# 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.

Model: Multi-Linear Regression

Top Attributes: Grade, Condition & Sqft Living

# Agenda

- Business & Data Understanding
- Modeling
- Regression Results
- Limitations
- Recommendations
- Path Forward

# Business and Data Understanding

- **Stakeholder:** Real Estate Agency
- **Use case:** Improvements Home Value Change
- **Dataset Filtered:** 20 → 11 Features
- **Target Variable:** Home Price
- **Main Coefficients:** Sqft Living, Condition & Grade
- **Modified Prediction Variables:** Condition & Grade

| Main Features | Description |
|---|---|
| **Home Price** | Price will be out target variable. Price is the amount of the house in context of the current attributes. |
| **Sqft Living** | The size of the livable space in the house |
| **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. |

| Feature | Description |
|---|---|
| **Bedrooms** | Number of bedrooms for the given home |
| **Bathrooms** | Number of bathrooms for the given home |
| **Sqft Lot** | The size of the lot |
| **Floors** | The Number of Floors. |
| **Waterfront** | Whether the house is on a waterfront |
| **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:
  - Price Prediction Error (Root Mean Square Error [RMSE Score]): ~$197,000
  - Model Goodness( R2 Score ): 66%

# Results

- Grade Feature: For every one-unit increase in the Grade, the home value increases by about 26%.
- Condition Feature: For every one-unit increase in the Condition, the home value increases by about 3%.
- Sqft_living Feature: For every 1% increase in the sqft, the home value increases by about 4%.

- Data Features Modified  for New Predictions*:

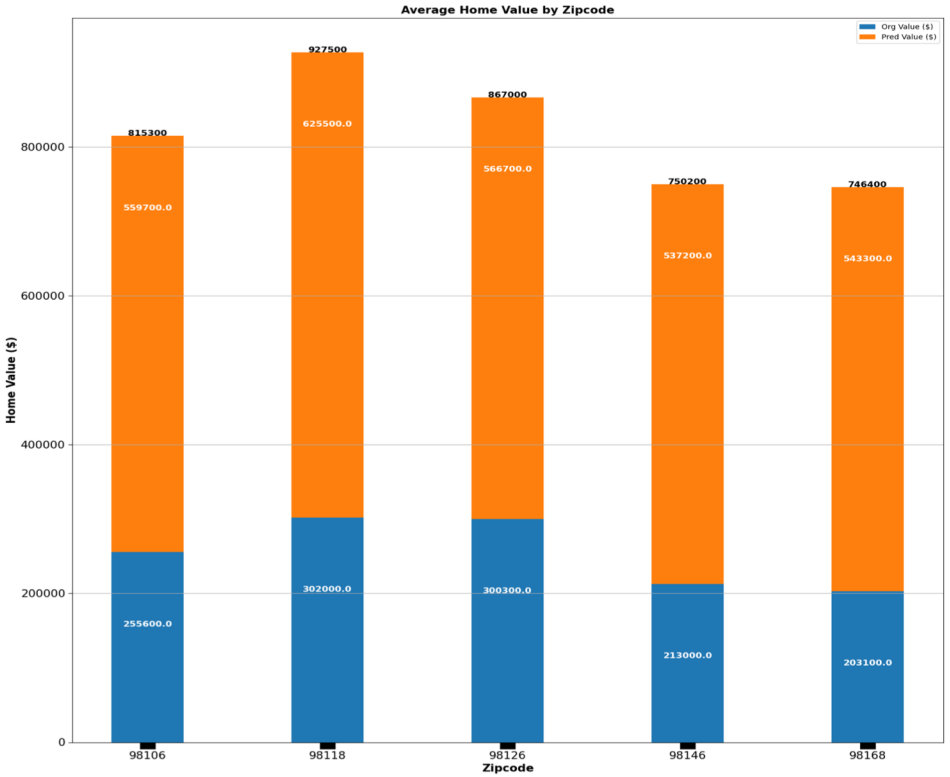
  **Condition: 3-Average or less → 4-Good**
  **Grade: 6-Low Average or less → 8-Good**

  ***Used Zipcodes to group homes**

# Top 5 Zipcode Averages

**Average home value differences :**

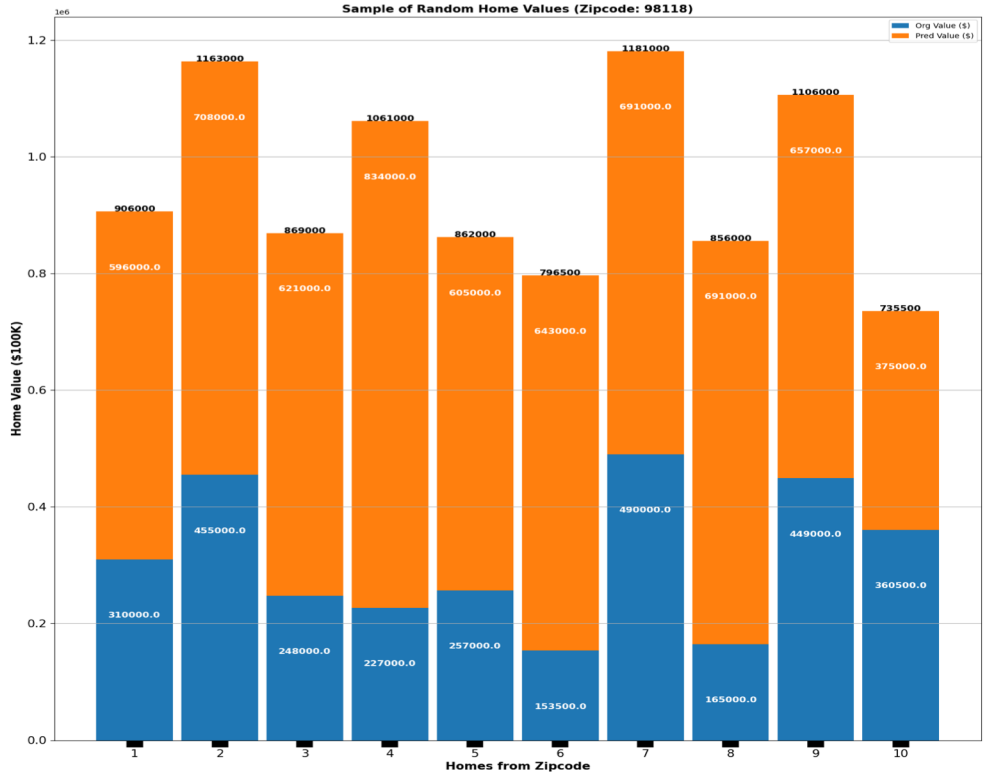
- Zipcode 98118: $ 323,500, 100% Increase
- Zipcode 98106: $ 304,100, 100% Increase
- Zipcode 98126: $ 266,500, 100% Increase
- Zipcode 98146: $ 324,300, 200% Increase
- Zipcode 98168: $340,300,  200% Increase



# Zipcode 98118

**Zipcode 98188 home value differences**:

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



# 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

**Low Hanging Fruit: Focus on zipcodes with major improvement needs to bring up home values.**

- At a minimum promote increasing conditions (e.g. basic maintenance) in the homes.
- Promote bring homes up to Average code (this increases the grade, which impacts home values the most).
- Promote ,increasing sqft living area by adding a small room (e.g. bathroom, small bedroom).

# Conclusion: Next Steps

- Choose and present  incentives and vision of home improvement within a community.
    - This incentivizes people and helps with accountability.
- 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.

# Thank You!

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