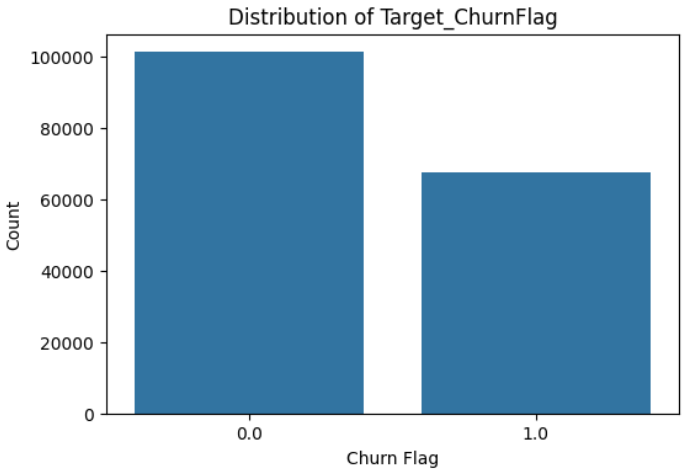
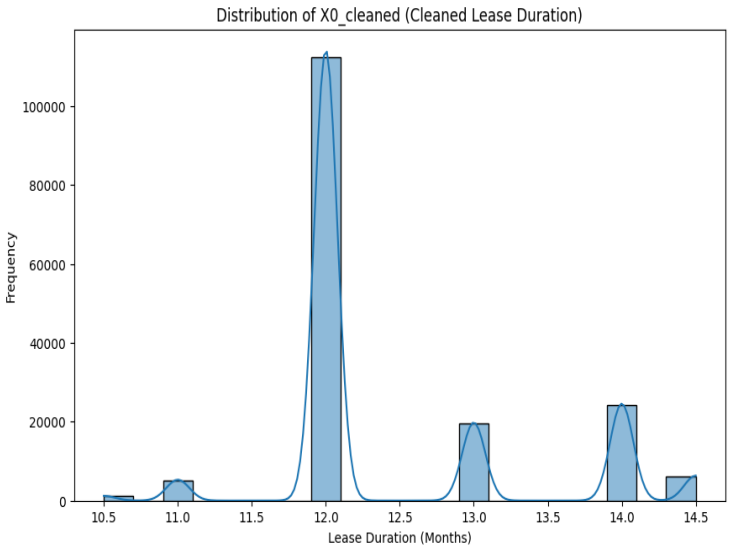
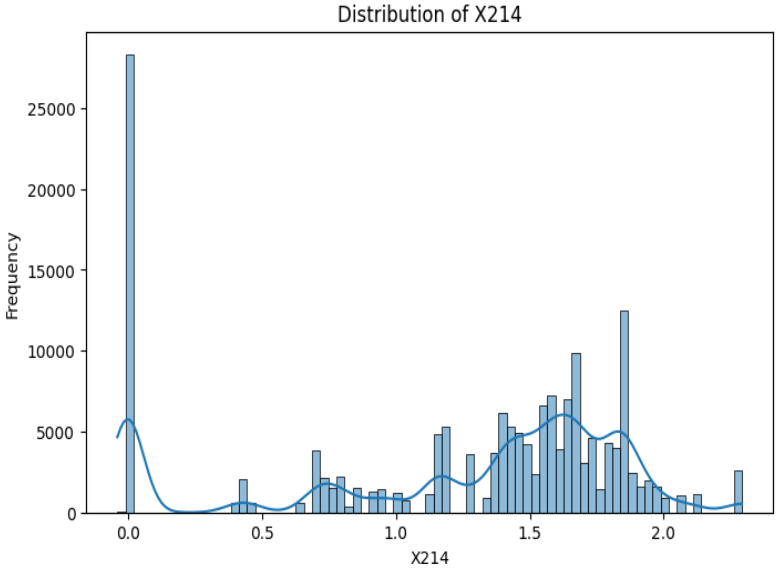
Distribution of Churn Flag in the Dataset:

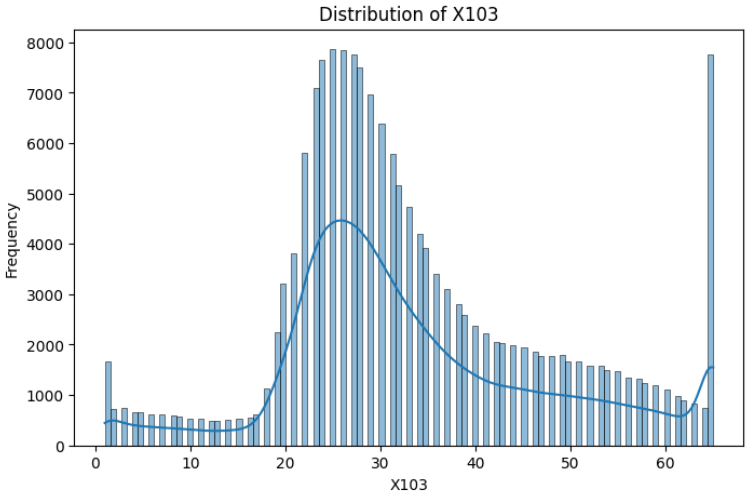
Before diving into complex analysis, it's crucial to understand our target: customer churn. This chart clearly shows the distribution of our customer base – those who remained with us versus those who churned. We can see that the number of non-churned customers significantly outweighs the churned customers. This highlights a common challenge in churn prediction: class imbalance. While our model needs to be accurate overall, it's particularly important that it's effective at identifying the smaller group of customers who are likely to churn, so we can intervene.

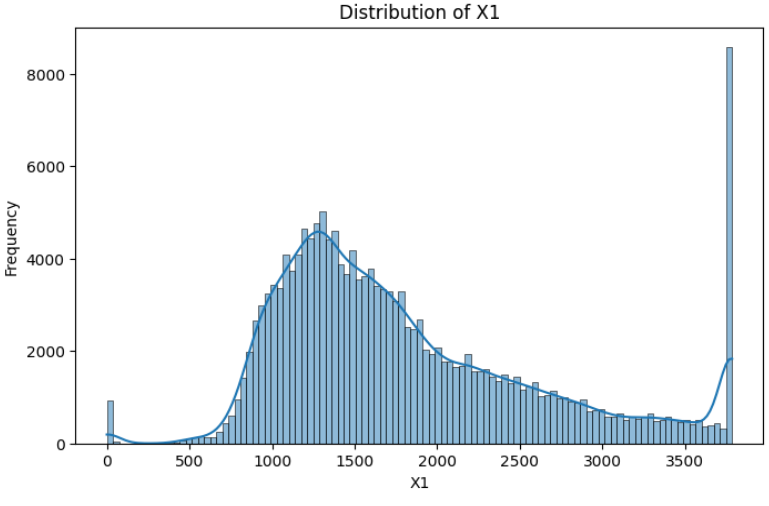


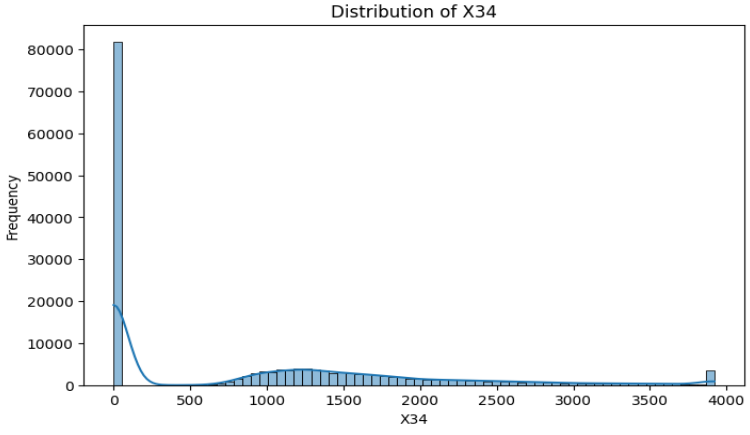
Distribution of Selected Numerical Features:

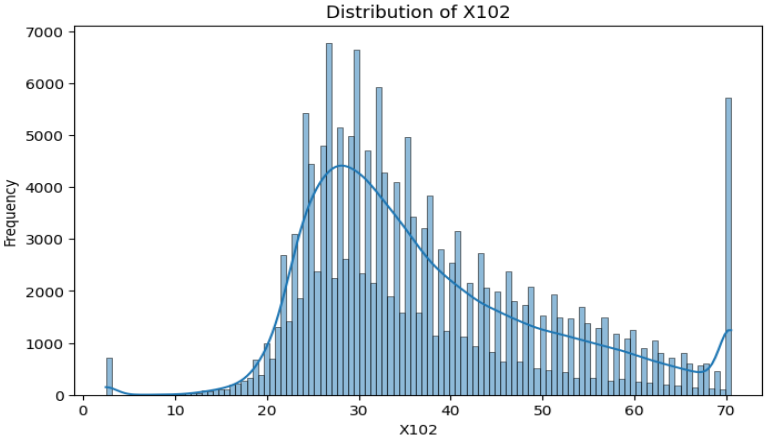




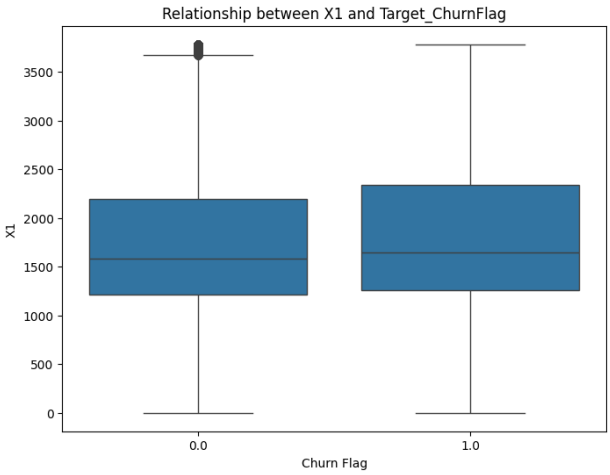


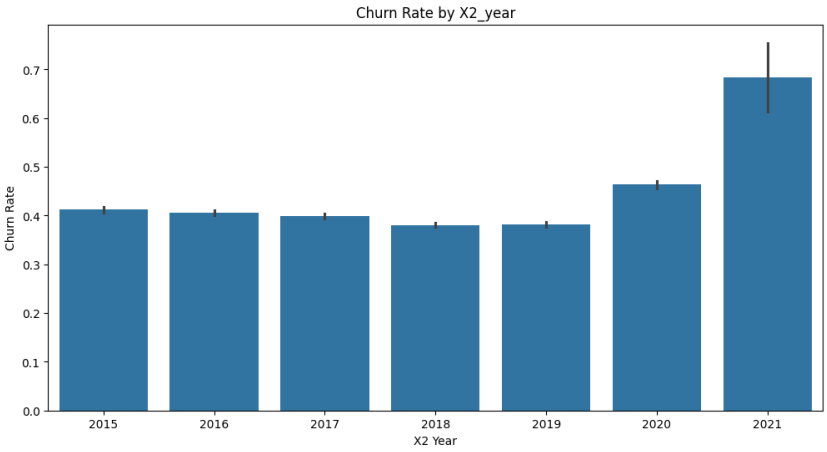




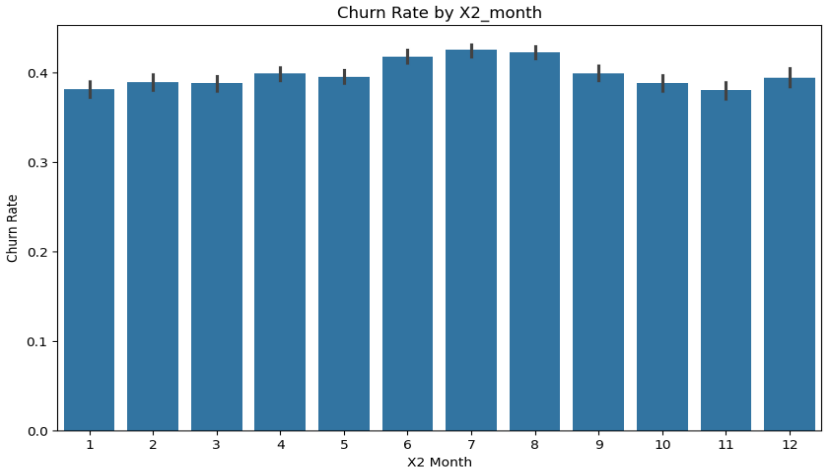


Let's look at how a key feature, `X1`, relates to churn. This box plot compares the distribution of `X1` values for customers who churned (1) and those who didn't (0). We can observe that customers who churned tend to have lower values of X1 compared to those who didn't. This suggests that `X1` is a potentially important indicator of churn risk. Understanding what `X1` represents in a business context could provide actionable

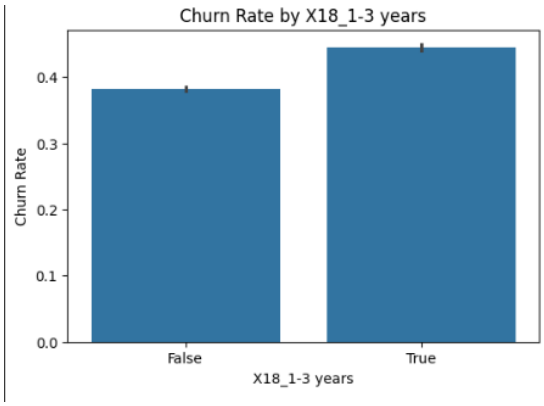


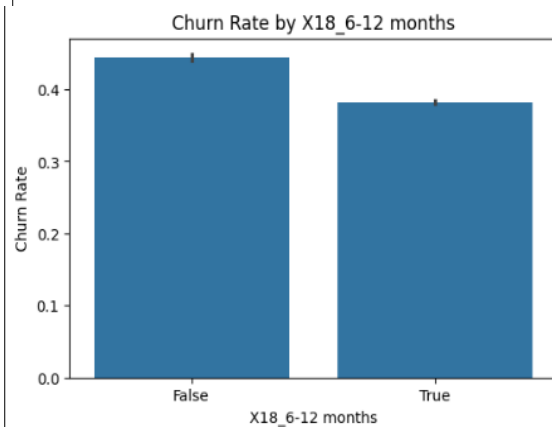


- We can observe a general increase in churn rate over the year



- In June to August months in the year churning rate is highest.



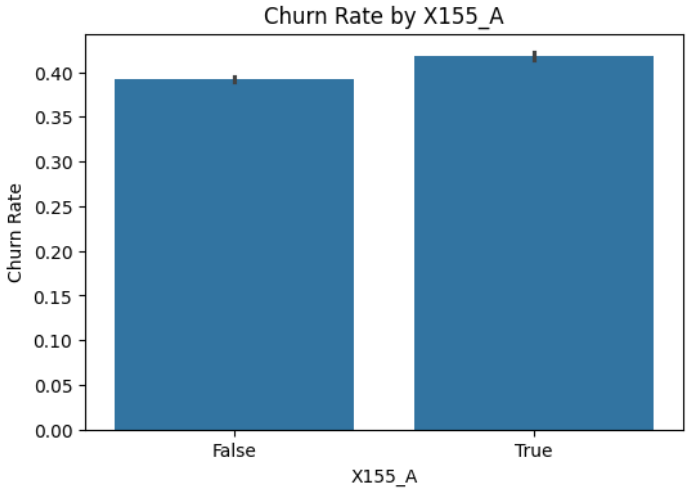


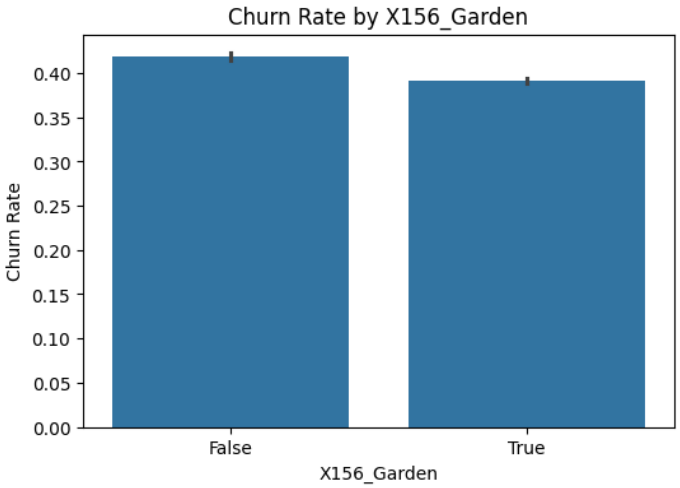
- Customers with 1–3 years tenure churn more, showing lesser engagement over time.
- Churn is lowest among customers with 6–12 months tenure.

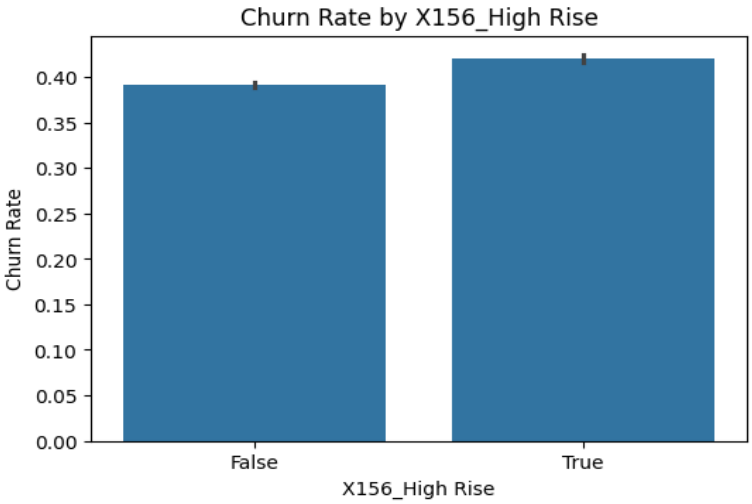
By looking at the bar charts we can conclude that:

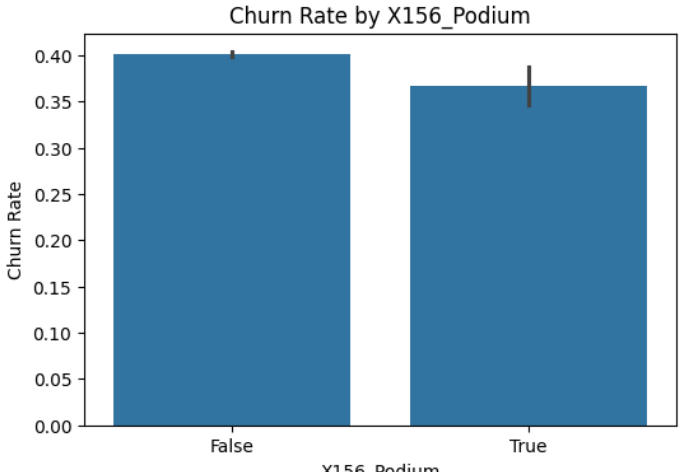
- X151_CA,X151_CO, X151_FL , X151_ NJ, , X151_ NY, , X151_ OR , X151_ TX , X151_ WA have a noticeably higher churn rate compared to X151_DC , X151_MA, X151_ MD , X151_PA , X151_TN , X151_VA.

- X155_A have a noticeably higher churn rate compared to X155_B

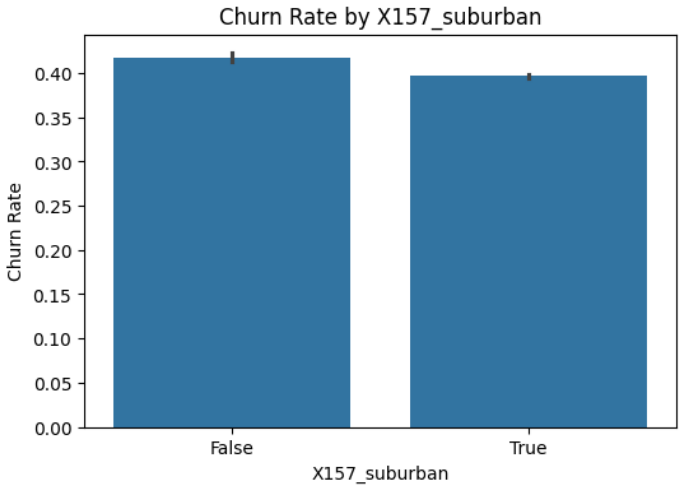


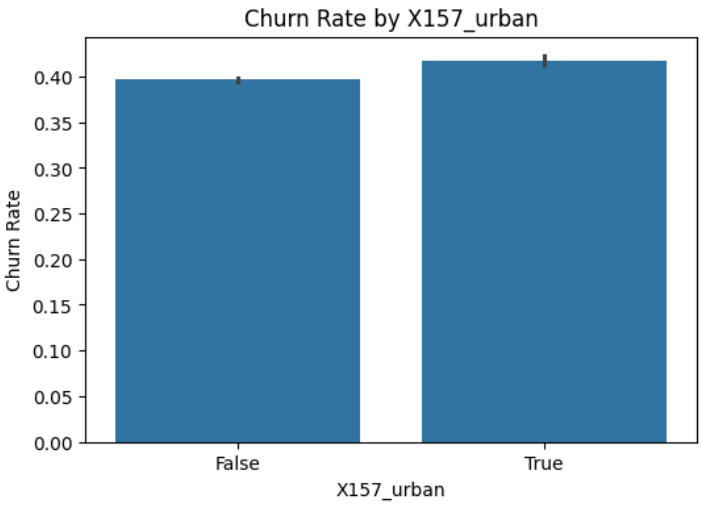




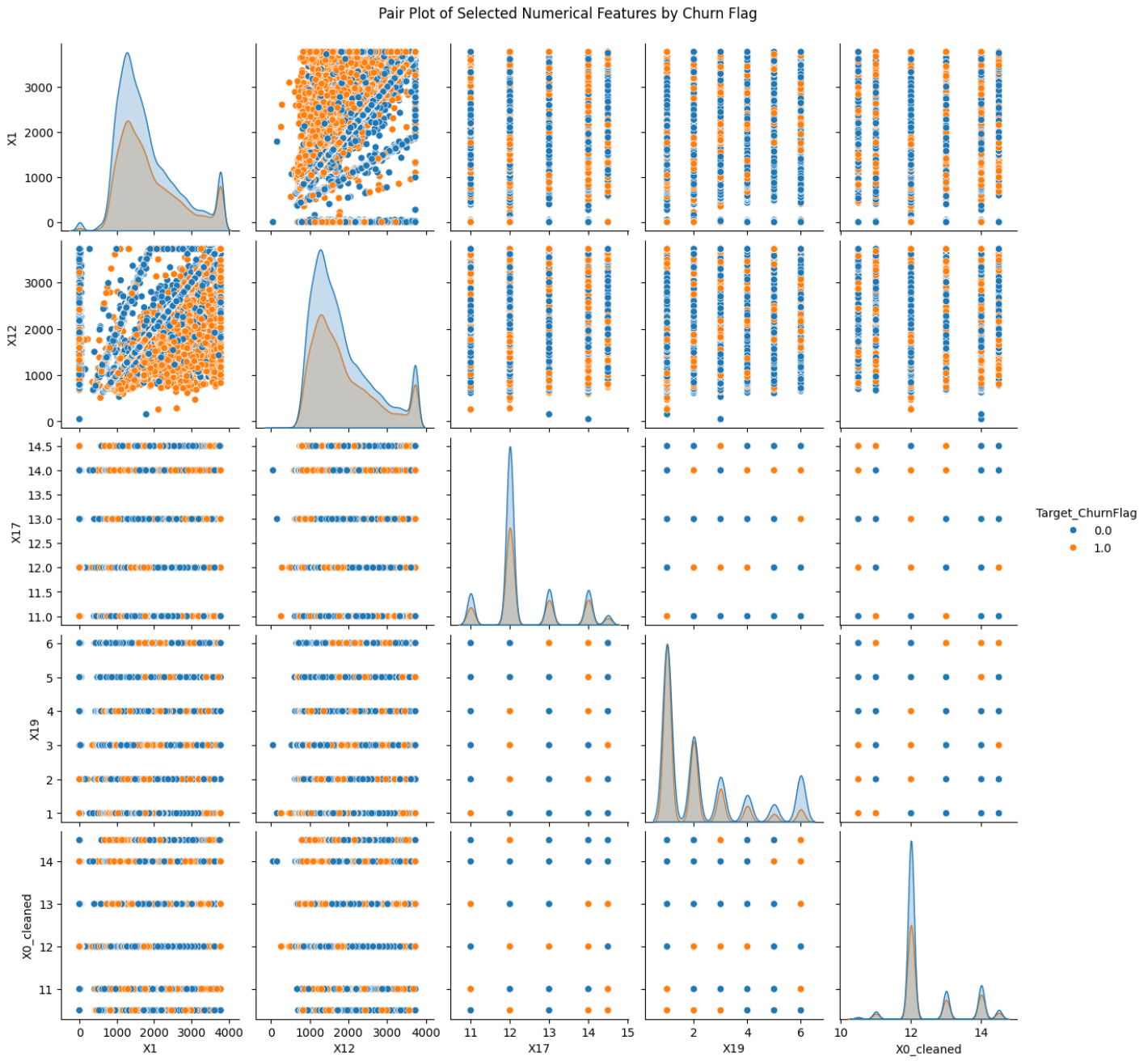


- X156_High Rise have a noticeably higher churn rate compared to X156_Garden and X156_Podium.

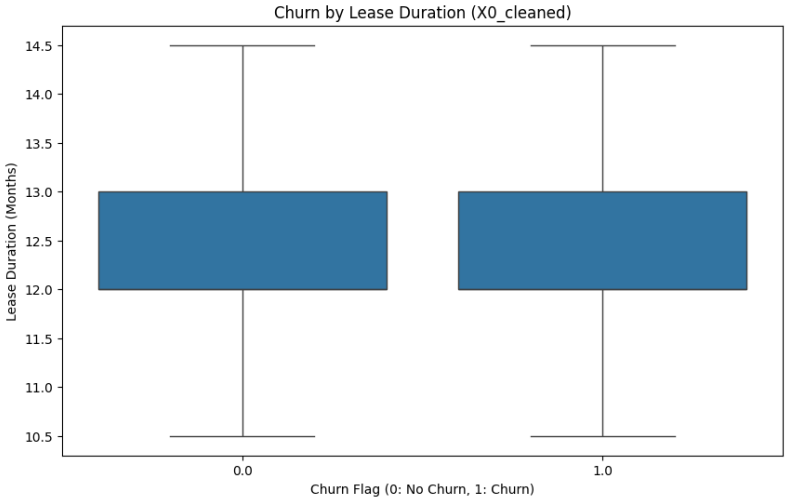




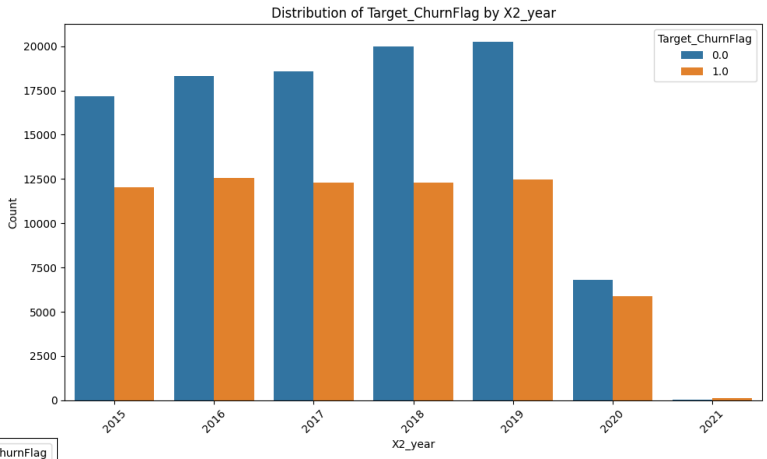
- X157_urban have a noticeably higher churn rate compared to X157_suburban.



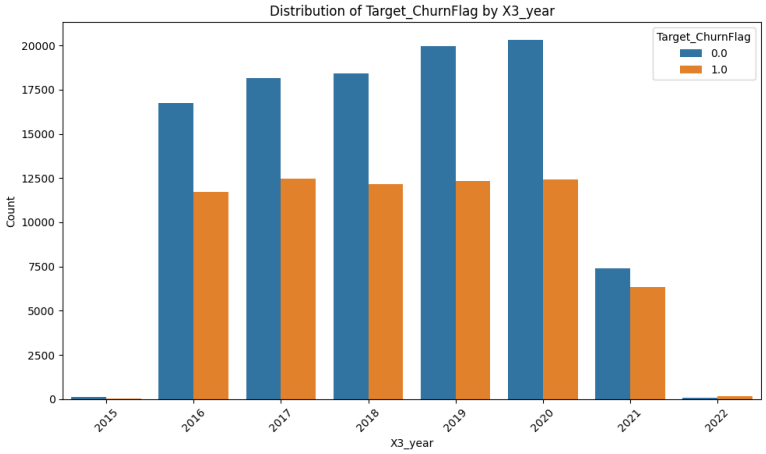
Moving beyond individual features, understanding how key numerical variables interact can provide deeper insights. This pair plot shows the relationships between several important numerical features (X1, X12, X17, X19, and the cleaned lease duration X0_cleaned). The diagonal plots show the distribution of each feature, allowing us to see how churned and non-churned customers are distributed along their ranges. The scatter plots in the off-diagonal provide a view of bivariate relationships. For example, we see a strong linear relationship between X1 and X12 .In some plots you can observe strong separation between churned and non-churned customers in the plots involving X0_cleaned'. This visualization helps us appreciate the multi-dimensional nature of the data and how combinations of features might be predictive of churn



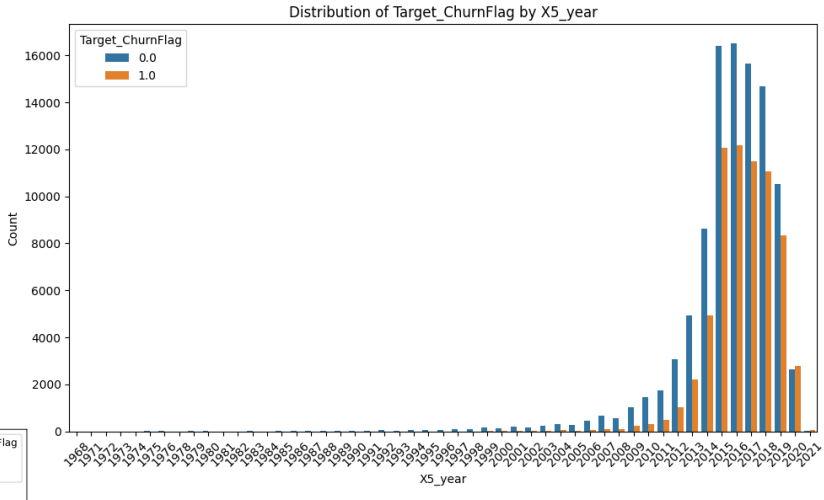
This histogram shows us the frequency of different lease durations in our dataset. We can observe the most common lease durations are around 12 and 14 months, with fewer customers having very short or very long leases. Understanding this distribution is important context when we look at how lease duration impacts churn.



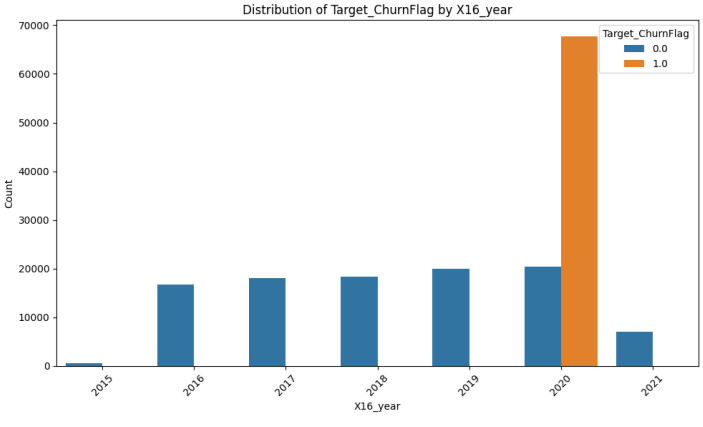
X2_year bar chart shows that the numbers of true(1) churn_flag are lesser than the false(0) churn_ flag except for the 2021 year.



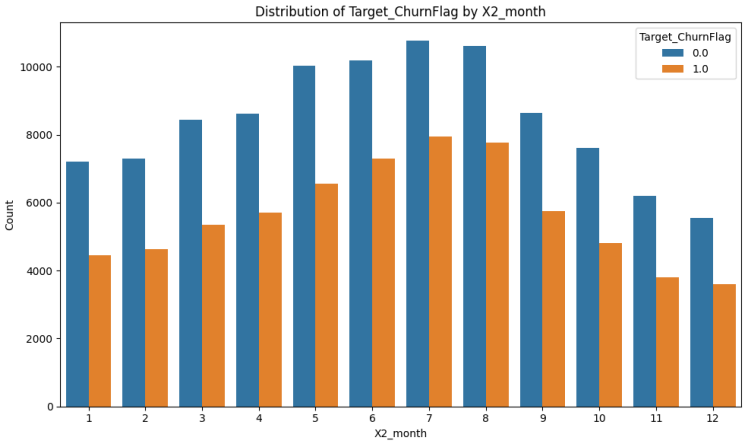
X3_year bar chart shows that the numbers of true churn_flag are lesser than the false churn_ flag except for the 2021 year.

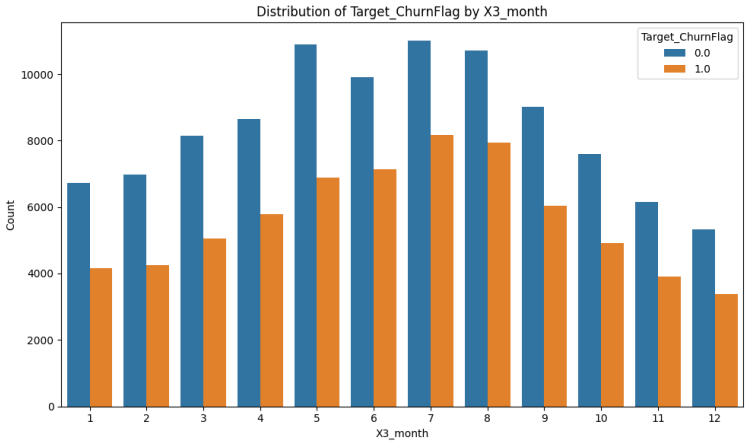


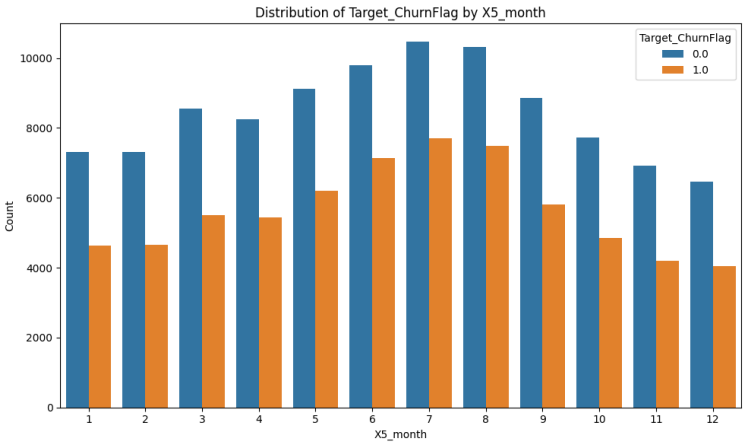
X5_year bar chart shows that the numbers of true churn_flag are lesser than the false churn_ flag except for the 2020 and 2021 year.



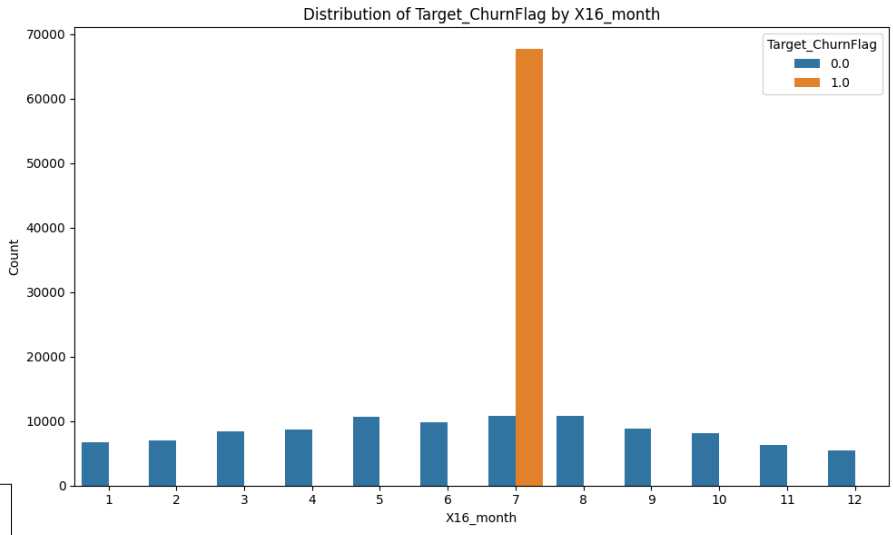
- X16_year bar chart shows that the numbers of true churn_flag are lesser than the false churn_ flag except for the 2020 year.
- This chart is not like other typical bar chart of years because here only year 2020 has false(0) churn_flag value.



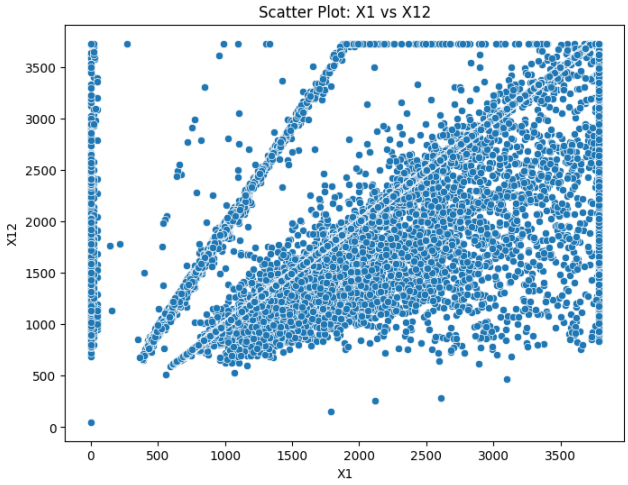




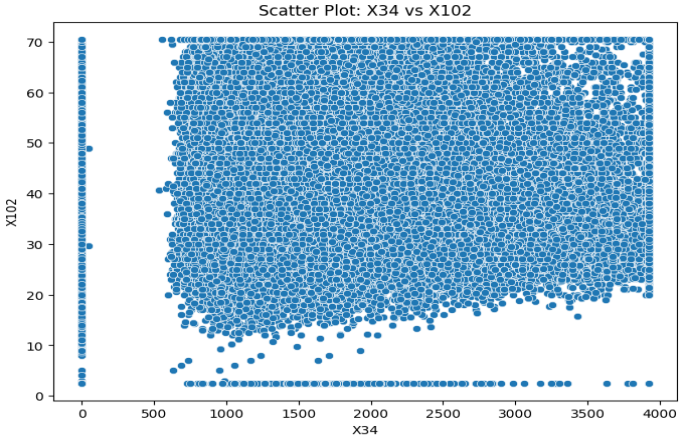
All three bar plots (X2_month,X3_month,X5_month)are showing almost similar trends with corresponding months.



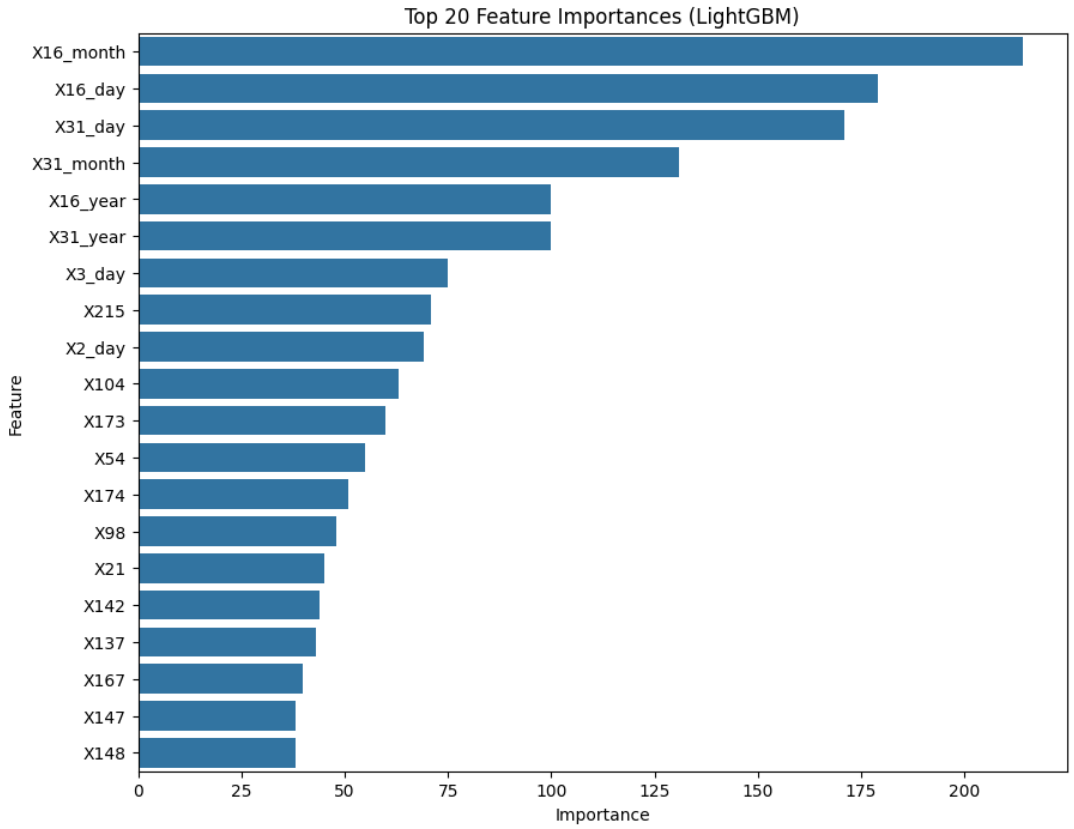
- The bar plot of X16_month shows irregular pattern.
- It has a high numbers of false churn_flag for month 7 and false churn_flag numbers of 0 in other months.



The plot of X1 vs X12 clearly shows a strong positive linear relationship



The plot of X34 vs X102 clearly shows a strong positive co-relationship



This are the columns with high correlation with the churn_flag arranged in decreasing order.