### **Task 6**

**Machine Learning**

Upload .py or Ipynb extension file on GitHub public repo “100DaysofBytewise" and share the link in the submission form by 2 July 2024.

* **Exercise:** Load a simple dataset (e.g., the Iris dataset from Scikit-Learn) and print the first 5 rows.

Code : def p\_head(x,n=5):

print (x.iloc[:n])

print(iris\_df.head())# pandas DataFrame that contains the Iris dataset.

* **Exercise**: Implement a function that takes a dataset and returns the number of features and samples.

Code:from sklearn.datasets import load\_iris

import pandas as pd

def func(x):

    num\_1=x.shape[0] # returns number of samples

    num\_2=x.shape[1] # returns number of features

    return num\_1,num\_2

num\_1,num\_2=func(iris\_df) # pandas DataFrame that contains the Iris dataset.

print(num\_1,num\_2)

* **Exercise**: Split a dataset into training and testing sets with an 80/20 split.

Code: from sklearn.model\_selection import train\_test\_split

import numpy as np

X = np.random.rand(10, 5)

y = np.random.randint(0, 2, 10)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

print("Training data size:", X\_train,y\_train)

print("Testing data size:", y\_test,X\_test)

* **Exercise**: Explore the basic statistics of a dataset, such as mean, median, and standard deviation for each feature.

Code: import pandas as pd

#data=pd.read\_csv('C:/Users/abdul/OneDrive/Desktop/ByteWise\_ML/USA\_Housing.csv')

#print(data.info())

data = {

    'feature1': [1, 2, 3, 4, 5],

    'feature2': [5, 6, 7, 8, 9],

    'feature3': [9, 10, 11, 12, 13]

}

data=pd.DataFrame(data)

m1=data.mean()

m2=data.median()

s\_d=data.std()

print('Mean : \n ',m1)

print('Median: \n ',m2)

print('Standard deviation : \n  ',s\_d)

* **Exercise**: Visualize the distribution of one of the features in the dataset using a histogram.

**Code**: import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

data = {

    'feature1': [1, 2, 3, 4, 5],

    'feature2': [5, 6, 7, 8, 9],

    'feature3': [9, 10, 11, 12, 13]

}

data=pd.DataFrame(data)

m1=data.mean()

m2=data.median()

s\_d=data.std()

print('Mean : \n ',m1)

print('Median: \n ',m2)

print('Standard deviation : \n  ',s\_d)

for median in df.columns:

    sns.histplot(df[median], kde=True)

    plt.title(f'Histogram of {median}')

    plt.xlabel(median)

    plt.ylabel('Frequency')

    plt.show()

* **Exercise**: Write a Python script to create a list of 10 numbers and compute their mean.

**Code**: import pandas as pd

list={1,2,3,4,5,6,7,8,9,10}

list=pd.DataFrame(list)

m1=list.mean()

print('Mean : \n ',m1)

* **Exercise**: Create a function that takes a list of numbers and returns a dictionary with the count, mean, median, and standard deviation.

**Code**: in progress

* **Exercise**: Generate a 5x5 matrix of random numbers and print it.

**Code**: import numpy as np

matrix=np.random.rand(5,5)

print(matrix)

* **Exercise**: Load a CSV file into a Pandas DataFrame and print summary statistics for each column.

**Code**: import pandas as pd

data=pd.read\_csv('C:/Users/abdul/OneDrive/Desktop/ByteWise\_ML/Day\_6/USA\_Housing.csv')

print(data.describe())

* **Exercise**: Implement a simple linear regression model using Scikit-Learn and print the model coefficients.

**Code**: import pandas as pd

from sklearn.linear\_model import LinearRegression

data = {

    'x':[1,2,3,4,5],

    'y':[3,5,9,6,0]

}

df=pd.DataFrame(data)

model = LinearRegression()

x=df[['x']]

y=df['y']

model.fit(x, y)

print("Model Coefficients:")

print(model.coef\_)