Tweet user Engagement Analysis

Project Overview

This project involves analyzing a JSON file containing tweets from a user's Twitter account. The goal is to extract key insights from the tweet data, including activity patterns, sentiment analysis, and engagement metrics. The analysis aims to optimize tweet strategies for better engagement and provide actionable recommendations based on the findings.

Data

The dataset is a JSON file containing tweets from a user's Twitter account. Each tweet includes various attributes such as:

- created_at: Timestamp of the tweet.
- full_text: Full text content of the tweet.
- favorite_count: Number of likes.
- retweet_count: Number of retweets.
- entities: Contains hashtags, user mentions, and URLs.

Data Preparation

1. **Loading Data**: The JSON file is loaded into a Python script.
2. **Extracting Relevant Information**: For each tweet, the following details are extracted:
    - Timestamp (created_at)
    - Full text (full_text)
    - Favorite count (favorite_count)
    - Retweet count (retweet_count)
    - Hashtags (entities["hashtags"])
    - Sentiment scores (polarity and subjectivity using TextBlob)
3. **Formatting Data**: The extracted data is formatted into a Pandas DataFrame, including:
    - datetime: Parsed timestamp.
    - day_of_week: Day of the week derived from the timestamp.
    - hour_of_day: Hour of the day derived from the timestamp.
    - hashtags: Comma-separated list of hashtags.
4. **Aggregating Data**: Aggregated counts of tweets, retweets, and favorites by day of the week and hour of the day.

5. **Saving Data**: The prepared DataFrame is saved as a CSV file for visualization in Tableau.

Visualizations

1. **Tweet Count by Day of Week and Hour of Day**: Bubble chart showing tweet activity, highlighting peak periods.
2. **Retweet Count by Day of Week and Hour of Day**: Bubble chart illustrating retweet activity, identifying high engagement times.
3. **Favorite Count by Day of Week and Hour of Day**: Bubble chart displaying favorite activity, showing when tweets are most liked.
4. **Most Retweeted Hashtags**: Bar chart listing the top 10 hashtags by retweet count.
5. **Most Favorited Hashtags**: Bar chart listing the top 10 hashtags by favorite count.
6. **Hashtag Connections through Cluster Analysis**: Bubble chart exploring the connections and engagement levels of various hashtags.
7. **Sentiment Analysis Over Time**: Line chart showing the trend of average sentiment polarity over time.
8. **Distribution of Sentiment Polarity**: Histogram displaying the distribution of sentiment scores.
9. **Average Sentiment by Day of the Week**: Bar chart comparing sentiment across different days.
10. **Sentiment Heatmap**: Heatmap visualizing average sentiment by day of the week and hour of the day.
11. **Polarity and Subjectivity by Hashtag**: Scatter plot showing the relationship between average polarity and subjectivity for different hashtags.