Conversion Rate Optimization for XYZ Company's Advertising Campaign.

Conversion rate optimization (CRO) is the process of increasing the percentage of conversions from a website or mobile app.

The conversion rate is always a challenging task for every company, here in xyz's company we will look into the every aspect of the conversion rate analysis.

Here is the data.

	ad_id	xyz_campaign_id	fb_campaign_id	age	gender	interest	Impressions	Clicks	Spent	Total_Conversion	Approved_Conversion
0	708746	916	103916	30-34	M	15	7350	1	1.43	2	1
1	708749	916	103917	30-34	M	16	17861	2	1.82	2	0
2	708771	916	103920	30-34	M	20	693	0	0.00	1	0
3	708815	916	103928	30-34	M	28	4259	1	1.25	1	0
4	708818	916	103928	30-34	M	28	4133	1	1.29	1	1

Data Description.

ad_id: a unique ID for each ad.
xyz_campaign_id: an ID associated with each ad campaign of XYZ company.
fb_campaign_id: an ID associated with how Facebook tracks each campaign.
age: age of the person to whom the ad is shown.
gender: gender of the person to whom the ad is shown.
interest: a code specifying the category to which the person's interest belongs (interests are as mentioned in the person's Facebook public profile).
Impressions: the number of times the ad was shown.
Clicks: number of clicks on for that ad.
Spent: Amount paid by XYZ company to Facebook, to show that ad.
Total_conversion: Total number of people who inquired about the product after seeing the ad.
Approved_conversion: Total number of people who bought the product after seeing the ad.

```
Column: Impressions
Mean: 186732.1329833771
Median: 51509.0
Mode: 152
StDev: 312762.1832081966

_____
Column: Clicks
Mean: 33.39020122484689
Median: 8.0
Mode: 0
StDev: 56.89243830355751

_____
Column: Spent
Mean: 51.36065613141295
Median: 12.36999989
Mode: 0.0
StDev: 86.90841793584877

_____
Column: Total_Conversion
Mean: 2.8556430446194225
Median: 1.0
Mode: 1
StDev: 4.483593472024835

_____
Column: Approved_Conversion
Mean: 0.9440069991251093
Median: 1.0
Mode: 0
StDev: 1.737708005904991

_____
Column: interest
Mean: 32.76640419947507
Median: 25.0
Mode: 16
StDev: 26.952130977053642
```

Here is the summary statistics of the data.
The average clicks - 33
The average money spent - $53
The average impressions - 186732
The average total_conversion - 2.8 almost 3
The average interest - 32.
Here we can understand the some of the descriptive statistics from the data.
Lets the how spread out the data points.
The std of Clicks - 56
The std of money spent - 86
The std of impressions - 312762
The std of total_conversion - 4.4
The std of interest - 26
We can see how spread out the data from the mean.

By seeing this information we can say that, our data has outliers.

**correlation of the data.**

**We Have the multicoliinearity Problem.**
- Clicks are highly correlated with the impressions.
- clicks are highly correalted with the Spent.
- Impressions are Highly correalated with the Spent.
- Impressions are Highly correalated with the clicks.
- Add_id is Highly correlated with the fb_campaign_id, and xyz_campaign_id.

All the features are Having a Good correlation with the target variable expect the interest.
- Here are the some of the features that are highly correlated.
  - total_conversion
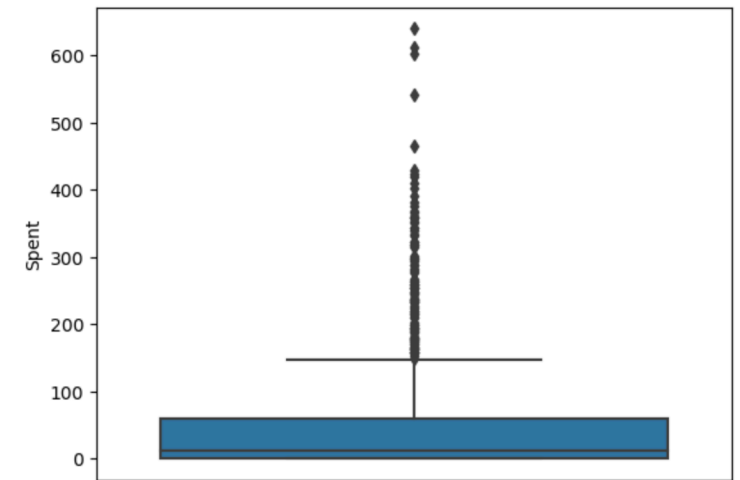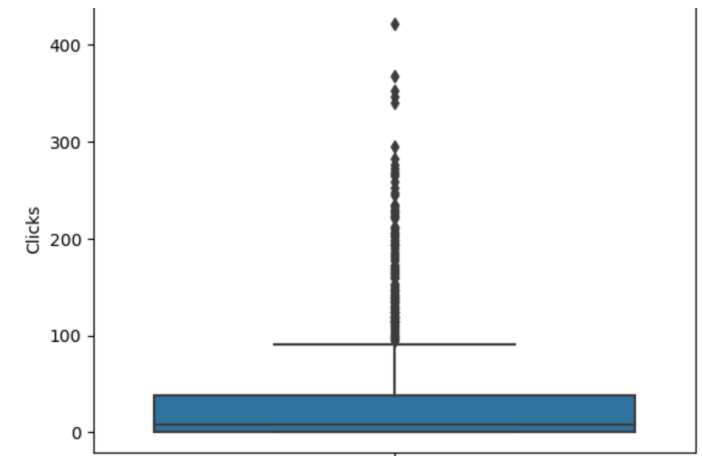 - Impressions
  - spent
  - Clicks
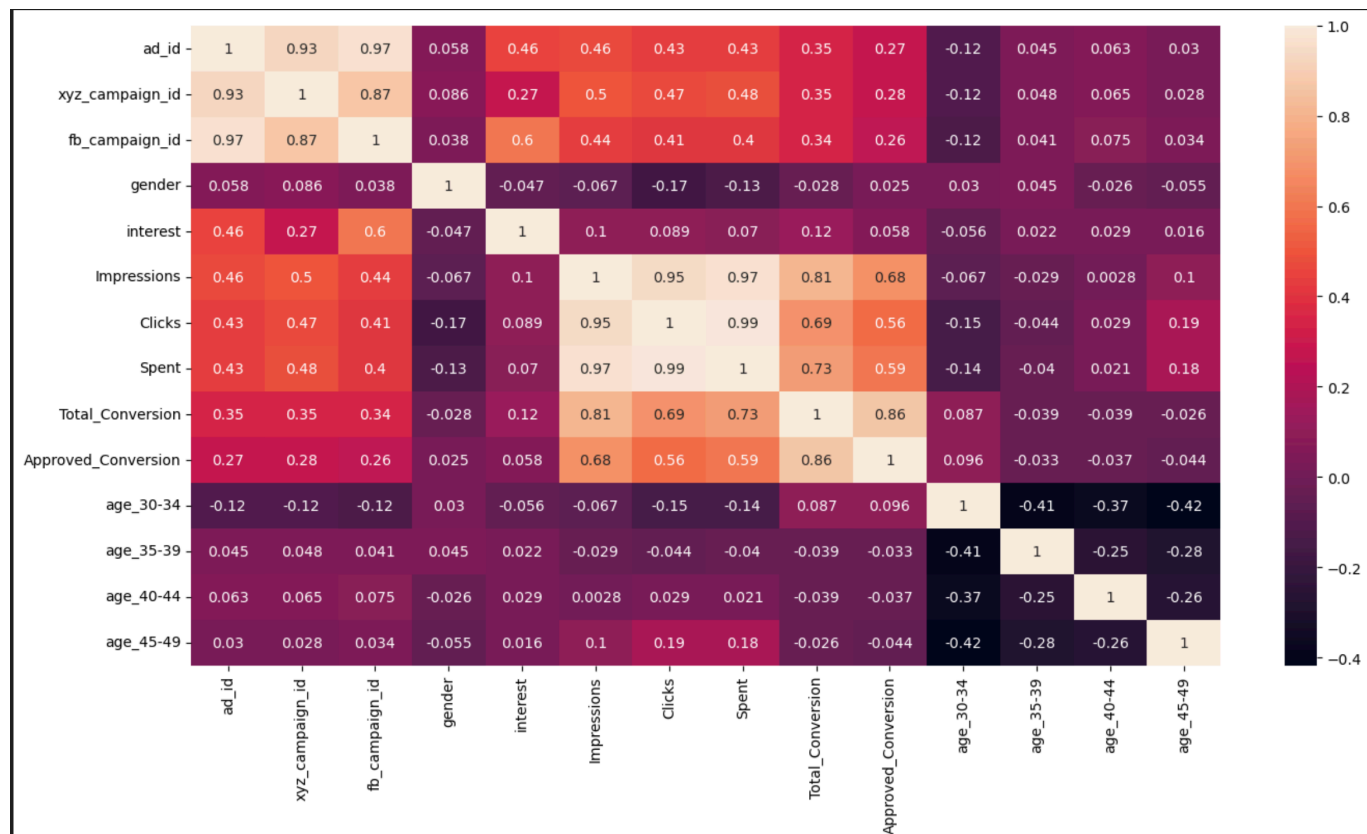
EDA, here some of the Graphs to understand the data, and to get better insights.
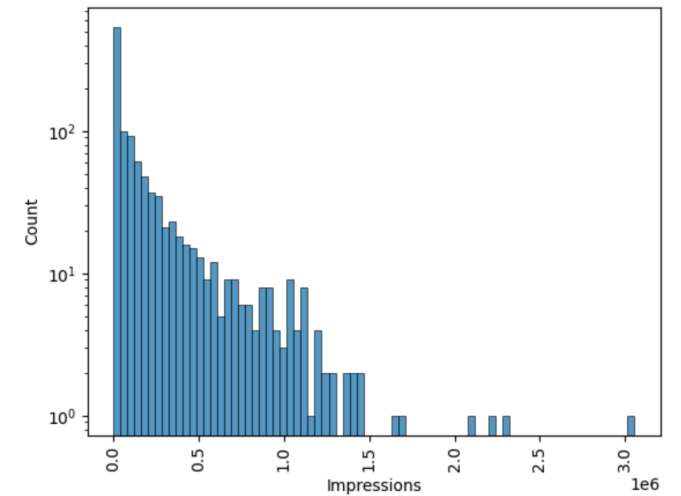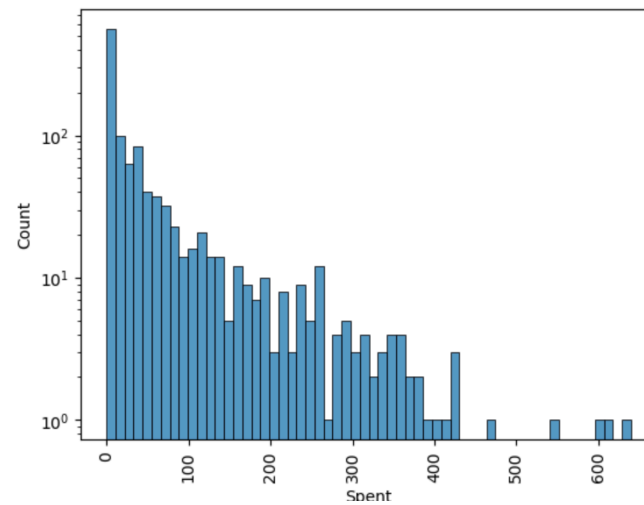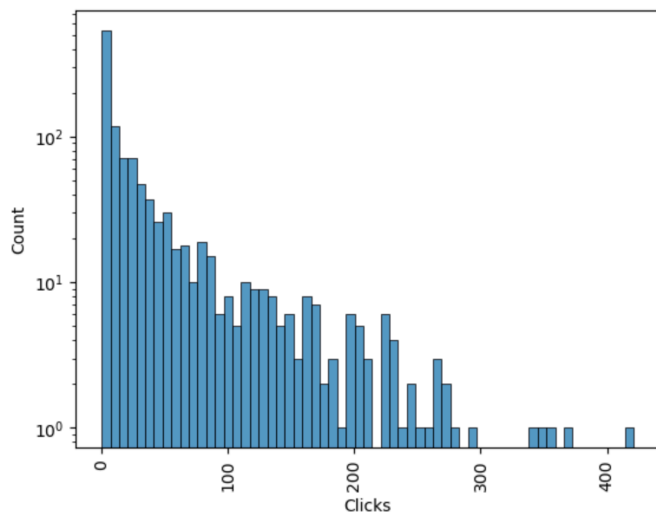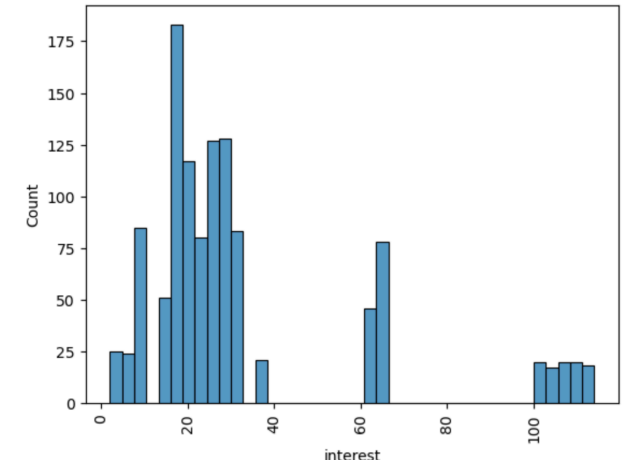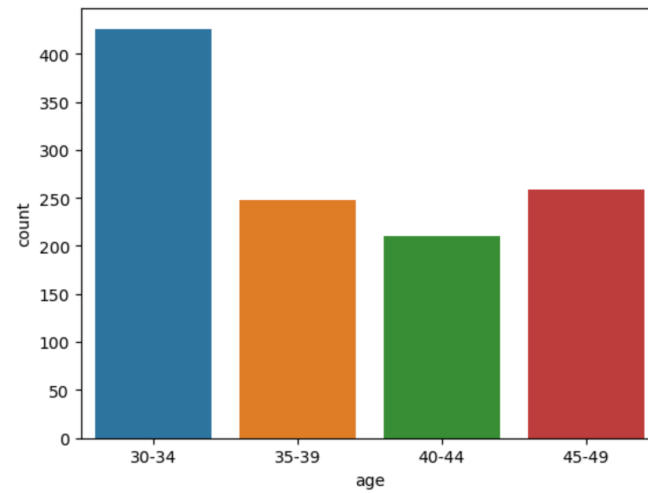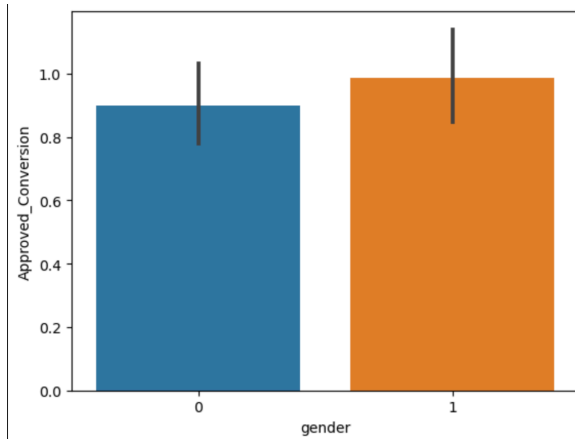
WE found some of the outliers, so Removed them by using IQR method.

```python
1  def remove_outliers_iqr(df):
2      # Calculate the first and third quartiles
3      Q1 = df.quantile(0.25)
4      Q3 = df.quantile(0.75)
5
6      # Calculate the interquartile range (IQR)
7      IQR = Q3 - Q1
8
9      # Determine the upper and lower bounds for outliers
10     lower_bound = Q1 - 1.5 * IQR
11     upper_bound = Q3 + 1.5 * IQR
12
13     # Filter out outliers from the dataframe
14     new_df = df[~((df < lower_bound) | (df > upper_bound)).any(axis=1)]
15
16     return pd.DataFrame(new_df)
```

Created 4 new features, from the existing. Data.
1. CTR - click through rate
2. CPC - cost per click
3. cost_per_conversion
4. Conversion_Rate
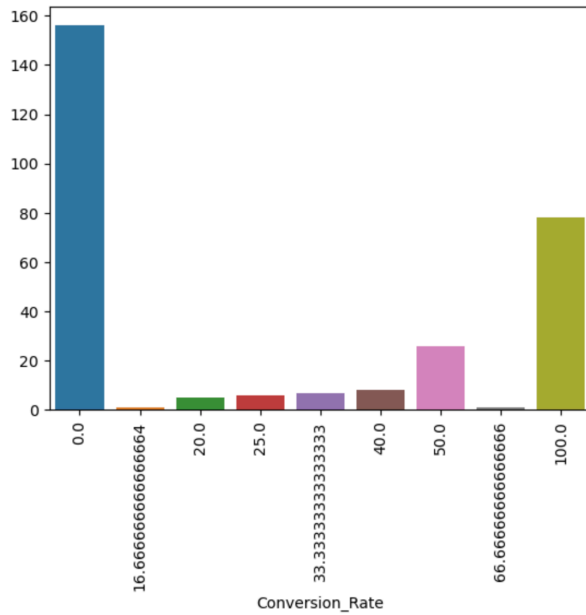
```
1  # Creating Click-Through Rate (CTR) feature
2  df['CTR'] = df['Clicks'] / df['Impressions'] * 100


1  # Creating Cost per Click (CPC) feature
2  df['CPC'] = df['Spent'] / df['Clicks']


1  # Creating Cost per Conversion feature
2  df['Cost_per_Conversion'] = df['Spent'] / df['Approved_Conversion']


1  # Creating Conversion Rate feature
2  df['Conversion_Rate'] = df['Approved_Conversion'] / df['Total_Conversion'] * 100
```
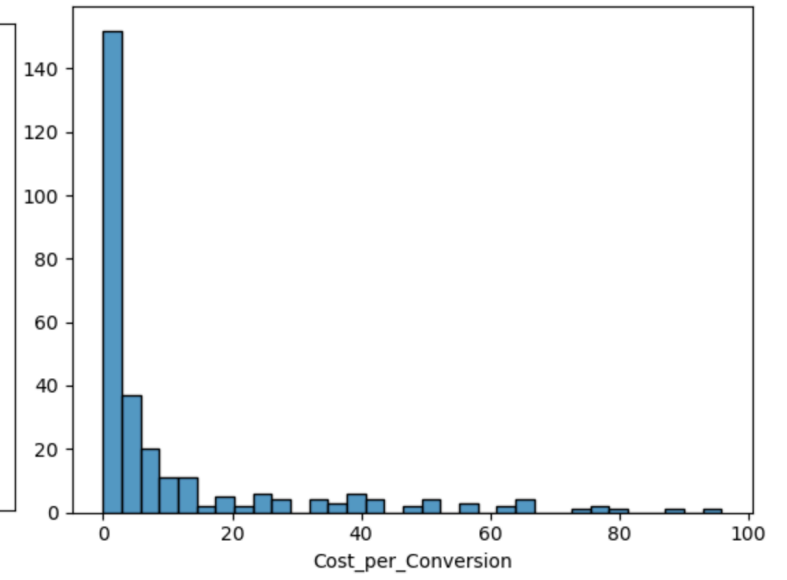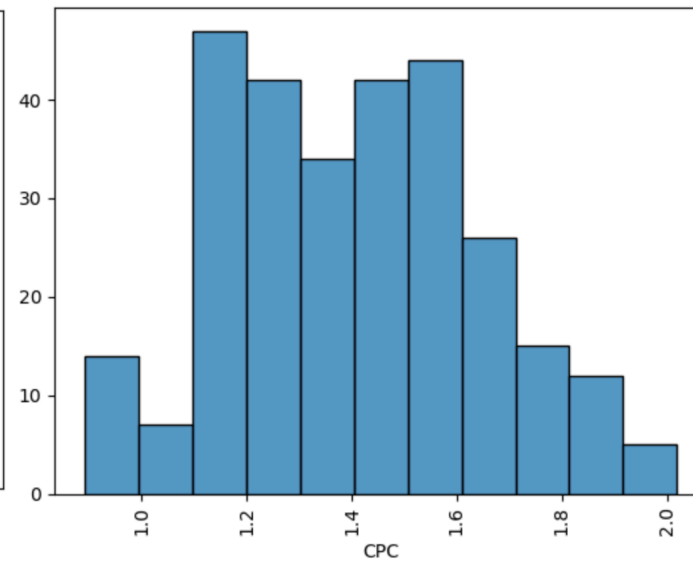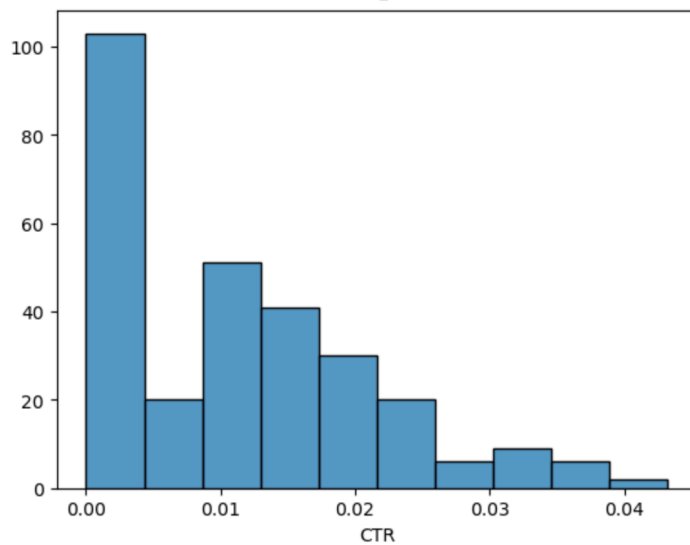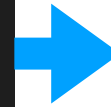
We fit four models, and their performance, are shown below.

1.linear Regression -

```
R2 score: 0.825944392537725
--------*---------
Mean Absolute Error: 0.1363531382808702
--------*---------
Root Mean Squared Error: 0.24424780512902408
```

2. Logistic Regression -

```
Accuracy: 0.9827586206896551
Precision: 0.9837164750957854
Recall: 0.9827586206896551
F1 Score: 0.9812807881773399
ROC AUC Score: 0.9725171369360753
```

```
Cassification Report:
              precision    recall  f1-score   support

         0.0       1.00      1.00      1.00        38
         1.0       0.89      1.00      0.94        17
         2.0       1.00      0.33      0.50         3

    accuracy                           0.97        58
   macro avg       0.96      0.78      0.81        58
weighted avg       0.97      0.97      0.96        58
```

3. Decision tree Regressor -

```
R2 score: 0.9279224875101245
--------*---------
Mean Absolute Error: 0.041536050156739814
--------*---------
Root Mean Squared Error: 0.15717611688306307
```

4. Decision tree classifier -

```
Accuracy: 0.9827586206896551
Precision: 0.9837164750957854
Recall: 0.9827586206896551
F1 Score: 0.9812807881773399
ROC AUC Score: 0.940379403794038
```

```
Classification Report:
              precision    recall  f1-score   support

         0.0       1.00      1.00      1.00        38
         1.0       0.94      1.00      0.97        17
         2.0       1.00      0.67      0.80         3

    accuracy                           0.98        58
   macro avg       0.98      0.89      0.92        58
weighted avg       0.98      0.98      0.98        58
```

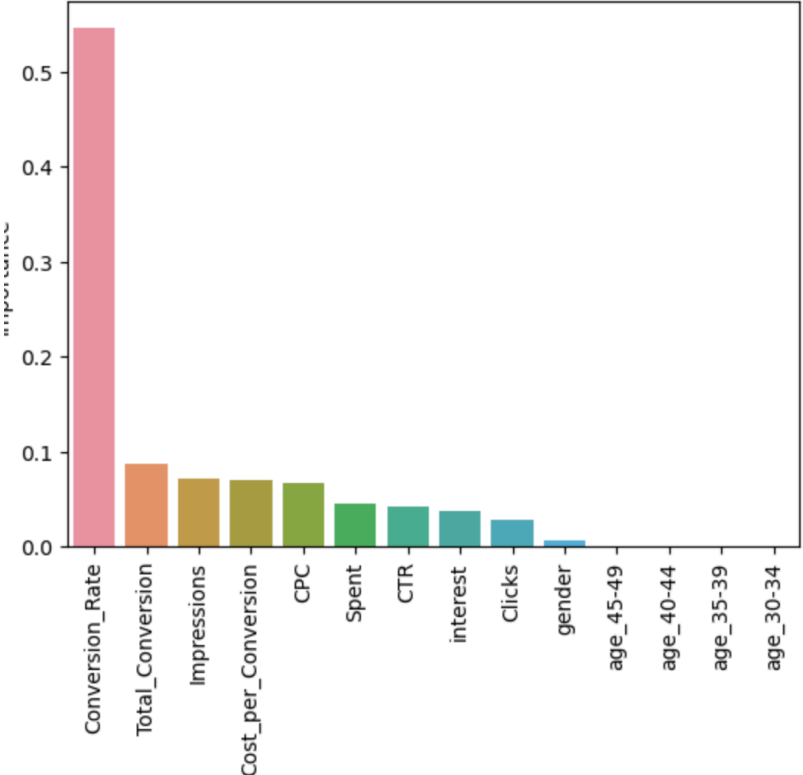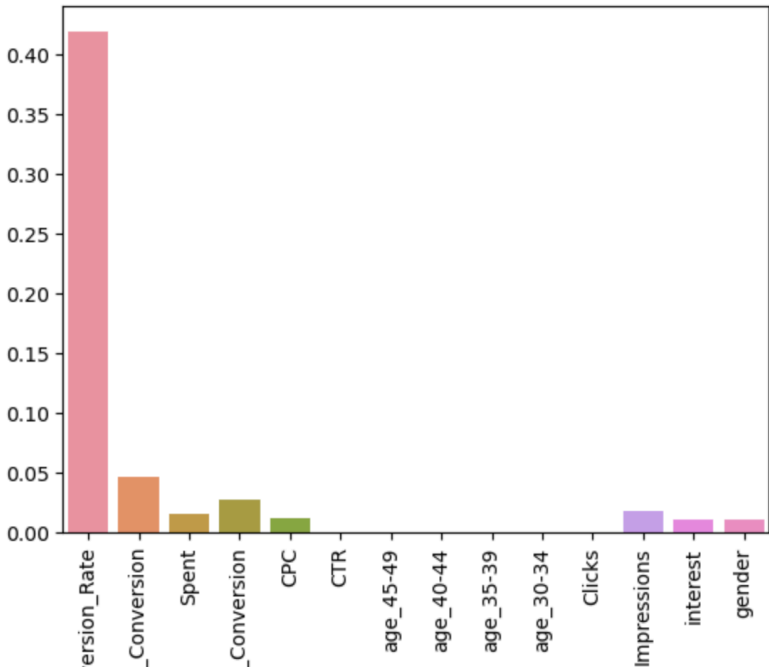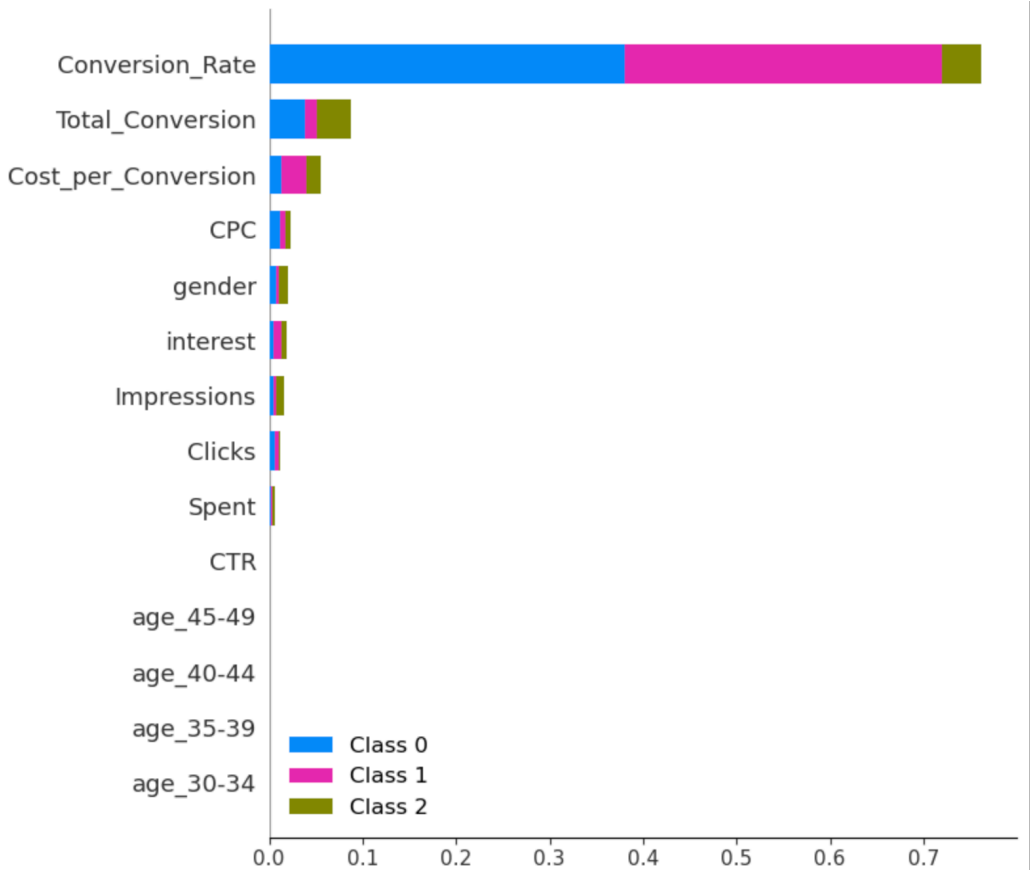Overall, Logistic Regression performs well than other models.

Feature importance analysis

Important Features
- Conversion_Rate
- total_conversion
- Cost_per_conversion
- impressions

Least important Features
- Age
- Id (removed)

Recommendations

Based on my Analysis, Here are some of the recommendations that I can give.

Focus on the most important features, and find the ways the to modify and increase the rate, as well as we should kept focus on the non-important features also, and why they are not performing well and come up with the solutions to solve them.

```
Our End Goal is to Increase the CONVERSION RATE.
```

1. **Targeting Aspects**
   - when I fit the models(lg,lr,dt), I found that, the age (groups), gender, and clicks have the significance affect on the conversion rate, so It would be better to focus on the different age groups, and come up with the good stratagey to boost the conversion rate.

2. **Ad Design Aspects.**
   - when we fit tree model for the feature importance, we can see that CPC and CTR having some importance in the conversion rate. so we should comeup with new technologies to attract and may be coming up with new designs, etc may works well.

3. **Budget allocation Aspects.**
   - From the Lr and Lg models, the amount Spent is the important features for the conversion rate, so we have to be encourage the amount spending inorder to increase the conversion rate, and also the budget should be reasonble so that we didn't end up in looses.

4. **Ad placement strategies Aspects.**
   - We have seen in the feature importance, the impressions are also one of the important feature for the conversion rate, so it always recommended to have a more attracting ads so that it wil increase the impressions, and then CTR and then conversion rate.and the ad should be placed in the more popular sites, so that It can attract auidence, which indirectly increase the conversion rate.

These are the some of the recommendations, I can give from my analysis, it doesn't mean that there are no pitfalls in the company, it always better to cross check everytime, and make a proper analysis for the increase and the decrease in the conversion rate.