### **Project: Impact Analysis of Twitter's Takeover on Mastodon's Popularity**

#### Goal:

* Investigate how the change in Twitter's ownership influenced user behavior and network growth on Mastodon

#### Sub-Goals

* User Growth Analysis
  + Compare rate of new user registrations on Mastodon 10 months before and after takeover
  + Timeframe we’re looking into:  
    01.01.2022 Beginning of Analysis  
    27.10.2022 Twitter Takeover  
    31.08.2023 End of Analysis
  + Analyze demographic changes in the user base, if possible.
* Engagement Trends, regarding xyz topic:
  + Study the change in posting frequency and user engagement.
  + Identify shifts in peak activity times or days.
* Content Shifts:
  + Analyze the change in popular topics and hashtags.
  + Perform sentiment analysis to understand user reactions to the Twitter takeover.
* Network Dynamics:
  + Examine how the user interaction network evolved.
  + Identify any new communities or clusters that formed post-takeover.

**To do**

* Finish setting up github
* ✅ Decide on which hashtags to filter (both capitalised and lowercase)
  + ✅ ‘twitter’
  + ✅ ‘Gaza’
  + ✅ ‘israel’
  + ✅ ‘thanksgiving’
  + ✅ ‘Milei’
  + ✅ ‘spotify’
  + ✅ ‘argentina’
  + ✅ ‘MissUniverseThailand2023’
  + ‘I\’m\’m’
  + ✅ ‘Trump’
* ✅ Decide on timeframe to filter → The last 30 days
* Data exploration: what to to focus on for data exploration, analysis and visualisation  
  → see suggested next page   
    
  Who does what?

| What | Who |
| --- | --- |
|  |  |

**Time frame for Thursday and Friday**

| When | What |
| --- | --- |
| Thursday  09:00 | Morning stand up/**login**   * Set goals for data exploration * Who does what? |
| Thursday  14:00 | First **meeting** to  **present** **results** and*optionally refine goals* |
| Thursday late evening  /  Friday  09:00 | Second **meeting** to  **show** **final** **results** |
| Friday  11:00 | **Decide** and **finalise** : **what** do we **present** (from our results)  Optional to do some last minute additional work |
| Friday  1400 | Prepare **presentation** |

**Suggested areas to focus on for**

**data exploration, analysis and visualisation of our dataset**

**Comparative Hashtag Frequency**

* Use your Mastodon data to count how often each hashtag appears.
* Create bar charts to compare the frequency of these hashtags between Twitter and Mastodon.

**Temporal Trends Analysis (created\_at):**

* Analyze when posts with these hashtags were most active.
* Create time series plots to show usage trends.

**User Engagement Analysis**

* Analyze 'replies\_count', 'reblogs\_count', and 'favourites\_count' to assess engagement.
* Compare engagement across the 10 hashtags.

**Language Distribution (language)**

* Analyze the 'language' column to see in which languages these hashtags are used.
* Create a pie chart or bar graph to show the language distribution.

**Content Analysis (content, spoiler\_text, media\_attachments)**

* Use a word cloud to visualize the most common words in posts with these hashtags.
* Examine the nature of content, including text, media, and spoilers.

**User Demographics and Behavior:**

* Use account.\* columns (like followers\_count, following\_count, statuses\_count) to understand the profile and behavior of users participating in these hashtags.
* Identify if accounts are bots (account.bot) or locked (account.locked

**Sentiment Analysis**

* Columns to focus on
  + Content (content): The main text of the post where sentiment will be analyzed.
  + Hashtags (tags): For filtering the posts related to specific trending hashtags.
  + Language (language): Important for language-specific sentiment analysis, as sentiment analysis tools are often language-dependent.
* Preprocessing:
  + Clean the 'content' column by removing URLs, special characters, and converting text to lowercase.
* Sentiment Analysis:
  + Use a sentiment analysis library like TextBlob or VADER in Python.
  + Apply the sentiment analysis tool to the 'content' column to get sentiment scores.
  + Classify each post as positive, negative, or neutral based on the sentiment score.
* Aggregating Sentiment by Hashtag:
  + Group the data by hashtags.
  + Calculate the average sentiment score for each hashtag.
* Visualization:
  + Create a bar chart or a pie chart for each hashtag showing the proportion of positive, negative, and neutral sentiments.

**User Network Analysis:**

* Columns to focus on
  + Mentions (mentions): Identifies users mentioned in each post. Essential for mapping interactions between users.
  + Account Information (account.username, account.id): To identify unique users and create nodes in the network.
  + Reblogs (reblog): Indicates re-shared posts. Useful for understanding content dissemination and interaction patterns.
  + Hashtags (tags): To filter or categorize interactions based on specific hashtags.
  + Visibility (visibility): Helps in understanding the scope of the post (public, private, etc.), which can be relevant for network dynamics.
* Data Preparation:
  + Extract mentions and reblogs information from your dataset.
  + Create a list of edges where each edge represents an interaction (mention or reblog) between two users.
* Network Construction:
  + Use a library like networkx in Python to create a graph.
  + Add nodes for each unique user.
  + Add edges based on the interactions extracted in the first step.
* Analyzing the Network:
  + Calculate network metrics like degree centrality to identify influential users.
  + Optionally, filter the network to only include interactions related to the specific hashtags.
* Visualization:
  + Use networkx or matplotlib to visualize the network.
  + Highlight different types of interactions (mentions, reblogs) using different colors or line styles.
  + Use node size to represent user influence (like followers count or centrality measure).
* Hashtag-Specific Networks:
  + Optionally, create separate network graphs for each hashtag to compare the interaction patterns between them.