**Problem Statement**

Online advertising platforms face a critical challenge in monitoring and enforcing ad integrity. Some ads are removed shortly after being displayed, often due to policy violations, misinformation, or deceptive content. This rapid removal not only signals harmful content slipping through but also impacts platform trust, advertiser costs, and user experience. The challenge we address is: Can we leverage BigQuery AI to detect and forecast ad removal patterns across regions and topics?

**Impact Statement**

Our solution applies BigQuery AI to the Google Ads Transparency Center dataset to analyze historical ad lifespans, identify removal indicators, and forecast the likelihood of future removals. By detecting “quick removals” (ads pulled within 7–30 days), our models surface patterns across regions, topics, and advertiser verification statuses.

The material impact is threefold:

* Policy Enforcement Efficiency – Regulators and platforms can prioritize high-risk ad categories for earlier review.
* Trust & Transparency – Stakeholders gain interpretable metrics (removal rates, topic/region risk profiles).
* Predictive Forecasting – Using BigQuery ML, we forecast which ad clusters are most likely to be removed soon, helping platforms intervene proactively.

**Code**: [here](https://github.com/pushyamiparitala/Bigquery-GoogleAds)This improves content moderation, reduces harmful exposure, and contributes to safer digital advertising ecosystems.

**1. Patterns Across Regions**

Data Extraction  
We first extract political and government-related ads:

simple\_extraction\_query = f"""

SELECT

creative\_id,

advertiser\_id,

advertiser\_legal\_name,

advertiser\_location,

advertiser\_verification\_status,

topic,

ad\_format\_type,

ad\_funded\_by,

is\_funded\_by\_google\_ad\_grants,

region\_stats

FROM `bigquery-public-data.google\_ads\_transparency\_center.creative\_stats`

WHERE topic IN ('Law & Government', 'Jobs & Education', 'News, Books & Publications')

OR REGEXP\_CONTAINS(UPPER(COALESCE(advertiser\_legal\_name, '')), r'POLITICAL|CAMPAIGN|ELECTION|GOVERNMENT')

OR ad\_funded\_by IS NOT NULL

LIMIT 10000

"""

df = pandas\_gbq.read\_gbq(simple\_extraction\_query, project\_id=proj\_id)

**Cross-Region Campaign Analysis**

We designed a function to detect simultaneous launches:

def analyze\_cross\_region\_patterns(df):

patterns = []

for creative\_id, group in df.groupby("creative\_id"):

advertiser\_name = group["advertiser\_legal\_name"].iloc[0]

regions = group["region\_code"].tolist()

start\_dates = group["first\_shown\_date"].dropna().tolist()

if len(start\_dates) > 1:

date\_spread = max(start\_dates) - min(start\_dates)

simultaneous\_launch = date\_spread.days <= 7

else:

simultaneous\_launch = False

patterns.append({

"creative\_id": creative\_id,

"advertiser\_name": advertiser\_name,

"regions": regions,

"region\_count": len(set(regions)),

"verification\_status": group["advertiser\_verification\_status"].iloc[0],

"simultaneous\_launch": simultaneous\_launch,

})

return pd.DataFrame(patterns)

We flagged ads with ≥5 regions, simultaneous launch, and unverifiable advertisers:

suspicious\_patterns = region\_patterns[

(region\_patterns['region\_count'] >= 5) &

(region\_patterns['simultaneous\_launch'] == True) &

(region\_patterns['verification\_status'] != 'VERIFIED')

]

These results were written back into BigQuery:

region\_patterns.to\_gbq(

'ads\_ml.cross\_region\_patterns',

project\_id='project\_123,

if\_exists='replace'

)

**2. Ads Removal Forecast**

Ad Removal Pattern Extraction

\*\*removal\_patterns\_query = """

WITH unnested\_data AS (

SELECT

advertiser\_id,

creative\_id,

ad\_format\_type,

advertiser\_verification\_status,

topic,

region.region\_code,

PARSE\_DATE('%Y-%m-%d', region.first\_shown) as first\_shown\_date,

PARSE\_DATE('%Y-%m-%d', region.last\_shown) as last\_shown\_date,

region.times\_shown\_lower\_bound,

region.times\_shown\_upper\_bound

FROM `bigquery-public-data.google\_ads\_transparency\_center.creative\_stats`,

UNNEST(region\_stats) AS region

WHERE region.first\_shown IS NOT NULL

AND region.last\_shown IS NOT NULL

AND PARSE\_DATE('%Y-%m-%d', region.first\_shown) >= DATE\_SUB(CURRENT\_DATE(), INTERVAL 365 DAY)

)

SELECT

first\_shown\_date,

last\_shown\_date,

topic,

ad\_format\_type,

advertiser\_verification\_status,

region\_code,

COUNT(\*) as total\_ads,

COUNTIF(DATE\_DIFF(last\_shown\_date, first\_shown\_date, DAY) <= 30) as removed\_within\_30\_days,

COUNTIF(DATE\_DIFF(last\_shown\_date, first\_shown\_date, DAY) <= 7) as removed\_within\_7\_days,

SAFE\_DIVIDE(

COUNTIF(DATE\_DIFF(last\_shown\_date, first\_shown\_date, DAY) <= 30),

COUNT(\*)

) \* 100 as removal\_rate\_pct,

AVG(times\_shown\_upper\_bound) as avg\_impressions

FROM unnested\_data

GROUP BY 1,2,3,4,5,6

HAVING total\_ads >= 5

ORDER BY removal\_rate\_pct DESC;

"""

removal\_patterns\_df = pandas\_gbq.read\_gbq(removal\_patterns\_query, project\_id=project\_123)

\*\*

This query computes:

* Total ads
* Quick removals (≤7 days) and early removals (≤30 days)
* Removal rate percentage
* Average impressions before removal

Forecasting with BigQuery ML

CREATE OR REPLACE MODEL `project\_123.ads\_ml.quick\_removal\_forecast`

OPTIONS(

model\_type = 'ARIMA',

time\_series\_timestamp\_col = 'first\_shown\_date',

time\_series\_data\_col = 'removal\_rate\_pct',

time\_series\_id\_col = 'region\_code'

) AS

SELECT

region\_code,

first\_shown\_date,

removal\_rate\_pct

FROM `project\_123.ads\_ml.removal\_patterns`;

We then forecast 30 days ahead:

SELECT \*

FROM ML.FORECAST(

MODEL `project\_123.ads\_ml.quick\_removal\_forecast`,

STRUCT(30 AS horizon)

);

**🗂 Patterns Across Regions**Dataset Selection

Dataset Used: bigquery-public-data.google\_ads\_transparency\_center.creative\_stats

This dataset contains ad creatives, advertiser details, and region-level statistics (e.g., impressions, first/last shown dates). We restricted our queries to political/government-related ads by filtering topic, advertiser\_legal\_name, and ad\_funded\_by.

Constraints Applied

* Time Constraint: Ads shown in the past 18 months.
* Verification Constraint: Only consider ads where advertiser\_legal\_name is not null.
* Topic Constraint: Focused on Law & Government, News & Politics, and Political Organizations.

Suspicious Pattern Rules:

* Ads shown in ≥5 regions.
* Simultaneous launches across regions (≤7 day spread).
* Advertiser not verified.

Input to the Analysis

Creative-level Data: creative\_id, advertiser\_id, advertiser\_legal\_name, verification\_status.

Region-level Data: region\_code, first\_shown\_date, last\_shown\_date, impressions.

Loading the Dataset

Authentication: Used google.cloud.bigquery with a service account key.

Query Execution:

df = pandas\_gbq.read\_gbq(political\_extraction\_query, project\_id="your\_project\_id")

Processing: Extracted into a pandas DataFrame (df), then grouped by creative\_id for cross-region pattern analysis.

Output: Results stored back into BigQuery (your\_project\_id.your\_dataset.cross\_region\_patterns).

**📊 Ads Removal Forecast**Dataset Selection

Dataset Used: bigquery-public-data.google\_ads\_transparency\_center.creative\_stats

Focused on all ad creatives (not just political) to analyze removal patterns across regions and topics.

Constraints Applied

* Time Constraint: Ads shown in the past 365 days.
* Statistical Significance Constraint: Only include groups with ≥5 ads (filters out noise).

Removal Definition:

* Quick Removal: Ads stopped running within 7 days.
* Early Removal: Ads stopped running within 30 days.

Metrics Computed:

* Total ads per group.
* Removal rates (%) by topic, ad format, and verification status.
* Average impressions before removal.

Input to the Analysis

* Creative-level Data: creative\_id, advertiser\_id, advertiser\_verification\_status.
* Region-level Data: region\_code, first\_shown\_date, last\_shown\_date, impressions.

Loading the Dataset

Authentication: Used pandas\_gbq to connect to BigQuery.

Query Execution:

removal\_patterns\_df = pandas\_gbq.read\_gbq(removal\_patterns\_query, project\_id="your\_project\_id")

Processing: Query results loaded into a pandas DataFrame (removal\_patterns\_df).

Forecasting:

* Built ARIMA models with BigQuery ML (ML.FORECAST).
* Forecast horizon = 30 days ahead.

Output: Predicted removal rates saved to BigQuery for visualization.

ARIMA stands for AutoRegressive Integrated Moving Average.

It’s a statistical model for time series forecasting that combines three ideas:  
AR (AutoRegressive): The current value depends on its past values (lags).  
Example: Today’s ad removal rate is influenced by yesterday’s and last week’s rates.

I (Integrated): Differencing is applied to make the series stationary (removes trends).  
Example: Instead of modeling raw removal counts, we model the change in counts between days.

MA (Moving Average): The model also depends on past forecast errors (noise).  
Example: If our forecast was too high yesterday, we adjust today’s prediction downward.

**Why ARIMA for Ads Removal Forecast?**

* Sequential data: Ad removals are logged day by day (time series).
* Trends & seasonality: Ads may have weekly patterns or seasonal spikes (e.g., elections).
* Short-term forecasting: ARIMA is effective for predicting the next 7–30 days.

In this project, we combined the power of Google BigQuery AI with large-scale data from the Google Ads Transparency Center to address two pressing challenges in digital advertising integrity:

Ads Removal Forecast

* By analyzing ad lifespans, we identified “quick removals” (ads taken down within 7–30 days).
* Using BigQuery ML’s ARIMA models, we forecasted removal rates across regions and topics, enabling early detection of potentially harmful or non-compliant campaigns.

Patterns Across Regions

* We examined political and government-related ads to detect cross-region coordination.
* Campaigns that launched simultaneously across multiple regions with unverifiable advertisers were flagged as suspicious patterns.

🌍 Real-World Impact

This solution provides a data-driven framework for trust and safety in online ads:

* Platforms can prioritize policy enforcement.
* Regulators and journalists can investigate suspicious coordination.
* Users benefit from greater transparency in digital advertising ecosystems.

Team Members:

1. Harshitha Dwaracherla
2. Pushyami Rajeswari Paritala