Group 1: Emma Doyle, Reid Hoffmeier, Matthew Lane, Mikhail Mikhaylov, Brandon Strong

Dr. Zakariya Qawaqneh

CSC/CIS 453: Data Analytics and Predictive Modeling

Phase 1

4/9/2021

**Dataset:** Kraggle Used Cars Market in Belarus

**Description:**

For our project we shall analyze the dataset and answer questions we have about our dataset. We will use R to find patterns in our data and use R Markdown to show our findings.

**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.

**Questions (At least 5):**

* What is the distribution of manufacturers for each region and whether manufacturer has a significant impact on the asking price of a vehicle? **Solved by: Pie graph**
* 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? **Solved by: Bar graph**
* What is the average age of each vehicle manufacturer and whether manufacturer has an impact on the impact of production year on the selling price? **Solved by: Group by, Summarize, Bar graph**
* 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? **Solved by:** **Scatter Plot**
* 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)? **Solved by: Scatter Plot**
* What is the relationship between engine type and engine capacity and what is the impact of this on the selling price? **Solved by: Scatter Plot**
* Manufacturer Origin Distribution **Solved by: Pie graph**
* Distribution of Engine type for each region and by car type? **Solved by: Mosaic Plot**
* Distribution of colors to body type. **Solved by: Bar Graph**
* Which variable has the largest impact on the selling price of a vehicle what impacts could explain the outliers for this correlation? **Solved by: Inspection of all previous graphs**