

# Customer Support on Twitter Analysis Proposal

Taha Ababou

## Customer Support on Twitter Analysis Proposal

### Personal Statement

My career goal is to become a software engineer, approaching problem-solving through a statistician's perspective. I have a background in computer engineering from my bachelor's studies, where I developed strong proficiency in tackling problems from an engineering and coding standpoint. However, I am now focused on enhancing my statistical approach to complement my engineering skills, allowing me to design solutions that are both technically sound and analytically driven. This project aligns with my goal by providing hands-on experience with statistical techniques in natural language processing and predictive modeling, essential for a data-centric approach to software engineering.

### Preliminary Question

**Can we predict the volume and timing of responses to customer support inquiries based on tweet characteristics, and how do sentiment and engagement timing affect customer support effectiveness?**

This question will guide my exploration of customer service engagement patterns and response quality on Twitter.

### Data Source

- **Dataset:** [Customer Support on Twitter Dataset](#), sourced from Kaggle.
- **Description:** The dataset includes ~2.8 million tweets from 2017, with columns such as `tweet_id`, `author_id`, `inbound`, `created_at`, `text`, `response_tweet_id`, and `in_response_to_tweet_id`, capturing conversations between customers and support agents.

## Proposed Timeline of Work

### 1. Exploratory Data Analysis (EDA)

- **Timeline:** November 10 - November 14
- **Tasks:** Review data structure, visualize distribution of response volumes and times, explore trends in engagement by time of day and sentiment, and identify patterns across brands.

### 2. Data Processing

- **Timeline:** November 15 - November 17
- **Tasks:** Clean and preprocess the dataset, including datetime conversion for `created_at`, handling missing values, running sentiment analysis on `text`, and conducting NLP preprocessing (tokenization, removing stop words) for topic modeling.

### 3. Modeling and Validation

- **Timeline:** November 18 - November 25
- **Tasks:**
  - **Predicting Reply Volume:** Use regression models (e.g., Poisson or Negative Binomial) to predict the number of follow-up responses based on features like `inbound`, `text`, `created_at`, and `author_id`.
  - **Response Time Analysis:** Build regression models to estimate the response time for inbound tweets, considering factors such as time of day, sentiment, and previous response patterns.
  - **Sentiment and Engagement Analysis:** Examine how sentiment affects response timing and frequency by categorizing tweets into positive, negative, and neutral, then observing engagement differences.
  - **Topic Modeling for Common Issues:** Apply clustering techniques (e.g., LDA) on `text` to identify frequent customer issues.
  - **Escalation Probability Prediction:** Develop a classification model to predict which tweets are likely to escalate into longer conversations, using initial sentiment and keyword features.
- **Validation:** Use Mean Absolute Error (MAE) and Mean Squared Error (MSE) for regression models and F1-score for classification tasks.

### 4. Write-up and Documentation

- **Timeline:** November 26 - November 30

- **Tasks:** Summarize findings, document modeling processes, evaluate key insights, and offer implications for customer support strategies, such as optimal response times and suggested response strategies based on sentiment.

## **Additional Experiments**

- **Engagement Analysis by Time of Day:** Analyze engagement patterns to determine optimal support staffing and scheduling, grouping tweets by hour and day.
- **Response Quality Assessment:** Measure response effectiveness by tracking if a support response ends the conversation or requires follow-up.