**Social Media Sentiment Analysis Report**

**Name: Karmveer Singh Gurjar**

**Domain: Data Science/ Data Analyst**

**Introduction**

Social media is a strong platform where people share their opinions, feelings, and experiences. Analyzing this data offers great value for businesses, organizations, and researchers. It helps them understand public attitudes toward products, services, policies, and current trends. This project aims to perform sentiment analysis on social media data using Python, Natural Language Processing (NLP), NLTK, and WordCloud. After cleaning and preparing the data, we integrate the results with Power BI. There, we create interactive dashboards, key performance indicators, and visual reports that provide real-time insights

The sentiment categories are classified into three types:

* **Positive**  – User expresses satisfaction, happiness, or agreement.
* **Negative**  – User expresses dissatisfaction, criticism, or frustration.
* **Neutral**  – User shares facts, questions, or non-emotional content.

**Objectives**

The main objectives of this project are:

* To preprocess and clean raw Twitter text by removing noise, such as stopwords, special characters, URLs, and hashtags.
* To apply NLP techniques for sentiment classification: Positive, Negative, and Neutral.
* To generate WordClouds for a clearer view of frequently used words.
* To calculate KPIs like Positive %, Negative %, Average Likes, and Average Retweets.
* To design an interactive Power BI dashboard with filters, slicers, and visualizations for sentiment insights.

**Tools and Technologies**

* **Python** -for data preprocessing and sentiment analysis.
* **NLTK (Natural Language Toolkit)-** for tokenization, stopword removal, and text cleaning.
* **WordCloud-** for visualizing the most common words in positive, negative, and neutral tweets.
* **Pandas and NumPy**- for data manipulation and structuring.
* **Matplotlib and Seaborn**- for static visualizations in Python.
* **Power BI -**for creating dashboards, tracking KPIs, and telling data stories.

**Dataset Details**

The dataset contains **Social Media posts**, with the following key columns:

* **UsrId** – Unique ID of the user.
* **User** – Twitter username.
* **Text** – Original tweet text.
* **Clean\_Text** – Processed text after cleaning and tokenization.
* **Sentiment** – Classified sentiment (Positive, Negative, Neutral).
* **Likes & Retweets** – Engagement metrics.
* **Country, Platform, Timestamp, Year, Month, Day, Hour** – Metadata for analysis.

**Methodology**

**Step 1: Data Cleaning**

* Remove URLs, mentions, hashtags, emojis, and punctuation.
* Convert text to lowercase. Remove stopwords using the NLTK stopword corpus.
* Tokenize the cleaned text into word lists.

**Step 2: Sentiment Classification**

* Label tweets as Positive, Negative, or Neutral based on their text content.
* Analyze sentiment distribution to identify user opinions and patterns.

**Step 3: Visualization with WordCloud**

* Create separate WordClouds for Positive, Negative, and Neutral tweets.
* Highlight frequent keywords that contribute to sentiment classification.

**Step 4: Power BI Dashboard**

* Import the cleaned dataset into Power BI for visualization.The dashboard includes:

**KPIs:**

* Positive % = (Positive Tweets / Total Tweets) × 100
* Negative % = (Negative Tweets / Total Tweets) × 100
* Average Likes = AVERAGE(Likes)
* Average Retweets = AVERAGE(Retweets)

**Charts:**

* Sentiment distribution (Pie/Bar Chart).
* Country-wise sentiment (Map visualization).
* Time-based sentiment patterns (Year/Month/Day slicers).

**Results and Visualizations**

**Positive Sentiment**

The WordCloud below shows the most common words linked to Positive tweets. Frequent terms like joy, laughter, love, new, and success highlight optimism, happiness, and encouragement.



**Neutral Sentiment**

The Neutral sentiment WordCloud includes words like life, beauty, soul, whisper, and reflect. These words are often descriptive, factual, or balanced, without any strong emotions.



**Negative Sentiment**

For negative tweets, common words are disappointment, grief, bitter, ruin, despair, and pain. These words express feelings of sadness, criticism, or frustration.



**Applications**

* **Business Intelligence**- Understanding customer feedback on products or services.
* **Social Media Monitoring**- Tracking brand reputation and public sentiment.
* **Political Analysis**- Measuring public opinion during elections.
* **Market Research-** Identifying trends, likes, and dislikes among users.

**Conclusion**

This project shows how NLP and data visualization work together to get valuable insights from social media data. By sorting tweets into Positive, Negative, and Neutral sentiments, businesses and researchers can gain a clearer view of user opinions and make informed decisions. Combining Python-based text analysis with Power BI dashboards provides solid technical support and makes data storytelling accessible.