**Maiden Century**

**Machine Learning Take Home Interview.**

**Assignment:**

**Input** [**https://github.com/ActiveConclusion/COVID19\_mobility/blob/master/apple\_reports/apple\_mobility\_report\_US.csv**](https://github.com/ActiveConclusion/COVID19_mobility/blob/master/apple_reports/apple_mobility_report_US.csv)

**Analize the covid situation across the various cities using the Apple mobility dataset. Showcase your ability to create relevant and interesting models to generate novel findings from the data. Account for seasonality. This problem is open ended by design.**

**The following are just ideas to consider, they are not meant as instructions for the direction that you are supposed to take the project.**

* **Univariate and multivariate time series analysis. Address seasonality etc. Possibly with changepoints, you could for instance determine which cities got Covid first and how it spread.**
* **Look for trend shifts.**
* **Possibly combine with US census data. From there you could look at factors such as age, income and many other demographic factors to augment your analysis.** 
  + If we want to join the mobility dataset to the U.S. Census data. The only way to do this is using geocoding, i.e.  converting AAPL mobility, at least on county level, but maybe all of them, to lat/longs. Then you can merge with census data, say using a 30km range and finally determine if income had any impact on how much people drive vs walk (this is the less important part to get right)

* + Here is some code to get you started, however these counties do not match with AAPL report counties, they have to be geocoded to be combined:

            censusIncome = censusdata.download('acs1', 2019,

                               censusdata.censusgeo([('state', state\_id),

                                                     ('county', county\_id),                                                     ]),

                              ['B19013\_001E']

                                   )

* **Possibly determine if mobility reductions had an impact on covid mortality rates by finding and pulling in external datasets.**

**How we evaluate:**

* Code quality – try to write enterprise level code, ideally using some OOP principals, error handling type hints etc. We are going to read every line of your code, so the quality of code even more important than the results.
* Clarity and communication – Each cell code should be labeled and include a short description of the intent of the code. Communicate that you understand the problem and explain how your code aims to help with the solution.
* We should not have to run your code; you should run your code and present all output in the notebook before submission
* As part of your workflow demonstrate OOP / classes, methods and well structured code.
  + I know that jupyter and OOP don’t always go together, but in this case you can take a cell and put a class into it with a few functions to achieve a specific part of the work.
* Print all charts and tables with clear and descriptive labels.
* Add comments to your code and output
* Show examples
* Highlight potential limitations and to-dos
* Remove/hide / suppress extraneous output if it makes your notebook difficult to read. The unhidden output which requires us to unnecessarily scroll too much or make your notebook confusing to read will be penalized.
* Use Notebook annotation / titling / output labeling to clearly describe each output, what it is, what the purpose is, and why it's important. Please label and title every cell.
* Creativity, want to see how you approach an analysis which is open ended and can be taken in multiple directions.
* Data science concepts like time series analysis will be highly valued.
* Show that you can transform, pivot and work the data with pandas.

**Submission:**

The core of the submission is a Jupiter notebook which is self-contained and fully executable. It should include all instillation of all dependencies ensuring it can function simply by running the entire notebook.

Create a GitHub repo and submit your notebook and any data files which you are using for your project.

Important: Run all your code and generate the required output before submission. Ideally we should not have to run any of your code since the output to your code should be clearly visible below every cell.