Practice Questions

I – Sales Dataset

- Read sales2.csv file into a dataframe.
- Display the first few records.
- Describe the dataset.
- Check for NA values.
- Plot a bar graph for ord_date Vs Purchase amount.
- Fetch the details of customer id 3001
- Update the purchase amount of order Id 70012 as 500.
- Retrieve the order no and customer id of the maximum purchase amount
- Retrieve the order no, customer id and purchase amount whose value is greater than 1000.
- Fetch the order details that are placed during the month of October.
- Split the dataset into groups on customer_id to summarize purch_amt with respect to each salesman id
- Calculate percentage of purch_amt in each group and rename purch_amt to purchaseAmount.

II Iris Dataset

- Load the iris.csv file into a dataframe.
- Displaying the shape of the dataset.
- Display a sample of 10 records.
- Display the records from 10th row to 20th row.
- Drop the duplicate records.

- Count the number of unique species
- Sort the data based on the sepal length
- Calculate sum, mean and mode of petal length.
- Extracting minimum and maximum value from sepal width
- Fetch the records with petal length > 4cm
- Split the data based on species type and find the mean sepal length
- Add a new attribute ‘total’ that stores the sum of sepal width and petal width
- Set the values of the rows 10 to 19 of the column 'petal_length' to NaN¶
- Fill the NA values with its mean
- Plot a histogram to show the species distribution