

DS4A DATA ENGINEERING : TEAM 3

# Analyzing the Impact of Super Bowl Ads



# OUR TEAM



**Ashley Tejada**

Data Governance  
Analyst at Blue  
Water Thinking  
Albany, NY



**Hawa Anthony**

Software Engineering  
Specialist at  
Accenture  
Clinton, MD



**Prianka Ball**

Analytics Manager at  
Big Village with  
Agency Media team,  
New York City



**Odesha Hill**

Operations Research  
Analyst at USCG  
Alexandria, Virginia



**Paula Bonasera**

Data Visualization  
New York City



**Adeola Adesoba**

Data Scientist II -  
Solutions Architect  
at NVIDIA  
Little Rock, AR





30

40

50

40

30

# What is Super Bowl known for?



Game



Audience



Advertisement

30

40

50

40

30

# The Super Bowl Ad



Draws in large number  
of viewers  
**~ 92M**



Very expensive  
**30s ~6.5M**

# The Issue

Hard to access the impact of ads



30

40

50

40

30

# The Proposal

How can we understand the effect of Super Bowl ads?



Brand sentiments through  
tweets on Twitter



Brand search interest on Google



Brand stock prices

# The Impact

- Can help uncover what matters to the audience
- Better direct marketing efforts
- Keep track of positive, negative and viral trends
- Understand effect on stock price



# Subject of Analysis

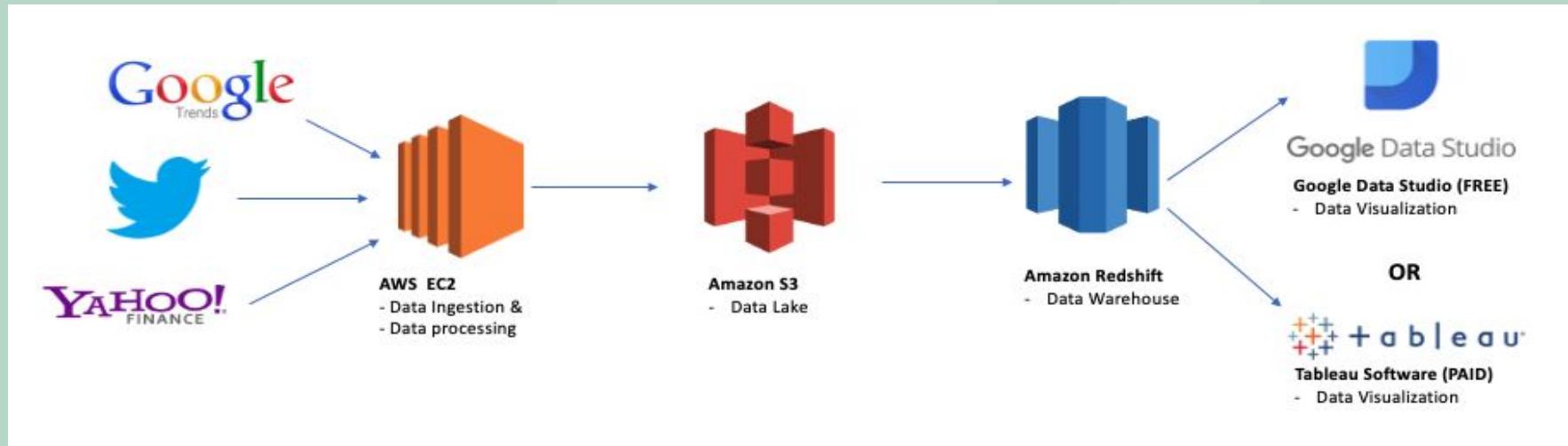
20 Companies with  
Super Bowl Ads

12 Industries

Timeline:  
2015 - 2022

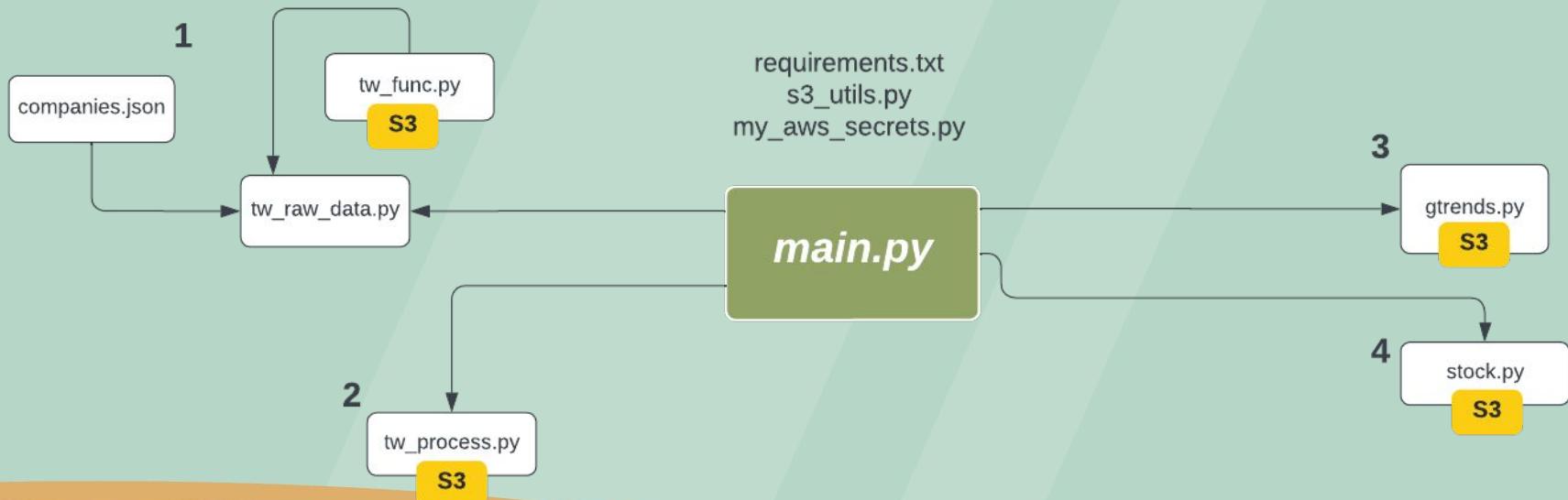


# DATA PIPELINE



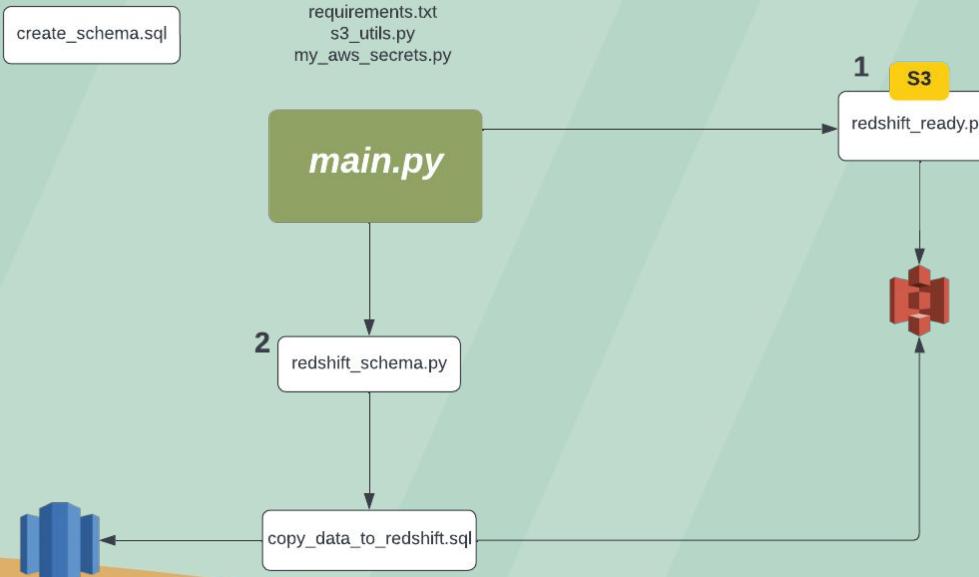
# AWS EC2 Python Scripts - STEP 1

Code: [https://github.com/Pball01/superbowl\\_analysis](https://github.com/Pball01/superbowl_analysis)



# AWS EC2 Python Scripts - STEP 2

Code: [https://github.com/Pball01/superbowl\\_analysis](https://github.com/Pball01/superbowl_analysis)

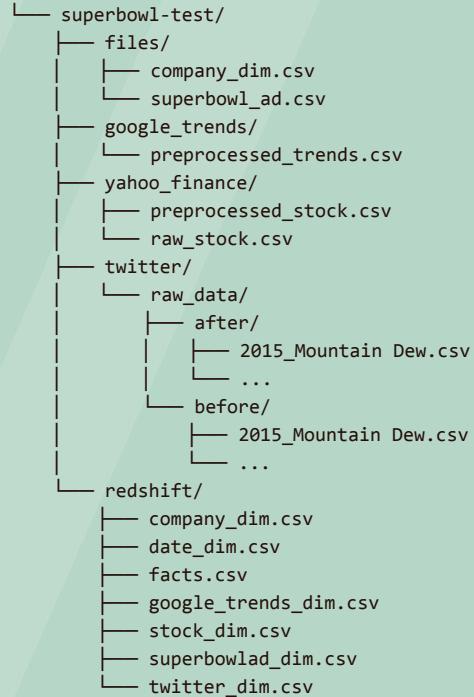


# AWS S3 File Structure

The screenshot shows the AWS S3 console interface. At the top, there's a navigation bar with 'Services' and a search bar. Below it, a sidebar lists services like S3, EC2, IAM, RDS, and Amazon Redshift. The main area is titled 'superbowl-test' and contains tabs for 'Objects', 'Properties', 'Permissions', 'Metrics', 'Management', and 'Access Points'. Under the 'Objects' tab, there's a heading 'Objects (5)'. A note says: 'Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)'.

Name	Type	Last modified	Size	Storage class
files/	Folder	-	-	-
google_trends/	Folder	-	-	-
redshift/	Folder	-	-	-
twitter/	Folder	-	-	-
yahoo_finance/	Folder	-	-	-

Screenshot of datasets within S3 bucket



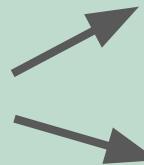
# Data Ingestion: Twitter

## Data Collection

Step 1



**Snsccape**  
a scraper for social networking services(SNS).



Brand related tweets 2 days before Super Bowl

Brand related tweets 2 days after Super Bowl

#brandbowl

Feb 13 Love hearing the extraordinary voice of Barbra Streisand on a Super Bowl commercial. Bravo Barbra and Bud Light! #SuperBowl @BarbraStreisand @Budweiser

Feb 13 Budweiser should have boycotted this #superbowl in solidarity with St.Louis

Feb 13 Companies that advertised in #SuperBowl I included Ford, Chrysler, RCA, R.J. Reynolds Tobacco, McDonald's, Budweiser and Goodyear. #sportsbiz

Feb 13 Bullfight Next went in with the production of that commercial! #Budweiser #SuperBowl

Super Bowl tweets about Budweiser ad

## Data Preprocess

Step 2

**Vader Sentiment Analysis** a model to detect the polarity of the tweets

3/0

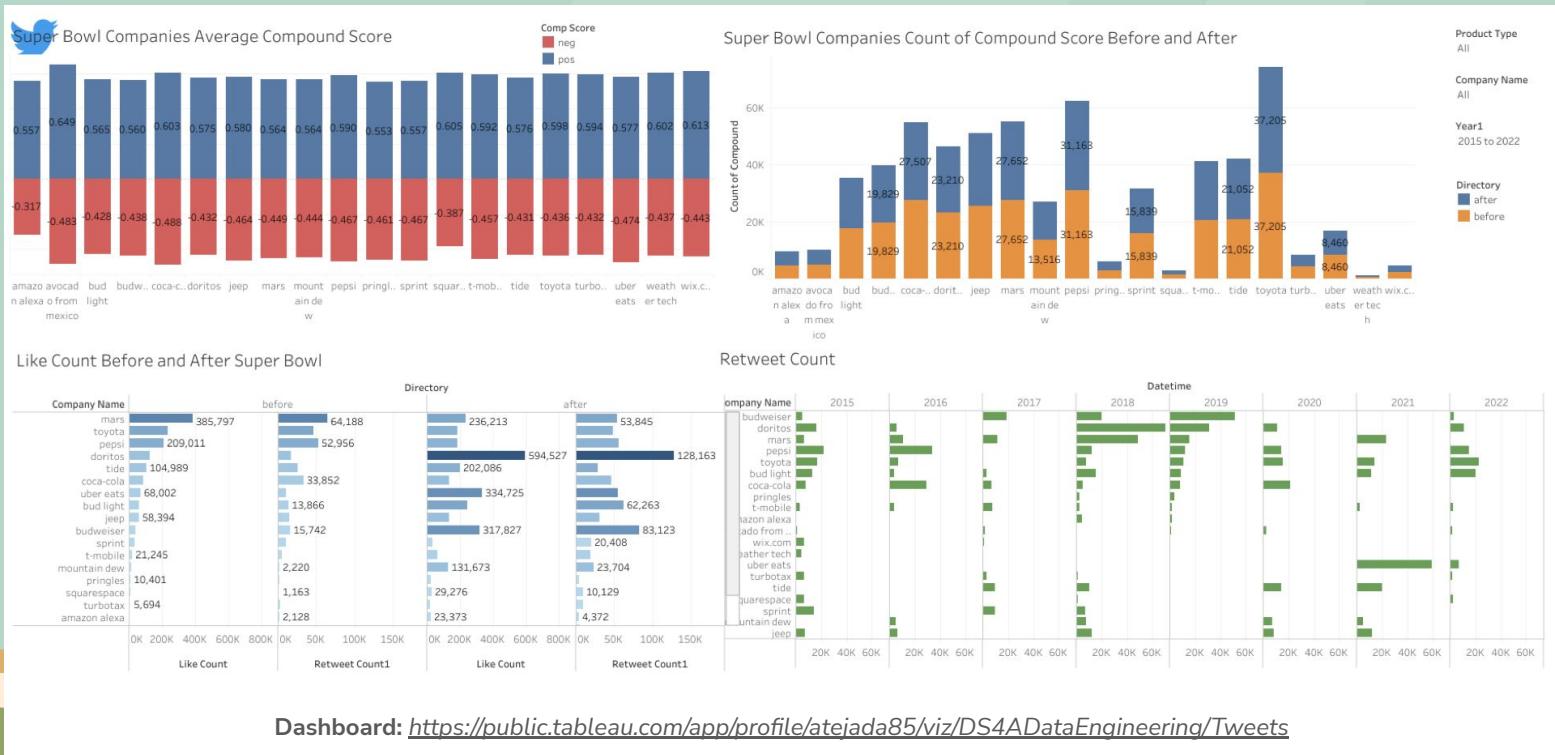
4/0

5/0

4/0

3/0

# Twitter



3/0

4/0

5/0

4/0

3/0

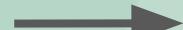
# Data Ingestion: Google Trends



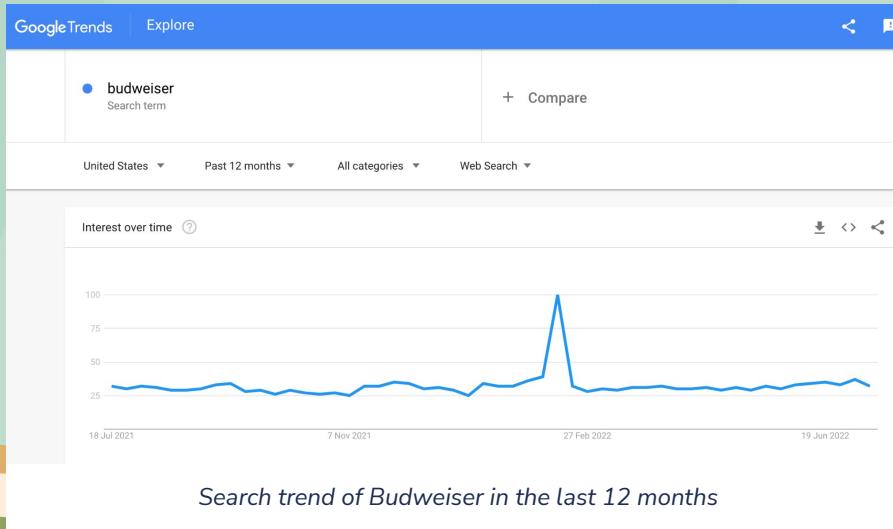
**Pytrends**

*API to automate downloading reports from Google Trends.*

an



**Steps:** List keyword of the company, the timeframe from January - March 2015 - 2022



# Google Trends



Month of Date Time  
All

Company Name  
All

Product Type  
All

## Google Trends

	amazon	alexa	avocado from m...	bud light	budweiser	coca-cola	doritos	jeep	mars	mountain dew	pepsi	pringles	sprint	squarespace	t-mobile	tide	toyota	turbotax	uber eats	weather tech	wix.com
2015	January	1.00	0.00	1.00	2.58	8.35	18.06	31.13	32.97	12.16	67.06	10.81	82.13	40.23	26.03	16.81	91.10	32.16	0.00	0.94	10.06
	February	1.11	0.04	1.11	3.43	9.43	19.32	34.82	42.86	20.93	69.11	11.29	85.79	52.64	24.39	14.64	92.21	57.21	0.18	1.07	10.18
	March	0.87	0.00	1.03	1.61	10.13	13.65	36.52	66.74	17.39	74.74	10.65	84.10	42.58	23.87	14.77	93.52	27.52	0.13	0.77	11.13
2016	January	3.42	0.00	0.77	1.16	7.61	5.19	34.81	22.13	9.61	51.81	3.48	74.39	38.06	17.65	15.94	87.84	28.90	3.26	0.87	6.23
	February	4.45	0.00	1.03	2.62	7.83	11.86	36.93	30.83	12.76	59.79	3.48	75.45	44.93	17.93	13.48	91.45	49.38	3.52	0.86	7.52
	March	5.19	0.00	0.94	1.03	7.74	5.39	36.71	51.19	10.61	56.32	3.52	75.61	40.16	17.71	13.39	90.65	18.77	4.68	0.61	7.23
2017	January	56.97	0.13	1.03	1.10	8.03	27.65	35.74	21.74	11.29	59.58	19.65	86.77	67.61	19.00	21.45	92.58	36.65	50.19	1.00	9.94
	February	34.25	0.11	0.89	2.64	8.32	32.54	35.96	30.11	11.36	66.39	19.00	85.93	72.82	18.04	18.57	90.11	59.54	55.86	1.00	10.68
	March	38.42	0.00	0.97	1.32	8.71	27.81	36.13	52.48	11.35	63.26	18.87	86.35	70.71	17.42	18.61	89.58	29.58	59.84	1.00	9.90
2018	January	44.32	0.00	1.06	1.58	6.90	11.32	37.35	26.81	9.52	53.90	8.10	83.81	39.26	17.29	45.87	88.55	23.74	65.90	0.90	6.71
	February	35.46	0.00	1.11	1.93	6.96	24.07	38.75	30.75	13.82	59.79	8.61	86.75	42.32	16.93	21.32	88.86	41.54	75.21	1.00	6.93
	March	33.97	0.00	1.00	1.42	7.68	12.48	38.97	54.19	9.10	58.19	7.74	86.06	37.68	16.48	15.84	90.35	20.39	81.16	0.90	6.81
2019	January	17.00	0.06	1.16	2.45	6.52	49.45	33.68	27.55	6.26	41.94	36.26	83.35	14.32	16.13	24.94	82.42	36.55	66.52	1.10	2.16
	February	11.43	0.11	1.93	3.04	10.89	56.18	35.21	36.79	7.14	49.54	36.29	84.25	14.89	16.21	22.00	83.50	64.89	70.46	1.25	2.11
	March	10.00	0.13	1.13	2.42	15.10	46.16	35.87	79.39	7.19	46.26	36.10	87.03	13.13	18.55	22.00	82.81	32.87	69.45	1.16	1.90
2020	January	12.71	0.00	1.81	2.29	8.16	13.71	35.84	23.19	9.55	47.06	10.26	57.00	11.52	13.65	22.77	89.26	37.42	60.16	1.06	1.81
	February	9.28	0.03	1.31	2.55	8.41	19.93	36.83	27.31	13.72	52.34	11.76	60.86	11.69	13.66	23.07	89.21	62.17	58.93	1.07	1.79
	March	7.87	0.00	1.03	1.81	8.13	14.84	28.45	68.00	8.68	42.29	10.03	52.77	11.94	13.97	19.68	69.35	32.58	68.00	0.94	1.68
2021	January	11.45	0.61	0.19	1.00	16.29	41.87	37.84	8.81	11.23	50.97	31.26	88.71	7.90	33.42	33.00	87.42	39.35	71.19	2.00	1.26
	February	9.36	0.21	0.21	1.11	17.04	46.68	38.54	23.75	13.00	54.14	32.36	86.54	7.93	30.86	30.18	86.75	67.96	71.86	2.11	1.29

## Trends Percent Change By Product



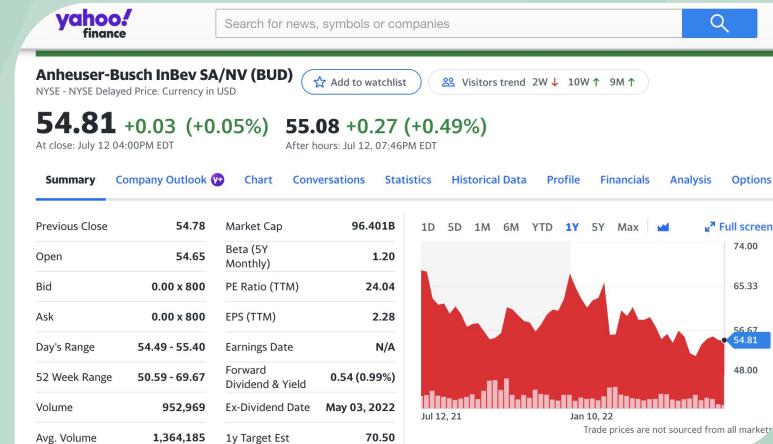
# Data Ingestion: Stock



yfinance

open-source tool that uses Yahoo's  
publicly available APIs to collect data  
from yahoo finance

Steps: List stock ticker names for each  
company, the data was pulled from 2015 -  
2022.



Stock data of Budweiser

# Stock



**Product Type**

Company Name  
All

Year of Date



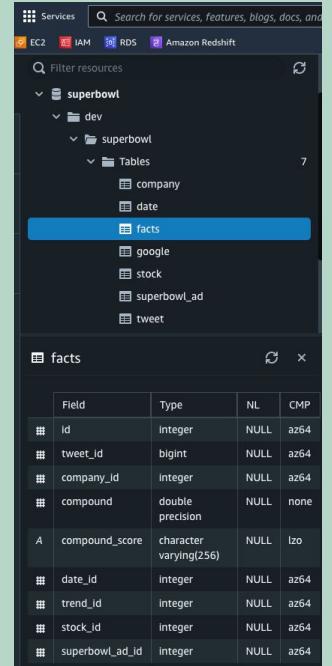
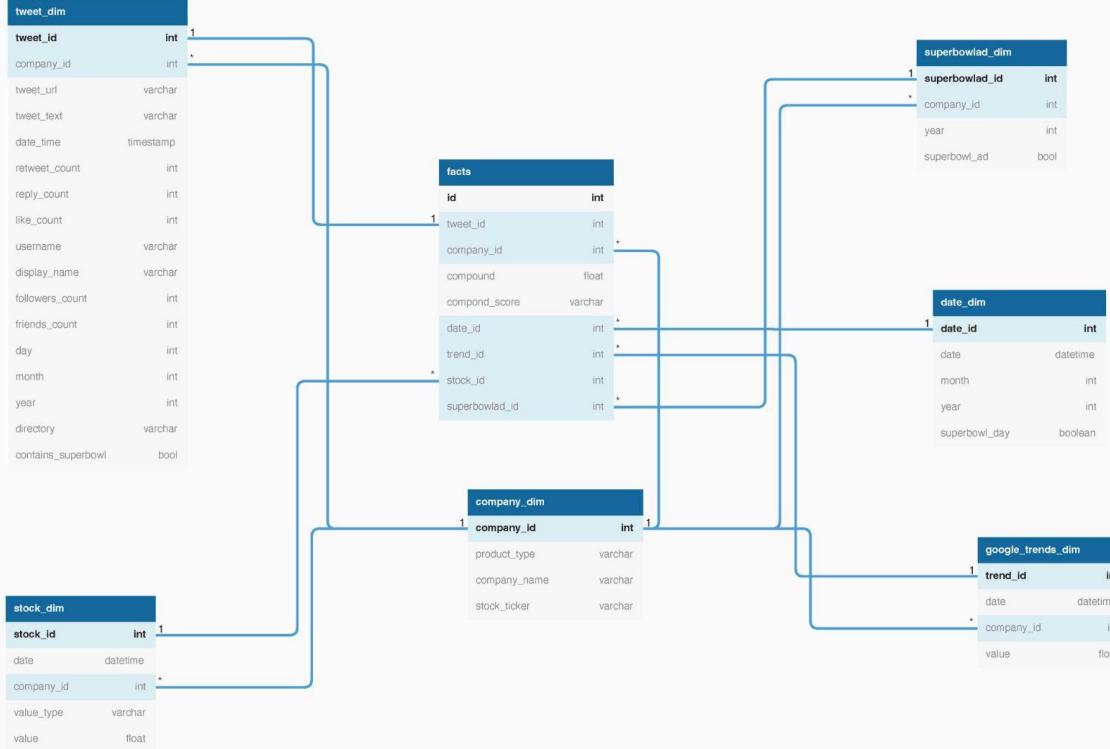
## Stocks Percent Change by Product



## Stocks

Dashboard: <https://public.tableau.com/app/profile/atejada85/viz/DS4ADataEngineering/Stocks>

# Data Schema



Screenshot of tables  
within Redshift

# Dashboard QR Code

Tableau



[https://public.tableau.com/app/profile/atejada85  
/viz/DS4ADataEngineering/SuperBowlImpact](https://public.tableau.com/app/profile/atejada85/viz/DS4ADataEngineering/SuperBowlImpact)

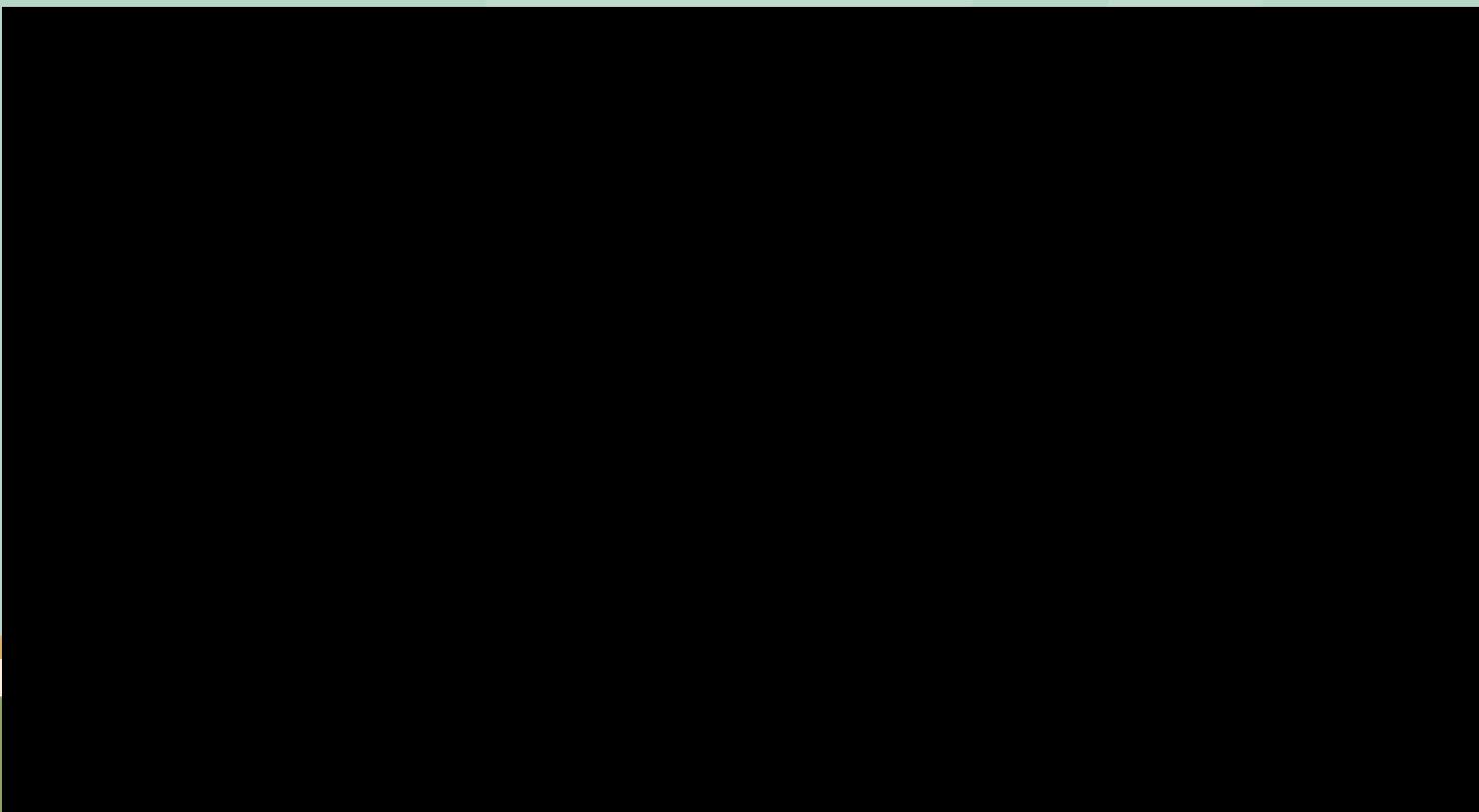
Data Studio



<https://datastudio.google.com/reporting/11fab1d4-cdf1-428c-bf25-0af3d69895a6>

(Data Studio dashboard might not work any more as we have deleted our Redshift account, see next slide to watch video of the Data Studio dashboard)

# Data Studio Dashboard



# Challenges



- Retrieving data from Twitter API took a long time
- Finding the appropriate variables to pull data for each company & stock
- Some companies don't have stock data as they are not public and other brands fall under same public company.
- Adapting the preprocessed data to fit into AWS Redshift.
- Figuring out how to connect time series data (stock and Google trends) with Twitter data.

# Improvements

1. Implement Pyspark for data processing when retrieving the data to reduce time & load.
2. Utilize AWS Glue to load data into Redshift for data integration to analyze the data quickly.
3. Create a Cron Job to pull data for future Super Bowl ads.

# DATAFOLIO



30

40

50

40

30



# OTHER LINKS



30

40

50

40

30

# Other Links

**Github:** [https://github.com/Pball01/superbowl\\_analysis](https://github.com/Pball01/superbowl_analysis)

**Tableau:** <https://public.tableau.com/app/profile/atejada85/viz/DS4ADataEngineering/Tweets>