Marketing Campaign Performance Insights

S.VINOTHKUMAR

Problem Statement:

In the highly competitive landscape of digital marketing, effectively evaluating the success of various marketing campaigns is essential for optimizing return on investment (ROI) and improving overall performance. Despite having extensive data on multiple campaigns, there is a need for a thorough analysis to assess and compare key metrics such as conversion rates, acquisition costs, and ROI across different campaign types, channels, and audience segments. This project aims to uncover actionable insights by examining temporal trends, geographical influences, and audience responses to identify factors driving campaign success and provide recommendations for enhancing future marketing strategies.

Objective

To create a presentation that effectively communicates the findings of the marketing campaign analysis, showcasing key insights through visualizations, reports, and data-driven storytelling for strategic decision-making.

Tools & Libraries Used

TOOLS USED

- PYTHON
- GOOGLE COLLAB
- PPT FOR PRESENTATION

LIBRARIES USED

- import numpy as np
- import pandas as pd
- import matplotlib.pyplot as plt
- import seaborn as sns
- import plotly.express as px

Project Overview:

In this project, I analyzed a marketing campaign dataset to evaluate the effectiveness of various channels, customer segments, and campaign types. The objective was to uncover actionable insights that could improve marketing performance.

Key tasks included:

- Data Cleaning & Preparation using pandas
- Exploratory Data Analysis (EDA) to study conversion rates, engagement scores, and ROI
- Visualization of trends using Seaborn, Matplotlib, and Plotly
- Identified the most cost-effective channels, top-performing customer segments, and ROI distribution by location
- Used groupby, agg(), and idxmax() to isolate key insights like the overall top customer segment and best-performing campaign types
- Created interactive visualizations using Plotly to make insights easier to explore and present

Data Dictionary

Column	Description
Campaign_ID	Unique identifier for each campaign.
Company	The organization running the campaign, represented by various fictional brands.
Campaign_Type	The type of marketing effort used, such as email, social media, influencer, display, or search.
Target_Audience	The specific demographic or audience segment targeted by the campaign (e.g., women aged 25-34).
Duration	The duration of the campaign, expressed in days.
Channels_Used	The platforms or mediums used to promote the campaign, including email, social media, YouTube, websites, or Google Ads.
Conversion_Rate	The percentage of impressions or leads that resulted in desired actions, reflecting campaign effectiveness.
Acquisition_Cost	The monetary expense incurred to acquire each customer through the campaign.

Column	Description
ROI	Return on Investment, indicating the profitability and success of the campaign.
Location	The geographical area where the campaign was executed (e.g., New York, LosAngeles).
Language	The language in which the campaign's content was delivered (e.g., English, Spanish).
Clicks	The total number of clicks generated by the campaign, showing user interaction.
Engagement_Score	A score from 1 to 10 representing the level of engagement and interaction generated by the campaign.
Customer_Segment	The specific group or category of customers targeted by the campaign (e.g., tech enthusiasts, fashionistas).
Date	The date on which the campaign occurred.

1.LOAD THE DATASET

```
df =pd.read_csv("/content/capstone_Python.csv")
Marketing_df = pd.DataFrame(df)
Marketing_df
```

2.Descriptive Analysis

```
Marketing_df.head(10)
```

```
Rows = Marketing_df.shape[0]
print("The No of rows in dataset :",Rows)

Columns = Marketing_df.shape[1]
print("The No of rows in dataset :",Columns)
```

```
Marketing_df.info()
```

```
Marketing_df.describe()
```

2.1)Data Exploration

➢ Print the number of unique Campaign_ID values in the dataset.

```
Unique_Campaign_ID = Marketing_df['Campaign_ID'].nunique()
print("Number of unique Campaign_ID's =",Unique_Campaign_ID)
```

➢ List the unique values of the Location and Customer_Segment columns.

```
Unique_Location = Marketing_df['Location'].nunique()
print("Number of unique Unique_Location's =",Unique_Location)

Customer_Segment = Marketing_df['Customer_Segment'].nunique()
print("Number of unique Customer_Segment's =",Customer_Segment)
```

➢ Count the occurrences of each category in the Campaign_Type and Channel_Used and columns.

```
occurences_of_Campaign_Type = Marketing_df['Campaign_Type'].value_counts()
print("The occurences of Campaign_Type = ",occurences_of_Campaign_Type)
```

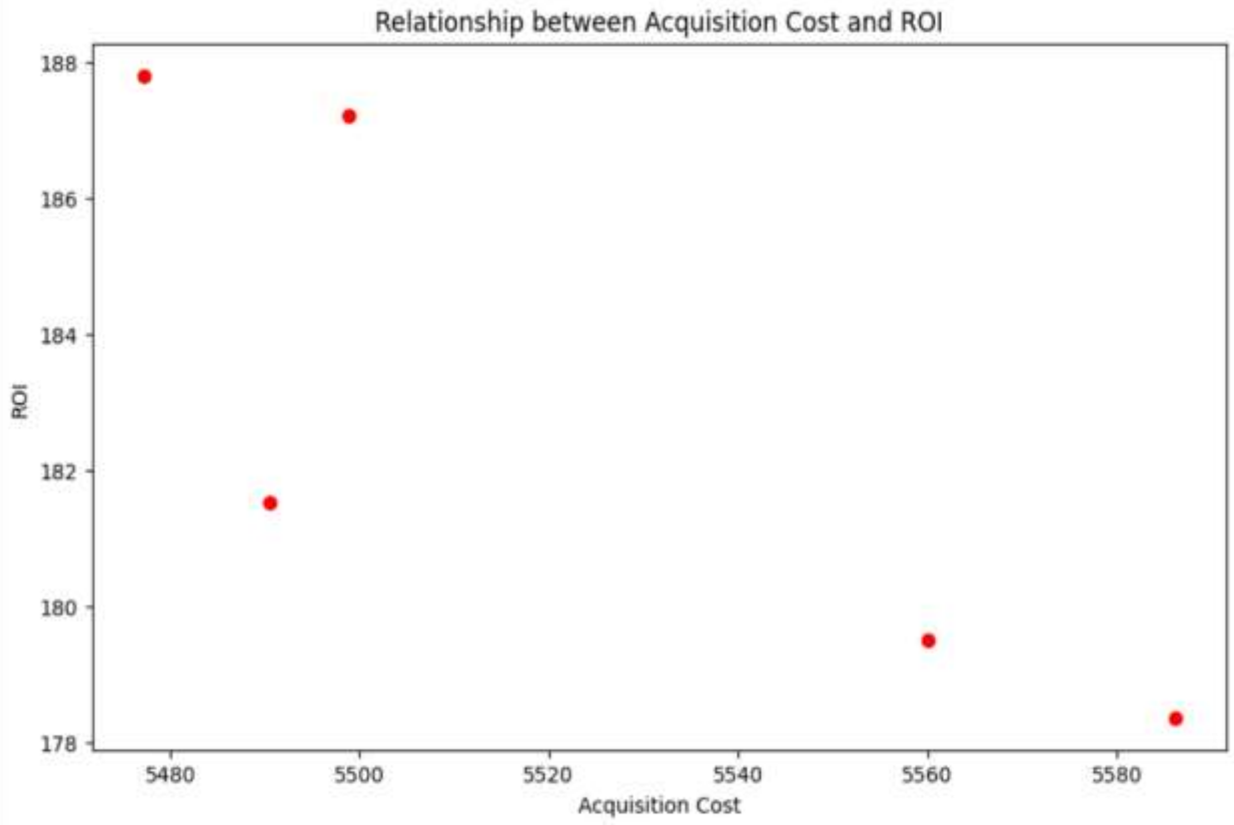
```
occurences_of_Channel_Used =Marketing_df['Channel_Used'].value_counts()
print("The occurences of Channel_Used = ",occurences_of_Channel_Used)
```

3) Exploratory Data Analysis (EDA) and Visualization

Campaign Performance:

➢ Plot a scatter plot to visualize the relationship between Acquisition_Cost and ROI.

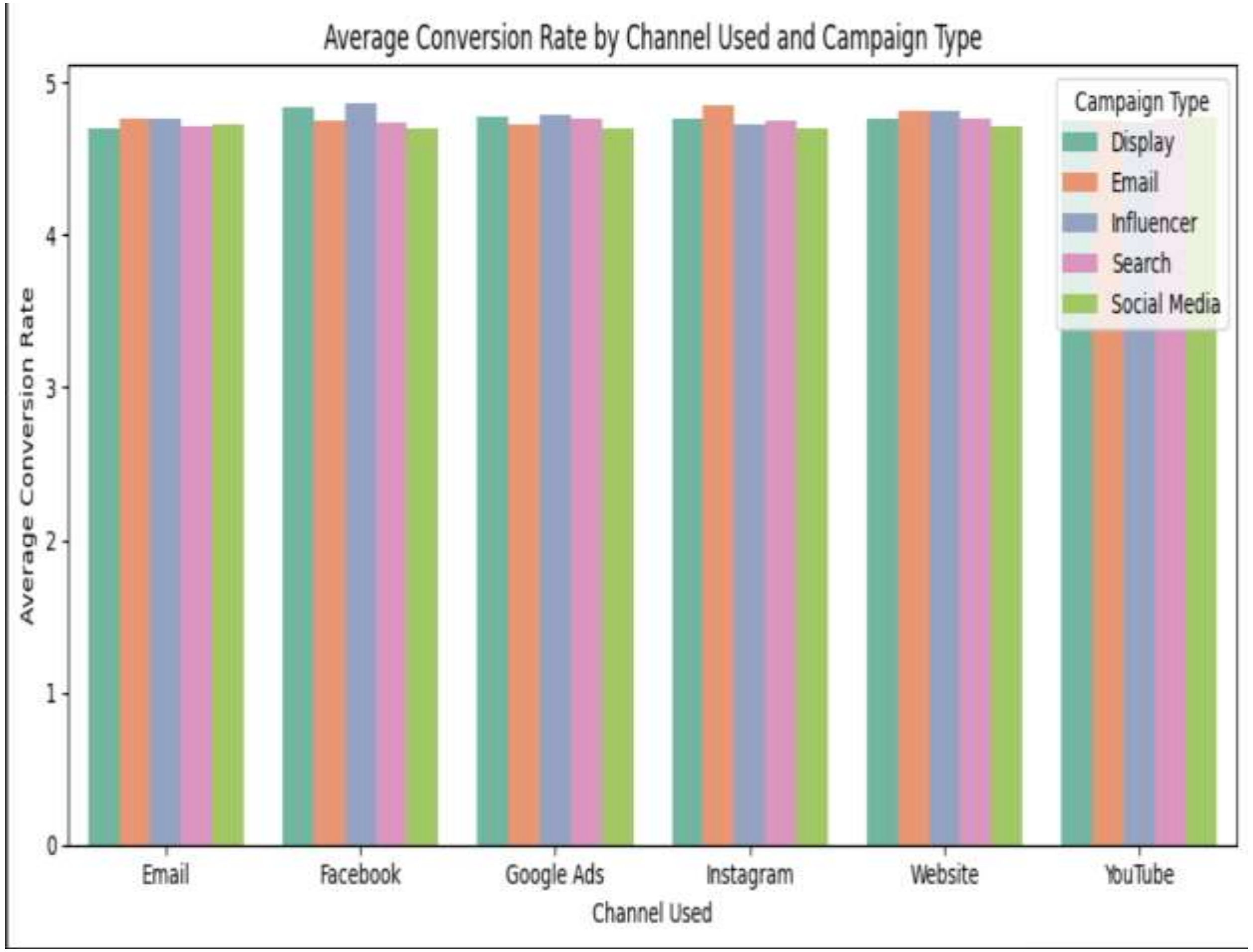
```
scatter_visual =
Marketing_df.groupby('Company')[['ROI','Acquisition_Cost']].
mean().reset_index()
plt.figure(figsize=(10,6))
plt.scatter(scatter_visual['Acquisition_Cost'],
scatter_visual['ROI'], color='red')
plt.title("Relationship between Acquisition Cost and ROI")
plt.xlabel("Acquisition Cost")
plt.ylabel("ROI")
plt.show()
```



Campaign Performance:

➢ Create a bar chart to visualize the average Conversion_Rate for different Channel_Used, categorized by Campaign_Type.

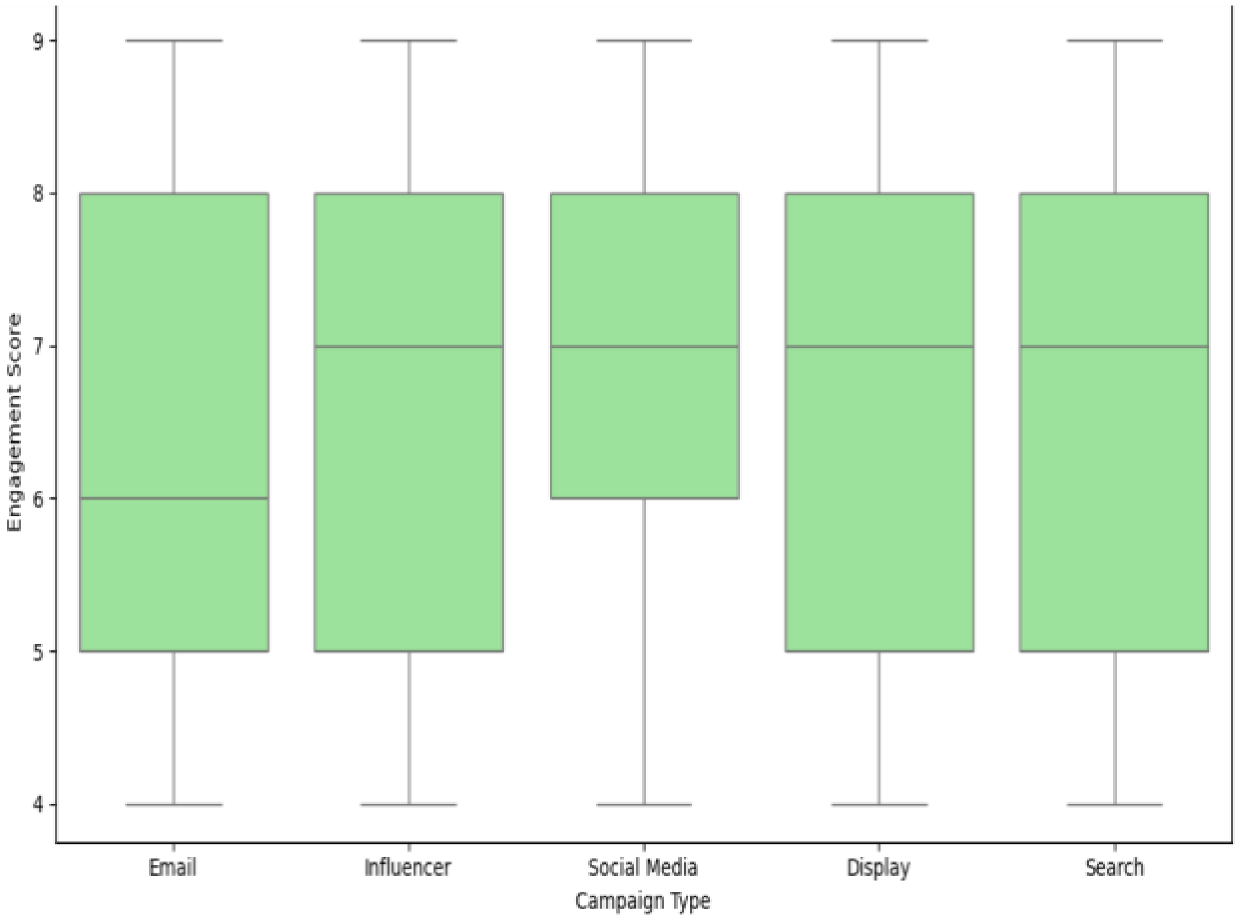
```
Bar_Visual = Marketing_df.groupby(['Channel_Used',
'Campaign_Type'])['Conversion_Rate'].mean().reset_index()
plt.figure(figsize=(12,6))
sns.barplot(data=Bar_Visual, x='Channel_Used',
y='Conversion_Rate', hue='Campaign_Type', palette='Set2')
plt.title("Average Conversion Rate by Channel Used and
Campaign Type")
plt.xlabel("Channel Used")
plt.ylabel("Average Conversion Rate")
plt.legend(title='Campaign Type')
plt.tight_layout()
plt.show()
```



Campaign Performance:

➢ Visualize the distribution of Engagement_Score across different Campaign_Type using a box plot.

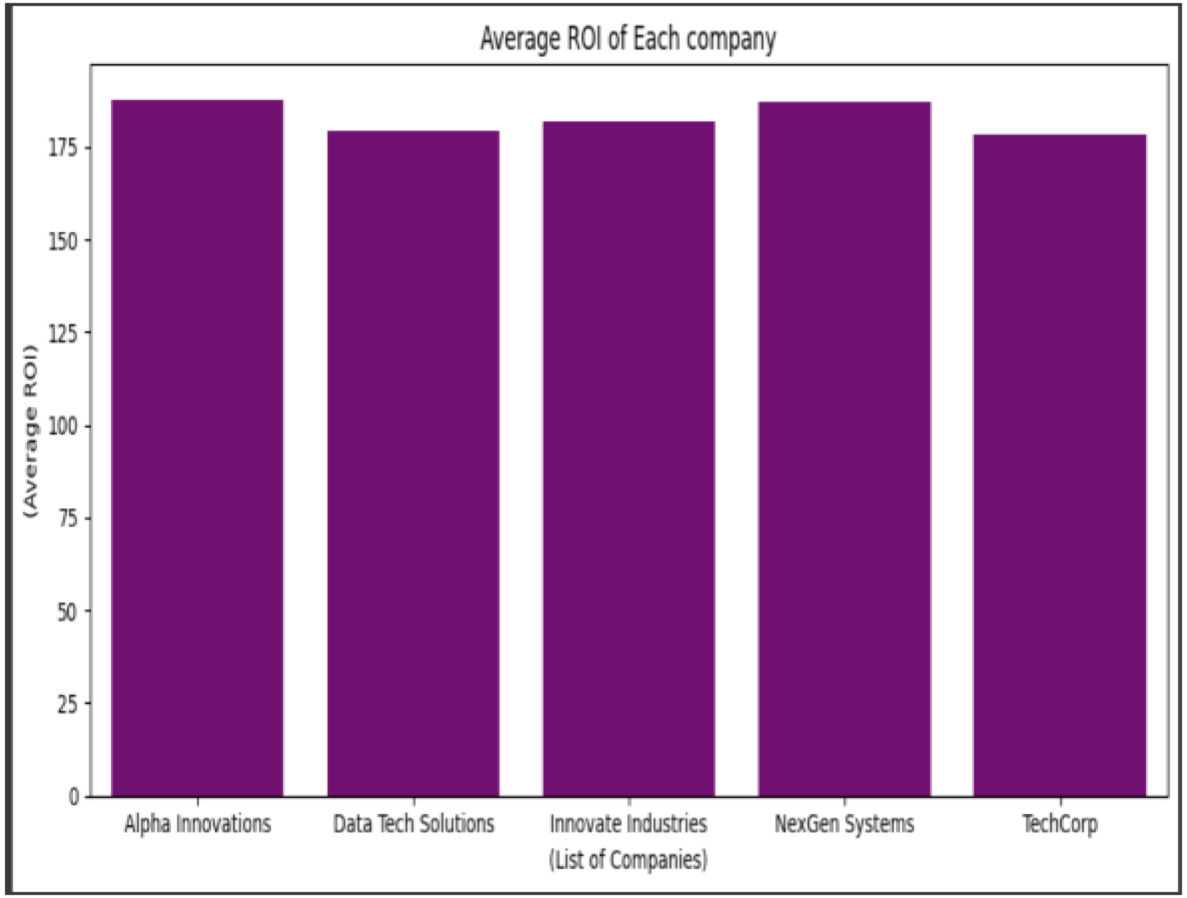
```
plt.figure(figsize=(10,6))
sns.boxplot(data=Marketing_df, x='Campaign_Type',
y='Engagement_Score',color = 'lightgreen')
plt.title("Distribution of Engagement Score by Campaign
Type")
plt.xlabel("Campaign Type")
plt.ylabel("Engagement Score")
plt.tight_layout()
plt.show()
```



Campaign Performance:

➢ Analyze the average ROI by Company using a bar chart to compare the profitability of campaigns conducted by different companies.

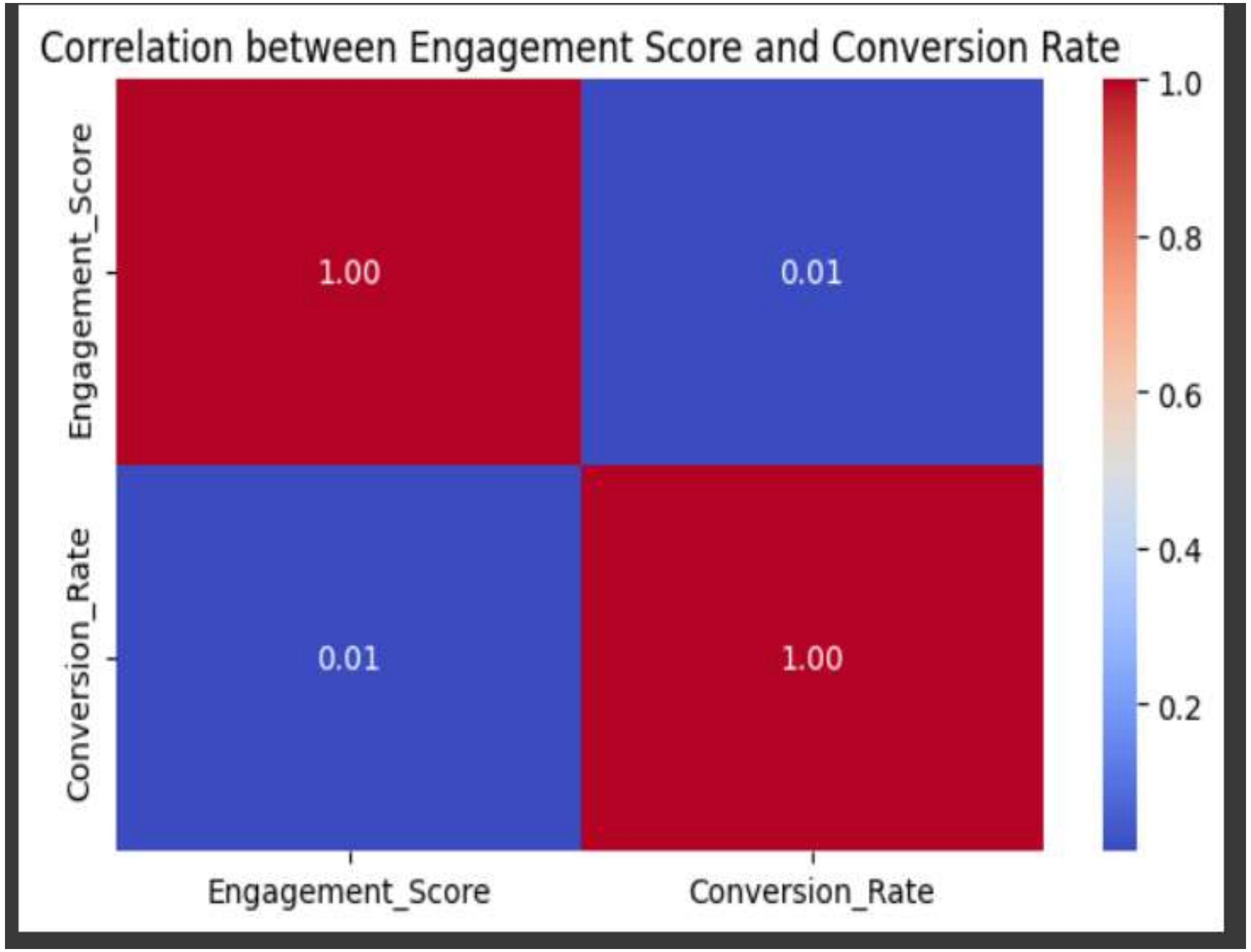
```
ROI_by_Company =
Marketing_df.groupby('Company')['ROI'].mean().round(2).reset
_index()
plt.figure(figsize=(10,6))
sns.barplot(data=ROI_by_Company,x='Company',y='ROI',color='
Purple')
plt.title('Average ROI of Each company')
plt.ylabel('(Average ROI)')
plt.xlabel('(List of Companies)')
plt.tight_layout()
plt.show()
```



Campaign Performance:

➢ Examine the correlation between Engagement_Score and Conversion_Rate using a heatmap.

```
correlation_matrix =
Marketing_df[['Engagement_Score',
'Conversion_Rate']].corr()
plt.figure(figsize=(6,4))
sns.heatmap(correlation_matrix, annot=True,
cmap='coolwarm', fmt=".2f")
plt.title("Correlation between Engagement Score and
Conversion Rate")
plt.tight_layout()
plt.show()
```



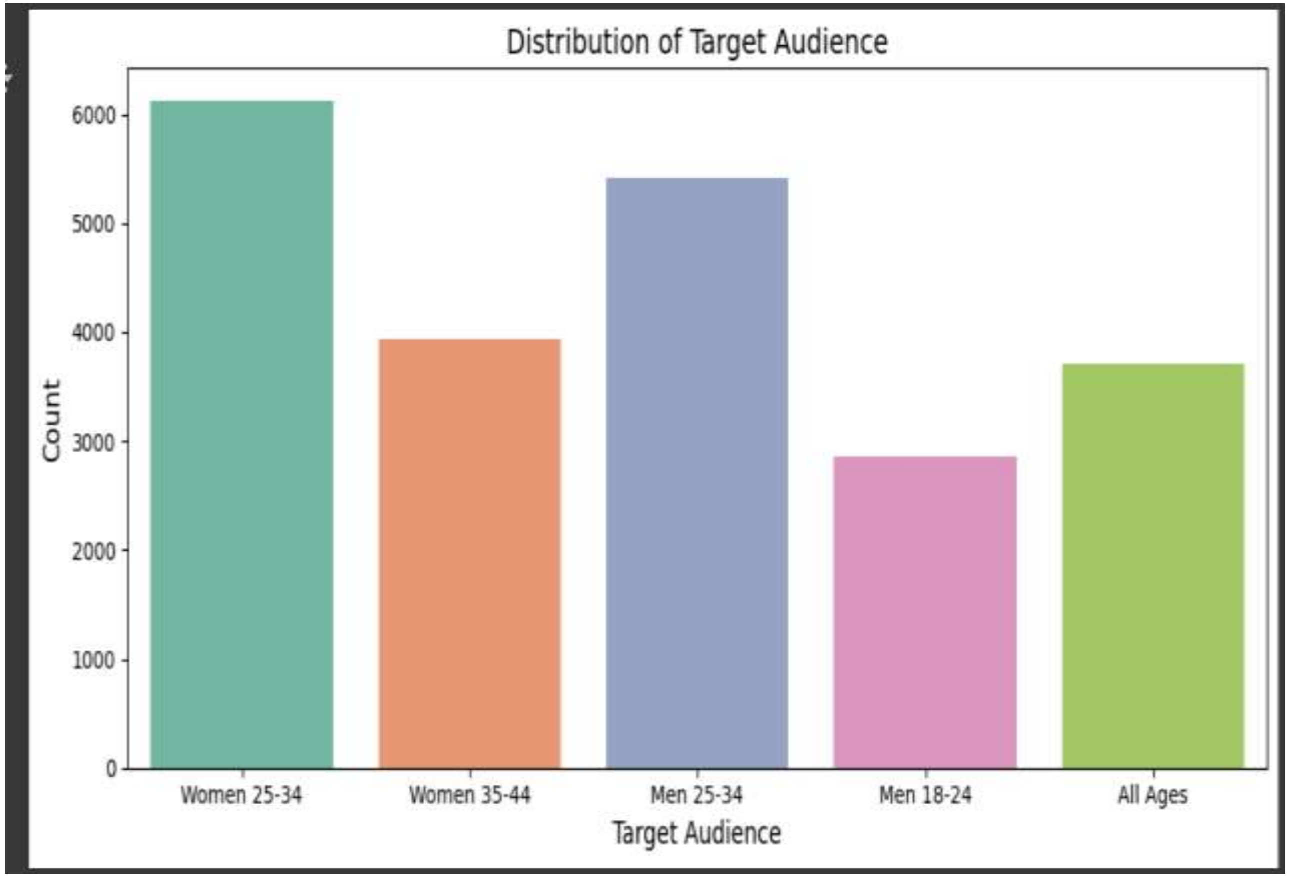
Inferences of Campaign Performance:

- "The ROI tends to be higher when the average acquisition cost is low, and it shows a steady decline as acquisition costs increase. This suggests an inverse relationship between acquisition cost and ROI, indicating that more cost-effective campaigns yield better returns."
- Email and Influencer campaigns show slightly higher conversion rates across most channels, indicating they are more effective in driving user actions.
- All marketing channels perform consistently well, with conversion rates ranging between 4.7 and 4.9, suggesting a balanced and optimized campaign strategy.
- Median Engagement Score is higher for Influencer, Social Media, Display, and Search campaigns (~7), while Email campaigns have a slightly lower median (~6).
- All campaign types have a similar spread and range, suggesting consistent audience engagement across different strategies.
- Alpha Innovations and NexGen Systems have the highest average ROI, indicating more effective campaign outcomes.
- The differences in ROI across companies are minimal, implying all companies are performing fairly well in terms of returns.
- The correlation coefficient is 0.01, indicating almost no linear relationship between Engagement Score and Conversion Rate.
- This suggests that even if a campaign has high engagement, it doesn’t necessarily lead to higher conversion — other factors may influence conversions.

Customer Segmentation:

➢ Create a count plot to visualize the distribution of Target_Audience.

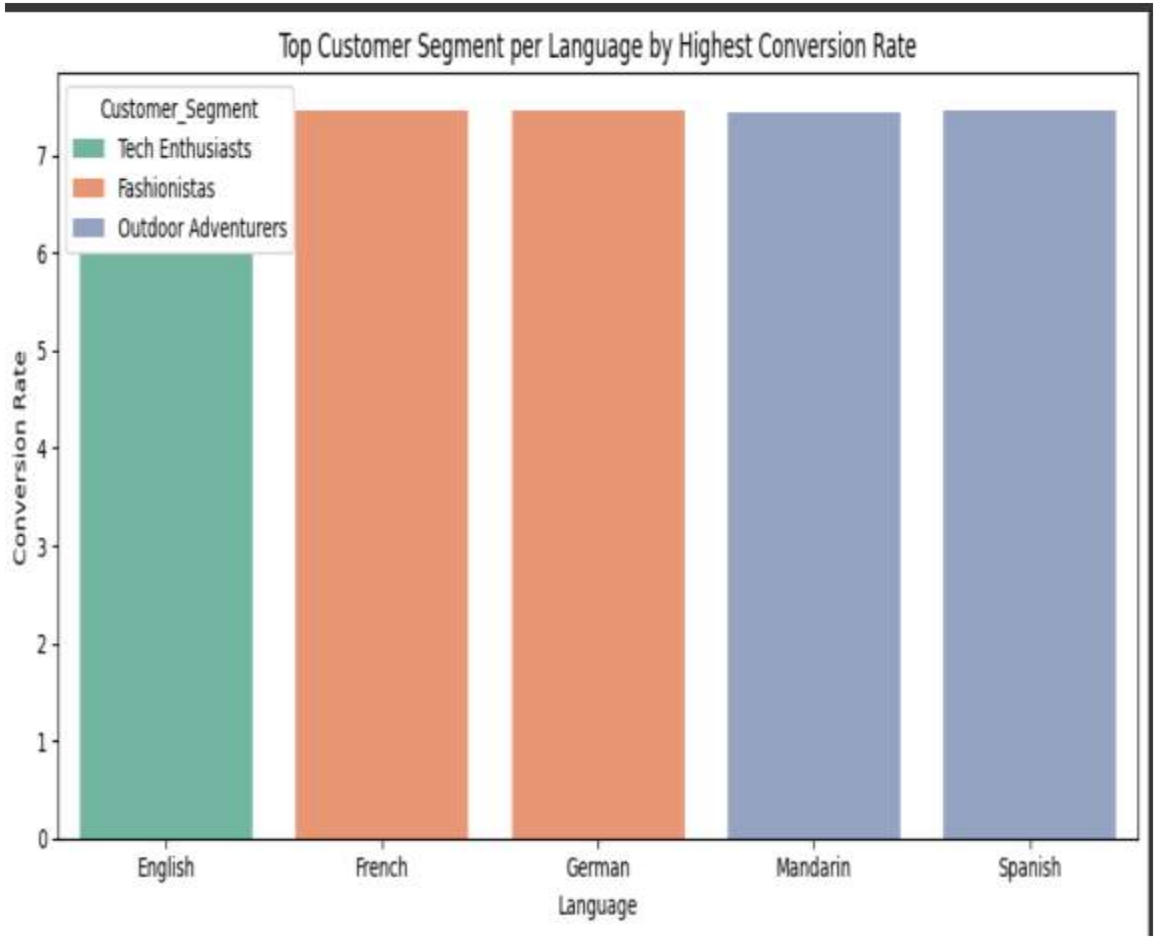
```
plt.figure(figsize=(10, 5))
sns.countplot(data=Marketing_df,
x='Target_Audience', hue='Target_Audience',
palette='Set2', legend=False)
plt.title("Distribution of Target Audience", fontsize=14)
plt.xlabel("Target Audience", fontsize=12)
plt.ylabel("Count",fontsize=12)
plt.tight_layout()
plt.show()
```



Customer Segmentation:

➢ Identify which Customer_Segment has the highest Conversion_Rate for each Language using a bar chart.

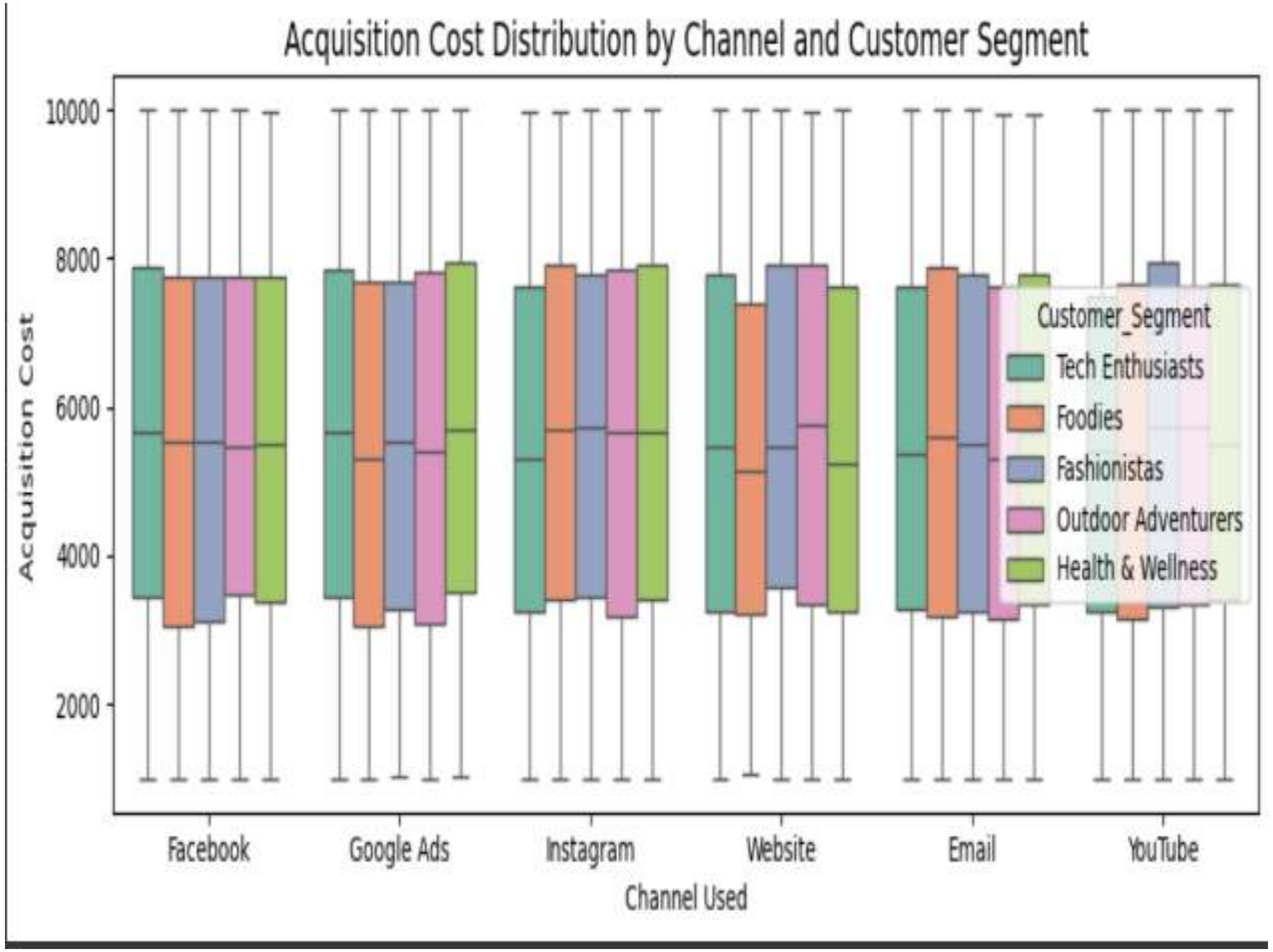
```
top_indices =
Marketing_df.groupby('Language')['Conversion_Rate'].idxmax()
top_segments = Marketing_df.loc[top_indices]
plt.figure(figsize=(10, 5))
sns.barplot(data=top_segments, x='Language',
y='Conversion_Rate', hue='Customer_Segment', palette='Set2')
plt.title("Top Customer Segment per Language by Highest
Conversion Rate")
plt.ylabel("Conversion Rate")
plt.tight_layout()
plt.show()
```



Customer Segmentation:

➢ Visualize the distribution of Acquisition_Cost across each Customer_Segment,categorized by Channel_Used, using a box plot.

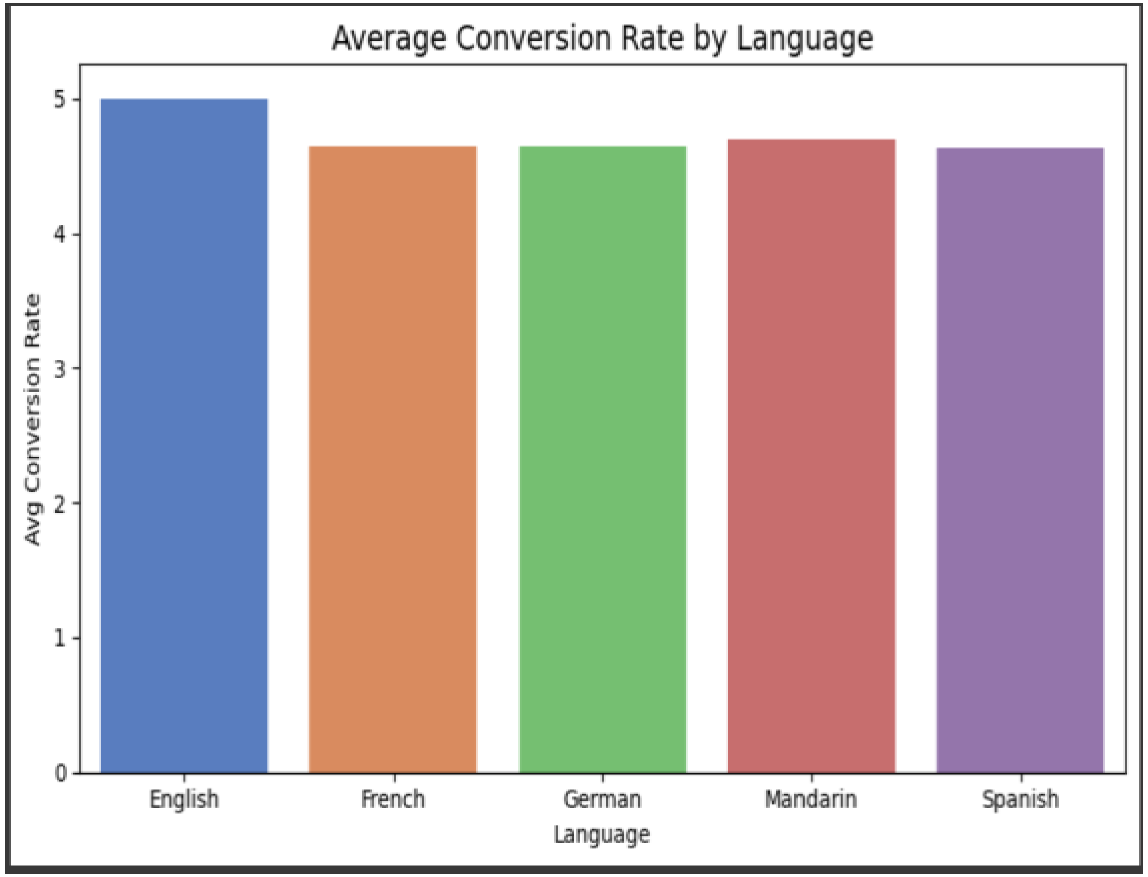
```
plt.figure(figsize=(10, 6))
sns.boxplot(
    data=Marketing_df,
    x='Channel_Used',
    y='Acquisition_Cost',
    hue='Customer_Segment',palette='Set2')
plt.title("Acquisition Cost Distribution by Channel and Customer Segment", fontsize=14)
plt.xlabel("Channel Used")
plt.ylabel("Acquisition Cost")
plt.tight_layout()
plt.show()
```



Customer Segmentation:

➢ Analyze average Conversion_Rate by Language using a bar chart to compare the effectiveness of campaigns conducted in different languages.

```
Conversion_Rate_by_Language =
Marketing_df.groupby('Language')['Conversion_Rate'].mean().reset
_index()
plt.figure(figsize=(10, 6))
sns.barplot(data=Conversion_Rate_by_Language, x='Language',
y='Conversion_Rate',hue='Language',palette='muted',legend=False)
plt.title("Average Conversion Rate by Language", fontsize=14)
plt.xlabel("Language")
plt.ylabel("Avg Conversion Rate")
plt.tight_layout()
plt.show()
```



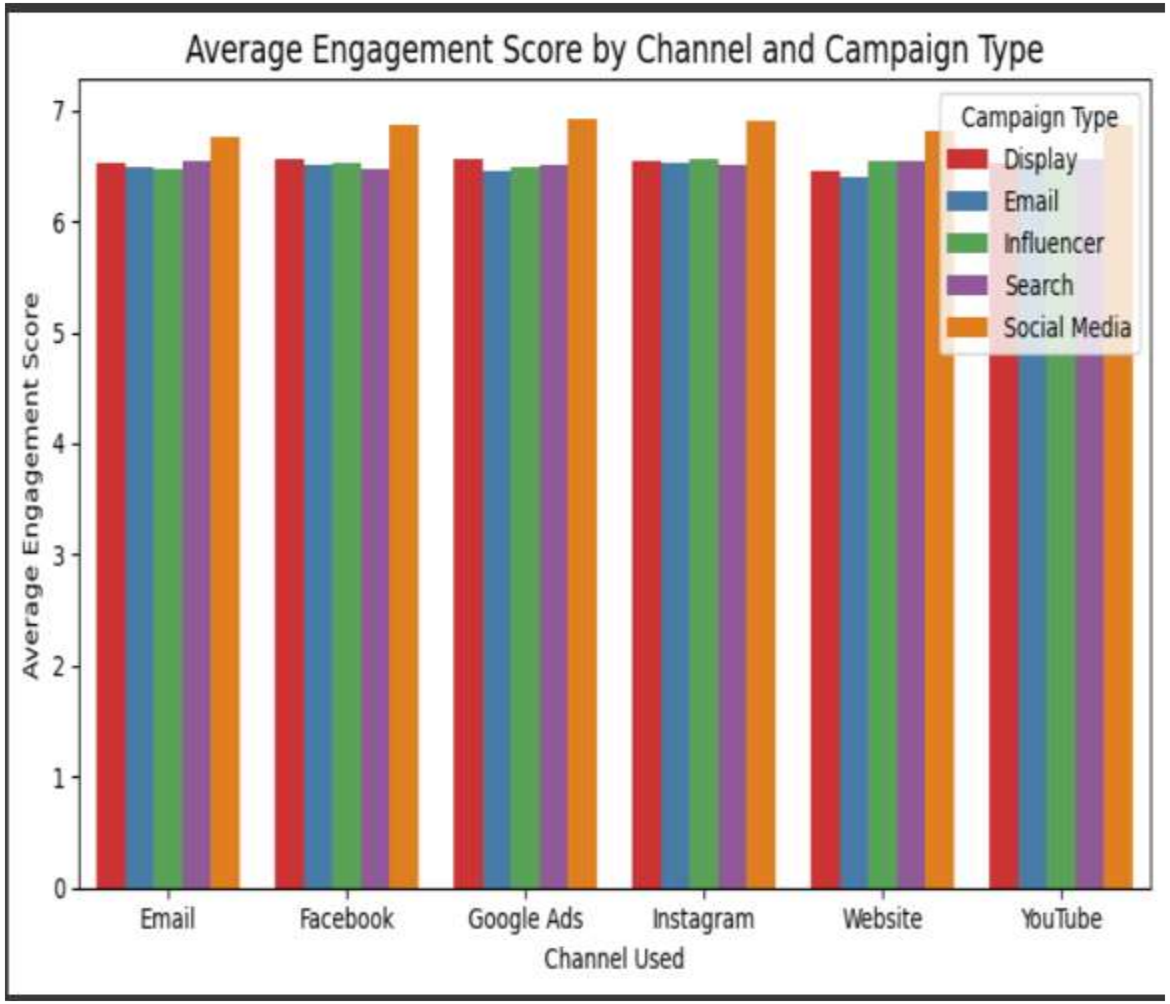
Inferences of Customer Segmentation:

- The audience is split by gender and age, with women making up the majority across categories.
- Women aged 25–34 are the top audience group (~6000), while men aged 18–24 are the lowest (~3000).
- All the Languages has the same Conversion rates as the highest for each language
- English-speaking audiences convert best among Tech Enthusiasts.
- Fashionistas dominate in French and German-speaking segments, while Outdoor Adventurers lead in Mandarin and Spanish.
- Acquisition costs are fairly consistent across channels, with some variation by customer segment.
- Google Ads and Instagram have slightly higher spread in acquisition cost across segments.
- English shows the highest average conversion rate among all languages.
- Other languages like French, German, Mandarin, and Spanish show stable but slightly lower conversion performance.

Channel Effectiveness:

➢ Compare the Engagement_Score for different Channels_Used, segmented by Campaign_Type, using a bar chart.

```
Channels_Engagement_score =
Marketing_df.groupby(['Channel_Used','Campaign_Type'])['Engagem
ent_Score'].mean().reset_index()
plt.figure(figsize=(10,5))
sns.barplot(data=Channels_Engagement_score,x='Channel_Used',y='
Engagement_Score',hue='Campaign_Type',palette='Set1')
plt.title("Average Engagement Score by Channel and Campaign
Type", fontsize=14)
plt.xlabel("Channel Used")
plt.ylabel("Average Engagement Score")
plt.legend(title='Campaign Type')
plt.tight_layout()
plt.show()
```

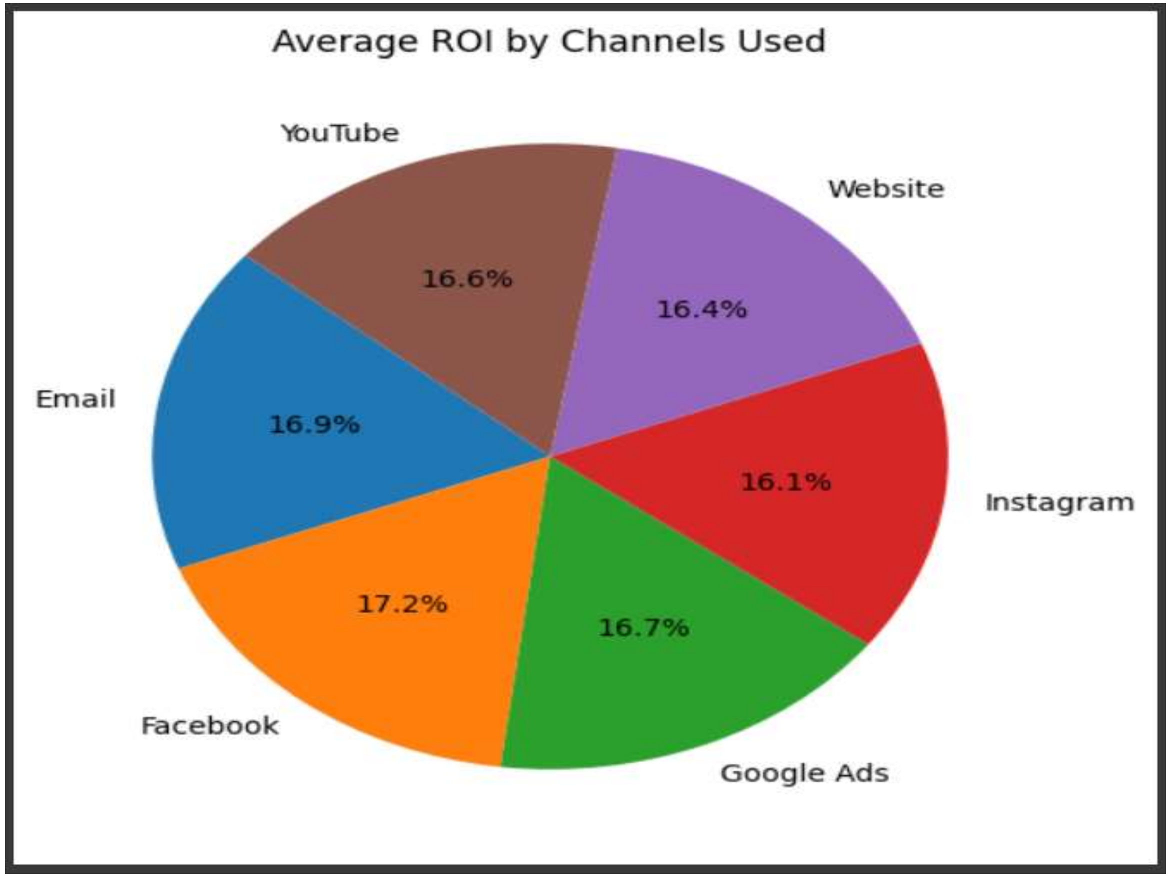


Channel Effectiveness:

➢ Show the distribution of total ROI across different Channels_Used using a pie chart.

```
roi_by_channel =
Marketing_df.groupby('Channel_Used')['ROI'].mean()

plt.figure(figsize=(8, 6))
plt.pie(roi_by_channel.values,
labels=roi_by_channel.index, autopct='%1.1f%%',
startangle=140)
plt.title("Average ROI by Channels Used")
plt.show()
```

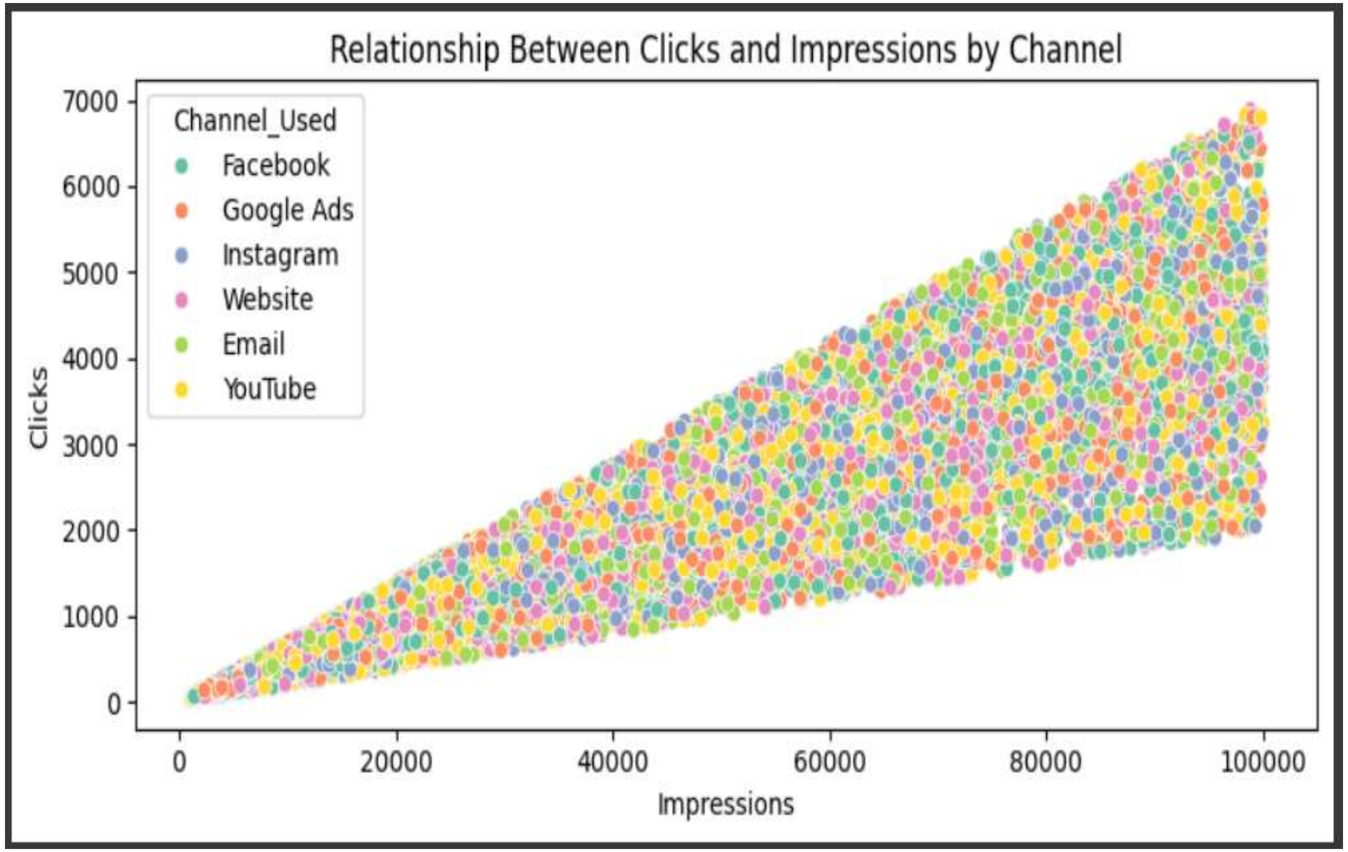


Channel Effectiveness:

➢ Plot a scatter plot to show the relationship between Clicks and Impressions for each Channel_Used.

```
plt.figure(figsize=(14, 6))


sns.scatterplot(data=Marketing_df,x='Impressions',y
='Clicks',hue='Channel_Used',palette='Set2')
plt.title("Relationship Between Clicks and
Impressions by Channel")
plt.xlabel("Impressions")
plt.ylabel("Clicks")
plt.tight_layout()
plt.show()
```



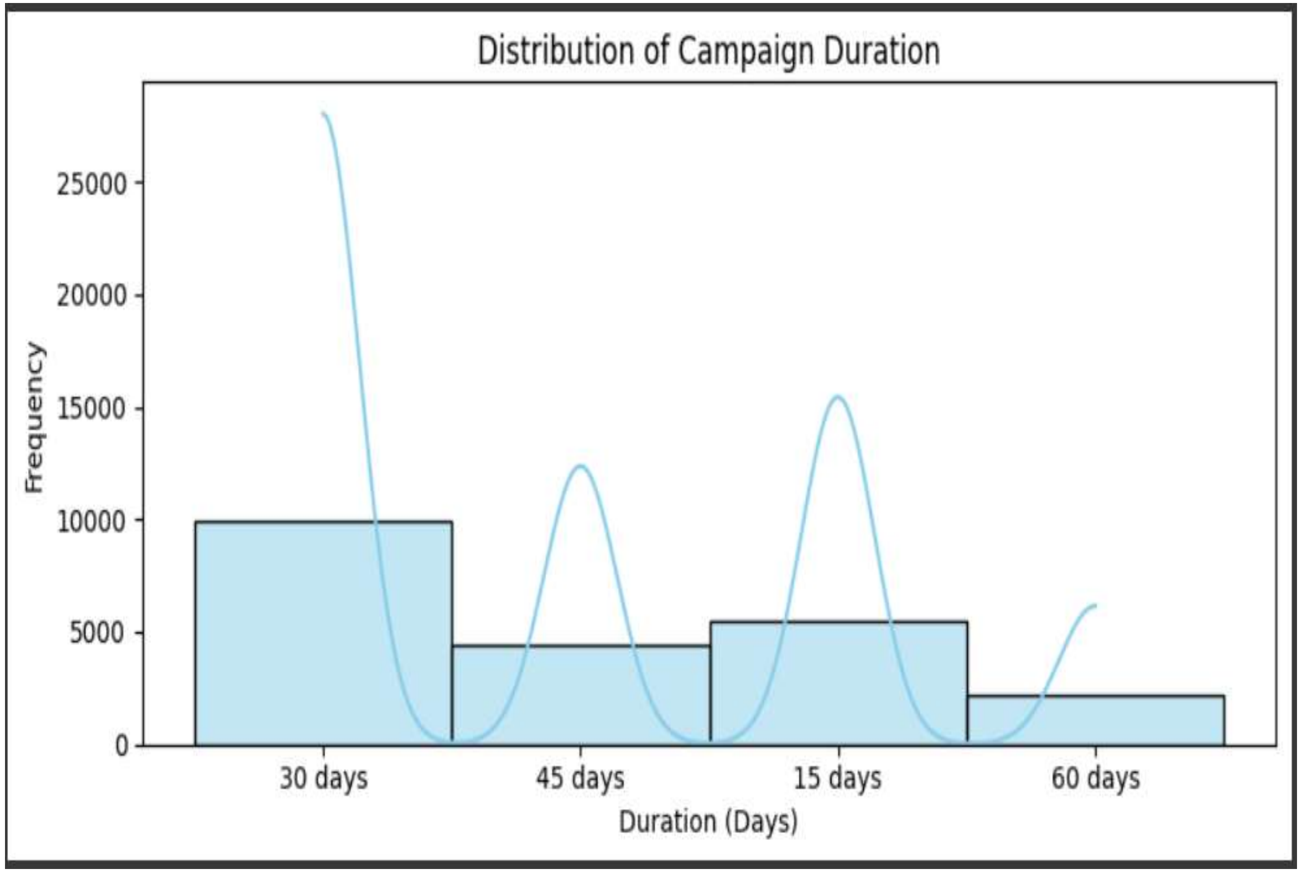
Inferences of Channel Effectiveness:

- ❖ A total of 6 marketing channels were used, and across all of them, the Social Media campaign consistently outperforms others, making it the most effective campaign type overall.
- ❖ The engagement scores remain fairly uniform across channels, with no major deviations, indicating consistent audience interaction.
- ❖ Average ROI across all channels ranges between 16% and 17.5%, with Facebook leading at 17.2% and Instagram trailing slightly at 16.1%.
- ❖ The remaining four channels fall within a narrow ROI band, suggesting stable and reliable returns across platforms.
- ❖ As observed in the pie chart, the total ROI contribution is evenly distributed among all channels, showing balanced performance.
- ❖ There is a strong positive correlation between impressions and clicks — more impressions consistently lead to more clicks.
- ❖ All channels exhibit a uniform trend in click behavior, reinforcing consistent user engagement across different platforms.

Time-Based Analysis:

➢ Plot the distribution of Duration using a histogram.

```
plt.figure(figsize=(8, 4))
sns.histplot(data=Marketing_df, x='Duration',
kde=True, bins=20, color='skyblue')
plt.title("Distribution of Campaign Duration")
plt.xlabel("Duration (Days)")
plt.ylabel("Frequency")
plt.tight_layout()
plt.show()
```

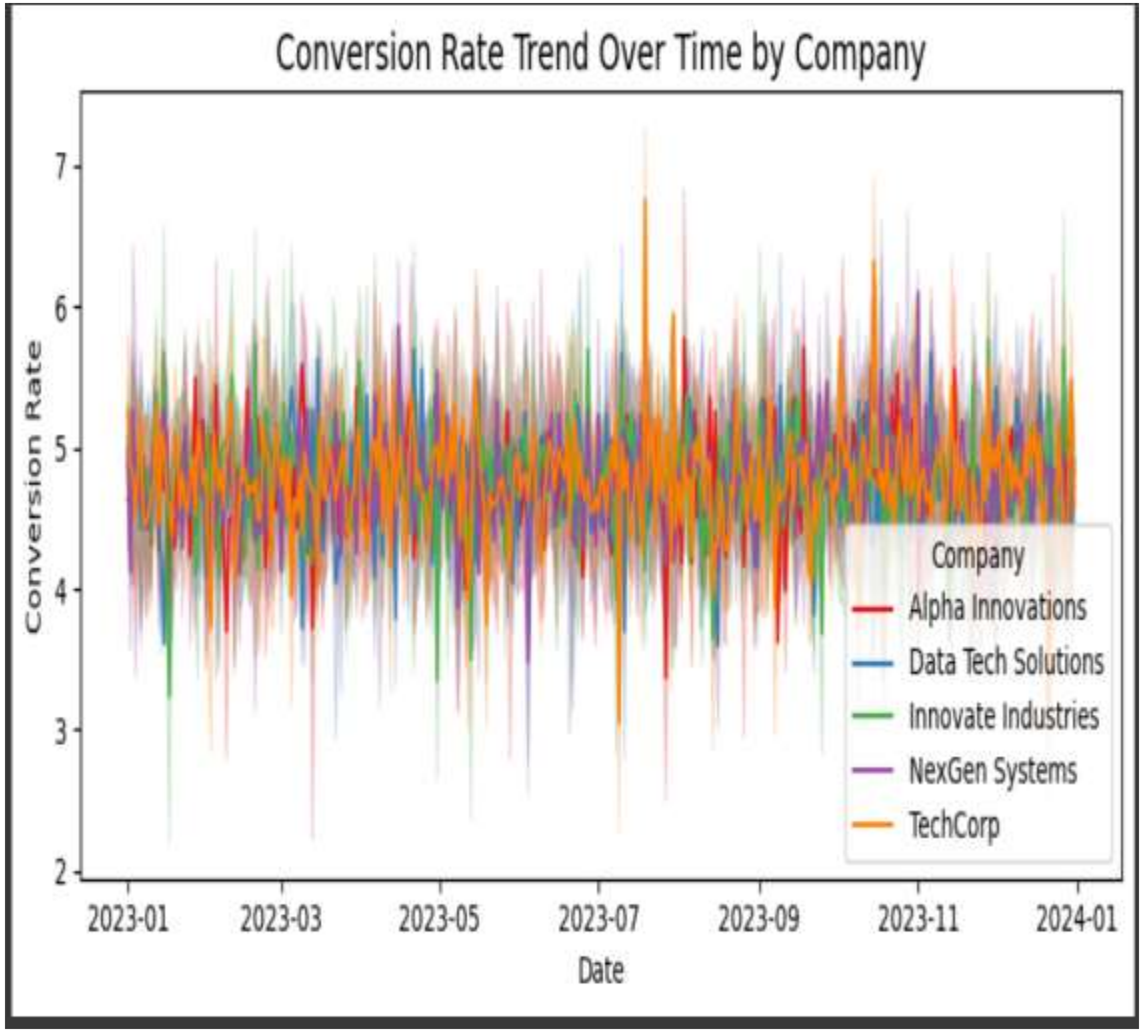


Time-Based Analysis:

➢ Analyze how the overall Conversion_Rate has changed over Date for each Company using a line chart.

```
Marketing_df['Date'] = pd.to_datetime(Marketing_df['Date'],
format='%d-%m-%Y')


Marketing_df = Marketing_df.sort_values(by=['Company', 'Date'])
plt.figure(figsize=(12, 6))
sns.lineplot(data=Marketing_df, x='Date', y='Conversion_Rate',
hue='Company', palette='Set1')
plt.title("Conversion Rate Trend Over Time by Company",
fontsize=14)
plt.xlabel("Date")
plt.ylabel("Conversion Rate")
plt.tight_layout()
plt.show()
```

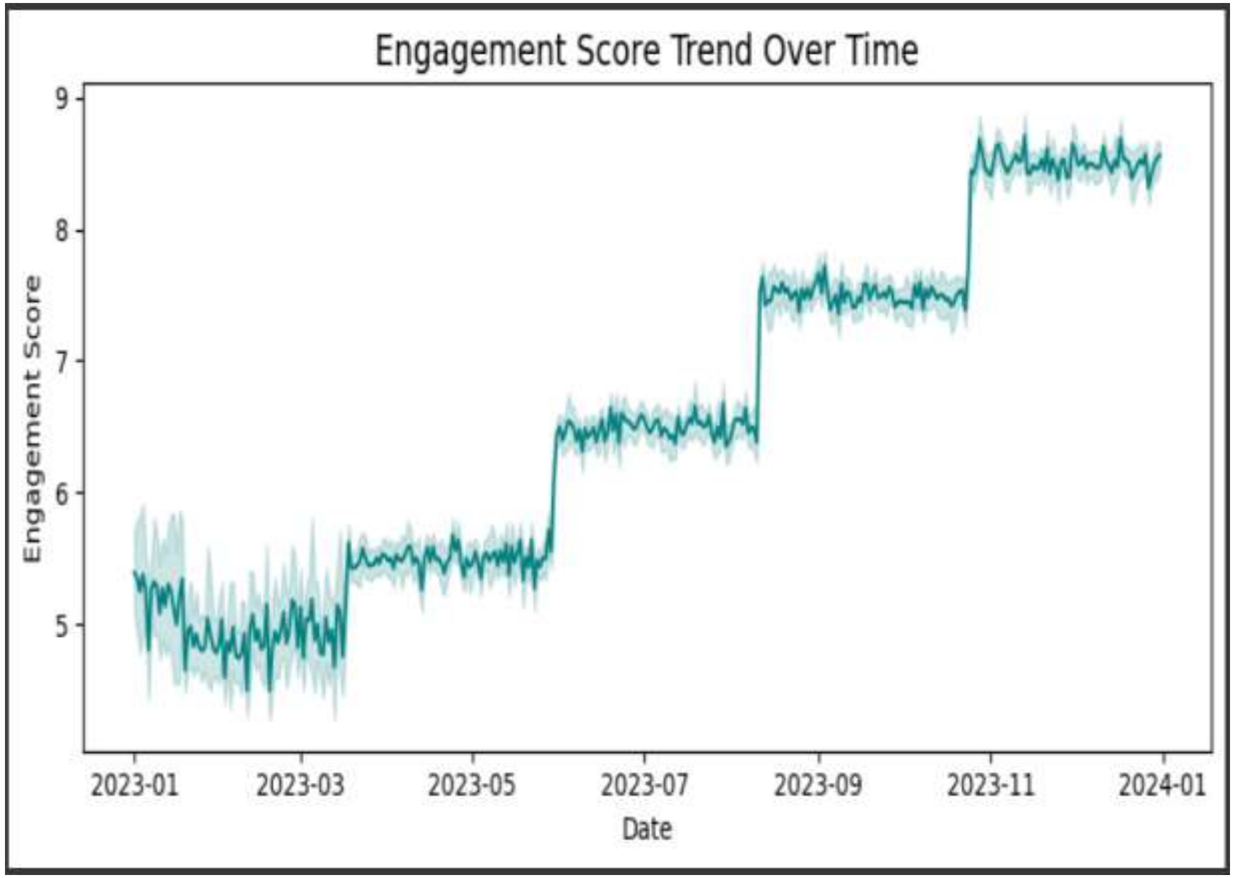


Time-Based Analysis:

➢ Examine the trend of Engagement_Score over Date with a line chart.

```
Marketing_df['Date'] = pd.to_datetime(Marketing_df['Date'],
format='%d-%m-%Y')


plt.figure(figsize=(8, 4))
sns.lineplot(data=Marketing_df, x='Date',
y='Engagement_Score', color='teal')
plt.title("Engagement Score Trend Over Time", fontsize=14)
plt.xlabel("Date")
plt.ylabel("Engagement Score")
plt.tight_layout()
plt.show()
```



Inference of Time-Based Analysis:

- ❖ Most marketing campaigns are run for 30 days, showing it's the most preferred duration, while 60-day campaigns are the least common.
- ❖ The conversion rate remains steady throughout the year for all companies, mostly staying between 4% and 6%.
- ❖ Engagement from users has increased over time, meaning more people are clicking, viewing, or interacting with the campaigns.
- ❖ There are noticeable jumps in engagement at certain points in the year, which may be due to better content, strategy changes, or special offers.
- ❖ All companies show a similar trend in conversion rate, meaning no company is performing much better or worse than the others.
- ❖ The steady rise in engagement scores shows that campaigns are slowly becoming more attractive and effective for users.
- ❖ Even though user engagement is rising, many users are still not taking the final step, like purchasing or signing up, suggesting a need to improve the last part of the user journey.

Geographic Analysis:

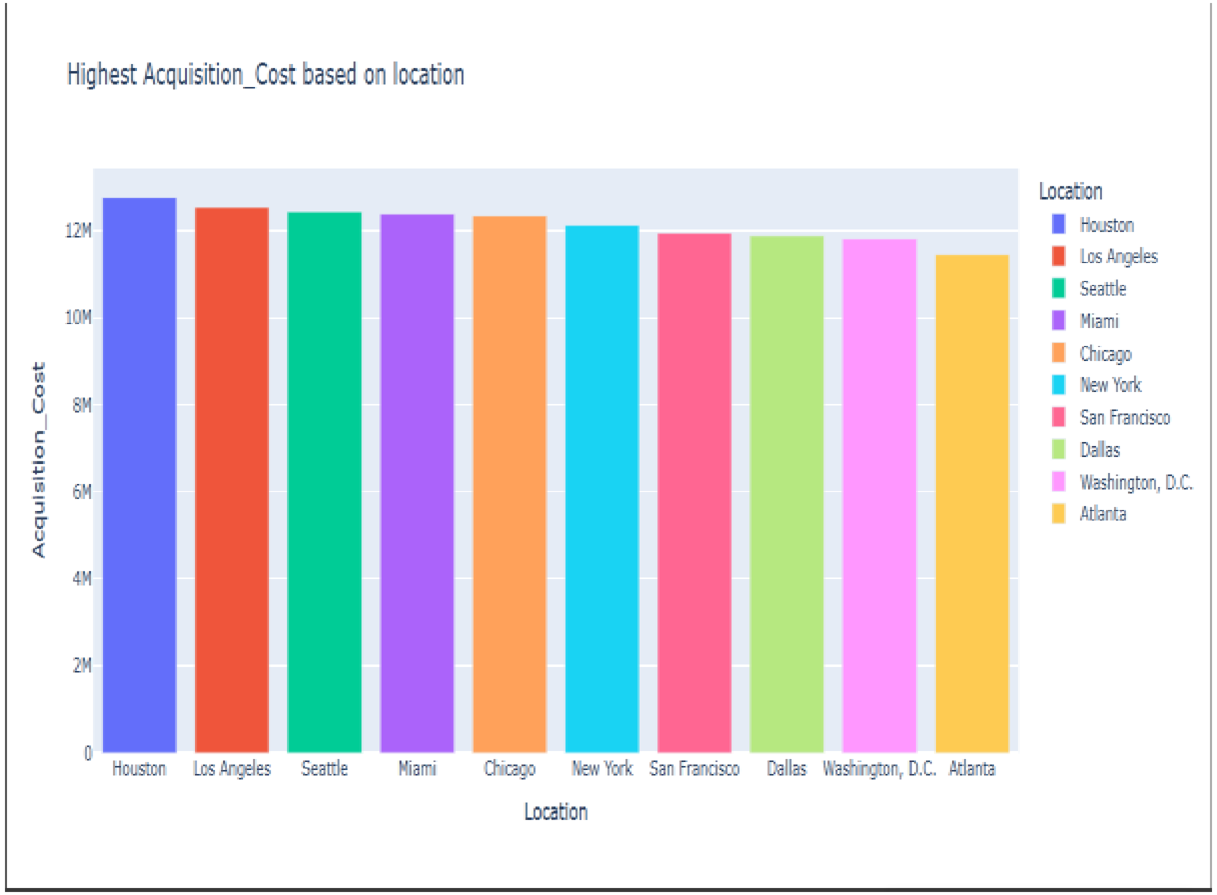
➢ Determine which location has the highest Acquisition_Cost using a bar chart.

```
import plotly.express as px

highest_cost_by_location =
Marketing_df.groupby('Location')['Acquisition_Cost'].sum().r
eset_index()

highest_cost_by_location =
highest_cost_by_location.sort_values('Acquisition_Cost',asc
ending=False)

fig=px.bar(highest_cost_by_location,x='Location',y='Acquisiti
on_Cost',title = 'Highest Acquisition_Cost based on
location',color = 'Location')
fig.show()
```

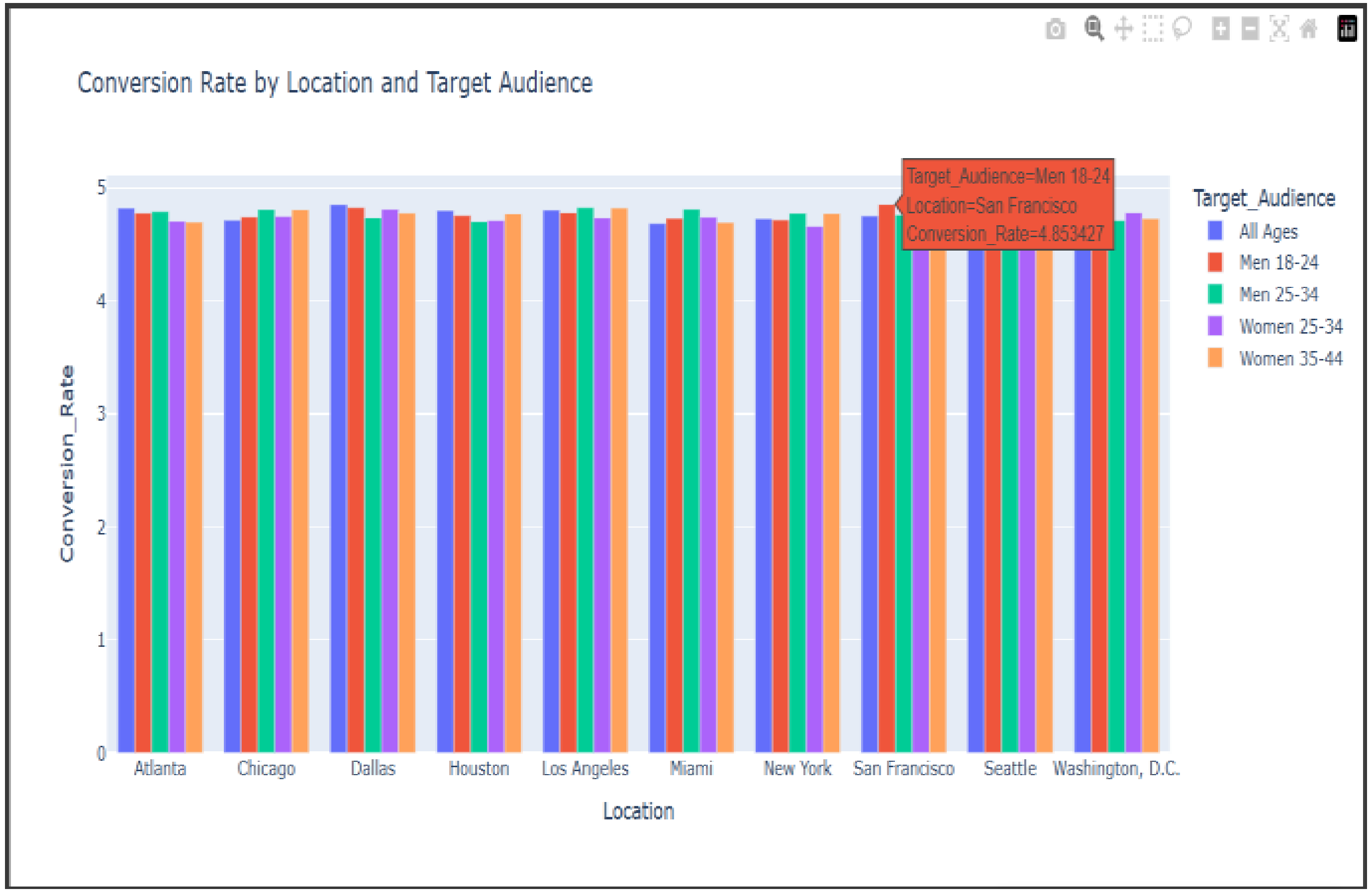


Geographic Analysis:

➢ Visualize the Conversion_Rate by different Location, categorized by Target_Audience, using a bar chart.

```
Converstion_rate_by_Location =
Marketing_df.groupby(['Location','Target_Audience'])['Conversion_Rate'].mean().reset_index()

fig=px.bar(Converstion_rate_by_Location,x='Location',y='Conversion_Rate',title = 'Conversion Rate by Location and Target Audience',barmode='group',color = 'Target_Audience')
fig.show()
```

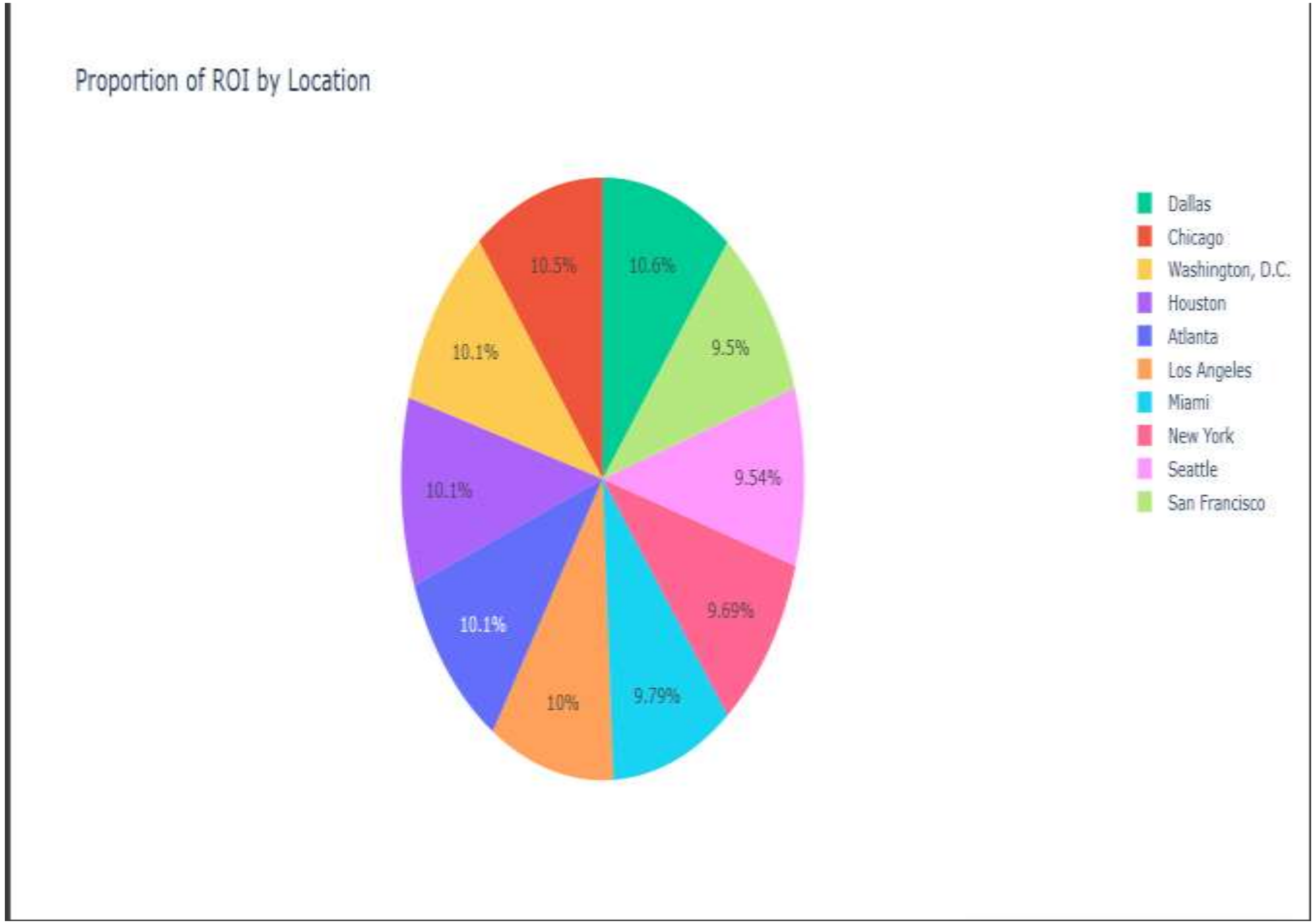


Geographic Analysis:

➢ Illustrate the proportion of ROI by Location using a pie chart.

```
ROI_WITH_LOCATION =
Marketing_df.groupby('Location')['ROI'].mean().reset_index()

fig =
px.pie(ROI_WITH_LOCATION,names='Location',values='ROI',title='Proportion of ROI by
Location',color='Location')
fig.show()
```



Inferences of Geographic Analysis:

- Houston has the highest customer acquisition cost among all locations, which means more money is spent there to get new customers compared to other cities.
- All other cities like Los Angeles, Seattle, and Miami also have high acquisition costs, but they are slightly lower than Houston's, showing that getting new customers is expensive in most major cities.
- The conversion rates across locations are quite similar, staying close to 4.7% to 4.9% for all age and gender groups, showing balanced performance in all cities.
- In San Francisco, the conversion rate for young men aged 18–24 is slightly higher than other groups, which may suggest targeted campaigns to this group could work better there.
- Even though Houston has the highest cost, it does not show the highest return on investment (ROI), meaning spending more does not always give better returns.
- Dallas and Chicago have the highest ROI among all cities, which means campaigns there bring back more value compared to the cost spent.
- Cities like San Francisco and Seattle have lower ROI compared to others, so marketers might need to improve their strategies there or consider reducing spending.

Project Summary :

This project aimed to analyze a marketing campaign dataset to uncover actionable insights into customer behavior, campaign performance, and channel effectiveness. We began by loading and exploring the dataset, followed by performing descriptive and exploratory data analysis using Python libraries like pandas, seaborn, and matplotlib.

Objectives 1

Analyze marketing campaigns across various channels and types to identify the most effective strategies in terms of conversion rate, engagement score, and ROI.

Objectives 2

Discover key audience demographics and customer segments that drive higher engagement and conversions, segmented by age, gender, and language preferences.

Objectives 3

Examine the relationship between acquisition cost and ROI to identify cost-effective marketing channels and uncover patterns in clicks, impressions, and returns across platforms.

THANK You