Project Report: Marketing Spend Effectiveness Analysis (Statistical & Regression Analysis)

Date: 19th July, 2025
To: Stakeholders
From: Shijin Ramesh, Data Analyst
Subject: Comprehensive Analysis of Marketing Channel Performance and Cost Efficiency
---

1. Executive Summary

This report summarizes a comprehensive statistical analysis undertaken to evaluate the effectiveness of various marketing channels and campaign performance using a dataset of approximately 200,000 records. The primary objective was to uncover statistically significant differences in key metrics such as conversion rates, click-through rates (CTR), and cost-efficiency across channels including Social, Email, and Search.

Key Findings Highlighted:

• **Consistent Conversion Rates:** Average conversion rates were found to be remarkably similar across all marketing channels, hovering around 16.6% to 16.7%.

• **ROI Not Driven by Spend:** Analysis showed a **weak correlation between acquisition spending and Return on Investment (ROI)**, suggesting that merely increasing spend does not guarantee better ROI. Campaign success appears to depend more on targeting strategy or quality.

• **Equal ROI Across Demographics:** ROI was found to be consistent across different **gender groups** (Men and Women) and **age groups**, even though the 25-34 age group had significantly higher engagement or targeting.

• **Clicks Don't Guarantee Conversions:** No strong correlation was observed between the number of clicks and conversion rates, with conversion rates remaining uniform around 8% regardless of click volume. This implies that **higher clicks alone do not necessarily lead to higher conversions**.

• **Cost Per Acquisition (CPA) Variances:** CPA demonstrated significant variation and a considerable number of **outliers (810 records)**, indicating campaigns with unusually high or low acquisition costs. This suggests potential inefficiencies or highly profitable campaigns that warrant deeper investigation.

• **Stable Click-Through Rate (CTR):** In contrast to CPA, CTR showed **no significant outliers** and remarkable stability across campaigns after necessary data transformations, suggesting consistent ad engagement.

• **Hypothesis Testing Insights:**
  ◦ The **average Cost Per Acquisition (CPA) was found to be significantly different from ₹300**.

This indicates a need for optimizing cost-related factors in marketing strategies.

◦ The **average Click-Through Rate (CTR) was found to be statistically aligned with an expected 10% benchmark**, suggesting campaigns are performing as anticipated regarding engagement.

In essence, while ad engagement (CTR) is stable and conversions are consistent across channels, **optimizing Cost Per Acquisition (CPA) should be a primary focus** for future marketing efforts, as current CPA values significantly deviate from expected benchmarks and show considerable variability.

2. Project Overview & Methodology

This project utilized Python with libraries such as Pandas, Numpy, Matplotlib, Seaborn, Jupyter Notebook, and Scipy for data manipulation, analysis and visualization. The analysis involved several stages:

• **Data Cleaning & Exploratory Data Analysis (EDA):** Initial processing and visualization to understand data characteristics.

• **Univariate & Bivariate Analysis:** Examining individual variable distributions and relationships between pairs of variables.

• **Statistical Analysis:** Applying advanced statistical techniques including distribution analysis, outlier detection, confidence interval estimation, and hypothesis testing.

3. Detailed Findings & Insights

3.1. Initial Exploratory Data Analysis (EDA) Insights

• **Channel Effectiveness:** The average conversion rate was consistent across all channels (Social, Email, Search), at approximately 16.6% to 16.7%.

• **Acquisition Cost vs. ROI:** A scatter plot with a regression line showed a nearly flat line and weak correlation between acquisition spending and ROI. This indicates that **increased spending does not strongly predict better ROI**, implying that campaign success is more dependent on factors like targeting strategy or quality rather than just budget.

• **Demographic ROI:** ROI appeared to be equal for both male and female groups, despite a slightly higher count of men in the dataset. Similarly, ROI was consistent across all age groups. While the 25-34 age group had almost double the number of people compared to other groups, their mean, median, and other ROI parameters were the same, suggesting either major targeting towards this group or their higher responsiveness.

• **Clicks vs. Conversions:** A scatter plot revealed no strong or visible correlation between the number of clicks and the conversion rate. The regression line indicated that most conversion

rates centered around 8% irrespective of click volume (e.g., whether 200 or 1000 clicks). This signifies that **simply increasing clicks does not necessarily translate to higher conversion rates**.

3.2. Statistical Analysis & Data Transformation

• **Data Distributions:** While Acquisition Cost, Engagement Score and Conversion Rate showed uniform distributions, Click-Through Rate (CTR) and Cost Per Acquisition (CPA) were **highly right-skewed**. This skewness meant that a few campaigns had unusually high values, pulling the distribution to the right.

• **Log Transformation:** To address the skewness and enable the use of parametric statistical methods (like z-scores and hypothesis testing), a **log transformation was applied to both CTR and CPA**. This compressed high values and spread-out low values, making the distributions more symmetrical and approximately normal.

• **Normality Confirmation:** After transformation, both Log CTR and Log CPA visually resembled a bell curve. Although Shapiro-Wilk tests on random samples indicated they weren't strictly normal (p-values close to zero), the distributions were **"approximately normal" or "normal enough"** for practical purposes. This was further confirmed by the Empirical Rule, where observed percentages (68.02%, 95.93%, 99.59% for CPA and 68.08%, 94.78%, 100% for CTR) closely matched the theoretical 68%, 95%, and 99.7% for normal distributions. This successful transformation allowed for confident progression with deeper statistical analysis.

3.3. Outlier Detection (Z-Score Method)

• **CPA Outliers:** Using the Z-score method (where Z-scores > +3 or < -3 indicate outliers), **Cost Per Acquisition (CPA) had 810 outliers**. This means a significant number of campaigns had unusually high or low acquisition costs compared to the average. These outliers might represent campaigns that are either not cost-effective or are highly profitable, warranting a deeper look.

• **CTR Stability:** In contrast, **Click-Through Rate (CTR) had 0 outliers**. This suggests that engagement rates (clicks on ads) are consistent and stable across campaigns after transformation, with no extreme behavior observed.

3.4. Confidence Level Estimation

Confidence intervals were calculated for the mean values of the log-transformed CPA and CTR metrics at both 95% and 99% confidence levels.

- **Log CPA:** The mean was 5.91.
  - **95% Confidence Interval: 5.906 to 5.915**
  - **99% Confidence Interval: 5.905 to 5.916**
- **Log CTR:** The mean was -2.30.

◦ **95% Confidence Interval: -2.308 to -2.301**
◦ **99% Confidence Interval: -2.310 to -2.300**

These very narrow intervals suggest a large, consistent dataset and indicate that the average values are stable and reliable. This allows for greater confidence in setting targets, spotting unusual patterns, or optimizing campaign strategies.

3.5. Hypothesis Testing

Two key business questions were addressed through one-sample hypothesis tests, using the log-transformed values due to the initial data skewness.

• **Hypothesis Test on Log CPA:**

◦ **Business Question:** Is the average Cost Per Acquisition (CPA) significantly different from ₹300?
◦ **Null Hypothesis (H_0):** Average Log CPA = log(300)
◦ **Alternative Hypothesis (H_1):** Average Log CPA ≠ log(300)
◦ **Result:** The p-value was extremely small (0.0), which is much less than 0.05. Therefore, the null hypothesis was **rejected**.
◦ **Insight:** This means the **average Cost Per Acquisition is significantly different from ₹300**. Campaigns are not performing around this benchmark; they are either much higher or lower, necessitating further business evaluation and optimization of cost-related factors.

• **Hypothesis Test on Log CTR:**

◦ **Business Question:** Is the average Click-Through Rate (CTR) significantly different from 10% (0.10)?
◦ **Null Hypothesis (H_0):** Average Log CTR = log(0.10)
◦ **Alternative Hypothesis (H_1):** Average Log CTR ≠ log(0.10)
◦ **Result:** The p-value was greater than 0.05. Therefore, the null hypothesis was **not rejected**.
◦ **Insight:** This means the **average CTR is not significantly different from 10%**. Campaigns are performing close to the expected CTR, and there is no strong evidence to suggest otherwise.

3.6 Regression Analysis

This project focused on understanding what drives the Cost Per Acquisition (CPA) in marketing campaigns. By analyzing past campaign data, we built a statistical model to predict CPA using three key factors:

- Cost Per Click (how much we pay for each ad click)
- Conversion Rate (how many clicks turn into customers)
- Number of Clicks

The model is able to explain almost 87% of the changes in CPA, which means it gives us a very reliable understanding of what affects our acquisition cost.

Key insights:

- Higher conversion rates significantly lower the CPA.
- Higher ad costs and more clicks slightly increase CPA, but their impact is smaller compared to conversion rates.

However, the model isn't perfect — some assumptions were not fully met, which means there might be patterns in the data that the model doesn't capture. In the future, we can test more advanced methods to improve accuracy.

Overall, this model provides valuable insights to optimize marketing budget, improve ROI, and focus efforts where they matter most — improving conversion rates.

4. Conclusion and Recommendations

This analysis has provided valuable insights into the performance and cost-efficiency of marketing campaigns. While overall conversion rates across channels are similar, and CTR is stable and aligned with expectations, **Cost Per Acquisition (CPA) stands out as an area requiring immediate attention**.

Key Recommendations:

1. **Investigate CPA Outliers:** Conduct a deeper dive into the 810 campaigns identified as CPA outliers. Understand the underlying reasons for unusually high or low acquisition costs, as this could reveal opportunities for cost optimization or successful strategies to replicate.

2. **Optimize CPA Strategies:** Given that the average CPA is significantly different from the ₹300 benchmark, re-evaluate and optimize strategies aimed at reducing acquisition costs. This could involve refining targeting, adjusting bids or improving ad relevance for higher conversion efficiency.

3. **Maintain CTR Performance:** Continue to monitor CTR, as its current stability suggests effective ad engagement. While not a primary focus for improvement based on this analysis, ensuring its continued performance is crucial.

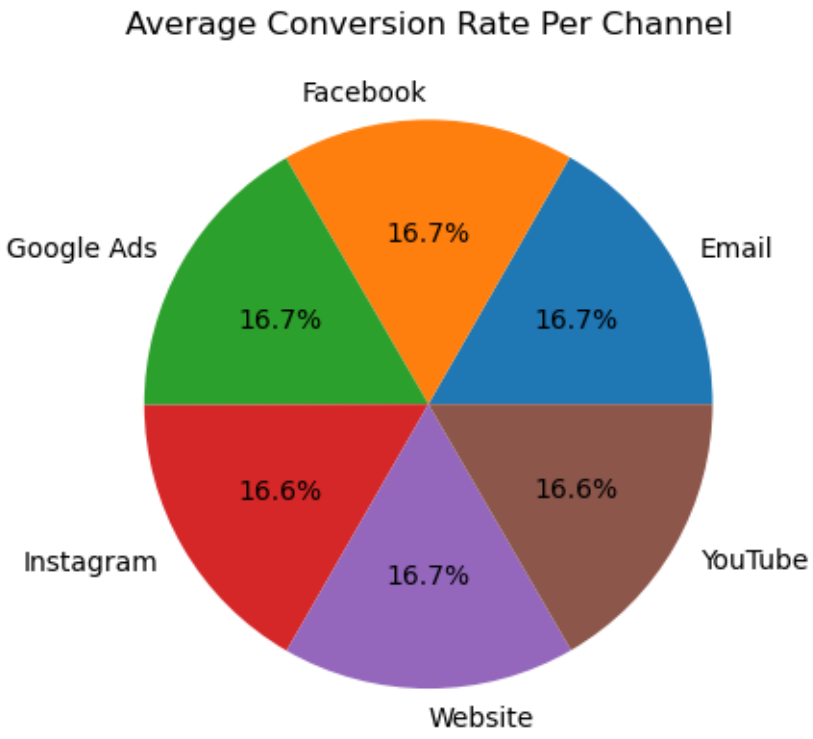
4. **Focus on Campaign Quality over Raw Spend/Clicks:** Reiterate that merely increasing acquisition spend or the number of clicks does not guarantee improved ROI or conversion rates. Emphasis should be placed on **quality of targeting, campaign design, and ad relevance**.

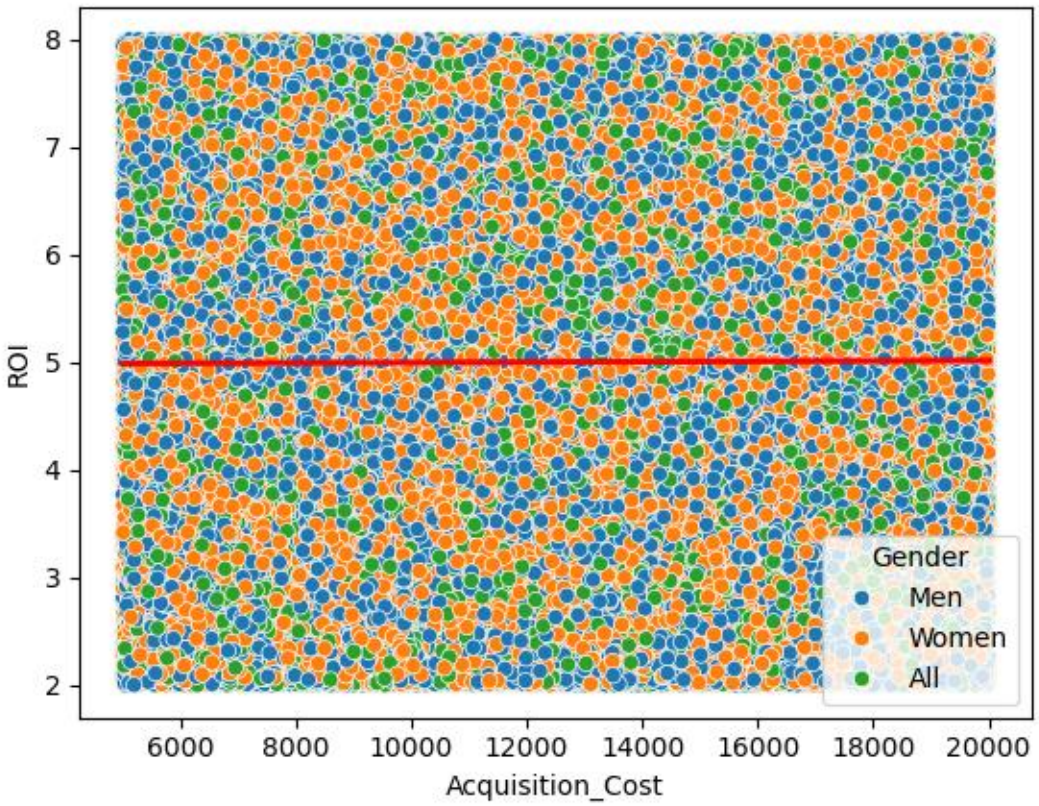
5. **Leverage Demographic Insights:** While ROI is consistent across genders and age groups, the higher engagement from the 25-34 age group suggests this demographic is highly responsive.

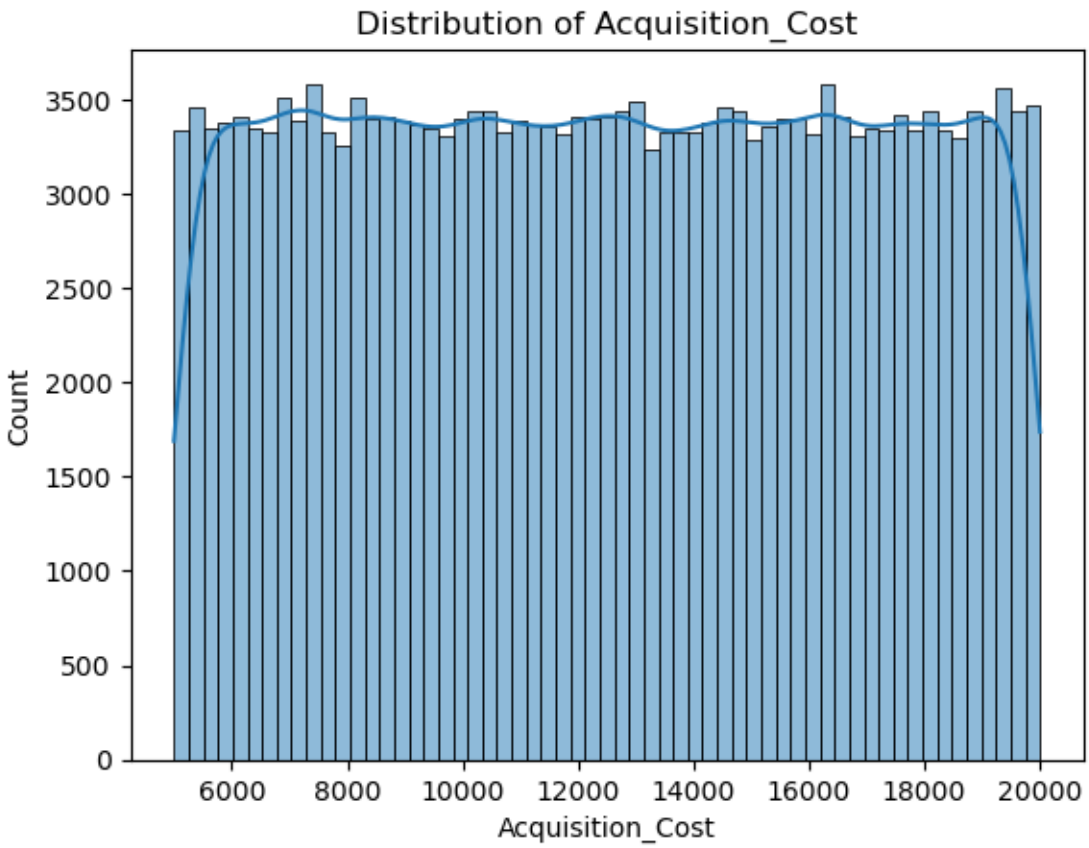
Consider further understanding and potentially leveraging this responsiveness in future campaigns.

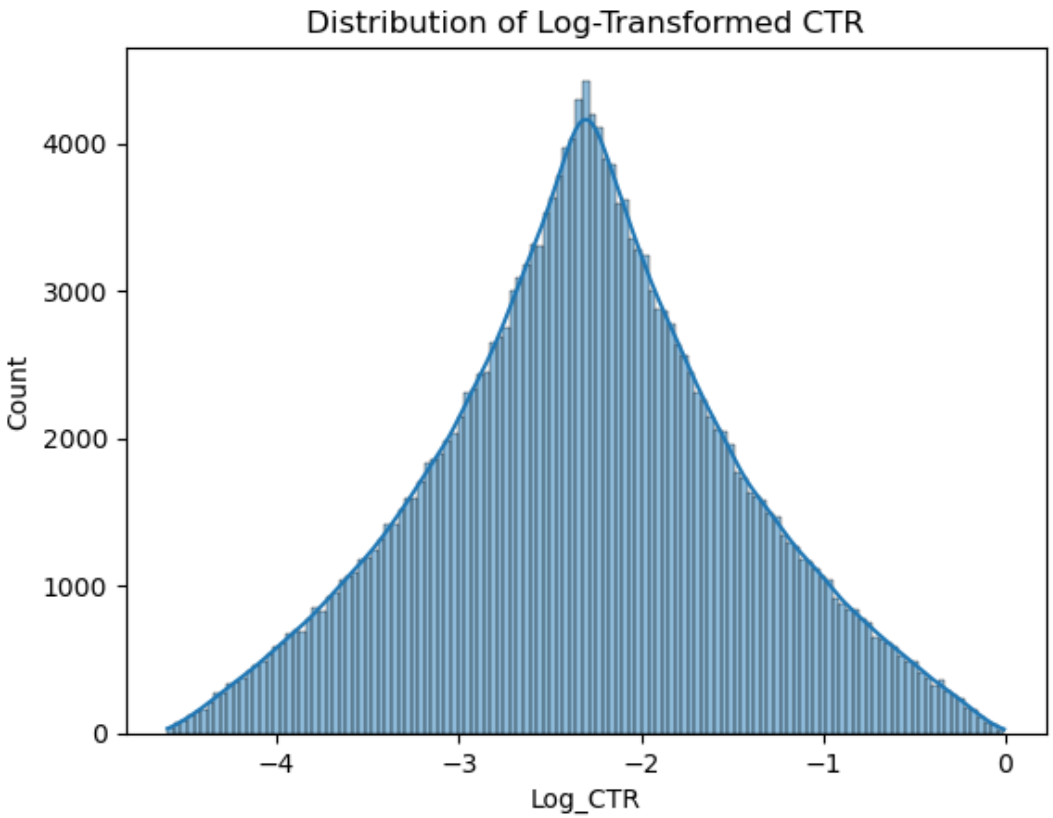
6. **Use Predictive Modeling to Guide Budget Decisions**: A regression model built on campaign data shows that over 86% of the variation in CPA can be explained by three factors: Cost Per Click, Conversion Rate, and Number of Clicks. The model highlights that improving conversion rates has the most significant impact on reducing CPA. This insight can help prioritize efforts where they yield the most return and inform future media planning and performance optimization strategies.
By focusing on these areas, particularly optimizing CPA, stakeholders can make data-driven decisions to enhance marketing campaign efficiency and maximize Return on Investment.

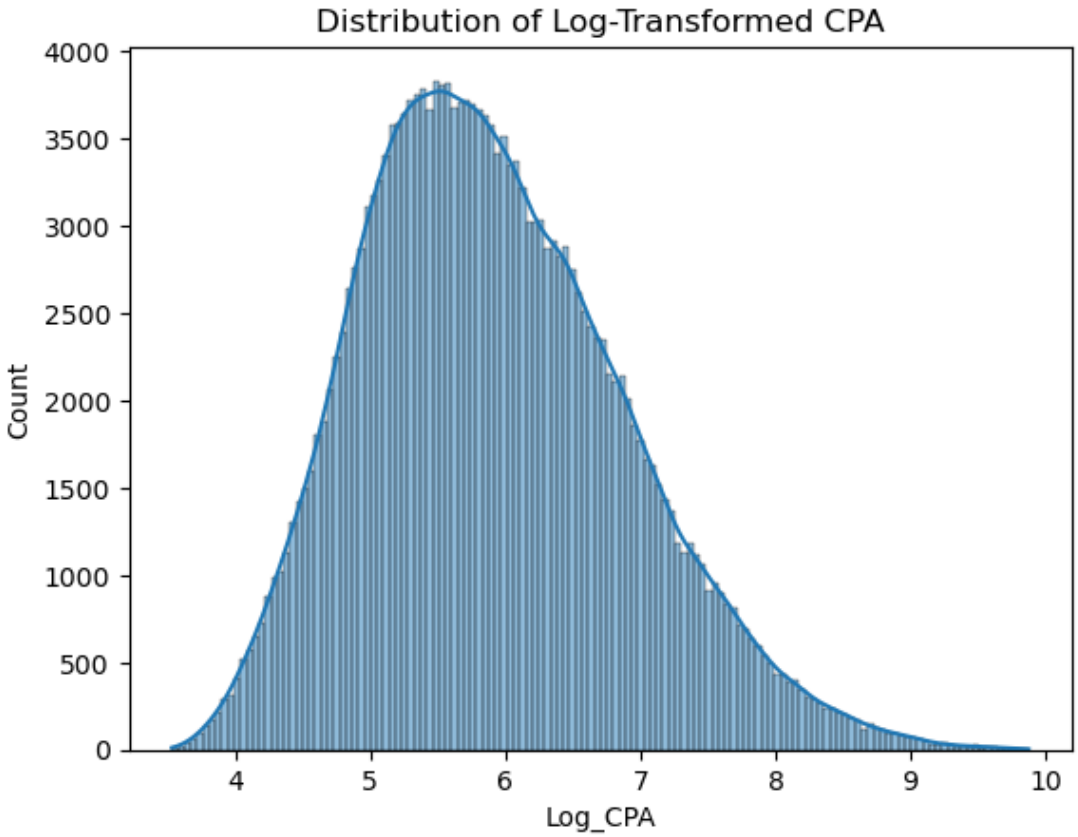
Visuals:




ROI
Acquisition_Cost
Gender
Men
Women
All
2
3
4
5
6
7
8
6000
8000
10000
12000
14000
16000
18000
20000


Distribution of Acquisition_Cost
Count
Acquisition_Cost
0
500
1000
1500
2000
2500
3000
3500
6000
8000
10000
12000
14000
16000
18000
20000


Distribution of Log-Transformed CTR
Count
Log_CTR
4000
3000
2000
1000
0
−4
−3
−2
−1
0


Distribution of Log-Transformed CPA
Count
Log_CPA
4000
3500
3000
2500
2000
1500
1000
500
0
4
5
6
7
8
9
10

Regression Analysis Visuals:

1. **Model Summary**

OLS Regression Results

Dep. Variable:	Log_CPA	**R-squared:**	0.869
Model:	OLS	**Adj. R-squared:**	0.869
Method:	Least Squares	**F-statistic:**	4.424e+05
Date:	Mon, 21 Jul 2025	**Prob (F-statistic):**	0.00
Time:	16:33:30	**Log-Likelihood:**	-70673.
No. Observations:	199190	**AIC:**	1.414e+05
Df Residuals:	199186	**BIC:**	1.414e+05
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	8.6356	0.016	550.280	0.000	8.605	8.666
Cost_Per_Click	0.0177	5.11e-05	346.460	0.000	0.018	0.018
Conversion_Rate	-15.7372	0.019	-823.162	0.000	-15.775	-15.700
Log_Clicks	-0.3300	0.002	-142.754	0.000	-0.334	-0.325

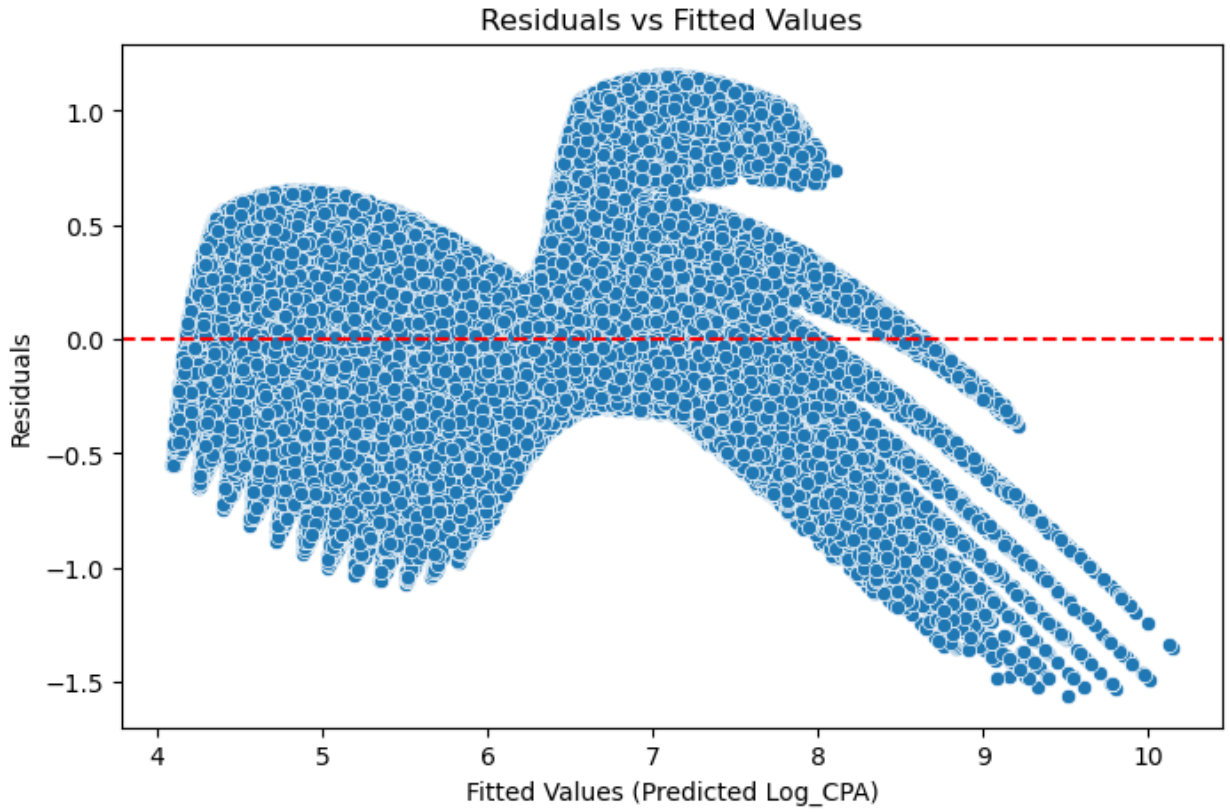
Omnibus:	3188.819	**Durbin-Watson:**	2.002
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	6222.455
Skew:	-0.050	**Prob(JB):**	0.00
Kurtosis:	3.860	**Cond. No.**	1.04e+03

Insights:

- When Cost_Per_Click, Conversion_Rate and Log_Clicks are used as independent variables X, the R-square appears to be 0.869. This means 86.9% of variance in Log_CPA is explained by Cost_Per_Click, Conversion_Rate and Log_Clicks indicating a strong fit.
- p-value for all coefficients are zero which means all coefficients are statistically significant.
- Regression Equation: Log_CPA = 8.63 + 0.017*Cost_Per_Click - 15.73*Conversion_Rate - 0.33*Log_Clicks

- For every unit increase in CPC, Log_CPA increases slightly by 0.0177 units, holding other variables constant.
- A strong inverse relationship; higher conversion rate substantially reduces Log CPA.
- As number of clicks increases (log-transformed), Log CPA decreases slightly.

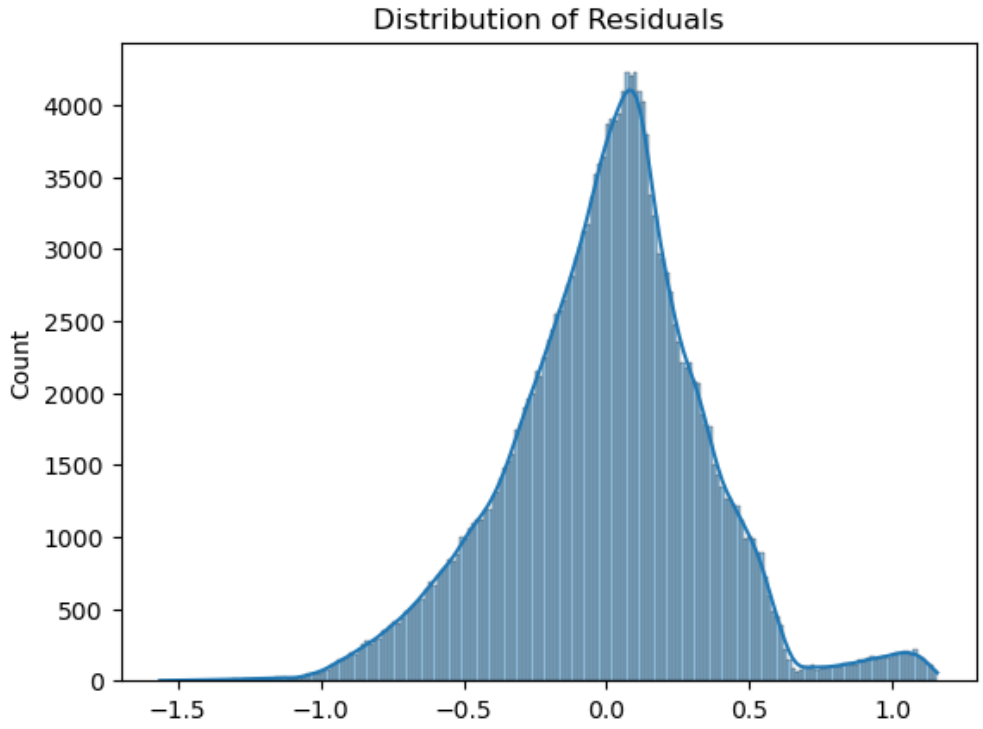
2. Checking Linearity

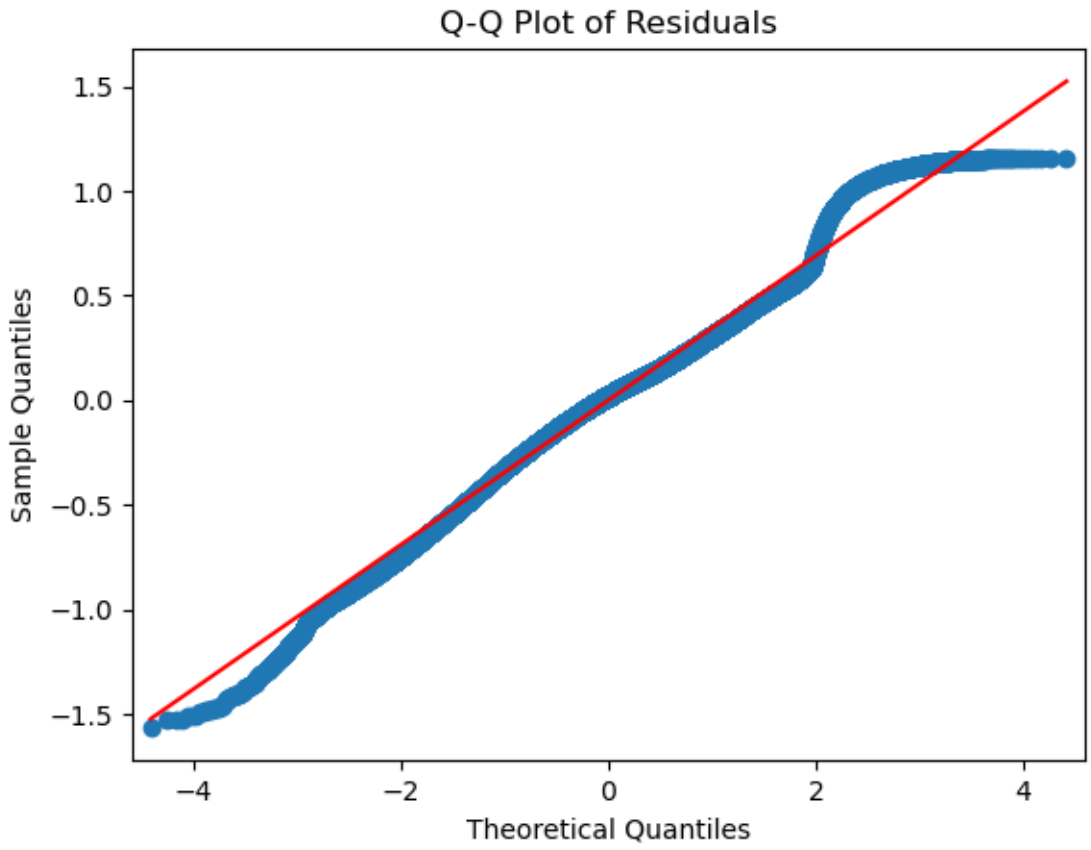


Insights:

- The residuals are not randomly scattered around the red horizontal line at zero.
- There's a clear and complex pattern, indicating non-linear relationships or interactions among predictors that the linear model is failing to capture.
- This plot violates the linearity assumption.

3. Checking Normality





Insights:

- Residuals are approximately normally distributed which meets our normality assumption.

4. Checking for Homoscedasticity