DS 203

Assignment4

Gyandev Satyaram Gupta, 190100051

1. 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.).

   Solution: In a way the python data types are that in what form is the data being stored in the variable in the computer memory like it is float, int, object, etc whereas the data type taught in class was about the type of value thats stored in the variable. In a single python data type we can have multiple data type taught in class. For e.g an integer python data type can be numerical,categorical, ordinal data type covered in class.

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:

   Solution:

   (a) Finding whether people from Bandra and Powai have different distance traveled distributions
   Descrptive

   (b) Analyzing net savings in carbon footprint if a new train station is added to Bandra versus Powai
   Prescriptive

   (c) Modeling distance traveled as a function of income, job type, and residence locality
   Predictive

   (d) Finding ranges of distance traveled variable in the data
   Exploratory

   (e) Finding the number of samples that have distance traveled variable missing in the data
   Exploratory

   (f) Finding whether the distribution of distance traveled by commuters is Gaussian or beta
   Descriptive

   (g) Plotting histograms of number of people by residence locality variable in your data
   Exploratory

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 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/

b. 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

c. 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

Solution:

(a) Exploratory Data Analysis:
We will reasearch and find out datasets related to neonatal mortality in the biggest city and also for the other states. We would do data pre processing like fixing the data if variables are missing and correcting them. Looking for similar variables in both sets which can help us. Finding coorelations between variables and plotting graphs of neonatal mortality with other variables

Descriptive Data Analysis:
We can look to answer here for our multiple questions like do the given factor/variable affect the mortality rate? Do literacy affect the rate compare to big cities and others? Do big cities have larger neonatal rate or small cities ? What are the major reasons for the deaths

Predictive Data Analysis:
We can predict what variables could decrease the neonatality rate. We could also predict which variables is the area of concern and where we could work on it.

Prescriptive Data Analysis:
As to whatever we had predicted we can provide different ways or methods to bring down neonatality rate and which areas to work in. Which regions of society to focus on can be conveyed.
Other link to refer: reference

(b) Exploratory Data Analysis:
We will research and find out datasets related to stocks. We would analyze the stocks high, low, open, close, volume and trends. We would do data pre-processing like fixing data for missing variables and discarding any unnecessary row like dividends, earnings,etc. We make plot to analyze them and probably opening and closing price would give us a better highlight.

Descriptive Data Analysis:
We would look for various stocks and market leading stocks and analyze the trends. We can make multiple plots like histograms and line graph and even candle sticks graph for stock markets. Coorelating variables and then finding which a better stock for investment? Which sector is dominating stocks? Which stocks affects industry average?

Predictive Data Analysis:
We can predict after the analysis the future values of the stocks. We can predict what could be the market trend probably unless some unnecessary disaster happens. Also check whether stock could be bearish so that one shouldn't go for it

Prescriptive Data Analysis:
After all the analysis we can prescribe which stocks can be considered in bellwether stock
Other link to refer: reference, one can looks for different stocks and news

(c) Exploratory Data Analysis:
We would perform data pre-processing tasks on the data set to fixed the variables and entry. Explore the past data for road development in the areas nearer to the region

Descriptive Data Analysis:
We can look to answer here for our multiple questions like do the given which road can be built up in manageable money/amount ? which road can be badly affected by external factors like rain, wind etc ? Status of past development of road if any was done or being done?

Predictive Data Analysis:
We can predict which road can have less effect on the environment and positive effect on the economy sector

Prescriptive Data Analysis:
As to whatever we had predicted and by our analyses we could prescribe with which road to go ahead
Other link to refer: reference