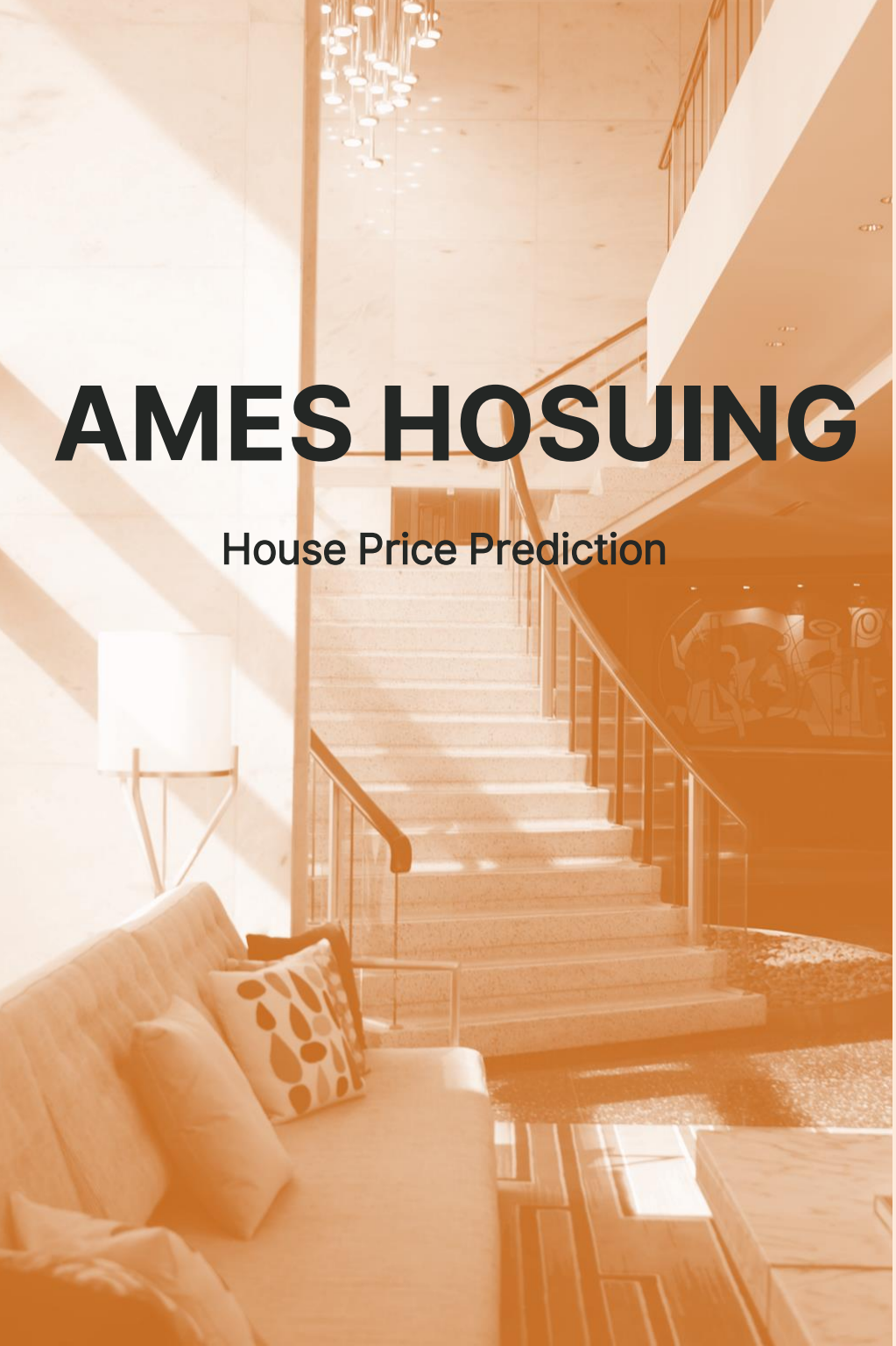


AMES HOSUING

House Price Prediction

5

With 79 variables describing many aspects of residential homes, this dataset can help me have a deeper understanding on how the house price is determined. In the future, if I want to buy a house, I would have a sense on the possibilities for price negotiation.

Click the links to view:

Tableau Storyboard | Github Repository

Project Overview

Housing Profile Analysis

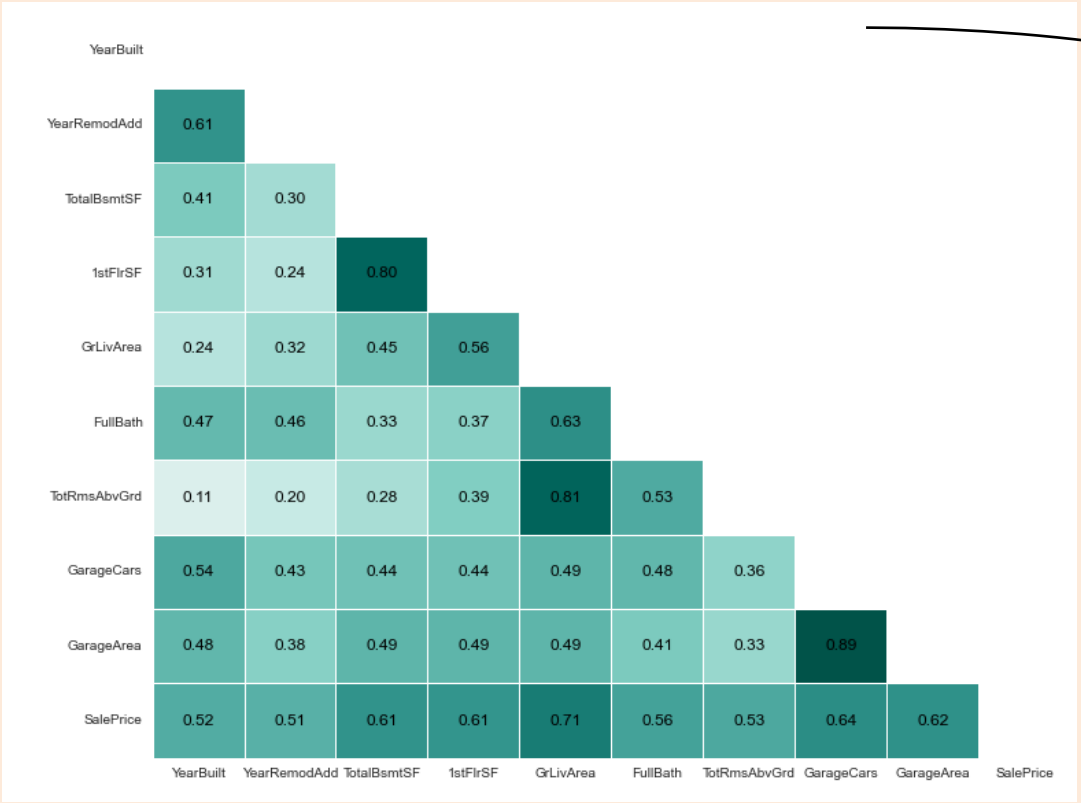
- Human housing purchasing preference investigation
- Key features influencing housing price identification
- Housing price prediction model establishment

Data

- Datasets: 80 features and 2930 observations from House Price Prediction
- Data source: Accessed from The Ames Housing dataset on July 2022
- Investigated factors: House capacity, setting, condition, location and sale type ...

Skills

• Python • Jupyter Notebook • Feature selection • Feature engineering • Feature scaling • Data Transformation • Label Encoding • Data wrangling • Data modeling • Reporting in Tableau



Feature Selection

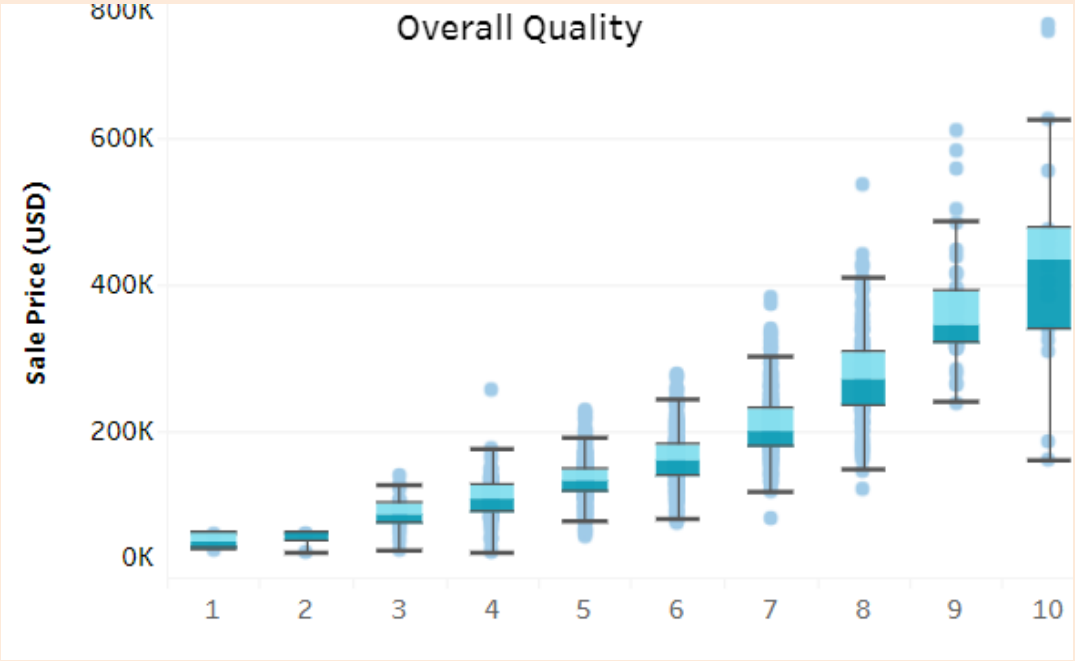
I have used correlation heatmap to determine the critical numeric features influencing housing price. We can see that living area above ground is the most correlated to the sale price. What is also highly correlated with living area is total room above ground, which we can remove to ease the pressure of too many features for modeling.

Feature Engineering

The year house was built and the remodel date are transformed to house age and renovation age. These two features are newly created to simplify and speed up data transformation and enhance model accuracy.

YearBuilt	YearRemodAdd
2003	2003
1976	1976
2001	2002
1915	1970
2000	2000

Age	RenovateAge
7	0
34	0
9	1
95	55
10	0

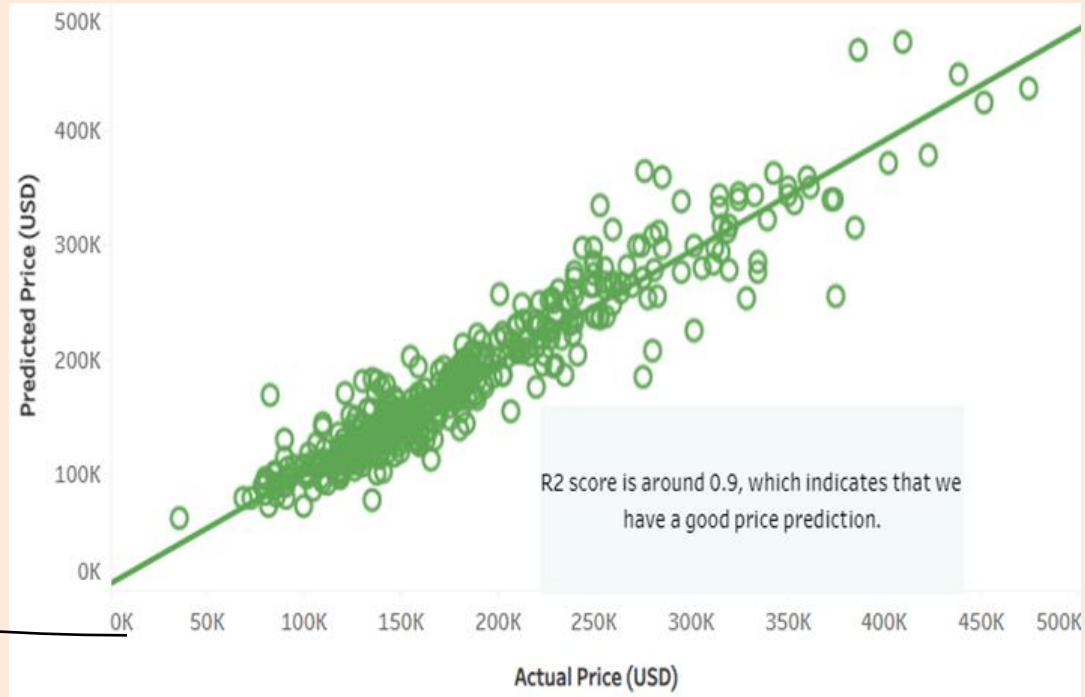


Feature Selection

I have used histogram and box plot chart to investigate categorical features. Data with similar pattern would be removed and the feature would be kept if the relationship between the feature and Sale price is significant. We can see that overall quality is positively related to the housing price.

Model Establishment

After feature selection and engineering, I have used training and test dataset to build a model for housing price prediction. R2 score is around 0,9, indicating that our prediction is 90% close to the actual data.



Conclusion

Insights and Recommendations

Insights

- Features contribute the most to housing price are **living area above ground, garage size, number of full bath, basement size, building year, renovation time, neighborhood and housing overall quality.**
- Log transformation for response variable, correlation heat map for numerical features, histogram and box plot chart for categorical features, regression algorithm are critical steps to select the most important features.
- Housing price has changed gradually over time. The lowest housing price in the early 2000 was already 3.5 fold of the lowest housing price in previous years.
- People prefer to buy houses with **large basement area, 1-2 full bathrooms, and the garage capacity for 2 cars.**
- The houses in **Northridge Heights, Northridge, Stone Brook, Timberland** are significantly pricier then the other neighborhoods, while the houses in Briardale, Meadow Village, Iowa DOT and Rail Road are priced the lowest.

Recommendations

- Although our prediction model has around 90% accuracy, the model can still be improved through **collecting more housing data and optimizing categorical feature selection.**