Group 1 Phase 1

Title: Analysis of Used Car Prices in Belarus

Dataset: Kraggle Dataset: Used Cars Market in Belarus

Detailed Description: During this project we will analyze the dataset to investigate the lead indicator of selling price for the Used Car Market in Belarus. Visualization, and ML methods will be incorporated to effectively gather insights into our data and predict the most effective method of gauging the selling price of a vehicle in Belarus.

Specification: The dataset was collected on December 2, 2019 from one of the most popular online catalogs in Belarus and contains entries parsed from used car ads.

At least 5 Questions:

1. What is the distribution of manufacturers for each region and whether a manufacturer has a significant impact on the asking price of a vehicle? (Pie graph to get distribution of Manufacturers for each region). To find the impact of manufacturer on Price we can do a Multi-Way ANOVA Test or use Linear Regression).

2. What is the most popular model by region and whether we can conclude that the popularity of a model has a direct impact on the price of a vehicle?

Bar Graph, Impact on Price: Multi-Way ANOVA Test or use Linear Regression).

3. What is the average age of each vehicle manufacturer and whether impact of production year varies for each manufacturer on the selling price? Group by,Summarize, bar graph: Multi-Way ANOVA Test or use Linear Regression

4. What is the average asking price for each region and what impact does a region have on price, does body\_type make a difference on the asking price for a region?

5. What is the relationship between odometer and price and whether sellers are more likely to accept exchanges as the odometer value increases (Fill in with whether its exchangeable)? Scatter Plot. (Simple Regression Analysis)

6. What is the relationship between engine type and engine capacity and what is the impact of engine type and capacity on the selling price? Scatter Plot

7. Distribution of Engine type for each region and by car type? Mosaic Plot

8. Distribution of colors and whether it significantly impacts price. Bar Graph

9. Which variable has the largest impact on the selling price of a vehicle what impacts could explain the outliers for this correlation?

We will divide the questions 1-8 to each member and use our data from those questions to derive an answer to the last question.

Our means of communication will be discord for common discussion.

Github for editing code together and Google Docs for editing documents and pooling ideas together.

We will be communicating daily on our progress, recapping difficulties and successes each week.