Real Estate Home Improvement Price Predictions

Deztany Jackson
Data Science Flex
Phase 2 Project
January 6, 2023

Overview

Real Estate agents in King County, Seattle are evaluating the neighborhoods to encourage current home owners of he benefits of improving and upgrading their property value.

Housing data from King County was used to develop linear regressions models to support future price predictions

Agenda

- Business Problem
- Data Understanding
- Modeling
- Regression Results
- Recommendations
- Path Forward

Business and Data Understanding

- **Primary Stakeholder:** Real Estate Agency
- **Secondary Stakeholder :** Homeowners, Homebuyers & Businesses
- **Primary Use case:** Price difference after home improvements
- **Dataset Filtered:** 20 → 11 Features
- **Target Variable:** Home Price
- **Modified Prediction Variables:** Condition & Grade

Feature	Description
Home Price	Price will be out target variable. Price is the amount of the house in context of the current attributes.
Bedrooms	Number of bedrooms for the given home
Bathrooms	Number of bathrooms for the given home
Sqft Living	The size of the livable space in the house
Sqft Lot	The size of the lot
Floors	The Number of Floors.
Waterfront	Whether the house is on a waterfront
Condition	How good the overall condition of the house is. Related to maintenance of house.
Grade	Overall grade of the house. Related to the construction and design of the house.
Yr Built	Year when house was built
Zipcode	ZIP Code used by the United States Postal Service

Modeling

- A multi-linear regression model was created predicting housing prices . The model created to do this used an initial subset of data (from King County database) to process, train and test for this problem.
  - Developed using data science and python language
  - Developed over several iterations to refine accuracy

- The final accuracy metrics are “good enough” to use the model for basic home value predictions.
  - Model Error(Root Mean Square Error [RMSE Score]): .31
  - Prediction Accuracy(R2 Score): .66

Results

- Subset of homes were chosen which had space for improvement.
- Dataset filtered by the top five zip codes that have the most homes with
- Data Features Modified:

  Condition 3-Good or less → 4-Good
  Grade 6-Low Average or less → 8-Good

Average home value differences :

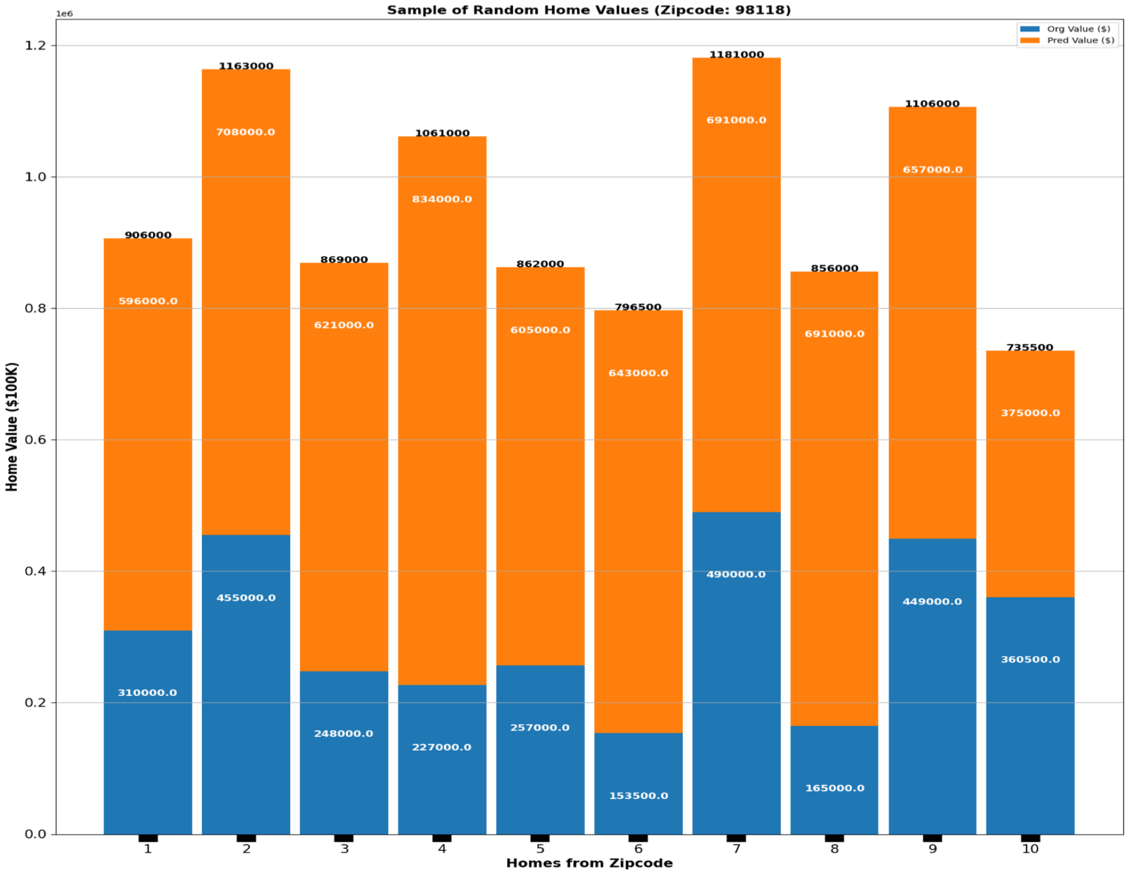
- Zipcode 98118: $ 323,500, %100 Increase
- Zipcode 98106: $ 304,100, %100 Increase
- Zipcode 98126: $ 266,500, %100 Increase
- Zipcode 98146: $ 324,300, %200 Increase
- Zipcode 98168: $340,300, %200 Increase

Zipcode 98188 home value differences:

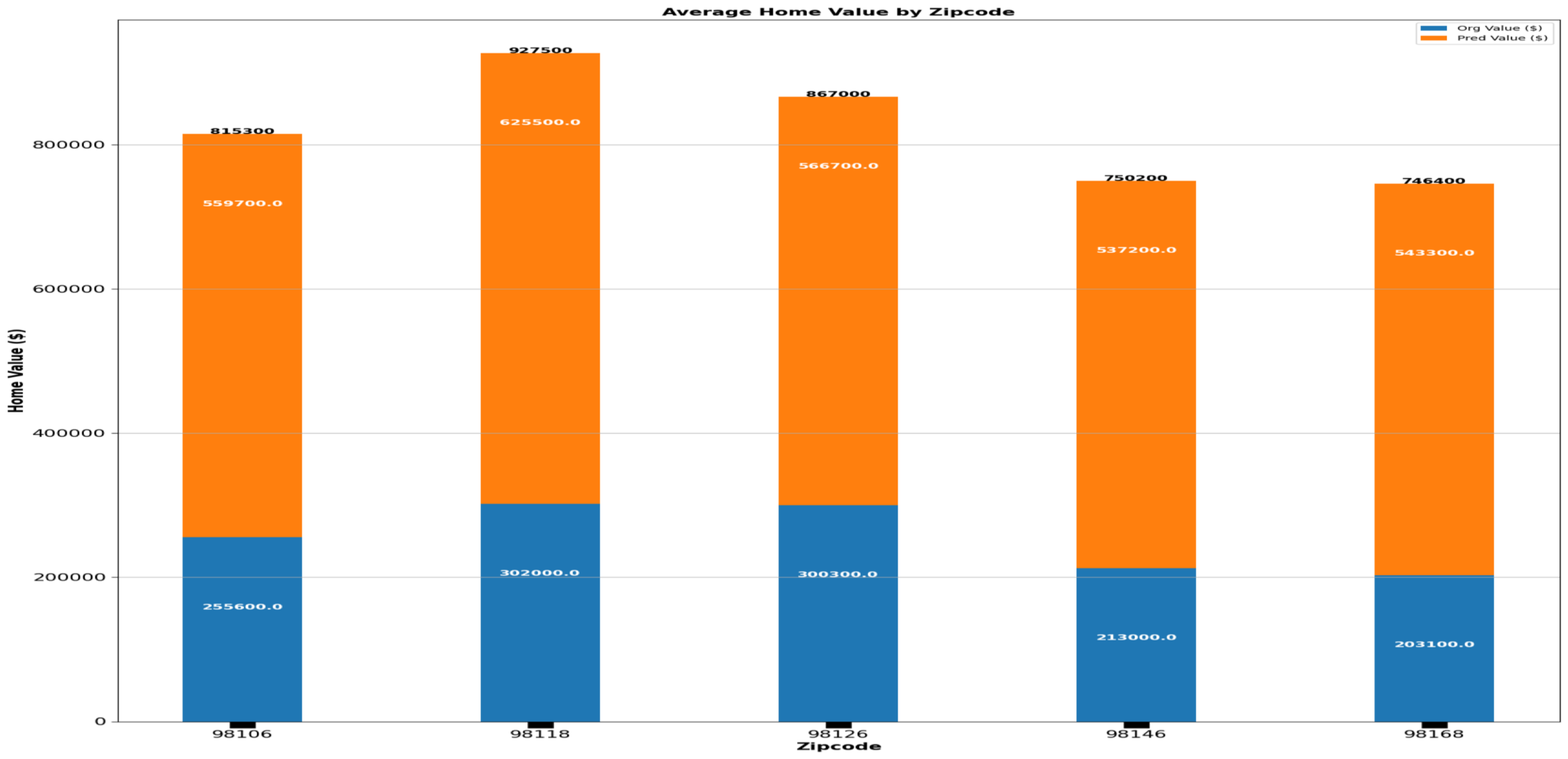
- Lowest $14,500 Increase*
- Highest $624,000 Increase

Zipcode 98118

- Displaying 10 of 106 entries meeting criteria



Top 5 Zipcode Averages



Conclusion: Limitations

- Limited dataset
- **Unknown realistic "Condition" and "Grade" values**
- **Unknown affects on other variables**
- Communal effects
- Model approach fit for specific problem
- Actual market culture
- External effects
- Time/Resources
- Data scientist skillset

Conclusion: Recommendations & Next Steps

- **Choose and present incentives and vision of home improvement within a community. This incentivizes people and helps with accountability.**
    - Choose a few zipcodes to try out and then offer feedback that would improve model accuracy or approach.
    - Focus on major areas with high needs to bring up. The communal effect will possibly increase prices even though the highest improvements may not have happened.
    - Present businesses(e.i. construction, remodeling) of the potential work to be done. Partnering with them and possibly offering discounts to the select communities would be a good incentive for communities to improve together

- **Market to potential homebuyers (individuals and investors) of the potential return on investment. These homebuyers may potentially buy the homes before the improvements and then fix them up.**

- **Increase consultation with the data scientist/analyst to improve our domain knowledge. As both parties educate each other the model solution has a better chance at being more accurate and robust.**
    - Feedback on realistic feature values after home improvement
    - Having examples and case studies of home improvements specifics would help give a realistic picture to all of the stakeholder supporting the predictions and work.

Thank You!

Email: dmvinedata@gmail.com
GitHub: @dmvinedata
LinkedIn: linkedin.com/in/deztanyjackson/