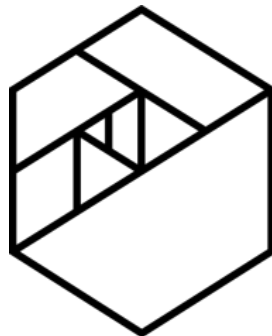# Regression Analysis of Craigslist (Used Car) Listings in Seattle

Chintan Sumaria

October 6, 2017

METIS

Motivation

Derive Market Insights From The Online Portal

Predict a Car's Value From Its Specifications

Identify Trends In The Market for Select Features

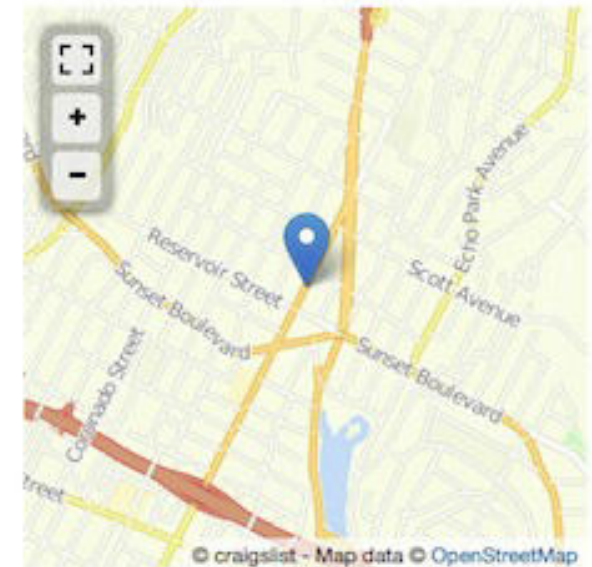Identify Cars that are Being Offered below their Market Value
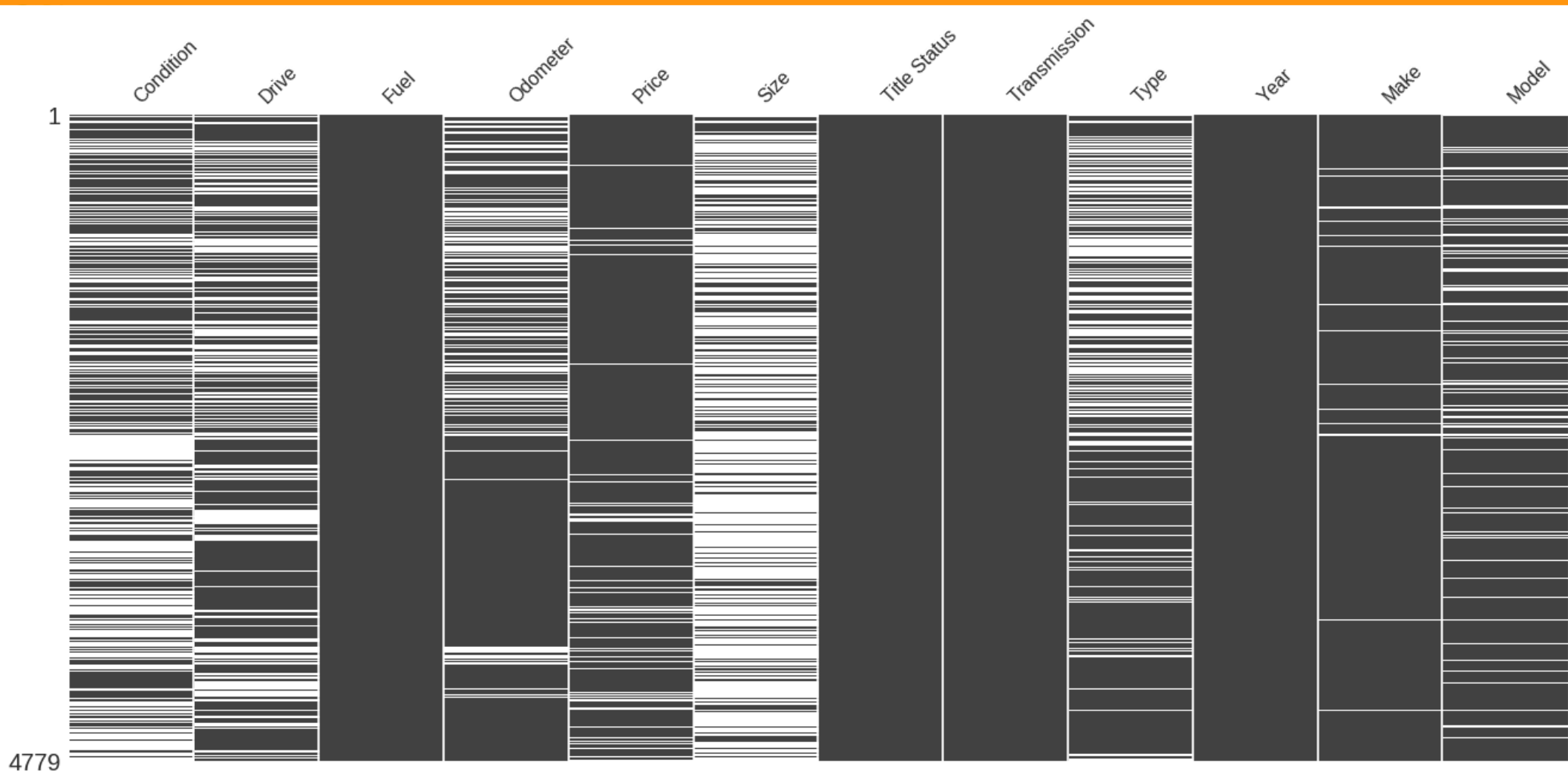
Data Collection

Web Scraping

- Data Scraped Using the Selenium Module in Python

- Web Elements Selected by Class Name, Link Text, Tag Name & XPATH

- 4,779 Pages Scraped

- 81 Makes and 1084 Models

- 59 Entries from Before 1960
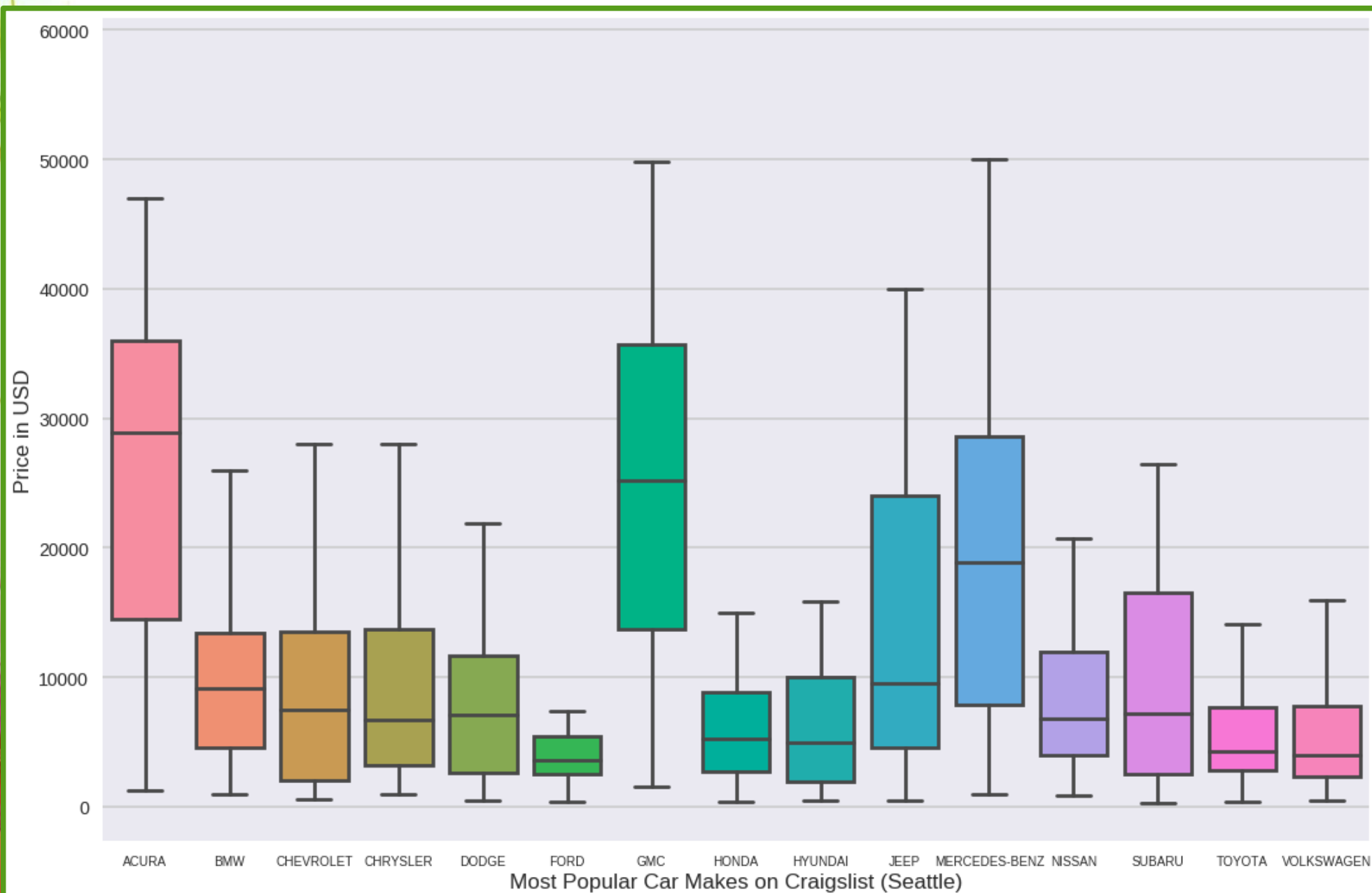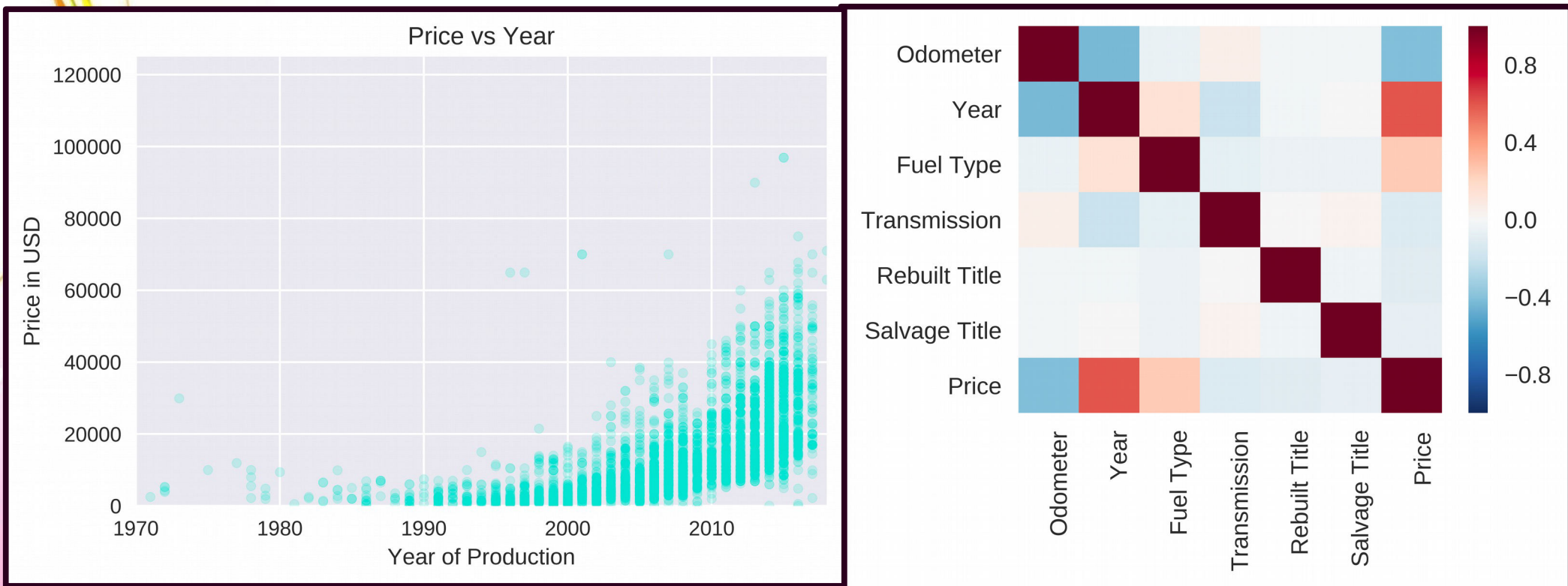

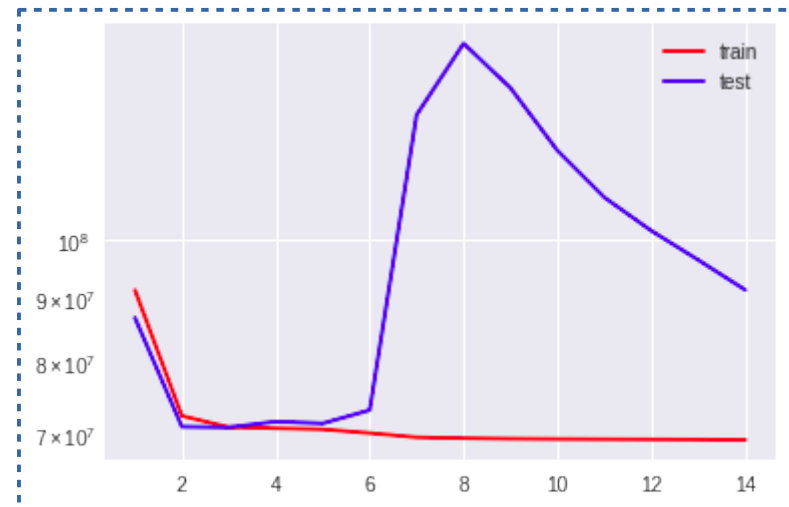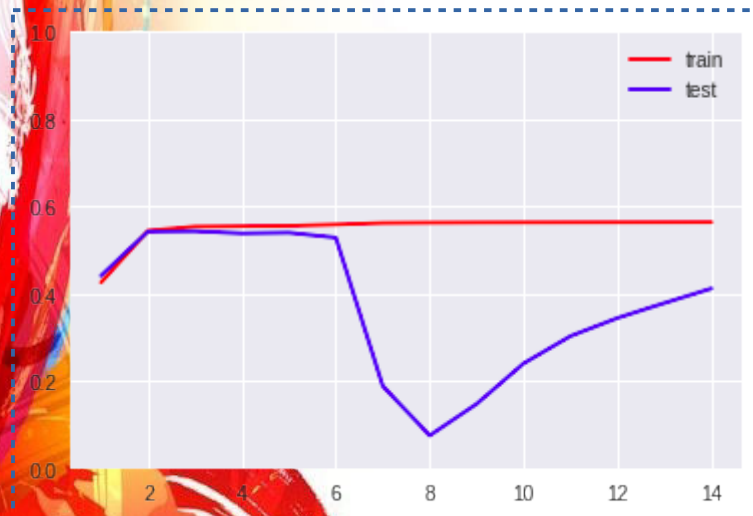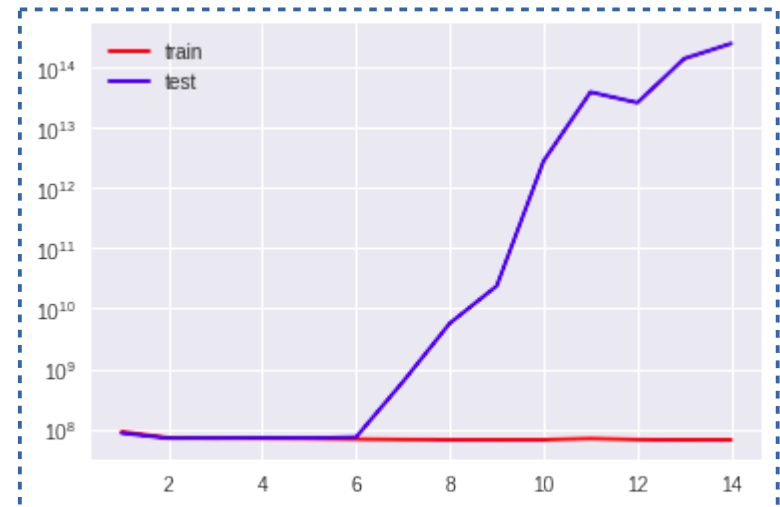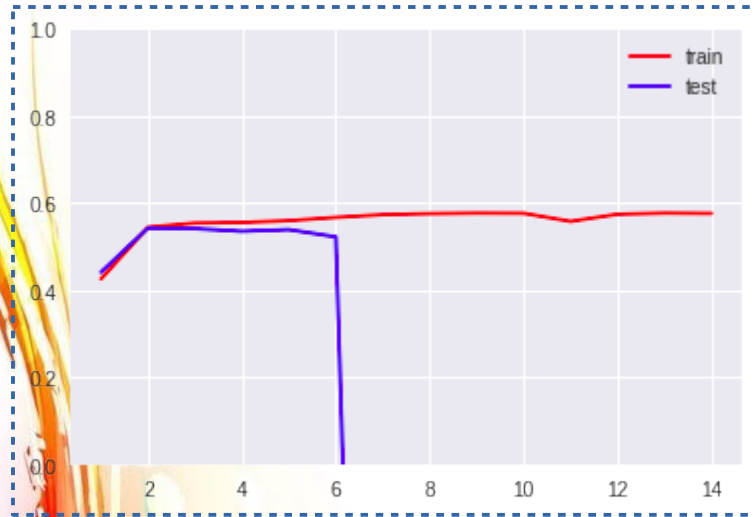*1928 Ford Model T*

Data Wrangling

Top 15 Makes

Trends Across All Make Model Combinations

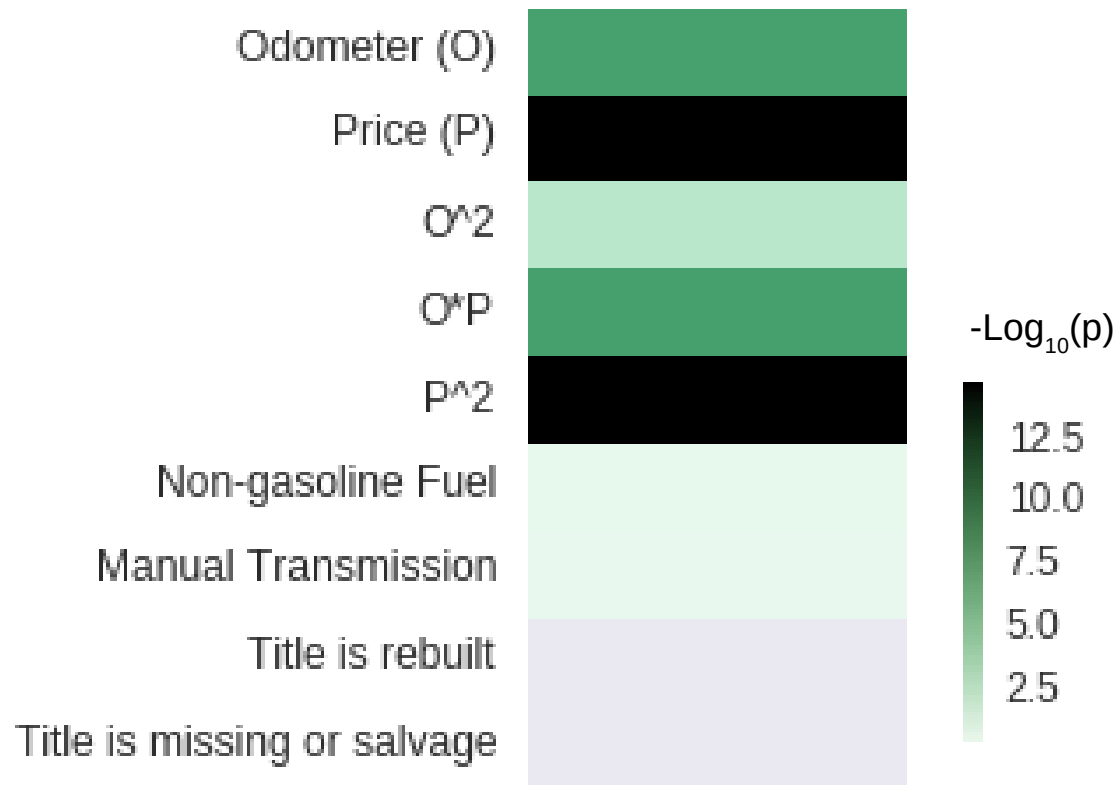# Applying Regression Models Across All Make Model Combinations

Applying Regression Models Across All Make Model Combinations

- R-squared = 0.545

- Deriving Mean Impact of a Feature from Coefficients

- Title is Rebuilt: $6,175 lost in value

- Title is Salvage or Missing: $6,629 lost in value

- Fuel Type is not Gasoline: Value increases by $4,883
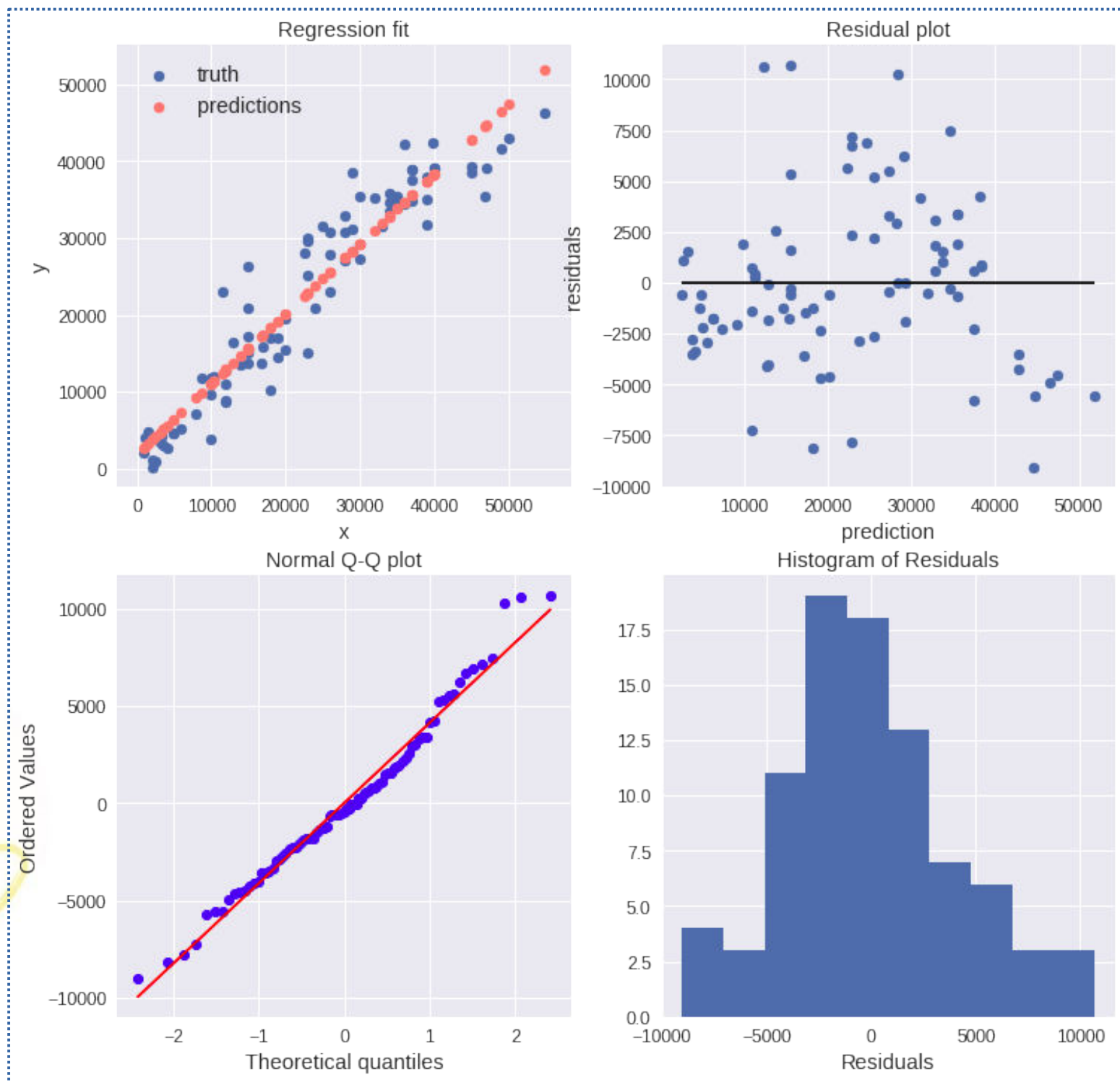
- Transmission Type: No significant impact
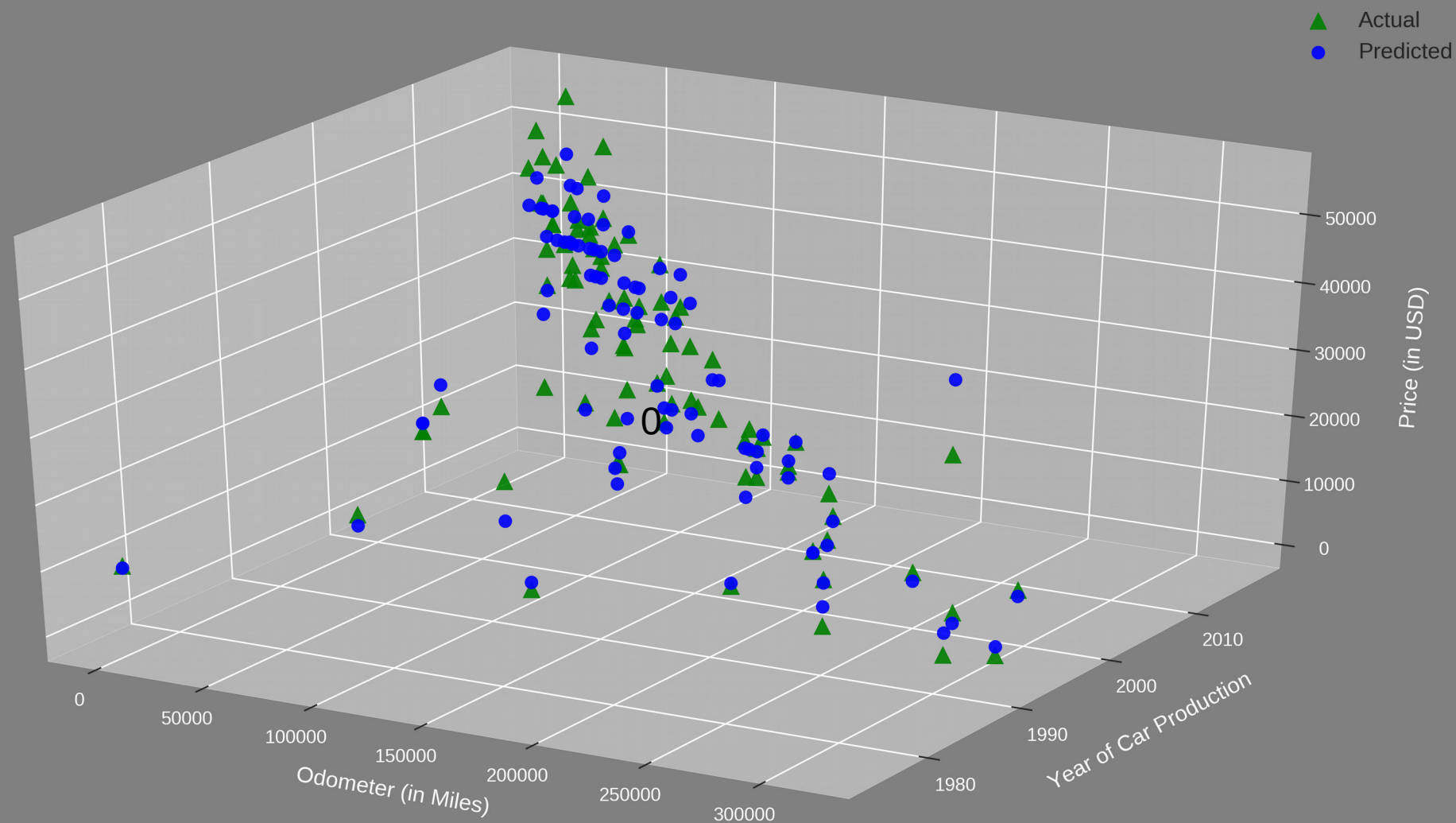
Modeling Specific Make Model Combinations

Chevrolet Suburban 1500

Method	OLS
R-squared	0.91
Adj. R-squared	0.90

Odometer (O)
Price (P)
O^2
O*P
P^2
Non-gasoline Fuel
Manual Transmission
Title is rebuilt
Title is missing or salvage
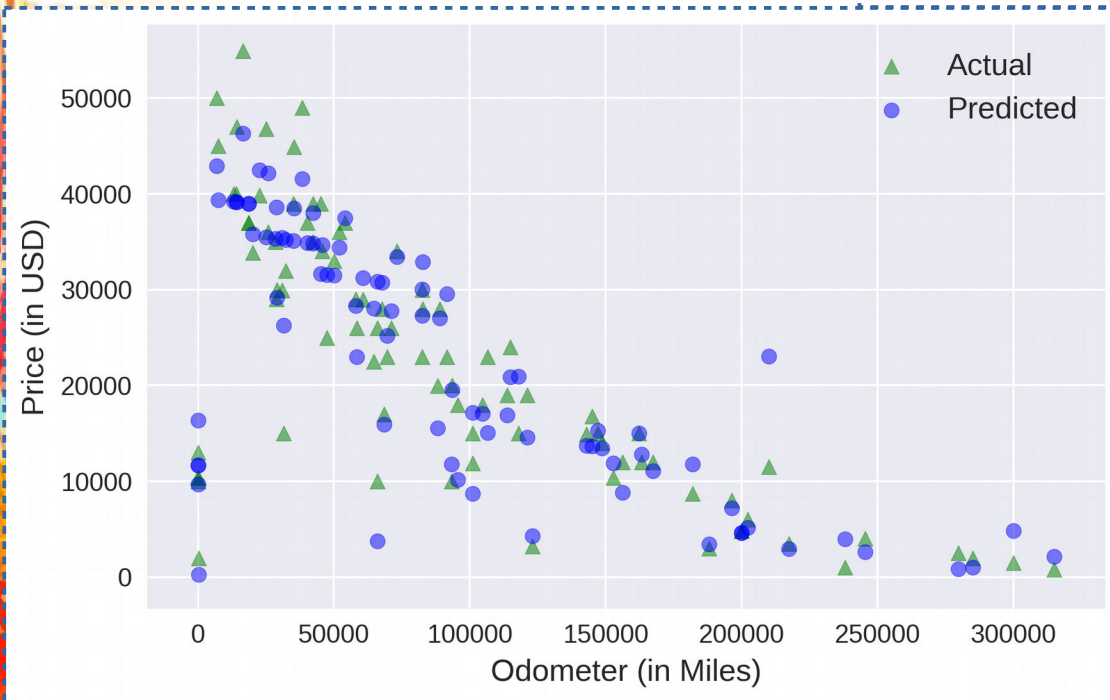
$-\text{Log}_{10}(p)$
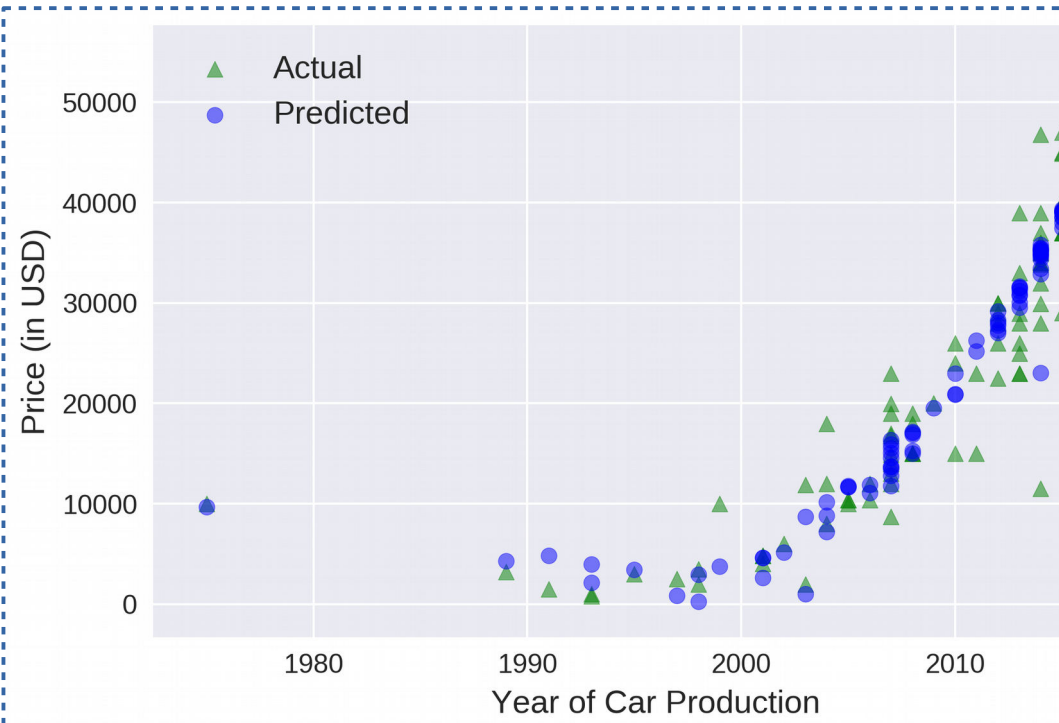
12.5
10.0
7.5
5.0
2.5

Diagnostic Plots

Actual vs Predicted Prices Plotted Against Quantitativea Features

2D Plots

Next Steps

- Improve Upon the Volume of Data Exponentially

- Include Additional Features in Model

- Develop Models Across All Major Make Model Combinations

- Identify Models That Lose the Least Value Over Time