Career Lab 2020

MAXX POTENTIAL

Part 1: Find Interesting Data

Great Data Sources:

- https://www.data.gov/
- https://www.kaggle.com/
- https://data.world/datasets/open-data
- Many others out there, feel free to find data with any other tool

Some Guidelines:

- No such thing as a perfect data set
    - Several will be missing data, this is normal. Sensors fail, collection points get missed, typos happen. Try to find the best possible data set with the least amount of missing data (< 5% per field missing)
    - Data sets vary in size, and for this activity you'll want to balance between large and small sets. The larger the data set, the more difficult it can be to work with. The smaller the data set, the less insight it can provide.
- Various forms exist - JSON, CSV, etc. Finding a CSV will be easier to work with for most beginners, to get through some of the challenge problems you may want JSON or to have both available.

Data Case Studies:

- https://www.sas.com/en_us/customers.html
- https://www03.ibm.com/software/businesscasestudies/us/en/corp
- https://www.informs.org/Impact/O.R.-Analytics-Success-Stories

Part 2: Prepare Your Environment

Useful Software:

- Excel / Numbers / Google Sheets
    - Useful to explore data, and complete activity if you're uncomfortable with code
- Recent Python Version
- Python Pandas
    - Intro to Pandas
- Other Data Visualization Tools
    - seaborn
    - Bokeh
    - plotly