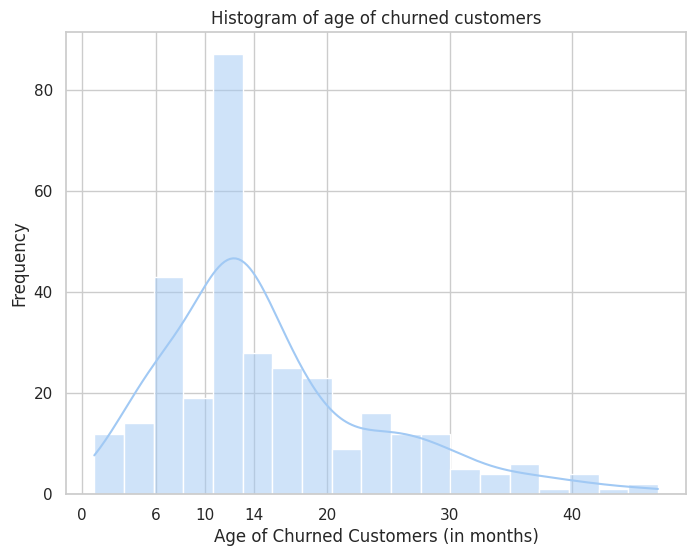
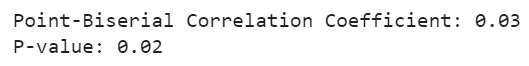
**QWE Case Study on Customer Churn Analysis**

**Questions & Answers:**

**1. Age and Churn Rate Correlation Analysis - Is Wall’s belief about the dependence of churn rates on customer age supported by the data?**





**Answer 1**

The analysis supports Wall's belief regarding customer age and churn. Plots, including a histogram and boxplot, reveal a concentration of churns for customers aged between 6 to 14 months, with lower rates for those below 6 months and above 14 months. The point-biserial coefficient further affirms a small positive correlation between customer age and churn, backed by a statistically significant p-value of 0.02. However, caution is advised, emphasizing that correlation does not imply causation, and the practical relevance of this correlation should be interpreted within the context of the specific domain.

**2. Regression Model Implementation - Run a regression model that best predicts the probability that a customer leaves.**

**Answer 2**

The logistic regression model attains a 95% accuracy score. However, the balanced accuracy score, at 0.5, signals difficulties in accurately predicting churn due to dataset imbalance (94% non-churned, 16% churned instances). The model's high accuracy may be skewed by the dominant class, underscoring the need for supplementary metrics like balanced accuracy for a more comprehensive evaluation in imbalanced scenarios.

**2.1 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**

**Answer 2.1**

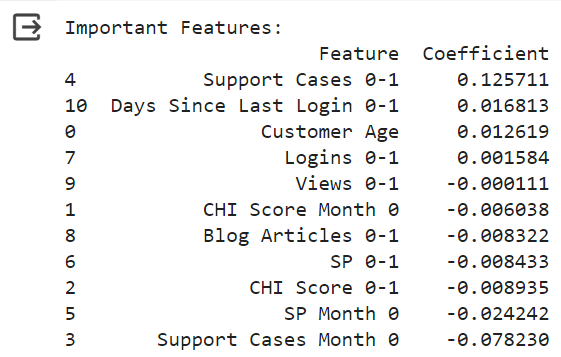
The probability that Customer 672 will leave between December 2011 and February 2012 is = 0.03 (less than 0.5). The obtained probability is low, indicating that customer 672 will stay. The actual churn indicator for Customer 672 = 0, which indicates that this customer stayed and our prediction is correct.

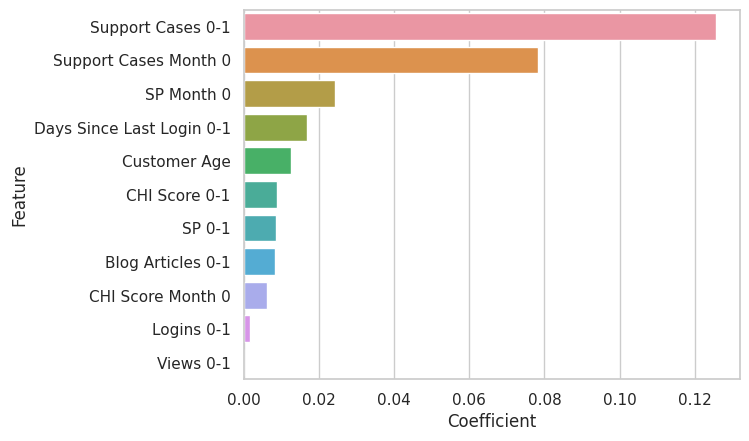
**2.2 What about Customers 354 and 5203?**

**Answer 2.2**

* The probability that Customer 354 will leave between December 2011 and February 2012 is = 0.04 (less than 0.5). The obtained probability is low, indicating that customer 354 will stay. The actual churn indicator for Customer 354 = 0, which indicates that this customer stayed and our prediction is correct.
* Similarly, the probability that Customer 5203 will leave between December 2011 and February 2012 is = 0.04 (less than 0.5) which is low, indicating that customer 5203 will stay and the actual churn indicator for Customer 5203 = 0, which indicates that this customer stayed and affirms our prediction.

**3. What factors contribute the most to the predicted probability that these customers will leave?**

****

****

**Answer 3**

The logistic regression model identifies the following most important features influencing customer churn predictions. Key findings include:

* Support Cases 0-1 (Coeff = 0.12): Significant positive impact on predicting churn.
* Support Cases Month 0 (Coeff = -0.07): Substantial negative impact on churn prediction.
* SP Month 0 (Coeff = - 0.02): Notable negative influence on churn prediction.

These insights underscore the importance of specific customer engagement metrics—Support Cases and SP—in shaping the model's churn predictions. Businesses should prioritize and monitor these features for effective customer retention strategies.

**4. Answer Wall’s “ultimate question”: provide the list of 100 customers with the highest**

**churn probabilities and the top three drivers of churn for each customer.**

**Answer 4**

Please refer to the attached Excel file - [Top 100 Customers.xlsx](https://docs.google.com/spreadsheets/d/1q4SbFlA9y8cJwRSWEZm2mt75TnM7gMs4/edit?usp=sharing&ouid=106814108103931879950&rtpof=true&sd=true)

*(Below is a snapshot of what the data for Top 100 customers with the Top 3 drivers influencing churn looks like)*

