**Project Report: Analysis of Twitter Sentiment**

1. **Introduction**

**Twitter comprises of over 1,600,000 tweets retrieved through the Twitter API. This dataset contains valuable real time data, including tweet IDs, timestamps, user details, and the text of the tweets. This assignment unlocks insights from this dataset by exploring the sentiment trends in contains.**

**1.1 Objectives**

* **To load, process and store the ProjectTweets.csv dataset into a suitable database for analysis.**
* **To employ Apache Spark, a distributed data processing environment, to process the dataset efficiently.**
* **To conduct sentiment analysis using Text\blob library and to explore sentiment trends over time.**
* **To present the findings through an interactive and dynamic dashboard.**
  1. **Description of Dataset**

**The dataset used contains a vast collection of Twitter posts, know as “tweets”. Each tweet provides insights like the emotions, opinions, and sentiments of the users. The dataset comprises of:**

* **ids: The unique identifiers for each tweet**
* **date: Timestamps reflecting the date and time when each tweet was posted.**
* **Flag: Describes whether the tweet was associated with a specific query or labelled as “NO\_QUERY”.**
* **User: The usernames of the users who authored each tweet, offering insights into the source of content.**
* **Text: The text of the tweets.**

1. **Data Preparation Storage**

**2.1 The dataset used in this assignment is titled “ProjectTweets.csv”. The dataset was collected from Twitter and contains a substantial number of tweets related to a specific topic or keyword, making it suitable for analysis.**

**2.2 Data Preprocessing**

**Before conducting ny analysis, it was essential to preprocess the provided Twitter dataset. Data preprocessing included:**

**Date \Column, Text Data talk about these**

* 1. **Data Storage**

**Talk about SQLite**

* 1. **Distributed Data Processing**

**Talk about Apache spark- faster computation**