YICHEN HU

# MODEL TO PREDICT HOUSING PRICES IN AMES, IA

# OBJECTIVE

- Create a model that can be used to predict home sale prices in Ames, Iowa
- Evaluate whether the model is successful at predicting prices

# DATASET

- Home sales from 2006 - 2010
- 2051 rows
- 79 columns or “features” for each sale
  - Expected features such as “Overall Quality”, number of bedrooms/bathrooms
  - More detailed features such as the quality of the material on the exterior of the house

# DATA ANALYSIS AND PROCESSING

- Missing values
- Analyzing / removing columns
- Categorical Data
- “Interactions”

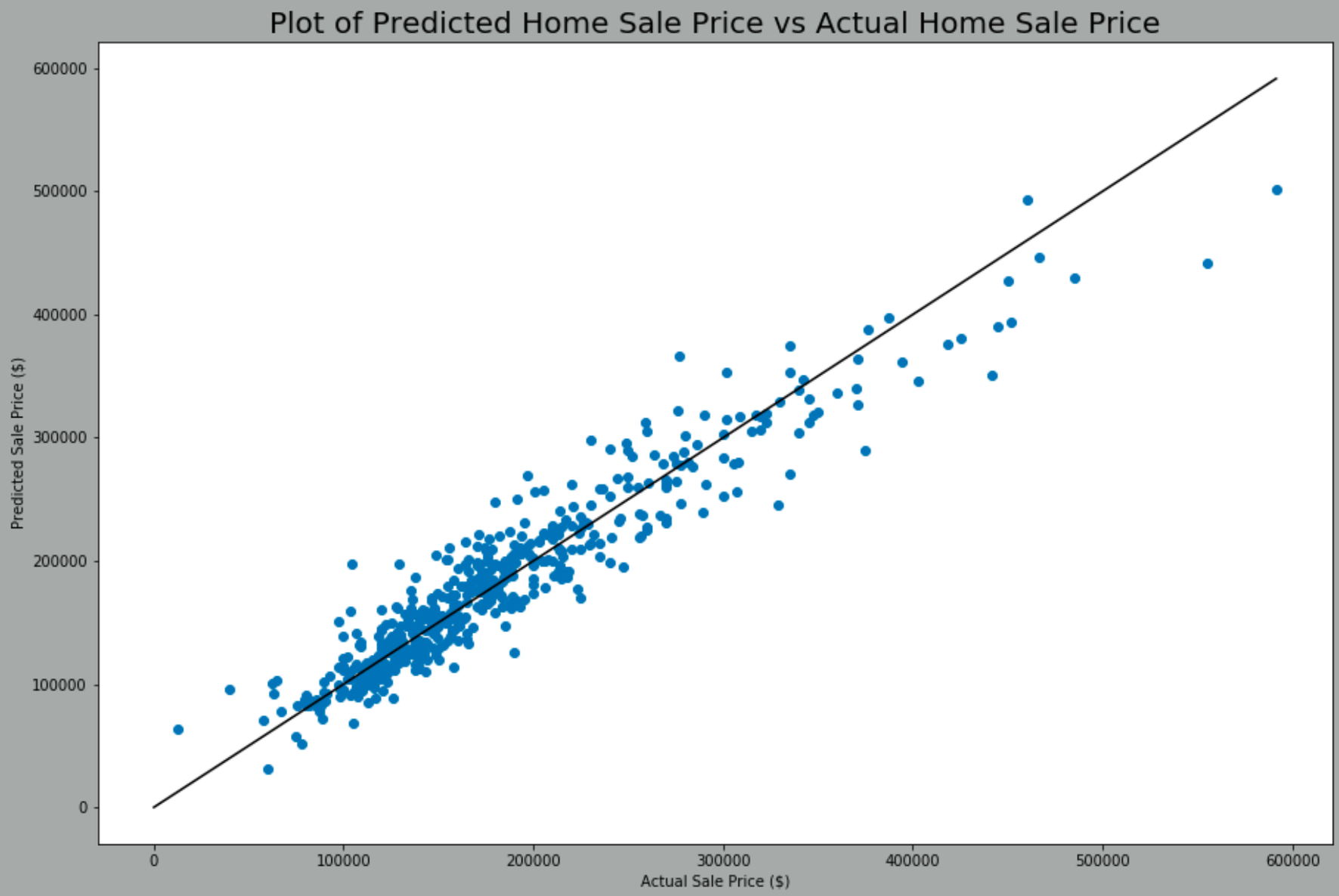
# SPLITTING OF DATA

- In order to properly evaluate the model, I split my dataset into two parts
- 75% of the dataset was used to build or “train” the model
- The remaining 25% was held out and then used for model evaluation

# EVALUATION OF MODEL

- Numeric metrics
  - $R^2$ score - .9018
  - RMSE - $24,993

# EVALUATION OF MODEL (CONTINUED)



# CONCLUSION

- I was successfully able to build a model to predict the home sale price with reasonable accuracy
- Small improvements to the model could be made by adding more features and identifying additional interactions to include