Name: Madhuri Ramakrishnan

Prn:119A3032 TE IT E2

EXPERIMENT 3

**Aim**: Data Visualization / Exploratory Data Analysis for the selected data set using Matplotlib

* Create a bar graph, contingency table using any 2 variables.
* Create normalized histogram.
* Describe what this graphs and tables indicates?

**Theory:**

Data visualization in python is perhaps one of the most utilized features for data science with python in today’s day and age. The libraries in python come with lots of different features that enable users to make highly customized, elegant, and interactive plots.

A contingency table is a table showing the distribution of one variable in rows and another variable in columns. It is used to study the correlation between the two variables. It is a multiway table which describes a dataset in which each observation belongs to one category for each of several variables

Matplotlib-Matplotlib is a visualization library in Python for 2D plots of arrays. Matplotlib is written in Python and makes use of the NumPy library. It can be used in Python and IPython shells, Jupyter notebook, and web application servers. Matplotlib comes with a wide variety of plots like line, bar, scatter, histogram, etc. which can help us, deep-dive, into understanding trends, patterns, correlations. It was introduced by John Hunter in 2002.

Histograms are visual representations of 1) the values that are present in a data set and 2) how frequently these values occur. The histogram shown above could represent many different types of information.

Histograms are extremely effective ways to summarize large quantities of data. By glancing at the histogram above, we can quickly find the frequency of individual values in the data set and identify trends or patterns that help us to understand the relationship between measured value and frequency.

**Implementation:**

# Create a bar graph, contingency table using any 2 variables.

**Bar graph:**

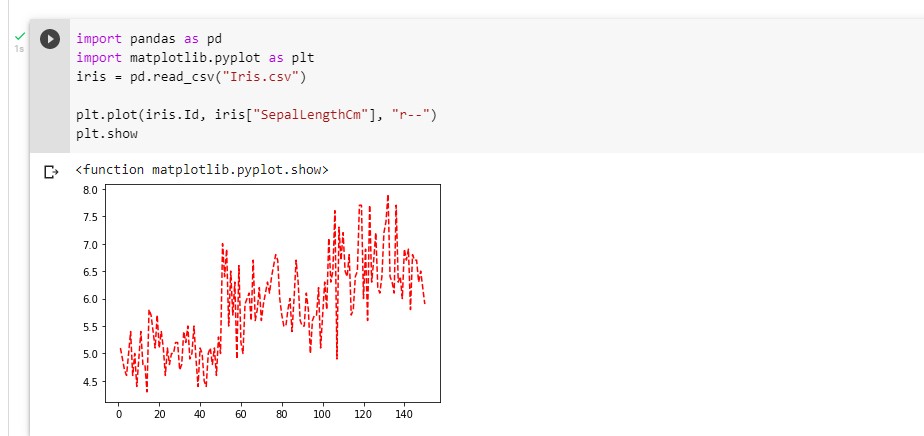
**Code:**

import pandas as pd

import matplotlib.pyplot as plt iris = pd.read\_csv("Iris.csv")

plt.plot(iris.Id, iris["SepalLengthCm"], "r--") plt.show

**Output:**

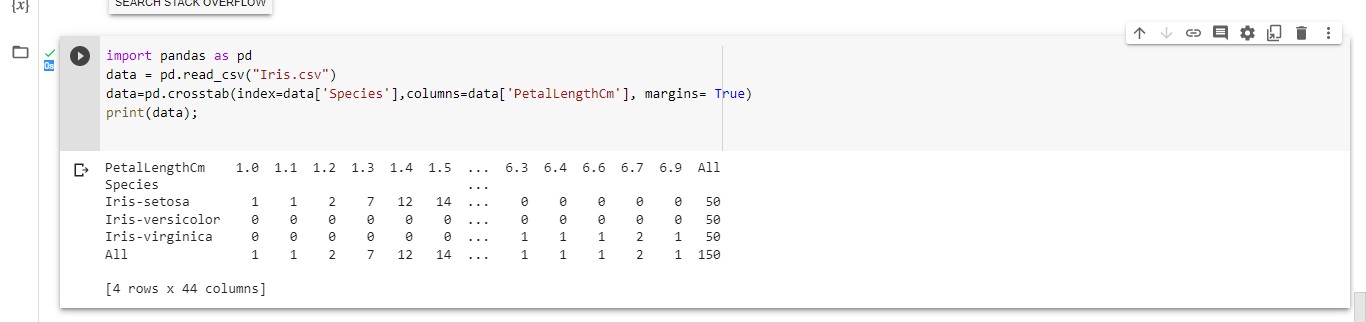


**Contingency Table Code:**

import pandas as pd

data = pd.read\_csv("Iris.csv") data=pd.crosstab(index=data['Species'],columns=data['PetalLengthCm'], margins= True) print(data);

**Output:**



# Create normalized histogram.

**Code:**

import numpy as np import pandas as pd

import matplotlib.pyplot as plt data = pd.read\_csv("Iris.csv") plt.figure(figsize = (10, 7))

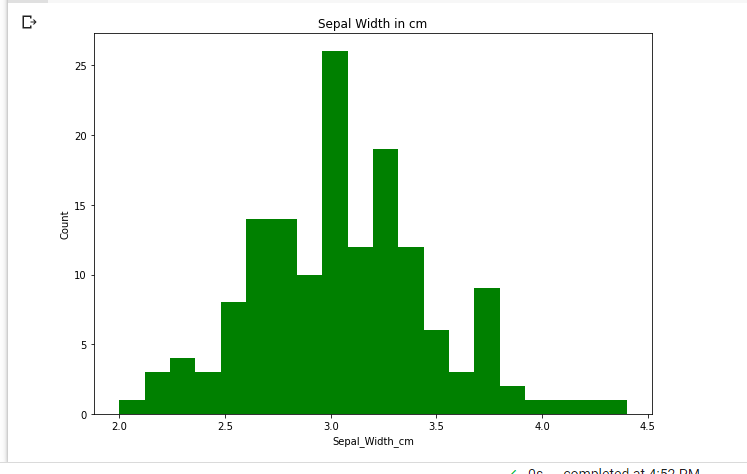
x = data.SepalWidthCm

plt.hist(x, bins = 20, color = "green")

plt.title("Sepal Width in cm") plt.xlabel("Sepal\_Width\_cm") plt.ylabel("Count")

plt.show()

**Output:**



# Describe what this graphs and tables indicates?

These are the charts/plots that are used to observe and display relationships between variables using Cartesian Coordinates. The values (x: first variable , y: second variable) of the variables are represented by dots. Scatter plots are also known as scattergrams, scatter graphs, scatter charts , or scatter diagrams. It is best suited for situations where the dependent variable can have multiple values for the independent variable.

A contingency table is a table showing the distribution of one variable in rows and another variable in columns. It is used to study the correlation between the two variables. It is a multiway table which describes a dataset in which each observation belongs to one category for each of several variables

**Conclusion:** Hence we successfully performed Data Visualization / Exploratory Data Analysis for the selected data set using Matplotlib