Customer Engagement, Retention, and Conversion Analysis for Acem.ai

Executive Summary:

This report is to present the analysis that explores key factors affecting customer engagement, retention, and conversion using detailed data insights for **Acem.ai**. It identifies an alarming decline in new customer acquisition over time, which impacts the conversion rate with the number of paying customers. The relationship between new customers and paying customers shows a strong positive correlation, suggesting that acquiring new customers directly influences the number of paying customers. Additionally, the conversion rate plays a significant role in this process, with a moderate to strong correlation to paying customers.

The regression model confirms that both new customer acquisition and conversion rates are crucial in driving the number of paying customers. This analysis also highlights two distinct customer clusters: one with high engagement and conversion rates, and another with low engagement and conversion rates, indicating the need for targeted strategies to improve engagement and retention.

Overall, the analysis provides clear recommendations focused on improving customer acquisition strategies, optimizing conversion processes, and enhancing customer engagement to drive growth in paying customers.

Introduction:

The purpose of this analysis is to gain a deeper understanding of the factors influencing customer engagement, retention, and conversion over time. By examining key metrics such as new customer acquisition, paying customers, churn rate, and conversion rates, the goal is to identify trends and relationships that can guide strategic decisions to improve customer retention and business growth.

About Acme.ai:

It's a company that uses artificial intelligence technology to create interactive experiences that allow users to converse with historical figures and gain valuable insights into their lives.

Data Sources: The analysis is based on key metrics that reflect user engagement and platform performance:

1. **Active Users:** Daily active user counts, tracking engagement on the platform.
2. **Landing Page Visitors:** Daily visitor counts to the platform's landing page.

3. **Subscription Churn:** Tracks the rate at which subscribers discontinue services, measuring retention and satisfaction.
4. **Subscription Conversion:** Measures the conversion rate from free users to paying subscribers, reflecting the effectiveness of our monetization strategies.

Methodology

1. Data Collection and Cleaning

The analysis began by loading and cleaning the four datasets using Pandas in Jupyter Notebook:

- **Parsing and Cleaning Metadata**: Irrelevant rows (such as those indicating start and end dates) were removed to ensure only meaningful data was extracted.
- **Active Users and Landing Page Visitors Datasets**: These datasets contained extra metadata rows, which were skipped, and values were adjusted using appropriate delimiters for proper parsing.
- **Handling Landing Page Visitors**: The dataset was cleaned by separating "Active" and "New" visitor types into distinct Data Frames for better analysis.
- **Date Conversion**:
  - Extracted and converted the start and end dates from the "Nth Day" column.
  - Mapped "Nth Day" to actual dates using the extracted start date, creating a new "Nth day_date" column.
- **Data Filtering**: Filtered records to include only data within the specified time range to maintain consistency.
- **Time-Based Alignment**: Additional columns were added to align datasets for monthly analysis.
- **Merging**: The datasets were merged on common columns to create a unified dataset for deeper insights.

2. Descriptive Analysis

- **Summary Statistics**: Key metrics such as mean, median, and distribution were explored to understand overall trends.
- **Time Series Analysis**:
  - Changes in new customers, active users, and conversions over time were examined through visual trends.

    - Line plots helped in identifying seasonal variations and long-term behavioural patterns of users.

3. Advanced Analysis

- **Correlation Analysis**: Pearson correlation was applied to measure the strength of relationships between key variables such as new customer acquisition, conversion rate, and paying customers. This helped in identifying which factors had the most influence on platform performance.
- **Regression Analysis (OLS & Linear Regression)**:

  Ordinary Least Squares (OLS) regression was used to model the relationship between paying customers and key predictors. It is a **standard and powerful method** for estimating relationships between variables, providing coefficients that quantify the expected change in paying customers for a **one-unit** change in each independent variable.

  A linear regression model was implemented to quantify the relationship between paying customers and key predictors—new customer acquisition and conversion rates.
- **Feature Engineering**: Moving averages and percentage changes were created to identify trends and patterns in customer engagement and retention.
- **Clustering & Segmentation (PCA + K-Means)**:
    - **Principal Component Analysis (PCA)** was performed for dimensionality reduction while preserving data variance.
    - **K-Means Clustering** was used to segment users into groups based on engagement and conversion behaviours.
    - The analysis revealed two major customer clusters—one highly engaged with strong conversion rates, and another with low engagement and poor conversion. This segmentation aids in designing targeted engagement strategies.

Results:

Key Findings:

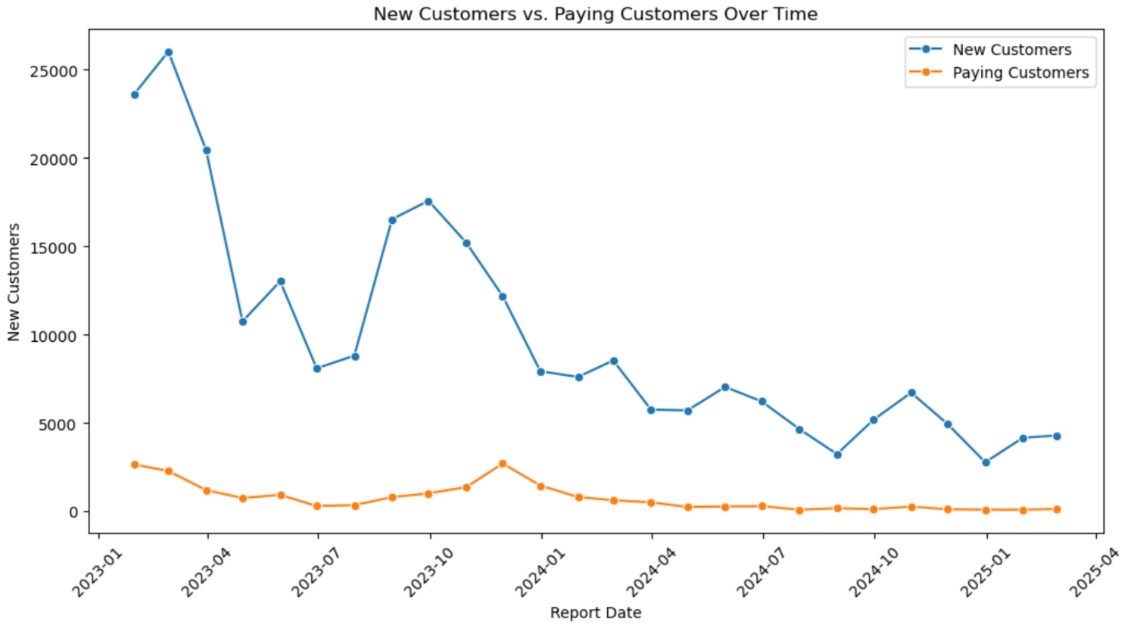
1. New Customers vs. Paying Customers Over Time

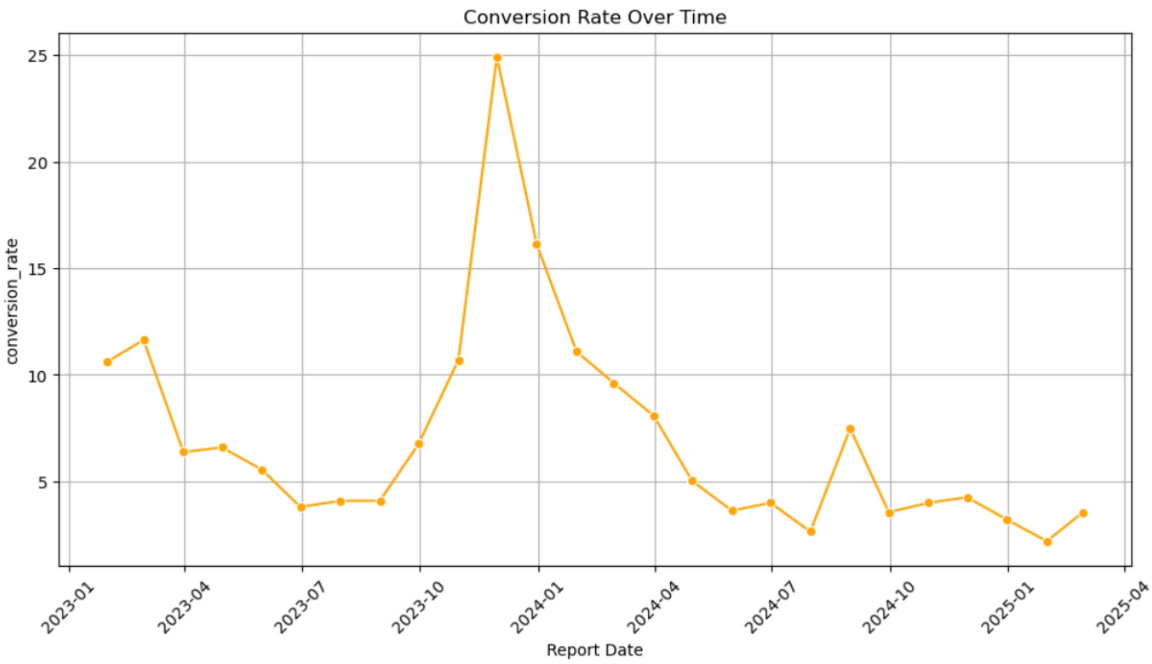
Insight:

- The New Customers count declined significantly from early 2023 to 2025, reflecting a decrease in customer acquisition effectiveness.

- Similarly, Paying Customers had a downward trend in the latter half of 2023 to 2025.

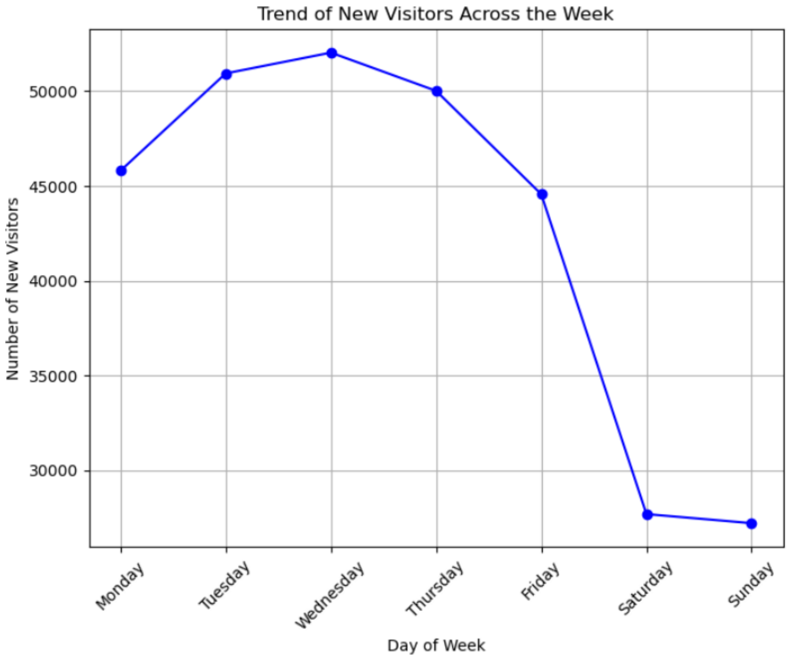
- The conversion dropped dramatically aligning with the decline in Paying Customers, indicating that fewer new customers were successfully converted into paying customers as time progressed.



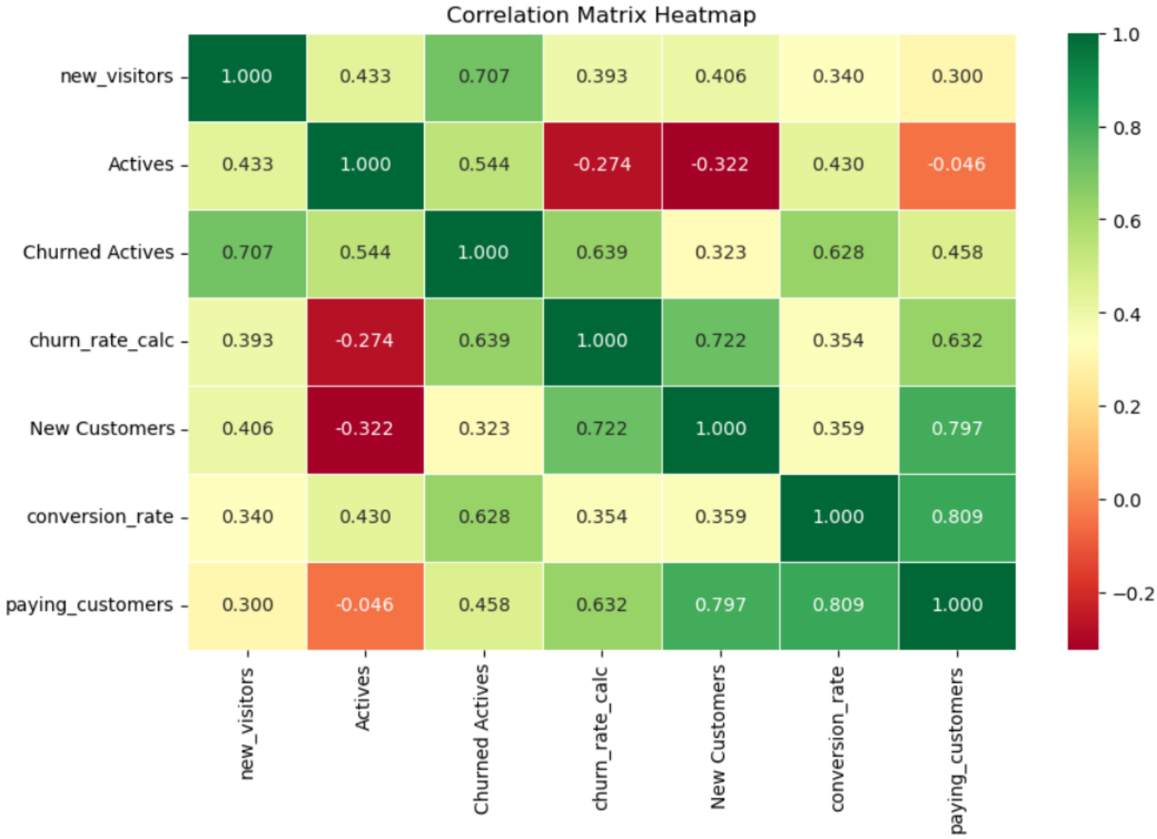


2. New Visitors across the week:

   The analysis reveals that **'Wednesday's'** experiences the highest number of new visitors. This indicates a potential peak in user engagement on this day over time. The trend of new visitors across the week shows fluctuations, suggesting that visitor activity varies significantly depending on the day.



3. Correlation between variables:



Strong Correlation:

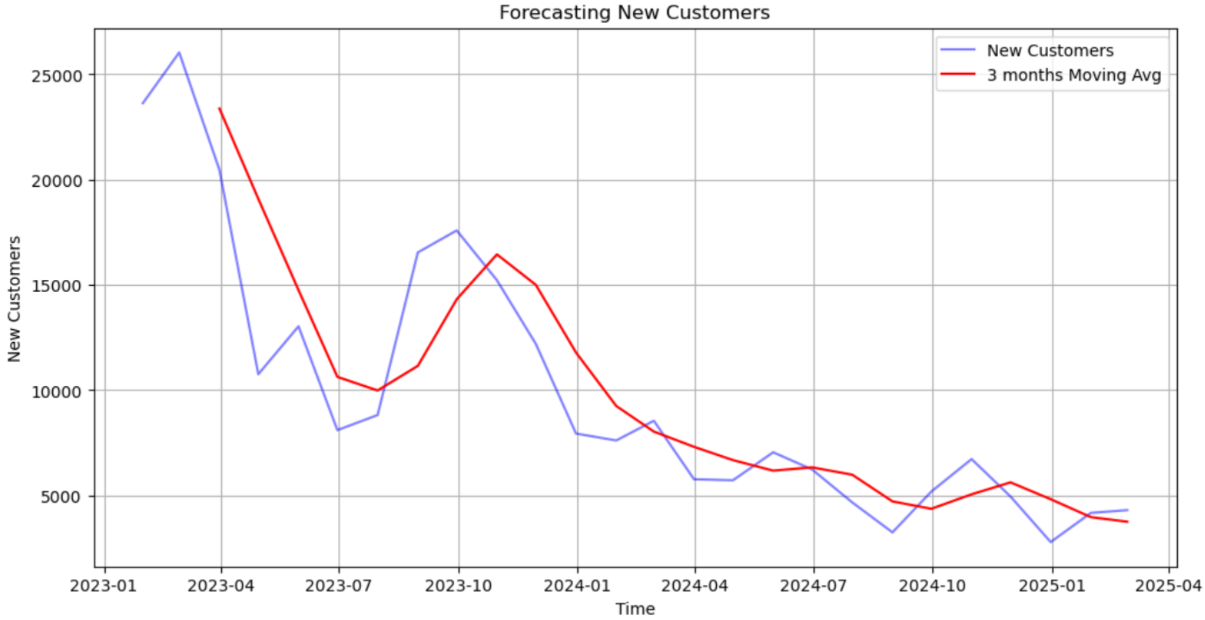
New Customers & Paying Customers: Correlation of 0.80, indicating a strong positive relationship. As the number of customers increases, paying customers also significantly rise.

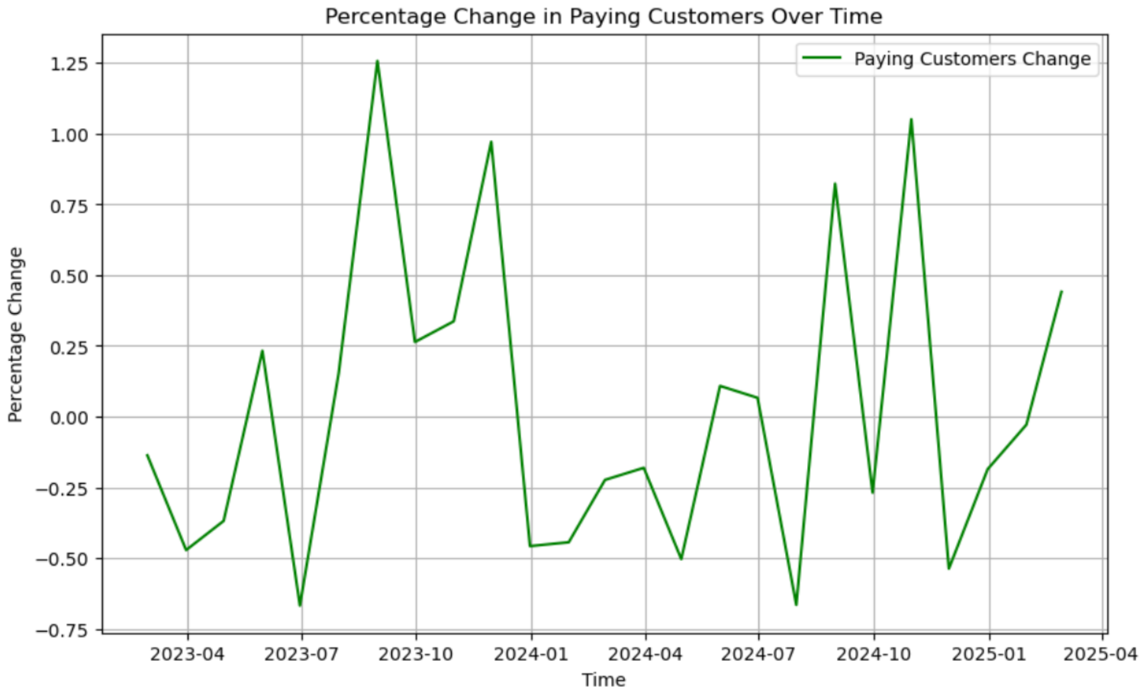
Moderate Correlation:
Churn Rate & Paying Customers: Correlation of 0.63, showing a moderate positive relationship.

Possible reasons:
- The company is focusing on acquiring new customers to increase the payin customers and conversion rate to replace churned ones.
- It could also be influenced by other factors such as product improvements, pricing changes, or targeted marketing.

Conclusion:
The strong correlation between new customers and paying customers remains the most impactful. However, the moderate correlation between churn rate and paying customers suggests that higher churn might be driving the need to focus more on paying customers.
Further analysis to better understand this relationship and determine how to reduce churn while still acquiring the number of customers can help to increase the retention rate.

4. Forecasting New Customers and Percentage change in Paying Customers over time :





a. New_Customers_Change:
Declining Trend: Looking at the New_Customers_Change column, we can observe a general decline in the number of new customers as the months progress. This suggests that fewer new customers are being acquired over time.

b. Paying_Customers_Change:
Rising Trend: On the other hand, the Paying_Customers_Change column shows more positive growth in the number of paying customers. This suggests that while the number of new customers is declining, the number of paying customers is rising, which could indicate that existing customers are converting to paying customers.

This forecast gives that the company focuses more on converting existing customers into paying ones, rather than focusing solely on acquiring new customers.

5. Multiple Linear Regression (OLS):

R-squared (0.986):
98.6% of the variation in the number of paying customers is explained by the combination of all the independent variables, indicating a very strong relationship.

Key Coefficients:
- New Customers (0.4515):
For each new customer acquired, the number of paying customers increases by 0.4515. This shows a strong positive and statistically significant relationship with paying customers, with a p-value of 0.000.

- Conversion Rate (0.7452):
Each unit increase in the conversion rate results in an increase of 0.7452 in paying customers. This also shows a significant and positive effect, with a p-value of 0.000, emphasizing the importance of conversion rate.

Features for Future Modelling:
New Customers (strong positive effect)
Conversion Rate (strong positive effect)

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:      paying_customers   R-squared:                       0.986
Model:                           OLS   Adj. R-squared:                  0.981
Method:                Least Squares   F-statistic:                     221.4
Date:               Wed, 19 Mar 2025   Prob (F-statistic):           1.54e-16
Time:                       09:12:07   Log-Likelihood:                 18.510
No. Observations:                 26   AIC:                            -23.02
Df Residuals:                     19   BIC:                            -14.21
Df Model:                          6                                         
Covariance Type:           nonrobust                                         
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const            6.939e-18      0.027   2.55e-16      1.000      -0.057       0.057
new_visitors       -0.0202      0.052     -0.390      0.701      -0.129       0.088
Actives             0.2246      0.129      1.743      0.098      -0.045       0.494
Churned Actives    -0.5653      0.186     -3.033      0.007      -0.955      -0.175
churn_rate_calc     0.4723      0.166      2.851      0.010       0.126       0.819
New Customers       0.4515      0.059      7.633      0.000       0.328       0.575
conversion_rate     0.7452      0.043     17.267      0.000       0.655       0.836
==============================================================================
Omnibus:                       14.572   Durbin-Watson:                   2.000
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               17.692
Skew:                          -1.225   Prob(JB):                     0.000144
Kurtosis:                       6.214   Cond. No.                         17.8
==============================================================================
```

6. Pearson correlation coefficient

Correlation between New Customers and Paying Customers: 0.797
P-value: 1.0984229506387437e-06

Correlation between Conversion Rate and Paying Customers: 0.809
P-value: 5.583034971089561e-07

<u>*New Customers and Paying Customers*</u>

- Correlation Coefficient: 0.79 (strongly positive): This proves that the increase in the number of customers will consequently improve the paying customers. P-value: 1.10e-06 (Less than the 0.05 significance level), which indicates the significant relation between the new customers and the paying customers.

<u>*Paying Customers and Conversion Rate*</u>

- Correlation between Conversion Rate and Paying Customers: 0.80: This proves that the increase in the number of paying customers will consequently improve the conversion rate. P-value: 5.583034971089561e-07 (Less than the 0.05 significance level), which indicates the significant relation between the paying customers and the conversion rate.

Conclusion:

Correlation between New Customers, Paying Customers and conversion Rate, indicates these 3 parameters are more significant and makes them the key features for further analysis and modelling.

7. Hypothesis Analysis:

Null Hypothesis (H_0):

The relationship between "new customers" and "paying customers" is not significant.

Alternative Hypothesis (H_1):

The relationship between "new customers" and "paying customers" is significant.

Conclusion:

<u>Correlation between New Customers and Paying Customers:</u> 0.79

<u>P-value:</u> 1.0984229506387437e-06

The p-value of 1.0984e-06 is very small (much less than the commonly used significance level of 0.05), indicating to reject the null hypothesis, indicating the relationship is statistically significant the evidence to suggest that an increase in the number of customers significantly increases the paying customers and in turn the conversion rate.

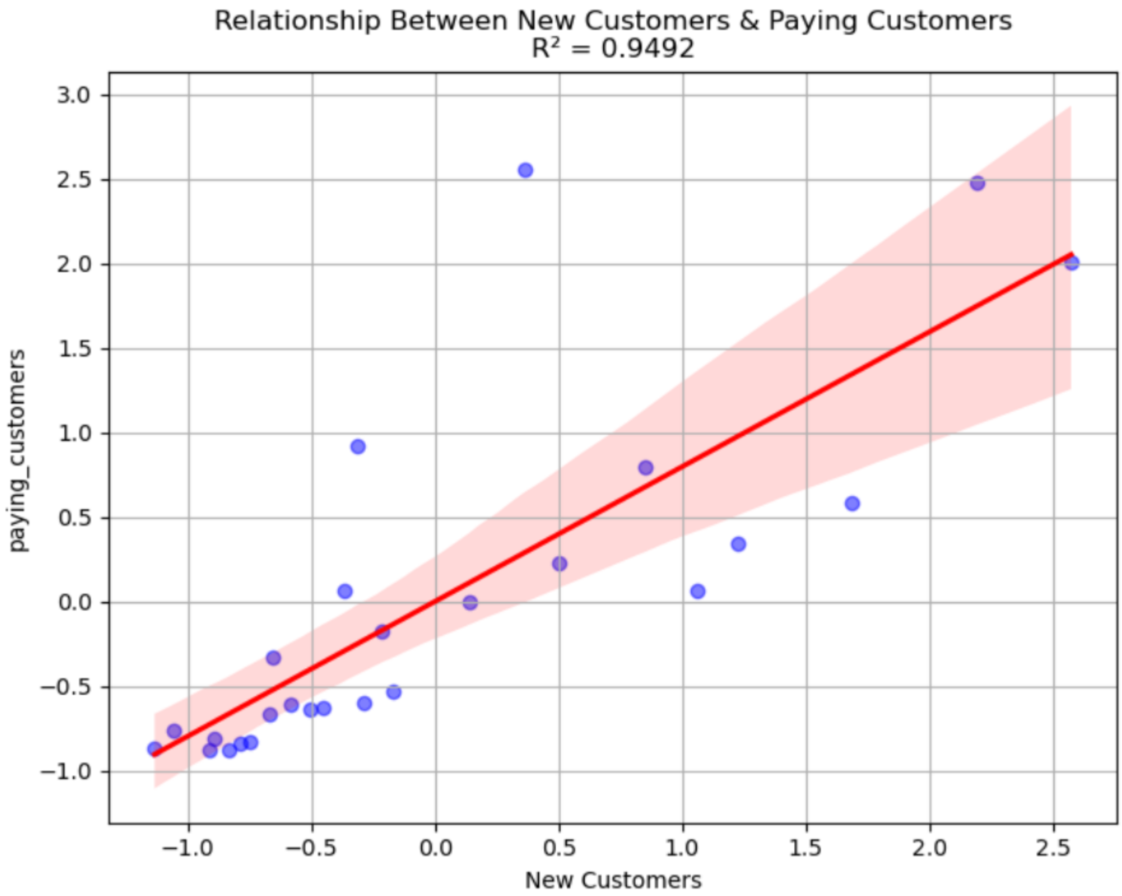
8. Customer Acquisition and Its effect on Paying Customers

Model Performance:

The model explains 94.92% of the variance in the number of paying customers, which indicates that the independent variables (New Customers and Conversio n Rate) are highly effective in predicting the number of paying customers.

For each unit increase in New Customers, the number of paying customers incr eases by 0.581366 (when the Conversion Rate is constant).

Similarly, for each unit increase in the Conversion Rate, the number of paying customers increases by 0.581366 (when the New Customers is constant).



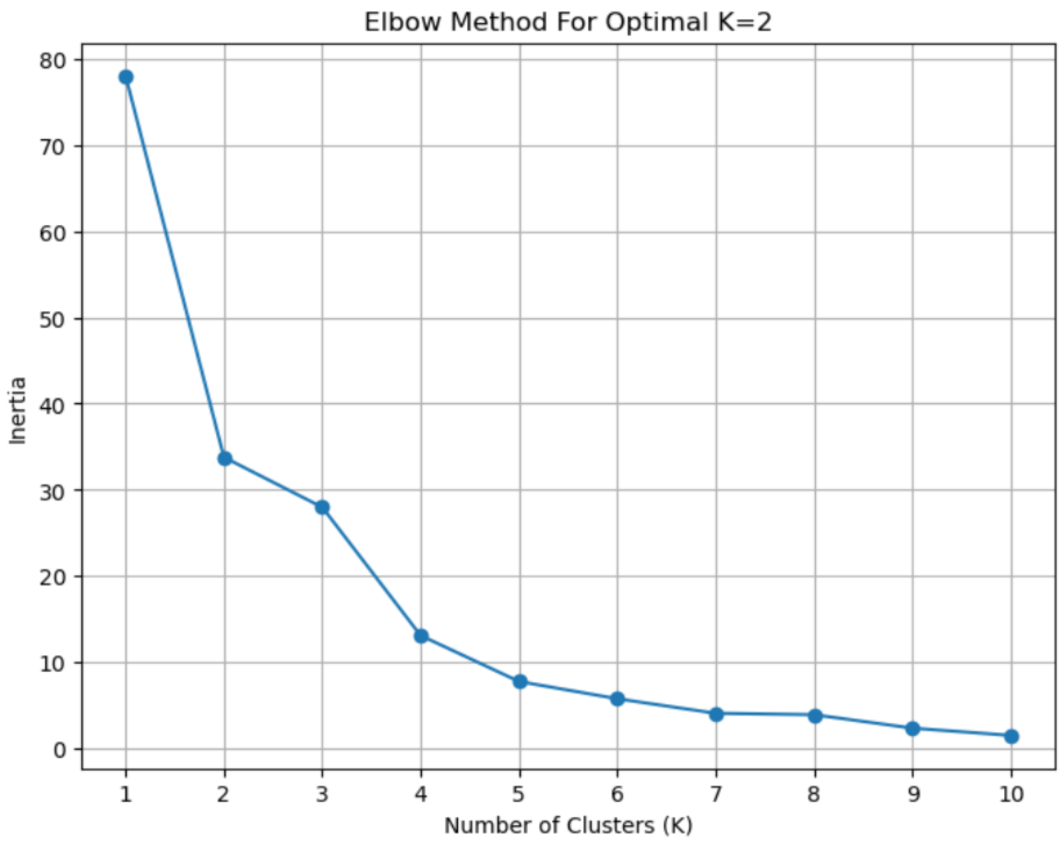
Model Evaluation:

This is a relatively low error, indicating that the model performs well in terms o f making predictions, not significantly large outliers and low prediction error.

9. Segmentation and Customer Analysis

Optimal Cluster Selection for Customer Analysis Using the Elbow Method:

Based on the **Elbow Method** for determining the optimal number of clusters, the optimal number of clusters (K) is **2**. This "elbow" point suggests that dividing the data into two clusters provides a meaningful segmentation of the data.



Exploring Customer Segmentation with PCA for Dimensionality Reduction:

Explained Variance Ratio: [0.77650173 0.21362627]

- PC1 explains about 77.65% of the variance in the data.
- PC2 explains about 21.36% of the variance.

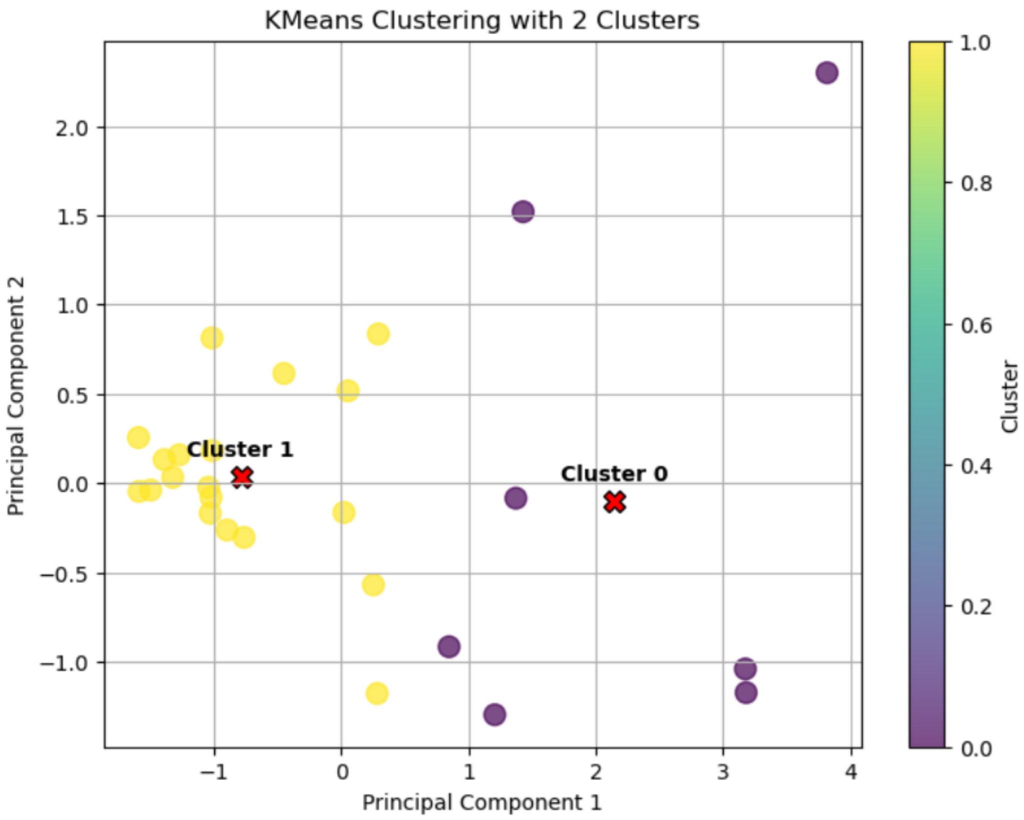
Together, they explain 99.01% of the total variance in the dataset.

PCA Components:

- PC1: This component is likely to represent a general trend where both New Customers and paying customers are positively correlated. It captures more of the variability in New Customers and paying customers.
- PC2: This component seems to capture the relationship between New Customers (negative correlation) and conversion rate (positive correlation). It suggests a second dimension that might highlight contrasts between the two.

Conclusion:

PC1 represents a general trend between New Customers and paying customers, while PC2 highlights a contrast between New Customers and conversion rate.



- Cluster 0: High engaged customers with high conversion rate. These customers show strong purchasing behaviours and they seem to be already benefiting from the product or service, indicating that they are active and responsive.

- Cluster 1: Low engaged customers with low conversion rate. These users may have shown initial interest but haven't followed through with purchases or have limited interaction with the platform. This group may be at risk of churn, as they haven't been effectively converted into paying customers or engaged users.

Recommendations

- Customer Retention and Engagement Strategies:
  Implement strategies like personalized offers or loyalty programs to improve customer retention and engagement.

- Marketing Campaigns & Acquisition Programs:
  A possible focus on conversion rates (getting existing/new customers to pay) rather than purely customer acquisition could cause this result. Increasing marketing campaigns and other factors to bring in more visitors can raise both the number of customers and the conversion rate.

- Conversion Rate Optimization:
  Focus on New Customers and Conversion Rate as key drivers of paying customers, given their significant positive relationships.
- Focus on New Customers & Conversion Rate:
  The number of paying customers is strongly influenced by New Customers and Conversion Rate, highlighting the importance of these two factors.
- Improving Customer Acquisition:
  Strategies to bring in more new customers should be prioritized, along with optimizing conversion strategies to turn more visitors into paying customers and improve overall customer retention.
- Cluster-Specific Strategies:
  - For Cluster 0 (High Engagement/High Conversion):
    Retain the customers with retention programs to fulfill their needs and reduce churn.
  - For Cluster 1 (Low Engagement/Low Conversion):
    Strategies to bring in more customers through paid advertisement or social media marketing. Plans to make the user experience more conversational and engaging. Additionally, target market campaigns with email campaigns and promotions to increase user engagement and optimization.

Conclusion:

This analysis provides an overview of the factors driving customer engagement, retention, and conversion at Acem.ai. The findings reveal a strong relationship between new customer acquisition, conversion rates, and the number of paying customers, highlighting the critical role these metrics play in driving growth. The regression and correlation analyses demonstrate that both new customer acquisition and conversion rates are significant predictors of paying customers. Additionally, the segmentation analysis identifies two distinct customer clusters, offering a clear opportunity to tailor strategies for high-engagement users and those at risk of churn.

By aligning marketing and engagement initiatives with the identified trends and behaviours, Acem.ai can foster stronger customer retention and drive sustainable growth in paying customers. Furthermore, continuously monitoring key metrics will help the company adjust strategies quickly in response to changes in customer behaviour and market trends.