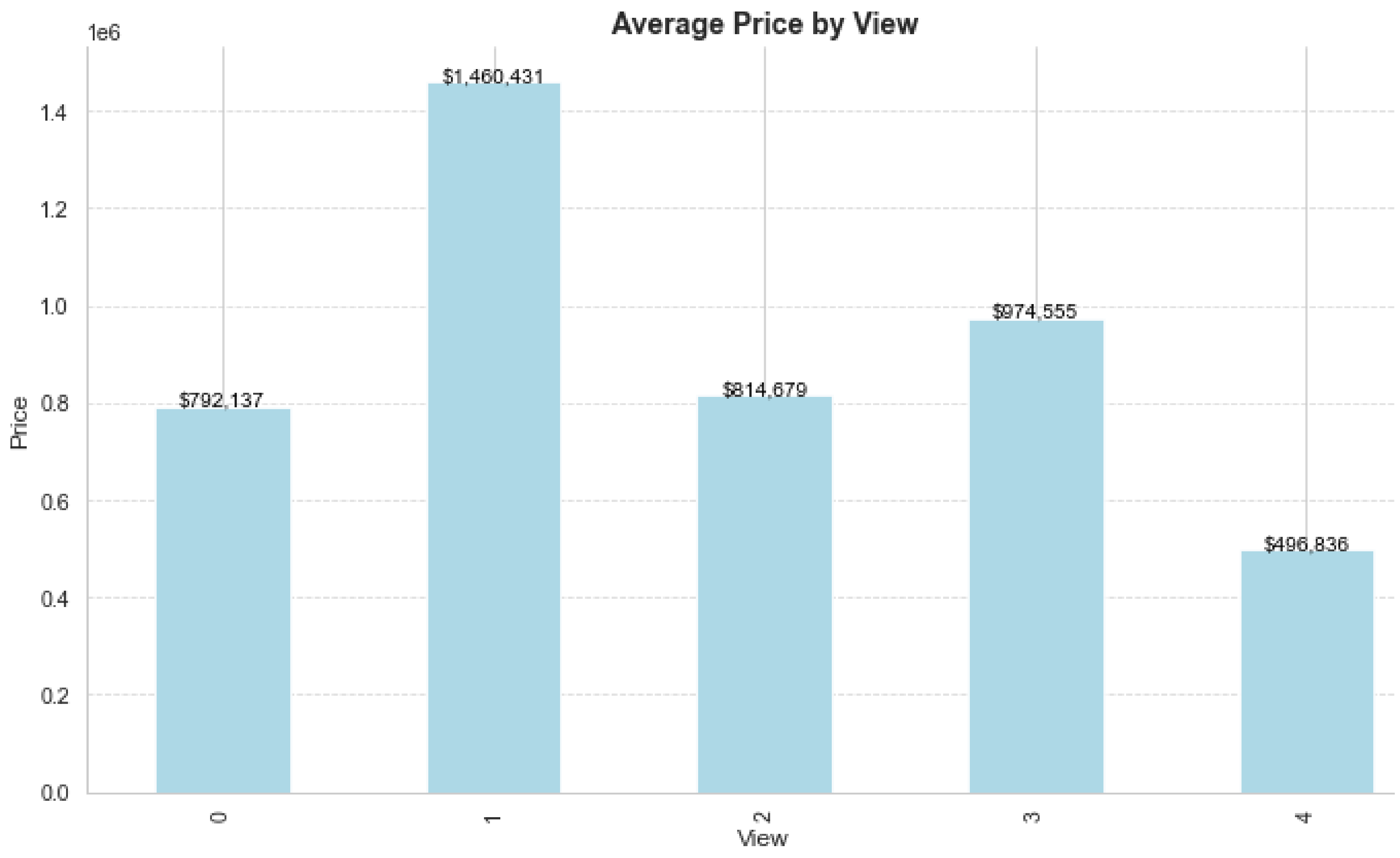
- This visualization helps to understand the impact of the waterfront on the price of the house. The boxplot shows that houses with waterfront tend to have higher prices compared to houses without

0 - YES
1 - NO

DATA EXPLORATION

Exploratory Data Analysis: View Feature

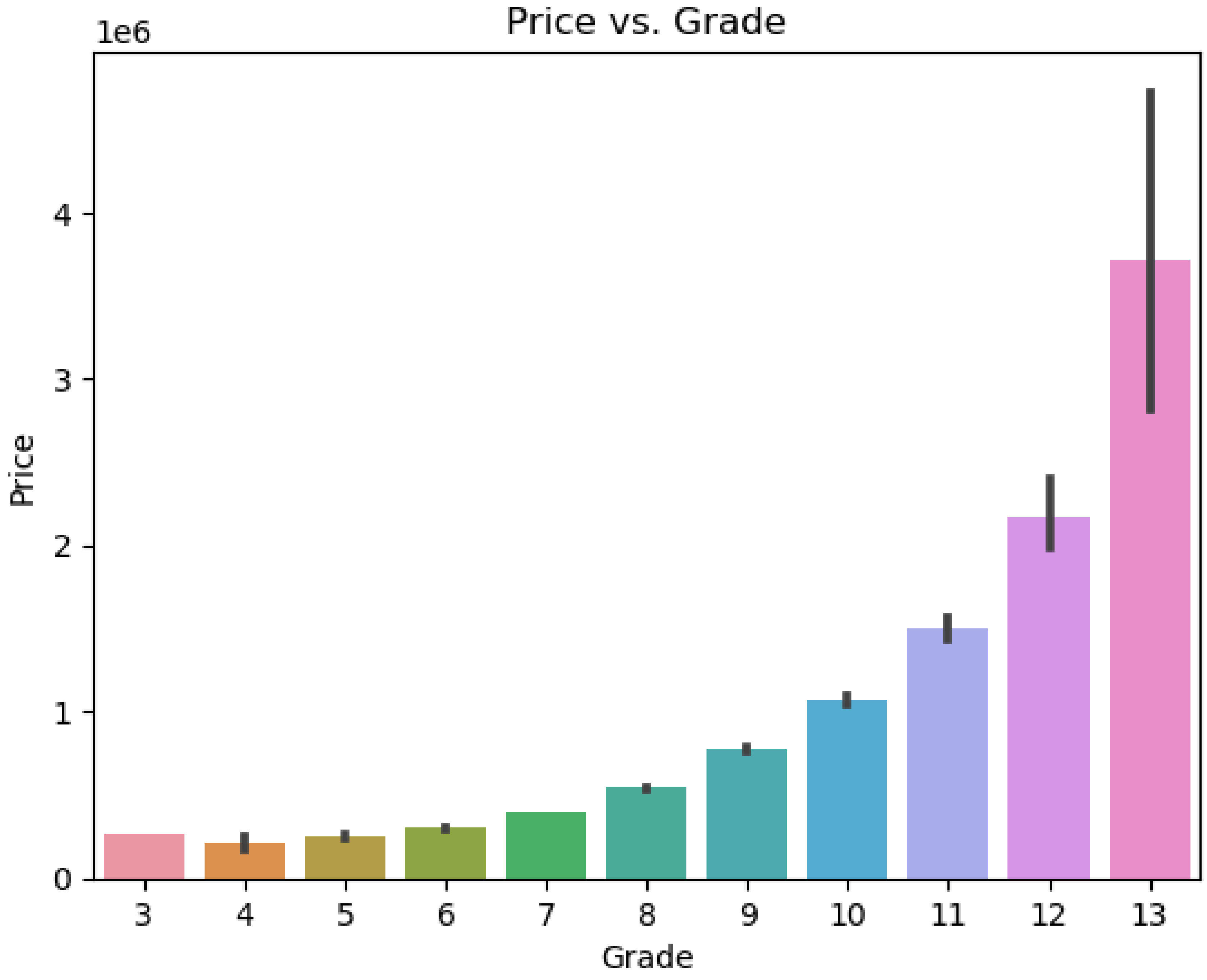
Average Price Vs View



The price increases as the view increases, however it seems houses with a good view 2 are priced higher than houses with a fair view

DATA EXPLORATION

Exploratory Data Analysis: Grade Feature



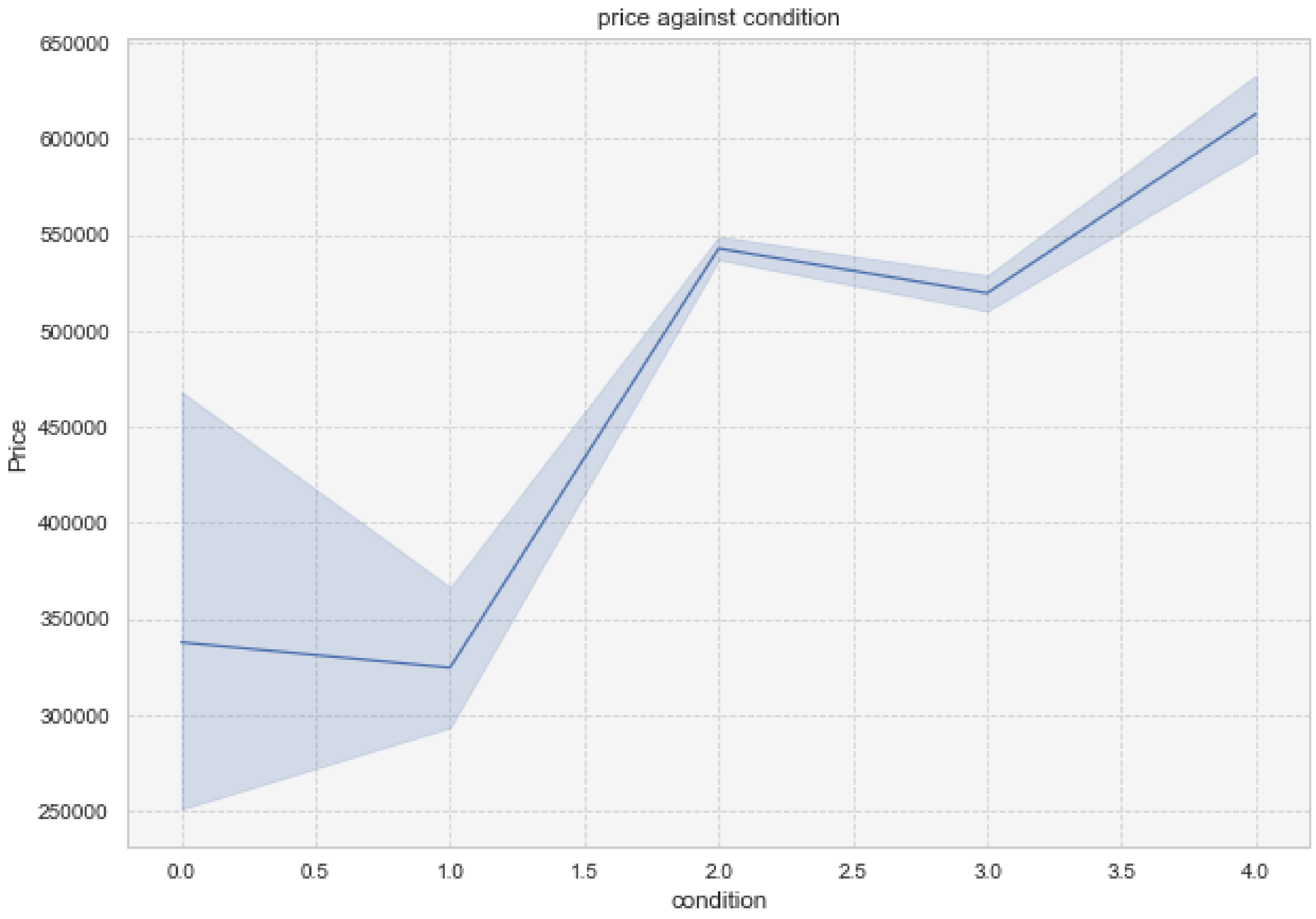
From this, we can see that as the grade of the house increases so does the sales price

The average price of a house with a grade of 13 is the highest

The average price of a house with a grade of 3 is the lowest

DATA EXPLORATION

Exploratory Data Analysis: Condition Feature

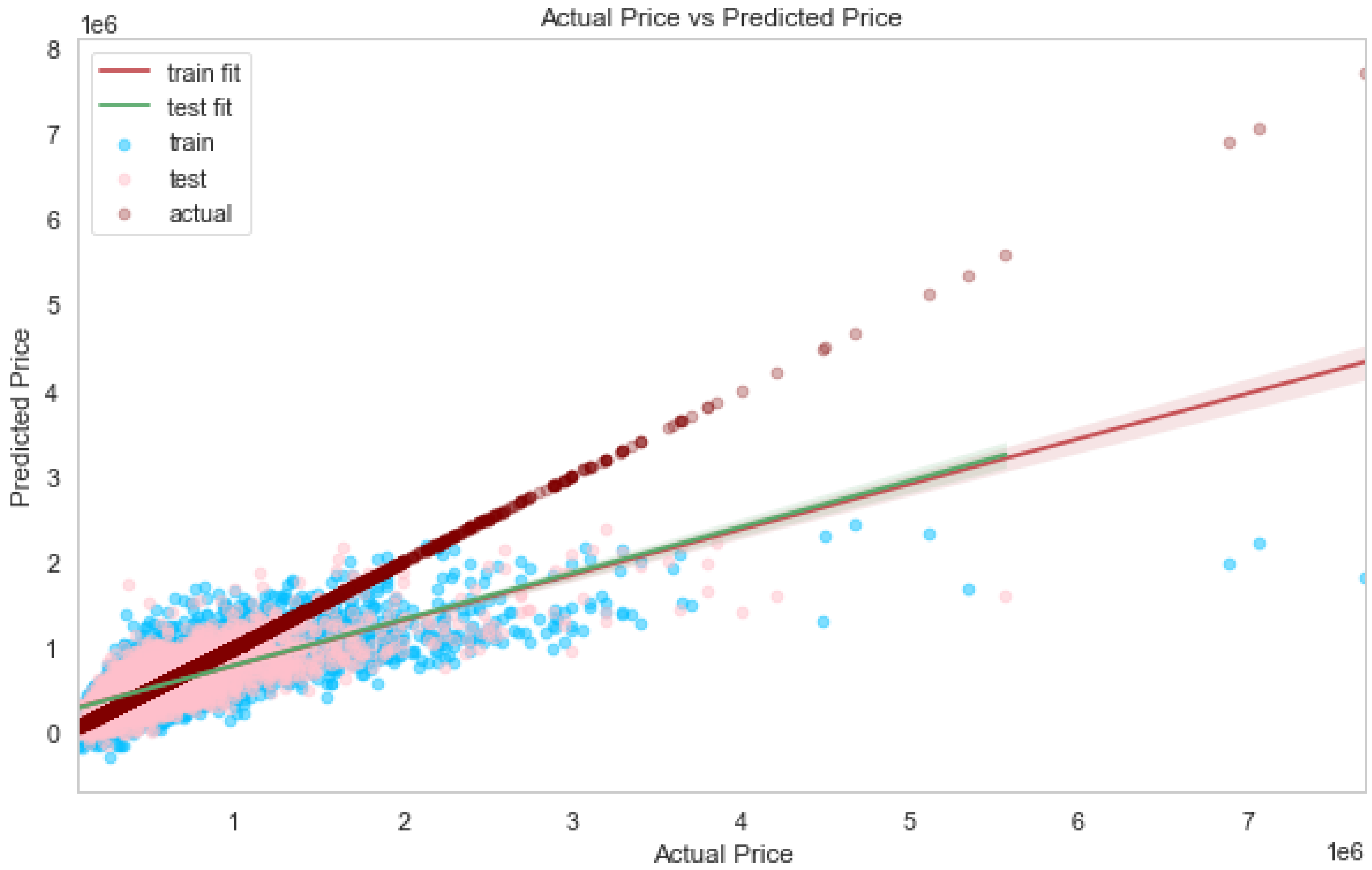


From this, we can see that price increases as the condition of the house increases

However, it appears that houses in poor condition are priced higher than those in fair condition

MODEL SELECTION

MODEL 1: Base Model



	Actual Values	Predicted Value	Difference	Percentage Difference
4000	282500.0	372241.105596	89741.105596	31.766763
13454	520000.0	399414.040805	120585.959195	23.189608
6911	546000.0	302448.202699	243551.797301	44.606556
12662	345000.0	468898.992303	123898.992303	35.912751
2924	265000.0	330622.165929	65622.165929	24.763081

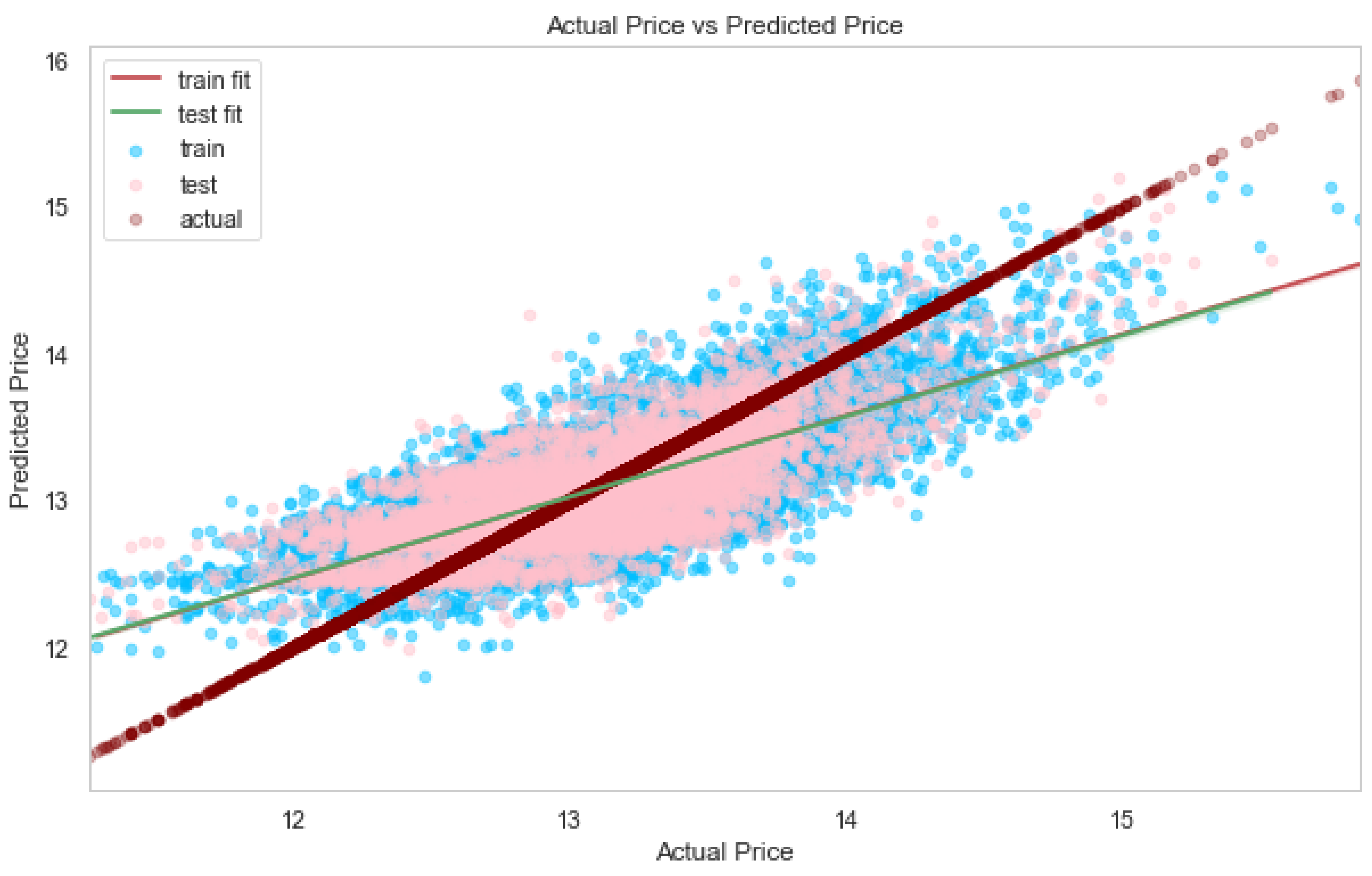
The model explains 54.92% of the variance in the target variable.

The error in the model's predictions lies between $163,711 and $157,450 respectively, which is quite large.

Overall, these metrics suggest that the linear regression model is not performing very well in making accurate predictions on the dataset.

MODEL SELECTION

MODEL 2: Log Model



55.06% of our features are explained by this model in relation to price

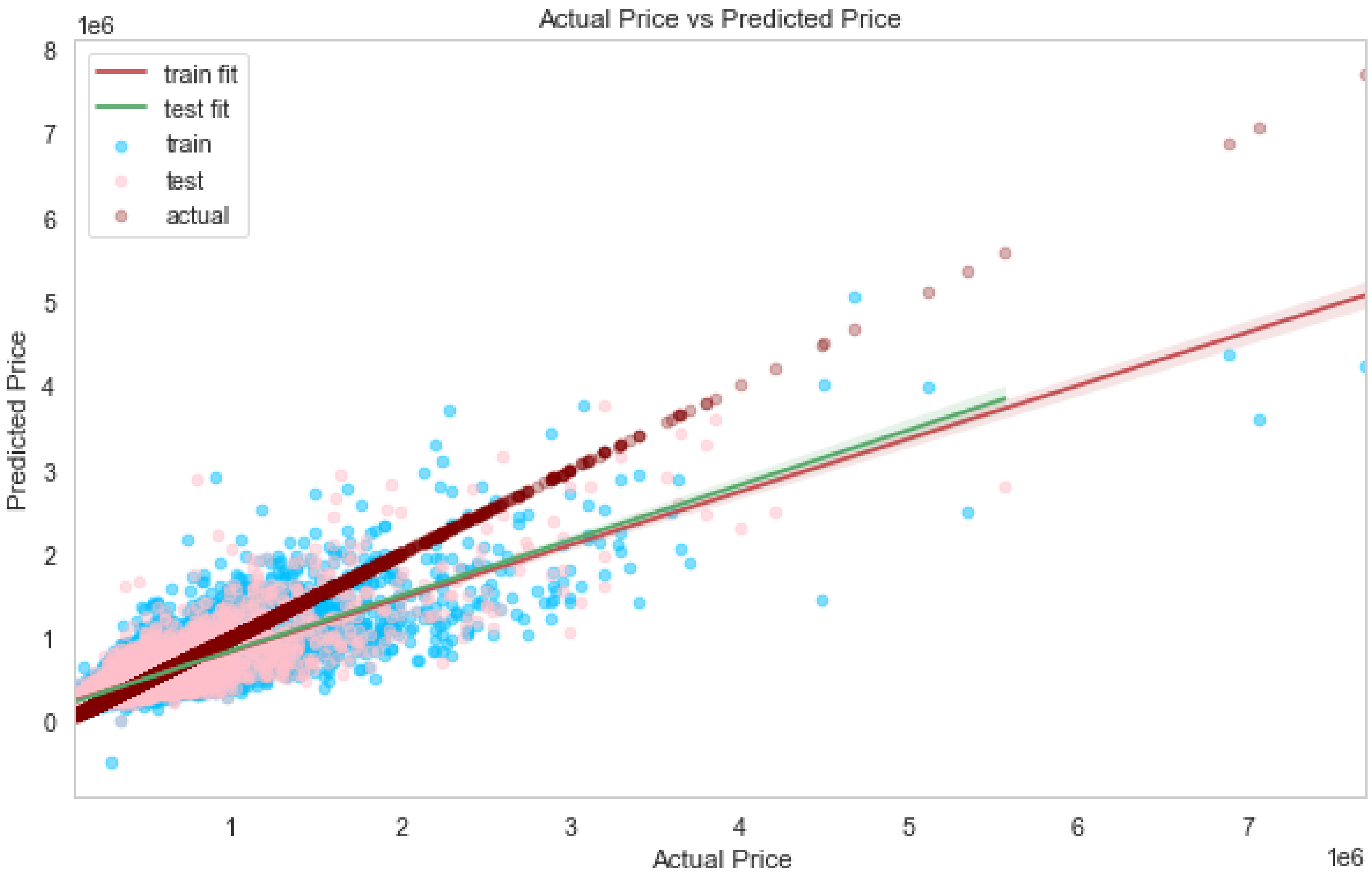
The model prediction is off by a range of 35.04% to 35.44% respectively from the logarithmic actual values.

based on our visual, we can see that train and actual regression are much closer compared to our initial base model.

	Actual Values	Predicted Value	Difference	Percentage Difference
4000	12.551434	12.831771	0.280337	2.233505
13454	13.161584	12.798907	0.362677	2.755574
6911	13.210374	12.624434	0.585940	4.435453
12662	12.751300	13.015178	0.263878	2.069419
2924	12.487485	12.733887	0.246402	1.973188

MODEL SELECTION

MODEL 3: Polynomial Model



- This model explains about 64.4% of the variation in the target variable which shows that the model is performing better
- The error of our predictions lie between \$147,506. 51 and \$149,612.02
- This is still significantly high but better than our previous two models

	Actual Values	Predicted Value	Difference	Percentage Difference
4000	282500.0	362263.664557	79763.664557	28.234926
13454	520000.0	356785.355940	163214.644060	31.387432
6911	546000.0	318987.634849	227012.365151	41.577356
12662	345000.0	428657.278879	83657.278879	24.248487
2924	265000.0	355090.340244	90090.340244	33.996355

CONCLUSIONS

- The pylonomial model in the given output performed reasonably well in predicting house prices, but there is room for improvement.
- The R-squared value indicates that the model explains 67% of the variance in the data, and on average, the model's predictions are off by $147,000.
- The high values of MSE, RMSE, and MAE suggest that the model's predictions may be quite far off for some instances. The Log Transformation Model and base Model were outperformed this model, but further improvements can be made by exploring other models or feature engineering.

RECOMMENDATIONS

- As a recommendation, it would be useful to perform further analysis to identify the features that are driving the model's predictions and consider incorporating new features that may improve the model's performance.
- Finally, it may be useful to gather more data to train the model and increase the size of the training set to get more accurate predictions
- Some features can also be considered when determining house prices. These features such as higher house grades, houses with better views, better condition, and the presence of a waterfront increase the value of the house. Kings


THANK YOU