

HOUSING MARKET

Click to add text

A SUPERVISED MACHINE LEARNING APPROACH

June 26th 2022

Mohamed MZAOUALI

OUTLINE

- Introduction
- Exploratory Data Analysis
- Feature Engineering
- Model Selection
- Data Analytics
- Conclusion
- References

INTRODUCTION

Objective:

- Prediction: House Sale Price
- Linear Regression Variations

Data Source

- AMES HOUSING DATA SET

EDA

DATA SET:

Shape:

- 1379 Lines by 80 Columns.

Null Values:

- None.

Missing Values:

- None.

NaN Values:

- None.

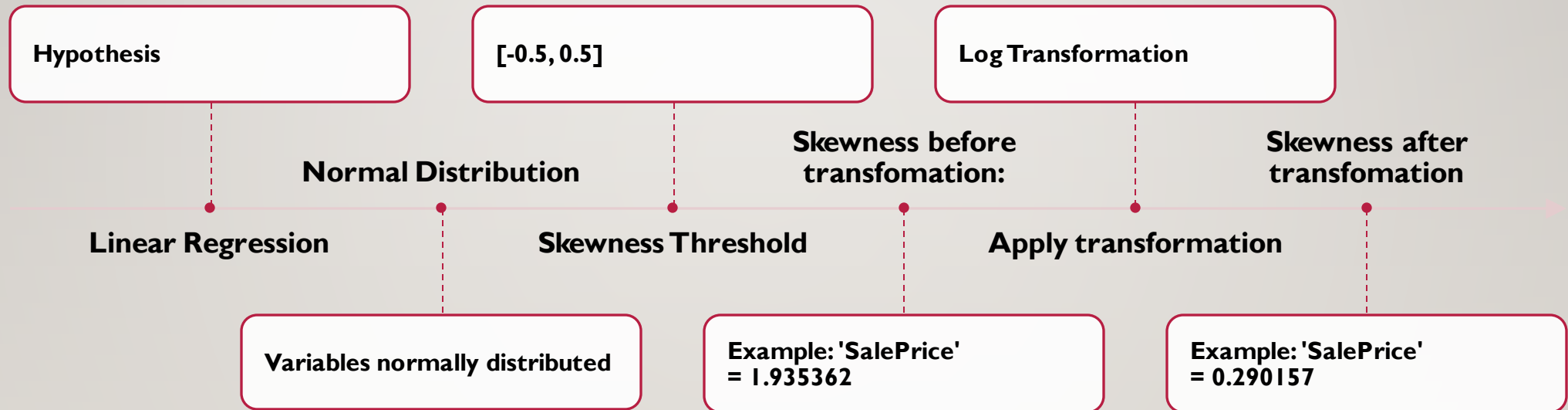
Independent Variables:

8 numerical features with high correlation with the Target Variable.

Dependent Variable:

'SalePrice' Column.

FEATURE ENGINEERING



MODEL SELECTION



Linear Models

Linear Regression
Ridge Regression
Lasso Regression

Best Score

- Training = 79.6% by Ridge Model
- Testing = 75.7% by Lasso Model

 **GridSearchCV**

CONCLUSION

- OverallQual: 50%
- GrLivArea: 25%

Most Important
Features

Future
Improvements:

- Categorical Features
- Polynomial interaction
- Ensemble models



REFERENCE

- The full approach is presented in the following ipynb:
- https://github.com/mzaoualim/Coursera_IBM_Machine_Learning_Professional_Certificate/blob/main/Supervised%20Machine%20Learning:%20Regression/Project.ipynb