About the Task

Create a scraper to get data from <u>one</u> of the following websites. The scrapper file should be in the .py format and scrapper must have a single python class which will be called to get the required data. The output should be in the csv format. Requirements:

- Only pick one of your trial tasks from the sources listed below
 Note: This is also a gauge of which type of data structures you are most comfortable with.
- Create scrapper
- Build clean standards, data should contain metadata along with all the values present in the dataset.
- Simple way to present your data in map, graphs or charts to provide synthesis and show analytical skills in a short report

The submission will be evaluated on the quality of the data output as well as code. Scrapper should be well optimized and able to handle large amounts of data. **The deadline for the task is 3 days.** Upload your code in your GitHub repo and push your code for us to evaluate.

Data Engineering and Backend Roles (Pick only ONE data source from below)

Time Series Data (Push your code to: https://github.com/Taiyo-ai/ts-mesh-pipeline)

- Apple Mobility Index
- Google Mobility Index
 Get the data for Top 50 Cities in the world for the above two sources on daily basis
- Git Commits for Keras, Tensorflow, MXNet, and other ML libraries
 - For time-series data aim is to get daily or weekly commits for the above git repos aggregated on country or city level for Apple or Google Mobility Index

Geospatial Data (Push your code to: https://github.com/Taiyo-ai/ts-mesh-pipeline)

- Pick one of the following Google Earth Engine API for the following locations:
 - The United States
 - Any daily measures like as <u>Terra Land Surface Temperature</u>

Projects and Tenders (Push your code to: https://github.com/Taiyo-ai/pt-mesh-pipeline)
Scrap data for the following sources by getting details of all the tenders present on the website:

- City of Sunnyvale Public Procurement
- UK Cabinet Contracts