Anmol Saraf

Email: 200070007@iitb.ac.in

Course: DS203 - Programming for Data Science Minor

Instructor: Dr. Amit Sethi

Assignment - 4

Roll No.: 200070007

Semester: Autumn 2021

Due Date: 5th September, 2021

Question 1:

The Python data types (int, float, object, etc.) are defined on the basis of numerical or non-numerical. Here, int and float are for numerical values and object for non-numerical values. Further, int denotes integers whereas float denotes numbers with a decimal point upto a certain precision.

While, the data types taught in class (categorical/nominal, ordinal, numerical(integer, quantized, continuous), etc.) are not defined on the basis of whether they are numerical or not. But, rather are based on the qualitative or quantitative nature of it.

For example, categorical/nominal data types are variables with no order or ranking, like, Gender or Roll No., it can be of both numerical or non-numerical type. Ordinal data type would be the variables which have an ordered series like, ratings. The Numerical data types can be used for quantitative representation like, length or temperature.

Question 2:

These analyses can be classified into exploratory, descriptive, predictive or prescriptive as follows:

a. Exploratory

b. Prescriptive

c. Predictive

d. Descriptive

e. Exploratory

f. Descriptive

g. Exploratory

Question 3:

Part a

Exploratory Extracting the information from 171,529 singleton live births using the data from the most recent round of the District Level Household Survey conducted in 20007-08.

Descriptive The odds of neonatal death were lower for neonates born to mothers with secondary level education compared to those born to illiterate mothers. Also, the odds were lower for infants born to unemployed mothers compared to those who worked as agricultural

worker/farmer/laborer.

Predictive Predicting approximate determinents which could possibly have had an impact on neonatal mortality. These could be mother's age at time of birth, sex of the infant, rest taken by mothers during pregnancy, family members to help during the pregnancy, etc.

Prescriptive Targeting vulnerable groups like young, first time and Scheduled Caste mothers, and improving overall household environment by increasing access to improved toilets, electricity, and pucca houses(solid or permanent) could also contribute to further reductions in neonatal mortality in rural India.

Dataset `http://rchiips.org/pdf/INDIA_REPORT_DLHS-3.pdf`

Part b

Exploratory Collecting various datasets that can predict the best bellwether stocks for a sector. This can be done for various sectors as well.

Descriptive How does the number of hours put in by employees affect the chances of a stock from becoming a bellwether stock.

Predictive How much more occupation, revune can be generated from a bellwether stock.

Prescriptive Since the bellwether stock changes reflect in the market, what is the gap of this reflection period and what can be done to minimize losses.

Dataset `https://www.atlantis-press.com/article/125925420.pdf`

Part c

Exploratory Collecting data related to areas or sectors which are affected by the construction of roads. And those areas which are causing the problems in the construction of roads.

Descriptive How many sharp turns will the road need which can be hazardous or cause accidents.

Predictive Due to the urbanization of the area surrounding the road, how many new alcohol related accidents will increase because of this.

Prescriptive Which area will have more wildlife disturbed during and after the construction of the road and thus which should be avoided.

Dataset `https://data.gov.sg/dataset/causes-of-road-accidents-causes-of-accidents-by-severity-of-injury-sustained`