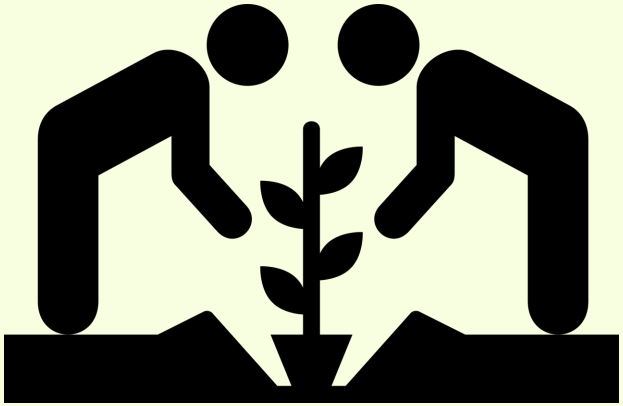
Growing King County

By: Michael Pozo & Nathaniel Green

Agenda

1. Business Objective
2. Data Understanding
3. Model Development
4. Regression Results
5. Recommendations
6. Conclusions
7. Next Steps



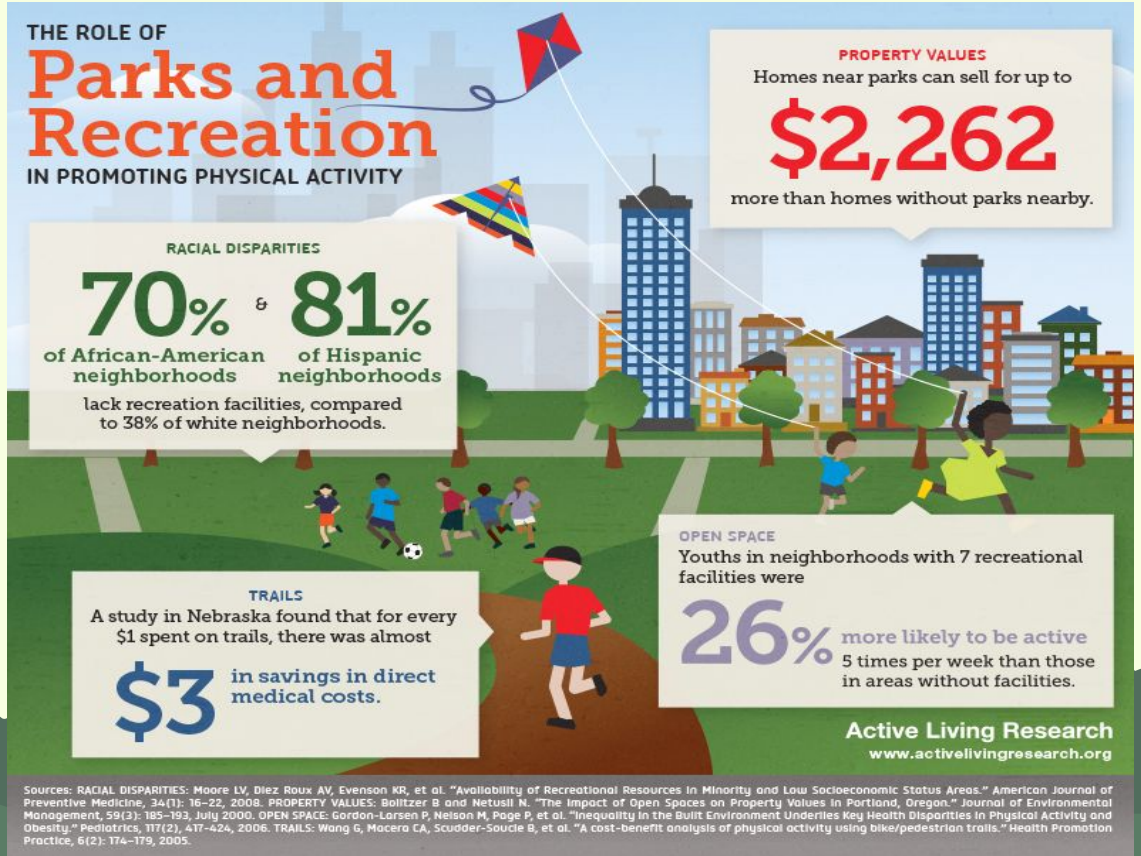
Business Objective

Problem: The King County parks department has approached our team to help build an argument for investing more in public green spaces and to identify opportunities to build parks/community gardens

Solution: Our proposed solution is to build a linear regression model to help the parks department evaluate home listings to find undervalued properties in priority neighborhoods that may be willing to work out a deal with the city/town

Why Does it Matter?

Parks, green spaces and community gardens improve home prices and daily life in towns and cities

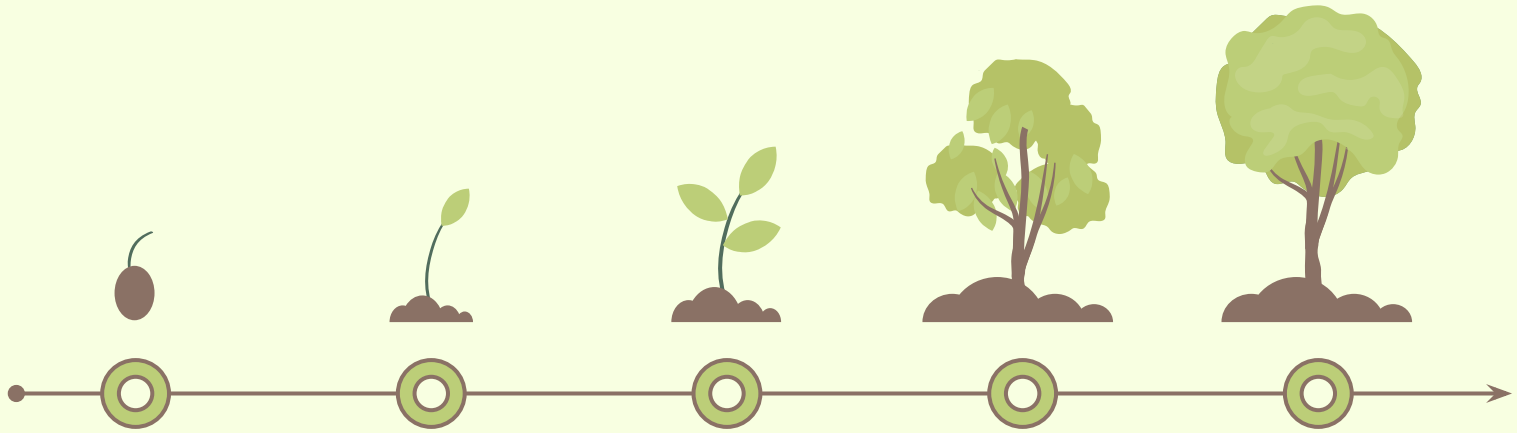


Data Understanding

Multiple Sources to Maximize Predictive Ability

Target Variable: House Price

Model Progression



EDA and Baseline Model

Add Numeric Interactions

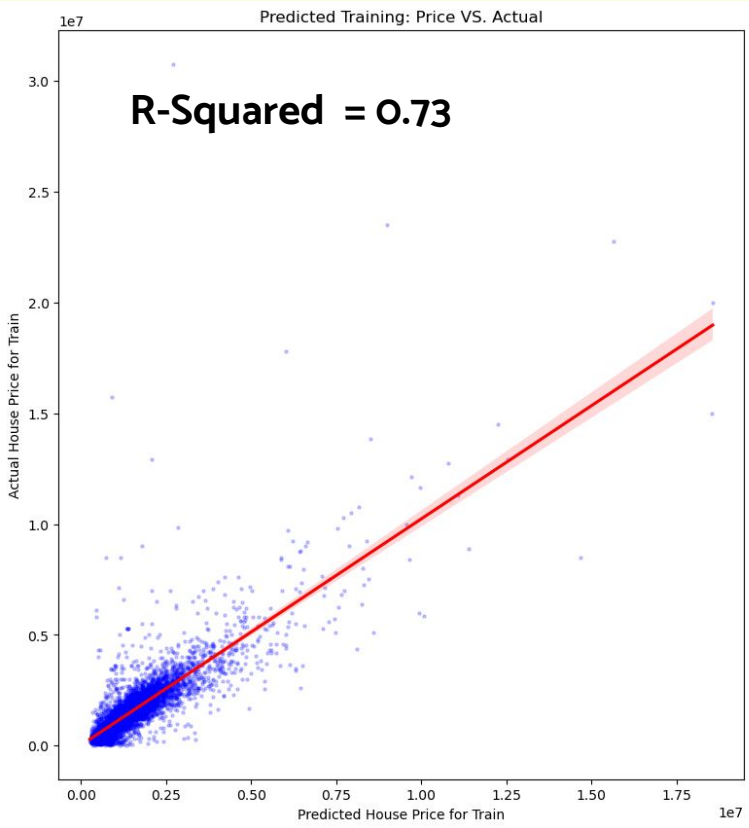
Encode Categorical Data

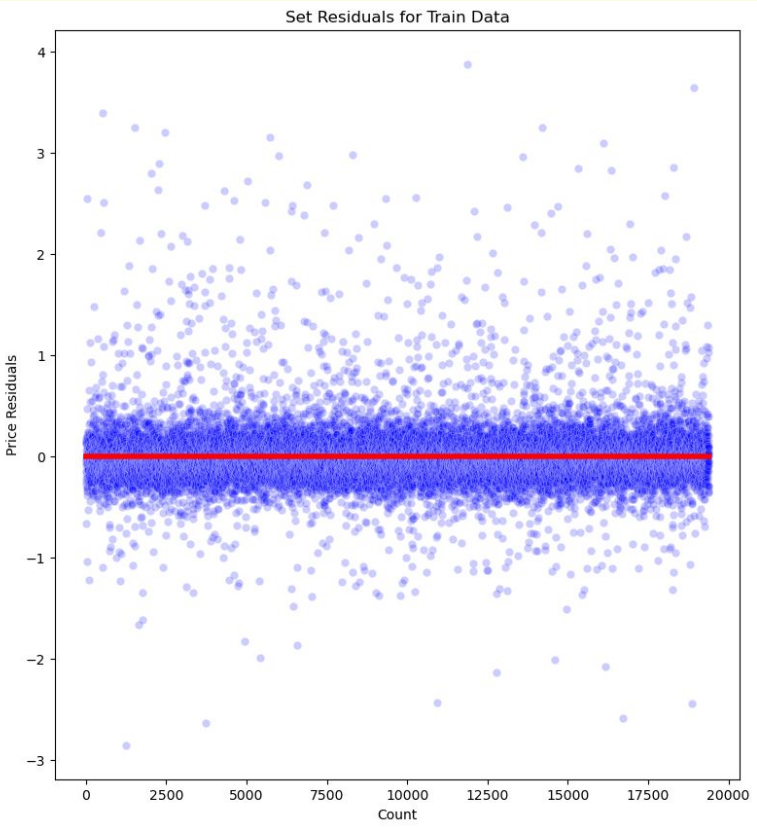
Add Education Data

Final Model

Model Training

Our training model was able to predict housing prices accurately





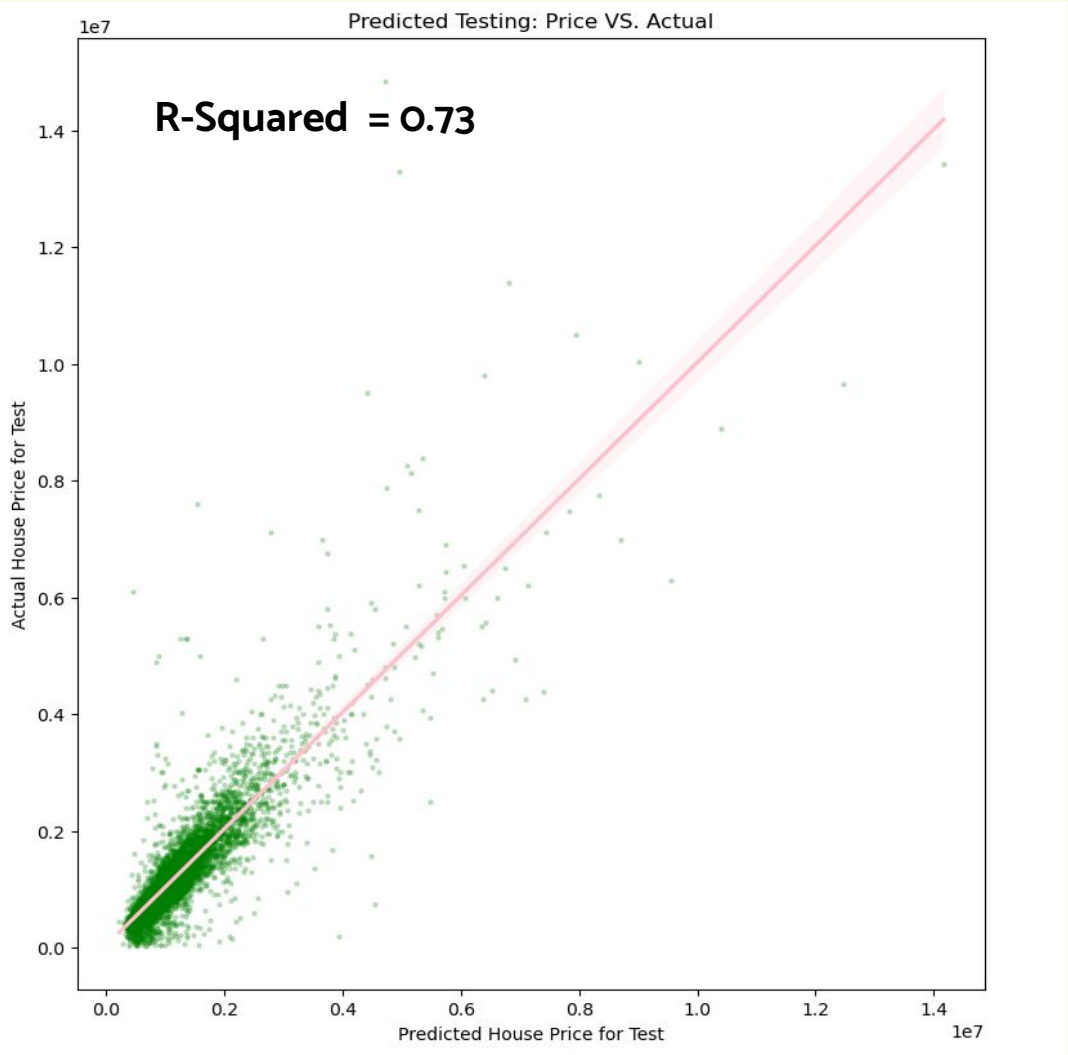
Final Model Results

The final model accounts for approximately **73%** of the variation of home sale prices in King County

Coefficient Highlights:

- **ZIP Code** is the single most influential feature in determining home prices
- Proximity to greenbelts also positively influences sale price

Takeaway: building green spaces in low access ZIP Codes can have a compounding effect building value in neighborhoods

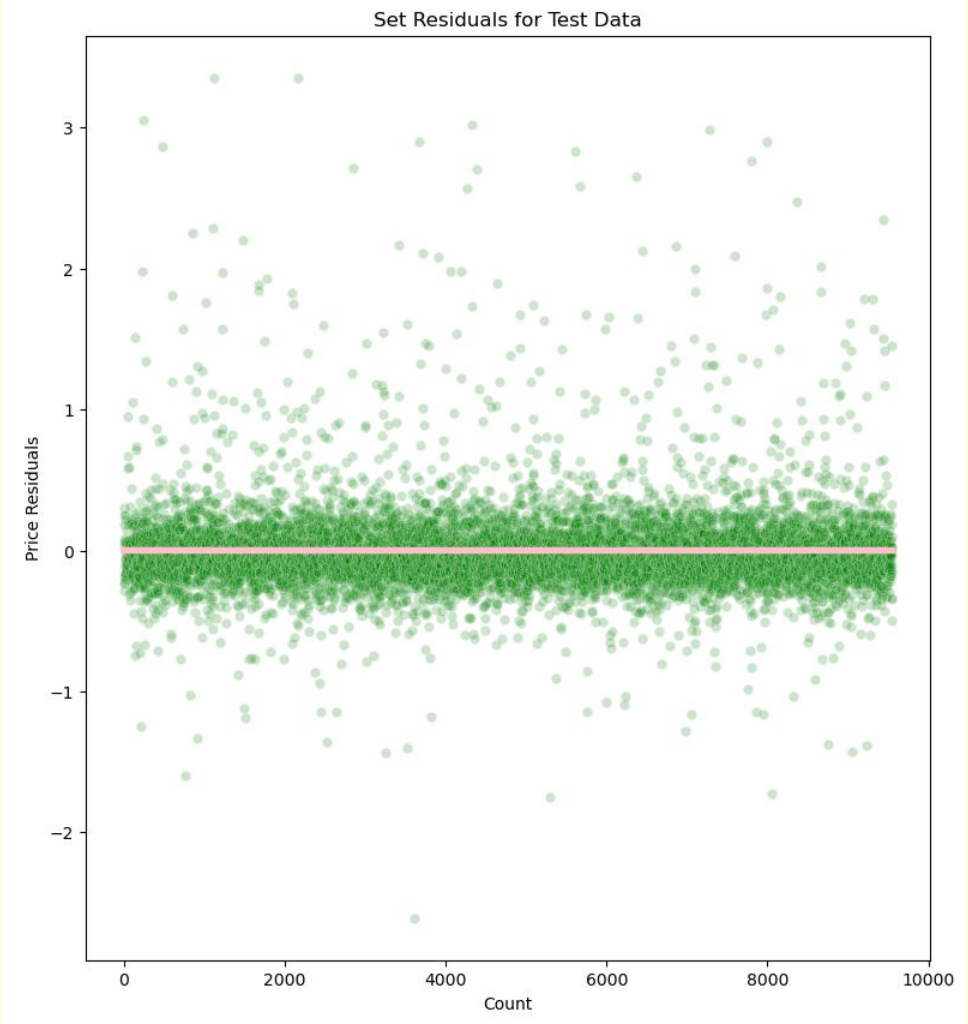


Final Model Accuracy

On average, the final model predicts home prices within $192, 704 of the actual test sale price

Based on the plot of the residuals, we can see that the majority of the residuals are clustered within the same area between -1 and positive 1 (units: log(price))

The spread tells us that the model consistently predicts price with the same accuracy across prices and other features

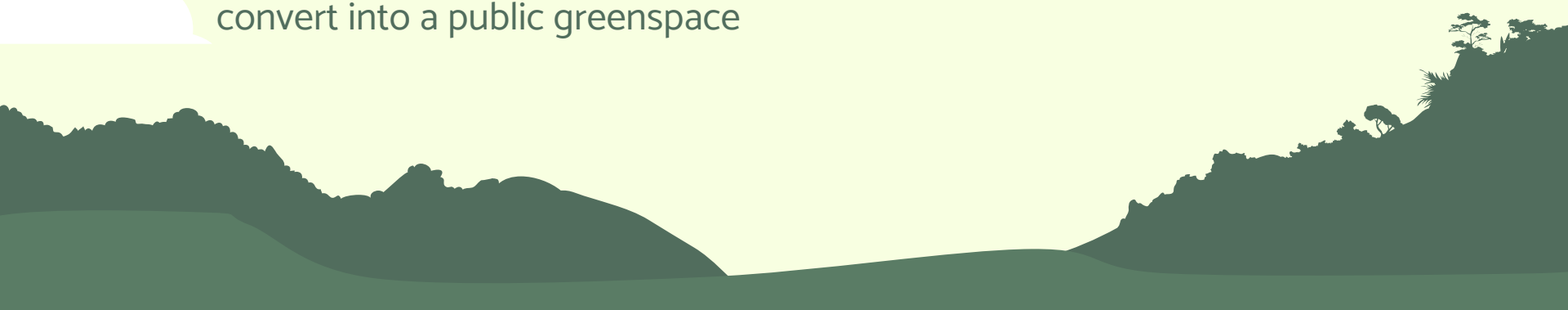


Recommendations

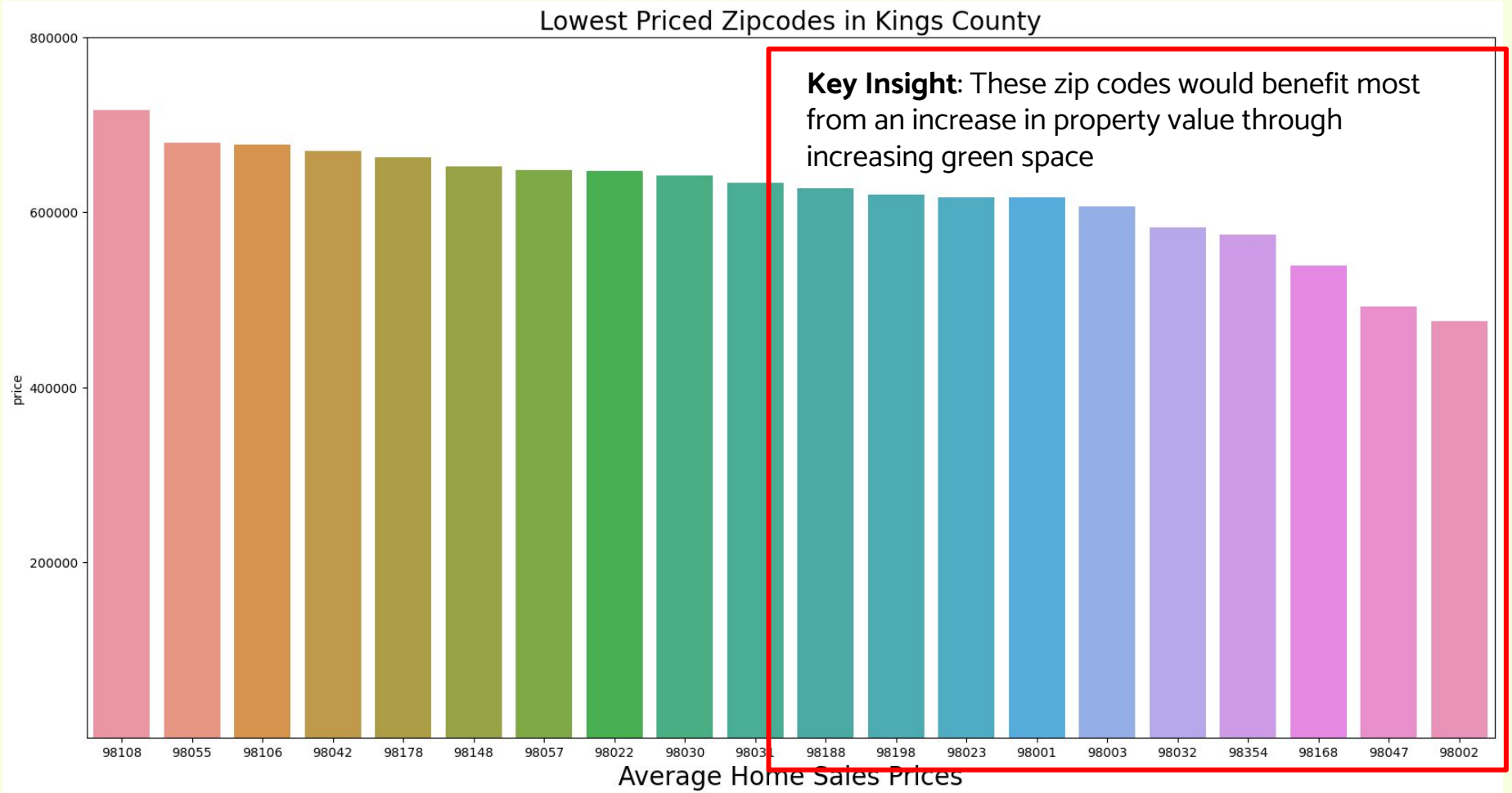
- Model results validate that proximity to green space increases property values. This can be used as a supporting piece of evidence when applying for funding

- The town's parks department should prioritize ZIP codes with low access to greenbelts and low property values relative to the rest of King County

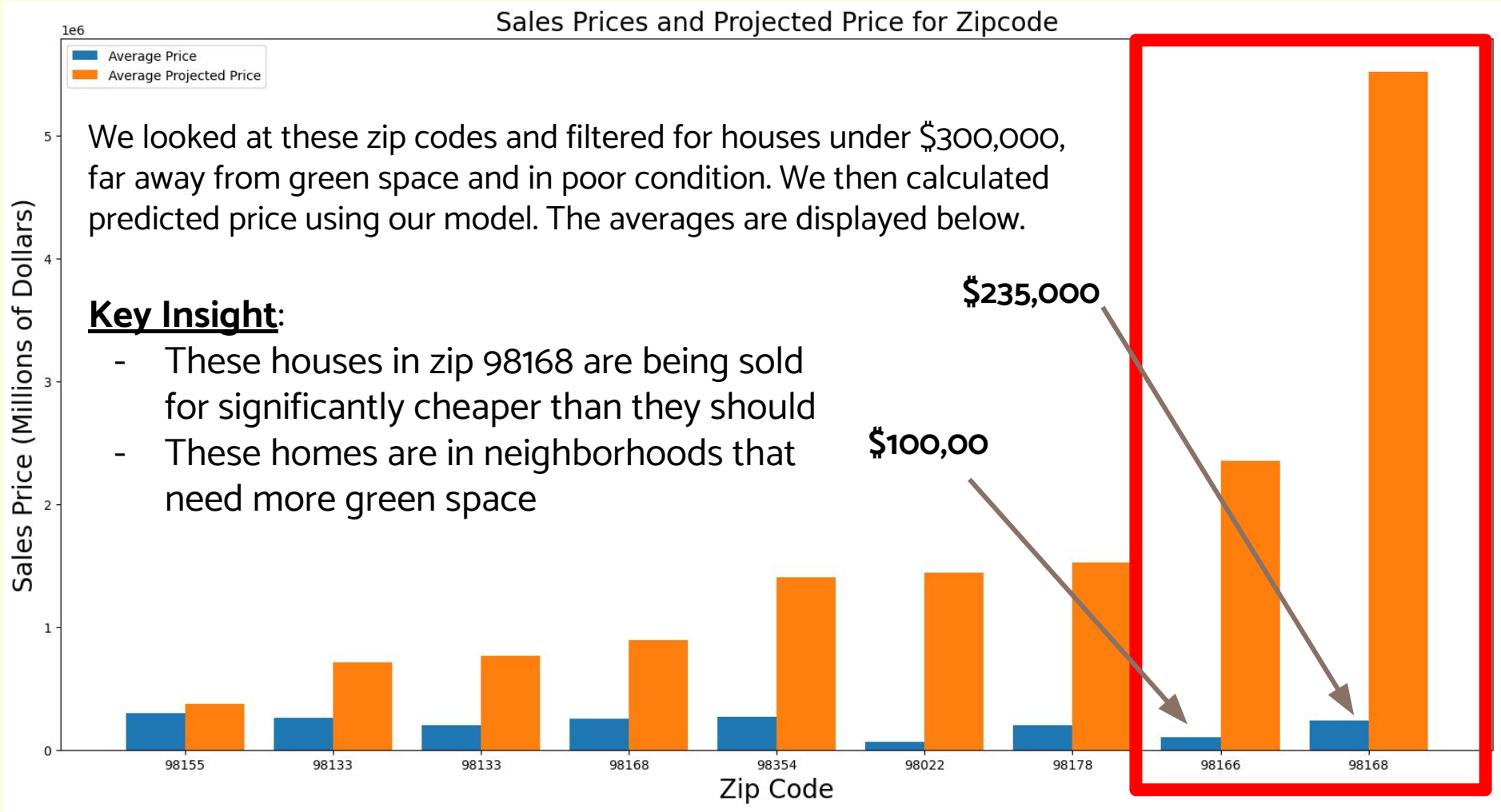
- Our model can be used to accurately evaluate home listings to determine whether a home is being sold below value and may be a potential target to convert into a public greenspace



ZIP Codes with the Lowest Average Property Values



Key Insight: These zip codes would benefit most from an increase in property value through increasing green space


Sales Prices and Projected Price for Zipcode
1e6
Average Price
Average Projected Price
Sales Price (Millions of Dollars)
0
1
2
3
4
5
98155
98133
98133
98168
98354
98022
98178
98166
98168
Zip Code
We looked at these zip codes and filtered for houses under $300,000, far away from green space and in poor condition. We then calculated predicted price using our model. The averages are displayed below.
Key Insight:
- These houses in zip 98168 are being sold for significantly cheaper than they should
- These homes are in neighborhoods that need more green space
$235,000
$100,00

Next Steps

1. Add other environmental data to our model with the hope of getting clearer insights on how greenspace affects price

2. Rework model using environmental data to predict other indicators of a healthy community (student test scores, medical costs, crime rates)

3. Expand EDA to categorize homes as empty lots or abandoned. This may require additional housing data

Thank You!

Email:
ngreen151@gmail.com
LinkedIn:
www.linkedin.com/in/nathaniel-green-14178a249

Email:
mikepozo17@gmail.com
LinkedIn:
www.linkedin.com/in/michael-pozo-546a10138