CYCLE II

1.) Given is a list of of words, wordlist, and a string, name. Write a Python function which takes wordlist and name as input and returns a tuple. The first element of the output tuple is the number of words in the wordlist which have name as a substring in it. The second element of the tuple is a list showing the index at which the name occurs in each of the words of the wordlist and a 0 if it doesn't occur.

2.) Download a file "auto.csv" of automobile data with the fields origin, company name, year, num-of-cylinders, displacement, mileage, horsepower, weight and acceleration etc from Kaggle/UCI/OpenML repository. Write Python codes using Pandas to

1) Clean and Update the CSV file

2)Print total cars of all companies

3) Find the average mileage of all companies

3.) The admission fee at a small fair is Rs 1.50 for children and Rs 4.00 for adults. On a certain day 2200 people enter the fair and Rs 5050 was collected. Calculate the number of children and adults who attended the fair by solving linear equations obtained from the given statement.

4.) Write Python program to write the data given below to a CSV file.

SN Name Country Contribution Year

1 Linus Torvalds Finland Linux Kernel 1991

2 Tim Berners-Lee England World Wide Web 1990

3 Guido van Rossum Netherlands Python 1991

Given the sales information of a company as CSV file with the following fields month_number, facecream, facewash, toothpaste, bathingsoap, shampoo, moisturizer, total_units, total_profit. Write Python codes to visualize the data as follows

a) Toothpaste sales data of each month and show it using a scatter plot

b) Face cream and face wash product sales data and show it using the bar chart

c) Calculate total sale data for last year for each product and show it using a Pie chart.

Create and use your own file with name company.csv

5.) Given a list of sentences, write a Python function that:

- Counts how many times a particular word appears in the entire list.

- Returns a dictionary where keys are sentence indices and values are the positions of a particular word in that sentence (or -1 if not present).

6.) Download a dataset students.csv with the fields: student_id, name, subject, marks, grade etc. from Kaggle/UCI/OpenML repository. Write Python codes using Pandas to:

- Remove duplicate rows and handle missing values.
- Find the student with the maximum marks in each subject.
- Group the students by grade and calculate the average marks for each grade.

7.) Write a Python program to store the following dataset into a CSV file tech_innovators.csv. Write python code to read this file and print the names of innovators who contributed before the year 2000.

SN	Name	Country	Innovation	Year
1	Geoffrey Hinton	Canada	Deep Learning	1986
2	Yann LeCun	France	Convolutional Neural Nets	1989
3	Andrew Ng	USA	Online ML Courses	2011
4	Phila Suoni	UK	Artificial Intelligence	2023
5	Yemehi Yathi	France	Data Analytics	2025

8.) Download a CSV file sales_data.csv with the following fields: month, electronics, clothing, groceries, furniture, profit etc. from Kaggle/UCI/OpenML repository. Write Python code to visualize:

a) Monthly profit using a line chart.
b) Electronics and Clothing sales in a grouped bar chart.
c) Total yearly sales for each category as a pie chart.

9.) Download Iris Dataset (with features: sepal_length, sepal_width, petal_length, petal_width, species). Perform the following:

- Load the dataset into Pandas and clean it if necessary.
- Split the dataset into training and testing sets (80–20 split).
- Display the number of samples for each species.
- Plot histograms of sepal length and petal length.
- Show the distribution of species using a bar chart.