|  |
| --- |
| **PYTHON : DATA VISUALIZATION QUESTIONS** |

1. **a)** Load the **‘Student Performance’** dataset into one of the data structures (NumPy or Pandas).

**b)**Display header rows and description of the loaded dataset.

**c)** Remove unnecessary features (E.g. drop unwanted columns) from the dataset such as ‘lunch’ and ‘test preparation course’ .

**d)** Manipulate data by replacing empty column values in ‘parental level of education’ with a default value.

**e)** Convert theattribute‘race/ethnicity’to have ‘groupA’ to be ‘Asian Students’, ‘groupB’ to be ‘African Students’ , ‘groupC’ to be ‘Afro-Asian Students’, ‘groupD’ to be ‘American Students’ and ‘groupE’ to be ‘European Students’.

**f)** Perform the following visualizations on the loaded dataset:

**i)** Tally of the Number of Male & Female students who took up the ‘test preparation course’ and those who did not.

**ii)** Total Number of Male & Female Students belonging to each student group

**iii**) No of students who ‘failed’(less than 40), ‘second class’(between 40 & 50).

‘first class’(between 60 & 75) and ‘distinction’(above 75) in ‘Maths’,

‘Reading’ and ‘Writing’.

**g)** Find the average Maths, Reading and Writing Score of each Group (Ethnicity)

1. **a)**Load the **‘Black Friday**’ dataset into one of the data structures (NumPy or Pandas).

**b)**Display header rows and description of the loaded dataset.

**c)** Remove unnecessary features (E.g. drop unwanted columns) from the dataset such as ‘User\_ID’, ‘Product\_ID ‘ ‘Stay\_In\_Current\_City\_Years’ .

**d)** Manipulate data by replacing empty column values in ‘City\_Category’ with a default value for the city.

**e)** Rename theattribute‘City\_Category’to have ‘A’ to be ‘Metro Cities’, ‘B’ to be ‘Small Towns’ , ‘C’ to be ‘Villages’.

**f)** Rename theattribute‘Product\_Category\_1’to have ‘Baseball Caps’, ‘Product\_Category\_2’to have ‘Wine Tumblers’ and ‘Product\_Category\_3’to have ‘Pet Raincoats’

**g)** Convert theattribute‘Marital\_Status’to have ‘1:Married’ and ‘0:Un-Married’

**h)** Perform the following visualizations on the loaded dataset:

**i)** Tally of the Number of Male & Female persons who bought ‘Product\_Category\_1’ and ‘Product\_Category\_2’.

**ii)** Total Number of Male & Female persons belonging to each city category

**Black Friday Dataset Description:**

A retail company “ABC Private Limited” wants to understand the customer purchase behaviour (specifically, purchase amount) against various products of different categories. They have shared purchase summary of various customers for a selected high volume products from last month.

|  |  |
| --- | --- |
| * User\_ID * Product\_ID * Gender * Age * Occupation * City\_Category | * Stay\_In\_Current\_City\_Years * Marital\_Status * Product\_Category\_1 (Baseball Cap) * Product\_Category\_2 (Wine Tumblers) * Product\_Category\_3 (Pet Raincoats) |

1. **Python** for Data Science - Perform Data Visualization on **Iris Dataset**

**a)**Load the Titanic dataset into one of the data structures (NumPy or Pandas).  
**b)**Display header rows and description of the loaded dataset.

**c)** Clean the data if applicable

**d)** Find the average petal width of each category of IRIS Species

**e)** Data Visualization for:

**(i)** How many flowers of each species exists for each value of sepal width

**(ii)** How many flowers are there whose petal width is <1, between 1 to 2 and >2

**(iii)** Tally the Iris-Versicolour and Iris-Virginica species according to the value of Sepal Width

1. **Python** for Data Science - Perform Data Visualization on **Titanic Dataset**  
   **a)**Load the Titanic dataset into one of the data structures (NumPy or Pandas).  
   **b)**Display header rows and description of the loaded dataset.  
   **c)** Remove unnecessary features (E.g. drop unwanted columns) from the dataset.  
   **d)** Manipulate data by replacing empty column values with a default value.  
   **e)** Perform the following visualizations on the loaded dataset:  
     **i)** Passenger status (Survived/Died) against Passenger Class  
     **ii)** Survival rate of male vs female  
     **iii)** No of passengers in each age group