

PROJECT TEAM 13

Team Members:

1. Niraj Sai Prasad - NUID 001006514
2. Sindhu Swaroop - NUID 001006558
3. Vatsal Doshi - NUID 002776613
4. Avani Kala - NUID 002772623

Topic: Twitter: Summing up the Pulse of the Internet

Data Model: Document (NewSQL)

Target Platform: Azure SQL Multi-Model

1. Sentiment analysis:

Sentiment analysis involves using natural language processing techniques to identify and extract subjective information from textual data, such as tweets, reviews, or news articles. The goal of sentiment analysis is to determine the polarity of the text (positive, negative, or neutral), and it can be used for various applications such as social media monitoring, brand reputation management, or customer feedback analysis.

To perform sentiment analysis on tweets, you would need to first collect a dataset of tweets, preprocess the text by removing stop words, punctuations, and URLs, and then apply a machine learning or deep learning model to classify the tweets into positive, negative, or neutral categories.

2. Trending tweets during a time period:

To analyze the trending tweets during a specific time period, you would need to collect a large dataset of tweets from that time period, and then perform text analysis to identify the most frequent keywords, hashtags, or mentions in the tweets. You can use tools like Twitter's API or third-party platforms like Tweepy or Twython to collect the tweets.

Once you have the dataset, you can use data visualization tools like Tableau or Power BI to create charts and graphs that display the trending topics over time.

3. Correlation matrix for users with the most tweets:

A correlation matrix is a table that displays the correlation coefficients between multiple variables. In the context of Twitter data, you can create a correlation matrix that shows the correlation between the users who have tweeted the most and the topics they have tweeted about.

To create the correlation matrix, you would need to collect a large dataset of tweets from the users who have tweeted the most, and then perform text analysis to extract the topics from the tweets.

4. Location-based analysis:

The location-based analysis involves analyzing the tweets from a specific geographic location or region. This can be useful for understanding the sentiment, topics, or events that are relevant to a particular location.

To perform location-based analysis, you must collect a dataset of tweets that contain geolocation information. You can use the Twitter API or third-party tools like Tweepy or Twython to collect the tweets.

Once you have the geolocated tweets, you can perform text analysis and sentiment analysis to identify the topics and sentiments associated with the tweets. You can also use data visualization tools like Tableau or Power BI to create maps and charts that display the distribution of tweets by location.

5. Analyze the most followed Twitter users:

To analyze the most followed Twitter users, you would need to collect a dataset of the top users by follower count.

Once you have the user data, you can perform text analysis to identify the topics and sentiments associated with the tweets of the most followed users. You can also use network analysis techniques to identify the followers and influencers of the most followed users.

Overall, these objectives involve collecting and analyzing large datasets of tweets and applying natural language processing, machine learning, and data visualization techniques to gain insights into the topics, sentiments, and relationships behind each tweet.