1. Age and Churn Rate Correlation Analysis (2 Points)

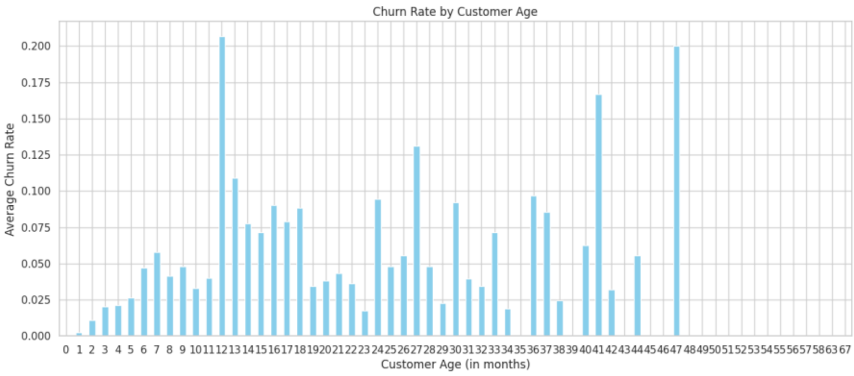

Diagram 1 reveals that customer age does influence churn rates, as evidenced by varying churn rates across age groups. However, the relationship is not straightforward or linear, suggesting that while age is a factor in churn, it is not the predominant or only determinant. This indicates a more complex interplay of factors influencing customer churn, with age being one of several contributors.

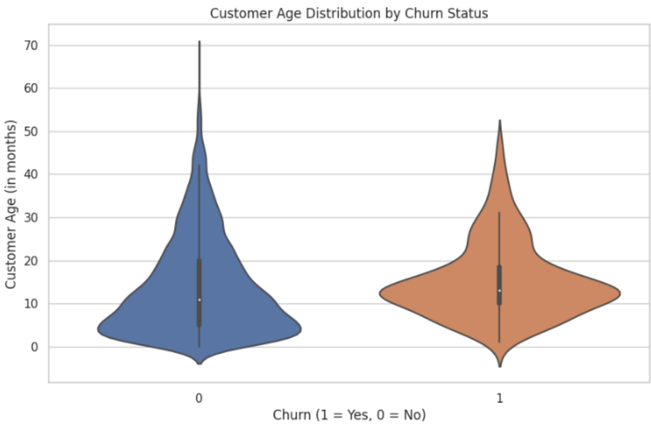

Diagram 2 shows the width of each "violin" indicates the density of data points at different ages. However, there is no stark contrast in the age distribution between the two groups, suggesting that customer age may not be a strongly distinguishing factor in churn.

These findings suggest that while there might be a slight difference in the average age of churned versus non-churned customers, the difference is not substantial. This implies that Wall's belief about a strong dependence of churn rates on customer age may not be fully supported by the data.

2. Regression Model Implementation (2 Points)

A. Customer 672:

The logistic regression model predicts an 8.81% probability of churn for Customer 672. This prediction indicates that it is low probability for customer 672 to leave during that period. In reality, Customer 672 did not leave.

B. Customer 354 and 5203:

For Customer 354, the model estimates a churn probability of approximately 5.72%. This indicates a low risk of churn. The actual data also shows that Customer 354 did not leave.

For Customer 5203 stands at a relatively low 3.07% probability of churn, the lowest among the three customers. This low probability suggests a minimal risk of this customer leaving the company. The actual data also confirms that Customer 5203 did not leave.

3. Key Contributing Factors (2 Points)

The feature "Support Cases 0-1" emerges as the most influential in predicting customer churn, with an importance score of 0.000179. This suggests that changes in the number of support cases over time are a critical indicator of churn risk. Following closely is "Views 0-1" with an importance score of 0.000158, indicating that variations in the number of views are also a significant predictor of churn, likely reflecting changes in customer engagement or satisfaction. The "CHI Score Month 0" and "Support Cases Month 0" are also notable predictors, albeit with lower importance scores of 0.000053 and 0.000021, respectively, suggesting that initial customer interactions and support needs are key indicators. Other factors like changes in CHI score, days since last login, and customer age show minimal to no impact on churn prediction in this model.

4. Addressing Wall’s Ultimate Question (2 Points)

Based on the data, the customer at the top of this list is ID 2287, with a predicted probability of approximately 40.00% of leaving the company.

Top three drivers:

1. Support Cases 0-1: Changes in the number of support cases over time are the most significant factor. This suggests that varying levels of customer service needs or satisfaction are critical in influencing a customer's likelihood of churning.
2. Days Since Last Login 0-1: The duration since a customer last logged in is another crucial factor. A longer period without customer engagement could indicate waning interest or satisfaction, thus increasing the risk of churn.
3. Customer Age (in months): The customer's tenure with the company, represented by their age in months, is also a key driver. This implies that both newer and longer-term customers have specific churn risks, possibly for different reasons.

For results of data and process of machine learning, please refer to the notebook.