Data Analyst L4 Hackathon

L4 September 21 Cohort

Introduction

The Data Analyst hackathon on **12th and 13th July** gives you an opportunity to work in a team and conduct exploratory data analysis on a new dataset. You will be required to deliver a 10 minute presentation about the outcome of your work.

You are with different datasets to conduct exploratory data analysis on Nigeria health facility data & covid19:

- Nigeria NMIS health facility data
- Population data for Administrative Zone 1 (states) areas in Nigeria
- Map boundaries for Nigerian states (for plotting and binning)
- Covid cases across Nigeria (as of May 20, 2020)

You are given a sample of possible questions to investigate and tasks to undertake. However, you are free to design your own tasks as a team and augment the initial datasets for your own analysis.

Schedule

Tuesday 12th July

- 9.30 am - 10 am - Induction
- 10 am - 5 pm - Working in your teams

Wednesday 13th July

- 9.30 am to 3 pm - Working in your teams
- 3 pm to 5 pm - Presentations

The dataset

- This document builds on the dataframes found in this notebook.
- There are 4 datasets we'll be primarily looking at:
- Population data of Nigerian states (file: nga_admpop_adm1_2020.csv), COVID cases in Nigeria (up through May 20th 2020), health facility data across Nigeria, and geodata for Nigeria to link states with geo-coordinates.

Getting started

Many interesting questions with this data relate to COVID cases over time. Some examples:

- Start by undertaking exploratory data analysis of the dataset, what useful insights can you find through summarisation and data visualisation?

- How are the 'previous week' and 'next week' columns related?

- What about geo-spatial information? You would expect that if a state had high COVID cases, that neighboring states would also see rising cases at some delay. How might you do this? One new direction to consider if you have time could be to explore how to use the Geoplot and Geopandas library to develop geospatial visualisation that are digestible to stakeholders: https://geopandas.org/gallery/plotting_with_geoplot.html

- Find another dataset from one of these https://nigeria.opendataforafrica.org/, https://data.humdata.org/group/nga) websites (or another!) that interests you and explore it. What kind of data is in it? What characteristics does it share with the given datasets that you might connect the two with (state, health facility name, etc.)? Is the data sparsely populated, with many missing values or is it relatively clean and complete? How can it help your analysis?

Output

You are expected to produce a jupyter notebook that can replicate your work. You will then prepare and present a 10 min presentation (PPT) with the following structure:

- Who you are (30 s)
- Elevator pitch of problem (30 s)
- Explain what relevant tools / techniques / relevant modules in the programme you have used (3 min)
- Describe how you applied them / what solution and WIP currently looks like (3 min)
- Any results (2 min)
- Further work (1 min)
- Q&A: 5 minutes

The presentation will be scheduled from **3pm onwards on Wednesday 13th July**