**CS 210 – Course Project**

**Sude Dok  26847**

**Instructor:**

**Onu Varol**

**Date: 18.01.2024**

**A blue sign with white text

Description automatically generated**

Explore Twitter's virtual environment by taking part in my data science project, "Unveiling Twitter Sentiments on 'Data Science'." In order to uncover the underlying feelings within the massive sea of tweets, I employed data scraping, strong preprocessing techniques, and perceptive visualizations in my investigation.

My main purpose in this project was to examine the tweets about 'Data Science' on Twitter and to conduct an analysis on this.

Data Gathering and Preprocessing: To start, I carefully scraped a treasure mine of tweets from Twitter's ever-changing feed. Strict preprocessing techniques were used to organize and clean the raw data, making it ready for further analysis. To tell you a little bit, I scraped 5000 tweets containing 'Data Science' using the ntscraper library. The data I obtained did not mean much in its raw form. For this reason, I had to make this data ready for use. Since things like emojis etc. are not suitable for processing, I had to first clean the tweets containing these contents. Then I checked whether there were duplicate values in my dataset and deleted these data. I also chose to delete the missing data instead of filling it with average values. Because the data size I currently had was sufficient for my use. After completing all these preprocessing, I saved my clean dataset for later use.

Exploratory Data Analysis: Equipped with a refined dataset, I explored the core of the Twitterverse. Through the discovery of fascinating patterns, trends, and anomalies, my exploratory data analysis provided a thorough knowledge of the dynamics at work. I made discoveries such as the size of the dataset, the contents of my columns, and how many different users there were. I even identified users who tweet frequently in this area. The data I had also included the number of likes and comments. I created popularity categories using this data and included this information in a new column for my later use. When I examined this data, I saw that the general density was medium popularity. Since I also had the sharing dates of the tweets, I wanted to use this information as well. I had the opportunity to see the months with the most tweets in this area. According to the data I have, although January 2024 was the month with the most tweets, when I included the popularity, I saw that the most popular month was April 2020.

Another thing I was curious about was the content of these tweets. Therefore, I removed English stop words from the tweet content and examined the most and least used words. I saw that some of the most used words are data, ai, python, analytics. This wasn't too surprising to me. However, it was very nice in terms of the accuracy of the dataset I obtained and in terms of seeing the concepts related to Data Science. And I observed that these frequently used terms do not change according to popularity categories. Although tweet popularity changes, the common words used in tweets related to 'Data Science' generally remain the same.

Visuals as the Compass: I used the power of visuals as my compass to make our results come to life. The subtleties of the data were brought to life through graphs, charts, and plots, which created a visual story that appealed to both data enthusiasts and casual viewers. As you can see in the content of my project, I included many visualizations to reveal various relationships and see the distribution of our dataset. While it revealed things like the correct relationship between likes and comments, it also gave me the opportunity to better examine the distributions in my data set by visualizing them.

Sentiment Analysis: I identified the sentimental undertones in the tweets by utilizing cutting-edge natural language processing techniques. I had subjectivity scales like Objective-Subjective and polarity scales like positive-negative-neutral. By visualizing these scalings, I gained information about their distribution. I saw that most of the tweets were positive and also most were objective. However, when we consider these two features together, the highest number of tweets were sent as Neutral-Objective.

In the end, “Unveiling Twitter Sentiments on 'Data Science'” wasn’t just a project for me. It was a data science experience that I can call basic and intermediate level, where I obtained the data myself, organized it in a way that was useful to me, learned what was thought about data science on one of the most important social media platforms.