Day 1: 23.03.2024

A weather station wants to know the most common types of weather in their area. They have a list of all the weather conditions that have occurred in the past year, along with the number of times each weather condition has occurred. Write a program that will calculate the frequency distribution of weather conditions and print out the most common weather type.

2. Write a Pandas program to replace the 'qualify' column contains the values 'yes' and 'no' with True and False.

Sample Python dictionary data and list labels:

exam_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'],
'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19],
'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],
'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}
labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']

3. You have trained a dataset to evaluate its performance with various metrics. Write a Python program that loads a dataset and trained model from scikit-learn. The program should ask the user to input the names of the features and the target variable they want to use for evaluation. The program should then calculate and display common evaluation metrics such as accuracy, precision, recall, and F1-score for the model's predictions on the test data.

4. A hospital wants to know the most common diseases among their patients. They have a list of all the diseases that their patients have been diagnosed with in the past year, along with the number of patients who have been diagnosed with each disease. Write a program that will calculate the frequency distribution of diseases and print out the most common disease using following dataset

DISEASE_NAME	IAGNOSED_PATIENTS
Common Cold	320
Diabetes	120
Bronchitis	100
Influenza	150
Kidney Stones	60

5. You are tasked with preparing a clean and consolidated dataset for a project at a tech startup. The company has collected data from multiple sources, including customer demographics, user activity logs, and customer support interactions. Your goal is to create a unified dataset that will identify all factors influencing customer satisfaction.

The datasets are

1. **Customer Demographics (`customer_demographics.csv`):**
   - Columns: `customer_id`, `age`, `gender`, `location`.
2. **User Activity Logs (`user_activity_logs.csv`):**
   - Columns: `customer_id`, `timestamp`, `page_views`, `interaction_duration`.
3. **Customer Support Interactions (`customer_support.csv`):**
   - Columns: `customer_id`, `support_tickets`, `satisfaction_score`.