DS203 ASSIGNMENT 4

Question 1.

Usual Python data types like int, float, etc. are used for a broad class of variables that take real values either discrete or continuous. Python also supports “objects,” a custom data type, which can be used to group several attributes of a thing (say, the name and age of a person) under a single data type.

On the other hand, a nominal/categorical variable is something that cannot be quantified. It can only be classified into different categories but does not possess any intrinsic ordering to the types. An ordinal variable is very similar to a categorical variable, with the exception that, there is an explicit ordering of the variables. Numerical variables are somewhat identical to the usual python types; for example, an integer is analogous to the int type in python. A quantized variable is something that only takes discrete values within some specified range, while a continuous variable takes a spectrum of values within a range.

Question 2.

1. Descriptive
2. Predictive
3. Exploratory
4. Exploratory
5. Exploratory
6. Descriptive
7. Exploratory

Question 3.

a) Firstly, I would find data quantifying the district-wise neonatal deaths in the state with the most probable causes associated with each one of them. Certain relevant studies on the determinants of neonatal mortality are - https://pubmed.ncbi.nlm.nih.gov/23734339/

Exploratory:-

I would broadly look at the data to get an idea about the data available. An example would be whether the data correlates the neonatal mortality in a particular district with factors such as the level of hospital sanitation, medicines available, etc.

Descriptive:-

I would plot the data to compare the district-wise situation and identify the districts where the correlation between poor hospital conditions and increased neonatal deaths is the highest.

Predictive:-

It will be possible to predict which hospitals in particular districts require the most urgent intervention efforts to bring down the state's neonatal mortality from the data.

Prescriptive:-

On the of the above data analysis, it is also possible to identify the corrective measures and reforms to be taken in the worst-hit hospitals, such as focusing on the sanitation conditions.

b) Firstly, I would gather stock price data of say, the biggest 100 companies in a particular sector, from the past 5-10 years to gauge the company's long-term performance in that specific sector. Also, to get a broad sense of the economy's state, I would again gather data depicting GDP trends over previous years. Furthermore, to get an estimate on the size of a company relative to others, I would also look at the market values of the different companies.

Some relevant data sources are -
https://finance.yahoo.com/quote/VTI/history?p=VTI

https://www.quandl.com/data/NSE/TATAGLOBAL-Tata-Global-Beverages-Limited

https://tradingeconomics.com/india/gdp

Exploratory :-

Initially, I would look at all the relevant data gathered from different companies over different years, searching for any missing values or discrepancies in the data, preventing a direct comparison. For each particular sector, I would plot the share price data of the top-performing companies. Further, I would also depict the average share prices of all the companies in a sector to better understand how the sector performs as a whole.

Descriptive :-

In a particular sector, I would see which company correlates best with the overall performance of the economy over the years using the GDP data as an indicator of the state of the economy. I would also see which company has the best correlation with a particular sector's average performance over different periods.

Predictive :-

With all this information, it is possible to predict the company whose stock price most closely correlates with the GDP trends and its particular sector, with the company also having a high market value.

Prescriptive :-

The stock from this predicted company can be viewed as the bell-weather stock of that sector, helping us meet our objective.

c) For the construction of a road at a particular location, plenty of factors come into play. The most important one is the impact of its construction on the environment in solid waste, toxic generation, air pollution, water pollution, etc. Supplementary factors can include other significant things like approval given through general public opinion in the area. So, to compare the feasibility of the two road construction projects, I would first collect the relevant information to each such as the number of materials to be invested, expected vehicular activity and corresponding air pollution levels, proximity to forests, or amount of area to be cleared. Also relevant, among other things, would be proper surveys so that the people getting affected can properly voice their opinions.

Exploratory :-

Firstly, I would like to take an overview of the obtained data and verify that there are no issues with it, such as missing values of a relevant factor for one of the roads. Further, I would also plot the values of quantifiable factors of one road with the other’s corresponding values.

Descriptive :-

On the basis of the plots, I would see which project is relatively better on the more important factors such as causing the least environmental pollution, farther from forested areas, requiring a lesser space to be cleared, etc. and also which project has a lesser public objection to go ahead.

Predictive :- Compiling all this information, it is possible to predict which project would be better to go ahead, focusing on the more significant issues.

Prescriptive:- As a result of analyzing all the data, we can recommend which projects should go ahead, helping us meet our objective.