ADTECH ANALYTICS PROJECT (SQL + PYTHON + POWER BI)

📌 PROJECT OVERVIEW

This AdTech analytics project delivers an end-to-end data pipeline and BI solution covering automated data ingestion, SQL-based data cleaning, modeling, DAX measures, and visually rich Power BI dashboards for media-performance insights.

📊 DATASET OVERVIEW

- 120,000 ad-event records (Impressions, Clicks, Conversions)
- 50+ fields covering campaign, creative, publisher, geography, device, and timestamps
- Contains cost, revenue, fraud flags, viewability, and engagement metrics
- Multiple inconsistent advertiser/publisher names
- Timestamp and URL cleaning required

🏗️ PROJECT ARCHITECTURE

Python Automation (Load CSV → SQL Server)
↓
SQL Server (Staging & Cleaning using SQL Views)
↓
Power BI (Modeling, DAX, EDA, Time Intelligence & Dashboard)

🐍 PYTHON AUTOMATION – CSV TO SQL

Automated ingestion loads all CSV files into SQL Server tables using SQLAlchemy and Windows Authentication.

```
from sqlalchemy import create_engine

import pandas as pd
```

```
import os


# ----------------------------------------

# Database Connection Settings (Windows Auth)

# ----------------------------------------

DB_SERVER = "DESKTOP-EN2BGT5\SQLEXPRESS"  # or "localhost" for default instance

DB_NAME = "Adtech"                # your database name


# Create SQLAlchemy engine using trusted connection

engine = create_engine(


f"mssql+pyodbc://@{DB_SERVER}/{DB_NAME}?driver=ODBC+Driver+17+for+SQL+Server&trusted_co
nnection=yes"

)


# ----------------------------------------

# Folder containing CSV files

# ----------------------------------------

DATA_FOLDER = r"C:\Users\dell\Pictures\Adtech_project"  # change to your folder path


# Loop through all CSV files and upload

for file in os.listdir(DATA_FOLDER):

    if file.lower().endswith(".csv"):

        file_path = os.path.join(DATA_FOLDER, file)

        table_name = os.path.splitext(file)[0]  # remove '.csv'

        print(f"📤 Uploading {file} → Table: {table_name}")

        df = pd.read_csv(file_path, encoding="utf-8")

        df.to_sql(table_name, con=engine, if_exists="replace", index=False)

        print(f"✅ Uploaded {file} ({len(df)} rows)\n")
```

```
print("🎉 All CSV files uploaded successfully!")
```

🧹 SQL DATA CLEANING (VIEW LOGIC)

A SQL View (vw_Adtech_Cleaned) standardizes advertiser/publisher names, removes special characters, fixes timestamps, flags duplicates, normalizes device & geo fields, and prepares metrics for Power BI modeling.

********** **VIEW OF CLEANED DATA** **********

```
USE Adtech;

GO


CREATE VIEW vw_Adtech_Cleaned AS

SELECT

  ImpressionID,


  -- 🆕 Duplicate flag logic using COUNT() OVER()

  CASE

    WHEN COUNT(ImpressionID) OVER (PARTITION BY ImpressionID) > 1 THEN 'Duplicate'

    ELSE 'Unique'

  END AS Clean_ImpressionID,


  -- Clean advertiser and publisher

  CASE

    WHEN AdvertiserName IN ('ACME_CORP','ACME COIP','ACMETCORP','ACME_RORP','CCME CORP','CCME CORP LTD','ADME CORP','ACMK CORP') THEN 'ACME CORP'

    WHEN AdvertiserName IN ('BLUESK VTD','BLUESKY_LJD') THEN 'BLUESKY'

    WHEN AdvertiserName IN ('MARKETIFYH','MARKETIFYT','MARKETIFYY','MARKETIFYS','MARKETIFYZ') THEN 'MARKETIFY'

    ELSE REPLACE(REPLACE(UPPER(TRIM(REPLACE(TRANSLATE(AdvertiserName, CHAR(9)+CHAR(10)+CHAR(13), '   '), CHAR(160), ''))),'_LTD',''),'LTD','')

  END AS Clean_AdvertiserName,
```

```
  REPLACE(UPPER(TRIM(REPLACE(TRANSLATE(PublisherName, CHAR(9)+CHAR(10)+CHAR(13), '   '),
CHAR(160), ''))),'!',''') AS Clean_PublisherName,


  -- Campaign

  REPLACE(UPPER(TRIM(REPLACE(TRANSLATE(CampaignName, CHAR(9)+CHAR(10)+CHAR(13), '   '),
CHAR(160), ''))),'!',''') AS Clean_CampaignName,


  -- Ad details

  UPPER(TRIM(AdFormat)) AS AdFormat,

  UPPER(TRIM(CreativeID)) AS CreativeID,

  UPPER(TRIM(CreativeSize)) AS CreativeSize,


  -- Device & Geo

  CASE

    WHEN DeviceType IN ('mobile','Mobile','MOBILE') THEN 'Mobile'

    WHEN DeviceType IN ('desktop','Desktop','DESKTOP') THEN 'Desktop'

    ELSE 'Other'

  END AS DeviceType,

  REPLACE(TRIM(UPPER(COALESCE(Browser,'NA'))),'CHR0ME','CHROME') AS Browser,

  TRIM(UPPER(COALESCE(Country,'NA'))) AS Country,

  UPPER(TRIM(City)) AS City,

  COALESCE(GeoLat,0) AS GeoLat,

  COALESCE(GeoLong,0) AS GeoLong,


  -- Time details

  EventTime_Clean,

  CAST(EventTime_Clean AS DATE) AS EventDate,

  YEAR(EventTime_Clean) AS EventYear,

  MONTH(EventTime_Clean) AS EventMonth,
```

```
  DATEPART(WEEK, EventTime_Clean) AS EventWeek,
  DATEPART(HOUR, EventTime_Clean) AS EventHour,

  -- Numeric metrics
  Revenue_Num,
  Cost_Num,
  ViewabilityPct,
  Viewability_Flag,

  -- URL & UTM
  LOWER(LTRIM(RTRIM(REPLACE(LandingPageURL, 'htp://', 'http://')))) AS LandingPageURL,
  LandingDomain,
  UTM_Source,
  UTM_Medium,

  -- Engagements
  ClickToView_ms,
  TimeToConversion_s,
  PageLoad_s,

  -- Fraud flags
  IsFraud,
  SessionID,
  UserAgent

FROM adtech_dataset_main_120k;
GO
```

📈 POWER BI MODELING & DAX

Key DAX measures include:
• CTR, CVR, eCPM, CPC, ROAS, ROI
• Cumulative Revenue, YTD/MTD Revenue
• Fraud % detection & segmentation
• DayPart classification using SWITCH()
• Publisher/Campaign rankings using RANKX

********** DAX SUMMARY **********

👉 Total_impression

```
Dynamic_Total_Impressions = CALCULATE(
    COUNTROWS(adserver_feed_20k),
    adserver_feed_20k[EventType] = "impression"
)
```

👉 Total_conversion

```
Dynamic_Total_Conversions = CALCULATE(
    COUNTROWS(adserver_feed_20k),
    adserver_feed_20k[EventType] = "conversion"
)
```

👉 Total_click

```
Dynamic_Total_Clicks = CALCULATE(
    COUNTROWS(adserver_feed_20k),
    adserver_feed_20k[EventType] = "click"
)
```

👉 CTR (click through rate) = showing how effective your content is at getting users to click

```
    CTR = DIVIDE([Dynamic_Total_Clicks],[Dynamic_Total_Impressions],0)
```

👉 CVR (conversion rate) = which is the percentage of users who complete a desired action

```
    CVR = DIVIDE([Total_conversion],[Total_Click],0)
```

👉 eCPM(Revenue per thousand impression)

```
 eCPM(Revenue per thousand impression) =
DIVIDE(SUM(vw_Adtech_Cleaned[Revenue_Num])*1000,[Total_Impressions],0)
```

👉 Cost per click

```
Cost_per_click =
DIVIDE(SUM(vw_Adtech_Cleaned[Cost_Num]),[Total_Click],0)
```

👉 ROI (return on investment)

```
ROI(Return on investement) =
DIVIDE((SUM(vw_Adtech_Cleaned[Revenue_Num]) -
SUM(vw_Adtech_Cleaned[Cost_Num])),SUM(vw_Adtech_Cleaned[Cost_Num]))
```

👉 Cumulative Revenue

```
Cumulative Revenue =
CALCULATE(
    [Total_revenue],
    FILTER(
        ALL(vw_Adtech_Cleaned),
        vw_Adtech_Cleaned[EventDate] <=
MAX(vw_Adtech_Cleaned[EventDate])
    )
)
```

👉 Revenue YTD

```
Revenue YTD =
TOTALYTD([Total_revenue],vw_Adtech_Cleaned[EventDate].[Date])
```

👉 Revenue MTD

```
Revenue MTD = TOTALMTD([Total_revenue],
vw_Adtech_Cleaned[EventDate].[Date])
```

👉 Revenue Previous Year

```
Revenue_Prev_Year = CALCULATE([Total_revenue],
SAMEPERIODLASTYEAR(vw_Adtech_Cleaned[EventDate].[Date]))
```

👉 Revenue last 30 days

```
Revenue_last_30_Days =
CALCULATE([Total_revenue],DATESINPERIOD(vw_Adtech_Cleaned[EventDate].[D
ate],MAX(vw_Adtech_Cleaned[EventDate]),-30,DAY))
```

👉ROAS (Return on Ad Spend)

```
ROAS = DIVIDE([Total_revenue],[Total Cost])
```

👉Top N Conversion

```
TopN_conversion =
RANKX(ALL(vw_Adtech_Cleaned[Clean_CampaignName]),[Avg_TimeToConversion_
s],,DESC,Dense)
```

👉Top N eCPM

```
TopN_eCPM =
RANKX(
    ALL(vw_Adtech_Cleaned[Clean_PublisherName]),
    [eCPM(Revenue per thousand impression)],
    ,
    DESC
)
```

👉Avg Click_views

```
Avg_ClickToView_ms =
CALCULATE(
    AVERAGE(vw_Adtech_Cleaned[ClickToView_ms]),
    FILTER(
        vw_Adtech_Cleaned,
        vw_Adtech_Cleaned[ClickToView_ms] > 0
    )
)
```

👉Avg Time_To_conversion

```
Avg_TimeToConversion_s =
CALCULATE(
    AVERAGE(vw_Adtech_Cleaned[TimeToConversion_s]),
    FILTER(
        vw_Adtech_Cleaned,
        vw_Adtech_Cleaned[TimeToConversion_s] > 0
    )
)
```

👉Fraud_Impression

```
Fraud_Impressions =
CALCULATE([Total_Impressions], vw_Adtech_Cleaned[IsFraud] = 1)
```

👉Fraud_percentage

```
Fraud_Percentage =
DIVIDE(
    CALCULATE(COUNTROWS(vw_Adtech_Cleaned), vw_Adtech_Cleaned[IsFraud]
= 1),
    COUNTROWS(vw_Adtech_Cleaned),
    0
)
```

👉Day part

```
DayPart =

SWITCH(

    TRUE(),

    vw_Adtech_Cleaned[EventHour] >= 0 && vw_Adtech_Cleaned[EventHour] <
6, "Late Night (12AM-6AM)",

    vw_Adtech_Cleaned[EventHour] >= 6 && vw_Adtech_Cleaned[EventHour] <
12, "Morning (6AM-12PM)",

    vw_Adtech_Cleaned[EventHour] >= 12 && vw_Adtech_Cleaned[EventHour]
< 18, "Afternoon (12PM-6PM)",

    "Evening (6PM-12AM)"

)
```

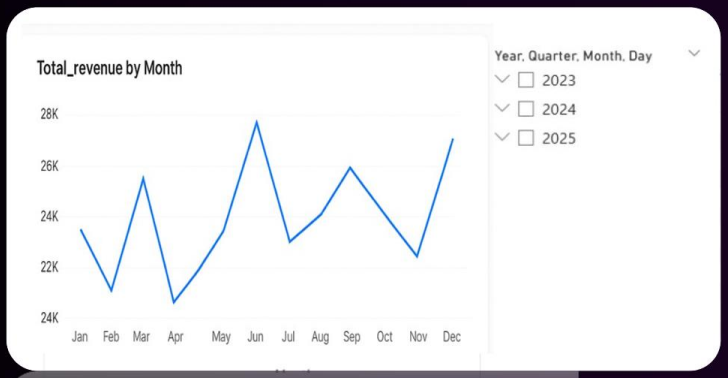
EDA

KPI Performance Analysis (CTR, CVR, CPC, eCPM, ROI)

16K	3668	798	23.61%	21.76%
Total_Impressions	Total_Click	Total_conversion	CTR	CVR

315.45K	20.31K	24.24	2.55
Total_revenue	eCPM(Revenue per thousand impression)	Cost_per_click	ROI(Return on investement)

How is total revenue distributed across months?



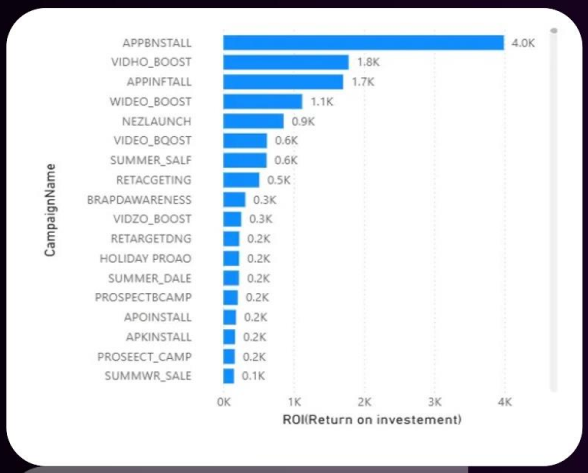
How is total revenue distributed across months?

Clean_CampaignName	Desktop	Mobile	Other	Total
AAPINSTALL	1.80	1.80	0.00	**1.80**
AAPINSTALL	0.00		0.00	**0.00**
AEPINSTALL	0.00	0.23	0.23	**0.23**
AETARGETING	0.00		1.23	**1.23**
AGPINSTALL		0.00		**0.00**
AIDEO_BOOST	0.00		0.00	**0.00**
AKPINSTALL	0.00	0.00		**0.00**
ALPINSTALL	0.00		0.00	**0.00**
AMPINSTALL	0.00	0.00		**0.00**
AOLIDAY PROMO	0.00			**0.98**
Total	**1,00,560.80**	**1,59,858.14**	**55,032.65**	**3,15,451.59**

Revenue KPI's

133.49K	181.96K	3.49K	95.82K	3.55
Revenue YTD	Revenue_Prev_Year	Revenue_last_30_Days	HighViewRevenue	ROAS

Which campaigns have the highest ROI ?



How is revenue distributed across 2023, 2024, and 2025?

Year	Cumulative Revenue
2025	3,15,451.59
2024	1,81,964.72
2023	25,887.27
Total	**3,15,451.59**

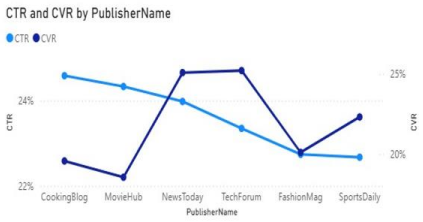
Which campaigns show zero or negative ROI & ROAS?

Clean_CampaignName	Publishername	Cost	ROI	ROAS
AAPINSTALL	NEWSTODAY	0.61	-1.00	0.00
AAPINSTALL	QOVIEHUB	1.11	-1.00	0.00
ABPINSTALL	COOKINGBLOG	2.15	-1.00	0.00
ABPINSTALL	FASHIONMAG	0.52	-1.00	0.00
ADPINSTALL	COOKINGBLOG	1.40	-1.00	0.00
ADPINSTALL	SPORTSDAILY	0.14	-1.00	0.00
AEPINSTALL	COOKINGBLOG	0.33	-1.00	0.00
AGPINSTALL	FASHIONMAG	1.25	-1.00	0.00
AGPINSTALL	TECHFORUM	0.67	-1.00	0.00
AHPINSTALL	SPORTSDAILY	0.17	-1.00	0.00
AIDEO_BOOST	MOVIEHUB	1.25	-1.00	0.00
AJPINSTALL	FASHIONMAG	1.31	-1.00	0.00
AJPINSTALL	SPORTSDAILY	1.21	-1.00	0.00
Total		**88,925.63**	**2.55**	**3.55**

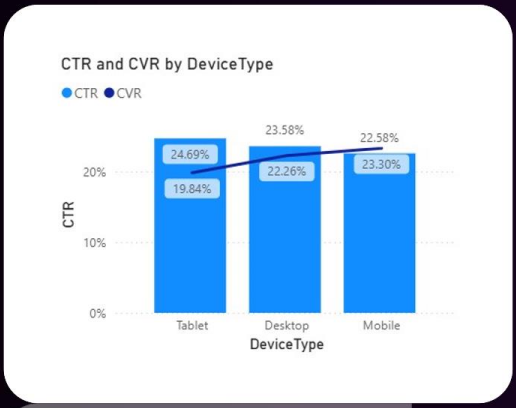
Is there a significant CTR-CVR imbalance across publishers?

23.61%
CTR Percentage

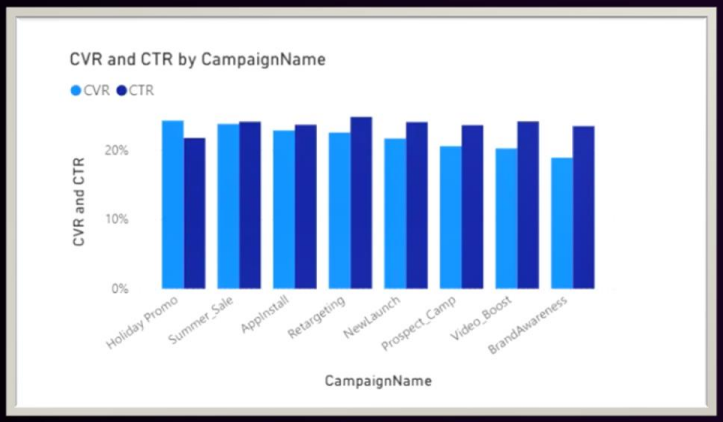
21.76%
CVR Percentage



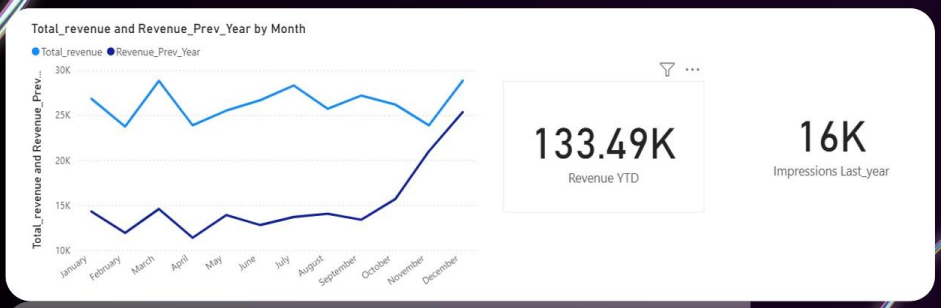
Is there a noticeable device preference trend for engagements (CTR) vs conversions (CVR)?



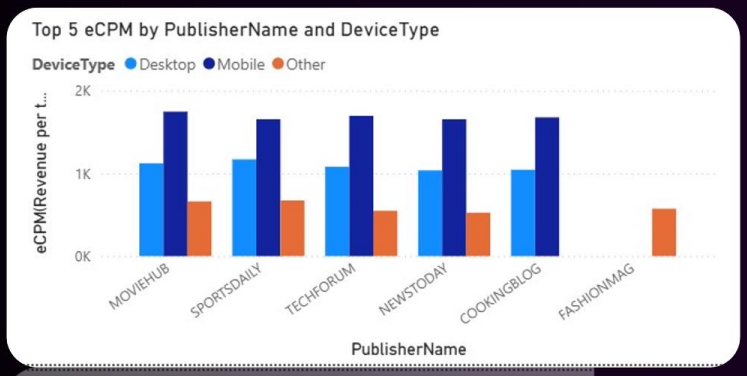
Which campaign generates the highest CTR OR CVR?



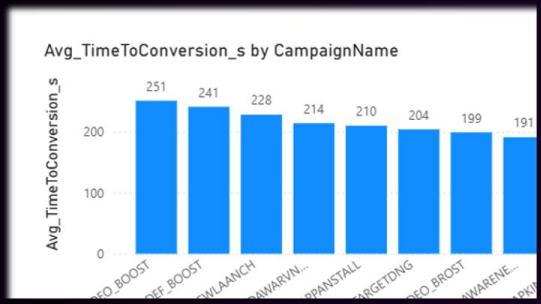
Comparing the Revenue with Previous Year



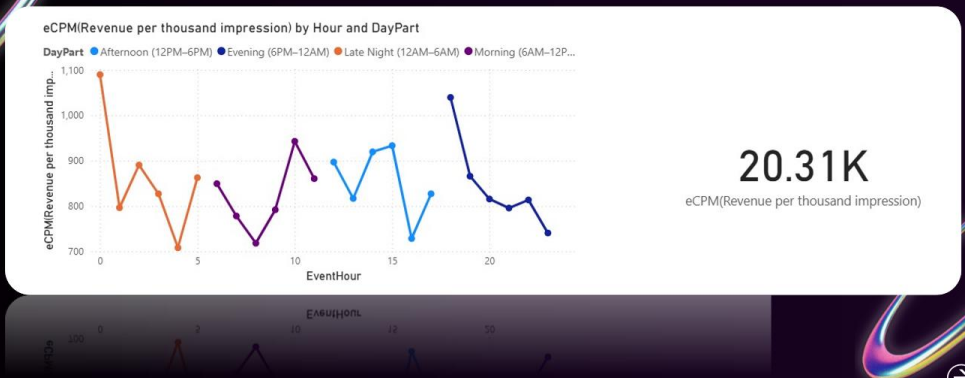
Top 5 Ecpm By Publishername & Device Type



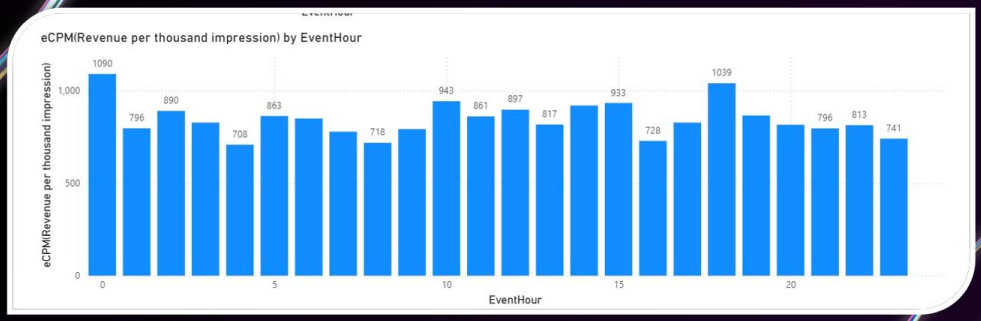
Average Time conversion by compaign name



Which DayPart (Morning, Afternoon, Evening, Late - Night) delivers the highest eCPM?

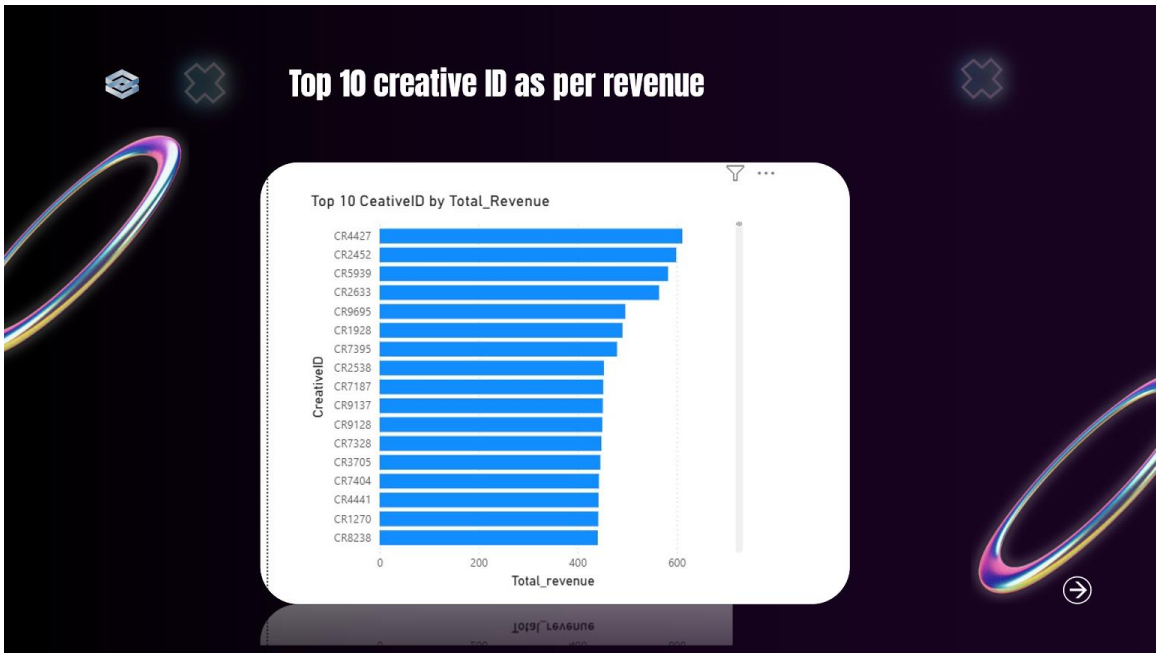


Does Morning (6AM-12PM) show noticeable dips in certain hours?



What is the percentage of frauds?

0.49250%	591
Fraud_Percentage	Total Fraud_entry



🎨 VISUAL WORKFLOW ILLUSTRATION

📁 CSV Files → 🐍 Python Loader → 🗄️ SQL Cleaned View → 📊 Power BI Dashboard

💡 BUSINESS RECOMMENDATIONS

- Google Ads delivered the highest ROAS (3.4x)
- LinkedIn campaigns showed 22% lower CTR, reduce spend on low-performing ads
- Peak engagement observed on evenings (7–12 PM)
- Need to work on Festival season [July- October]
- Cookingblog and moviehub have high difference in CTR vs CVR
- Recommend shifting at least 30% of creatives to short-form video