



Twitter Sentiment Analysis

Tesla Case Study

Data Analytics Boot Camp Capstone Project 2021
University of California, Berkeley Extension

Chelsea Fusco, Samantha Villanueva, Haolun Wei

1

Overview

2

Data Gathering

3

Data Cleaning

4

Machine Learning

5

Scoring Tweets

6

Database

7

Visualization

8

Limitations &
Recommendations

Overview

Monitor Tesla's online perception and frequency of mentions on Twitter.

- 200+ million active users
- Vast and ideal for creating benchmarks
- Monitoring brand mentions to detect public perception

Why Twitter?

What Metrics?

- Tesla's twitter presence on an average day
- Frequency of mention
- Positive or negative sentiment

Sentiment Analysis

Sentiment Analysis is a procedure used to determine if a chunk of text is positive, negative or neutral. In text analytics, natural language processing (NLP) and machine learning (ML) techniques are combined to assign sentiment scores to the topics, categories or entities within a phrase.

Data analysts make use of sentiment analysis to extract information for market research as well as monitor brand conversation. Monitoring the online conversation surrounding a brand can help predict shifts in reputation and allow companies to better the consumer experience.



DATA GATHERING



Data Gathering

1

Training Dataset

- Sem-Eval 40,000 tweets
- Categorized *positive, negative, neutral* spanning a range of topics and time

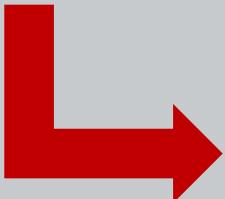
2

Tesla Sample Data

- Twitter APICall
- 67,200 tweets
- 200 per hour over a two-week range

Sem-Eval Training Data

	A
1	Raw Sem-Eval Data
2	260097528899452929neutralWon the match #getin . Plus\u002c tomorrow is a very busy day\u002c with Awareness Day\u2019s and debates. Gulp. Debates...
3	263791921753882624neutralSome areas of New England could see the first flakes of the season Tuesday.
4	264194578381410304negative@francesco_con40 2nd worst QB. DEFINITELY Tony Romo. The man who likes to share the ball with everyone. Including the other team.
5	264041328420204544neutral#Thailand Washington - US President Barack Obama vowed Wednesday as he visited storm-ravaged New Jersey shore to... http://t.co/XzI4LFhs
6	263816256640126976neutralDid y\u2019u2019all hear what Tony Romo dressed up as for Halloween? A Giants quaterback! Cause that\u2019s all he could throw to sunday night.
7	263722278712393728positiveTim Tebow may be available ! Wow Jerry \u002c what the heck you waiting for ! http://t.co/a7z9FBL4
8	262936443171778560positive@mariakaykay aga tayo tomorrow ah. :) Good night\u002c Ces. Love you! >D<
9	260486470828171265neutralTina Fey & Amy Poehler are hosting the Golden Globe awards on January 13. What do you think?
10	262968617233162240positiveLunch from my new Lil spot ...THE COTTON BOWLpretty good#1st#time#will be going back# http://t.co/Dbbj8xLZ
11	263790847424880641positiveSNC Halloween Pr. Pumped. Let\u2019s work it for Sunday....Packers vs....who knows or caresn. #SNC #cheerpracticeonhalloween 264255236435243008negative@jacquelinemegan I\u2019m sorry\u002c I Heart Paris is no longer available at the Rockwell branch! You may call 8587000 to get a copy transferred! :)



Tesla Data –Twitter API v2 Standard Basic

Search Tweets

Recent search endpoint allows you to programmatically access filtered public Tweets posted over the last week.

Date Range



START: 2021-11-01

END: 2021-11-07



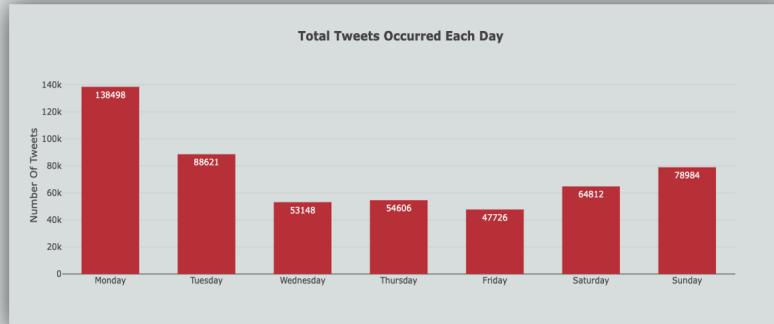
Total Tweets



WEEKLY TOTAL: 526395

Tweet Counts

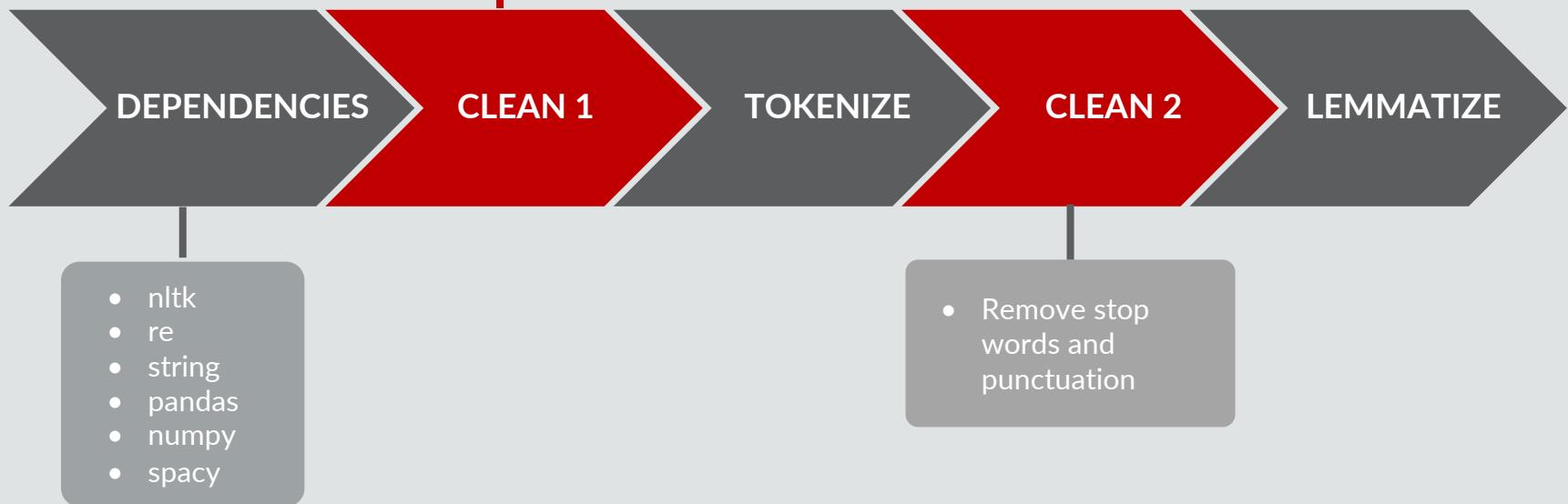
The recent Tweet counts endpoint allows you to programmatically retrieve the numerical count of Tweets for a query, over the last seven days.



DATA CLEANING



Data Cleaning



Data After Preprocessing



Text

Score

['dear', 'driving', 'newcastle', 'sunday', 'glad', 'liked', 'great']

1

["i'm", 'sad', "can't", 'watch', 'messi', 'argentina', 'squad']

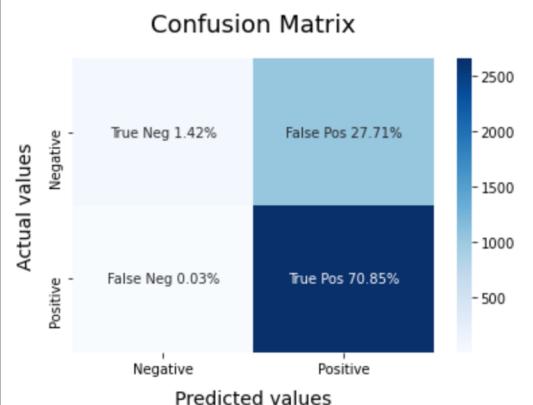
0

💡 MACHINE LEARNING



```
#Bernoulli Naive Bayes Model  
BNBmodel = BernoulliNB()  
BNBmodel.fit(X_train, y_train)  
model_Evaluate(BNBmodel)  
y_pred = BNBmodel.predict(X_test)
```

	precision	recall	f1-score	support
0	0.98	0.05	0.09	1090
1	0.72	1.00	0.84	2653
accuracy			0.72	3743
macro avg	0.85	0.52	0.46	3743
weighted avg	0.80	0.72	0.62	3743

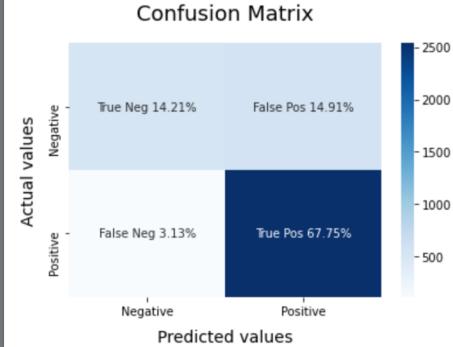


Bernoulli Naïve Bayes Model

Logistic Regression Model

```
#LR Model  
LRmodel = LogisticRegression(C = 2, max_iter = 1000, n_jobs=-1)  
LRmodel.fit(X_train, y_train)  
model_Evaluate(LRmodel)  
y_pred3 = LRmodel.predict(X_test)
```

	precision	recall	f1-score	support
0	0.82	0.49	0.61	1090
1	0.82	0.96	0.88	2653
accuracy			0.82	3743
macro avg	0.82	0.72	0.75	3743
weighted avg	0.82	0.82	0.80	3743

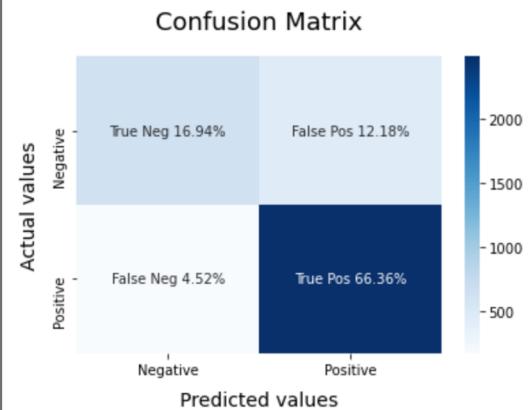


Linear Support Vector Classifier Model

#LinearSVC Model

```
SVCmodel = LinearSVC()  
SVCmodel.fit(X_train, y_train)  
model_Evaluate(SVCmodel)  
y_pred2 = SVCmodel.predict(X_test)
```

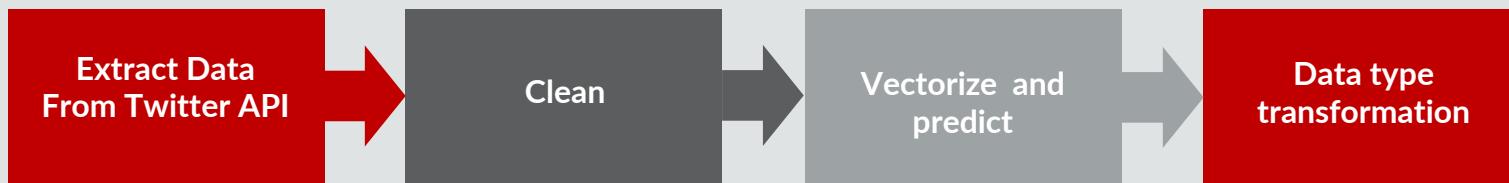
	precision	recall	f1-score	support
0	0.79	0.58	0.67	1090
1	0.84	0.94	0.89	2653
accuracy			0.83	3743
macro avg	0.82	0.76	0.78	3743
weighted avg	0.83	0.83	0.82	3743





SCORING TWEETS

Scoring Tweets

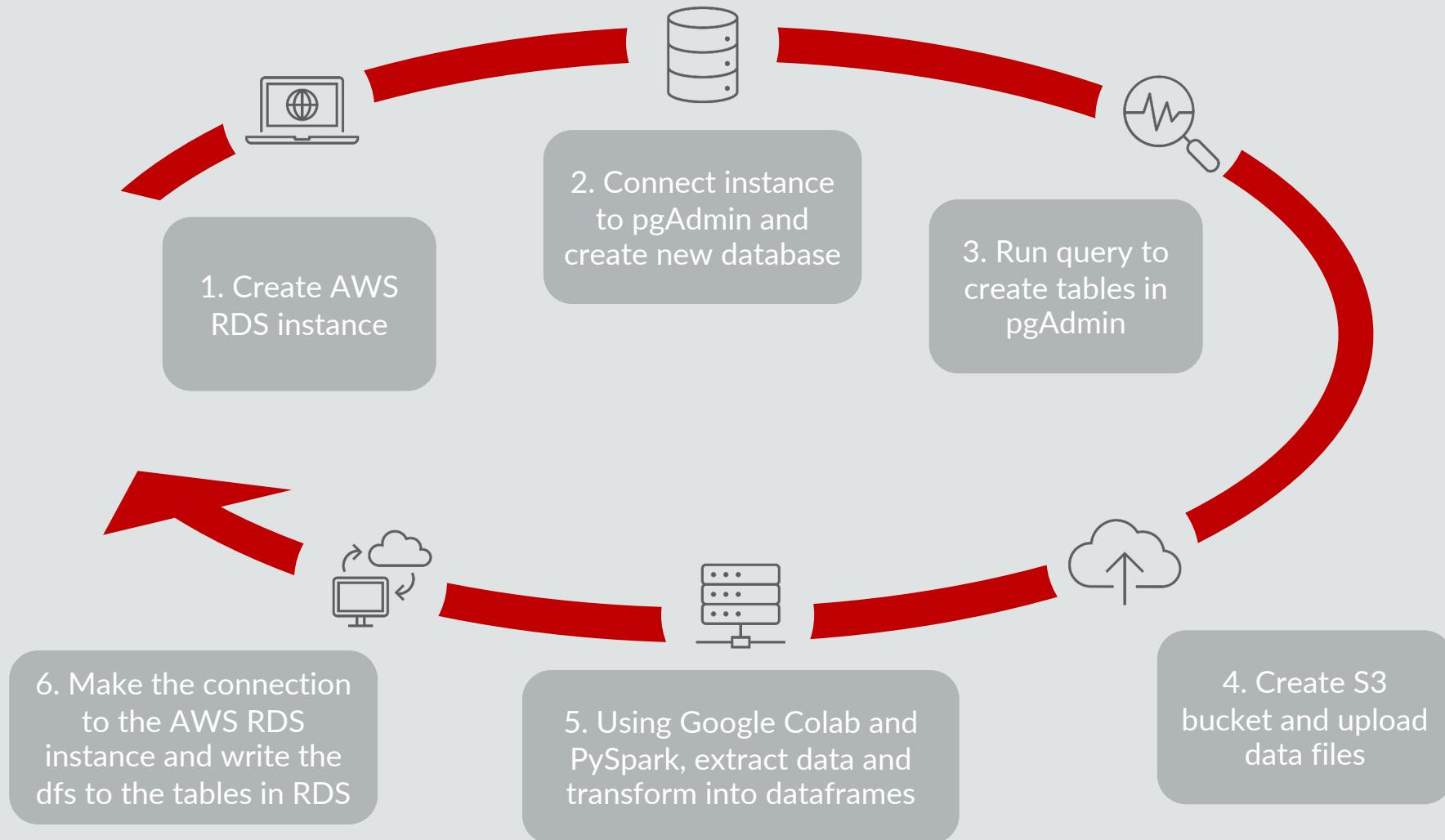


created_at	created_date	weekday	hours	tweet	text	score
2021-11-01 00:59:56+00:00	11/1/21	Monday	15	@SECGov Tesla is crazy. Something wrong with the stocks.	['tesla', 'crazy', 'wrong', 'stock']	0
2021-11-01 06:59:08+00:00	11/1/21	Monday	6	@WatcherGuru @elonmusk @Tesla @Shibtoken If this is confirmed it would be a great development.	['confirmed', 'great', 'development']	1

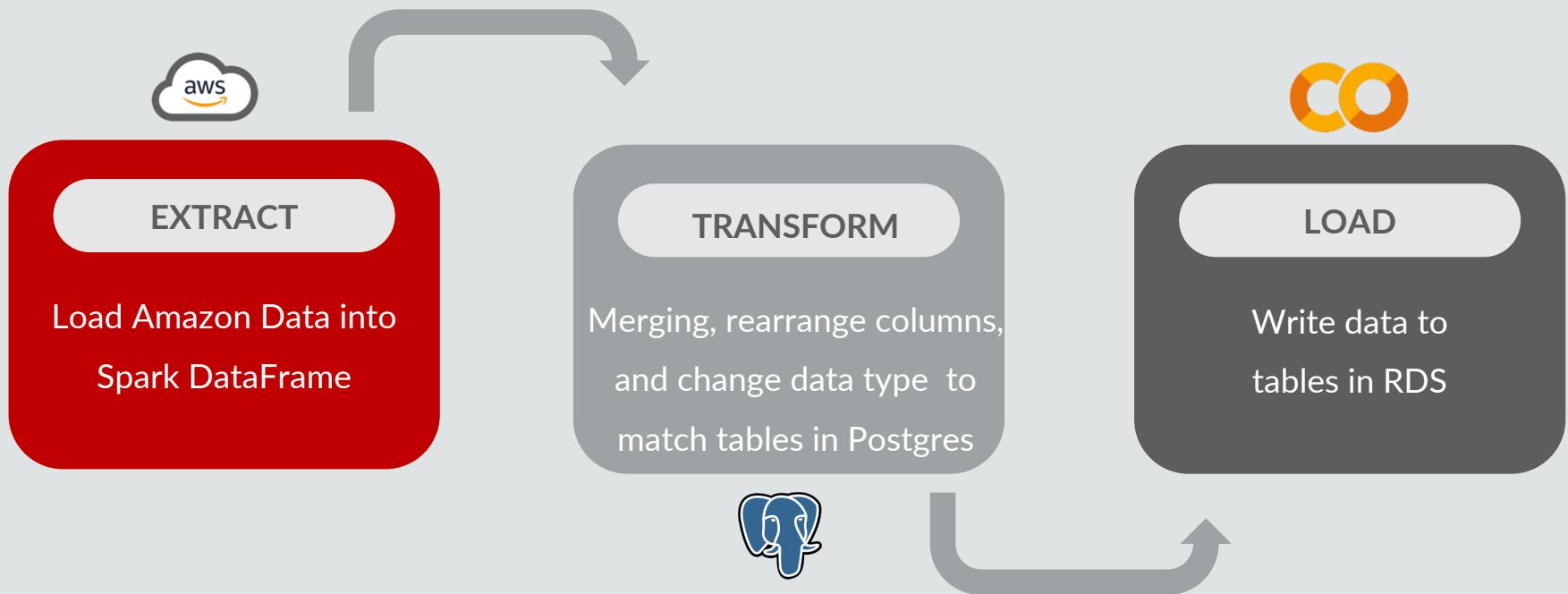


DATABASE





ETL Sample Code



Database Connection

The screenshot shows the pgAdmin interface for managing database connections. On the left, there's a tree view of servers and databases. Under 'twitterproject', three databases are listed: 'postgres', 'rdsadmin', and 'twitterproject'. The 'twitterproject' database is selected. To the right, a 'Connection' panel displays the following details:

- Connected?: True
- Host name/address: twitterproject.cgrabvwy3u7o.us-west-1.rds.amazonaws.com
- Port: 5432
- Maintenance database: postgres
- Username: postgres

Tables

The screenshot shows the pgAdmin interface displaying a list of tables under the 'api_count_table' database. The table 'api_tweets_table' is expanded, showing its structure. Other visible tables include 'api_count_table' and 'initial_tweets_table'.

1 `SELECT * FROM api_count_table;`

Data Output Explain Messages

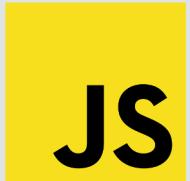
	end_time	start_time	created_date	weekday	tweet_count
	character varying	character varying	[PK] date	text	integer
1	2021-11-02T00:00:00.000Z	2021-11-01T00:00:00.000Z	2021-11-01	Monday	138498
2	2021-11-03T00:00:00.000Z	2021-11-02T00:00:00.000Z	2021-11-02	Tuesday	88621
3	2021-11-04T00:00:00.000Z	2021-11-03T00:00:00.000Z	2021-11-03	Wednesday	53148

Query Sample (join)

	created_date	weekday	total_count	collected_count	percentage_collected
	[PK] date	text	integer	bigint	numeric
1	2021-11-01	Monday	138498	5881	4.25
2	2021-11-02	Tuesday	88621	5392	6.08
3	2021-11-03	Wednesday	53148	5676	10.68
4	2021-11-04	Thursday	54606	5489	10.05
5	2021-11-05	Friday	47726	5195	10.89



VISUALIZATION



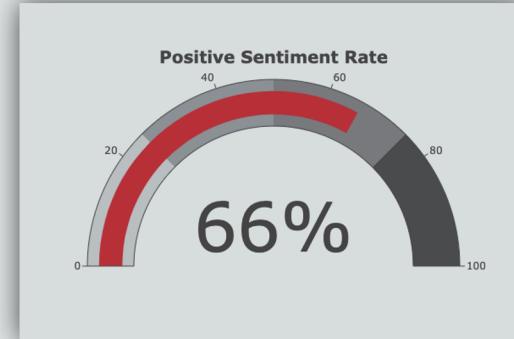
created_date

Date Range



START: 2021-11-01
END: 2021-11-07

weekday

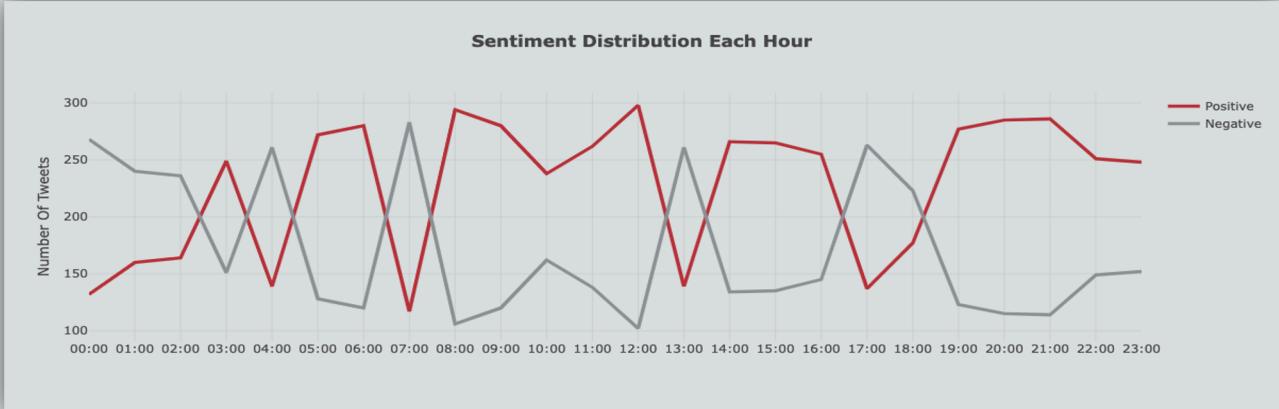


Sentiment Percentage

POSITIVE: 66%

NEGATIVE: 34%

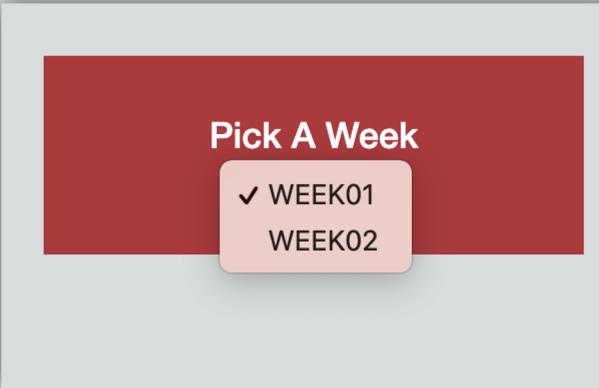
hours



Data For Visualization

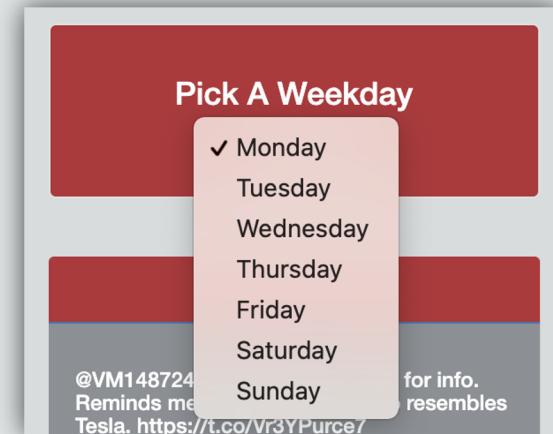
WEEKLY

- Total tweets @tesla occurred, per day
- Assign each day a week No.



WEEKDAY

- Random Sample collect
- 200 tweets per hour
- 4800 tweets per day
- 2 sets of weekdays: 9600 samples per day



Tesla Twitter Sentiment Dashboard

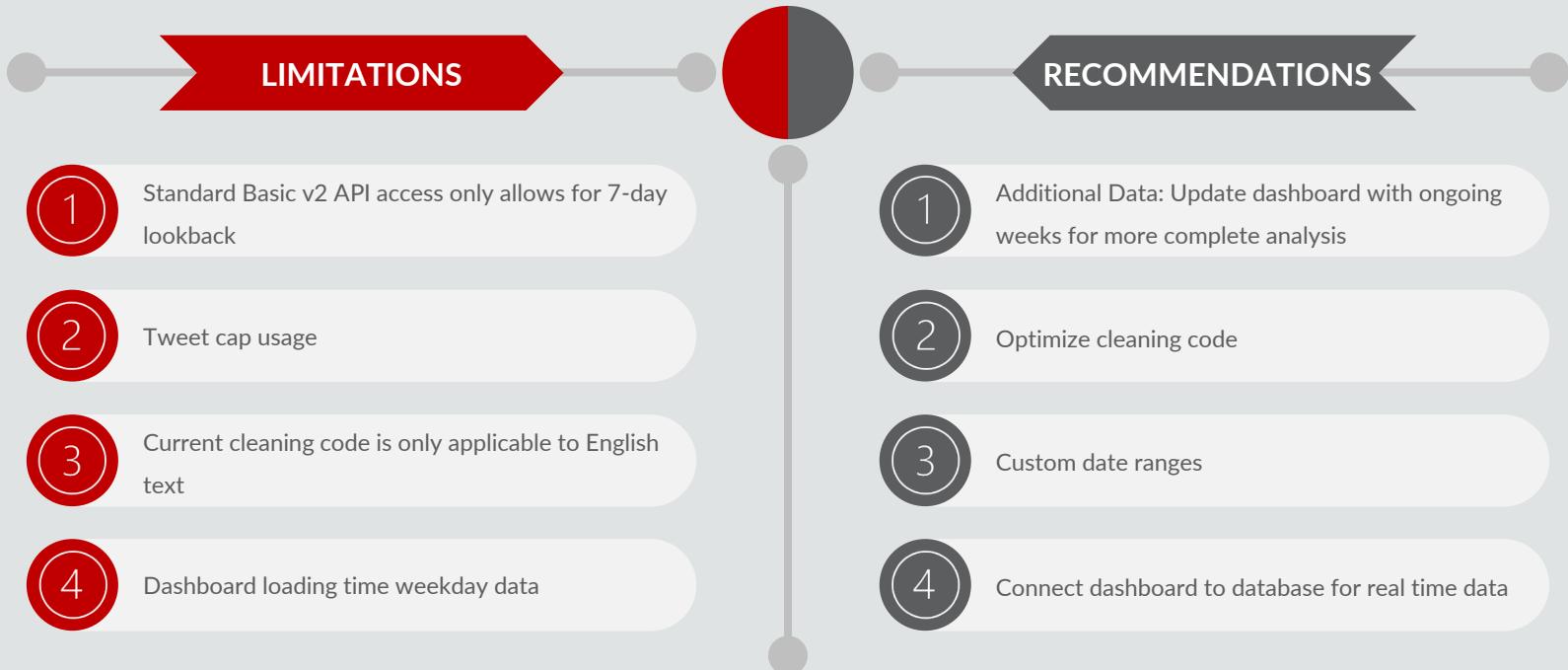
Use the interactive charts below to explore the sentiment data



 Weekly Data Dashboard

 Weekday Data Dashboard

Final Thoughts



REFERENCES

- Rosenthal, S. et al. (n.d.). *SemEval-2017 Task 4*. ALT QCRI.
<https://alt.qcri.org/semeval2017/task4/>
- Edward, A. (2021, June 13). *An Extensive Guide to collecting tweets from Twitter API v2 for academic research using Python 3*. Towards Data Science.
<https://towardsdatascience.com/an-extensive-guide-to-collecting-tweets-from-twitter-api-v2-for-academic-research-using-python-3-518fcb71df2a>
- Goyal, G.(2021, June 11). *Twitter Sentiment Analysis- A NLP Use-Case for Beginners*. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2021/06/twitter-sentiment-analysis-a-nlp-use-case-for-beginners/>
- Effrosynidis, D. (2019). *Text-preprocessing-techniques*. Github.
<https://github.com/Deffro/text-preprocessing-techniques>
- Other useful articles: <https://developer.twitter.com/en/docs/twitter-api/early-access>

THANK YOU

