

# Data Analyst Internship Project Documentation

## Project 3: Twitter Sentiment Analysis

### Overview:

The Twitter Sentiment Analysis project focuses on analyzing a dataset of tweets to classify their sentiment as positive, negative, or neutral. The dataset is sourced from Kaggle, and this analysis aims to provide insights into public sentiment using data analytics and machine learning.

### Project Objectives:

1. Data Exploration: Explore the structure and features of the dataset.
2. Data Cleaning: Handle missing values, duplicate entries, and irrelevant content.
3. Exploratory Data Analysis (EDA): Visualize sentiment patterns and distributions.
4. Sentiment Distribution: Understand the distribution of positive, negative, and neutral tweets.
5. Word Frequency Analysis: Identify common words associated with each sentiment.
6. Temporal Analysis: Analyze sentiment trends over time.
7. Text Preprocessing: Prepare the text data for machine learning.
8. Sentiment Prediction Model: Implement a model to predict tweet sentiment.
9. Feature Importance: Identify key words that contribute to sentiment prediction.
10. User Interface (Optional): Build a simple interface for sentiment prediction.

### Key Findings:

- Sentiment Trends: The dataset showed an imbalance in sentiment, with negative tweets being more frequent.
- Important Terms: Common terms were identified for both positive and negative sentiments.
- Temporal Insights: Specific time frames showed spikes in sentiment, often aligned with significant

events.

## Data Visualization with Power BI

Power BI was used for interactive visualizations. Key visuals included:

- Sentiment Distribution: Displaying the proportion of positive, negative, and neutral tweets using bar charts or pie charts.
- Word Frequency: Word clouds visualized the most common words in each sentiment category.
- Temporal Trends: Time series charts highlighted changes in sentiment over specific time periods.

## Machine Learning Implementation:

Data Preprocessing: Tweets were tokenized, cleaned, and lemmatized to prepare for machine learning.

Sentiment Prediction Model: A logistic regression model was trained using TF-IDF features and evaluated using accuracy and F1 score.

## Conclusion:

The Twitter Sentiment Analysis project provided key insights into public sentiment on Twitter using machine learning and data visualization techniques. Power BI helped create interactive and visually engaging representations of the analysis.