# TWITTER SENTIMENT ANALYSIS

## 2022 AI TWEETS

Liangyuwang

# BACKGROUND & OBJECTIVE

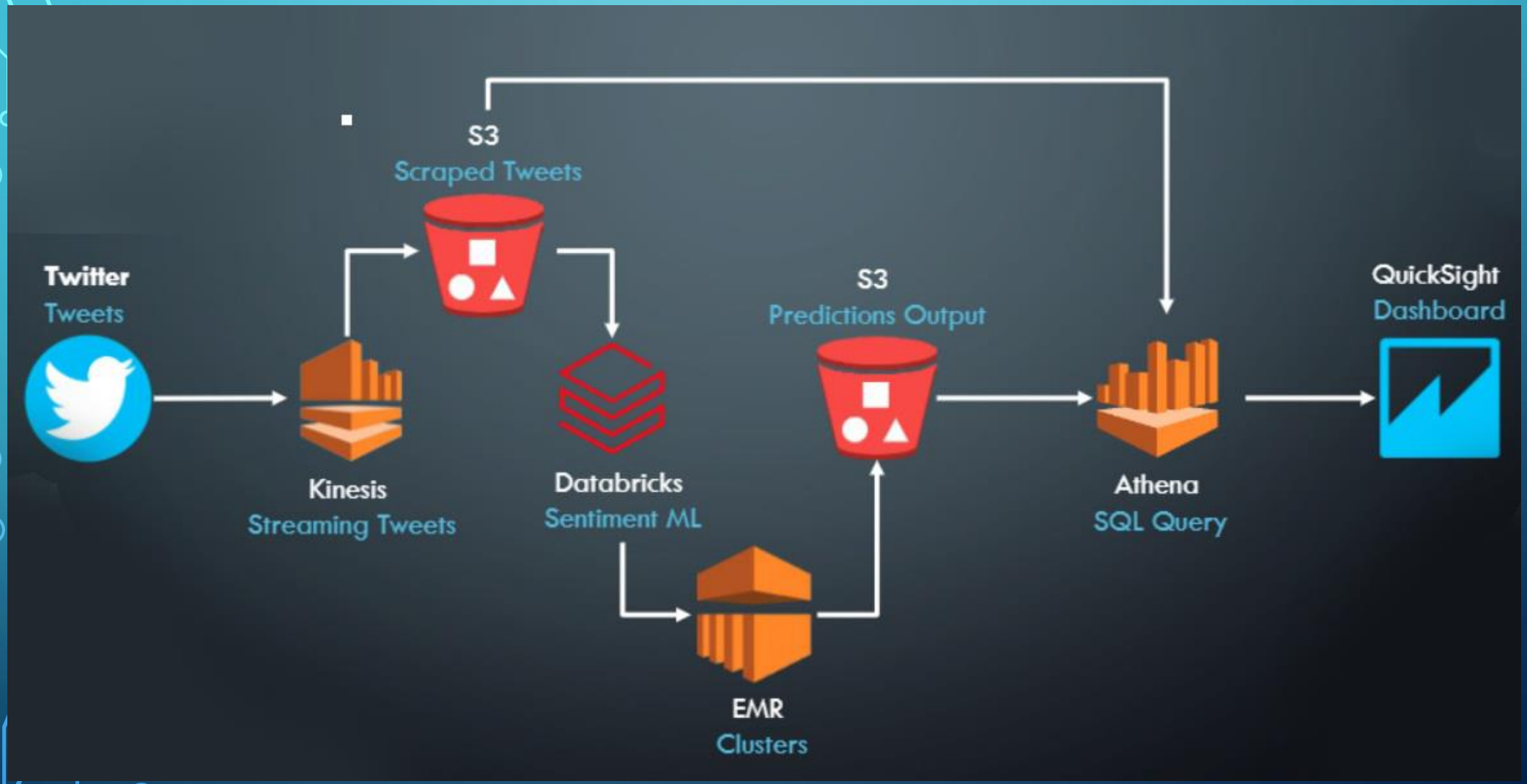
## Background

- AI became a hot topic in 2022 because of introduction of Large Language Models and diffusion models
- Some people were positive towards this development, while others were negative

## Objective

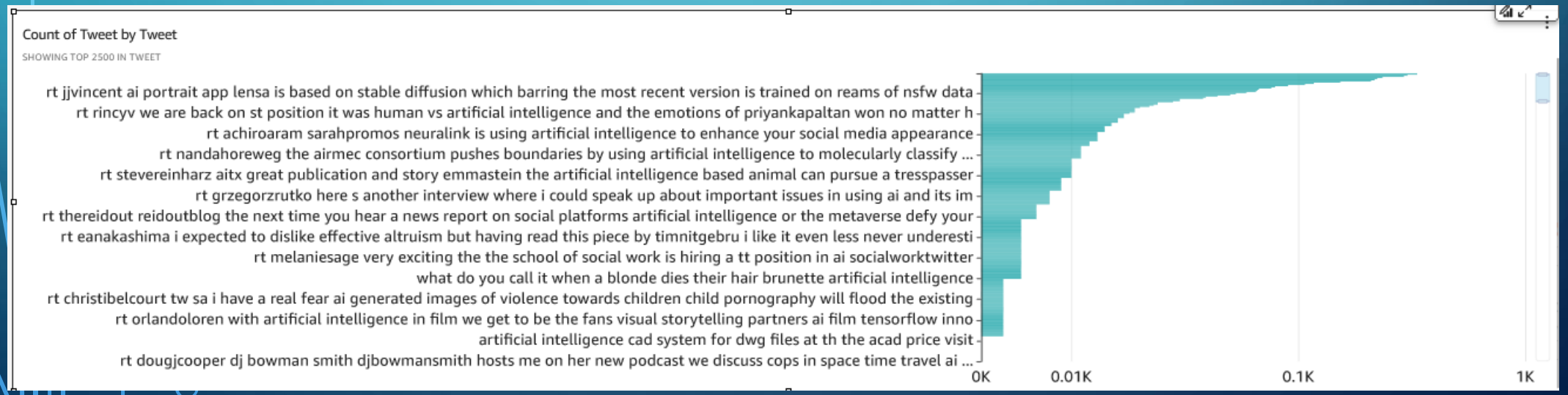
- Scrap #AI tweets
- Perform sentiment analysis
- Present findings

# DATA


S3
Scraped Tweets
Twitter
Tweets
Kinesis
Streaming Tweets
Databricks
Sentiment ML
S3
Predictions Output
EMR
Clusters
Athena
SQL Query
QuickSight
Dashboard

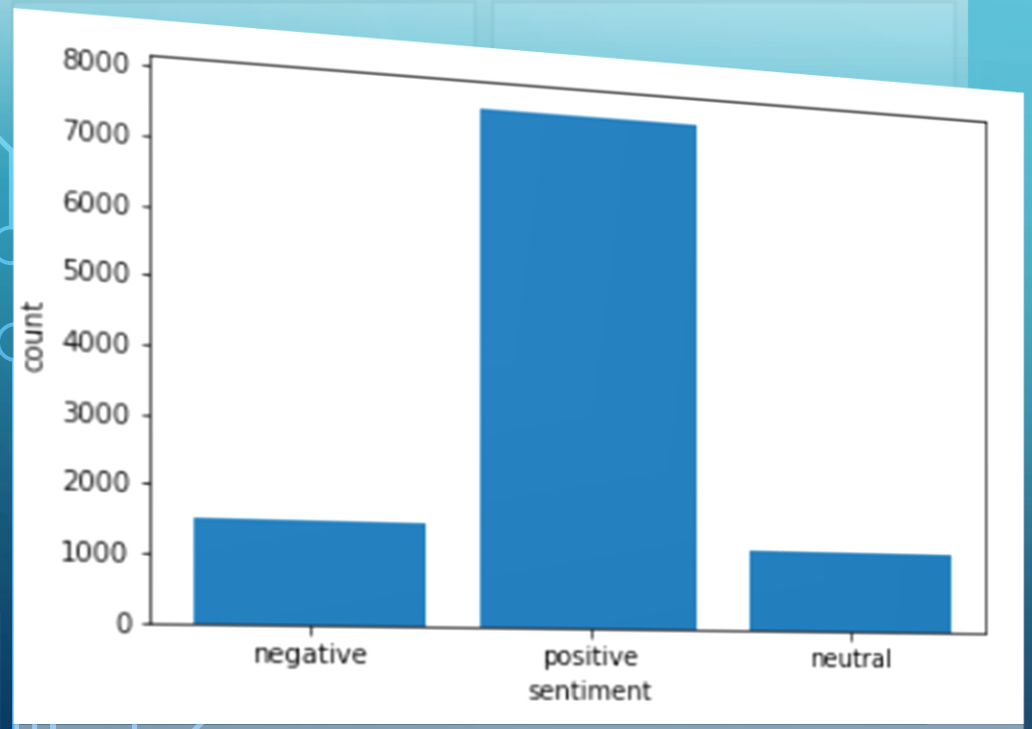
# TWEET ANALYSIS

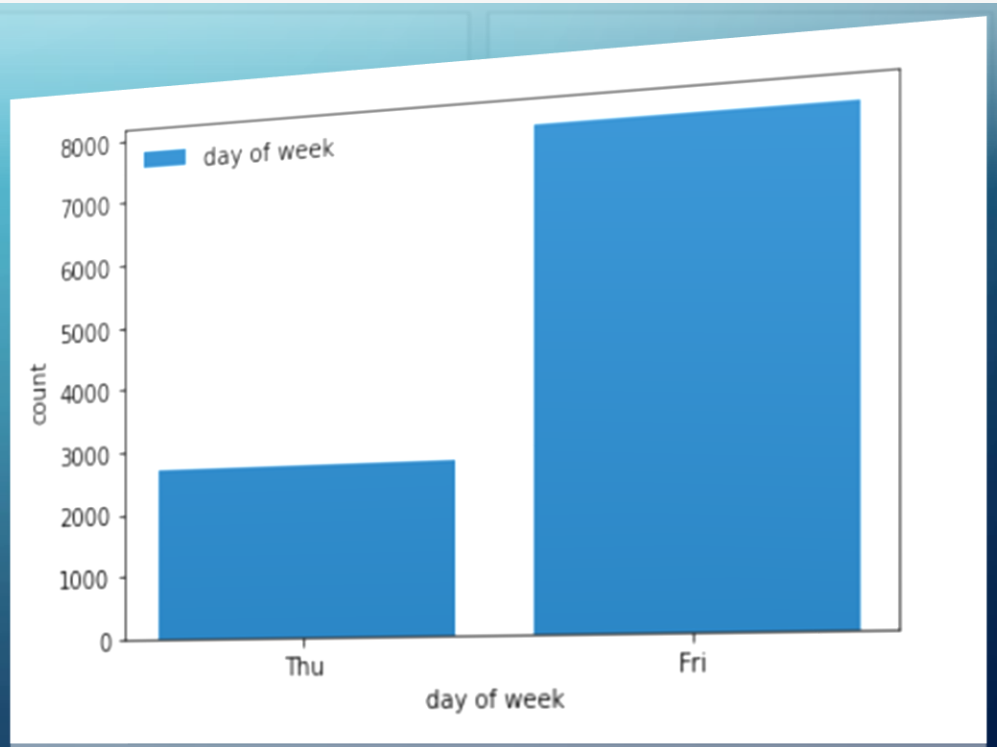
| id | name | username | tweet | followers_count | location | geo | created_at |
|---|---|---|---|---|---|---|---|
| 1601172099045158912 | YUNUS HANBAL | HanbalYunus | @CryptoEmdarks Th... | 21 | None | None | Fri Dec 09 11:09:... |
| 1601172122730041344 | ASLI HANBAL | HanbalAsli | @Crypto__Diva #GP... | 152 | None | None | Fri Dec 09 11:09:... |
| 1601172161372491778 | YUNUS HANBAL | HanbalYunus | @CryptoThro There... | 21 | None | None | Fri Dec 09 11:09:... |
| 1601172171602419712 | ASLI HANBAL | HanbalAsli | @belufrancese #GP... | 152 | None | None | Fri Dec 09 11:09:... |
| 1601172214056767489 | YUNUS HANBAL | HanbalYunus | @cryptojack There... | 21 | None | None | Fri Dec 09 11:09:... |
| 1601172226631311362 | ASLI HANBAL | HanbalAsli | @CryptoThro #GPLE... | 152 | None | None | Fri Dec 09 11:09:... |
| 1601172266460454914 | ASLI HANBAL | HanbalAsli | @cryptoworld202 #... | 152 | None | None | Fri Dec 09 11:09:... |
| 1601172313017581568 | ASLI HANBAL | HanbalAsli | @pascualprincipe ... | 152 | None | None | Fri Dec 09 11:10:... |



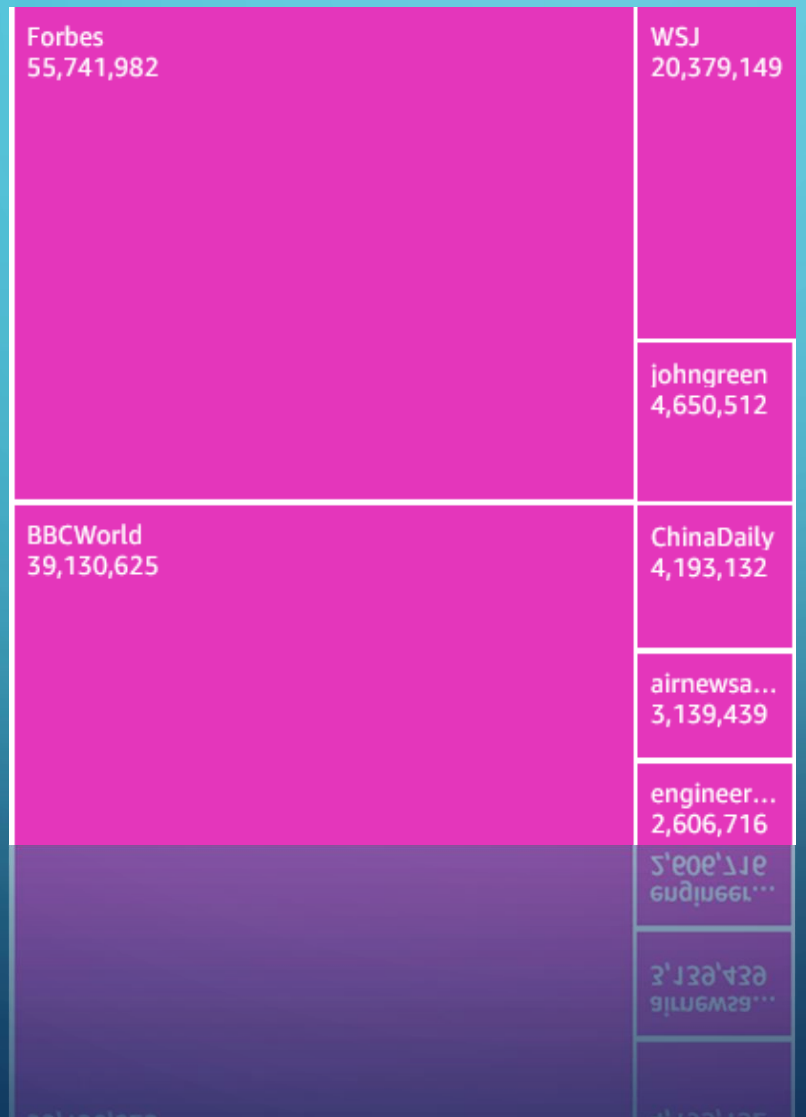
# TWEET ANALYSIS

| Total Number of Tweets | Positive Sentiment Tweets | Negative Sentiment Tweets | Neutral Sentiment Tweets |
|---|---|---|---|
| 10491 | 7753 | 1509 | 1229 |

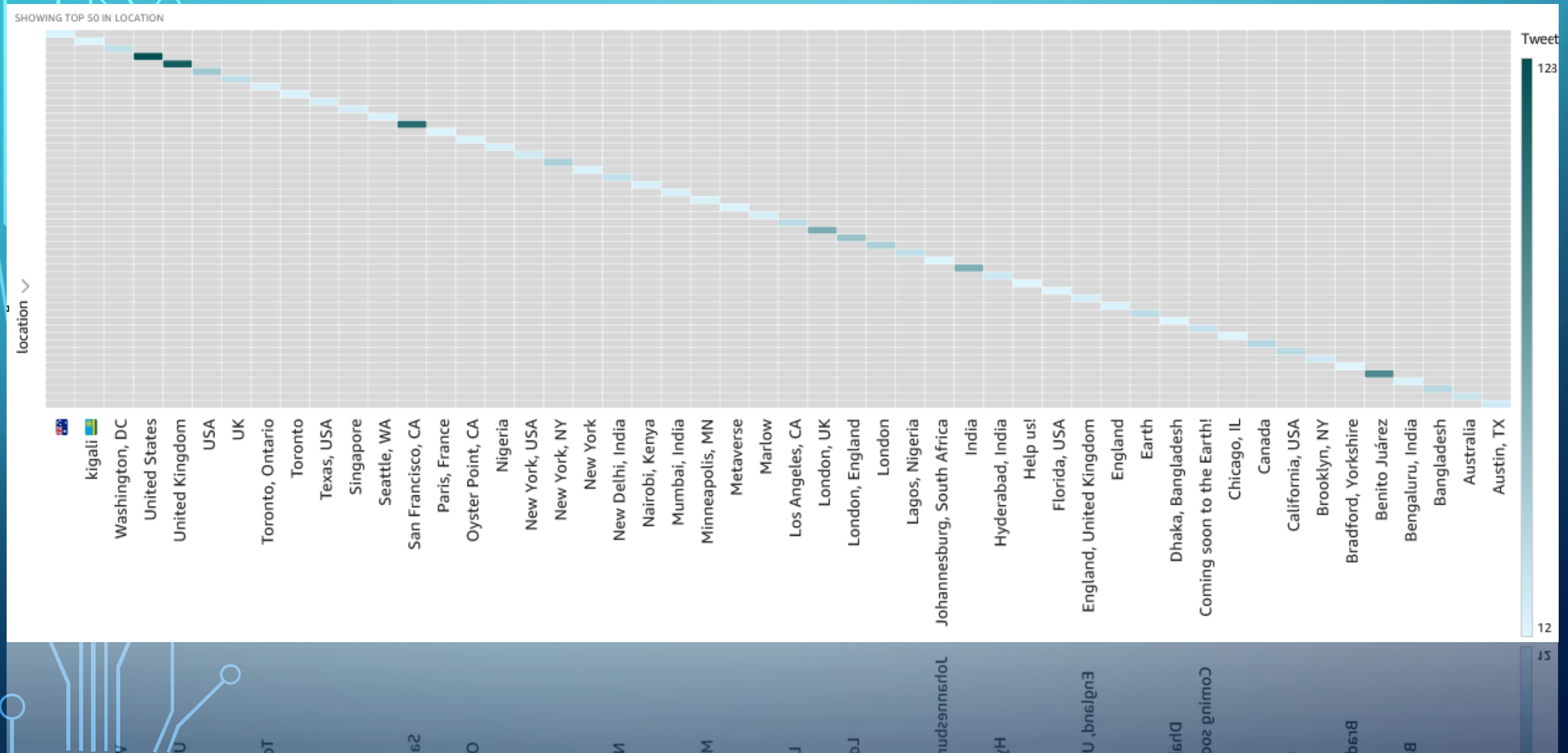


# TWEET ANALYSIS


# TWEET ANALYSIS



# TWEET ANALYSIS

## Logistic Regression

Accuracy Score: 0.8977
ROC-AUC: 0.8976

## Decision Tree

Accuracy Score: 0.7974
ROC-AUC: 0.7426

# CONCLUSION

## Challenges

- Databricks community edition cluster has 2 hours limit
- Databricks community edition cluster is very slow when working with large data sets
- 59% of tweets are retweets

## Conclusion

- Most people are positive towards developments in AI
- Logistic regression is better than Decision Tree model
- Logistic Regression model accuracy is 90%