Real Estate Price Prediction

BY: SONAM MEHTA

CONTENTS:

Introduction:

1. In this project, I have built a machine learning model to predict the house prices.
2. House sale prediction is a crucial application of data science and machine learning in the real estate industry.
3. By leveraging historical data and advanced analytical techniques, we can forecast property prices with increasing accuracy, benefiting buyers, sellers, investors, and policymakers alike.

Objective:

1. Main aim is to accurately estimate the future selling price of residential properties using various data analysis and machine learning techniques for better transparency.
2. It will serve various purposes:-
   - For Buyers and Sellers
   - For Real-State Industry
   - For Investors

3. To consider factors influencing house prices, including property traits, location, economic data, and social sentiments.

Problem Statement:

There are many significant challenges in accurately determining the fair market value of residential properties due to the complex interplay of numerous factors affecting house prices.

1. Inefficiencies in the market, with properties often mispriced.
2. Real estate professionals without data-driven insights cannot advise clients properly.
3. Increased risk for buyers, sellers, and investors.
4. Less transparency in real estate market.

Data Preprocessing:

The dataset contains raw housing data with 5,000 records and 16 features. It includes various features such as property characteristics, location details, pricing information, house age to represent historical records.

To prepare the dataset for analysis or modeling here are the steps:

- Data Cleaning
- Feature Engineering
- Outlier Detection
- Data Splitting

Correlation Heatmap

1. sqrt_ft & total_rooms (0.74): Strong positive correlation between square footage and total number of rooms

2. sold_price & sqrt_ft (0.57): Strong positive correlation between house price and square footage

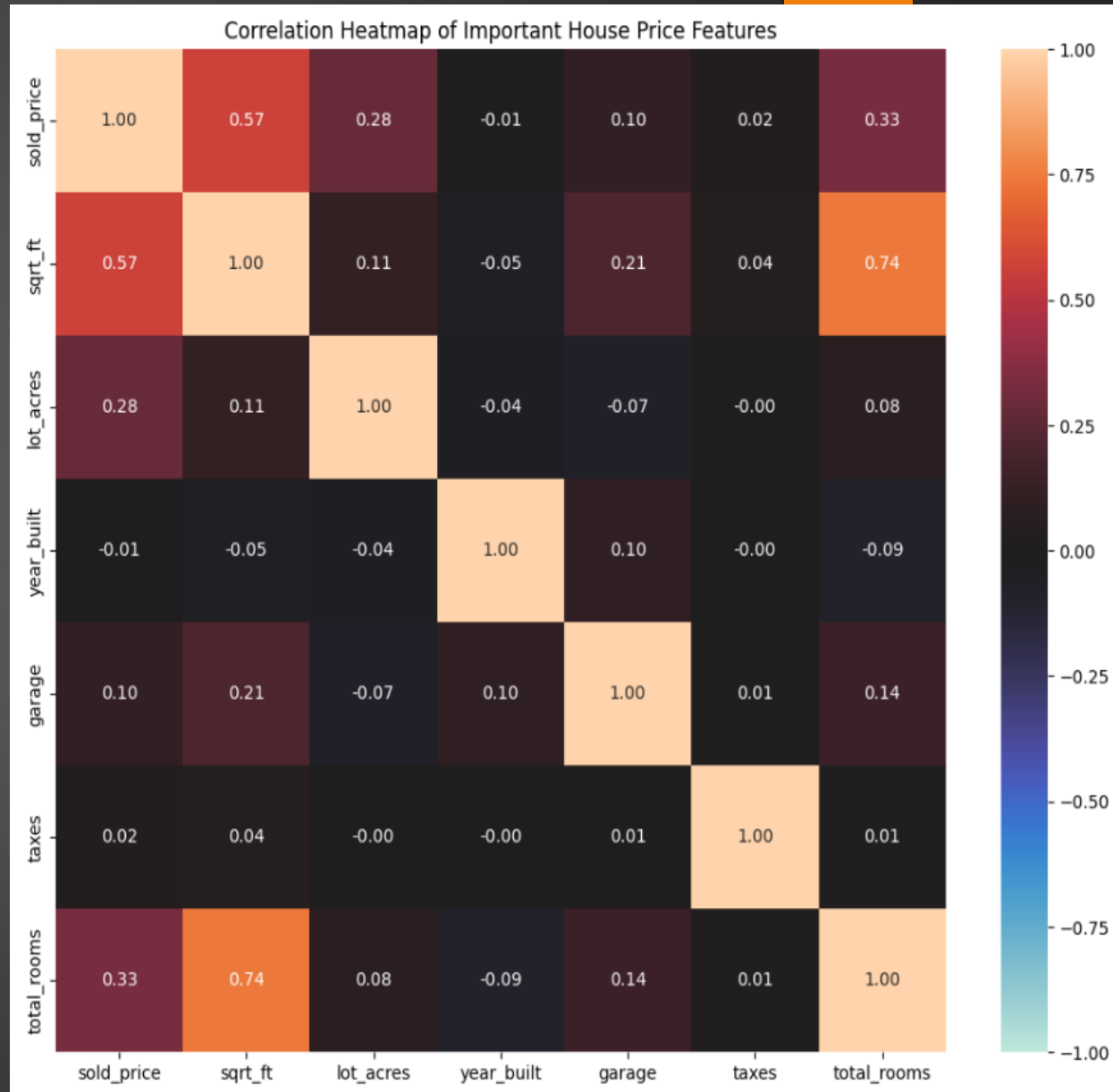
3. sold_price & total_rooms (0.33): Moderate positive correlation between price and number of rooms

<u>Key Insights</u>

- Square footage is the strongest predictor of house price

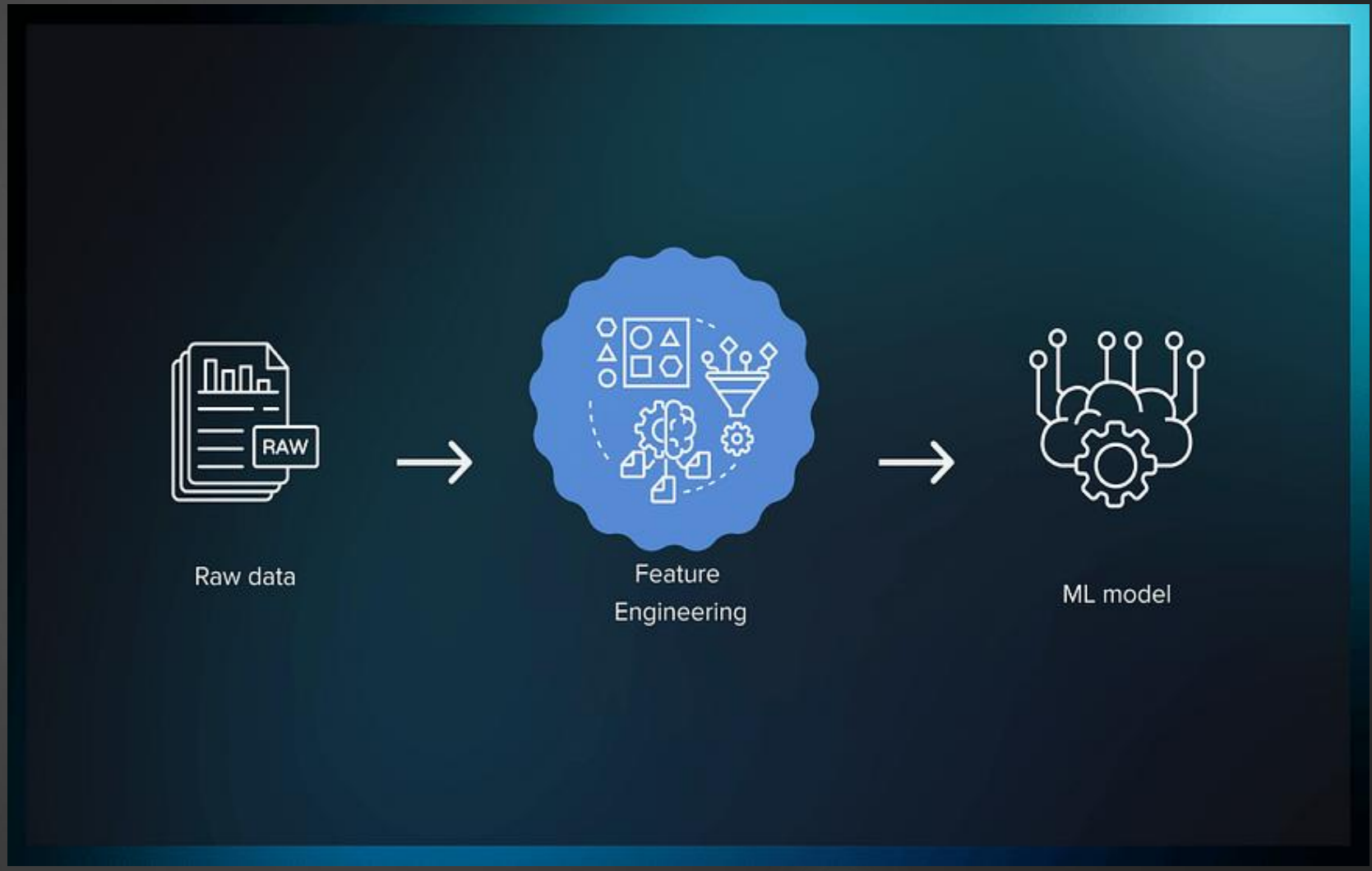
- Total rooms and lot size have moderate influence on price

- Taxes and year built have minimal correlation with price



Feature Engineering:

- Feature engineering is the process of using domain knowledge to create or transform variables (features) that make machine learning algorithms work more efficiently. It involves selecting, modifying, and creating new features from raw data to improve the model's accuracy and performance.



Steps performed in project:

1. **Data Type Conversion**
2. **Feature Creation**

   - A new 'rooms' feature created summing 'bedrooms' and 'bathrooms'

   - 'Price per square feet' calculated using 'price' and 'square feet'

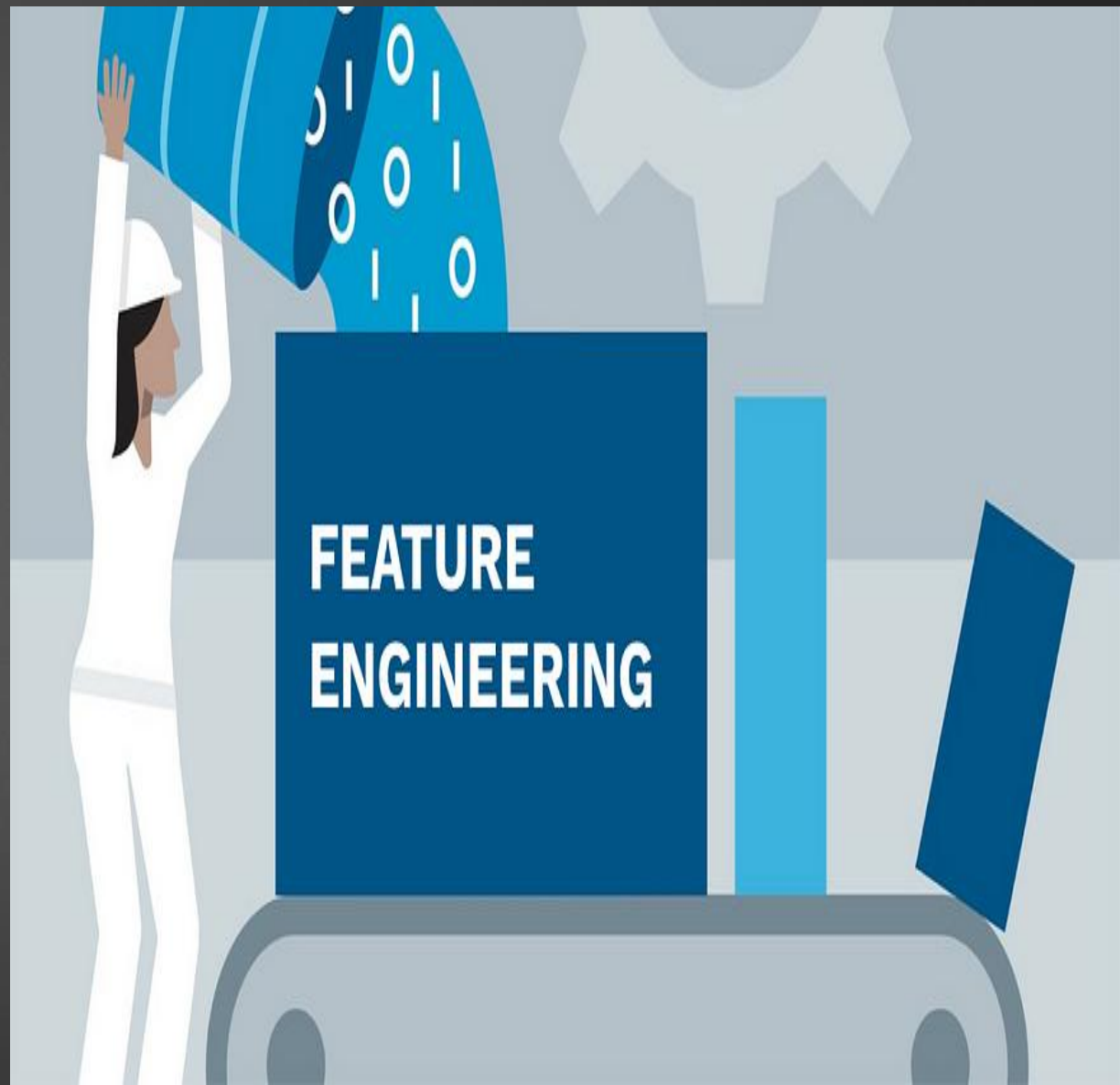
3. **Feature Removal**

   - 'bedrooms' and 'bathrooms' feature dropped after creating 'rooms'

4. **Binning**

   - 'Price per square feet' binned into 20 quantities

5. **Scaling**



Model Selection:

A model helps us to measure patterns and relationships within data. Major usage of models are:

- **Predictive power**
- **Pattern Identification**
- **Decision Making**

Models used in my project are-

1. **KNN Classifier**
2. **Gradient Descent**

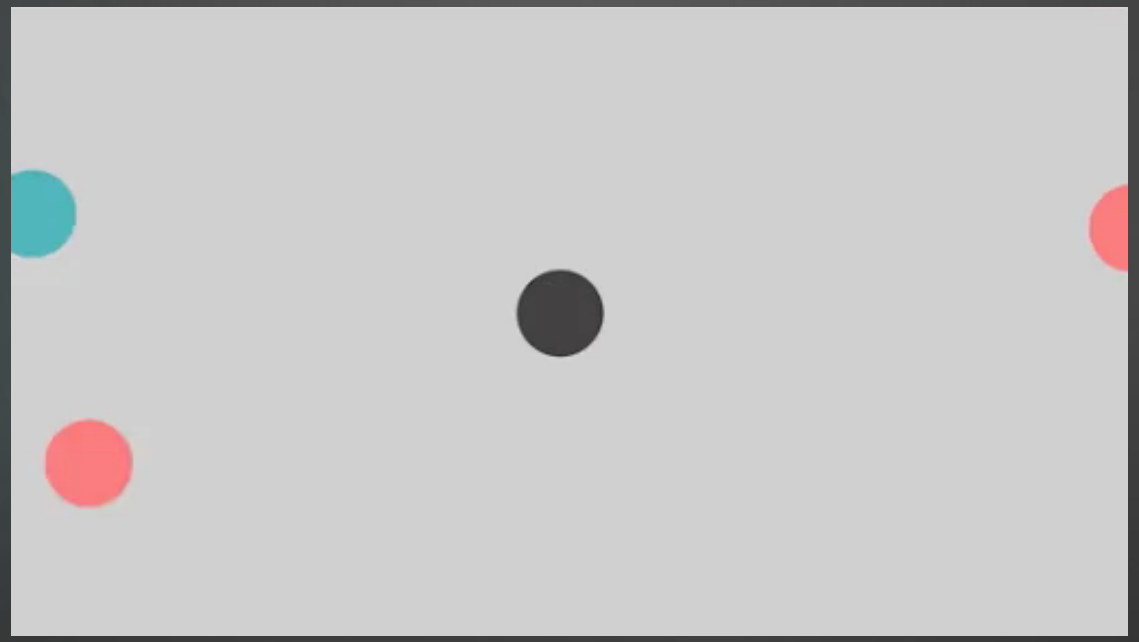
Let's Predict

KNN Classifier:

K-Nearest Neighbor (KNN)

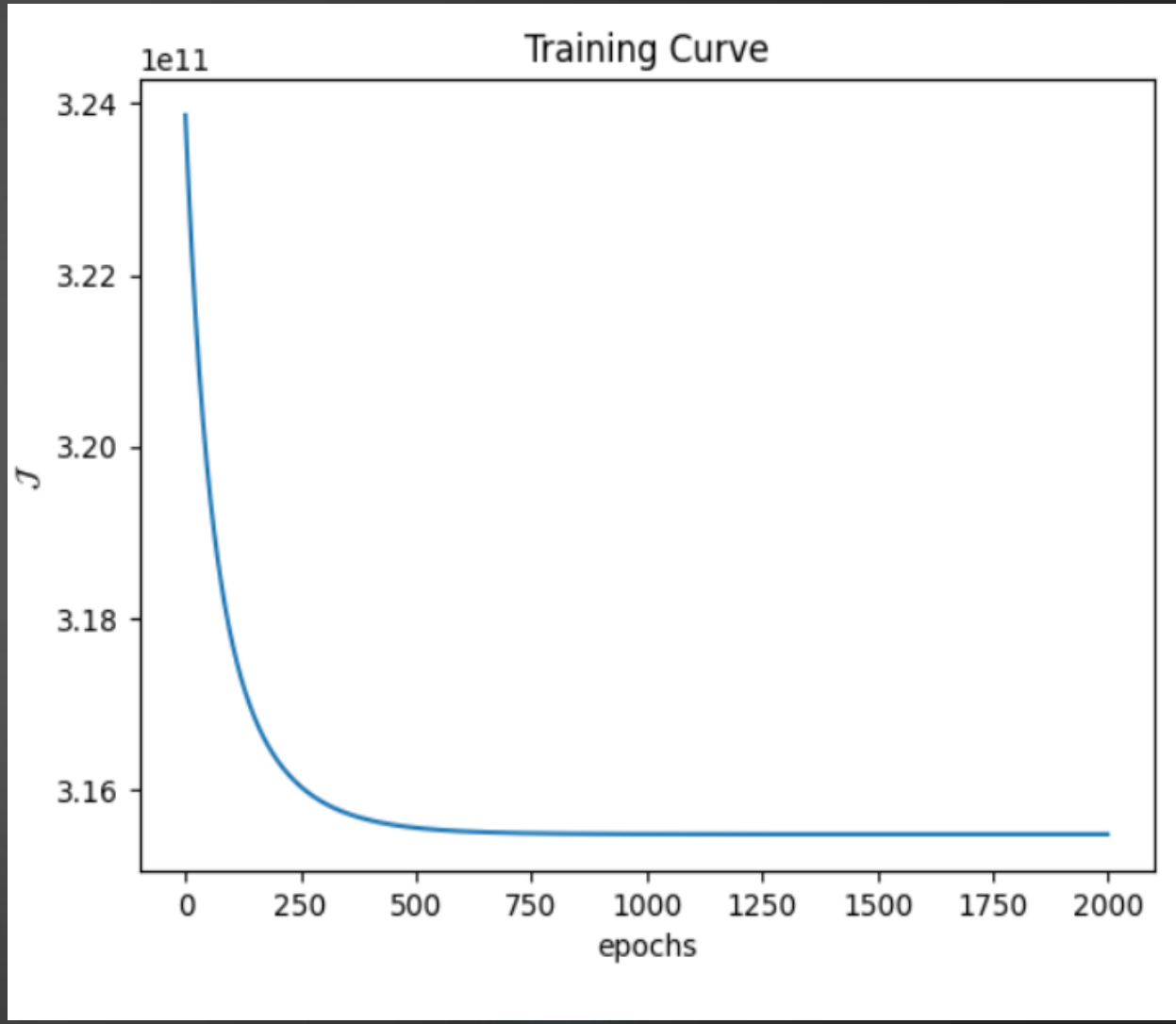
- KNN is a supervised learning algorithm capable of performing both classification and regression tasks.

- The algorithm stores the entire dataset and classifies each new data point based on the existing data points that are similar to it.



OLS Regressor

- Gradient Descent is an optimization algorithm for finding a local minimum of a differentiable function.
- Used to find the values of a function's parameters (coefficients) that minimize a cost function as far as possible.
- Some major hyper parameter like epochs and eta which can be used for optimal performance.
- MAE: 1652.6499
- R2: 0.0258



Conclusion:

Current Findings

- Our OLS (Ordinary Least Squares) regression model shows limitations in predicting house prices accurately
- The high loss indicates need for more sophisticated modeling approaches

Next Steps

Alternative Models to Explore:

- Neural Networks
- Random Forest Regression

THANK YOU!