Sample Big Query Experiment and result

> **Objective**
>
> Sentiment Analysis of a company or product in any location, or worldwide using GDELT and BigQuery.

Experiment

Overview

- **What type of experiment is this?**
  SQL query to extract Tone data for a given company.
- **SQL query:**
  The following query tries to extract all the V2Tone data for organization names like Abbott for dates > `2018-01-01`

```
SELECT
  EXTRACT (date
  FROM
      PARSE_TIMESTAMP('%Y%m%d%H%M%S',CAST(date AS string))) AS Date,
  CAST(SPLIT(V2Tone, ",") [
  OFFSET
      (0)] AS FLOAT64) AS tone,
  CAST(SPLIT(V2Tone, ",") [
  OFFSET
      (1)] AS FLOAT64) AS pos_score,
  CAST(SPLIT(V2Tone, ",") [
  OFFSET
      (2)] AS FLOAT64) AS neg_score,
  CAST(SPLIT(V2Tone, ",") [
  OFFSET
      (3)] AS FLOAT64) AS polarity,
  CAST(SPLIT(V2Tone, ",") [
  OFFSET
      (4)] AS FLOAT64) AS arf,
  CAST(SPLIT(V2Tone, ",") [
  OFFSET
      (5)] AS FLOAT64) AS sg_rf,
  CAST(SPLIT(V2Tone, ",") [
  OFFSET
      (6)] AS FLOAT64) AS wc,
  V2Organizations
  FROM
  `gdelt-bq.gdeltv2.gkg_partitioned`
  WHERE
  DATE(_PARTITIONTIME) >= "2018-01-01"
  AND lower(V2Organizations) LIKE '%abbott%'
```

Results

The query results are available in two formats: **CSV** and **JSON**

Date	tone	pos_score	neg_score	polarity	arf	sg_rf	wc	V2Organizations
16/06/2022	-0.1928640309	2.700096432	2.892960463	5.593056895	22.56509161	0.4821600771	959	Howard Coalition,3577;Renewable Energy,2927;Renewable Energy,3228;Renewable Energy,3296;Renewable Energy,4016;Renewable Energy,4148
16/06/2022	0	2.757158006	2.757158006	5.514316013	22.26935313	0.5302226935	870	Howard Coalition,3269;Abbott Coalition,3372
16/06/2022	-0.4926108374	5.418719212	5.911330049	11.33004926	25.12315271	0.9852216749	183	Abbott Lab,93
16/06/2022	-5.970149254	2.23880597	8.208955224	10.44776119	18.65671642	0	121	Abbott Laboratories,565;United States,706
16/06/2022	-1.510574018	1.208459215	2.719033233	3.927492447	21.45015106	0	310	Drug Administration,277;Abbott Laboratories,29;Reuters,8;United States,701
16/06/2022	-5	2.5	7.5	10	25	0	37	Abbott Laboratories,19
16/06/2022	-2.118644068	0.4237288136	2.542372881	2.966101695	12.28813559	0.4237288136	220	Walmart,153;Walmart,219;Walmart,357;Walmart,494;Walmart,667;Albertsons,959;Kroger,946;Abbott Laboratories,1179;Reuters,8;United States,
16/06/2022	-5.970149254	2.23880597	8.208955224	10.44776119	18.65671642	0	121	Abbott Laboratories,565;United States,706
16/06/2022	0	2.760084926	2.760084926	5.520169851	22.29299363	0.5307855626	870	Howard Coalition,3268;Abbott Coalition,3371
16/06/2022	-0.3629764065	2.722323049	3.085299456	5.807622505	27.04174229	0.5444646098	510	Veterans Affairs,3189;Drug Administration,2431;Abbott Laboratories,2185
16/06/2022	1.489361702	2.127659574	0.6382978723	2.765957447	22.44680851	0.4255319149	710	Glaxosmithkline,899;Research Methodology,4868;Research Methodology,5732;Lonza Group,870;t Telecom,5613;Algoro Research Consultants F
16/06/2022	0	2.760084926	2.760084926	5.520169851	22.29299363	0.5307855626	870	Howard Coalition,3268;Abbott Coalition,3371
16/06/2022	-1.317715959	1.02489019	2.342606149	3.36749634	21.37628111	1.317715959	621	Twitter,2249;Drug Administration,1611;Drug Administration,2184;Drug Administration,3233;Drug Administration,3463;Abbott Laboratories,23;Ab
16/06/2022	-0.7462686567	2.23880597	2.985074627	5.223880597	17.1641791	0	121	Drug Administration,272;Abbott Laboratories,29;Reuters,8
16/06/2022	0.8561643836	1.54109589	0.6849315068	2.226027397	24.65753425	1.54109589	530	Abbott Labs,330
16/06/2022	-6.25	0	6.25	6.25	31.25	0	29	Abbott Laboratories,19
16/06/2022	-2.118644068	0.4237288136	2.542372881	2.966101695	12.28813559	0.4237288136	220	Walmart,153;Walmart,219;Walmart,357;Walmart,494;Walmart,667;Albertsons,959;Kroger,946;Abbott Laboratories,1179;Reuters,8;United States,
16/06/2022	-5.529953917	1.843317972	7.373271889	9.216589862	20.2764977	0	197	Abbott Laboratories,235;United States,370
16/06/2022	0.1028806584	2.777777778	2.674897119	5.452674897	22.42798354	0.5144032922	894	Howard Coalition,3287;Abbott Coalition,3390
16/06/2022	-1.317715959	1.02489019	2.342606149	3.36749634	21.37628111	1.317715959	621	Twitter,2249;Drug Administration,1611;Drug Administration,2184;Drug Administration,3233;Drug Administration,3463;Abbott Laboratories,23;Ab

Feature Highlights

- This is a public request to the Gdelt Project through BigQuery. It is a simple search looking for a company called *abbott*.

  Tone= average tone [-100,+100]. common values between -10 and 10.
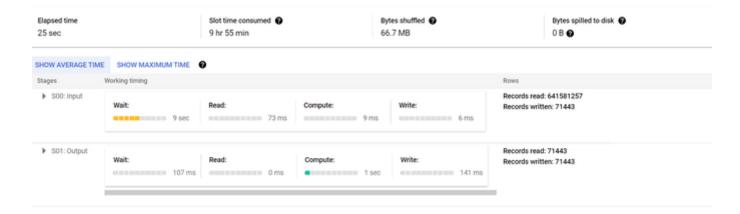
  SGRF = self/group reference density. % of all words in the article that are pronouns, capturing a combination of self-references and group-based discourse.

  ARF = activity reference density. % of words that were active words.

  polarity = % of words that had matches in the tonal dictionary as an indicator of how emotionally polarized or charged the text is. If Polarity is high but Tone neutral, this suggests the text was highly emotionally charged but had roughly equivalent numbers of positively and negatively charged emotional words

👾 Job Details

Job ID	eco-tape-347918:US.bquxjob_708faed0_1816bfc0660
User	rohit@visionbox.ai
Location	US
Creation time	16 Jun 2022, 15:38:09 UTC+5:30
Start time	16 Jun 2022, 15:38:10 UTC+5:30
End time	16 Jun 2022, 15:38:36 UTC+5:30
Duration	25 sec
Bytes processed	105.25 GB
Bytes billed	105.25 GB
Job priority	INTERACTIVE
Use legacy SQL	false
Destination table	Temporary table

🕵 Execution Details

| Elapsed time | Slot time consumed ❓ | Bytes shuffled ❓ | Bytes spilled to disk ❓ |
| 25 sec | 9 hr 55 min | 66.7 MB | 0 B ❓ |

SHOW AVERAGE TIME   SHOW MAXIMUM TIME ❓

Stages	Working timing				Rows
▶ S00: Input	Wait: ▪▪▪▪▪▫▫▫▫ 9 sec	Read: ▫▫▫▫▫▫▫▫▫▫ 73 ms	Compute: ▫▫▫▫▫▫▫▫▫▫ 9 ms	Write: ▫▫▫▫▫▫▫▫▫▫ 6 ms	Records read: 641581257 Records written: 71443
▶ S01: Output	Wait: ▫▫▫▫▫▫▫▫▫▫ 107 ms	Read: ▫▫▫▫▫▫▫▫▫▫ 0 ms	Compute: ▪▫▫▫▫▫▫▫▫▫ 1 sec	Write: ▫▫▫▫▫▫▫▫▫▫ 141 ms	Records read: 71443 Records written: 71443

✨ Conclusions

Follow-up

- We can query and obtain results in two formats(JSON and CSV)
- We can then write a script to generate BSS using the tone data for a date range (obtained from step 1)
- To-Do: Have to create a flow for the above scenario.