Day-10 Morning Assessment

Pandas

Data = {“name” : [“Monica”,”Phoebe”, “Joey”,”Chandler”,”Rachel”,”Rebecca”,”Sammy”,”Kevin”,”Ohm”,”Sandy”,”Emily”,”Taylor”,”Shiva”,” Joseph”,”Abhi”],”Technology” : [“Java”,”Python”,”Power BI”,”Ruby”,”Swift”,”C#”,”SQL”,”Mongo DB”,”Kotlin”,”C”,”JavaScript”,”PlSQL”,”R”,”React”,”Shell”],”marks”:[16,45,75,84,36,64,95,70,24,57,65,38,76,96,68]}

Df = pd.DataFrame(Data)

Print(Df)

Teacher = [“Ramesh”,”Suresh”,”Pallavi”,”Geetha”,”Rakesh”,”Hansika”,”Lalitha”,”Pankaj”,”Ravi”,”Manjula”,”lilly”,”William”,”Harish”,”Mamatha”,”Brishti”]

Df[“teacher”] = Teacher

Print(Df)

Df.drop(“Technology”)

Print(Df)

Print(Df.head(10))

Print(Df.tail(5))

Print(Df.describe())

Print(Df.shape())

Sorted\_df = df.sort\_values(by = “marks”, ascending = False)

Print(sorted\_df)

Data Loading and Overview

1. Reading a csv file into a pandas dataframe:

Lets assume the csv file is “data.csv”

Data = pd.read\_csv(‘data.csv’)

Print(Data)

1. The shape of the dataframe is

Let the DataFrame be “Data”

Print(Data.shape)

1. Getting Summary Statistics

Let the dataframe be “Data”

Print(Data.describe())

Data Selection and Filtering

1. Getting the “Age” column

Let the dataframe be “Data”

Print(Data[‘Age’])

1. Filtering rows based on condition

Let the dataframe be “Data”

Print(Data[‘salary’]>50000)

1. Filtering multiple conditions

Let the dataframe be “Data”

Print(Data[(Data[‘Department’] == “HR”) & (Data[‘Age’] > 30)])

Data Cleaning

1. Checking for missing values

Let the dataframe be “Data”

Print(data.isnull().sum())

1. Replacing the missing values

Let the dataframe be “Data”

Data[‘salary’].fillna(0,inplace = True)

Print(Data)

1. Removing Duplicate rows

Let the dataframe be “Data”

Data.drop\_duplicates().reset\_index(drop = True)

Print(Data)

Data Aggregation and Sorting

1. Sorting the dataframe by a column

Let the dataframe be “Data”

Sorted\_values = Data.sort\_values(by = ‘Age’ ascending = False)

1. Grouping and finding the average

Let the dataframe be “Data”

Avg\_salary = Data.groupby(“Department”)[‘salary].mean()

Print(Avg\_salary)

1. Counting Unique values

Let the dataframe be ‘Data’

D1= Data[‘Department’].drop\_duplicates()

Print(D1.count())