

Business Problem

As a marketing agency, our primary objective is to maximize the return on investment (ROI) for our clients' advertising campaigns. We have conducted two ad campaigns, one on Facebook and the other on AdWords, and we need to determine which platform yields better results in terms of clicks, conversions, and overall cost-effectiveness. By identifying the most effective platform, we can allocate our resources more efficiently and optimize our advertising strategies to deliver better outcomes for our clients.

Research Question

Which ad platform is more effective in terms of conversions, clicks, and overall cost-effectiveness?

Importing Libraries

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import datetime as dt
import math
from sklearn.metrics import r2_score, mean_squared_error
from sklearn.metrics import mean_squared_log_error
from sklearn.metrics import mean_squared_error
import warnings
warnings.filterwarnings('ignore')
print("Everything works fine")

everything works fine
```

Data Description

The dataset comprises a collection of data comparing the performance of two separate ad campaigns conducted throughout the year 2019. Specifically, the data covers a Facebook Ad campaign and an AdWords Ad campaign. For each day of the year 2019, there is a corresponding row in the dataset, resulting in a total of 365 lines of campaign data to analyze. The dataset includes various performance metrics for each ad campaign, providing insights into their effectiveness and efficiency over time.

Key features included in the dataset are as follows:

- **Date:** The date corresponding to each row of campaign data, ranging from January 1st, 2019, to December 31st, 2019.
- **Ad Views:** The number of times the ad was viewed.
- **Ad Clicks:** The number of clicks received on the ad.
- **Ad Conversions:** The number of conversions resulting from the ad.
- **Cost per Ad:** The cost associated with running the Facebook ad campaign.
- **Click-Through Rate (CTR):** The ratio of clicks to views, indicating the effectiveness of the ad in generating clicks.
- **Conversion Rate:** The ratio of conversions to clicks, reflecting the effectiveness of the ad in driving desired actions.
- **Cost per Click (CPC):** The average cost incurred per click on the ad.

```
In [2]: # loading the dataset
df = pd.read_csv("advertising_campaign.csv")

In [3]: # data overview
df

Out[3]:
   Date      Facebook Ad Views  Facebook Ad Clicks  Facebook Ad Conversions  Cost per Facebook Ad  Facebook Click-Through Rate (Clicks / View)  Facebook Conversion Rate (Conversions / Click)  AdWords Ad Views  AdWords Ad Clicks  AdWords Ad Conversions  Cost per AdWords Ad  AdWords Click-Through Rate (Clicks / View)  AdWords Conversion Rate (Conversions / Click)  AdWords Cost per Click (Ad Cost / Click)
0  2019-01-01      2156         16          18          $104         0.83%         42.73%         $7.14  4864         59         5         $194         1.18%         6.47%         $3.30
1  2019-01-02      3106         36          12          $102         1.15%         34.04%         $6.38562         4022         71         6         $75         1.77%         7.80%         $1.05
2  2019-01-03      3105         28          8          $102         0.84%         37.45%         $3.69  3863         44         4         $141         1.13%         9.59%         $2.23
3  2019-01-04      1107         27          9          $71         2.45%         34.76%         $2.62  3911         49         5         $141         1.26%         11.08%         $2.86
4  2019-01-05      1317         15          7          $78         1.10%         47.59%         $5.38  4070         55         7         $133         1.36%         12.22%         $2.40

In [4]: # retrieving columns count of the dataset
df.shape

Out[4]:
(365, 17)

In [5]: # datatypes of the columns
df.dtypes

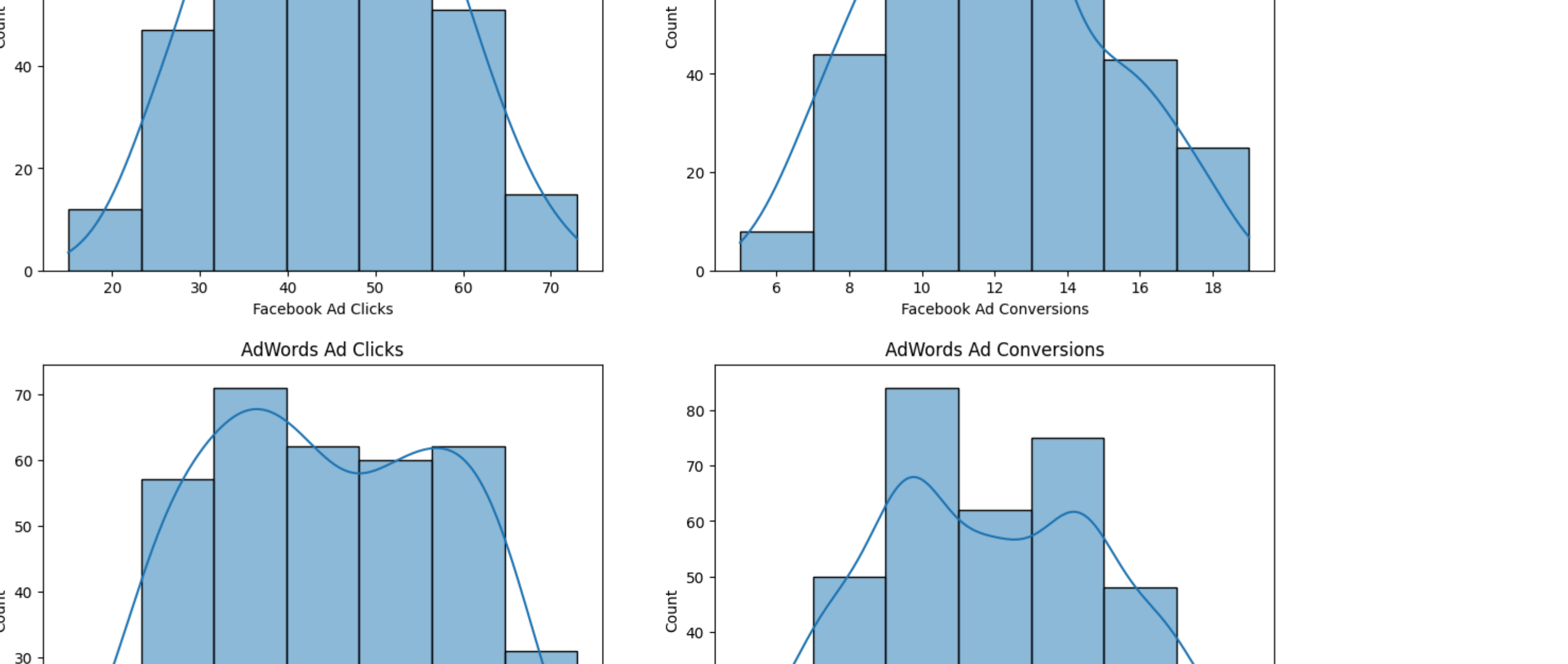
Out[5]:
Date                                object
Facebook Ad Campaign                int64
Facebook Ad Views                   int64
Facebook Ad Clicks                  int64
Facebook Ad Conversions              int64
Cost per Facebook Ad                object
Facebook Click-Through Rate (Clicks / View)  object
Facebook Conversion Rate (Conversions / Clicks)  object
Facebook Cost per Click (Ad Cost / Clicks)      object
AdWords Ad Views                    int64
AdWords Ad Clicks                   int64
AdWords Ad Conversions              int64
AdWords Cost per Click (Ad Cost / Clicks)      object
AdWords Click-Through Rate (Clicks / View)      object
AdWords Conversion Rate (Conversions / Click)    object
AdWords Cost per Click (Ad Cost / Clicks)      object
dtype: object

In [6]: # converting date to datetime
df["date"] = pd.to_datetime(df["date"])

In [7]: # descriptive stats of the campaigns
df.describe()
```

```
Out[7]:
   Date      Facebook Ad Views  Facebook Ad Clicks  Facebook Ad Conversions  AdWords Ad Views  AdWords Ad Clicks  AdWords Ad Conversions
count  365                365.000000          365.000000          365.000000          365.000000          365.000000          365.000000
mean    2019-07-02 00:00:00          2179.687671          44.048315          11.742466          4717.19726         60.383562          5.980222
min     2019-01-01 00:00:00          1050.000000          15.000000          5.000000          3714.000000          31.000000          3.000000
25%    2019-04-02 00:00:00          1656.000000          35.000000          10.000000          4247.000000          49.000000          5.000000
50%    2019-07-02 00:00:00          2020.000000          43.000000          12.000000          4711.000000          60.000000          6.000000
75%    2019-10-01 00:00:00          2717.000000          54.000000          13.000000          5180.000000          73.000000          7.000000
max     2019-12-31 00:00:00          3520.000000          73.000000          19.000000          5760.000000          89.000000          9.000000
std      NaN                618.074659          12.140559          2.624788          561.11406         14.366225          1.628106

In [8]: # distribution of the clicks and conversions
plt.figure(figsize=(15,6))
plt.subplot(1,2,1)
plt.title('Facebook Ad Clicks')
sns.histplot(df['Facebook Ad Clicks'], bins = 7, edgecolor = 'k', kde = True)
plt.subplot(1,2,2)
plt.title('Facebook Ad Conversions')
sns.histplot(df['Facebook Ad Conversions'], bins = 7, edgecolor = 'k', kde = True)
plt.show()
```



All the histograms are showing somewhat symmetrical shape. This symmetrical shape suggests that the number of clicks and conversions is relatively evenly distributed. In other words, there are not many clicks or conversions that are outliers on either the high or low end, even though we have outliers we can't remove them as this database contains data of 365 days for analysis.

How frequently do we observe days with high numbers of conversions compared to days with low numbers of conversions?

```
In [9]: # creating function to calculate the category for the conversions
def create_conversion_category(conversion_count):
    category = 1
    if conversion < 6:
        category.append("less than 6")
    elif 6 <= conversion < 11:
        category.append("6 - 10")
    elif 11 <= conversion < 16:
        category.append("more than 15")
    else:
        category.append("more than 15")
    return category

# applying function of different campaign's conversions
df["Facebook Conversion Category"] = create_conversion_category("Facebook Ad Conversions")
df["AdWords Conversion Category"] = create_conversion_category("AdWords Ad Conversions")

In [10]: df[["Facebook Ad Conversions", "Facebook Conversion Category", "AdWords Ad Conversions", "AdWords Conversion Category"]].head()
```

	Facebook Ad Conversions	Facebook Conversion Category	AdWords Ad Conversions	AdWords Conversion Category
0	8	6 - 10	5	less than 6
1	12	10 - 15	6	6 - 10
2	8	6 - 10	4	less than 6
3	9	6 - 10	5	less than 6
4	7	6 - 10	7	6 - 10

```
In [11]: # here we are counting the number days when conversion happened in w
df["Facebook Conversion Category"].value_counts()
```

Facebook Conversion Category	count
6 - 10	189
6 - 10	128
more than 15	47
less than 6	1
more count, dtype: int64	

```
Out [12]: # dataFrame(df["Facebook Conversion Category"].value_counts())
facebook

0      6 - 10  189
1      6 - 10  128
2    more than 15   47
3    less than 6    1
```

```
In [13]: df[["AdWords Conversion Category"].value_counts()]

Out [13]:
AdWords Conversion Category
6 - 10      209
less than 6    1
more count, dtype: int64
```

```
In [14]: # dataFrame(df["AdWords Conversion Category"].value_counts())
adwords

Out [14]:
Category    count
0      6 - 10    209
1    less than 6    156
```

```
In [15]: category_df = pd.merge(facebook, adwords, on = 'Category', how = 'outer', fillna=0)
category_df
```

Category	count_x	count_y
0	10 - 15 189	0.0
1	6 - 10 128	209.0
2	less than 6 1	156.0
3	more than 15 47	0.0

```
In [16]: # we can't present data to stakeholder like this
# we need to show it in correct way and using graphs so first will sort the data in asc order
category_df = category_df.sort_values(["x", "y"])
category_df
```

Category	count_x	count_y
3	more than 15 47	0.0
1	6 - 10 128	209.0
0	10 - 15 189	0.0
2	less than 6 1	156.0

```
In [17]: x_axis = np.arange(len(category_df))
x_axis
```

```
Out [17]: array([0, 1, 2, 3])

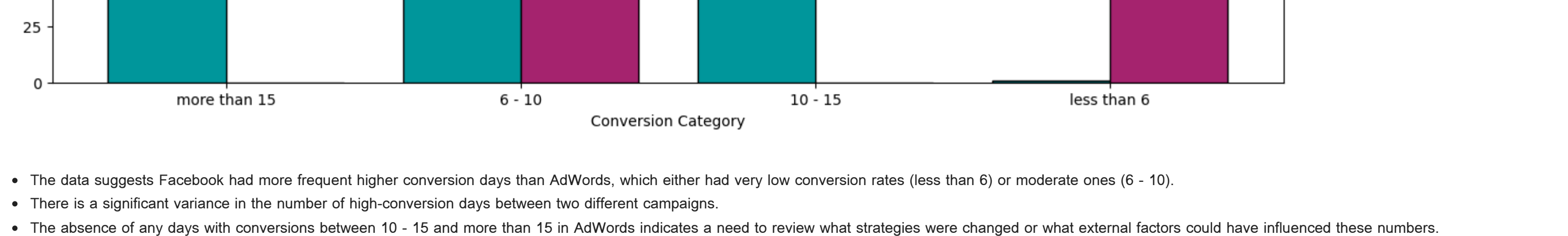
In [18]: x_axis = np.arange(len(category_df))
plt.figure(figsize=(15,6))
plt.bar(x_axis, y = category_df["count_x"], 0.4, label = "Facebook", color = "#039898", linewidth = 1, edgecolor = "k")
plt.bar(x_axis, y = category_df["count_y"], 0.4, label = "Adwords", color = "#A6237A", linewidth = 1, edgecolor = "k")
plt.xticks(x_axis, category_df["Category"])
plt.xlabel("Conversion Category")
plt.ylabel("Number of days")
plt.title("Frequency of Daily Conversions by Conversion Categories", fontsize = 15)
plt.legend(fontsize = 15)
plt.show()
```



- The data suggests Facebook had more frequent higher conversion days than AdWords, which either had very low conversion rates (less than 6) or moderate ones (6 - 10).
- There is a significant variance in the number of high-conversion days between two different campaigns.
- The absence of any days with conversions between 10 - 15 and more than 15 in AdWords indicates a need to review what strategies were changed or what external factors could have influenced these numbers.

Do more clicks on the ad really lead to more sales?

```
In [19]: plt.figure(figsize=(15,6))
plt.subplot(1,2,1)
plt.title('Facebook')
sns.scatterplot(df['Facebook Ad Clicks'], y = df['Facebook Ad Conversions'], color = "#039898")
plt.xlabel('Clicks')
plt.ylabel('Conversions')
plt.subplot(1,2,2)
plt.title('AdWords')
sns.scatterplot(df['AdWords Ad Clicks'], y = df['AdWords Ad Conversions'], color = "#039898")
plt.xlabel('Clicks')
plt.ylabel('Conversions')
plt.show()
```



```
In [20]: facebook_corr = df[["Facebook Ad Conversions", "Facebook Ad Clicks"]].corr()
facebook_corr
```

	Facebook Ad Conversions	Facebook Ad Clicks
Facebook Ad Conversions	1.000000	0.873775
Facebook Ad Clicks	0.873775	1.000000

```
In [21]: adwords_corr = df[["AdWords Ad Conversions", "AdWords Ad Clicks"]].corr()
adwords_corr
```

	AdWords Ad Conversions	AdWords Ad Clicks
AdWords Ad Conversions	1.000000	0.447993
AdWords Ad Clicks	0.447993	1.000000

```
In [22]: print("Correlation Coeff: {}".format(facebook_corr.values[0,1,2]))
print("Facebook: ", round(facebook_corr.values[0,1,2]))
print("AdWords: ", round(adwords_corr.values[0,1,2]))

Correlation Coeff:
facebook: 0.87
adwords: 0.45
```

- A correlation coefficient of 0.87 indicates a strong positive linear relationship between clicks on Facebook ads and sales. This suggests that as the number of clicks on Facebook ads increases, sales tend to increase as well.
- This strong correlation suggests that Facebook ads are highly effective in driving sales, as a large portion of the variation in sales can be explained by the variation in clicks on Facebook ads.
- A correlation coefficient of 0.45 indicates a moderate positive linear relationship between clicks on AdWords ads and sales. While there is still a positive relationship, it is not as strong as with Facebook ads.
- The moderate correlation between clicks on AdWords ads and sales indicates that while AdWords advertising does contribute to sales, its effectiveness may be influenced by other factors. Further analysis is needed to identify these factors and optimize AdWords campaigns accordingly.

Hypothesis Testing

Hypothesis: Advertising on Facebook will result in a greater number of conversions compared to advertising on AdWords.

Null Hypothesis (H0): There is no difference in the number of conversions between Facebook and AdWords, or the number of conversions from AdWords is greater than or equal to those from Facebook.

H0: $\mu_{Facebook} \leq \mu_{AdWords}$

Alternate Hypothesis (H1): The number of conversions from Facebook is greater than the number of conversions from AdWords.

H1: $\mu_{Facebook} > \mu_{AdWords}$

```
In [23]: print("Mean Conversion: {}".format(facebook_corr.values[0,1,2]))
print("Facebook: ", round(facebook_corr.values[0,1,2]))
print("AdWords: ", round(adwords_corr.values[0,1,2]))

t_stat, p_value = st.ttest_ind(df["Facebook Ad Conversions"], df["AdWords Ad Conversions"], equal_var = False)
print("t-statistic: ", t_stat, "p-value: ", p_value)

# comparing the p-value with the significance level of 0.05
if p_value < 0.05:
    print("p-value is less than significance value, Reject the null hypothesis")
else:
    print("p-value is greater than significance value, Accept the null hypothesis")

Mean Conversion
Facebook: 11.74
AdWords: 5.98
T statistic: 32.8842046798184
p-value: 3.498312433049e-136
p-value is less than significance value, Reject the null hypothesis
```

- The mean number of conversions from Facebook ads (11.74) is substantially higher than the mean number of conversions from AdWords ads (5.98). This suggests that, on average, Facebook advertising is more effective in generating conversions compared to AdWords advertising.
- The T statistic (32.88) is a measure of the difference between the means of the two groups relative to the variation within the groups. A larger T statistic indicates a greater difference between the means of the two groups.
- The p-value (9.35e-136) is extremely small, indicating strong evidence against the null hypothesis.
- The results strongly support the alternate hypothesis, indicating that the number of conversions from Facebook advertising is indeed greater than the number of conversions from AdWords advertising.
- Facebook advertising appears to be a more effective channel for generating conversions compared to AdWords advertising, based on the sample data analyzed.
- Given the significant difference in conversion rates between Facebook and AdWords, consider reallocating resources towards Facebook advertising efforts. This could involve increasing ad spend, expanding targeting efforts, or experimenting with different ad formats to capitalize on the platform's effectiveness in driving conversions.

Regression Analysis

What will happen when I do go with the Facebook Ad? How many facebook ad conversions can I expect given a certain number of facebook ad clicks?

```
In [24]: # independent variable
x = df[["Facebook Ad Clicks"]]

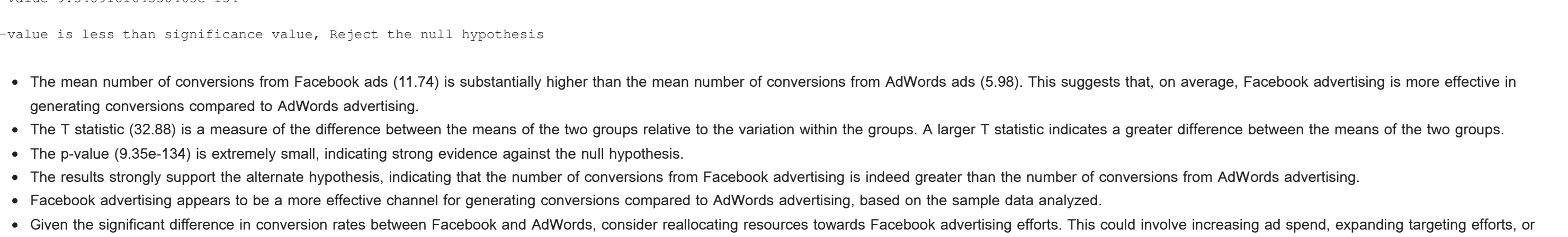
# dependent variable
y = df[["Facebook Ad Conversions"]]

# initializing and fitting linear Regression model
reg_model = LinearRegression()
reg_model.fit(x,y)
prediction = reg_model.predict(x)
```

```
# model evaluation
r2 = r2_score(y, prediction)*100
mse = mean_squared_error(y, prediction)
print("Accuracy (R2 score): {}".format(r2))
print("Mean Squared Error: {}".format(mse))

Accuracy (R2 score): 76.35 %
Mean Squared Error: 2.02
```

```
In [25]: plt.figure(figsize=(8,6))
plt.scatter(x=df["Facebook Ad Clicks"], y=df["Facebook Ad Conversions"], color = "#039898", label = "Actual data points")
plt.plot(df["Facebook Ad Clicks"], prediction, color = "#A6237A", label = "Best fit line")
plt.legend()
plt.show()
```



```
In [26]: print("For 10 Clicks, Expected Conversion: {}".format(reg_model.predict([10])[0][0,2]))
print("For 18 Clicks, Expected Conversion: {}".format(reg_model.predict([18])[0][0,2]))

For 10 Clicks, Expected Conversion: 13.0
For 18 Clicks, Expected Conversion: 19.3
```

- The model has a reasonably good predictive power, with an R2 score of 76.35%. This suggests that it can effectively predict Facebook ad conversions based on the number of Facebook ad clicks.
- The regression line provided by the Linear Regression model, businesses can make informed decisions about resource allocation, budget planning, and campaign optimization.
- For instance, knowing the expected number of Facebook ad conversions based on a certain number of Facebook ad clicks can help in setting realistic campaign goals, optimizing ad spend, and assessing the ROI of Facebook advertising efforts.

Analyzing Facebook Campaign metrics over time.

```
In [27]: # cleaning data / removing unwanted symbols from the columns and converting them to numeric columns
df["Facebook Click-Through Rate (Clicks / View)"] = df["Facebook Click-Through Rate (Clicks / View)"].apply(lambda x: float(x[1:-1]))
df["Facebook Conversion Rate (Conversions / Clicks)"] = df["Facebook Conversion Rate (Conversions / Clicks)"].apply(lambda x: float(x[1:-1]))
df["Facebook Cost per Click (Ad Cost / Clicks)"] = df["Facebook Cost per Click (Ad Cost / Clicks)"].apply(lambda x: float(x[1:-1]))

In [28]: # filtering for Facebook campaign
df = df[["Date", "Facebook Ad Views",
        "Facebook Ad Clicks", "Facebook Ad Conversions", "Cost per Facebook Ad",
        "Facebook Click-Through Rate (Clicks / View)", "Facebook Conversion Rate (Conversions / Clicks)",
        "Facebook Cost per Click (Ad Cost / Clicks)"]]
df.head()
```

```
Out [28]:
   Date      Facebook Ad Views  Facebook Ad Clicks  Facebook Ad Conversions  Cost per Facebook Ad  Facebook Click-Through Rate (Clicks / View)  Facebook Conversion Rate (Conversions / Click)  Facebook Cost per Click (Ad Cost / Click)
0  2019-01-01      2156         16          18          $104         0.83%         42.73%         $7.14
1  2019-01-02      3106         36          12          $102         1.15%         34.04%         $6.38562
2  2019-01-03      3105         28          8          $102         0.84%         37.45%         $3.69
3  2019-01-04      1107         27          9          $71         2.45%         34.76%         $2.62
4  2019-01-05      1317         15          7          $78         1.10%         47.59%         $5.38
```

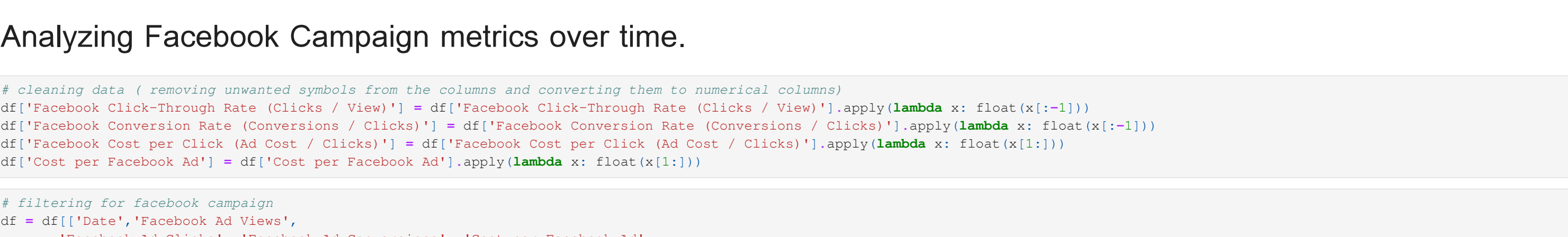
At what times of the week or days of the week do we observe the conversions?

```
In [29]: # extracting month and week day from the date column
df["month"] = df["date"].dt.month
df["week"] = df["date"].dt.weekday
```

```
In [30]: plt.figure(figsize=(8,5))
monthly_conversion = df.groupby("month")["Facebook Ad Conversions"].sum()
monthly_names = ["Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"]
plt.plot(monthly_names, monthly_conversion["Facebook Ad Conversions"], color = "#039898", edgecolor = "k")
plt.show()
```



```
In [31]: plt.figure(figsize=(8,5))
plt.title("Monthly Conversions")
monthly_conversion = df.groupby("month")["Facebook Ad Conversions"].sum()
monthly_names = ["Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"]
plt.plot(monthly_names, monthly_conversion["Facebook Ad Conversions"], color = "#039898", edgecolor = "k")
plt.show()
```



- Across the weekdays over a year, the total number of conversions remains relatively consistent, indicating a consistent level of engagement throughout the week. However, Mondays and Tuesdays consistently exhibit the highest conversion rates compared to other days of the week, suggesting that the beginning of the workweek sees heightened user engagement or responsiveness to marketing efforts.
- Examining the monthly trend in conversions reveals an overall upward trajectory, indicating a general increase in conversions over time. However, certain months stand out with variations in conversion rates. February, April, May, June, August, and November experience a decline in conversions compared to neighboring months. These periods of decreased conversion rates could be influenced by factors such as seasonal fluctuations, changes in consumer behavior, or adjustments in marketing strategies.

How does the Cost Per Conversion (CPC) trend over time?

Cost Per Conversion (CPC): This metric is used to evaluate the cost effectiveness and profitability of an online advertising campaign. This metric helps marketers understand how much they are spending to obtain each conversion, allowing them to optimize their spending and targeting strategies effectively.

```
In [32]: plt.figure(figsize=(8,5))
plt.title("Monthly Cost Per Conversion (CPC)")
monthly_cpc = df.groupby("month")["Facebook Ad Conversions", "Cost per Facebook Ad"].sum()
monthly_df["Cost per Conversion"] = monthly_df["Cost per Facebook Ad"] / monthly_df["Facebook Ad Conversions"]
monthly_names = ["Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"]
plt.plot(monthly_names, monthly_cpc["Cost per Conversion"], color = "#A6237A", edgecolor = "k")
plt.show()
```



- The CPC trend over the 12-month period shows some fluctuations but overall maintains a relatively stable range.
- May and November have the lowest CPC values, indicating potentially more cost-effective advertising or higher conversion rates during these periods.
- February has the highest CPC value, suggesting that advertising costs may be relatively higher during the month compared to others.
- Lower CPC values in certain months (e.g., May and November) could indicate periods of higher advertising effectiveness or more favorable market conditions.
- Consider allocating more advertising budget to months with historically lower CPC values (e.g., May and November) to maximize ROI.

Is there a long-term equilibrium relationship between advertising spend and conversion rates that suggests a stable, proportional impact of budget changes on conversions over time?

```
In [33]: score_p_value, _ = stats.ttest_ind(df["Cost per Facebook Ad"], df["Facebook Ad Conversions"])
print("Correlation test score: {}".format(score))
print("P-value: ", p_value)
if p_value < 0.05:
    print("p-value is less than significance value, Reject the null hypothesis")
else:
    print("p-value is greater than significance value, Accept the null hypothesis")

Correlation test score: -14.75542838510219
P-value: 2.13373757901323e-26
p-value is less than significance value, Reject the null hypothesis
```

- Since the p-value is significantly lower than the chosen significance level, we reject the null hypothesis. This indicates that there is a long-term equilibrium relationship between advertising spend (cost) and conversions.
- Businesses can use the understanding of the stable relationship between advertising spend and conversions to optimize their advertising strategies. By investing in campaigns that demonstrate a strong return on investment (ROI) and adjusting spending based on performance, companies can maximize conversions while minimizing costs.

```
In [ ]:
```