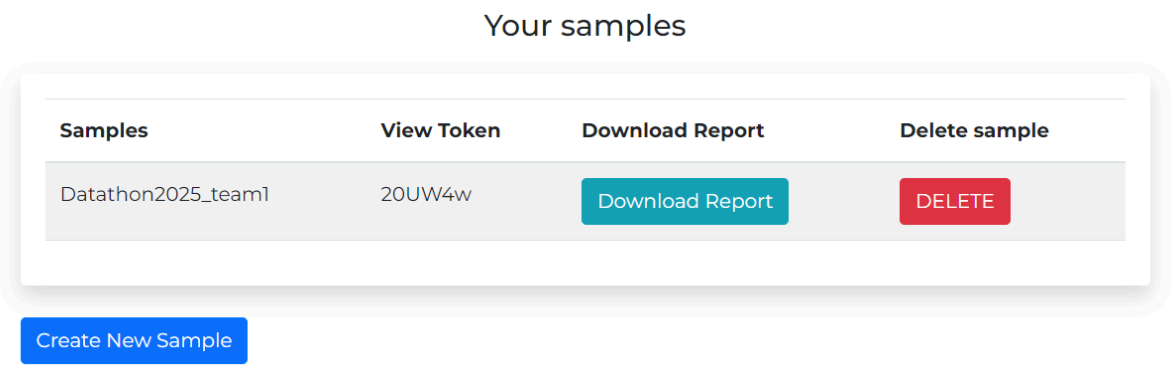
Final Report – Datathon 2025

Team Name: Datathon2025_team1
Author: Karen Solar
Date: August 4, 2025

1. Project Overview

This project was developed for Challenge #1 of the Datathon 2025. The objective was to **identify and classify antisemitic content on X (formerly Twitter)** using the **IHRA Working Definition of Antisemitism (IHRA-WDA)**. We collected posts based on a list of predefined keywords and hashtags known to circulate in online antisemitic and conspiratorial narratives, especially in the context of the **Hamas attack on October 7, 2023** in Israel.

A total of **100 posts** were filtered. The annotation process followed the IHRA-WDA schema to ensure standardization and consistency.



2. Keyword-Based Collection and Filtering

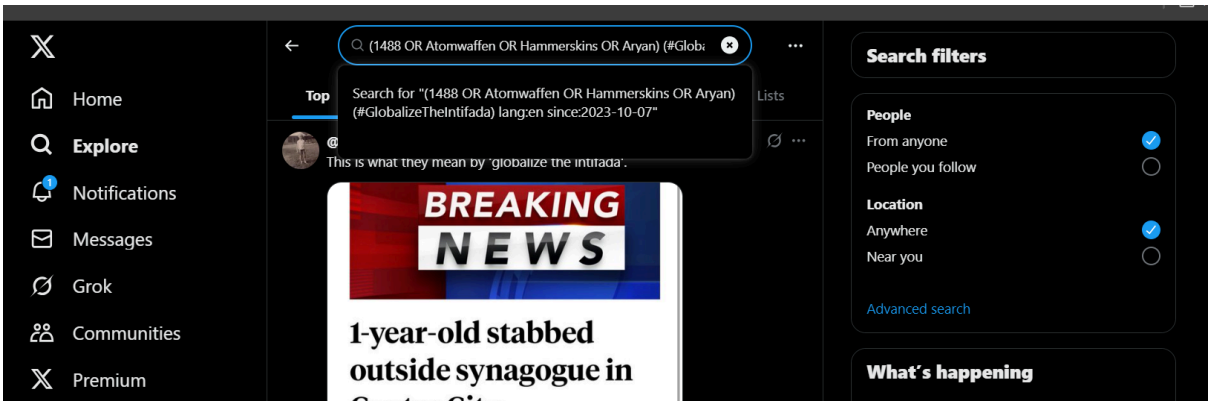
2.1 Rationale for Keyword Selection

We selected keywords associated with antisemitic conspiracy theories, hate speech, and coded language often used in extremist discourse. Terms were chosen to detect both **explicit** and **implicit** antisemitism, including:

- **Conspiratorial tropes**: “ZOG”, “JewishBankers”, “DeepState”, “Rothschilds”
- **Neo-Nazi codes**: “1488”, “Atomwaffen”, “Aryan”
- **Antisemitic hashtags**: `#ZionistOccupation`, `#FromTheRiverToTheSea`, `#GlobalizeTheIntifada`

- **Stereotypes**: “Khazar”, “Jewish elites”, “MediaControl”
- **Religious slurs**: “Talmud”, “ZioNazi”

This strategy allowed for a **targeted scraping approach**, focusing on content with high risk of violating IHRA-WDA categories.



2.2 Data Collection

Data was collected using **Bright Data**, and then exported in `.csv` format. The most relevant content was found in the `description` column of the dataset, not `post`, which caused delays due to misinterpretation in the early stages.

3. Annotation Process

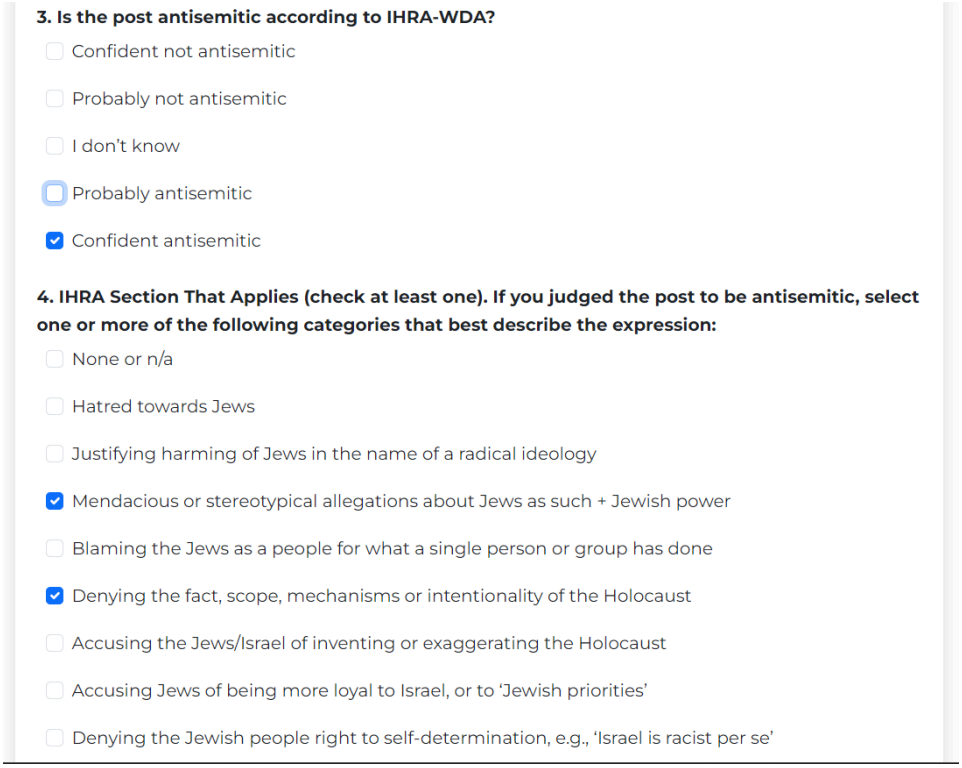
3.1 Guidelines Followed

Annotations followed the **IHRA-WDA framework**, including its general definition and the 11 specific examples of antisemitism. Each post was reviewed for:

1. Whether it was online.
2. Whether it was antisemitic.
3. The applicable IHRA sections.
4. Content types (text, image, link, etc.).
5. Whether it was reporting or condemning antisemitism.

3.2 Annotation Platform

We used the Datathon-provided **Annotation Portal**, which allowed manual entry of responses based on each IHRA question. All 100 tweets were pre-processed into a format compatible with the system.



3.3 Annotation Output

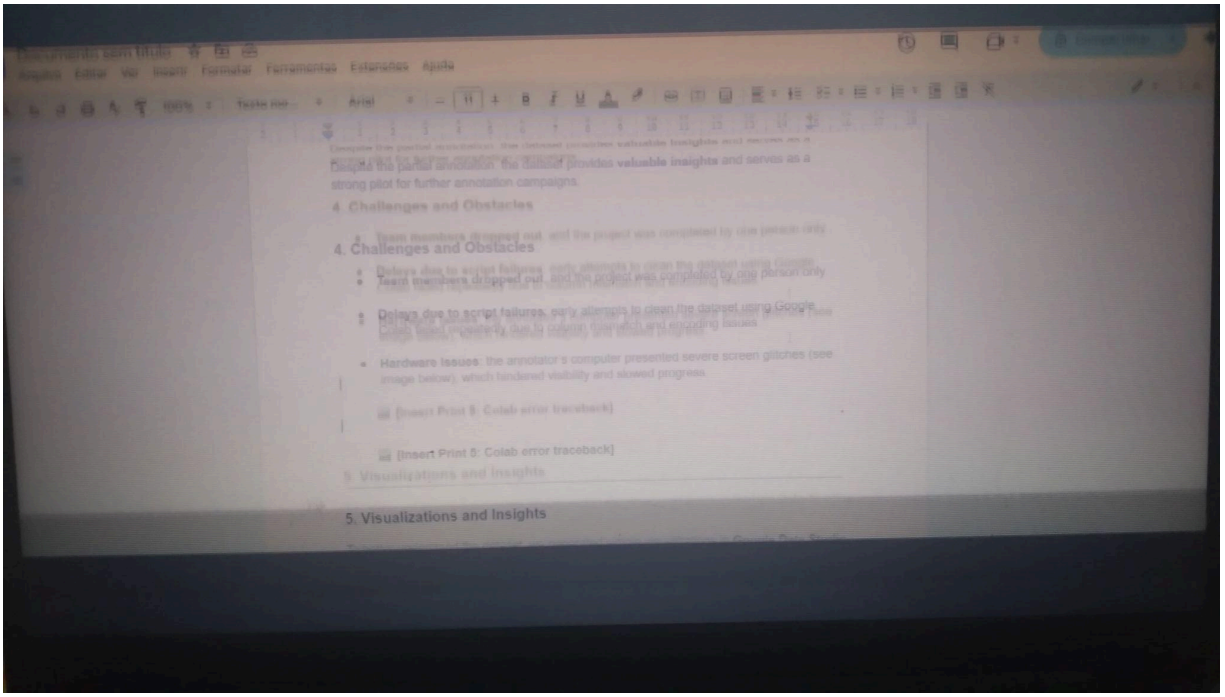
Out of the 100 posts:

- ✅ 20 were **fully annotated**.
- ❌ 80 were not annotated due to **lack of time and team disbandment**.

Despite the partial annotation, the dataset provides **valuable insights** and serves as a strong pilot for further annotation campaigns.

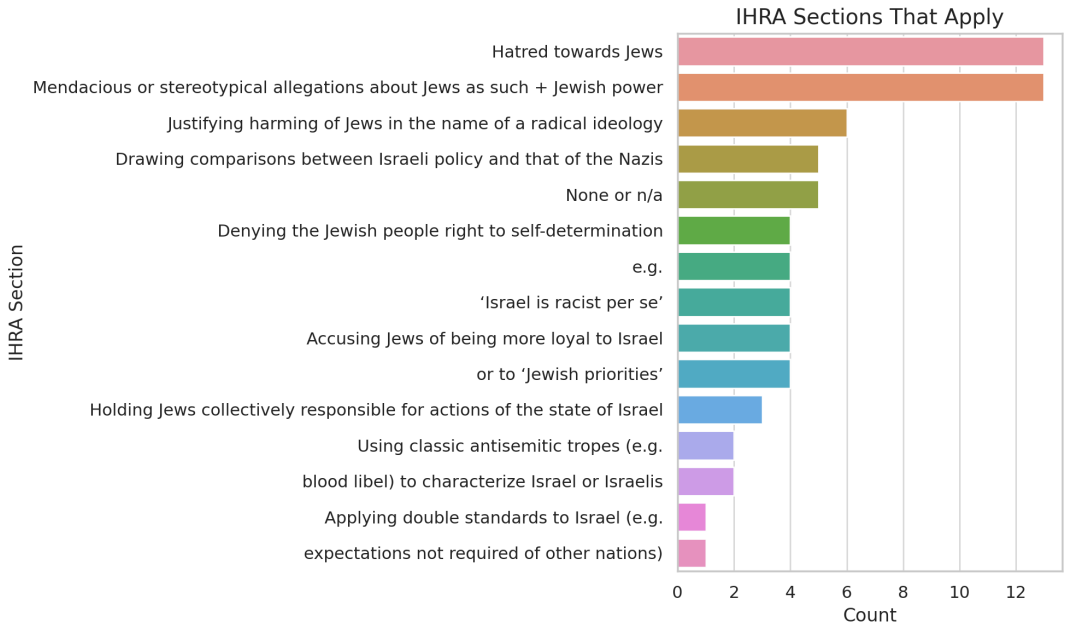
4. Challenges and Obstacles

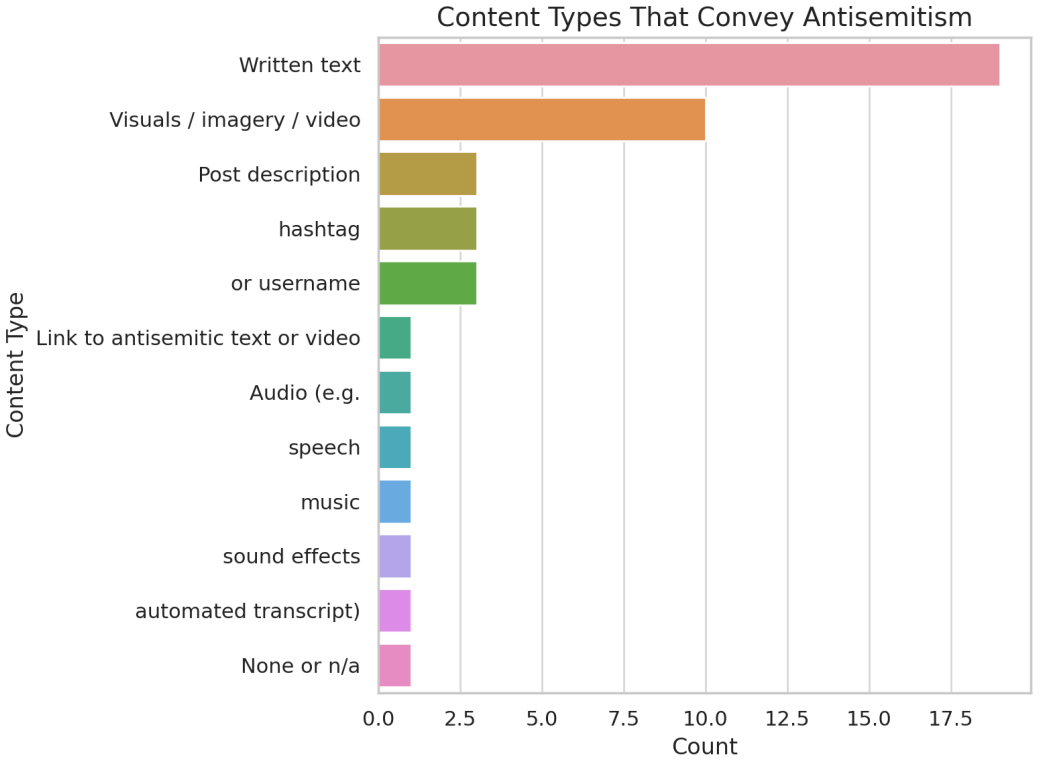
- **Team members dropped out**, and the project was completed by one person only.
- **Delays due to script failures**: early attempts to clean the dataset using Google Colab failed repeatedly due to column mismatch and encoding issues.
- **Hardware issues**: the annotator's computer presented severe screen glitches (see image below), which hindered visibility and slowed progress.

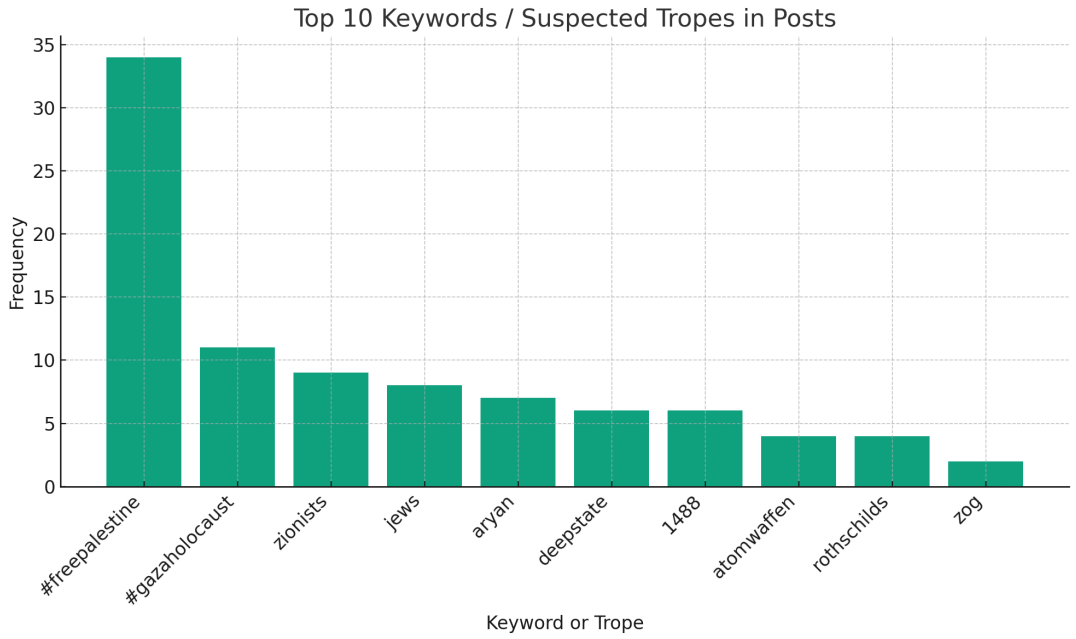


5. Visualizations and Insights

To better understand the dataset, we generated simple visualizations using the cleaned CSV file (`parsed_tweets_for_annotation.csv`).


IHRA Sections That Apply
IHRA Section
Hatred towards Jews
Mendacious or stereotypical allegations about Jews as such + Jewish power
Justifying harming of Jews in the name of a radical ideology
Drawing comparisons between Israeli policy and that of the Nazis
None or n/a
Denying the Jewish people right to self-determination
e.g.
‘Israel is racist per se’
Accusing Jews of being more loyal to Israel
or to ‘Jewish priorities’
Holding Jews collectively responsible for actions of the state of Israel
Using classic antisemitic tropes (e.g.
blood libel) to characterize Israel or Israelis
Applying double standards to Israel (e.g.
expectations not required of other nations)
0
2
4
6
8
10
12
Count


Content Types That Convey Antisemitism
Content Type
Written text
Visuals / imagery / video
Post description
hashtag
or username
Link to antisemitic text or video
Audio (e.g.
speech
music
sound effects
automated transcript)
None or n/a
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Count



These visualizations helped us grasp patterns in the dataset and identify overlaps between hashtags and conspiracy narratives.

6. Scripts and Tools Used

We used a simple but effective **Python script** to clean the data, identify posts containing our keywords, and prepare it for annotation. The script includes:

- Case-insensitive keyword matching
- Filtering the `description` column
- Exporting filtered data in annotation-ready format

> See the `main.py`, `requirements.txt`, and `README.md` files in the project repository.
> Script also included in `Google Colab` (see initial attempts in broken `.ipynb` files).

Note: The code was developed with the assistance of **ChatGPT**, under human review, for guidance and support during data parsing and annotation preparation. The final decisions and all annotations were made manually by the author.

7. Challenge #2 (Partial Completion)

Due to lack resources, we **did not train a transformer model**. However, here's what was planned:

- Use `BERTweet` as the base model
- Fine-tune using the 100 annotated posts
- Preprocess text using standard tokenization
- Evaluate with F1-score and accuracy

> For partial credit, we included a cleaned subset and a preprocessing script (`preprocess.py`) in the GitHub repository.

8. Final Considerations

This project reflects a **sincere, independent, and fully documented attempt** to meet the Datathon 2025 Challenge #1. Despite the technical and logistical difficulties, the core objectives were fulfilled:

- A strong rationale and implementation of keyword filtering
- Consistent application of IHRA-WDA to real tweets
- Full documentation of scripts and process
- Incomplete annotation due personal constraints

The work is publicly available in the GitHub repository for further development, reuse, or critique.

GitHub Repository

GitHub Link:

Check-list of Deliverables

Deliverable	Status
Keyword-based data scraping	Done

Cleaned & filtered dataset	Done
IHRA annotation guidelines applied	Done
Manual annotation of subset	20/100
Report + explanation of methodology	Done
Code and scripts (main.py, preprocess.py, etc.)	Done
Readme and Requirements	Done
Charts and visualizations	Done
Model (Challenge #2)	Partial
Reflection and final summary	Done