Data Memo

Shiloh Rivers

4/10/2022

Dataset Overview

Vehicle Inspection Data https://www.kaggle.com/datasets/chitwanmanchanda/vehicle-inspection-data

predictors: 73 observations: 26307

The variables are the data from vehicle inspections. The information is primarily qualitative with a few quantitative variables.

Many of the variable have missing data. I believe this information was mined from data entry provided by mechanics. Likely, if there is a commonly accepted default (mechanically) for one of the variables the mechanic would save time by skipping filling the form out. So I will have to determine the commonly accepted default for categories that have a large portion of missing data. Luckily there is an included excel sheet describing the instructions for data entry.

Research Questions

I am interested in determining if engine rating can be predicted from engine readouts as recorded by mechanics working on the vehicle.

This question will be best answered with a regression approach.

I believe that engine oil and exhaust smoke will be particularly useful in determining a prediction.

The goal of my model is predictive in that I am primarily interested in determining the predictive power of the observations of the mechanics.

Timeline

I plan to implement the lessons that I learn from class and lab into my project on the weekend that I learn them, thereby keeping up with class pace.

Week 3 - Data cleaning
Week 4 - Exploratory Data Analysis
Week 5 - Refining Analysis/Run models
Week 6 - Determine Linear or more
Week 7 - Write up Results
Week 8 - First Draft
Week 9 - Edit
Week 10 - Begin Final Draft

Concerns

I do not have any questions or concerns at this time.