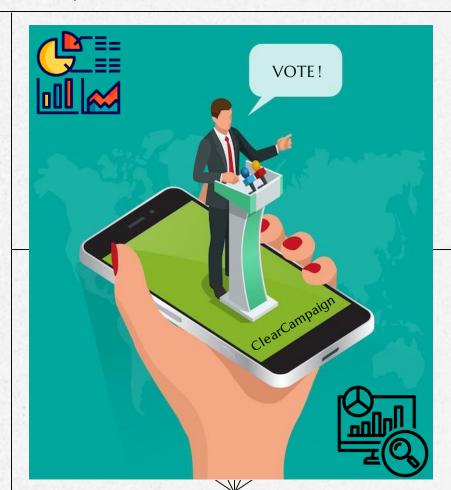
Data Layer Analysis of "ClearCampaign" An App For Transparent Political Ad Data

By Group 3 Ravi Teja Gattu - 017503746 Poorna Ravuri - 017510883 Jacob Culaton - 014706172



Statement Of The Problem

Can You Access Political Ad Campaign Data? **DEPENDS. NOT AN EMPHATIC YES!**

WHERE?





& AdImpact Roku Advertising

Easily Accessible?

Data Quality?

All Information At One Place?

Customizations?

Last Updated?

Mode Of Access?

INSUFFICIENT!

SOLUTION: ClearCampaign

ClearCampaign



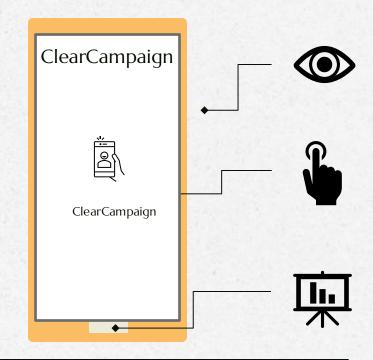
Vision: to create transparency in the arena of politics for a better world and a better future

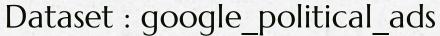


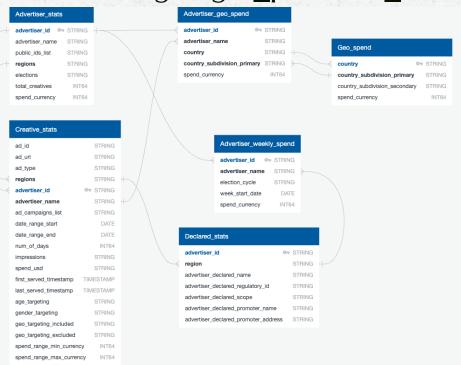
Value : Political Ad Campaign Data at user's fingertip



Versatility: Wide range of attributes to filter from, due to vast data stored.







Feature 1: All Time Top 10 Political Ad Campaigns of USA by Spending

```
SFI FCT
WITH FilteredCreativeStats AS (
                                                                  a.regions,
SFI FCT
                                                                  c.year,
advertiser id.
                                                                  a.advertiser name.
EXTRACT (YEAR FROM last served timestamp) AS year.
                                                                  SUM(c.spend_range_max_usd) AS Spending,
spend_range_max_usd,
                                                                  c.impressions.
impressions.
                                                                  c.ad_type,
ad_type,
                                                                  c.age_targeting,
age targeting.
                                                                  c.gender_targeting,
gender targeting.
                                                                  c.geo targeting included
geo_targeting_included,
                                                                  FROM
EXTRACT (YEAR FROM date_range_start) AS start_year
                                                                  FilteredCreativeStats c
FROM
                                                                  JOTN
`bigguery-public-data.google_political_ads.creative_stats`
                                                                  FilteredAdvertiserStats a ON c.advertiser_id = a.advertiser_id
WHERE
                                                                  GROUP BY
age_targeting IS NOT NULL AND
                                                                  a.regions.
EXTRACT (YEAR FROM date_range_start) = 2024
                                                                  c.vear.
                                                                  a.advertiser name.
FilteredAdvertiserStats AS (
                                                                  c.ad type.
SELECT
                                                                  c.impressions,
advertiser_id.
                                                                  c.age_targeting,
regions.
                                                                  c.gender_targeting,
advertiser_name,
                                                                  c.geo_targeting_included
spend_usd
                                                                  ORDER BY
FROM
                                                                  Spending DESC
`bigguery-public-data.google_political_ads.advertiser_stats`
                                                                  LIMIT
WHERE
                                                                  10;
regions = 'US' ANDspend_usd <> 0)
```

Feature 2: Weekly Political Ad Spends in California, 2024

```
WITH FilteredGeoSpend AS (
SELECT
advertiser_id,
country_subdivision_primary
FROM
`bigquery-public-
data.google_political_ads.advertise
r_geo_spend`
WHERE
country = 'US'
AND
country_subdivision_primary = 'CA'
)
```

```
SELECT
g.country_subdivision_primary,
w.week_start_date,
w.advertiser_name,
SUM(w.spend_usd) AS total_weekly_spend_usd
FROM
`bigquery-public-
data.google_political_ads.advertiser_weekl
y_spend` AS w
JOIN
FilteredGeoSpend AS g
ON
w.advertiser_id = g.advertiser_id
GROUP BY
1, 2, 3
ORDER BY
1, 2 DESC LIMIT 100;
```

Feature 3 : Demographics Targetted by Top 10 US Political Ad Spends & It's Impact in 2024

```
WITH FilteredCreativeStats AS (
                                                                     SELECT
SELECT
advertiser_id,
                                                                    a.regions,
EXTRACT(YEAR FROM last_served_timestamp) AS year,
                                                                     c.year,
spend_range_max_usd,
                                                                     a.advertiser_name,
impressions,
                                                                     SUM(c.spend_range_max_usd) AS Spending,
ad_type.
                                                                    c.impressions.
age_targeting
                                                                     c.ad_type,
gender_targeting,
                                                                     c.age_targeting,
geo_targeting_included,
EXTRACT(YEAR FROM date_range_start) AS start_year
                                                                    c.gender targeting
FROM
                                                                     c.geo_targeting_included
`bigquery-public-data.google_political_ads.creative_stats`
                                                                     FROM
WHERE
                                                                    FilteredCreativeStats c
age_targeting IS NOT NULL AND
                                                                     JOTN
EXTRACT(YEAR FROM date_range_start) = 2024
                                                                     FilteredAdvertiserStats a ON c.advertiser id = a.advertiser id
                                                                     GROUP BY
FilteredAdvertiserStats AS (
SELECT
                                                                     a.regions,
advertiser_id,
                                                                     c.year,
regions,
                                                                    a.advertiser name.
advertiser name.
                                                                     c.ad_type,
spend_usd
                                                                     c.impressions.
FROM
                                                                     c.age_targeting,
`bigguery-public-data.google_political_ads.advertiser_stats
                                                                     c.gender_targeting,
WHFRF
                                                                     c.geo_targeting_included
regions = 'US' AND
spend usd <> 0
                                                                     ORDER BY
                                                                     Spending DESC LIMIT 100;
```

Feature 4: State Wise Total Political Ad Spending in US

```
SELECT
  a.advertiser_id,
  a.advertiser_name,
  g.country_subdivision_primary,
  SUM(g.spend_usd) AS total_spend_usd
FROM
  `bigguery-public-
data.google_political_ads.advertiser_stats` AS
JOTN
  `bigguery-public-
data.google_political_ads.advertiser_geo_spend`
AS g
ON
  a.advertiser_id = g.advertiser_id
WHERE
  g.country = 'US'
GROUP BY
  a.advertiser_id,
  a.advertiser_name,
  g.country_subdivision_primary
ORDER BY
  total_spend_usd DESC LIMIT
  100:
```

Row	advertiser_id ▼	advertiser_name ▼	country_subdivision_primary ▼	total_spend_usd 🔻
1	AR07182956219827486721	WARNOCK FOR GEORGIA	GA	14346100
2	AR12365610929977556993	BIDEN FOR PRESIDENT	FL	13752300
3	AR12365610929977556993	BIDEN FOR PRESIDENT	PA	13557900
4	AR11825178974693097473	DONALD J. TRUMP FOR PRESI	FL	11550000
5	AR00764926253714112513	Yes on 27 - Californians for Solutions to Homelessness and Mental Health Support, a Coalition of Housing and Mental Health Experts,	CA	10247100

Other Features:

- Impressions of Ads
- Country Wise Political Ad Spending
 - US Ad Spending By Gender
- Weekly Total Political Ad Spending

DEMO

Before Optimization

After Optimization

Elapsed time	Slot time consumed Q	Bytes shuffled 🛭	Bytes spilled to disk. Q	Elapsed time	Slot time consumed 0	Bytes shuffled Q	Bytes spilled to disk ②
1 sec	5 980	28.38 MB	080	1 sec	4 sec	28.38 MB	080

Bytes shuffled: BigQuery shuffles data in order to support large-scale computation.

Slot time: A slot is a unit of computational capacity required to execute SQL

queries.

Before optimization:

Bytes shuffled/ slot time = 28.83MB/5sec =

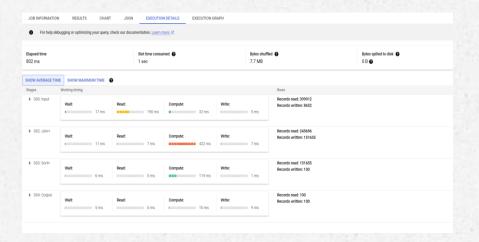
5.6MB/sec

After optimization:

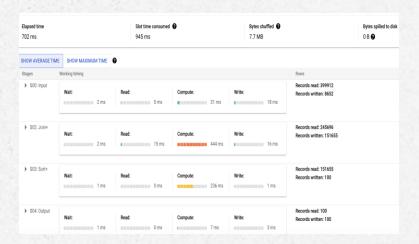
Bytes shuffled/ slot time = 28.83MB/4sec =

7.2MB/sec

Before Optimization

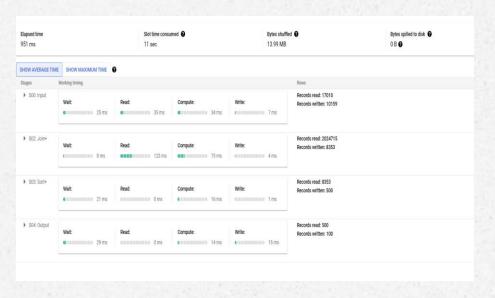


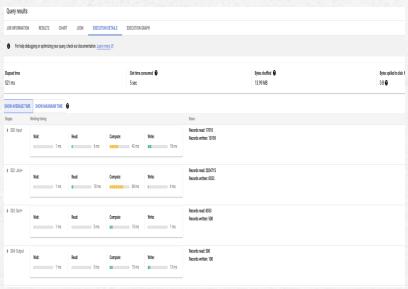
After Optimization



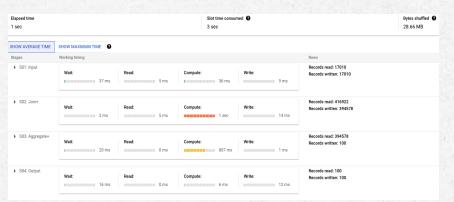
Before Optimization

After Optimization





Before Optimization

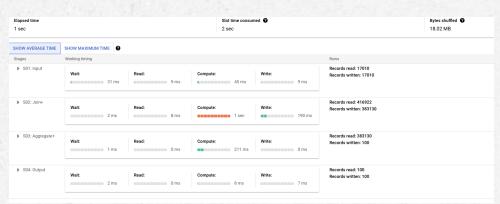


Read cost per GB: 0.01\$
Write cost per GB: 0.02\$
Total average size of each record is 100bytes

Read cost before optimization =
(17061+417711+395205)*100*0.01 = 0.0082\$
Write cost before optimization =
(17061+395205+395205)*100*0.02 = 0.016\$

Total VO cost before optimization = (Read cost + write cost) = (0.0082+0.016) = 0.0242\$

After Optimization



Read cost after optimization = (17061+417711+383725)*100*0.01 = 0.0081\$

Write cost after optimization =

(17061+383725+383725)*100*0.02 = 0.0156\$

Total VO cost after optimization = (Read cost + write cost) = (0.0081+0.0156) = 0.0237\$

Elapsed time Slot time consumed ② Bytes shuffled @ 28.66 MB 1 sec SHOW AVERAGE TIME SHOW MAXIMUM TIME Stages Working timing Rows S01: Input Records read: 17010 Wait: Read: Compute: Write: Records written: 17010 37 ms 5 ms 36 ms 9 ms Before Optimization ▶ S02: Join+ Records read: 416922 Wait: Read: Compute: Write: Records written: 394578 2 ms 5 ms 14 ms ▶ S03: Aggregate+ Records read: 394578 Compute: Records written: 100 20 ms 0 ms 807 ms 1 ms S04: Output Records read: 100 Wait: Read: Compute: Write: Records written: 100 16 ms 0 ms 6 ms 12 ms Elapsed time Slot time consumed @ Bytes shuffled @ 1 sec 2 sec 18.02 MB SHOW AVERAGE TIME SHOW MAXIMUM TIME Stages Working timing After Optimization ▶ S01: Input Records read: 17010 Read: Compute: Write: Records written: 17010 31 ms 9 ms 9 ms Records read: 416922 S02: Join+ Read: Compute: Records written: 383130 2 ms 8 ms 1 sec 190 ms ▶ S03: Aggregate+ Records read: 383130 Compute: Records written: 100 1 ms 0 ms 211 ms 0 ms S04: Output Records read: 100 Wait: Read: Compute: Write: Records written: 100

0 ms

2 ms

7 ms

8 ms

Results & Experience

Conclusion



BigQuery Data Accuracy & Insights

- + Data Variety
- + Accessibility
- Deprecated Tables Limiting Scope
- Index creation hindered by permissions



Learning

- + Optimization
- + SQL Exploration : Views, CTE, Indexes
- + Team Collaboration



Future Improvements & Scope

- + More Global Data
- + Revival of Deprecated Tables
 - -> More Features of based application
- + Prediction based on historical data using AI

Thank You