



PREDICTING CAR PRICES USING LINEAR REGRESSION

by Sultanah Aldossari & Rawan Alharbi

CONTENT

1 Introduction

2 Dataset
Exploratory Data Analysis

3 Pre-Processing

4 Modeling

5 Conclusion

INTRODUCTION

PROBLEM STATEMENT

- The purchase of a car can take a considerable amount of time.
- It is possible to determine the actual worth of a car based on many factors.
- The aim is to develop Linear regression model based on our data

DATASET

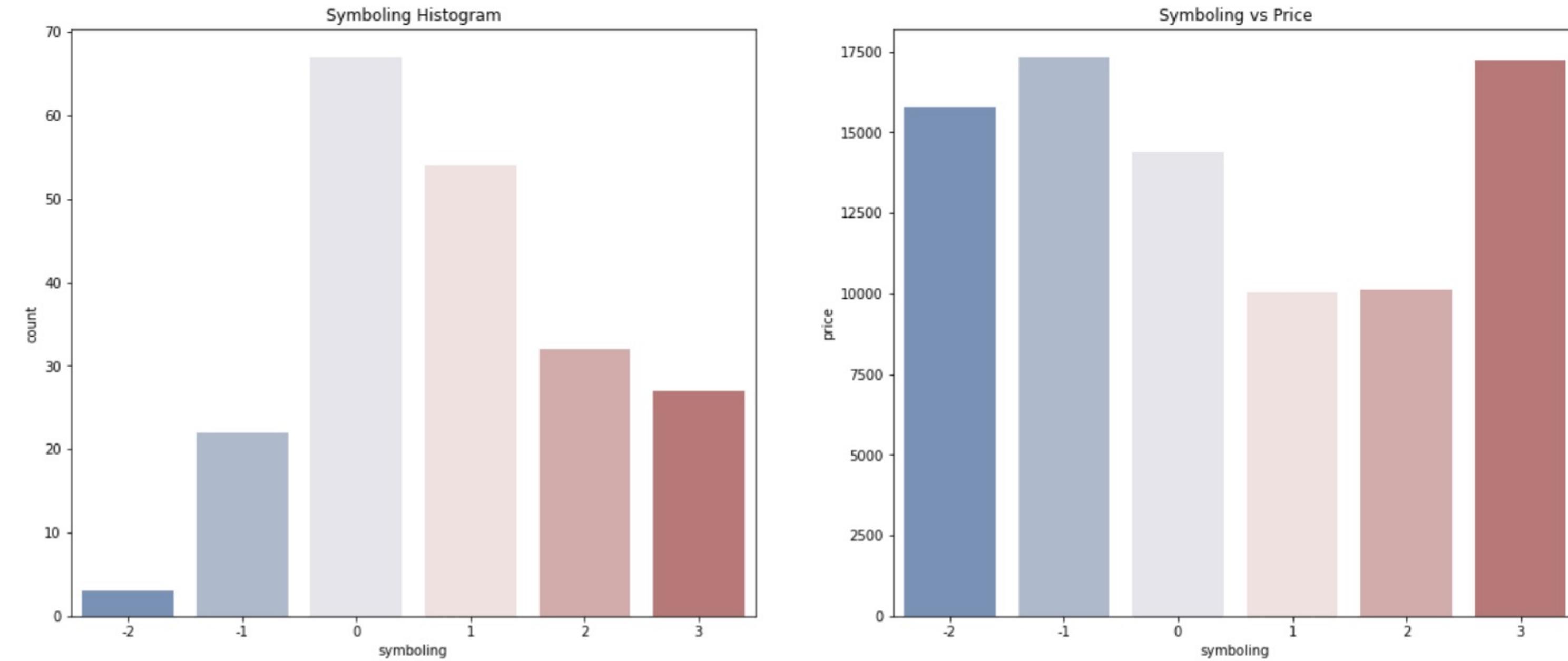
DATA

- Data scraped using Selenium.
- 205 observations, 26 features.
- Price is our target
- The independent variables are divided into Categorical and Numerical variables.

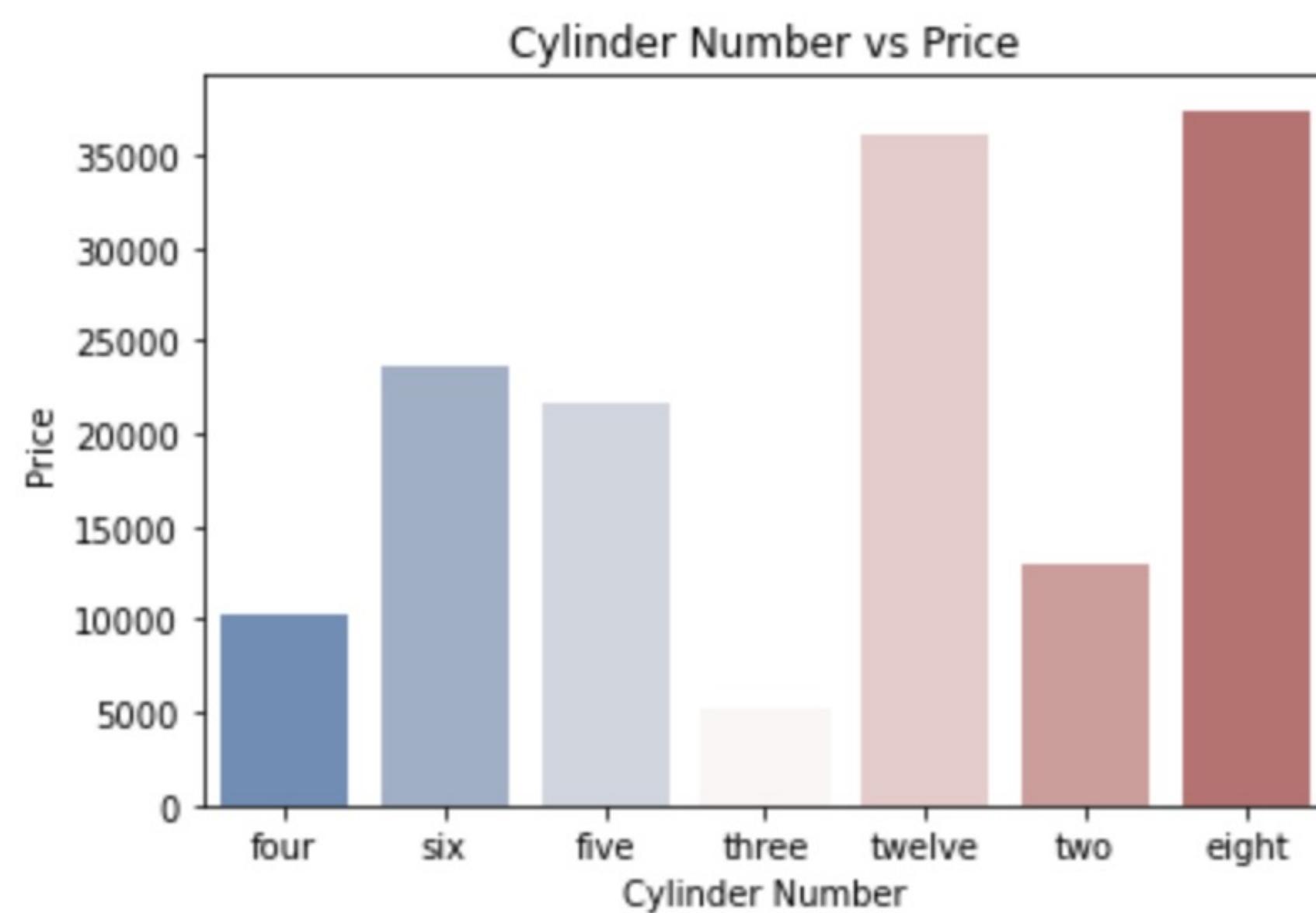
Our Dataset

- 1- Car Id
- 2- Symboling
- 3- Car Name
- 4- Fuel Type
- 5-Aspiration
- 6- Door Number
- 7- Car Body
- 8- Drive Wheel
- 9- Engine Location
- 10-Wheel Base
- 11- Car Length
- 12- Car Width
- 13- Car Height
- 14- Car Weight
- 15-Engine Type
- 16- Clinder Number
- 17-Engine Size
- 18- Fuel System
- 19- Bore Ratio
- 20- Stroke
- 21-Compression Ration
- 22-Horse Number
- 23-Peak Number
- 24-CityMpg
- 25-HighwayMpg
- 26-Price

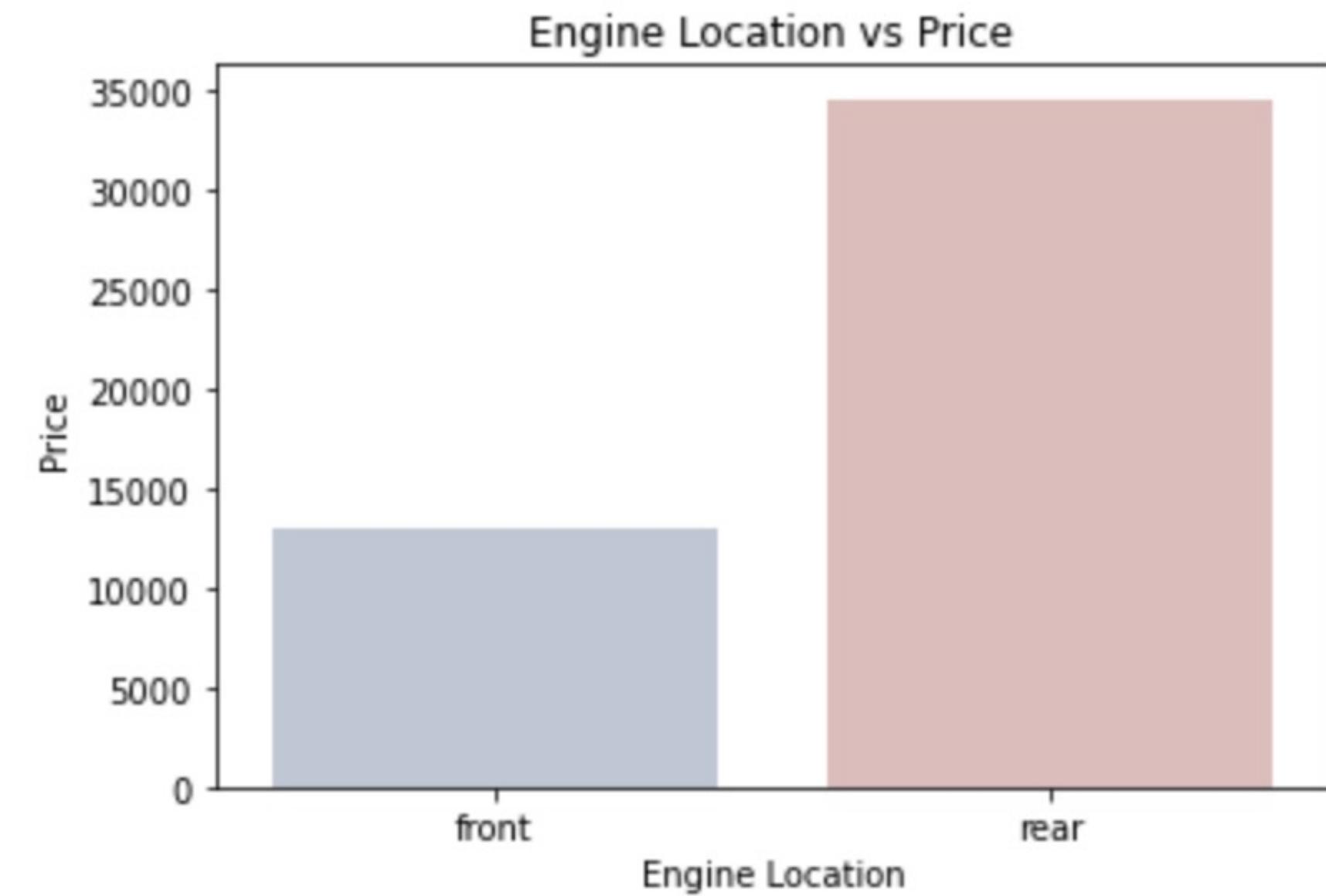
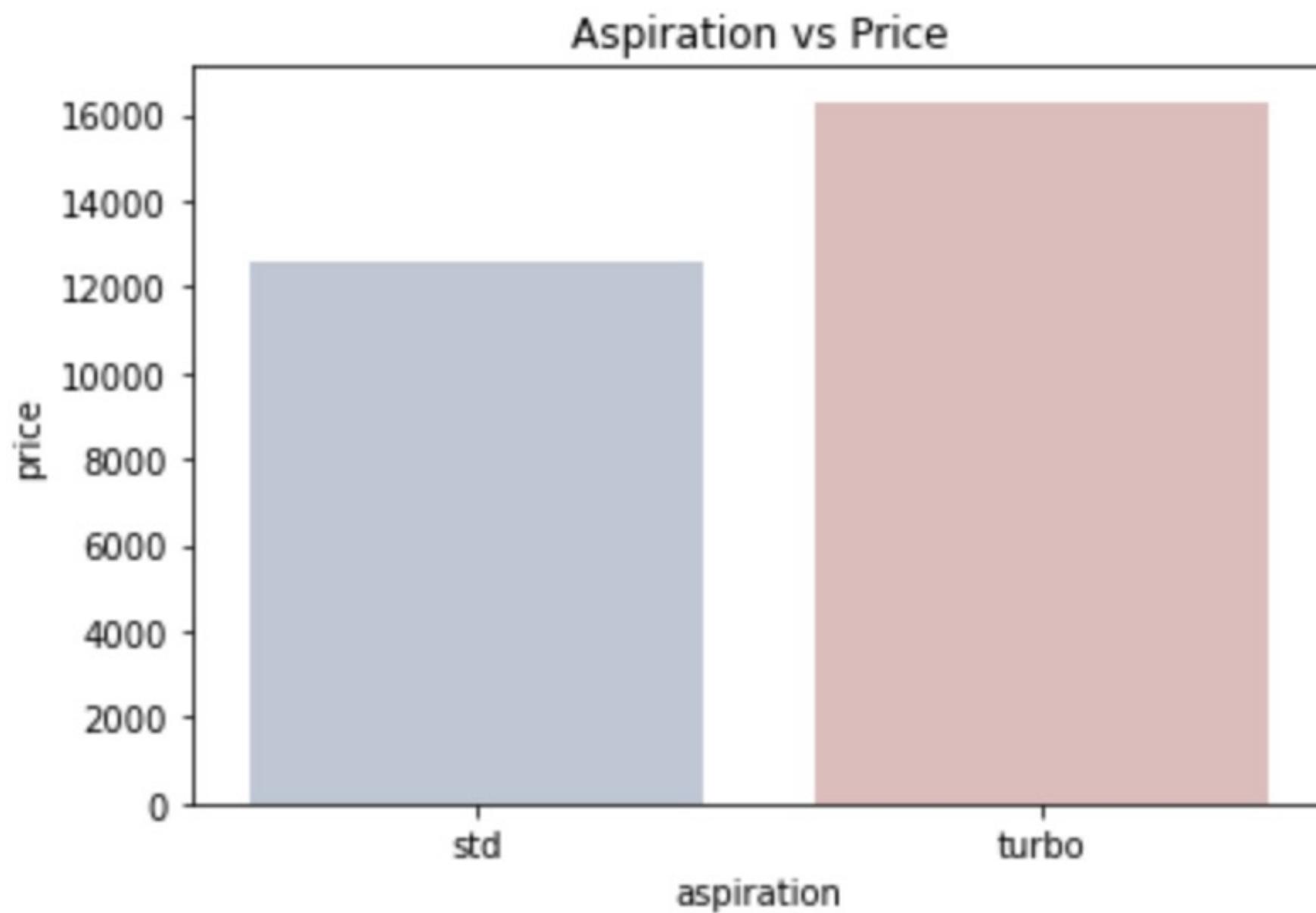
EXPOLARITY DATA ANALYSIS



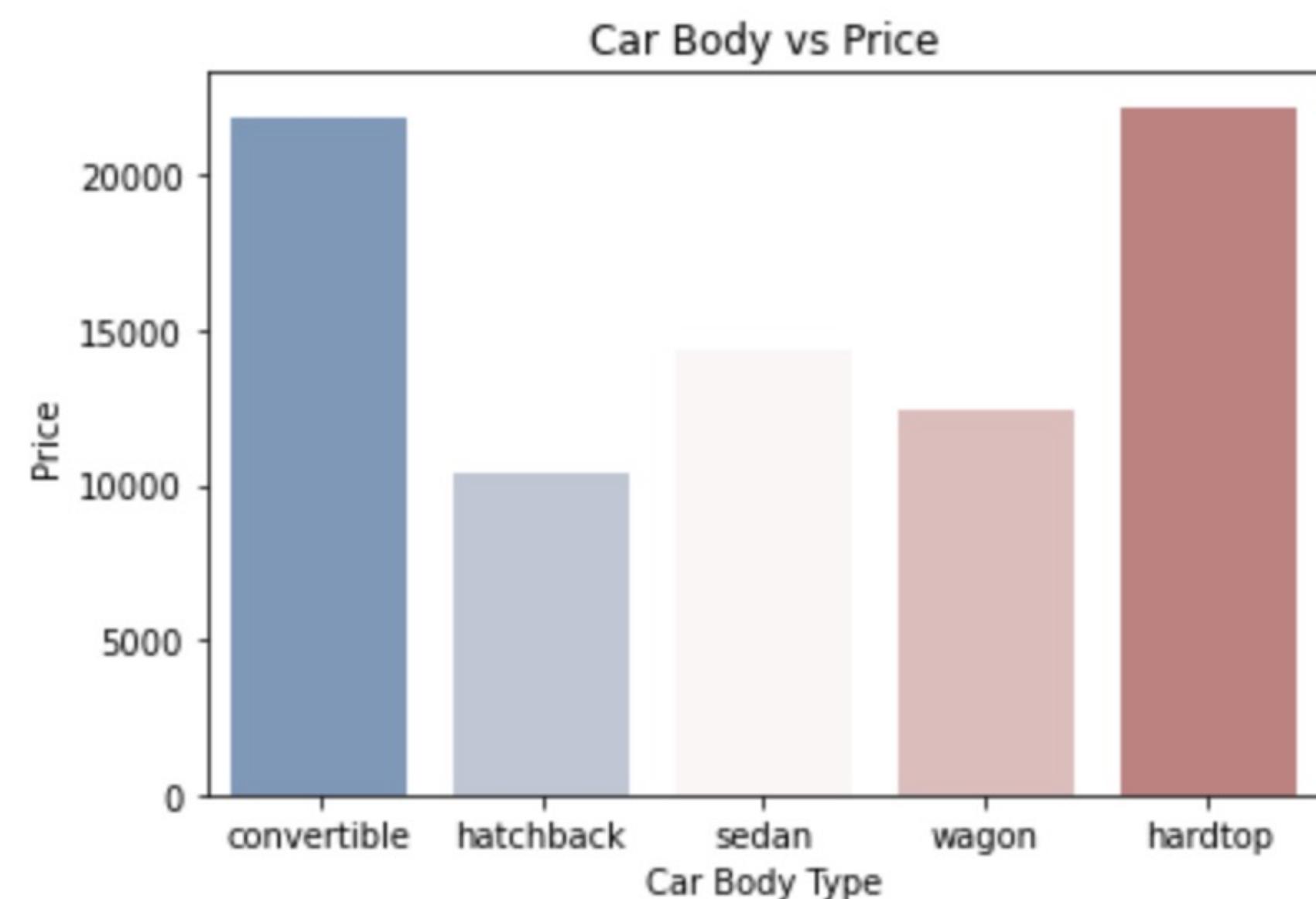
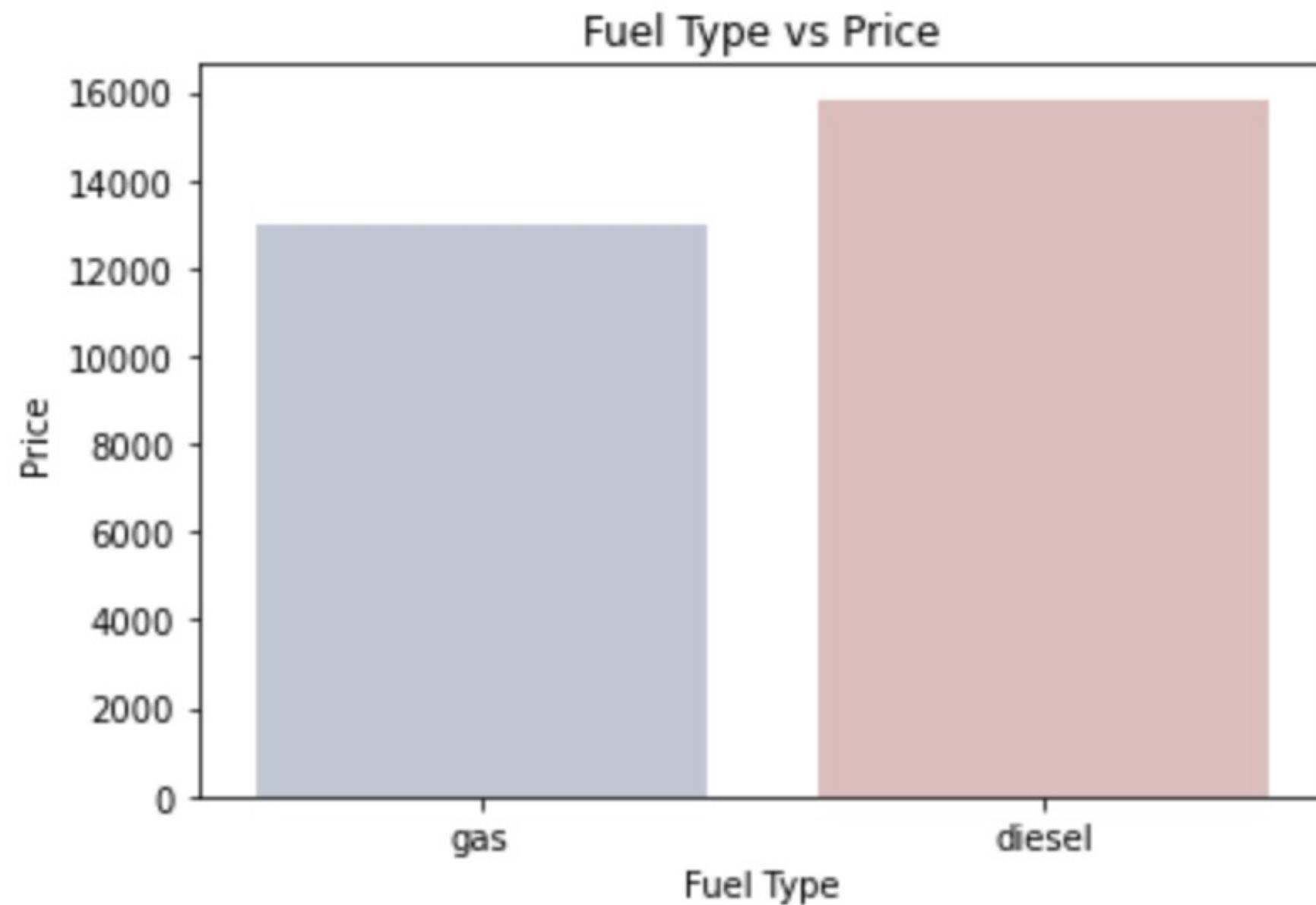
CONT: EXPOLARITY DATA ANALYSIS



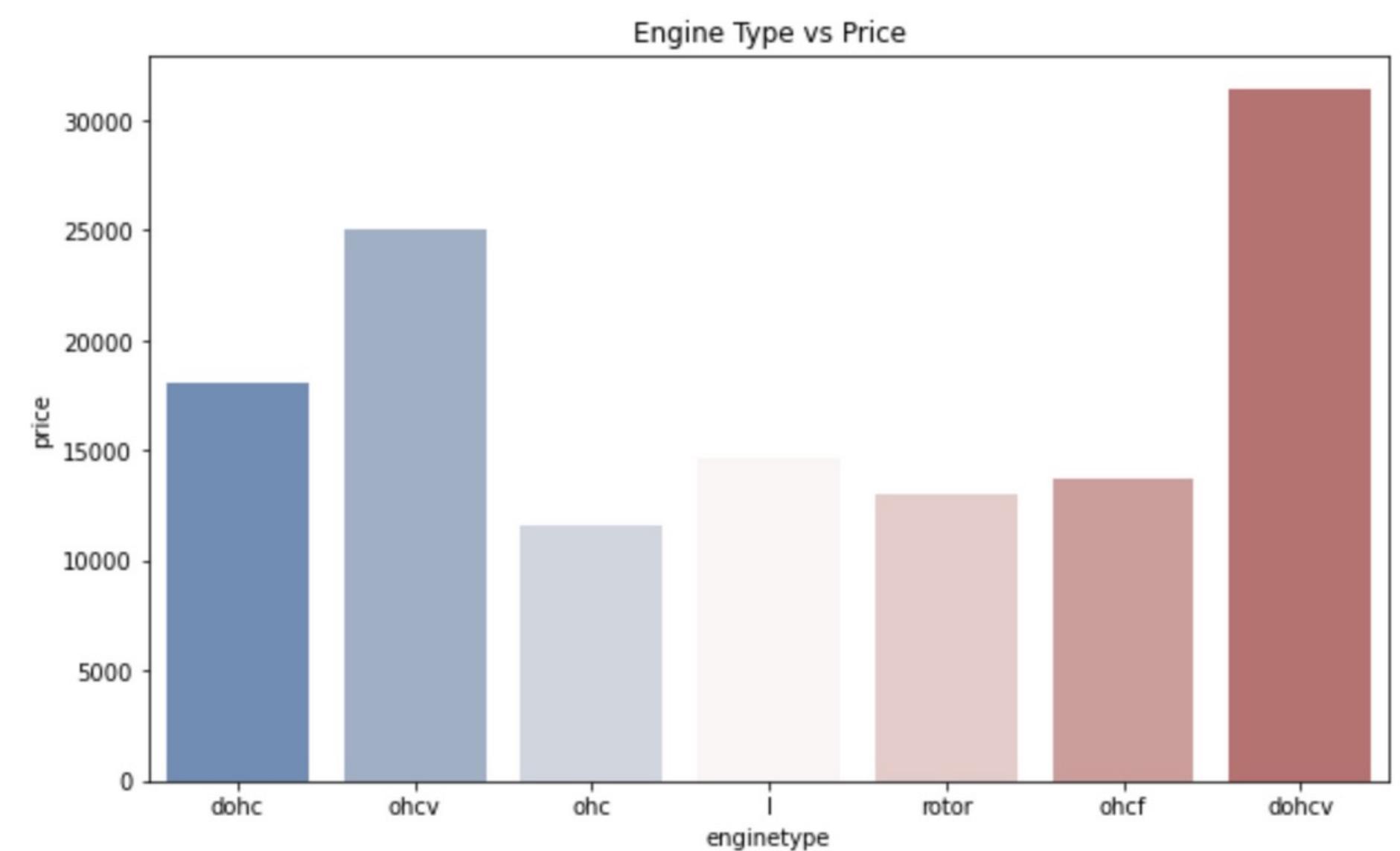
CONT: EXPOLARITY DATA ANALYSIS



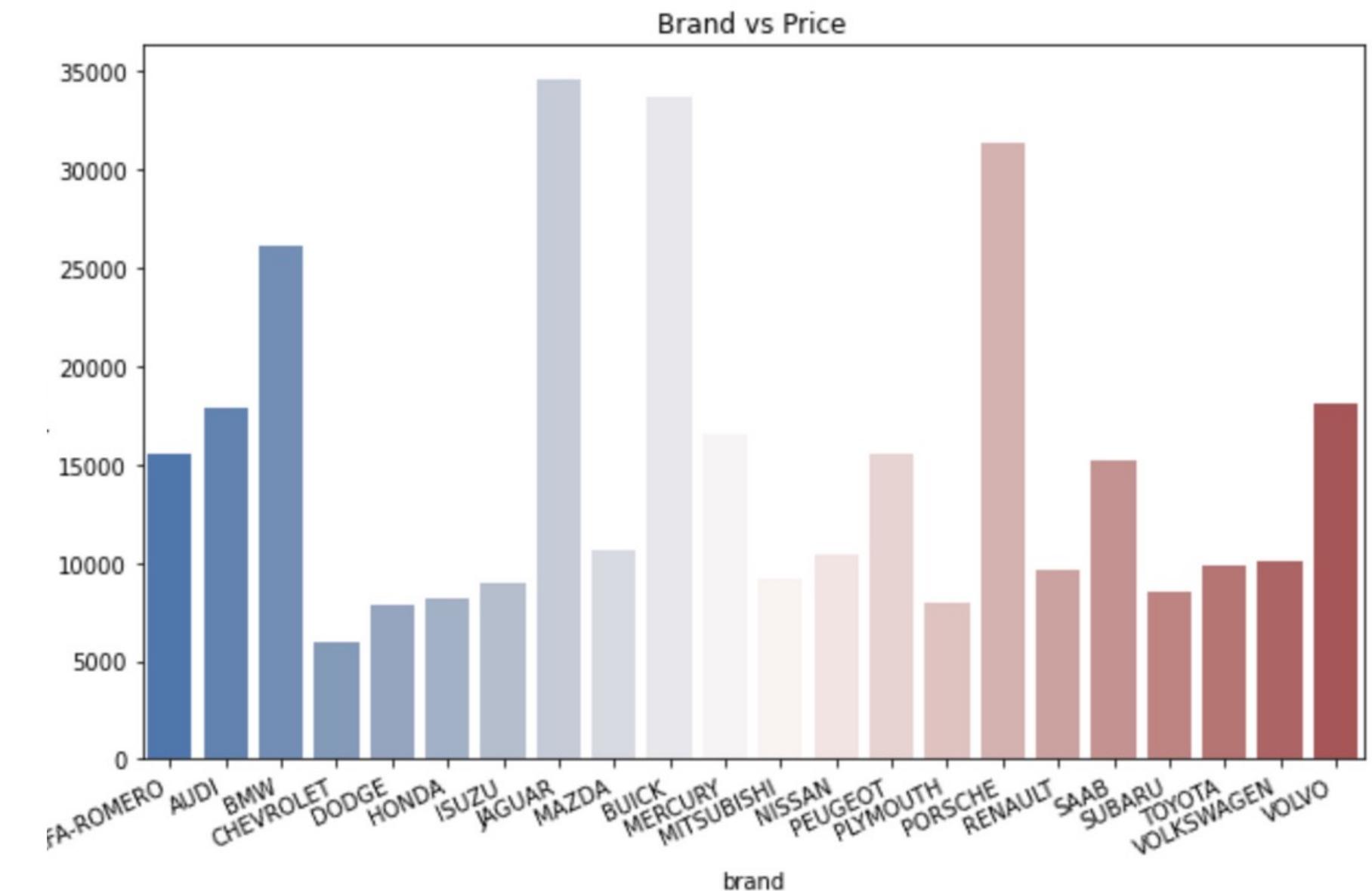
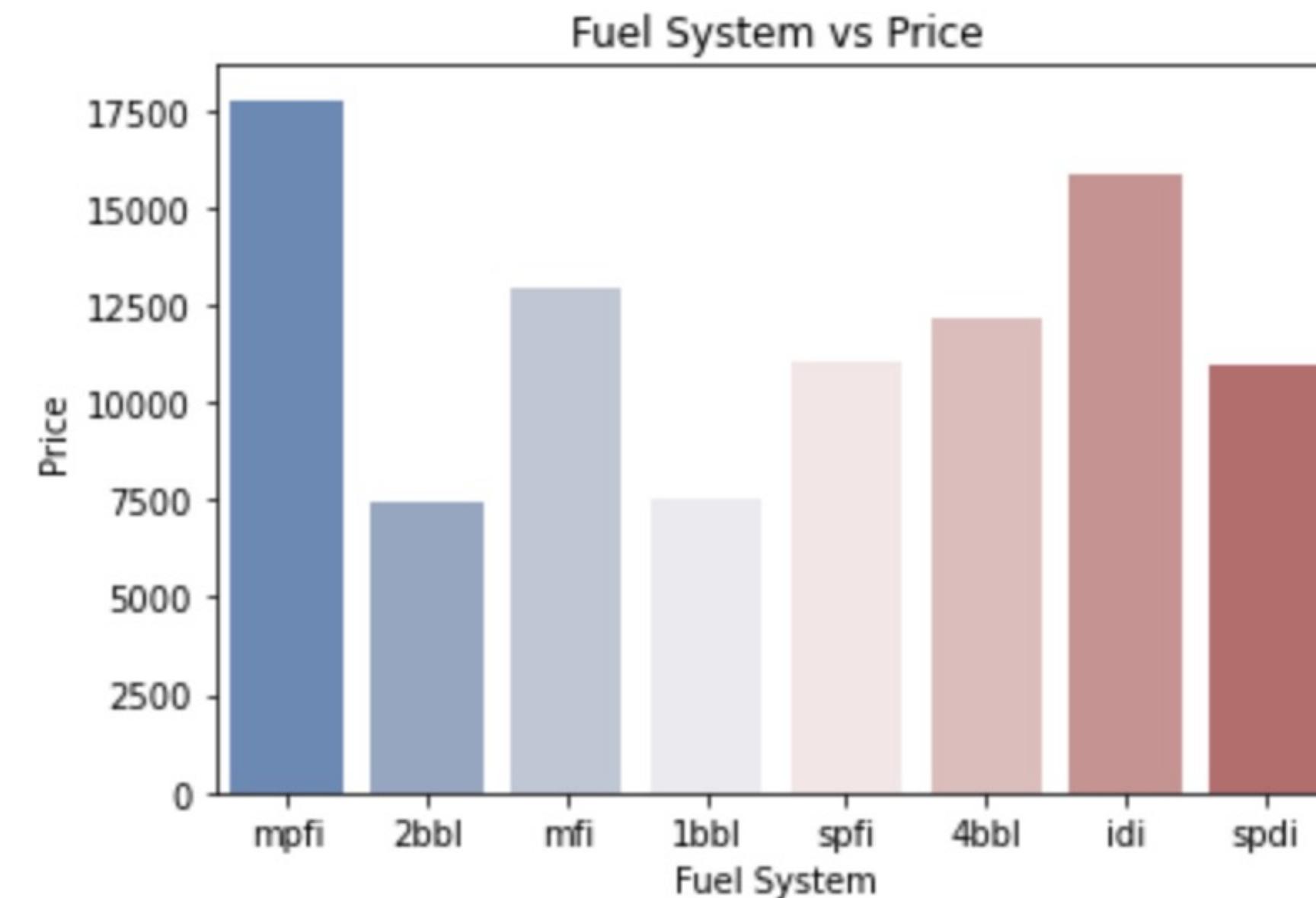
CONT: EXPOLARITY DATA ANALYSIS



CONT: EXPOLARITY DATA ANALYSIS



CONT: EXPOLARITY DATA ANALYSIS



CONT: EXPOLARITY DATA ANALYSIS



PRE-PROCESSING

- Converting categorical data into dummy variables
- Derive a new features from our data
- Scaled our data into the same scale
- Created a new dataframe

Cars Dataset

- | | |
|----------------|--------------|
| 1- Wheelbase | 15-Turbo |
| 2- Car length | 16-HardTop |
| 3- Car width | 17-Hatchback |
| 4- Curb weight | 18-Sedan |
| 5- Bore Ratio | 19-Wagon |
| 6- Engine Size | 20-fwd |
| 7- Mileage | 21-rwd |
| 8-Horse Number | 22-dohcv |
| 9-Five | 23-l |
| 10-Four | 24-ohc |
| 11-Six | 25-ohcv |
| 12-Three | 26-rotor |
| 13-Twelve | 27-Price |
| 14-Gas | |

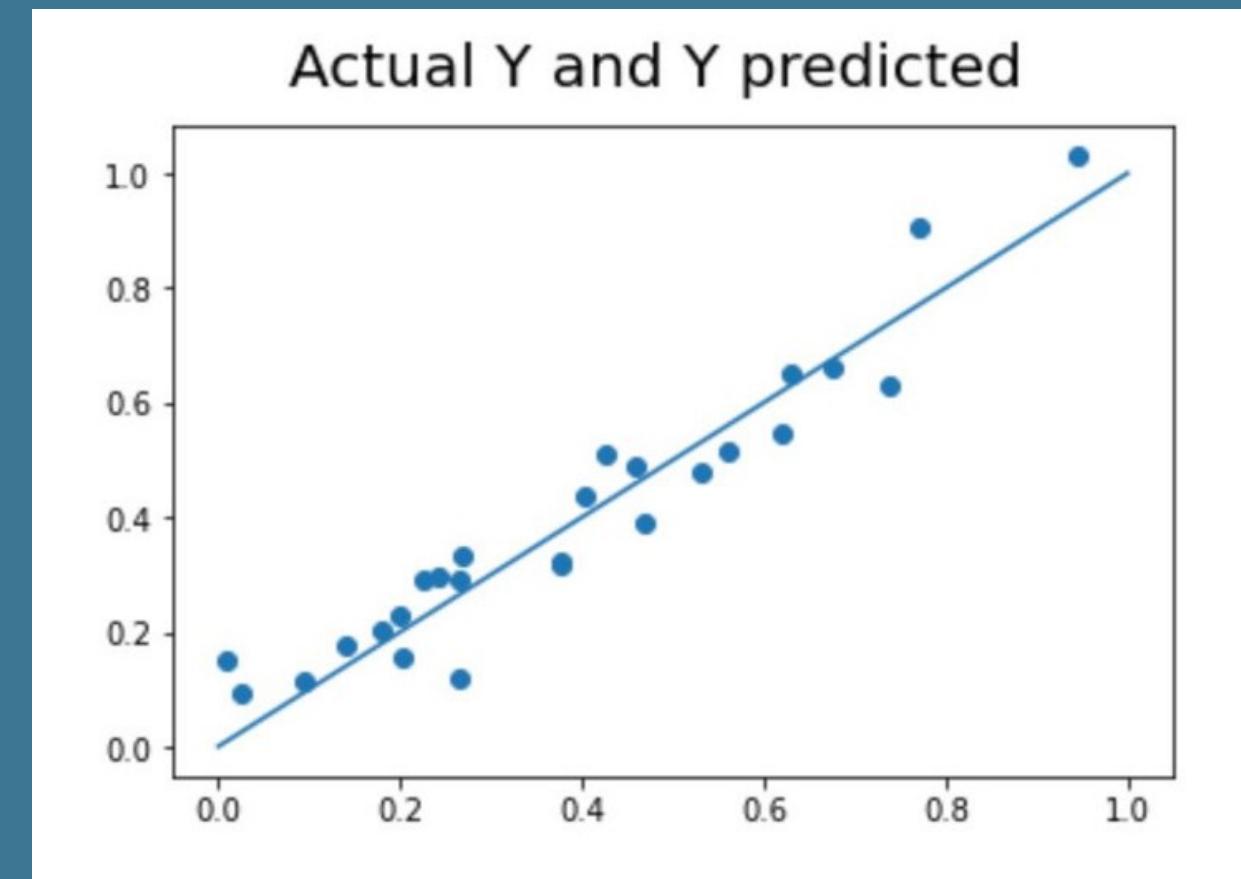
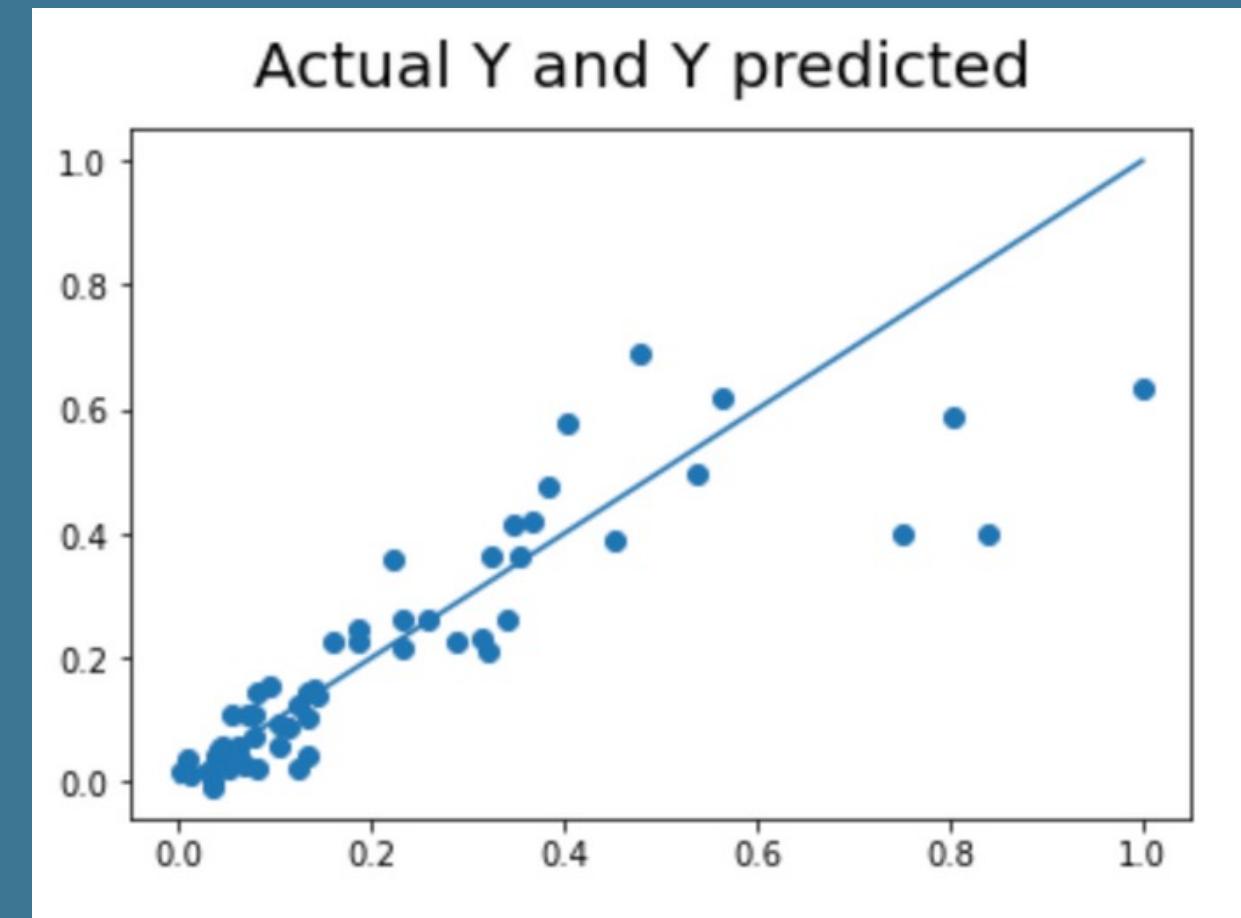
MODELING

Linear Regression

- Split the data into train and test
- Sklearn library were used

1- MSE Test Score = 0.012

2- MSE Test Score = 0.005

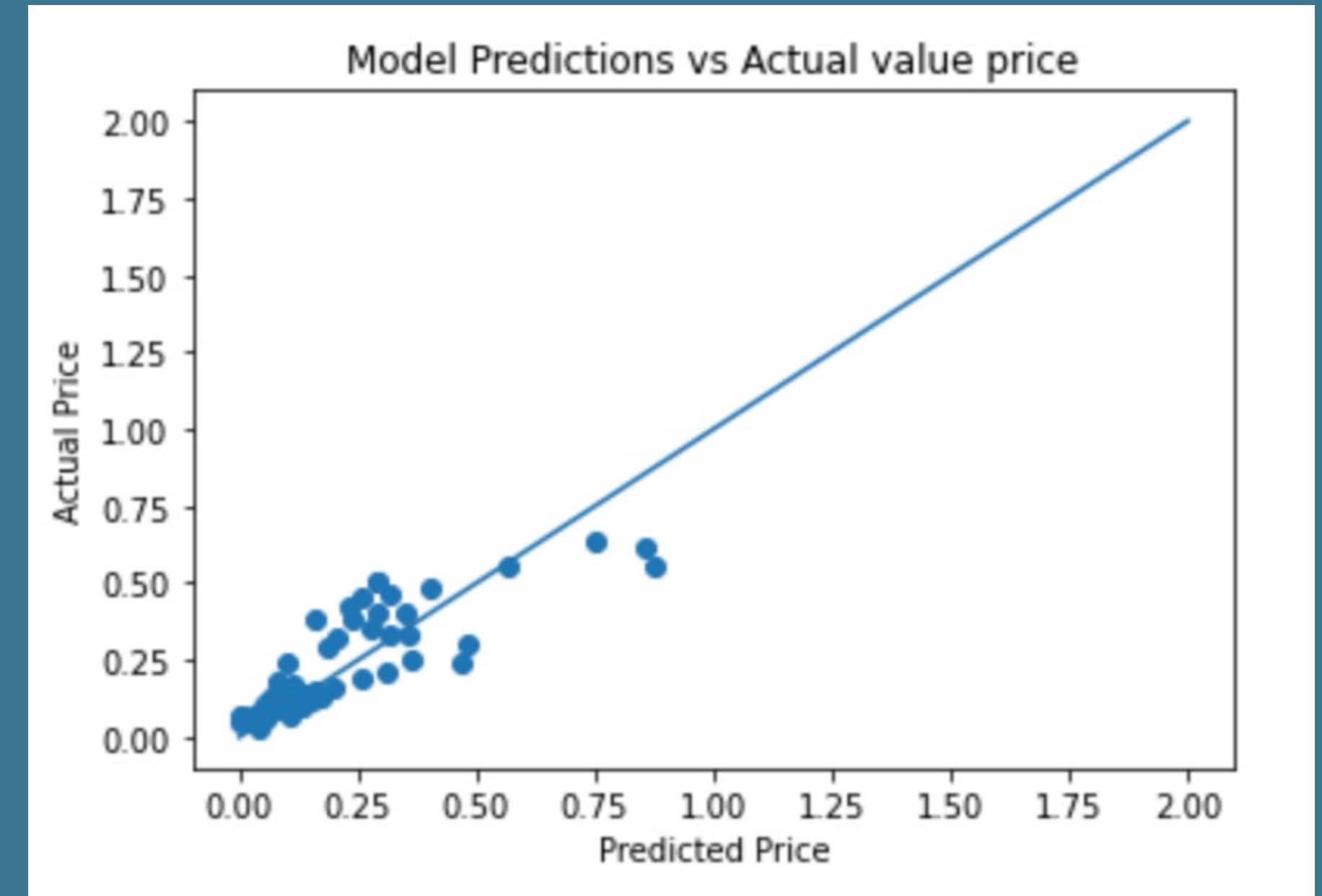


MODELING

Lasso Regression

- Split data into train and test
- The best Alpha value were 0.01

MSE Test Score = 0.012

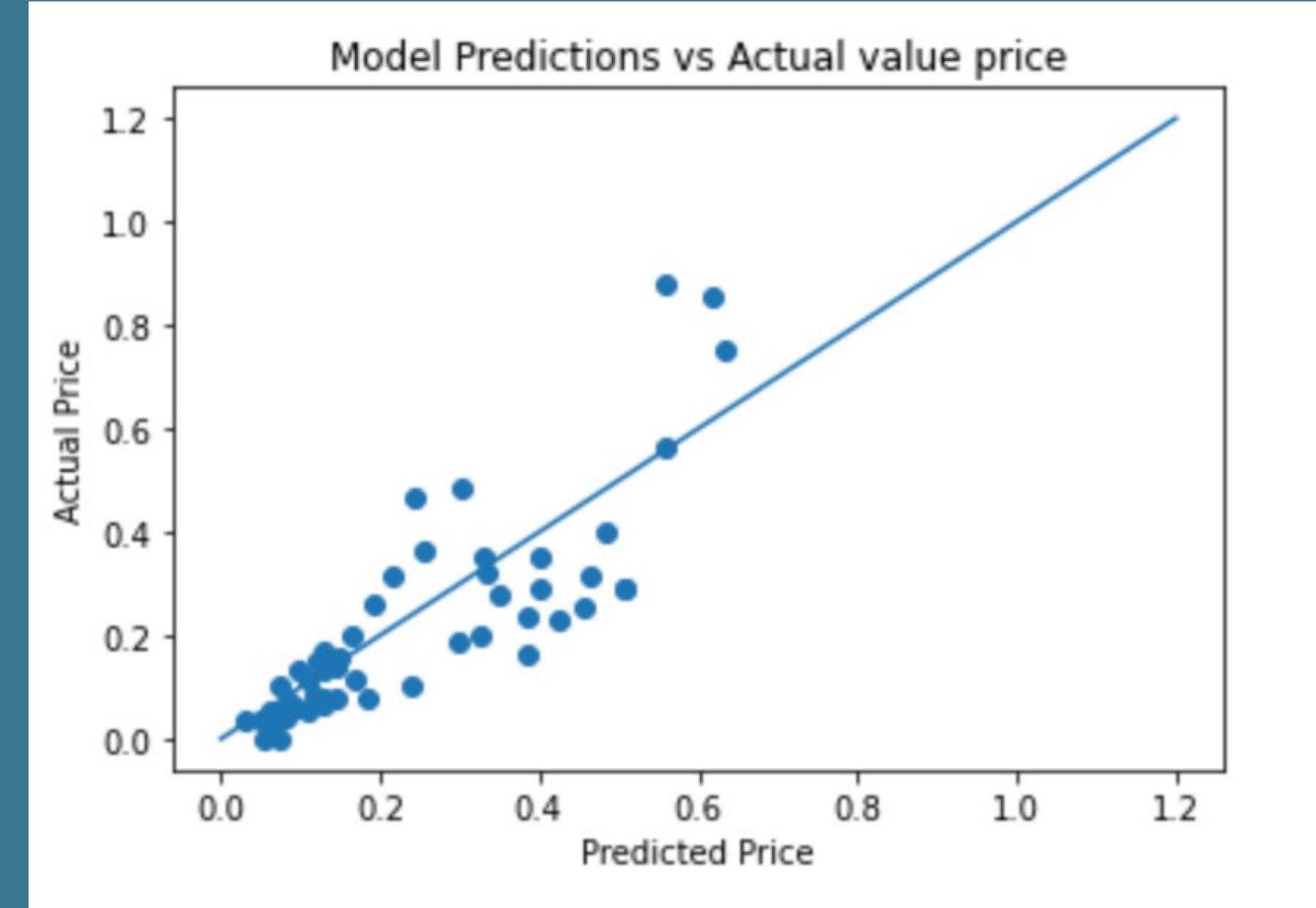


MODELING

Lasso Regression

- After Feature Selection

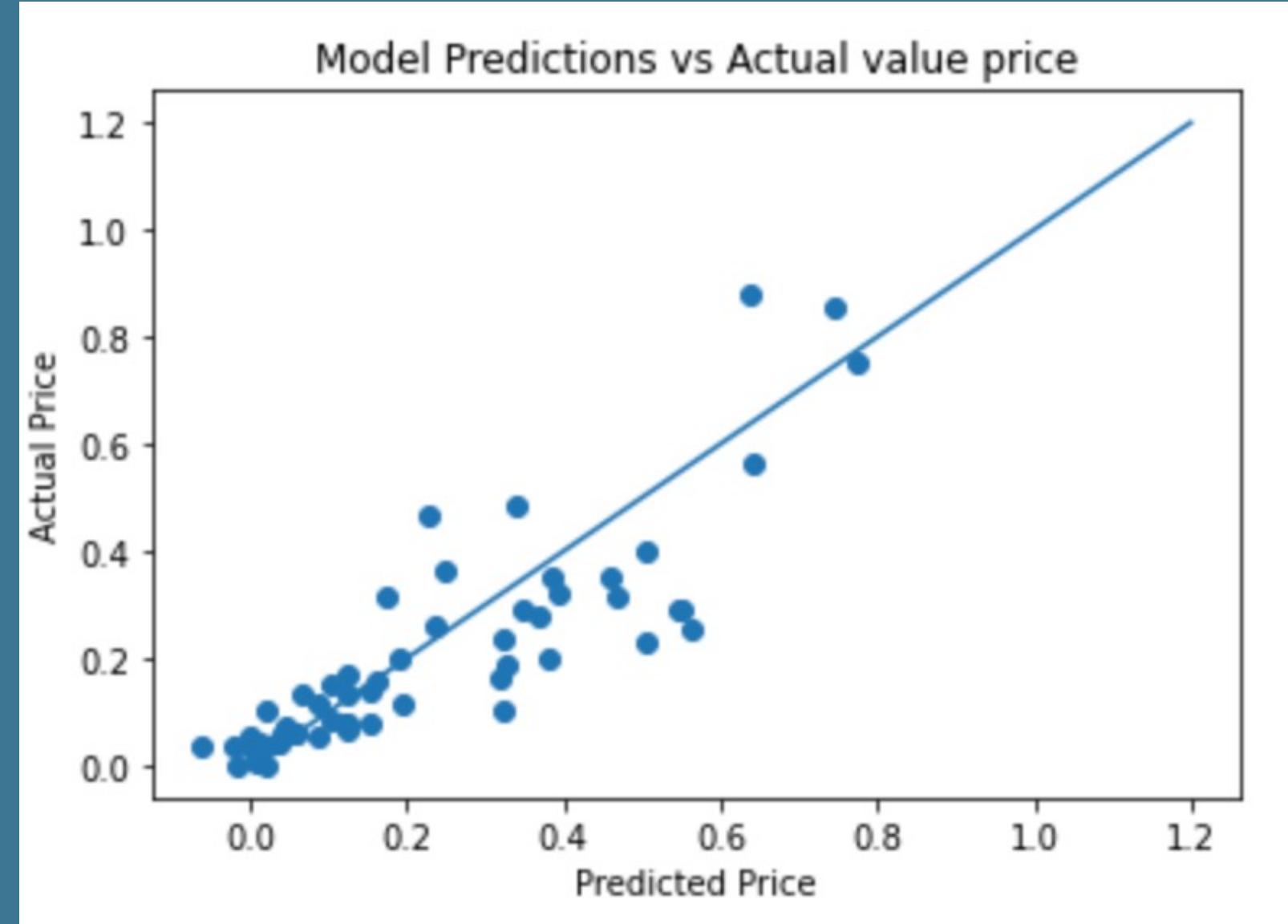
MSE Test Score = 0.011



MODELING

Ridge Regression

MSE Test Score = 0.012



CONCLUSION

The background features a dark teal color with abstract white shapes. On the left, there are two large, semi-transparent circles: one is light blue and the other is a darker shade of teal. Overlaid on these circles are several thin, black, wavy lines that intersect and curve across the frame.

THANK YOU!