Project Proposal

Problem statement: How to estimate the car price for customized cars with selecting features?

● Context: Liftit, a startup company focusing on car merchandise. The market team and management team want to create a product that can predict the MSRP of a customized car. The company saw it as a revolution to the market. Customers can get an idea what kind of features they can get within budget. Dealers can set the selling price with support from data.

● Criteria for success: Determine the price for a car with a selection of features that provides an accuracy above 80%.

● Scope of solution space: Based on the market performance of existing car models to decide what features contribute more to the MSRP. Create a prediction model that can give more than 80% accuracy.

● Constraints: The data only includes models from 1990 to 2017. More recent data is not available.

The cost behind features is unknown. How much time and effort up-front affect the price as well.

● Stakeholders: Marketing Lead, Executive Director, Director of Development

● Data sources: CSV file scraped from Edmund and Twitter with features including make, model, year, engine and other properties of car.

The purpose of this data science project is to come up with a pricing model for the startup company, Liftit. Lifeit's management team wants to develop a product that can predict the car price based on selected features. The team does not have a strong sense of what car features matter most for the customers and bring the most profit. This project aims to build a predictive model for car price based on a number of car features for different car models from well-known brands.

Exploring the data is the first step for the analysis. It is important to validate the dataset before using it for our prediction model. Then an exploratory data analysis shall be performed to identify what are the most correlated features with car price. Whether to isolate a portion of data or use the whole set of data for prediction? Grouping by market category or study only on non-luxury cars might give some different ideas on the car price. It is important to understand the brand value as well. When the data is ready, it is time to pick which modeling method works better. It would be a good idea to consider implementing Machine learning into the price prediction modeling. The final findings would be presented with a slide deck and a project report.