PROJECT-1 (TWITTER SENTIMENT ANALYSIS)

The following project is about analyzing the sentiments of tweets on social networking website 'Twitter'. The dataset for this project is scraped from Twitter. It contains 1,600,000 tweets extracted using Twitter API. It is a labeled dataset with tweets annotated with the sentiment (0 = negative, *2* = neutral, 4 = positive).

It contains the following 6 fields:

1. target: the polarity of the tweet (*0* = negative, *2* = neutral, *4* = positive)
2. ids: The id of the tweet .
3. date: The date of the tweet (*Sat May 16 23:58:44 UTC 2009*)
4. flag: The query. If there is no query, then this value is NO_QUERY.
5. user: The user that tweeted
6. text: The text of the tweet.

Design a classification model that correctly predicts the polarity of the tweets provided in the dataset.

PROJECT-2 (RATINGS OF GUVI COURSES)

The following project is about Guvi Courses. The dataset for this project contains information about Guvi courses in various categories, including course title, URL, price, number of subscribers, number of reviews, number of lectures, course level, rating, content duration, published timestamp, and subject. With this dataset, we can track the performance of courses and uncover opportunities to generate revenue.

The data is organized into columns with the following information:

Columns	Description
course_title	The title of the Guvi course. (String)
url	The URL of the Guvi course. (String)
price	The price of the Guvi course. (Float)
num_subscribers	The number of subscribers for the Guvi course. (Integer)

num_reviews	The number of reviews for the Guvi course. (Integer)
num_lectures	The number of lectures in the Guvi course. (Integer)
level	The level of the Guvi course. (String)
Rating	The rating of the Guvi course. (Float)
content_duration	The content duration of the Guvi course. (Float)
published_timestamp	The timestamp of when the Guvi course was published. (Datetime)
subject	The subject of the Guvi course. (String)

Design a regression model to predict the ratings given by the learners to the course.

PROJECT-3 (INSTAGRAM INFLUENCERS)

Instagram is an American photo and video sharing social networking service founded in 2010 by Kevin Systrom and Mike Krieger, and later acquired by American company Facebook Inc., now known as Meta Platforms. The app allows users to share posts that can be shared publicly or with pre-approved followers.

Instagram is very much used to influence people in a particular way for a specific issue - which can impact the order in some ways. The following dataset is about such influencers. The fields in the given dataset are as follows:

Columns	Description
rank	Rank of the Influencer
channel_info	Username of the Instagrammer
influence_score	Influence score of the users
posts	Number of posts they have made so far

followers	Number of followers of the user
avg_likes	Average likes on instagrammer posts
60*day*eng_rate	Last 60 days engagement rate of instagrammer as faction of engagements they have done so far
new*post*avg_like	Average likes they have on new posts
total_likes	Total likes the user has got on their posts. (in Billion)
country	Country or region of origin of the user

Answer the following questions based on the given data set:

1. Are there any correlated features in the given dataset? If yes, state the correlation coefficient of the pair of features which are highly correlated.
2. What is the frequency distribution of the following features?
    - Influence Score
    - Followers
    - Posts

3. Which country houses the highest number of Instagram Influencers? Please show the count of Instagram influencers in different countries using barchart.
4. Who are the top 10 influencers in the given dataset based on the following features
    - Followers
    - Average likes
    - Total Likes

5. Describe the relationship between the following pairs of features using a suitable graph

    - Followers and Total Likes
    - Followers and Influence Score
    - Posts and Average likes
    - Posts and Influence Score

PROJECT-4 (IMAGE CLASSIFICATION)

Image classification refers to grouping the images based on similar features. It is a supervised learning approach in which you are given a labeled dataset. In this particular problem, you need to perform a binary classification of the given images. The binary classes identified are cats and dogs. After extracting the suitable features from the images, design a binary classifier that performs classification on the given data set. Also, evaluate the given classifier using all the evaluation metrics.

You can download the data set from the given data source.

https://www.kaggle.com/datasets/chetankv/dogs-cats-images?resource=download