Capstone Title: What Drives the Price of a Car

**Summary of Findings**

**Business understanding of the problem**

In this capstone project, I am trying to understand which used car features will impact the sell price of used car. We are focus on numeric features, including “years” and “odometers” among all used car features to understand if they are correlate to the “price”.

To address my question, I applied PolynomialFeatures and Lasso in a machine learning pipeline to build up a model and use the model to understand the relationship between “years”/“odometers”/“years and odometers combination” and “price”.

**Data Cleaning**In data cleaning step, the following process are applied to create cleaned dataset:

1. Drop all rows contains ‘NaN’ data;
2. Drop all rows with 0 price value;
3. Drop all non-numeric features;

Split cleaned dataset to training and testing dataset.

**Model Building Process**

In order to understand the impact factor of used car price, I use PolynomialFeatures to introduce years and odometer features and Lasso to shrink irrelevant coefficients and leave the most important features in the model. Both PolynomialFeatures and Lasso are included in a ML pipeline to build model.

**Finding**

We predicted test dataset with the model built above and print out the feature’s coef to target value (a.k.a used car price).

|  |  |
| --- | --- |
| feature | coef |
| year | -16128.65 |
| odometer | -6783.81 |
| year^2 | 19229.41 |
| year \* odometer | 1058.22 |
| odometer ^ 2 | 4235.23 |

A graph with blue squares

AI-generated content may be incorrect.

The above data indicates:

1. The ‘year’ or ‘odometer’ feature of used cars are crucial in determining the used car price. Older cars or cars with high milage lead to lower price.
2. The quadratic terms (year^2 and odometer^2) suggest that the car price decreases with age and mileage, but at a diminishing rate. This may reflect the fact that while a new car depreciates rapidly, older cars stabilize in value once they reach a certain age or mileage.
3. The ‘year and odometer’ feature indicates that the relationship between car age and mileage is complex. A newer car with high mileage might be priced higher than an older car with the same mileage due to its newer model, despite the mileage.

**Recommendations**

I would recommend used car seller to sell used car when either it has newer year or less milage. The seller might be able to target premium buyer for newer cars buyer and target budget buyers for older and high milage car. Also, seller is able to consider to apply a better trade-in offer for newer used car and reduce offer for an older or high-milage car.

**Link of the Jupyter Notebook**

https://github.com/wjd20170117/WhatDrivesThePriceOfACar