TEXAS
The University of Texas at Austin

# MILESTONE 10

CS 327E ELEMENTS OF DATABASES FALL 2018

**PREET POPAT, HAI LAN**
CATALYSER

# Datasets

| Tweets mentioning stocks | |
|---|---|
| Source | Kaggle |
| Format | CSV |
| Tables | tweets_data<br>company_lookup |
| Link | |

| S&P 5 year stock price data | |
|---|---|
| Source | Kaggle |
| Format | CSV |
| Tables | 5yr_stock_data |
| Link | |

| id | tweet | timestamp | source | symbols | url | verified |
|---|---|---|---|---|---|---|

**Fig 1: tweet_data**

| | | |
|---|---|---|
| 36 | BIDU | Baidu |
| 37 | CX | CEMEX |
| 38 | CI | Cigna |
| 39 | DOV | Dover |
| 40 | ETN | Eaton |

**Fig 2: company_lookup**

| Row | date | open | high | low | close | volume | Name |
|---|---|---|---|---|---|---|---|
| 1 | 2013-02-08 | 15.07 | 15.12 | 14.63 | 14.75 | 8407500 | AAL |
| 2 | 2013-02-11 | 14.89 | 15.01 | 14.26 | 14.46 | 8882000 | AAL |
| 3 | 2013-02-12 | 14.45 | 14.51 | 14.1 | 14.27 | 8126000 | AAL |

**Fig 3: 5yr_stock_data**

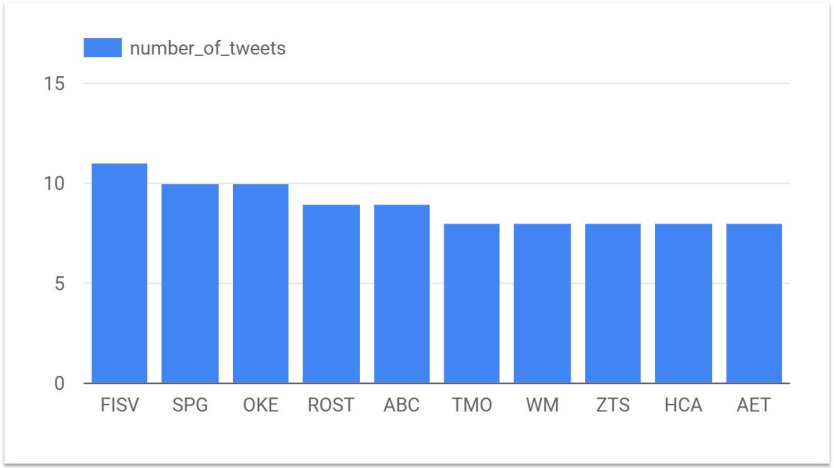
# Problem and Solution

# SQL

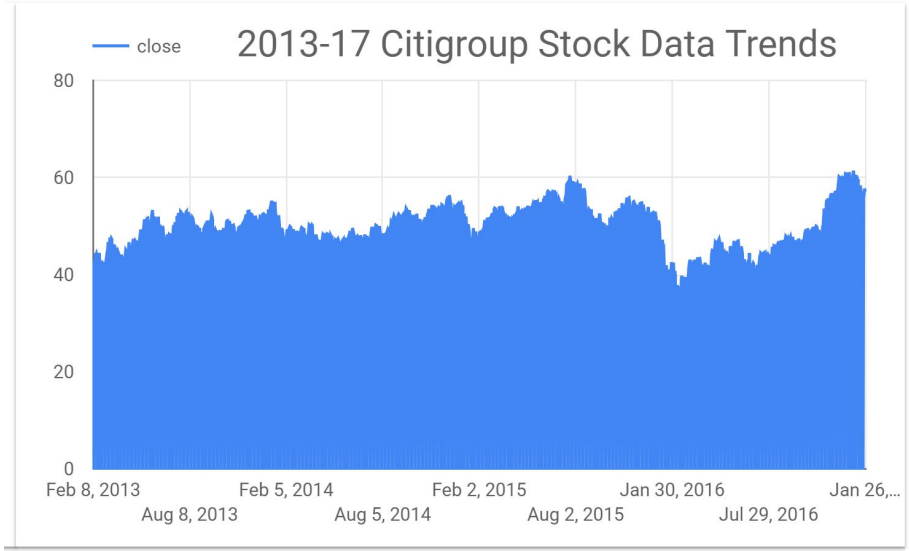
```
SELECT * FROM tweets.tweets_data WHERE symbols="FB" OR symbols="TWTR"

SELECT *, CAST(PARSE_DATETIME('%a %b %e %X +0000 %E4Y', timestamp) AS TIMESTAMP) AS new_timestamp FROM `avian-force-216105.tweets.tweets_data`
```

# Apache Beam: Link1 Link2

# Visualizations




**Most tweets of user bibeypost_stock by ticker**



# FUTURE IMPROVEMENTS

# SUGGESTIONS

# THANK YOU