

```

# -*- coding: utf-8 -*-
"""Assignment-01:Infra-IBM

Automatically generated by Colaboratory.

Original file is located at
    https://colab.research.google.com/drive/1M5cMkE50_G2eNqOWjxM1M7d_Lvke516S
"""

"""Regd-no:2071

Name:Ch.Suryavamsi

Date:11-10-2022

importing required libraries:
"""

import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt

df=pd.read_csv("/content/Enrollments_28092022.csv")

df

"""1.Identify variables and their types?"""

#for quantitative it gives numerical data type
#for qualitative it gives descriptive datatyp
df.info()

df.dtypes

"""2.Size of data?"""

#No.of Rows and coloumns of data in form(rows,coloumns)
df.shape

#To find number rows in a dataset
nrows=df.shape[0]
nrows

#To find number of columns in a dataset
ncols=df.shape[0]
ncols

"""3.Prepare histogram for DEGREE,INTER,and10TH CLASS?"""

#Histogram of Degree column in data
df.DEGREE.hist()

# Histogram of Intermediate coulum in data
df.INTERMEDIATE.hist()

#Histogram of SSC column in data
df.SSC.hist()

"""4.Create piechart to represent the enrollements for each internship program"""

#Piechart for given data
plt.pie(dataFrame["REPRITITIONS"], labels = dataFrame["INTERNSHIPS"],autopct="%1.2f%%")
plt.show()

"""5.To find number of enrollements for each internship program?"""

dataFrame = pd.DataFrame({
    "INTERNSHIPS": ['Data Science','Cloud Computing Services','Mean Stack Web Development'], "REPRITITIONS": [156,90,51]
})
dataFrame

"""6.Find measure of central tendency (MEAN,MEDIAN,MODE) for Degree,SSC,10th class?"""

degreemean=df.DEGREE.mean()
degreemean

df.DEGREE.mode()

df.DEGREE.median()

df.INTERMEDIATE.mean()

df.INTERMEDIATE.mode()

df.INTERMEDIATE.median()

df.SSC.mean()

df.SSC.mode()

df.SSC.median()

"""7.Find measure of variance (MINIMUM,MAXIMUM,RANGE,MEAN-DEVIATION,STANDARD-DEVIATION,CO-EFFICIENT OF VARIANCE),for SSC,Intermediate,10th class?"""

```

```

df.describe()

sd=df.DEGREE.std()
sd

cv = lambda x: np.std(x, ddof=1) / np.mean(x) * 100
cv(df.DEGREE)

cv(df.INTERMEDIATE)

cv(df.SSC)

rd=(df.DEGREE.max()-df.DEGREE.min())
rd

ri=(df.INTERMEDIATE.max()-df.INTERMEDIATE.min())
ri

rs=(df.SSC.max()-df.SSC.min())
rs

"""8.Measures of position (STANDARD SCORES,INTER-QUARTILE RANGE OF DEGREE,INTER,10THCLASS?"""

import scipy.stats as stats

print("Standard scores of DEGREE:")
print(stats.zscore(df['DEGREE']))

print("Standard scores of INTERMEDIATE:")
print(stats.zscore(df['INTERMEDIATE']))

print("Standard scores of SSC:")
print(stats.zscore(df['SSC']))

"""9.Creating a box-plot and identify the outliers for DEGREE,SSC,INTERMEDIATE?"""

plt.boxplot(df['DEGREE'])
plt.show

plt.boxplot(df['INTERMEDIATE'])
plt.show

plt.boxplot(df['SSC'])
plt.show

"""NOW FINDING OUTLIERS?"""

#Outlier Function
def outlier(a):
    q1=np.quantile(a,0.25)
    q3=np.quantile(a,0.75)
    iqr=q3-q1
    upper_bound=q3+(1.5*iqr)
    lower_bound=q1-(1.5*iqr)
    print("Inter-Quartile Range:",iqr)
    outliers=a[(a<=lower_bound)|(a>=upper_bound)]
    print("the following are outliers in box plot:\n{}".format(outliers))

outlier(df['DEGREE'])

outlier(df['INTERMEDIATE'])

outlier(df['SSC'])

"""10.Identify number of students with 90% percentile for DEGREE,10TH CLASS,INTERMEDIATE?"""

#90% percentile in cgpa is 9.4 so it is taken as greater than(>) 9.4
df.DEGREE[df.DEGREE>9.4].count()

df.INTERMEDIATE[df.INTERMEDIATE>90].count()

df.SSC[df.SSC>90].count()

```