PROJECT - 1 Fundamentals of data(mean,median,mode)

#importing python libraries

Line no 1- import pandas as pd

=> !pip install pandas

Line no 2- import numpy as np

=> !pip install numpy

Line no 3- import matplotlib.pyplot as plt

=>!pip install matplotlib

Line no 4- import seaborn as sns

=> !pip install seaborn

 # Sample dataset: Student Scores
Line no 5- data = [“RAMESH”,“SURESH”,“MILIND”,“RINKU”
“NARESH”,“MUKUND”,“MANOJ”,“VIJAY”]
Line no 6- math_score = [78,85,92,88,76,95,89,83,91,87]
Line no 7- df=pd.DataFrame(data)
=> student variable name for value stat concepts
math_score=> variable name for value (stat concepts)
The Dictionary Data we have convert into pandas DataFrame using
Pd.DataFrame()

 # Compute descriptive statistics
Line no 8- mean_score = np.mean(df[‘math_score’])
=> Mean_score -> Variable name
=> = > Assignment Operator
=> np.mean -> Function used to calculate mean
=> (-> Tuple Data Type
=> df -> calling the df variable
=> [-> List Data Type
=> ‘Math_Score’=> Calling the variable name ‘Math_Score’
=>] => List Data Type
=>) => Tuple Data Type
Line no 9- Median_score = np.median(df[‘Math_Score’])
=> Median_score -> Variable name
=> => Assignment Operator
=> np.median-> Function used to calculate median
=> (-> Tuple Data Type
=> df-> calling the df variable
=> [=> List Data Type
=> ‘Math_Score’=> Calling the variable name ‘Math_Score’
=>]=> List Data Type
=>)=> Tuple Data Type
Line no 10- Mode_score = df[‘Math_Score’]).Mode()[0]
=> Mode_score -> Variable name
=> => Assignment Operator
=> df-> calling the df variable
=> [-> List Data Type

=> [=> List Data Type
=> ‘Math_Score’=> Calling the variable name ‘Math_Score’
=>]=> List Data Type
=>)=> Tuple Data Type

=> mode() -> Ready Made Function to calculate Mode Value
=> ->[0] => we have to calculate most frequent value which is start with Index zero

Line no 11- print(f'Mean Score: {Mean_Score}")

=> print -> used to get the output on the console.
=> (-> Function
=> f-> formatting string
=> Mean_Score:{'Mean_Score''}, —> Console Output value

Line no 12- Same
Line no 13- Same

 # Visualizing the data

Line no 14- plt.Figure(Figsize=(8,5))

=> plt.Figure() -> plt stands for plot, Figure stands for Diagram
(it is the predefined function)

=> Figsize-> it is used for requirement of Fig where
8 is width in inch
5 is Height in inch

Line no 15- sns.Histplot(df['Math_Score'],kde=True,bins=5,color='blue',alpha=0.5)

=> sns.Histplot-> is used to create Histogram for seaborn
(it is the predefined function)

=> df['Math_Score'] -> used to plot math score on histogram

=> Figsize-> it is used for requirement of Fig where

=> df['Math_Score']

=> kde -> Kernel density Distributioni used for smooth distribution

=> bins -> shown in diagram

=> color -> Blue

=> alpha -> Transparency of bar (always in between 0 to 1)

Line no- 16

plt.aXVline(Mean_Score,color='blue',linstyle='dashed',linewidth=2,label=f'Mean:{Mean_Score}')

=> plt.axvline=> ax-> Axis
v-> Vertical
l-Line

=> Mean_Score-> Mean value of math

=> color -> red

=> instyle -> dashed

=> linewidth-> 2 (it is always in between 1 to 3)

=> label -> legend

Line no- 17 Same

Line no- 18 Same

Line no- 19 plt.legend()

=> used to show legend(plot)

Line no- 20 plt.title('Distribution of Math score with measure of central Tendency')

Line no- 21 plt.xlabel('MathScore')

=> we have to used for plotting in x_axis with name of mathscore

Line no- 22 plt.ylabel('Frequency)

=> we have to used for plotting on y-axis with the name of frequency

Line no- 23 plt.show()

=> used plot on screen

Graph Explanation:

Histogram => Blue
Kde => Blue
Mean => Redline
Median => Greenline
Mode => orange
Frequency =>represent the count of student who achieve mathscore
MathScore => it is a mark math subject.