day-8-LiC

Pandas

```
import pandas as pd
# standard way to import and use an alais for a library
#reading data
df = pd.read_excel("Employee Sample Data.xlsx")

#Viewing the first few rows
print(df.head())

#Checking dataset info
print(df.info())

#Basic statistics of data
print(df.describe())

#column names
print(df.columns)

#selecting columns
df["ColumnName", "Age"]
#how to include several columns?

#selecting rows by index or by condition
df.iloc[0]  # First row
df.iloc[0:5]  # First 5 rows

df.loc[df["Age"] > 30]  # Get rows where Age > 30

#Handeling missing values
df.isnull().sum()  # Count missing values in each column
df = df.dropna()  # Remove rows with missing values
df.fillna("Unknown", inplace=True)  # Fill missing values


#duplicates

df = df.drop_duplicates()

#renaming and changing data types
```

```
df.rename(columns={"OldName": "NewName"}, inplace=True)
df["Age"] = df["Age"].astype(int)


#sorting data
df.sort_values(by="Age", ascending=False)


#grouping data
df.groupby("Department")["Annual Salary"].mean()  # Average salary per
department


#excercise! '''
Load a dataset (Employee Sample Data.excel).
Display the first 5 rows.
Find and remove missing values.
Find the average salary per department.
'''
```

Numpy

```
import numpy as np
# creating a numpy array from a python list
# creating an array of zeros or ones
# creating an array with a range of zeros
# creating a random array
#basic mathematical operations arr1 = np.array([1, 2, 3])
arr2 = np.array([4, 5, 6])


#applying functions


#Basic statistical Analysis and finding the max &min values data =
np.array([10, 20, 30, 40, 50])

#reshaping arrays

#Indexing and slicing
arr = np.array([10, 20, 30, 40, 50])
```

```
#selecting specific rows and columns in a 2d array matrix = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])


'''
Exercise!
Create a NumPy array with numbers from 1 to 20.
Reshape it into a 4x5 matrix.
Find the mean, median, and standard deviation.
Extract the second column.
Find the index of the maximum value in the matrix.
'''
```

Matplotlib

```
import matplotlib.pyplot as plt

#Basic line plot

x = [1, 2, 3, 4, 5]
y = [10, 20, 25, 30, 40]

plt.plot(x, y)  # Plot the data
plt.xlabel("X-axis")
plt.ylabel("Y-axis")
plt.title("Simple Line Plot")
plt.show()  # Display the plot


plt.plot(x, y, marker='o', linestyle='--', color='r', label="Data Line")
plt.legend()  # Add a legend
plt.grid(True)  # Show grid
plt.show()

#rest of the features.....
```