Queen City Hackathon 2020: Kaggle Track Prompt

21st February 2020

BACKGROUND

City planners regularly face the problem of limited budgets and have to be highly selective about the geographical areas that they choose to spend money on development. They will often have access to aggregated time series socio-economic data about those geographical areas, and they need to leverage that data in order to make the best possible decisions for the city.

TEAM OBJECTIVE

Accurately predict the target variable - which is a blended metric quantifying growth/decline in the economic health of an area over the last decade.

AUDIENCE

City planners deciding where to allocate resources.

DATASET(S)

You will find the following files at on Slack in the #kaggle-track channel

- **training.csv** - contains aggregate economic indicator data (416 rows, 496 columns), including the target variable. This data is to be used for training a machine learning model to predict the target.
- **testing.csv** - contains aggregate economic indicator data (45 rows, 495 columns), **does not** include the target. This data is to be used for generating predictions that will be evaluated for accuracy.
- **data description.csv** - data dictionary containing information about the columns.

JUDGING CRITERIA

The models will be evaluated using a weighted MSE metric. These errors will be weighted by the column 'Population _2018'. If you are unfamiliar with MSE, see the following resources:

- https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_squared_error.html
- https://en.wikipedia.org/wiki/Mean_squared_error

CODE

All code used must be open source and available for review by judges.

EQUIPMENT

Automated machine learning tools, such as DataRobot or SageMaker, are not permitted. Participants are required to bring their own equipment, but all final code must be posted for submission in source control for consideration. Judges will use their discretion in determining which tools are permissible. Participants should check-in prior to the start of the event to determine if the tools they plan to use are permissible.

SUBMISSIONS

1. Predictions should be submitted via Leaderboard website, which is provided in the Kaggle slack channel - http://qc-hackathon-2020.s3-website-us-east-1.amazonaws.com/.
   Note:- you MUST be on the Red Ventures network to access this URL

2. You need to take the following steps to submit -
   a. Register your team on the scoreboard app. Save your submission key.
   b. Create a folder called 'submission'.
   c. Include your file with predictions named 'predictions.csv'. This file should contain 1 column with **no header** with numerical predictions (45 rows - order determined by data in testing.csv).
   d. Inside the submission folder, create a folder called 'code'. This should contain the code you used to build your model and generate predictions. Submissions that are not reproducible will be disqualified.
   e. Zip the submission folder and submit using your team's submission key that you will get when you register on the scoreboard.
      Note: You can use this command if you are having issues with zipping:- "zip -r submission.zip submission"
3. Rules
   a. We will only consider the LATEST submission as your current score. So, you are responsible for keeping track of your own best models.
   b. There are mini prizes for the leading team at 12:30 am and at 7:00 am.
   c. You are only allowed to submit 8 times per hour.
   d. Your last submission time is 12:00 pm on Saturday.
   e. The leaderboard will stop updating in real time at 11:00 am for extreme suspense and intrigue.