**Clean and analyse social media data(twitter) using python**

**Introduction**

Social media has become a ubiquitous part of modern life, with platforms such as Instagram, Twitter, and Facebook serving as essential communication channels. Social media data sets are vast and complex, making analysis a challenging task for businesses and researchers alike. In this project, we explore a simulated social media, for example Tweets, data set to understand trends in likes across different categories.

**Project Scope**

The objective of this project is to analyze tweets (or other social media data) and gain insights into user engagement. We will explore the data set using visualization techniques to understand the distribution of likes across different categories. Finally, we will analyze the data to draw conclusions about the most popular categories and the overall engagement on the platform.

**Importing required libraries**

importing the following required libraries:

• pandas for creating the dataframe

• numpy for forming a random number from a range

• Matplotlib.pyplot for displaying graphs

• seaborn for plotting the data

• random for making a choice from a list of items

**Generating random data for the social media data**

Defining a ***list*** of categories for the social media experiment. Namely: Food, Travel, 'Fashion, Fitness, Music, Culture, Family, Movies and Health.

Generating a Python data dictionary with fields Date, Category, and number of likes, all with random data.

**Loading the data into a Pandas DataFrame and Explore the data**

Loading the randomly generated data into the pandas dataframe and print the data.

**Cleaning the data**

Removing all the null data using the dataframe drop method.

Next, Removing all the duplicate values using remove method.

Converting the datatypes of date from string to datetime format.

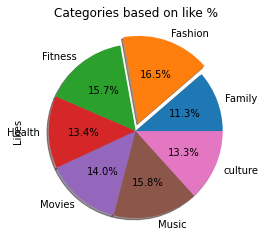
**Visualizing and Analyse the data**

Analysing the data using statistical functions such as min, max, sum, and mean

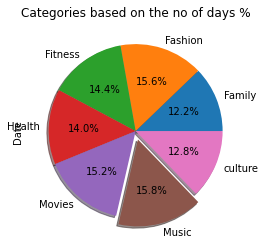
Visualizing the data using pie chart, bar graph and boxplot.

**Conclusion**

1. **The Fitness category has the maximum number of likes**
2. **The Family category has the minimum number of likes**
3. **culture and Heath has almost the same percentage of likes**



1. **.One the basis of number of days the music has the higher numbers**



1. **The average likes are higher in Fitness category and the lowest in the movies category**

