2025 Chevron Rice Datathon:

Predicting Vehicle Inventory for 2025

**Description**

Which vehicles will be on the road tomorrow? How many miles will be traveled by all of these vehicles? To understand future demand of fuels, we need to predict how many miles people will be driving and what vehicles people will be driving.

For this challenge, we will focus on predicting vehicle population. Vehicle population is simply an inventory of vehicles by features such as model year and type of vehicle. Vehicle inventories are available but most of these do not give information on the current year and do not generate insights into future years. So, we can look to data science to uncover insights and to make predictions for future years.

In this challenge, we will use data from 2019 to 2024 to predict vehicle population in 2025. The features include model year, vehicle type, fuel type, and number of vehicles registered to the same address.

**Data**

**DISCLAIMER**: The data provided for this challenge can be used only to develop a model for the challenge and should be discarded after the end of the challenge.

The challenge involves several files for your project. The files required to build and test your predictive model will be made available at the start of the Datathon.  The following files will be available at the start of the Datathon:

* **training.csv**: contains raw data used to build, train and test your model.
* **data\_dictionary.xlsx**: provides description of the variables in training.csv
* **submission\_format.csv**: gives the required format of the model predictions file obtained by applying your model to the scoring.csv file.
* **scoring.csv**: used to generate model predictions required for us to score your model accuracy
* **submission\_file.xlsx**: used to submit model predictions and calculate model accuracy. Input your model predictions into the “Predictions” sheet. After predictions are inserted, your model’s predictive accuracy will calculate automatically so ***do not make any changes to the “RMSE” sheet***.

**Project Submission**

Please submit your Datathon project for judging. At minimum, the submission should include a link to a github repo that contains:

* your project code
* a summary of your methodology and findings
* submission\_file.xlsx: the file in which you inserted the predictions outputted from running your model on the **scoring.xlsx** dataset

**Evaluation Criteria**

Projects will be evaluated holistically based on model accuracy the criteria outlined by Rice Datathon:

* **Technical Difficulty**: We are looking for technically advanced solutions to difficult problems that make use of a diverse set of modeling and data science techniques. That being said, if you can solve a challenging problem with a simple solution, we will be very impressed!
* **Analysis & Exploration**: We are looking for projects that take time to analyze and explore the nuances of whatever data they are working with.
* **Creativity**: We are looking for original ideas or new angles on existing ideas.
* **Predictive Accuracy**: predictive accuracy will be measured by the Root Mean Squared Error (RMSE) in **vehicle inventory** in the scoring.csv file:

Other areas that can be looked at include, but are not limited to feature generation, feature selection, and model selection/building.

**Rules**

1. Contestants must respect the privacy of the data and remove it from their computers upon completion of the competition.
2. Contestants' solution must be a model that is repeatable, adjusting model results manually (including “arbitrary” factors/constants) to tune model predictions is not permitted.
3. Contestants' prediction models can use only that data from the training.csv dataset provided to build the model.

**Prizes**

**FIRST PLACE:** TBD, thinking Sony headphones

**SECOND PLACE:** Physical $50 Amazon Gift Cards in Premium Gift Box for each winning team member