2023 Chevron Rice Datathon:

Predicting Renewable Energy Investments

## Description

Chevron is continuously looking for renewable energy businesses, initiatives, and start-ups to potentially invest in as part of their energy transition strategy. With each year, this task of narrowing down prospective collaborators becomes trickier and more difficult as the national investment in clean energy explodes throughout the United States. With the idea that the greater the investment in a state, the more likely Chevron is to find collaborators and companies of interest within that state, Chevron will narrow down the list of states to look for potential investments within using predictions for the amount of renewable energy investment dollars each state will be given.

## Objective

Provided that it is the year 2019, Chevron is planning for the next year and needs to know which US states seem most promising with regards to renewable investments. In this Data Science Challenge, you are tasked with developing a model to predict:

*The Renewable Energy Investments ($) of each US state for 2020.*

## Data

The challenge involves a couple of files for you to use. The files required to build and test your predictive model will be made available at the start of the Datathon. However, 30 minutes before the judging period starts, we will also release a scoring file that contains additional data for which you will run your final model and provide corresponding predictions.

The following files will be available on Piazza at the start of the Datathon:

* **Investment\_Data\_train.xlsx**: contains raw data used to build, train, and test your model.
* **MSN Codes.xlsx**: provides description of the codes in Investment\_Data\_train.xlsx

The following file will be available on Piazza 30 minutes before the Datathon judging begins:

* **Investment\_Data\_2020.xlsx**: used to generate model predictions required for us to score your model accuracy

## Project Submission

Please submit your Datathon project to devpost 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 **Investment\_Data\_2020.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 Gross Sold Quantity predictions for the stores in the scoring.csv file:

## where is the actual Gross Sold Quantity corresponding to the -th row in scoring.csv, and is the corresponding model prediction.

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:** Fujifilm Instant Film Cameras

**SECOND PLACE:** $50 Amazon Gift Cards