Advertisement Targeting
Business Report

Problem Statement

Context

About Datasets

Social Media Advertising Dataset:

Description:

The Social Media Advertising dataset is a comprehensive collection of data related to various social media advertising campaigns. It includes information such as ad impressions, clicks, spend, demographic targeting, and conversion rates. The dataset encompasses multiple social media platforms such as Facebook, Instagram, Pinterest, and Twitter, providing a diverse range of advertising campaign data.

Potential Uses for Data Analysis:

Campaign Performance Analysis: Analyze the performance of advertising campaigns across different social media platforms to identify the most effective channels and strategies.
Audience Segmentation: Utilize demographic targeting data to segment the audience and tailor advertising campaigns to specific demographic groups.
ROI Calculation: Calculate the return on investment (ROI) for advertising campaigns by comparing ad spend to conversion rates and revenue generated.
Optimization Strategies: Identify optimization opportunities by analyzing click-through rates, engagement metrics, and conversion funnels to improve campaign effectiveness.
Predictive Modeling: Build predictive models to forecast future campaign performance and optimize advertising strategies for maximum impact.

Marketing Campaign Performance Dataset

Description:
The Marketing Campaign Performance Dataset provides valuable insights into the effectiveness of various marketing campaigns. This dataset captures the performance metrics, target audience, duration, channels used, and other essential factors that contribute to the success of marketing initiatives. With 200000 unique rows of data spanning two years, this dataset offers a comprehensive view of campaign performance across diverse companies and customer segments.

Columns:
Company: The company responsible for the campaign, representing a mix of fictional brands.
Campaign_Type: The type of campaign employed, including email, social media, influencer, display, or search.
Target_Audience: The specific audience segment targeted by the campaign, such as women aged 25-34, men aged 18-24, or all age groups.
Duration: The duration of the campaign, expressed in days.

Channels_Used: The channels utilized to promote the campaign, which may include email, social media platforms, YouTube, websites, or Google Ads.
Conversion_Rate: The percentage of leads or impressions that converted into desired actions, indicating campaign effectiveness.
Acquisition_Cost: The cost incurred by the company to acquire customers, presented in monetary format.
ROI: Return on Investment, representing the profitability and success of the campaign.
Location: The geographical location where the campaign was conducted, encompassing major cities like New York, Los Angeles, Chicago, Houston, or Miami.
Language: The language used in the campaign communication, including English, Spanish, French, German, or Mandarin.
Clicks: The number of clicks generated by the campaign, indicating user engagement.
Impressions: The total number of times the campaign was displayed or viewed by the target audience.
Engagement_Score: A score ranging from 1 to 10 that measures the level of engagement generated by the campaign.
Customer_Segment: The specific customer segment or audience category that the campaign was tailored for, such as tech enthusiasts, fashionistas, health and wellness enthusiasts, foodies, or outdoor adventurers.
Date: The date on which the campaign occurred, providing a chronological perspective to analyze trends and patterns.

Scope:
By leveraging this dataset, marketers and data analysts can uncover valuable insights regarding campaign performance, audience preferences, channel effectiveness, and ROI. This dataset serves as a valuable resource for market research, campaign optimization, and data-driven decision-making, enabling businesses to refine their marketing strategies and drive targeted growth.

Data Overview:

- dataset has 200000 rows and 16 columns

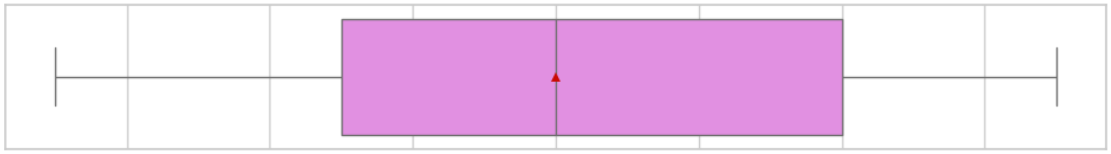
The statistical summary :

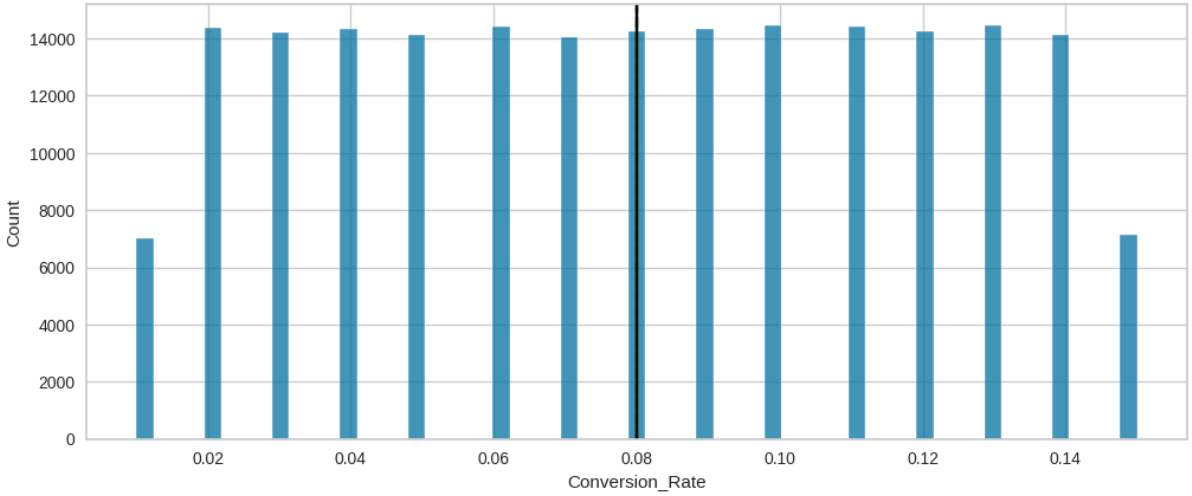
	Conversion_Rate	ROI	Clicks	Impressions	Engagement_Score
count	200000.000000	200000.000000	200000.000000	200000.000000	200000.000000
mean	0.080070	5.002438	549.772030	5507.301520	5.494710
std	0.040602	1.734488	260.019056	2596.864286	2.872581
min	0.010000	2.000000	100.000000	1000.000000	1.000000
25%	0.050000	3.500000	325.000000	3266.000000	3.000000
50%	0.080000	5.010000	550.000000	5517.500000	5.000000
75%	0.120000	6.510000	775.000000	7753.000000	8.000000
max	0.150000	8.000000	1000.000000	10000.000000	10.000000

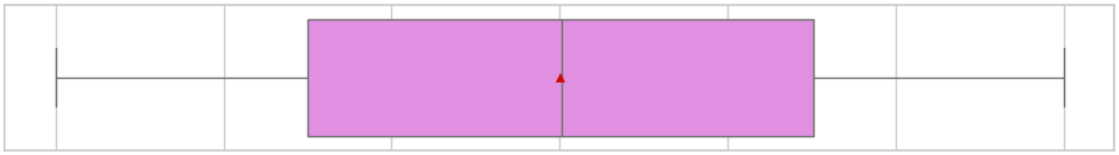
Exploratory Data Analysis (EDA):

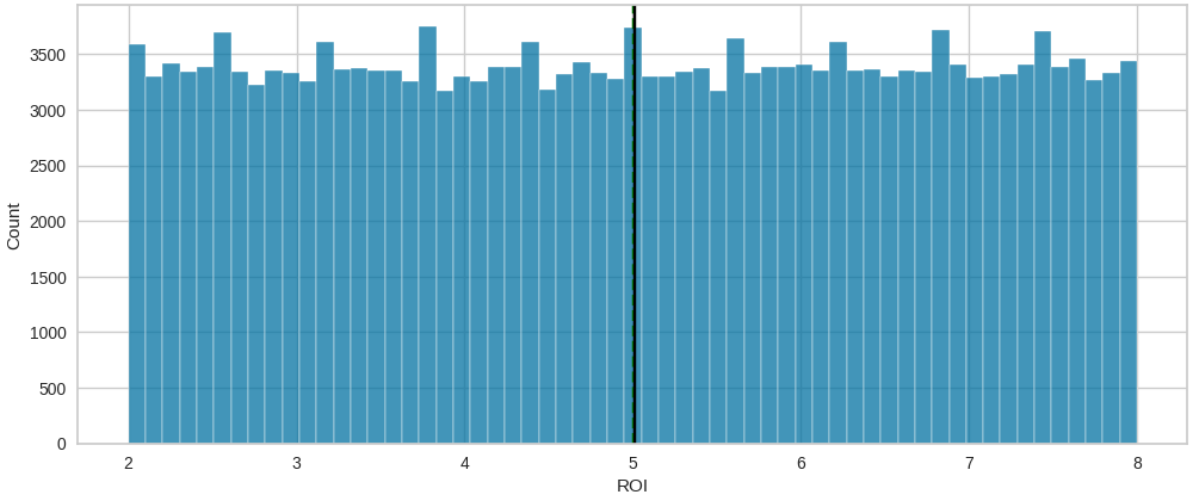
Let's perform univariate analysis for the numerical columns. Since SL_key and customer key don't add anything to the insights, lets remove these two columns from out data set

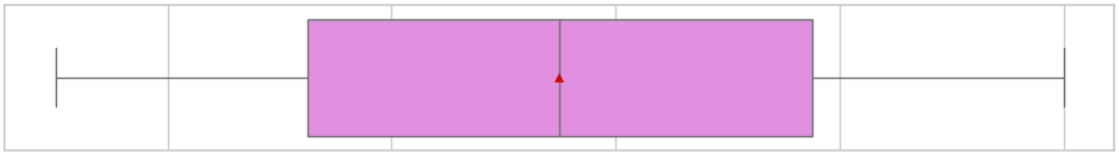
Univariate Analysis:

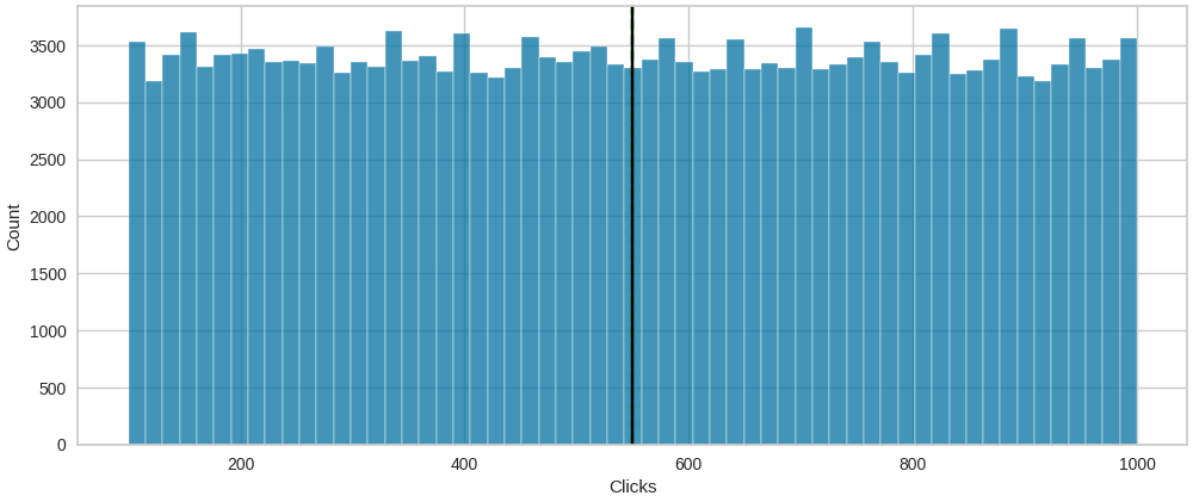


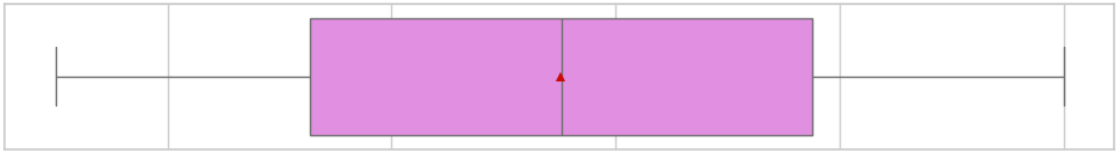

Count
14000
12000
10000
8000
6000
4000
2000
0
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Conversion_Rate

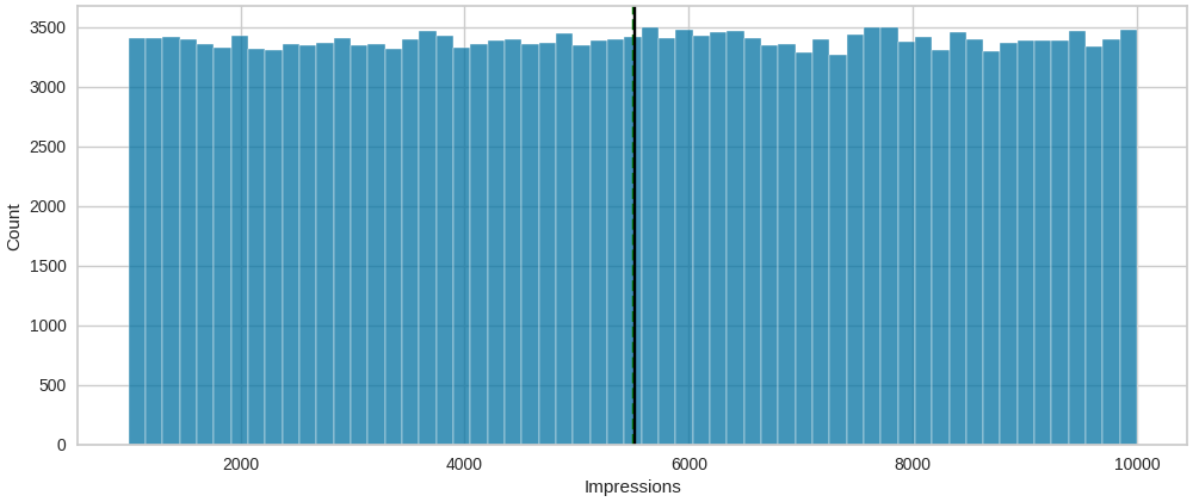


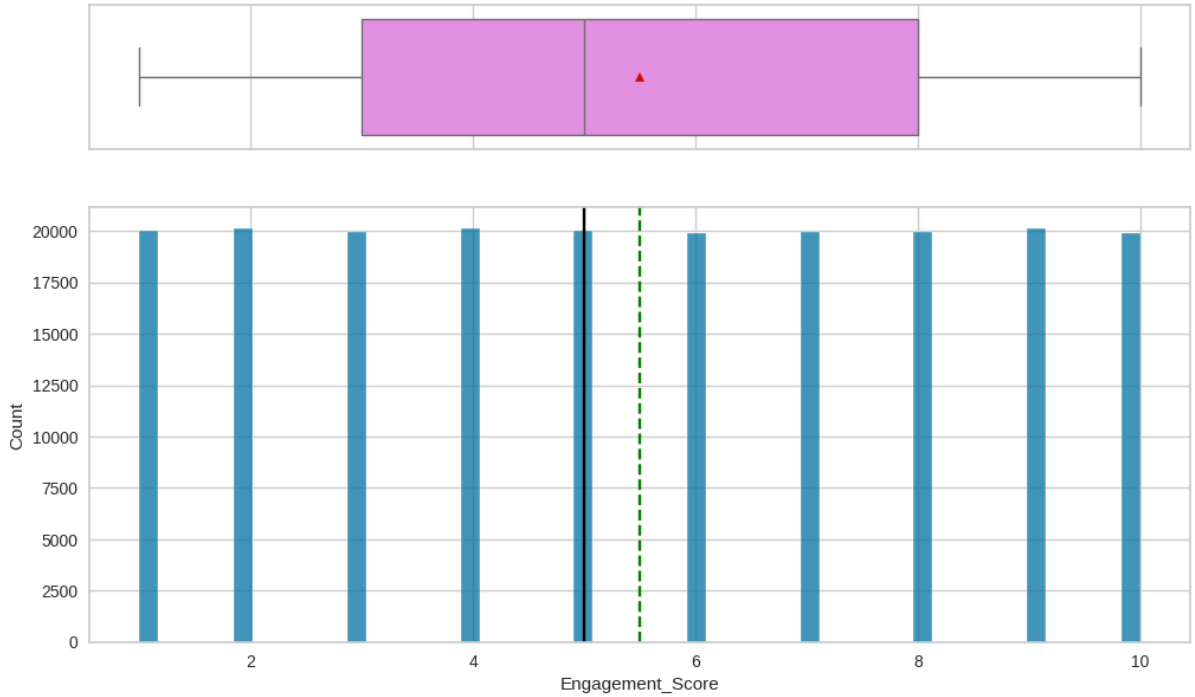

Count
3500
3000
2500
2000
1500
1000
500
0
2
3
4
5
6
7
8
ROI



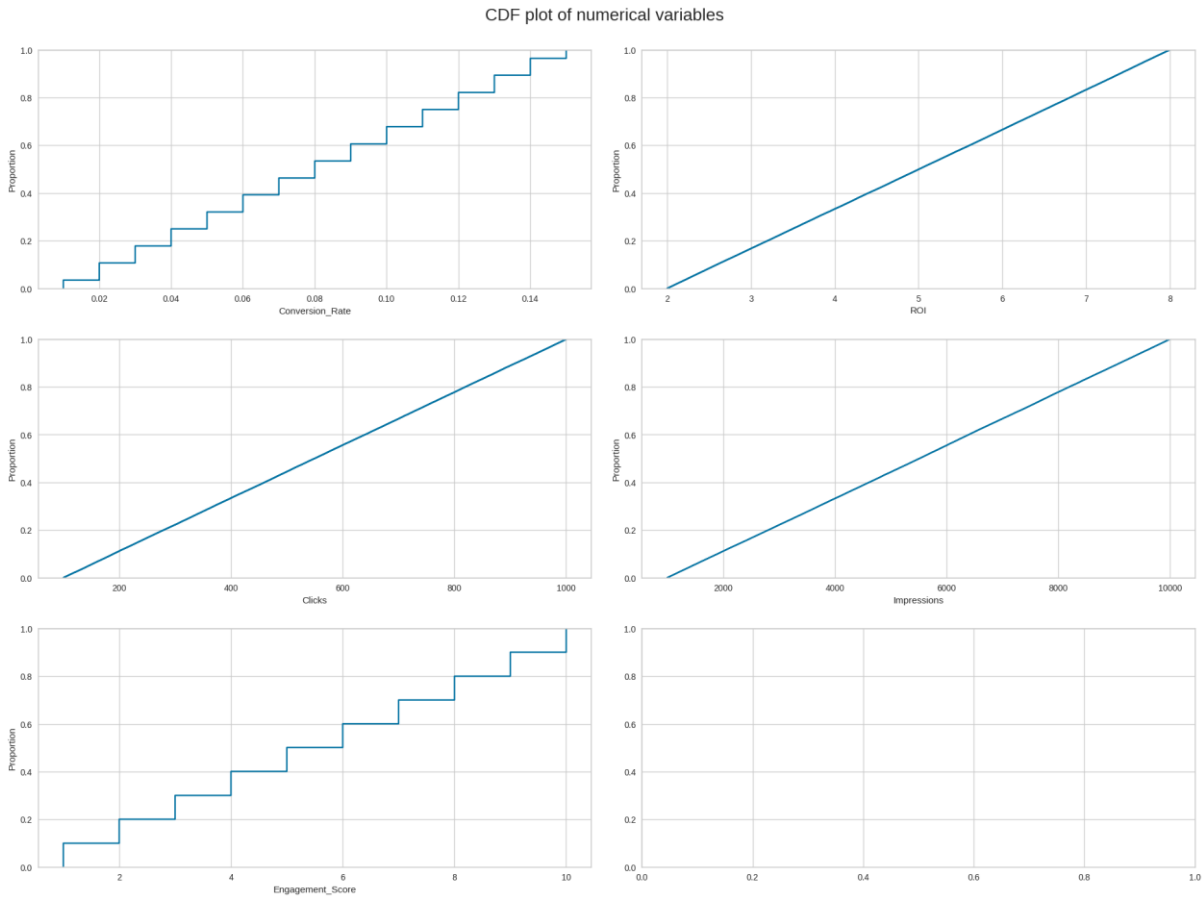

Count
3500
3000
2500
2000
1500
1000
500
0
200
400
600
800
1000
Clicks



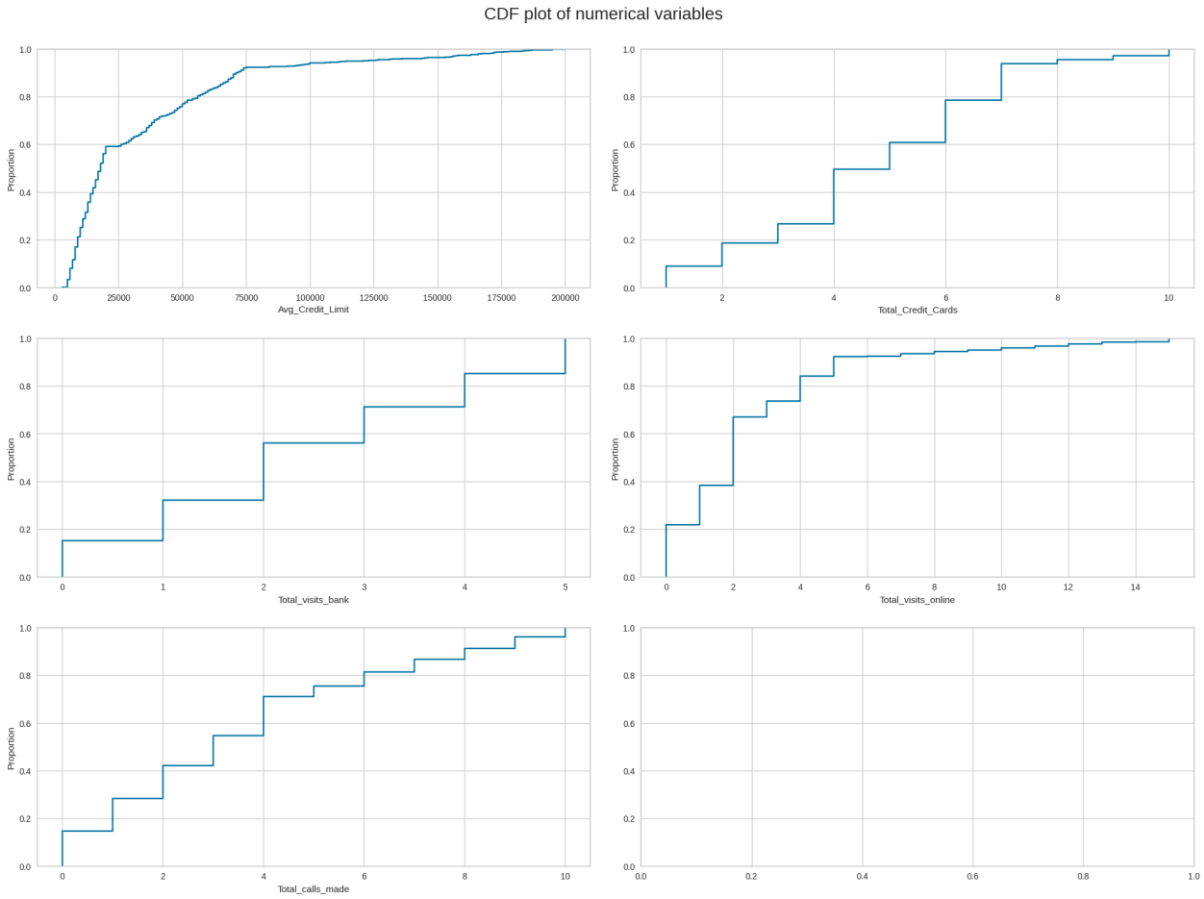

Count
3500
3000
2500
2000
1500
1000
500
0
2000
4000
6000
8000
10000
Impressions


Count
Engagement_Score

ECDF Plots:

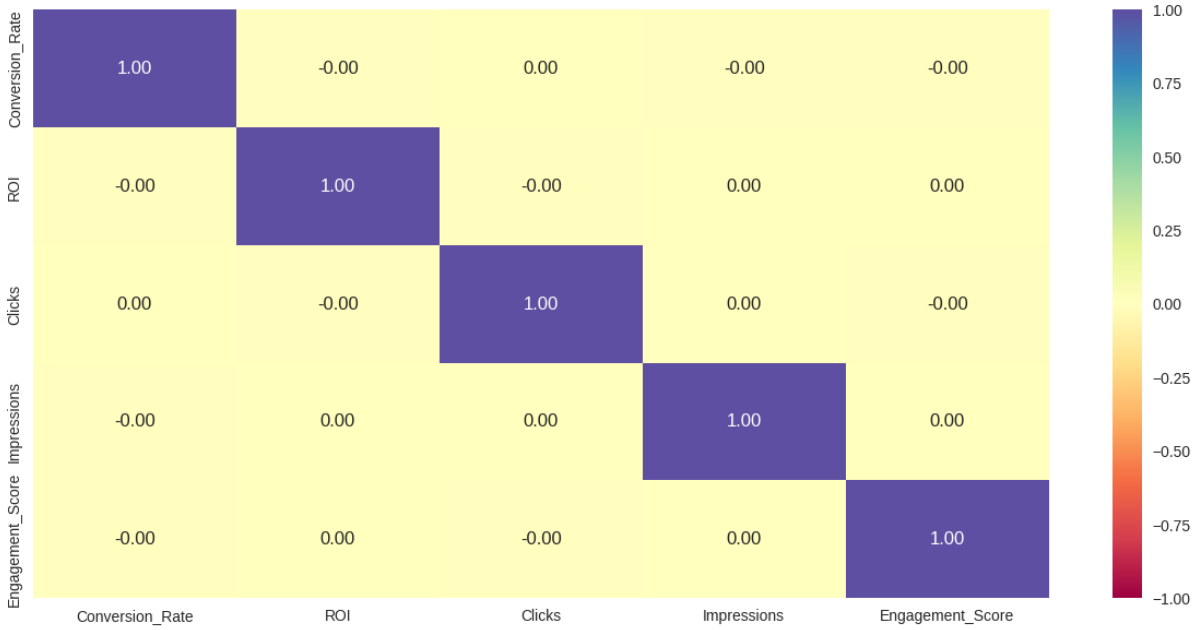

CDF plot of numerical variables
Proportion
Conversion_Rate
ROI
Clicks
Impressions
Engagement_Score

The ecdf plots:

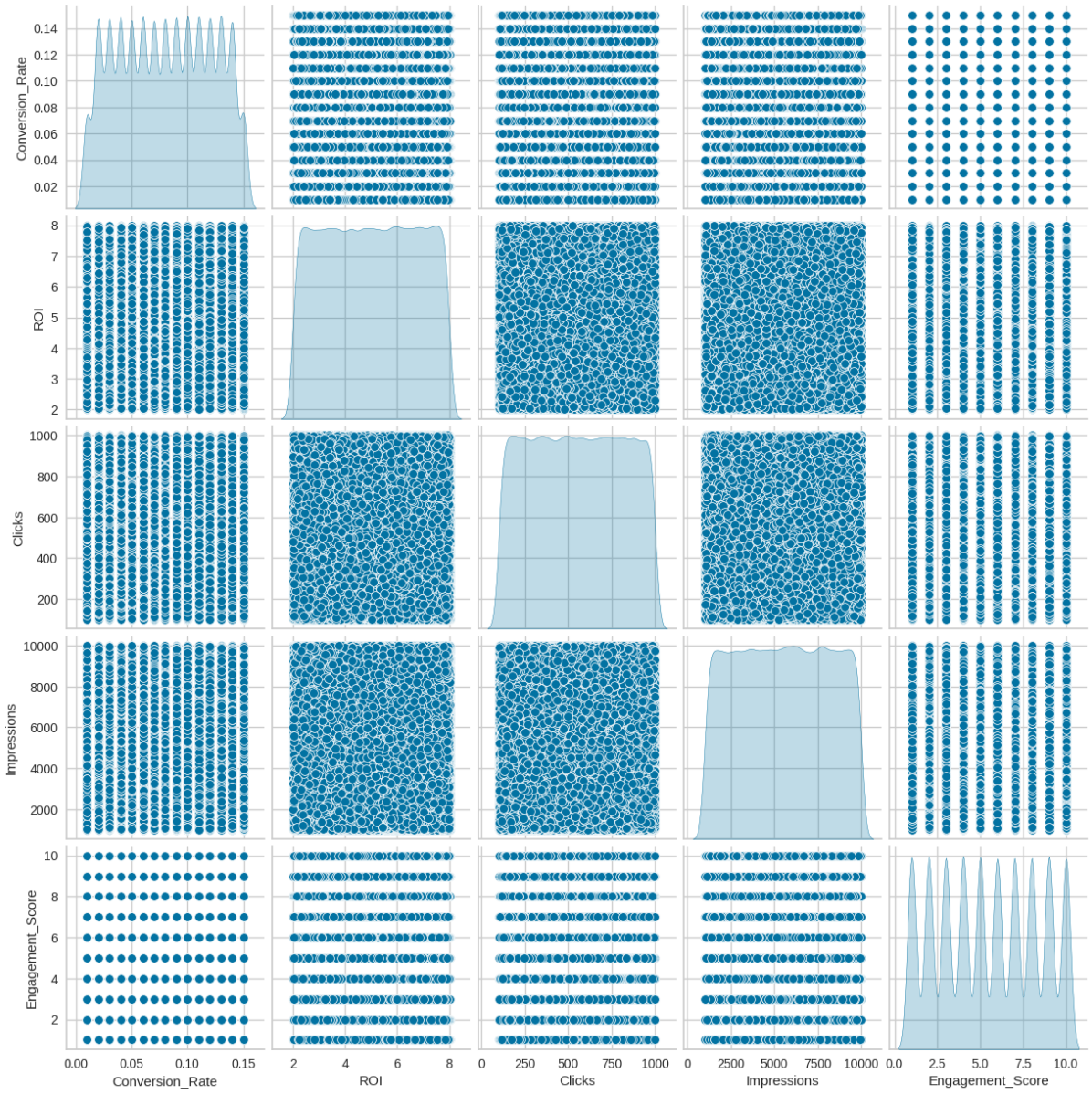


- The cummulative density function plots above also shows the proportion (or percentage) of data .

Correlation check:


Conversion_Rate
ROI
Clicks
Impressions
Engagement_Score
1.00
-0.00
0.00
-0.00
-0.00
-0.00
1.00
-0.00
0.00
0.00
0.00
-0.00
1.00
0.00
-0.00
-0.00
0.00
0.00
1.00
0.00
-0.00
0.00
-0.00
0.00
1.00
Conversion_Rate
ROI
Clicks
Impressions
Engagement_Score
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

Pairplot:



From the above plots, we can find that there is no correlation between the numeric fields.

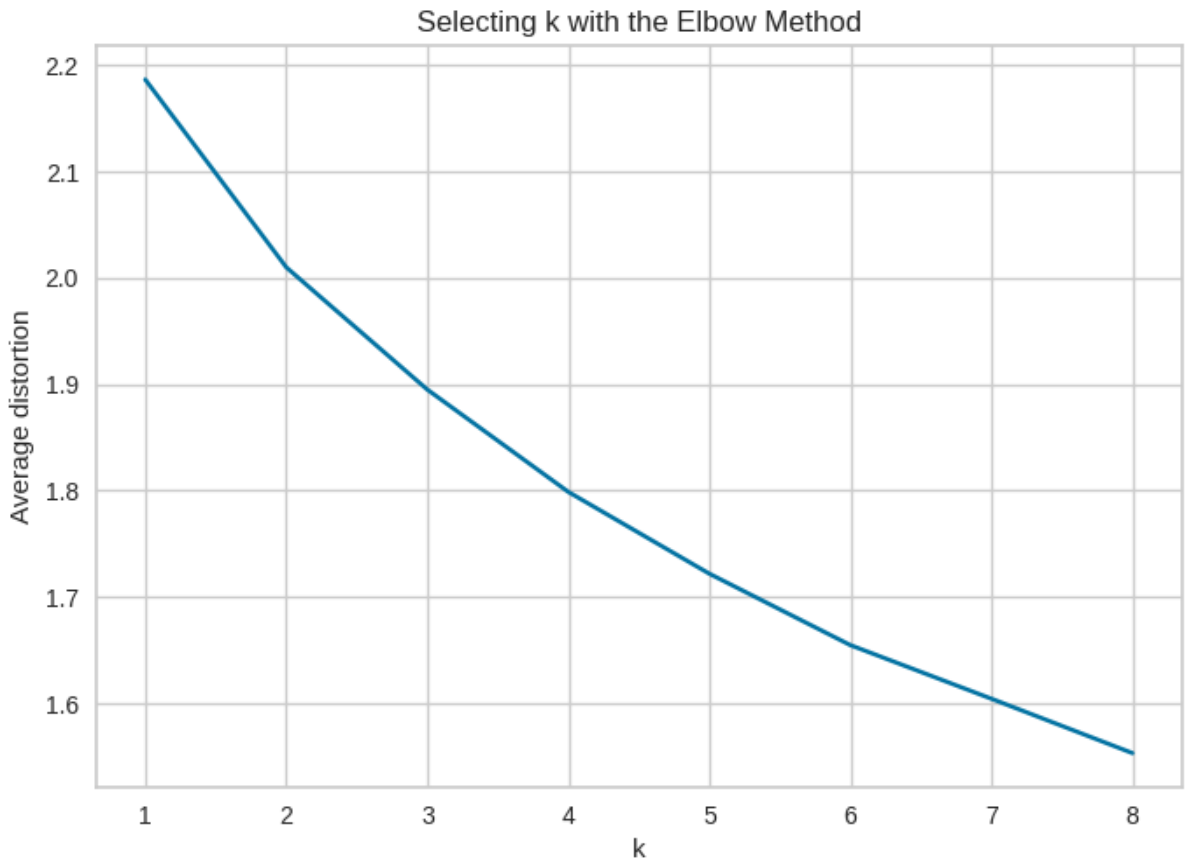
Data Preprocessing:

In clustering, the first step is to ensure that the data is scaled.

K – Means clustering:

We use Euclidean distance measure.

Elbow scores to calculate k value:



The average distortion values of each k value are:

Number of Clusters: 1 Average Distortion: 2.1864953935536944

Number of Clusters: 2 Average Distortion: 2.0098709015585894

Number of Clusters: 3 Average Distortion: 1.8948849588693109

Number of Clusters: 4 Average Distortion: 1.7986483914692637

Number of Clusters: 5 Average Distortion: 1.7217135366635163

Number of Clusters: 6 Average Distortion: 1.6545345204408501

Number of Clusters: 7 Average Distortion: 1.6041294777215505

Number of Clusters: 8 Average Distortion: 1.5530029872865165

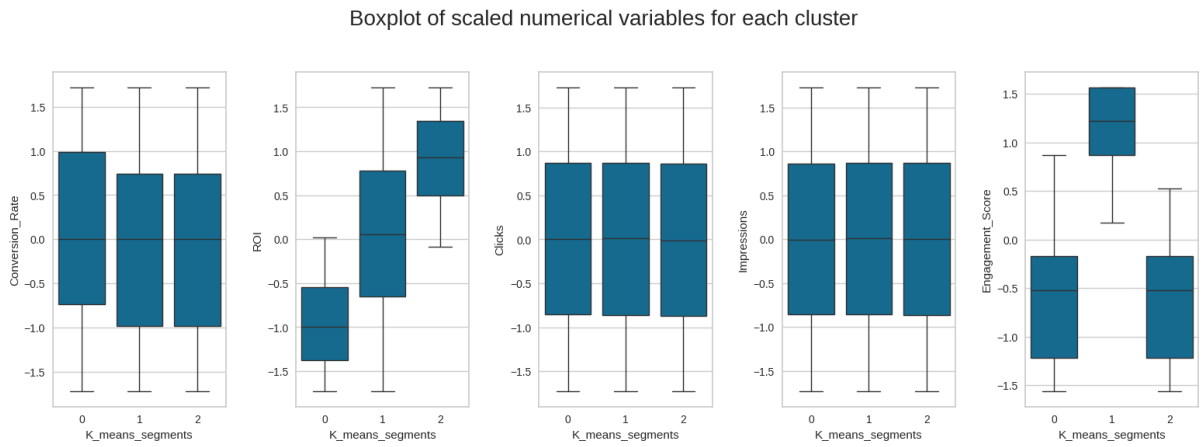
Based on the Elbow method plot generated by the code, the best k value to choose is 3. The plot shows a significant decrease in distortion between k=1 and k=3, then the rate of decrease slows considerably, creating an "elbow" shape at k=3. While there might be a slight further decrease in distortion with larger k values, the improvement is marginal, and selecting a larger k risks overfitting. Therefore, the point of diminishing returns (the "elbow") suggests that k=3 is the optimal number of clusters.

Cluster Profiles:

K_means_segments	Conversion_Rate	ROI	Clicks	Impressions	Engagement_Score
0	0.080682	3.344508	550.369697	5498.332479	3.789756
1	0.079906	5.089670	550.216813	5519.884615	8.512838
2	0.079648	6.577539	548.626750	5501.136575	3.563512

The above data shows different segments and the profile.

The boxplots for each segment:



From the above boxplots, we can come up with the following cluster profiles:

Cluster 0:

It has the highest conversation rate, but the least ROI

The clicks and impressions are high, but the engagement score is low.

Recommendations:

- Focus on converting this engaged audience with stronger call-to-actions (CTAs) and targeted offers.
- Consider A/B testing landing pages to increase conversion.
- Possibly retarget users from this cluster with conversion-focused messaging.

Cluster 1:

It has slightly higher conversation rate, and a high ROI

The clicks and impressions are high, and the engagement score is the highest.

Recommendations:

- Optimize campaigns to boost performance by:
- Enhancing personalization
- Increasing ad frequency
- Improving email or content targeting
- Monitor this group as they may show potential with slight tuning.

Cluster 2:

It has good conversation rate, with the highest ROI.

The clicks and impressions are high, but the engagement score is low.

Recommendations:

- These users convert well despite low engagement — focus on scaling acquisition of similar users.
- Explore which marketing channels or campaigns are driving this behaviour.
