DS 203 : Programming for Data Science
Tutorial and Assignment Sheet – 3

Submission guidelines:

- Prepare an ipython notebook for Q1, and name it <roll no.>.ipynb
- Prepare a pdf file answering rest of Q1 and all of Q2 and Q3 and name it <roll no.>.pdf
- Zip the two files into a single file and upload on Moodle before 11:59pm on September 9, 2020

1. Download, read, and display the data at the following URLs in in python using Google CoLab and pandas, and print the type of each variable. Comment on the difference between python data types (float, int, object etc.) and the data types taught in class (categorical/nominal, ordinal, numerical (integer, quantized, continuous) etc.).
    a. http://download.macrotrends.net/assets/php/stock_data_export.php?t=GOOG . You can replace GOOG with any other stock symbol, such as MSFT.
    b. https://archive.ics.uci.edu/ml/datasets/Amphibians
    c. https://data.gov.in/resources/quarterly-range-wise-performance-public-facilities-operation-minor-no-or-local-937

2. Assume that you are analyzing work travel habits of people from various localities of Mumbai. Classify the following into types of analyses into exploratory, descriptive, predictive, or prescriptive:
    a. Finding the number of samples that have *distance traveled* variable missing in the data
    b. Finding whether the distribution of *distance traveled* by commuters is Gaussian or beta
    c. Plotting histograms of number of people by *residence locality* variable in your data
    d. Finding whether people from *Bandra* and *Powai* have different *distance traveled* distributions
    e. Analyzing net savings in carbon footprint if a new train station is added to *Bandra* versus *Powai*
    f. Modeling *distance traveled* as a function of income, job type, and residence locality
    g. Finding ranges of *distance traveled* variable in the data

3. For each of the following scenarios, search for datasets related to the problem domain (even if the data is not pertaining to the exact situation) for a few minutes to get a sense of what data is collected around the world. Then exercise your imagination to write down reasonable exploratory, descriptive, predictive, and prescriptive data analyses to be done in case of each of the following hypothetical scenarios. Indicate sources of a few other data sets that you find related to each theme. Feel free to indicate if some of these categories of analyses do not apply to a particular scenario. Some loosely related links are provided:
    a. As an analyst for a stock market newsletter, you want to recommend bell-weather stocks for different sectors. See: https://www.investopedia.com/terms/b/bellwether-stock.asp
    b. As an intern at the Ministry of Environment, you are under pressure to approve one of the two roads that have been proposed, and you want to recommend the lesser of the two evils.See: https://www.nbmcw.com/tech-articles/roads-and-pavements/18263-clearances-required-under-environment-acts-for-highway-projects.html
    c. As an advisor to a state government, you want to close the gap between the neonatal mortality in the biggest city versus rest of the state, but you have limited resources to work on only a few hospitals. See: https://pubmed.ncbi.nlm.nih.gov/23734339/