Gasoline Case Study Rubric

DS 4002 - Spring 2024 - Samuel Rea

Due: TBD

Submission Format: Upload your GitHub link to UVa Canvas

Individual Assignment

Executive Summary:

This document outlines the rubric for a case study focused on building predictive models for gasoline prices. The primary objective is to produce a GitHub link containing your code, conclusions, and any additional materials used to develop your predictive model for gasoline.

Why Am I Doing This?

This case study offers an opportunity to showcase your technical data science skills while also practicing soft skills such as presentation and effective communication of results. The case study will introduce you to various techniques in machine learning, data wrangling, and exploratory graph analysis.

What Am I Going to Do?

The materials for the case study can be accessed at this link: GitHub Link. Your goal is to work with the provided data and deliver a comprehensive conclusion. This process includes utilizing the Caret package in R to clean the data, performing exploratory graph analysis, constructing a predictive model, and presenting the model in an accessible manner. The GitHub link includes examples of code and models, as well as a sample final deliverable. The final deliverable can be presented in a PDF or presentation format, but the information remains consistent across both options.

What Will Be Turned In

- A PDF or presentation summarizing your conclusions and significant findings.
- A GitHub link containing all the materials you used.

How will I know If I succeeded?

You will meet expectations on this case study if the following rubric is successfully followed and completed.

Spec Category	Spec Details

Formatting	• One GitHub repository (submitted via link on Canvas)
	  ○ Create a GitHub Repository for this assignment containing the following
	  ▪ README.md file
	  ▪ LICENSE.md file
	  ▪ Source code
	  ▪ Data files
	  ▪ Final Deliverable
	  ▪ References
	• Final Deliverable
	  ○ A PDF or PowerPoint containing the conclusion

README.md	• Goal: this file orients readers to the GitHub and what it contains. ○ A reader should be able to navigate as well as know what each folder in the GitHub contains
LICENSE.md	• A basic MIT license should be used when creating the GitHub repository

Source code	• This should be a file in your GitHub that contains all of the code necessary to reproduce the results of case study
	  ○ Code should be commented out and easy to understand
	  ○ Code should produce a predictive model for gas prices

Data Files	• This contains all the data used to build the predictive model. The data can be the same as the one presented in the start of this case study but outside data can be used to make a better model as well • The data should be able to be downloaded and used if wanted

Final Deliverable	• This is a PDF document or PowerPoint containing a digestible conclusion for your predictive model
	  o There is an example of a final deliverable in the case study github
	  o Should contain summary statistics about the model, what model was used, how accurate the model was, challenges and ways to improve the model

References	• This is a PDF document containing all the outside references used to complete this case study. o Format references in APA