DATS 6103 Project Proposal

Team B

Hanzhong Gao

Zhuohan Yu

Aaron Yang

**Background:**

Second-hand vehicle or used vehicle has a huge market base. As it is feasible and cheaper, more and more people consider buying a used vehicle instead of buying a new one. However, the main reason for the huge used car market is that vehicle cannot maintain its value. The price of one vehicle will reduce because of its age, mileage, damage and so on.

**Questions:**

To better understand this market, our team has collected data and decided to do statistical analysis on the used car market. The following questions are the main subjects that our team focus on:

* Understand and analyze what factors have impact on the value of a used vehicle most.

1. The influence of a vehicle’s brand, body type, fuel type and gearbox type on the value of the vehicle.
2. The influence of a vehicle’s milage, engine power and whether the car has been damaged or not on the value of the vehicle.

* Predict the value of each used vehicle in our dataset.
* Give suggestions on how to buy a used vehicle with high cost effectiveness.

**Data:**

The data in this project contains 31 variables and more than 200,000 observations. 15 variables are anonymous we will drop them later. 150,000 observations will be train set and 50,000 observation will be test.

**Methods:**

In this project, we are planning to use logit regression or KNN for questions that involves categorical data. Also, we plan to use linear models and other tests for statistics. Finally, we plan to use graphs such as plots and histograms to visualize data.

GitHub Link: <https://github.com/Zhuohan-Data/TeamB-FinalProject.git>

Data Resource : <https://tianchi.aliyun.com/competition/entrance/231784/information>