india-shelter-case-study-project

September 13, 2023

1 INDIA SHELTER CASE STUDY

- To present the Summary of the approach that how we will be dealing with a large dataset of "raw_tweets_queen_death1"
- The datasets containing raw tweets,surrounding the immediate time frame of the death of Queen Elizabeth II, **We have to read the dataset in python and answer the below the following mentioned questions:-**

1. Find the user of the most retweets.
2. Find the most effective tweet (create a measure of your own based on parameters such as retweets, time from death, etc.).
3. Show: Language distribution place distribution
4. Visualise and explain a relationship between likes, retweets and replies.
5. Does a video in the tweet make it more likeable? Support your answer with factual data from the given dataset

2 TYPICALLY A CASE STUDY INVOLVES SEVERAL STEPS INCLUDING DATA EXPLORATION,CLEANING,ANALYSIS AND INTERPRETATION

1. **Step 1: Understanding the Data:** Describing the "raw_tweets_queens_death1" data set. What does it contain? What are the columns/variables?
2. **Step 2: Data Cleaning and Preprocessing:** checking for missing values and handling them appropriately (remove, or ignore). Cleaning and formatting text data if necessary (e.g., removing special characters, lowercasing). Handling duplicate records if they exist. Converting data types as needed (e.g., dates to datetime objects).
3. **Step 3: Exploratory Data Analysis (EDA)** Performing basic statistical analysis (e.g., summary statistics {MEAN,MEDIAN,STANDARD DEVIATION}, distributions). Creating visualizations (e.g., histograms, word clouds, time series plots) to explore data patterns. Identifying any outliers or anomalies.
4. **Data Analysis and Modeling:** Applying appropriate statistical models and techniques to address and solve the problem.evaluating models if it's a predictive modeling task. 5.**Interpretation and Visualization:** Visualized the results using charts, graphs, and tables to make them more accessible and understandable. 6.**Drawing Conclusions:** Summarizing the findings and conclusions from the analysis. 7.**Review and Validation:**Validate your analysis by reviewing it with peers or domain experts to ensure accuracy and reliability.

3 LET US OUTLINE THE APPROACH TO DEAL WITH THE CASE STUDY BASED ON THE PROVIDED QUESTIONS REGARDING THE "raw_tweets_queens_death1" DATASET.

- **Find the user of the most retweets:** Imported the dataset into Python using appropriate libraries (e.g., pandas.matplotib,numpy etc..). Identified the tweet with the highest number of retweets. Retrieve the user who authored that tweet.
- **Find the most effective tweet:** Created a measure of tweet effectiveness based on parameters such as retweets, time from death, and possibly other factors. Analyzed the dataset to find the tweet that scores the highest on this measure.
- **highlighted the Language distribution and place distribution:** Analyzed the dataset to determine the distribution of languages used in the tweets. Analyzed the dataset to determine the distribution of places mentioned in the tweets.
- **Visualize and explain the relationship between likes, retweets, and replies:** Created scatterplots and other relevant visualizations to explore the relationships between likes, retweets, and replies and Explained the findings and any insights into user engagement.
- **Does a video in the tweet make it more likeable?** Identifed tweets in the dataset that contain videos. Compared the average number of likes for tweets with videos to tweets without videos. Supported my answer with factual data from the dataset.

4 SUMMARY OF ANALYSIS:

1. The analysis will provide insights into user engagement on Twitter during the time frame of Queen Elizabeth II's death.
2. It will identify the most influential tweet, user with the most retweets, and analyze language and place distribution.
3. The analysis will also explore the relationship between likes, retweets, and replies, shedding light on user engagement dynamics.
4. Lastly, it will investigate whether tweets with videos are more likeable. So here I provided a clear visualizations, explanations, and statistical analysis to support the findings in the case study report.