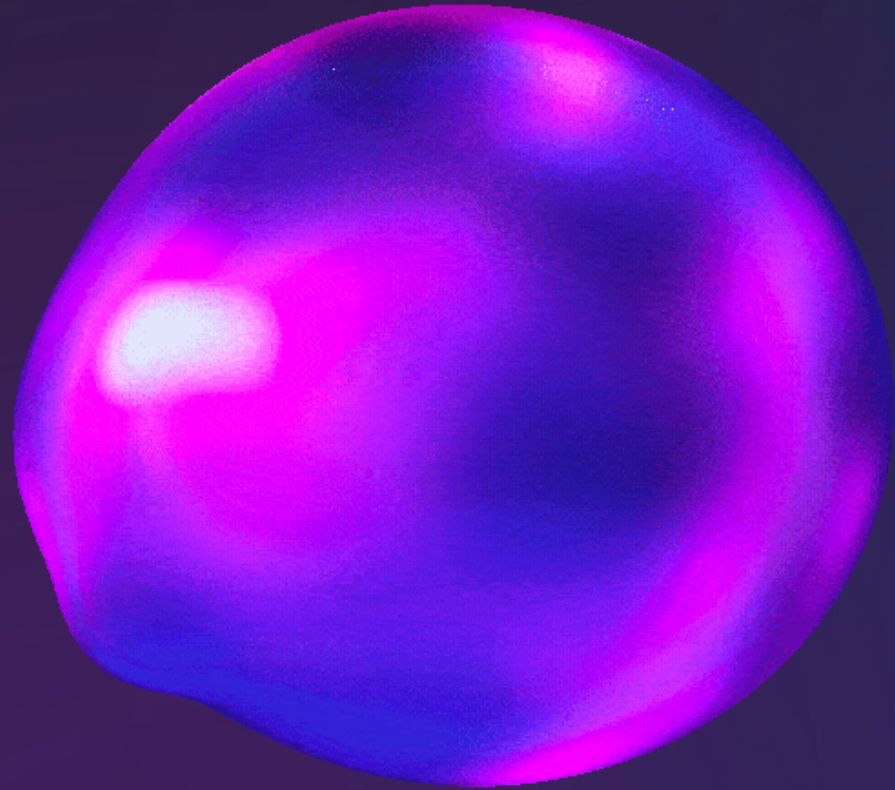




REDDIT SCRAPING

EXPLORING THE PALESTINE SUBREDDIT

Rihab Id m'hand
Asmaa Bazighe
Lina mouissou
Yahia Bourraoui



About Our Project

Project Overview:

- Our project delves into the heart of the Palestine subreddit, meticulously collecting and analyzing data through the Reddit API.

Significance of NLP:

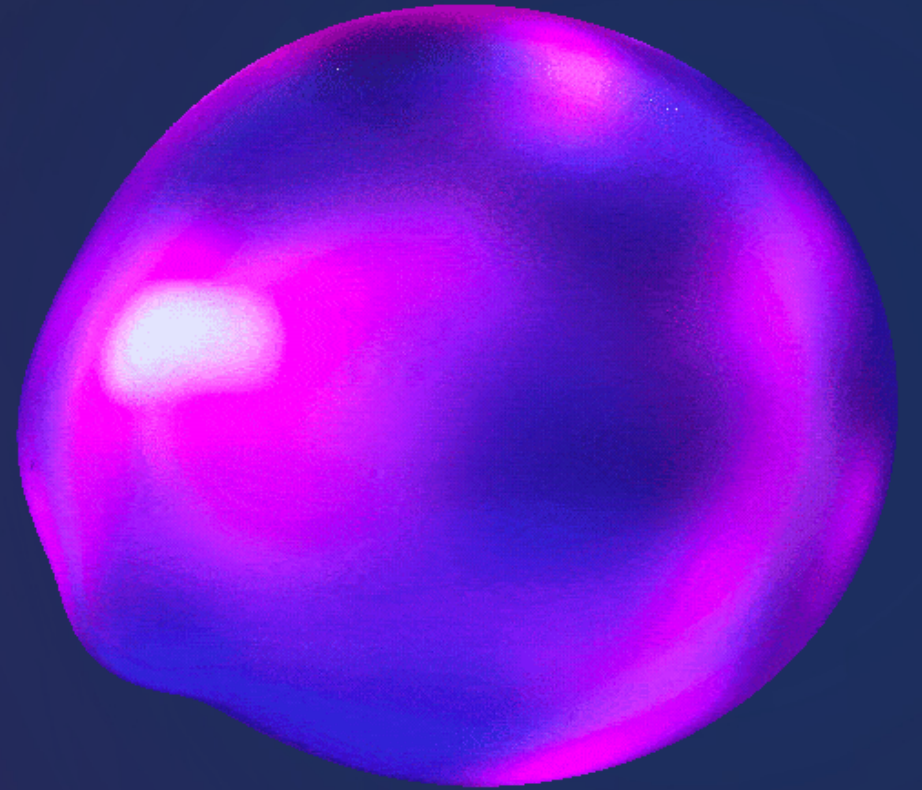
- Applying advanced Natural Language Processing (NLP) techniques, we unravel user sentiments and decode key themes, providing unprecedented depth to discussions.

Neo4j Implementation:

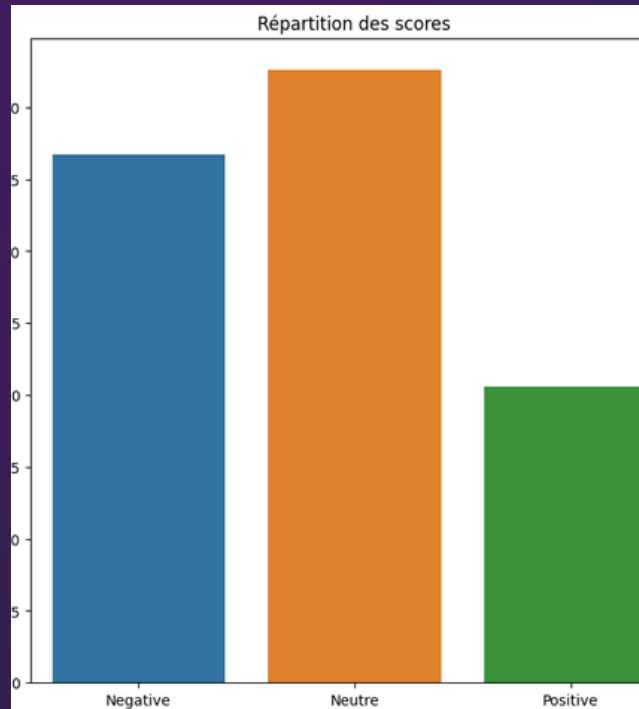
- Elevating our approach, we leverage Neo4j, a powerful graph database, to map intricate relationships, revealing profound insights into headlines, dates, scores, and upvotes.

The Data collection process

In our meticulous data collection process, we leveraged the robust capabilities of the Reddit API, utilizing unique client identifiers and secrets to seamlessly access and retrieve diverse content from the Palestine subreddit. Through the strategic use of ``reddit.subreddit().hot``, ``reddit.subreddit().top``, and ``reddit.subreddit().controversial``, we systematically gathered data to ensure a comprehensive representation of the subreddit's dynamics. This dynamic approach allowed us to capture trending, top-rated, and controversial content, providing a well-rounded dataset for our subsequent analyses. The use of API, client identifiers, and subreddit-specific methods has been instrumental in obtaining a rich and varied collection of headlines, text submissions, and comments, forming the foundation for our in-depth exploration of the Palestine subreddit.

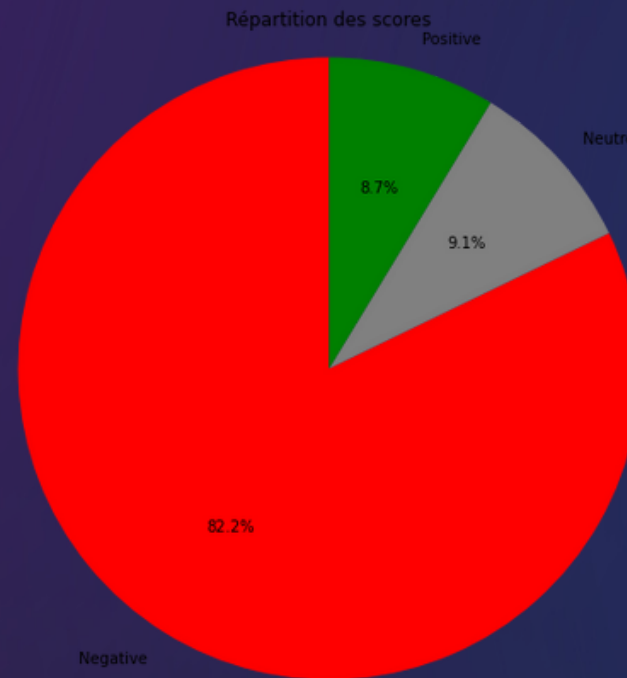


The NLP process



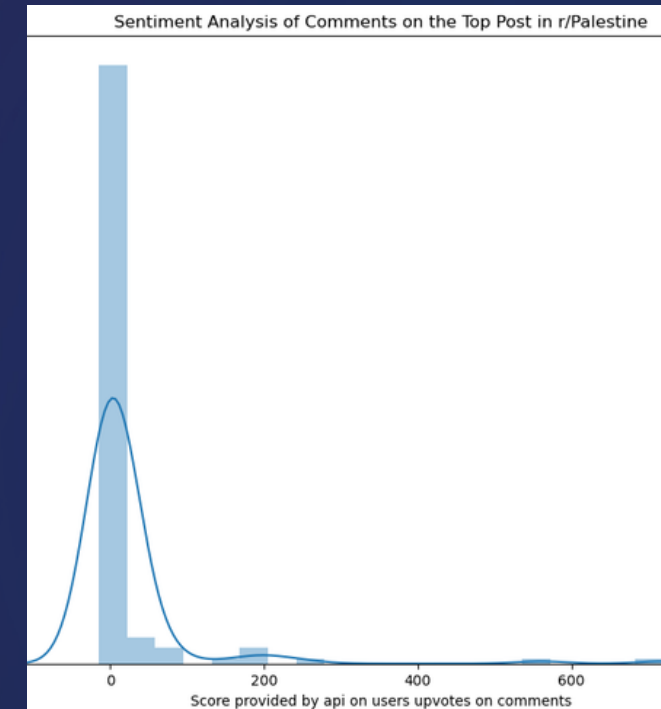
On the headlines

We gain a nuanced understanding of the community's sentiments and unveil underlying patterns



On the post

It provides a comprehensive analysis, allowing us to uncover deeper insights and sentiment trends.



On the comments

We extract insights, dissecting user sentiments and uncovering the underlying narrative that fueled its engagement

IMPLEMENTING NEO4J

Neo4j Overview:

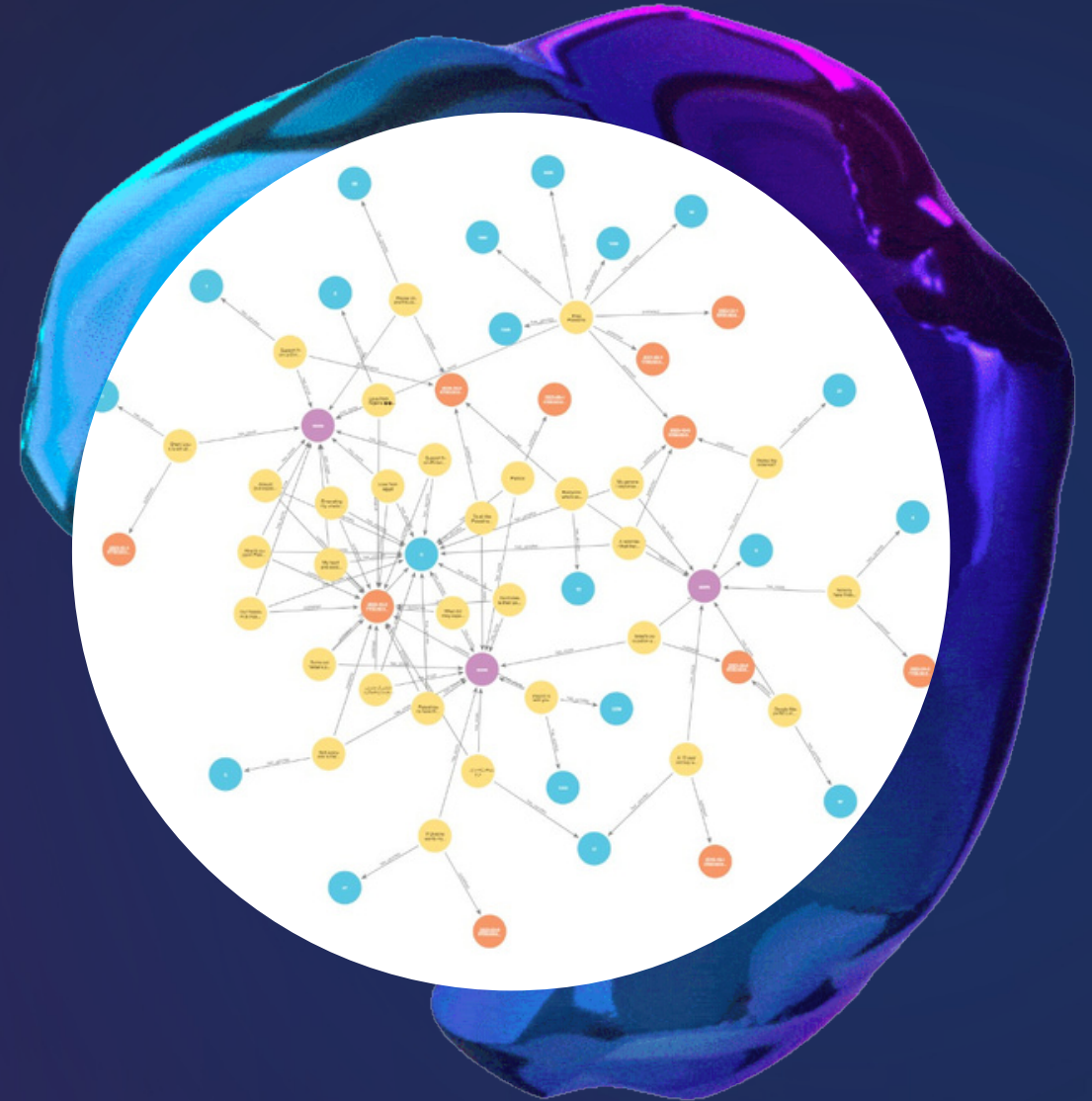
Neo4j, a powerful graph database, excels in capturing intricate relationships within datasets, making it ideal for nuanced project insights.

*Why Neo4j:

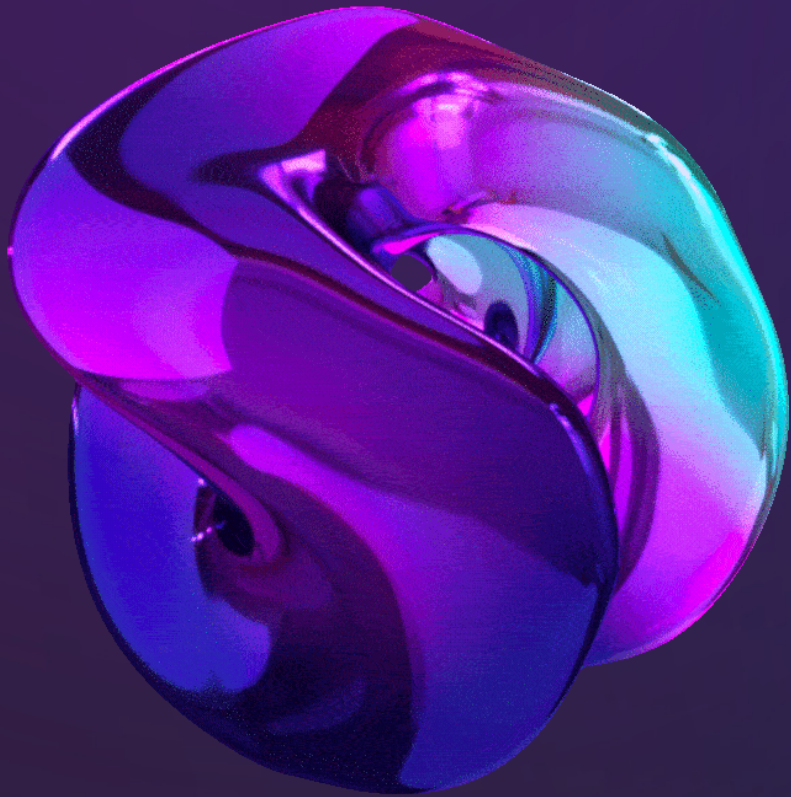
Chosen for its efficiency in handling interconnected data, Neo4j aligns seamlessly with our dataset's relational nature—headlines, dates, scores, and upvotes—providing intuitive exploration.

Graph Components:

In the Neo4j graph, headlines serve as nodes, relationships convey temporal aspects, popularity, and engagement, offering a concise yet comprehensive view of the Palestine subreddit dynamics.



Future Developments



Advanced NLP Techniques

Explore sophisticated NLP methods to deepen sentiment analysis and extract richer context from user interactions

Real-Time Data Integration

Incorporate real-time updates into our data collection process to ensure our insights reflect the most current trends and discussions.

Dataset Expansion

Enrich the dataset by including a broader range of content, providing a more diverse and representative sample for analysis

CONCLUSION

Our project has uncovered valuable insights from the Palestine subreddit. We invite your questions and discussions on our findings, methodology, and potential avenues for future exploration. Thank you for your attention.