QWE Case Study on Customer Churn Analysis

11/29/2023
BA815 Competing With Analytics

By:
Sricharan Mahavadi
Submitted to:
Prof.Oliver Bandte

Executive Summary:

QWE, Inc., a subscription-based platform dedicated to empowering small and medium-sized businesses in managing their online presence, conducted an in-depth analysis focused on predicting customer churn. Leveraging a 2-month customer dataset, our exclusive utilization of the Logistic Regression Model enabled the classification of customers into segments, offering insights into the probability of churn from the QWE platform.

The chosen Logistic Regression Model demonstrated its effectiveness, boasting a higher true positive rate when compared to alternative methods. Our investigation identified the top drivers influencing customer churn, prioritizing the following factors:

1. **CHI Score Month 0:** Reflecting the current happiness level of customers.
2. **Views 0-1:** Indicating changes in customer engagement within the last month.
3. **Days Since Last Login:** Highlighting the duration since the customer's last interaction with the platform.
4. **Chi Score 0-1:** Signifying changes in happiness levels compared to the previous month.

This strategic analysis provides valuable insights for QWE, suggesting that monitoring and enhancing current customer happiness levels, along with optimizing engagement strategies and addressing support cases promptly, will significantly contribute to customer retention efforts. As the QWE platform continues to support businesses in autonomously managing their online presence, focusing on these key drivers is integral to fostering lasting customer satisfaction and loyalty.

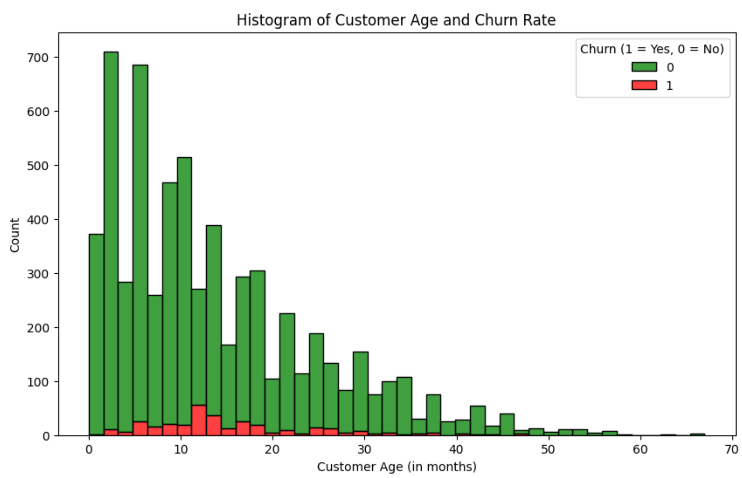
Introduction

Greetings! I'm stepping into the shoes of V.J. Aggrawal, the data scientist on a mission to crack the code of customer churn at company QWE. Armed with a dataset curated by Aggrawal, packed with state and "delta" variables, my goal is to whip up a logistic regression model. This model isn't just about numbers; it's the key to unraveling patterns and understanding why customers bid farewell. Together, we're diving into the data to unearth insights that will arm company QWE with the knowledge to tackle customer departures head-on. Let's decode the churn puzzle and set the stage for strategic decisions!

1. Age and Churn Rate Correlation Analysis

Is Wall's belief about the dependence of churn rates on customer age supported by the data?

To analyze the relation between Age and Churn rate, we plot the relation of Age and Churn.



Analysis 1: Histogram Visualization
Based on the histogram visualization, there is no clear and strong correlation between customer age and churn rate. The histogram illustrates that customers, regardless of age, exhibit both churn and non-churn behaviors. The distribution of churn and non-churn instances is spread across various age groups, and there is no evident pattern suggesting that customer age alone is a decisive factor in predicting churn. The lack of a concentrated trend in the histogram indicates that age, as a singular variable, may not be a strong predictor of churn in this dataset.

Analysis 2: Violin Plot Visualization
The Violin plot further supports the observation that there is nonclear and strong correlation between customer age and Churn rate. The plot reveals the distribution of churn and non-churn instances across different age groups.

Similar to the histogram, the violin plot indicates that customers across various age groups exhibit both Churn and non-churn behaviors.

Answer :

In conclusion, both the histogram and violin plot visualizations consistently depict a lack of a strong correlation between customer age and churn rate. The absence of a concentrated trend across different age groups suggests that other factors or a combination of factors might play a more significant role in influencing churn within the dataset. Therefore, the data visualizations do not strongly support Wall's belief regarding a direct dependence of churn rates on customer age.

Further statistical analysis and modeling may be required to uncover more nuanced insights and identify the key drivers of churn.

2. Regression Model Implementation :

Run a regression model that best predicts the probability that a customer leaves.

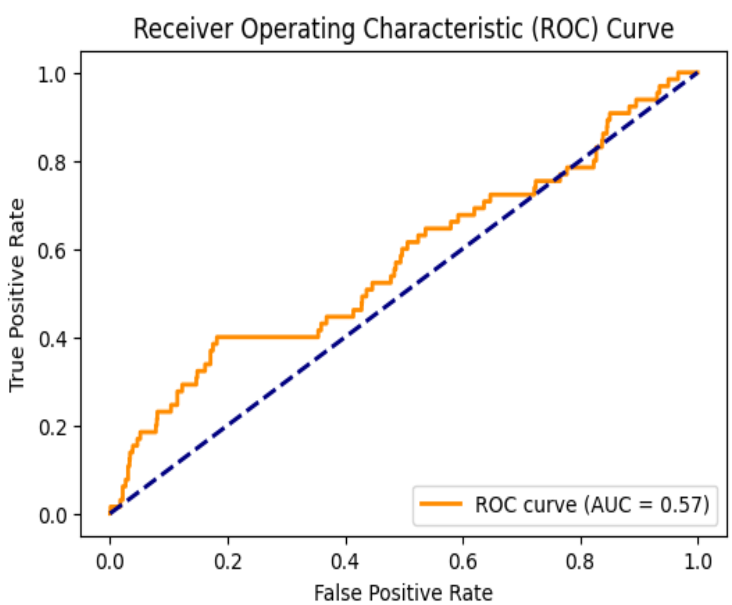
Answer:

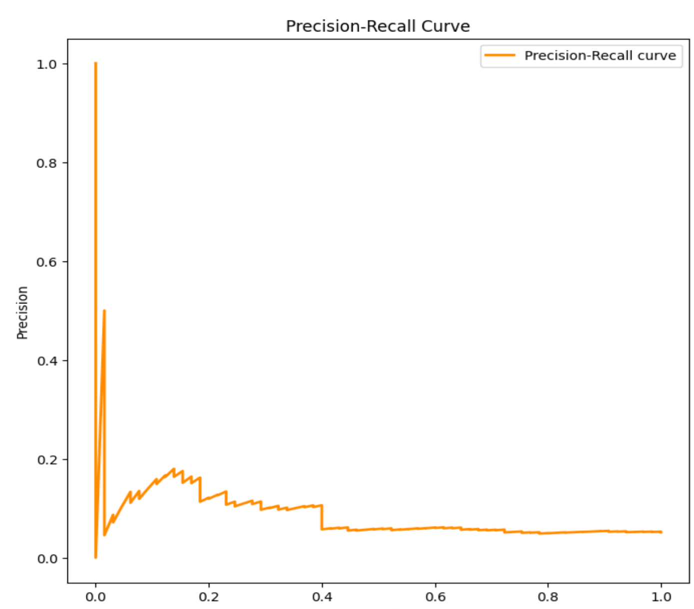
In our analysis, we have employed the Logistic Regression model to examine the likelihood of customer churn. Logistic Regression is a valuable tool for binary classification tasks, making it well-suited for predicting whether a customer will churn (class 1) or not (class 0). Our objective is to delve into the model's performance and effectiveness in capturing the probability of a customer leaving the platform. The following results showcase key metrics and insights derived from the Logistic Regression model, shedding light on its ability to accurately predict customer churn probabilities.

Results :

The results from the regression model implementation indicate a balanced accuracy of 0.5000. However, a closer examination of the confusion matrix and the classification report reveals a significant imbalance in the dataset. The model achieves high accuracy (95%) primarily due to accurately predicting the non-churn class (class 0), but it struggles to identify instances of the churn class (class 1). The precision, recall, and F1-score for the churn class are all very low, indicating that the model has difficulty distinguishing and correctly predicting instances of churn.

In the context of predicting the probability that a customer leaves, the regression model appears to face challenges, particularly in identifying customers who actually churn (class 1). The precision, recall, and F1-score metrics for the churn class suggest that the model's performance in predicting churn is limited. Further model refinement, feature engineering, or addressing the class imbalance may be necessary to improve its ability to predict the probability of customer churn more accurately.





2.1 Customer 672 Probability Analysis

What is the predicted probability that Customer 672 will leave between December 2011 and February 2012? Is that high or low? Did that customer actually leave?

Analysis:

The model predicts a probability of 0.0342, corresponding to a 3.42% likelihood that Customer 672 will churn. As the predicted probability is below the model's threshold (0.5), it indicates a low chance of the customer leaving, reflecting a high confidence in the prediction.

Comparison to Actual Dataset:
Upon inspecting the dataset, it's evident that the customer has not left, as the actual churn status is 0.

Output:
- Predicted Probability of Customer 672 Churning: 0.0342
- Actual Churn Status: 0
- Predicted Churn Status: 0
- The model prediction is correct.

2.2 Probability Assessment for Customers 354 and 5203

What about Customers 354 and 5203?

Analysis:

The Logistic Regression model predicts a low churn probability of 4.38% for Customer 354 and 4.28% for Customer 5203. Both predictions are below the 0.5 threshold, indicating high confidence in the model's forecast of customers not leaving.

Comparison to Actual Dataset:
Upon checking the dataset, both customers have an actual churn status of 0, aligning with the model's predictions.

Output:

- **Predicted Probability of Customer 5203 Churning: 0.0428**
- **Actual Churn Status: 0**
- **Predicted Churn Status: 0**
- **The model prediction is correct.**

Output:

- **Predicted Probability of Customer 354 Churning: 0.0438**
- **Actual Churn Status: 0**
- **Predicted Churn Status: 0**
- **The model prediction is correct.**

The model accurately predicts that these customers are unlikely to churn, showcasing its reliability in making correct predictions.

3. Key Contributing Factors

What factors contribute the most to the predicted probabilities that these customers will leave?

Analysis:

The provided output shows the churn probabilities and the top contributing factors for Customers 672, 354, and 5203. Let's analyze the information:

1. **Customer 672:**
   - **Churn Probability:** 0.0342
   - **Top Driver:** Days Since Last Login 0-1
   - **Second Driver:** Support Cases 0-1
   - **Third Driver:** Customer Age (in months)

   Analysis: Customer 672 has a very low churn probability of 0.0342. The top contributing factor indicating a low likelihood of churn is "Days Since Last Login 0-1." Additionally, "Support Cases 0-1" and "Customer Age (in months)" are also contributing factors suggesting stability in this customer's relationship with the service.

2. **Customer 354:**
   - **Churn Probability:** 0.0438
   - **Top Driver:** Days Since Last Login 0-1
   - **Second Driver:** Support Cases 0-1
   - **Third Driver:** Customer Age (in months)

   Analysis: Customer 354 has a slightly higher churn probability compared to Customer 672 (0.0438). Similar to Customer 672, the top contributing factor is "Days Since Last Login 0-1," indicating recent engagement with the service. "Support Cases 0-1" and "Customer Age (in months)" also contribute to the prediction of low churn likelihood.

3. **Customer 5203:**
   - **Churn Probability:** 0.0428
   - **Top Driver:** Days Since Last Login 0-1
   - **Second Driver:** Support Cases 0-1
   - **Third Driver:** Customer Age (in months)

   Analysis: Customer 5203 has a churn probability of 0.0428, which is similar to Customer 354. The top drivers, including "Days Since Last Login 0-1," "Support Cases 0-1," and "Customer Age (in months)," suggest that recent activity, support cases, and the overall duration of the customer relationship contribute to the model's prediction of a low likelihood of churn.

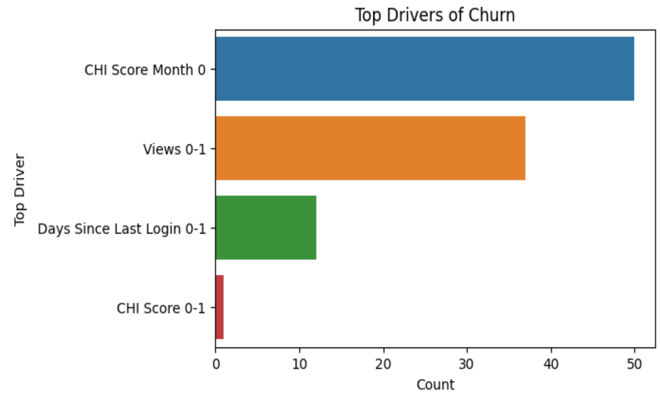
In summary, all three customers have low churn probabilities, and the top contributing factors highlight the importance of recent engagement ("Days Since Last Login 0-1"), support interactions ("Support Cases 0-1"), and the overall duration of the customer relationship ("Customer Age (in months)"). These factors collectively contribute to the model's prediction that these customers are not likely to leave.

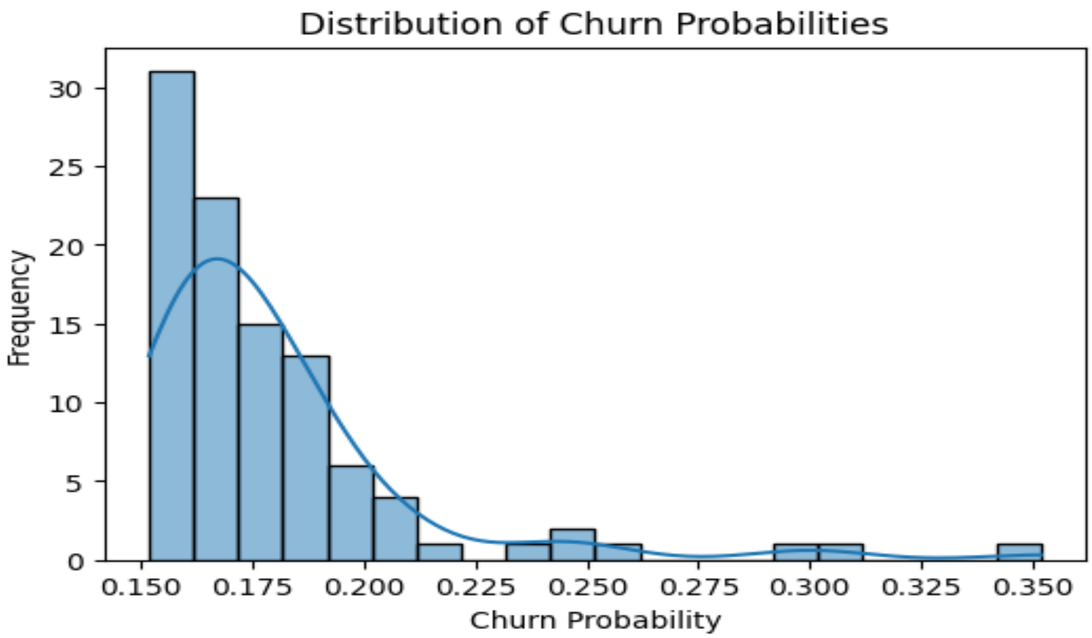
4. Top 100 Highest Churn Probability Customer

List of 100 customers with the highest churn probabilities and the top three drivers of churn for each customer.

My Analysis identified the top drivers influencing customer churn, prioritizing the following factors:

1. **CHI Score Month 0:** Reflecting the current happiness level of customers.
2. **Views 0-1:** Indicating changes in customer engagement within the last month.
3. **Days Since Last Login:** Highlighting the duration since the customer's last interaction with the platform.
4. **Chi Score 0-1:** Signifying changes in happiness levels compared to the previous month.





Business Implications

In summary, the analysis underscores the profound impact of key factors on customer churn within the QWE platform. Among these factors, the Customer Happiness Index (CHI) emerges as the most influential in shaping a customer's likelihood to churn. It becomes evident that, irrespective of the duration a customer has been with us, dissatisfaction with our service significantly heightens the probability of churn.

The strategic implication for QWE is clear: to mitigate churn effectively, the organization should prioritize initiatives aimed at enhancing customer satisfaction. The focus should extend beyond mere customer longevity, recognizing that sustained customer confidence in our service is paramount. As QWE continues to empower businesses in managing their online presence, the emphasis on customer satisfaction aligns seamlessly with the core mission of instilling confidence and trust among our valued customers.