Reporting: wragle_report

- Create a **300-600 word written report** called "wrangle_report.pdf" or "wrangle_report.html" that briefly describes your wrangling efforts. This is to be framed as an internal document.

This report details the wrangling efforts undertaken to clean and prepare a Twitter archive dataset for further analysis. The wrangling process aimed to improve data quality, consistency, and usability for subsequent exploration of user behavior and content.

Data Assessment The initial assessment revealed the following observations about the data:

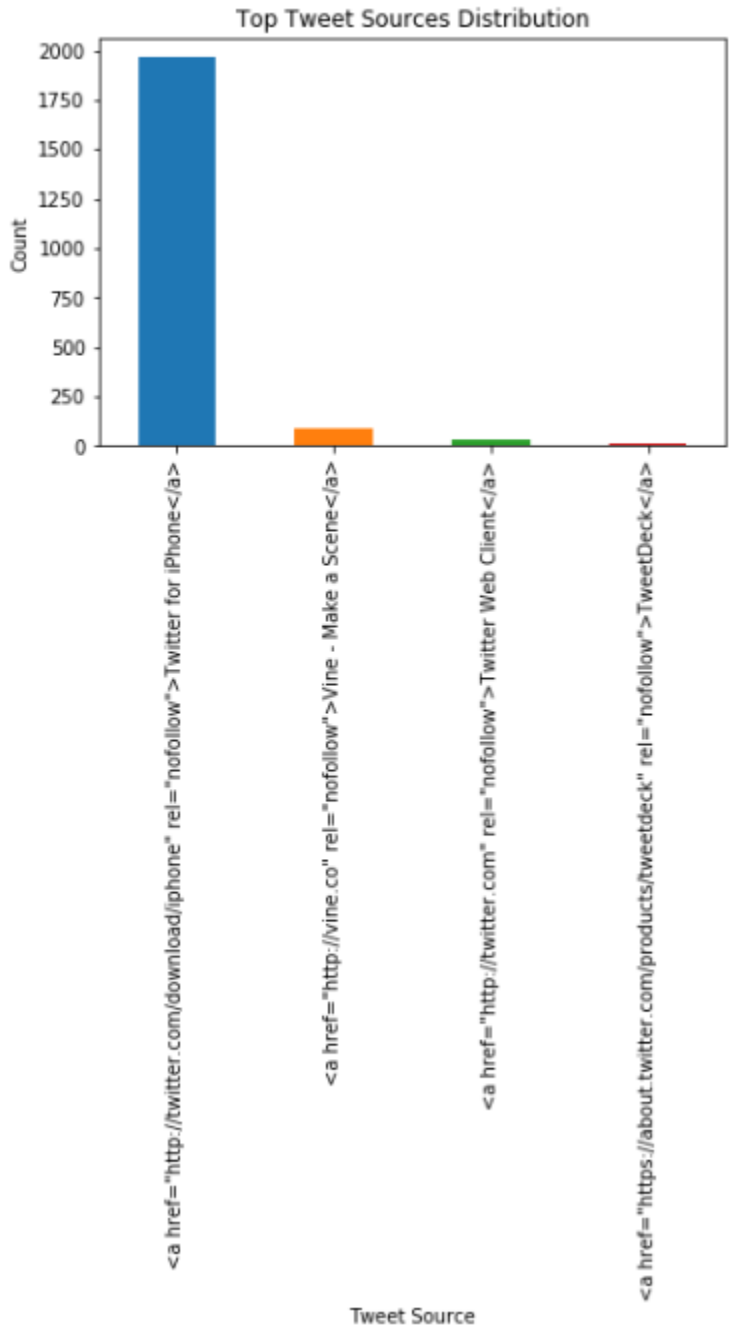
Presence of missing values in some columns (e.g., expanded_urls) Inconsistent formatting in the timestamp column (likely object type representing date and time) Categorical data represented by multiple columns (doggo, floofer, pupper, and puppo) with unclear meaning and potential redundancy

Data Cleaning and Storage The wrangling process resulted in a cleaned and improved version of the Twitter archive data stored in a CSV file named "twitter_archive_master.csv." This file excludes the row index and ensures data consistency for further analysis.

Recommendations for Further Analysis While the wrangling process addressed initial data quality issues, further exploration might be beneficial:

Advanced Sentiment Analysis: Employing more sophisticated sentiment analysis tools can provide deeper insights into the emotional tone of tweets beyond the basic approach used here. User Information Analysis (if available): If user profile data is included, analyzing demographics like location or interests alongside tweet content can reveal additional patterns. Text Preprocessing: Techniques like stemming, lemmatization, and stop-word removal can improve the quality of text data for advanced analysis tasks. By leveraging the cleaned data and exploring these avenues, we can gain valuable insights into user behavior, content trends, and sentiment within the Twitter archive.

I used gemini to generate for me the report


Top Tweet Sources Distribution
Count
2000
1750
1500
1250
1000
750
500
250
0
Twitter for iPhone
Vine - Make a Scene
Twitter Web Client
TweetDeck
Tweet Source