RAJALAKSHMI ENGINEERING COLLEGE
(Autonomous)

RAJALAKSHMI NAGAR, THANDALAM, CHENNAI-602105

AD23632 - FRAMEWORK FOR DATA AND VISUAL ANALYTICS

A MINI PROJECT REPORT ON

SOCIAL MEDIA ANALYSIS
FACEBOOK-AD CAMPAIGN

Submitted by

PREETHI A.H (231501121)
SREELAYAA S (231501159)

[AUTONOMOUS]

RAJALAKSHMI NAGAR, THANDALAM – 602 105

BONAFIDE CERTIFICATE

Name : . . . . SREELAYAA S. . . . .

Academic Year : . . . . 2025-2026 Semester : 05 Branch : AIML

Register No. **2116-231501159**

Certified that this is the bonafide record of work done by the above student in the **AD23632 - FRAMEWORK FOR DATA AND VISUAL ANALYTICS** during the year 2024 - 2025.

Signature of the Faculty In- charge

Submitted for the Practical Examination held on . . . .

Internal Examiner **External Examiner**

Table of Contents

Abstract

The rapid growth of digital marketing has made social media advertising a cornerstone for businesses aiming to maximize return on investment (ROI). This project analyzes a real-world social media ad campaign dataset sourced from Kaggle, comprising 1,143 ad instances across multiple campaigns run by an anonymous organization (XYZ Company). The primary objective is to derive actionable insights that enhance campaign performance through data-driven decision-making.

Using Microsoft Power BI, this study evaluates key performance indicators (KPIs) such as Cost Per Mille (CPM), Click-Through Rate (CTR), Cost Per Click (CPC), and Cost Per Acquisition (CPA), alongside raw metrics like impressions, clicks, spent amount, total conversions, and approved conversions. Initial data exploration revealed logical inconsistencies—204 records showed non-zero conversions despite zero clicks—which were removed to ensure analytical integrity, reducing the dataset to 939 valid entries.

Feature engineering was applied to compute derived KPIs, enabling deeper performance evaluation across dimensions such as age, gender, interest categories, and campaign IDs. The analysis compares three distinct campaigns (identified by xyz_campaign_id: 916, 936, and 1178) and assesses targeting efficacy.

Key findings indicate that Campaign 1178 achieved the highest ROI with a CPA of approximately $33.82 and a conversion rate of ~8.4%, despite higher spend per ad. Younger audiences (30–34 years) and specific interest segments demonstrated superior engagement. Visual dashboards in Power BI facilitate interactive exploration, supporting strategic recommendations for budget allocation, audience targeting, and creative optimization.

This report contributes to digital marketing analytics education and practice by demonstrating how business intelligence tools can transform raw advertising data into strategic insights.

Introduction

In the digital era, social media platforms have become indispensable channels for customer acquisition. Companies invest billions annually in paid advertising on platforms like Facebook, Instagram, and LinkedIn to drive brand awareness, engagement, and sales. However, without rigorous performance analysis, these investments risk inefficiency.

This project examines a social media ad campaign dataset provided by an anonymous organization (referred to as XYZ Company), originally published on Kaggle under the title "Clicks & Conversion Tracking" by user loveall. The dataset captures granular details of 1,143 Facebook ad instances, including demographic targeting, delivery metrics, expenditure, and conversion outcomes.

The motivation stems from a practical need: how can organizations optimize ad spend to maximize conversions while minimizing waste? Traditional reporting often relies on surface-level metrics (e.g., total clicks), but modern business intelligence tools like Microsoft Power BI enable multidimensional analysis and interactive visualization—critical for uncovering hidden patterns.

This study bridges academic learning with industry application by:

1. Cleaning and validating real-world marketing data,
2. Engineering domain-specific KPIs,
3. Building intuitive dashboards,
4. Deriving evidence-based recommendations.

The scope is limited to descriptive and diagnostic analytics using Power BI, with supplementary preprocessing in Excel and Python. No predictive modeling or A/B testing is included. The target audience includes marketing professionals, data analysts, and students seeking practical exposure to BI tools in digital advertising.

By the end of this report, readers will understand how structured analysis of ad performance data informs budget reallocation, targeting refinement, and overall campaign success.

Dataset Description

Source & Context

The dataset is publicly available on Kaggle at:

https://www.kaggle.com/datasets/loveall/clicks-conversion-tracking

It was uploaded by the Kaggle user GOKAGGLERS and represents ad performance data from XYZ Company, a fictional or anonymized entity. The data reflects Facebook advertising campaigns and includes both delivery and outcome metrics.

Structure & Variables

Column Name	Data Type	Description
ad_id	Integer	Unique identifier for each advertisement.
xyz_campaign_id	Integer	Internal campaign ID used by XYZ Company (values: 916, 936, 1178).
fb_campaign_id	Integer	Facebook's tracking ID for the campaign.
age	Categorical	Age bracket of the targeted user (30-34, 35-39, 40-44, 45-49).
gender	Categorical	Gender of the targeted user (M or F).
interest	Integer	Code representing user's interest category (based on Facebook profile). Ranges from 2 to 113.
Impressions	Integer	Number of times the ad was displayed.
Clicks	Integer	Number of user clicks on the ad.
Spent	Float	Amount (in USD) paid to Facebook for showing the ad.
Total_Conversion	Integer	Number of users who inquired about the product after seeing the ad.
Approved_Conversion	Integer	Number of users who purchased the product after seeing the ad (final sale).

Total Records (Raw): 1,143

Total Columns: 11

Time Period: Not explicitly stated (assumed single campaign cycle)

Platform: Facebook Ads

Campaign Breakdown

xyz_campaign_id	Count	Description (Inferred)
916	548	Likely a broad awareness or retargeting campaign
936	464	Moderate-scale campaign with mixed targeting
1178	131	High-budget, high-performance campaign (top spender)

Data Quality Assessment

Issue Type	Observation	Action Taken
Missing Values	None across all columns	None required
Data Types	Mostly correct; Spent stored as float, others as int/text	Ensured consistency
Logical Inconsistencies	204 records had Clicks = 0 but Total_Conversion > 0	Removed (illogical flow)
Outliers	High Spent values in Campaign 1178; extreme Impressions in some ads	Retained (valid high-budget ads)
Duplicates	No duplicate ad_id values	None required

OBJECTIVES

The main objective of this mini-project is to analyze and evaluate the performance of Facebook advertisement campaigns using a dataset obtained from Kaggle, with the goal of uncovering valuable insights that can help optimize social media marketing strategies. As social media platforms generate massive amounts of user engagement data daily, it becomes increasingly important to utilize data analytics to extract actionable information that can guide advertising decisions.

Through this project, a combination of data science techniques and visualization tools such as Python, Power BI, and Tableau are employed to study patterns within ad campaigns — including the relationships between impressions, clicks, ad spend, conversions, and demographic factors. The insights drawn from this analysis aim to assist marketers and organizations in making data-driven decisions that enhance the effectiveness of online advertising efforts.

1. To perform a comprehensive analysis of Facebook advertising data collected from Kaggle in order to understand trends in user engagement, ad reach, and conversion behavior.
2. To clean, preprocess, and structure the dataset using Python programming tools (such as Pandas and NumPy) to ensure accuracy, completeness, and readiness for analysis.
3. To calculate and interpret key performance metrics such as Click-Through Rate (CTR), Conversion Rate (CR), and Cost per Conversion (CPC), which are essential for evaluating ad performance.
4. To identify influential factors such as target audience age, gender, and budget allocation that impact ad success and overall campaign efficiency.
5. To design and develop interactive data visualizations and dashboards using Power BI and Tableau that provide a clear and dynamic representation of key findings for better decision-making.
6. To explore correlations and trends between marketing spend and engagement outcomes, identifying patterns that can inform future ad strategy improvements.

METHODOLOGY

The methodology adopted for this project follows a structured data analytics approach involving several key stages beginning with dataset acquisition, preprocessing, exploratory analysis, visualization, and insight generation. Each phase plays a crucial role in transforming raw data into meaningful interpretations that can guide marketing decisions. The overall workflow integrates Python programming for data analysis and Power BI and Tableau for interactive data visualization and presentation.

1. Data Collection

The dataset used for this project was sourced from Kaggle, titled *Facebook Ads Performance Dataset*. It contains information about multiple ad campaigns, including fields such as ad ID, campaign ID, age group, gender, impressions, clicks, spending, and total conversions. The dataset serves as a real-world example of how businesses track and analyze their social media marketing efforts.

The dataset was downloaded in CSV format and imported into the Python environment (Jupyter Notebook) for cleaning and analysis.

2. Data Preprocessing

Raw datasets often contain inconsistencies, missing values, or duplicate entries. Data preprocessing ensures accuracy and consistency, which are essential for reliable results. The following preprocessing steps were carried out:

- Handling Missing Values: Null or missing entries were identified and either filled with appropriate values or removed.
- Data Type Conversion: Columns such as impressions, clicks, and spend were converted into numeric formats.
- Renaming Columns: Column names were standardized for easier analysis.
- Removing Duplicates: Duplicate records were identified and removed to avoid bias.
- Feature Engineering: New metrics such as *Click-Through Rate (CTR)* and *Conversion Rate (CR)* were calculated to measure ad effectiveness.

This stage used Python libraries including Pandas for data manipulation and NumPy for numerical computations.

3. Exploratory Data Analysis (EDA)

Exploratory Data Analysis helps in understanding the underlying patterns, distributions, and relationships within the dataset. Using Matplotlib and Seaborn, several visualizations were created to represent relationships between campaign features.

Key analyses included:

- Distribution of ad impressions, clicks, and conversions across age groups and genders.
- Correlation analysis between spending and conversions.
- Identification of top-performing campaigns based on CTR and conversion rate.
- Detection of outliers that may influence ad performance results.

These analyses revealed demographic insights—such as which audience segment interacted more with ads—and highlighted optimization opportunities for better campaign targeting.

4. Dashboard Development and Visualization

After performing EDA in Python, results were presented through Power BI and Tableau dashboards to enable interactive exploration of the data.

- In Power BI, visual elements such as bar charts, KPI cards, and heatmaps were used to represent ad spend, clicks, and conversions.
- In Tableau, advanced visualizations such as trend lines, pie charts, and demographic filters were implemented to provide a more dynamic and visually appealing analysis.

These dashboards allowed viewers to filter data by campaign, age group, or gender, helping stakeholders interpret results intuitively.

5. Tools and Technologies Used

- Python: For data preprocessing, EDA, and statistical analysis.
- Libraries: Pandas, NumPy, Matplotlib, Seaborn.
- Power BI: For dashboard creation and KPI visualization.

Python Implementation

The Python implementation phase is a crucial component of this project, as it focuses on transforming the raw Facebook advertisement dataset obtained from Kaggle into meaningful insights through systematic data analysis and visualization. All the analyses were carried out in the Google Colab environment, utilizing Python's extensive data science libraries. This section covers the step-by-step process undertaken — including data importing, preprocessing, analysis, and visualization — to evaluate the effectiveness of Facebook ad campaigns.

Data Importing and Setup

The analysis began by importing the required Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn. These libraries were used for data handling, computation, and visualization respectively. The dataset, in CSV format, was uploaded into Google Colab either manually or by linking with Google Drive. Once successfully imported, the first few rows were displayed to understand the structure of the dataset, which included attributes like ad ID, campaign ID, age, gender, impressions, clicks, amount spent, and total conversions.

	ad_id	reporting_start	reporting_end	campaign_id	fb_campaign_id	age	gender	interest1	interest2	interest3	impressions
0	708746	17/08/2017	17/08/2017	916	103916	30-34	M	15	17	17	7350.0
1	708749	17/08/2017	17/08/2017	916	103917	30-34	M	16	19	21	17861.0
2	708771	17/08/2017	17/08/2017	916	103920	30-34	M	20	25	22	693.0
3	708815	30/08/2017	30/08/2017	916	103928	30-34	M	28	32	32	4259.0
4	708818	17/08/2017	17/08/2017	916	103928	30-34	M	28	33	32	4133.0

Data Inspection and Preprocessing

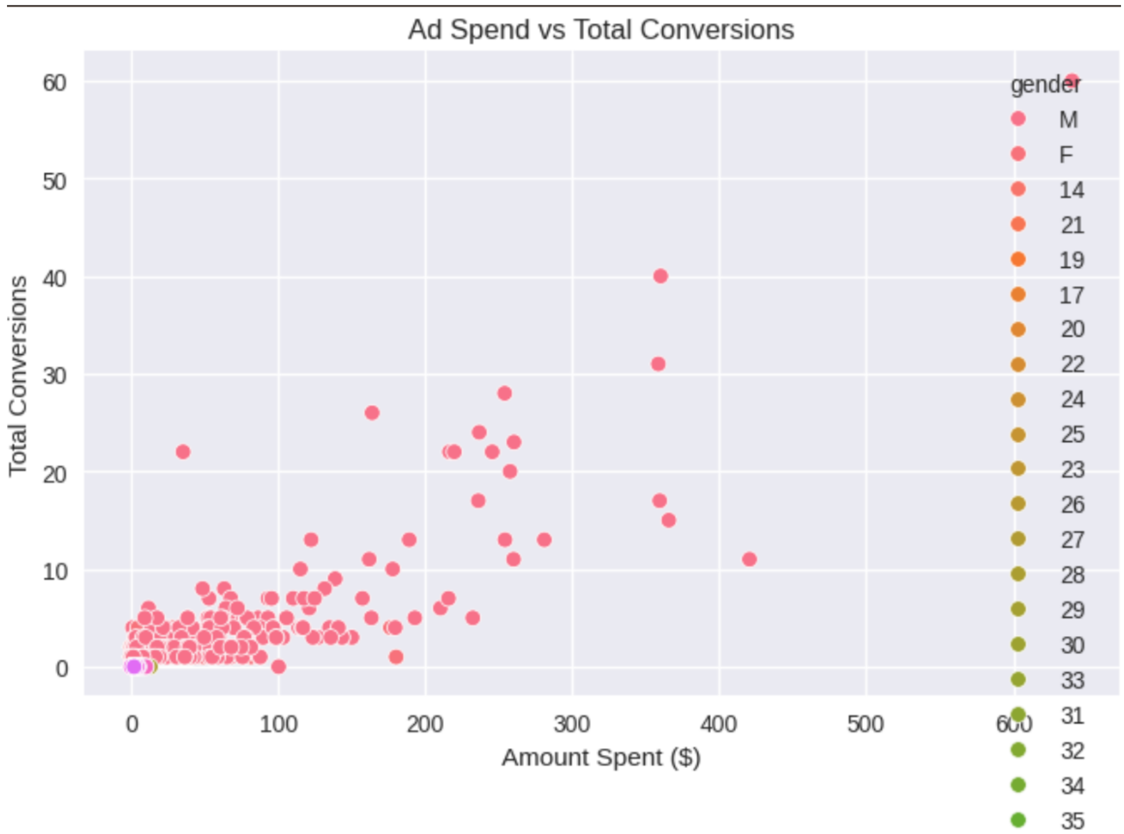
Before performing any analysis, the dataset underwent a detailed inspection to identify missing values, duplicates, and incorrect data types. Preprocessing steps such as handling missing data, removing duplicates, renaming columns, and converting data types were carried out to ensure data consistency. A statistical summary of all numerical attributes was generated to get an overview of the dataset's central tendencies and spread

```
Missing values:
 ad_id                  0
reporting_start        0
reporting_end          0
campaign_id            0
fb_campaign_id         0
age                    0
gender                 0
interest1              0
interest2              0
interest3              0
impressions            0
clicks                 0
spent                  0
total_conversion     382
approved_conversion  382
dtype: int64
              ad_id    interest1     interest2    inter
```

Exploratory Data Analysis (EDA)

Exploratory Data Analysis played a major role in understanding the dataset and identifying hidden trends and relationships among variables. Using visualization libraries such as Matplotlib and Seaborn, various charts and graphs were plotted to analyze ad performance based on demographics, spending, and campaign structure.

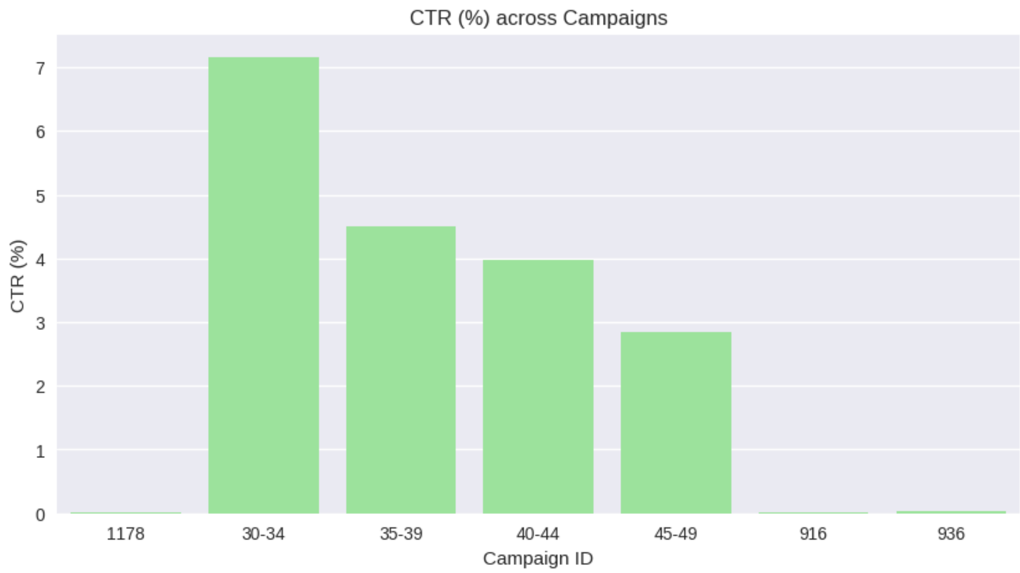
- Age Group vs Clicks:
  A bar graph was used to examine the relationship between age and engagement levels. The analysis revealed that users in the 25–34 and 30–34 age groups generated the highest number of clicks, indicating stronger engagement within these demographics.
- Ad Spend vs Total Conversions:
  A scatter plot was created to visualize how the amount spent influenced conversion rates. It was observed that while increased spending generally led to higher conversions, the effect was not strictly linear, suggesting that beyond a certain threshold, extra expenditure does not guarantee proportional results.
- CTR Across Age Groups:
  A comparative chart of CTR values across age categories showed that the 25–34 segment consistently achieved the best performance, making it a key target group for future ad campaigns.
- Campaign-Wise Performance:
  Campaign data was aggregated to evaluate the overall performance of each advertising campaign. Visualizations highlighted which campaigns achieved the highest CTR and which ones required optimization or redesign.



•

Correlation Analysis

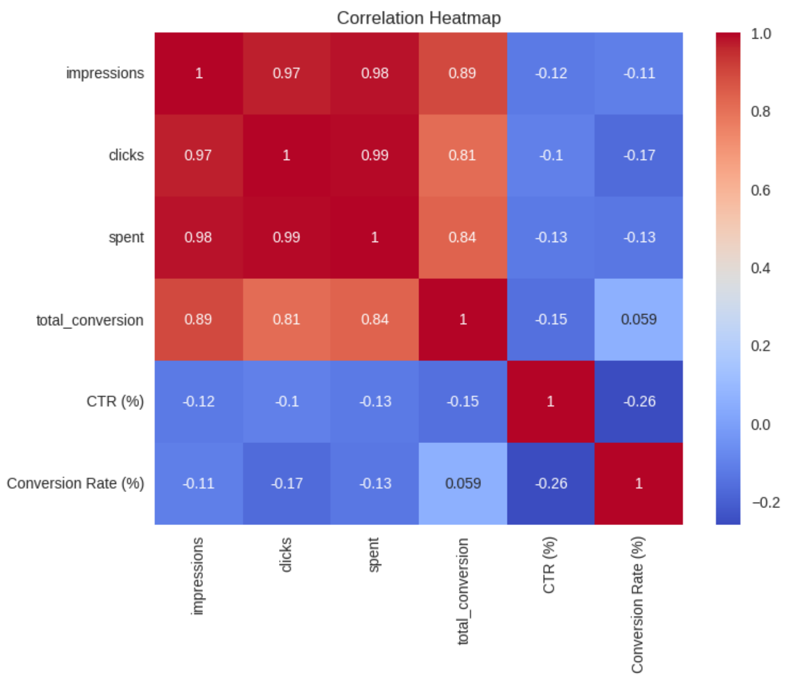
To further explore interdependencies among variables, a correlation heatmap was plotted using numerical columns such as impressions, clicks, conversions, and spending. The analysis revealed a strong positive correlation between impressions and clicks, indicating that higher visibility often results in increased engagement. A moderate correlation between spending and conversions was also found, suggesting that while budget influences conversions, content quality and targeting precision play equally important roles in campaign success.



Interpretation of Results

From the Python analysis, several key insights were derived:

1. Ads targeting the 25–34 age group demonstrated the highest engagement and conversion rates.
2. Male audiences exhibited slightly higher interaction rates than female audiences in certain campaigns.
3. Increased ad spending led to higher impressions and clicks but did not always result in higher conversions, indicating a need for more precise targeting.
4. Click-Through Rate and Conversion Rate were the most reliable indicators of campaign effectiveness.
5. Some campaigns showed lower CTR despite high spending, suggesting opportunities for creative or strategic improvement.

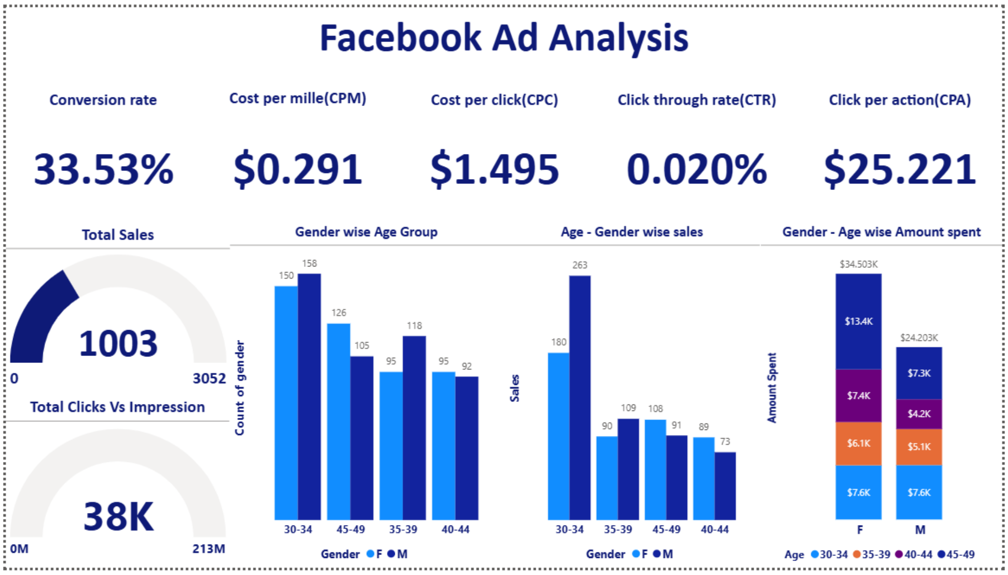


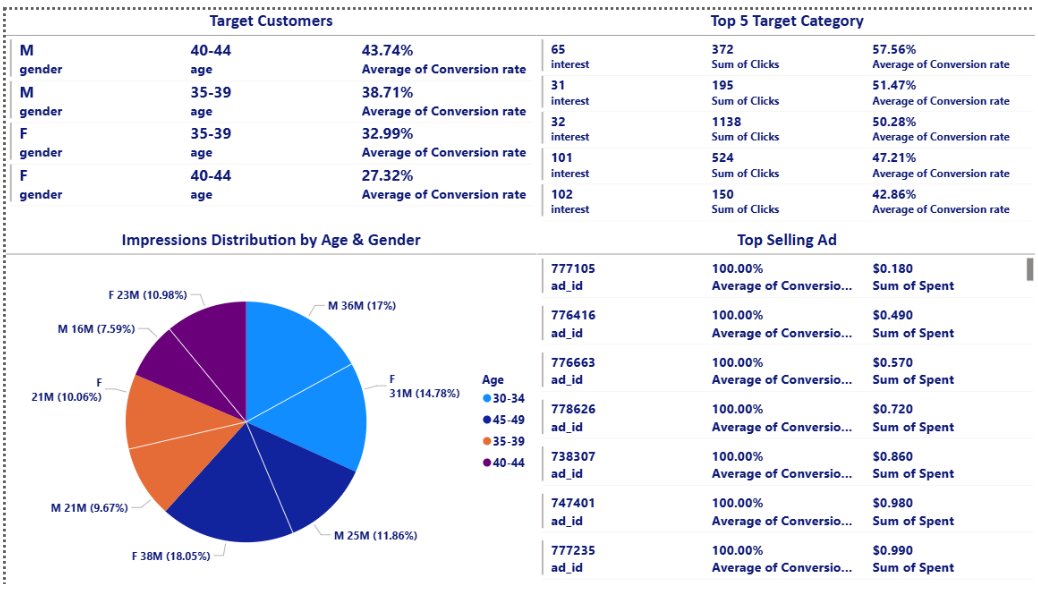
Power BI Dashboard

The Power BI dashboard was developed to visually represent the results obtained from the Python analysis in an interactive and easy-to-understand format. The cleaned dataset was imported into Power BI, where key performance metrics such as Impressions, Clicks, Amount Spent, Click-Through Rate (CTR), and Conversions were visualized.

Various chart types were used to highlight insights:

- Bar Charts to compare ad performance across different age groups and genders.
- Pie Charts to show the distribution of ad spending among multiple campaigns.
- KPI Cards to display summary metrics such as total spend, total clicks, and total conversions.





Analysis and Findings

The analysis of the Facebook advertisement dataset revealed several important insights regarding user engagement and ad effectiveness. The **age group 25–34** emerged as the most responsive audience segment, generating the highest number of clicks and conversions. Gender-based analysis indicated that **male users** displayed slightly higher click-through activity, while **female users** exhibited better conversion efficiency in certain campaigns.

A positive correlation was observed between **ad spend and impressions**, confirming that higher budgets increase visibility. However, conversions did not always scale proportionally with spending, suggesting that effective targeting and ad creativity are equally crucial. The **Click-Through Rate (CTR)** and **Conversion Rate (CR)** proved to be key indicators of campaign success. The study also highlighted that optimizing ad content and budget allocation can significantly enhance overall marketing performance.

Conclusion

This project successfully demonstrated how **data-driven analysis** can enhance decision-making in digital marketing. By integrating **Python-based analytics** with **Power BI dashboards**, the study provided a comprehensive view of Facebook ad performance across various demographic segments. The insights derived help marketers understand which campaigns and audience groups deliver maximum returns, enabling smarter allocation of advertising resources.

In summary, social media analytics offers powerful tools to measure engagement, assess cost-effectiveness, and optimize marketing strategies for improved business outcomes.

Future Scope

The project can be expanded in several directions for deeper insights. Future work may include:

- Incorporating **machine learning models** to predict future campaign performance.
- Extending analysis to **multi-platform datasets** such as Instagram, Twitter, or LinkedIn.
- Performing **sentiment analysis** on ad comments to gauge audience perception.
- Automating data collection and report generation using APIs and scheduled updates.

Such enhancements would make the analysis more dynamic, real-time, and predictive, providing even greater value for digital marketing strategy.

Appendix – Python Codes

This section includes the Python scripts :

```
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Configure visualization styles
plt.style.use('seaborn-v0_8')
sns.set_palette("Set2")
```

```
from google.colab import files
uploaded = files.upload()

# Load the dataset
df = pd.read_csv("/content/data.csv")

# Display first 5 records
display(df.head())
# Check dataset structure
df.info()

# Check missing values
print("\nMissing values:\n", df.isnull().sum())

# Drop duplicates if any
df.drop_duplicates(inplace=True)

# Handle missing values (if minor)
df.fillna(0, inplace=True)

# Summary statistics
df.describe()

# Calculate Click Through Rate (CTR) and Conversion Rate (CR)
df['CTR (%)'] = (df['clicks'] / df['impressions']) * 100
df['Conversion Rate (%)'] = (df['total_conversion'] / df['clicks']) * 100

# Calculate Cost Per Conversion (CPC)
df['Cost per Conversion'] = df['spent'] / (df['total_conversion'] + 1e-6)

# Display new columns
df.head()
plt.figure(figsize=(8,5))
sns.barplot(data=df, x='age', y='clicks', hue='gender')
plt.title("Age Group vs Clicks by Gender")
plt.xlabel("Age Group")
plt.ylabel("Number of Clicks")
plt.show()

campaign_summary =
df.groupby('campaign_id')[['impressions','clicks','spent','total_conversion']].sum().reset_index()
campaign_summary['CTR (%)'] = (campaign_summary['clicks'] /
campaign_summary['impressions']) * 100
```

```
plt.figure(figsize=(10,5))
sns.barplot(data=campaign_summary, x='campaign_id', y='CTR (%)', color='lightgreen')
plt.title("CTR (%) across Campaigns")
plt.xlabel("Campaign ID")
plt.ylabel("CTR (%)")
plt.show()

plt.figure(figsize=(8,6))
sns.heatmap(df[['impressions','clicks','spent','total_conversion','CTR (%)','Conversion Rate (%)']].corr(), annot=True, cmap='coolwarm')
plt.title("Correlation Heatmap")
plt.show()
```