

PHASE TWO FINAL PROJECT



Herbert LLC: Business Problem

Henbart LLC, a real estate developer in the Seattle area, has purchased undeveloped property outside of the Seattle metropolitan area. They are responding to the pandemic-era trend where there has been more demand for single-family homes, however, with inflation they want to make sure to build houses that will maximize their ROI

Top Three Features That Affect Price



Data & Methods

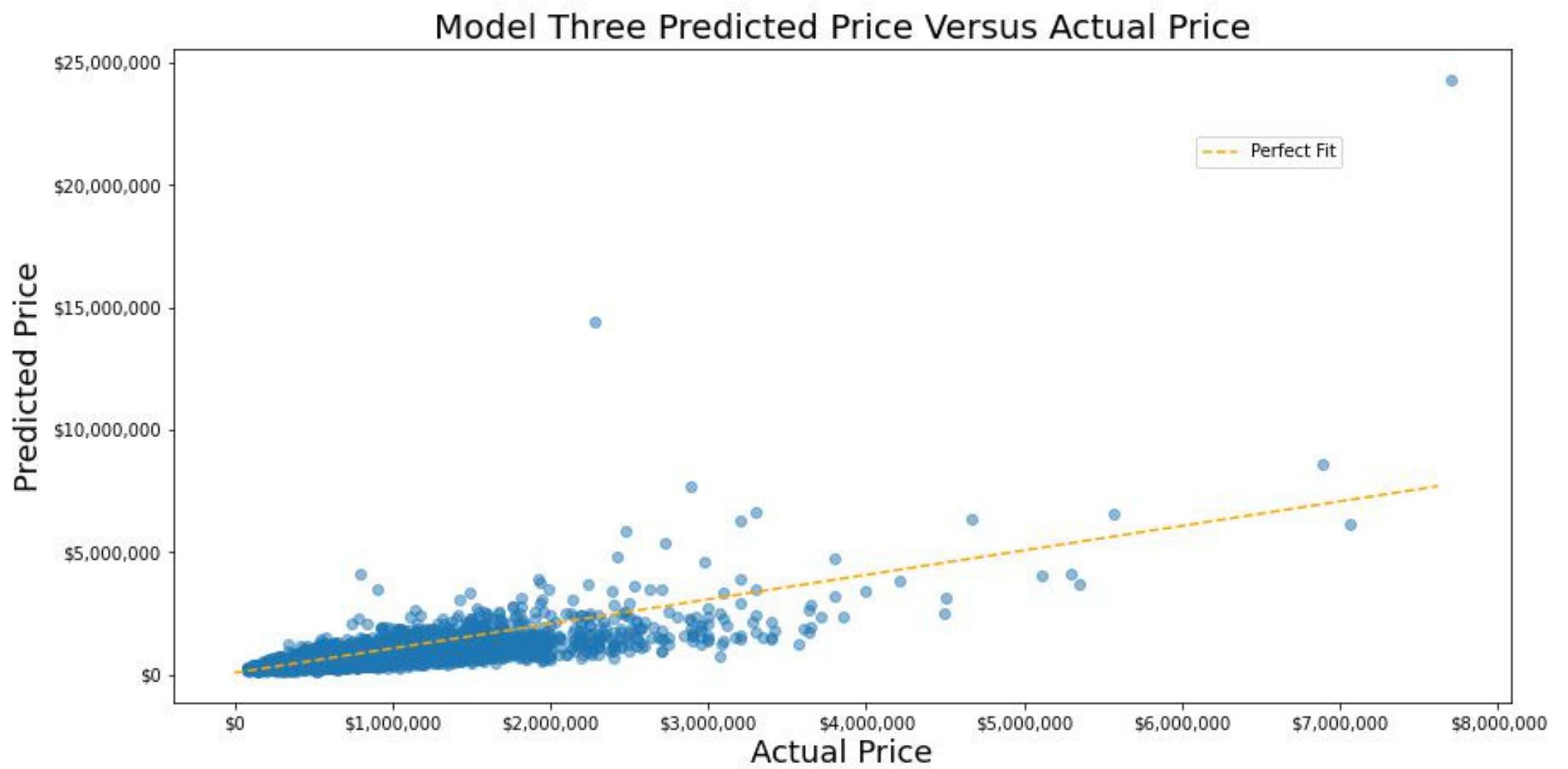
The Kings County House Sales Dataset is a dataset that contains the sale price and features of over 20,000 houses. Each house has 20 possible features

A linear regression model was developed to help predict house sale prices based on the features they possess.

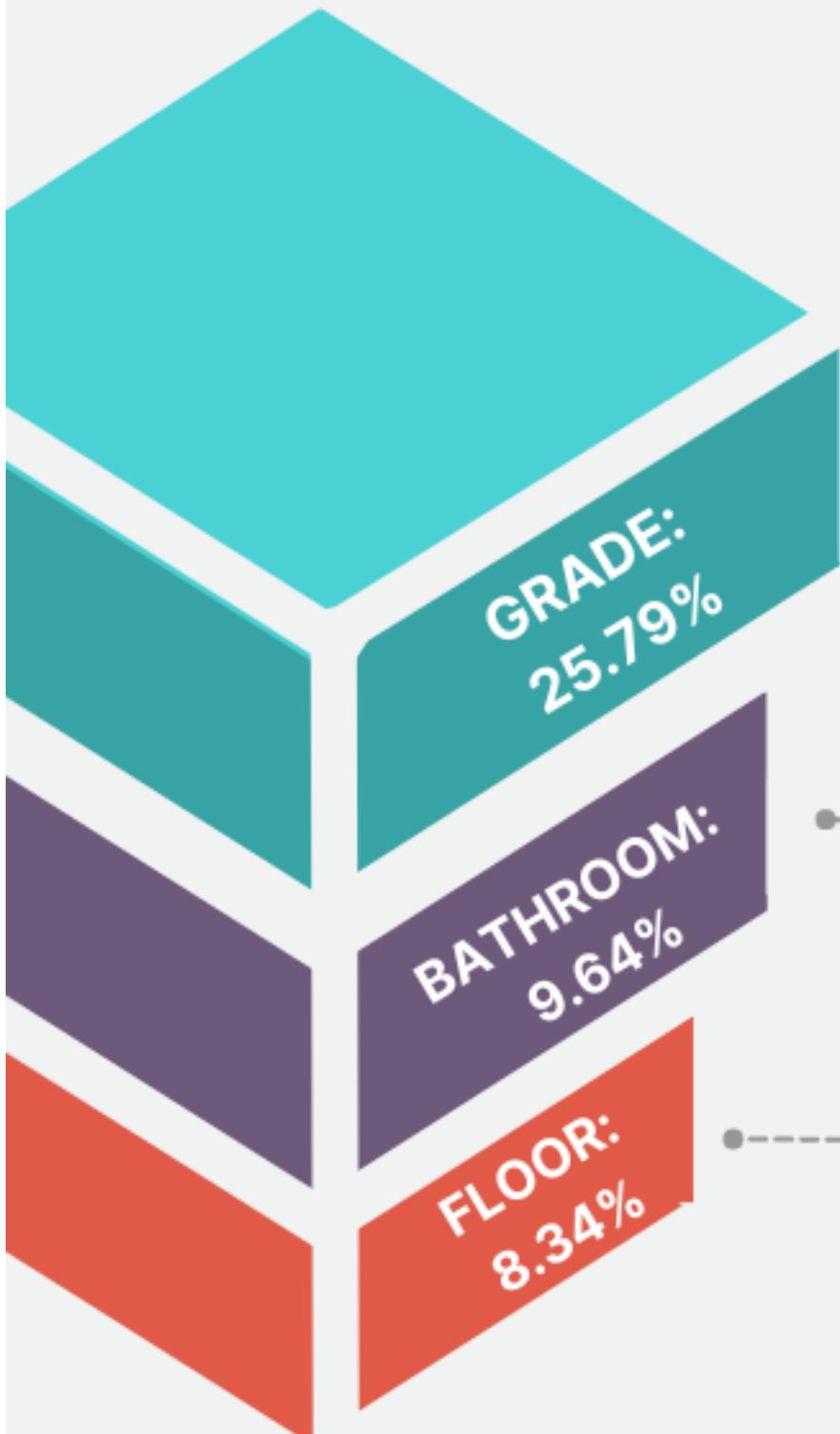


Results

R Squared Value was 63.5% with an error of \$260,920 when it comes to predicting the price of the house.



Top Three Features That Affect Price



According to the model developed a unit increase in the **Grade** i.e materials that go into the building of the house will yield an increase of **25.79%**



By adding an **Extra Bathroom**, the sale price increases by **9.64%**



By adding an **Extra Floor**, the sale price increases by **8.34%**

Conclusion

The model is not very accurate for this dataset. Not only is the error of the model very high, \$260,920, but the dataset does not fit well in a linear regression model.

Build quality, number of bathrooms, and number of floors in a house are the most statistically significant factors when it comes to increasing the sale price of a house.



Next Steps

- 1 Exploration of other modeling techniques that may be a better fit for the dataset
- 2 Further exploration of possible errors in outliers in the dataset
- 3 Add additional house entries to see if findings still hold true in a bigger dataset



ANY QUESTIONS?

Email: marshemily2@gmail.com

GitHub: [@marshemily2](https://github.com/marshemily2)

Linkedin: [linkedin.com/in/emily-k-marsh](https://www.linkedin.com/in/emily-k-marsh)

