PROBLEM STATEMENT

- Developing a machine learning model that can be used to determine the probability of a customer to leave(churn) a service or cancel a subscription to a service and looking at what factors are mostly attached to customers churning.

BUSINESS UNDERSTANDING

- Customer attrition/churn is the loss of clients or customers. Telephone service companies, internet service providers, Tv companies ,insurance firms and such are among the ones who are majorly concerned with customer attrition.
- The aim of this project is to retain more customers by looking into what factors are mostly attached to customers unsubscribing to a service ,in this case it is telecommunication service.
- The telecommunication company aims to reduce the amount of money that is lost because of customers who don't stick around very long.

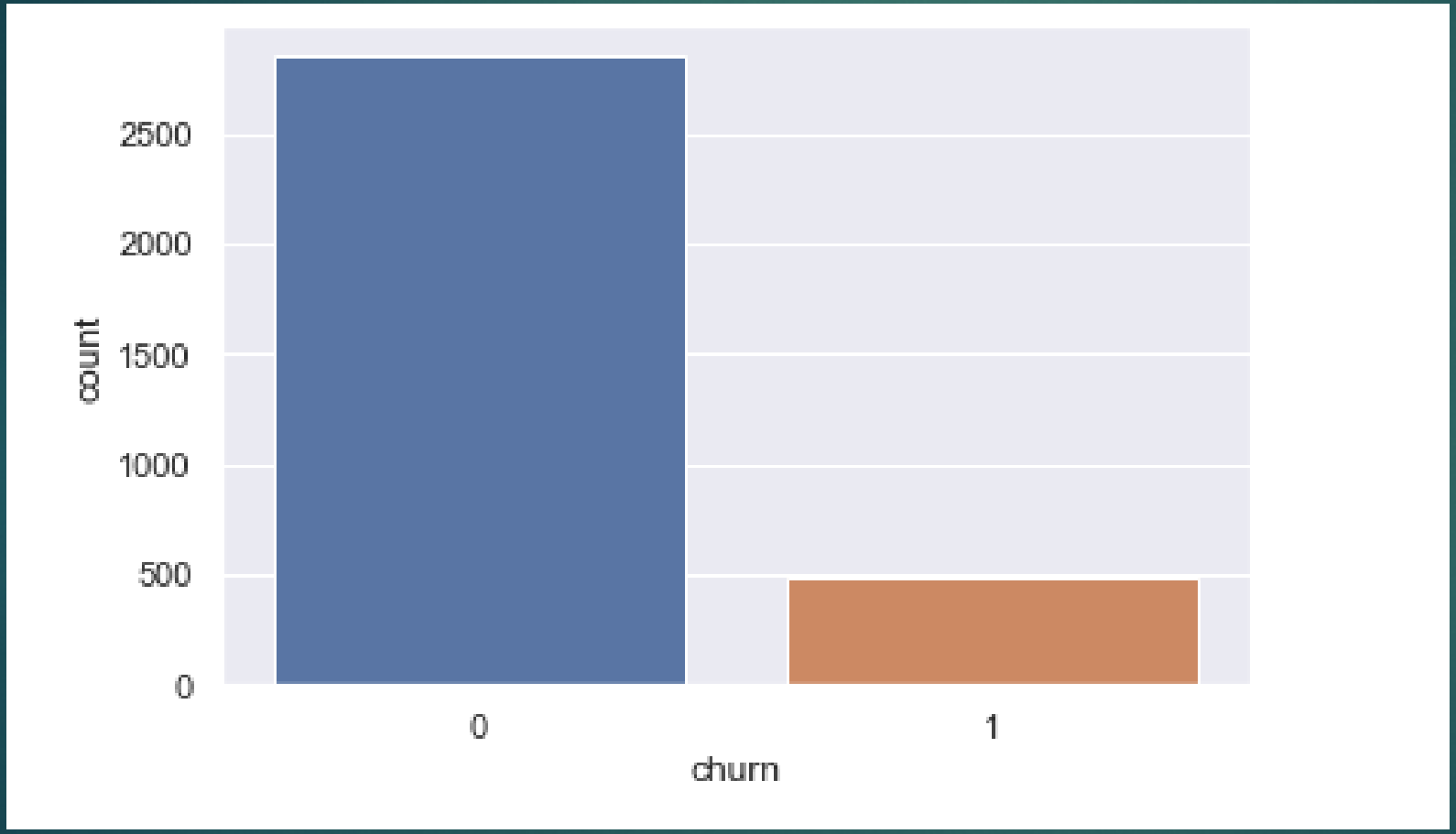
DATA UNDERSTANDING

- In this project, we will predict whether a customer will pull out of the telecommunication service or not based on many factors The factors are:

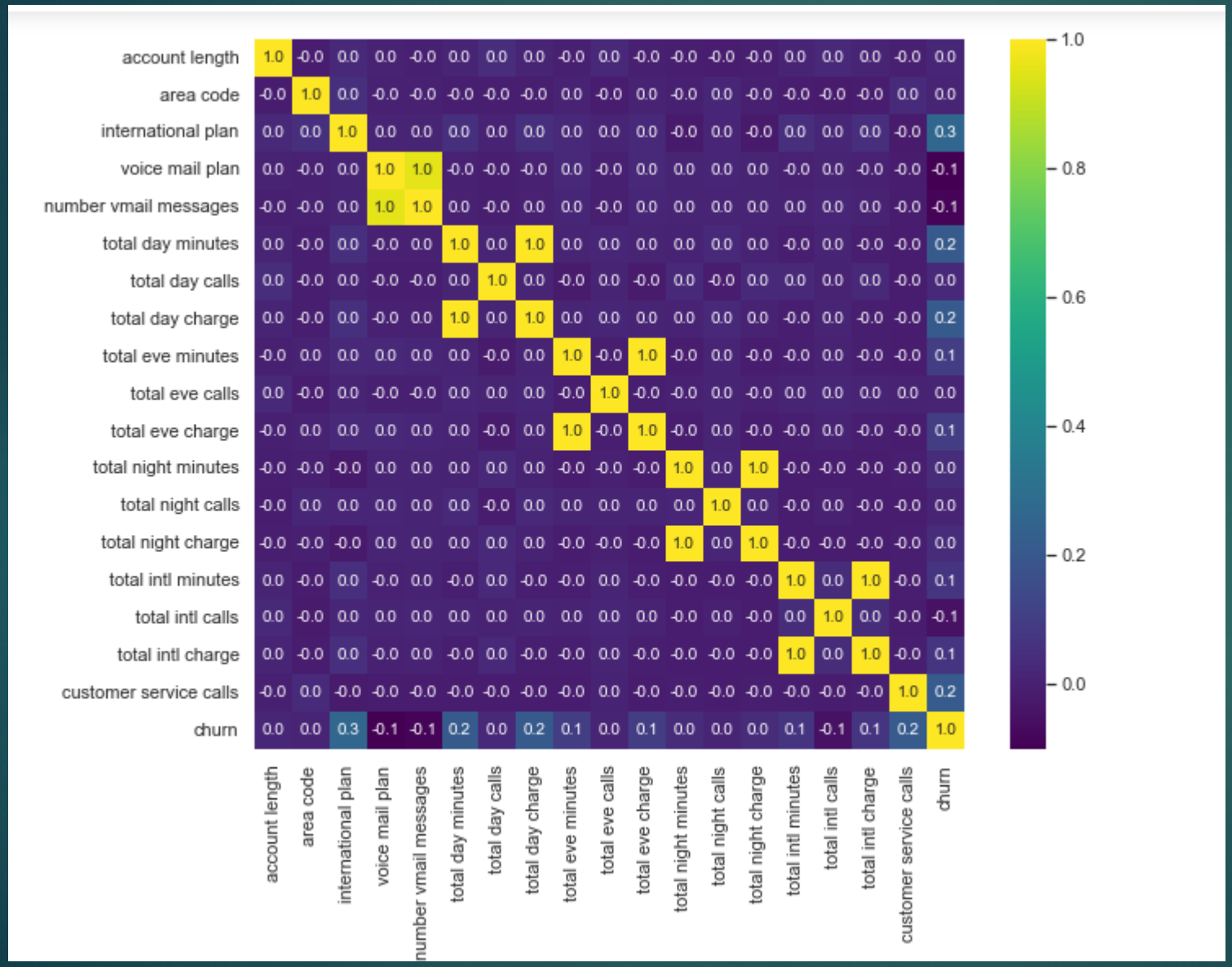
- ✓ State
- ✓ Area code
- ✓ International plan
- ✓ Number vmail messages
- ✓ Total day calls
- ✓ Total eve calls
- ✓ Total eve charge
- ✓ Total night calls
- ✓ Total intl minutes
- ✓ Total intl charge
- ✓ Churn(target variable)

Account length
Phone number
Voice mail plan
Total day minutes
Total day charge
Total eve minutes
Total night minutes
Total night charge
Total intl calls
Customer service calls

Count of churn(target variable)

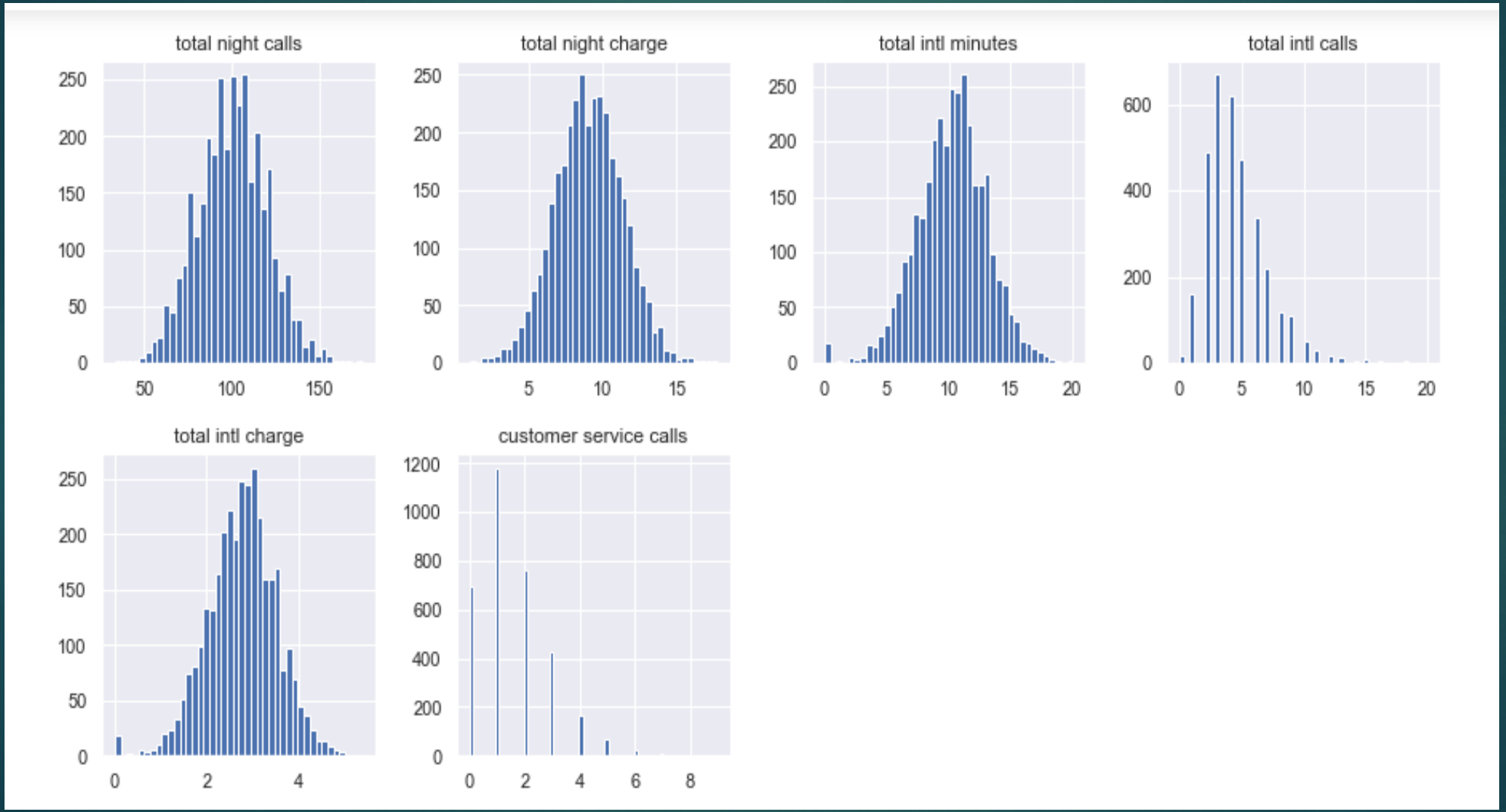


Heat map to show correlation of features

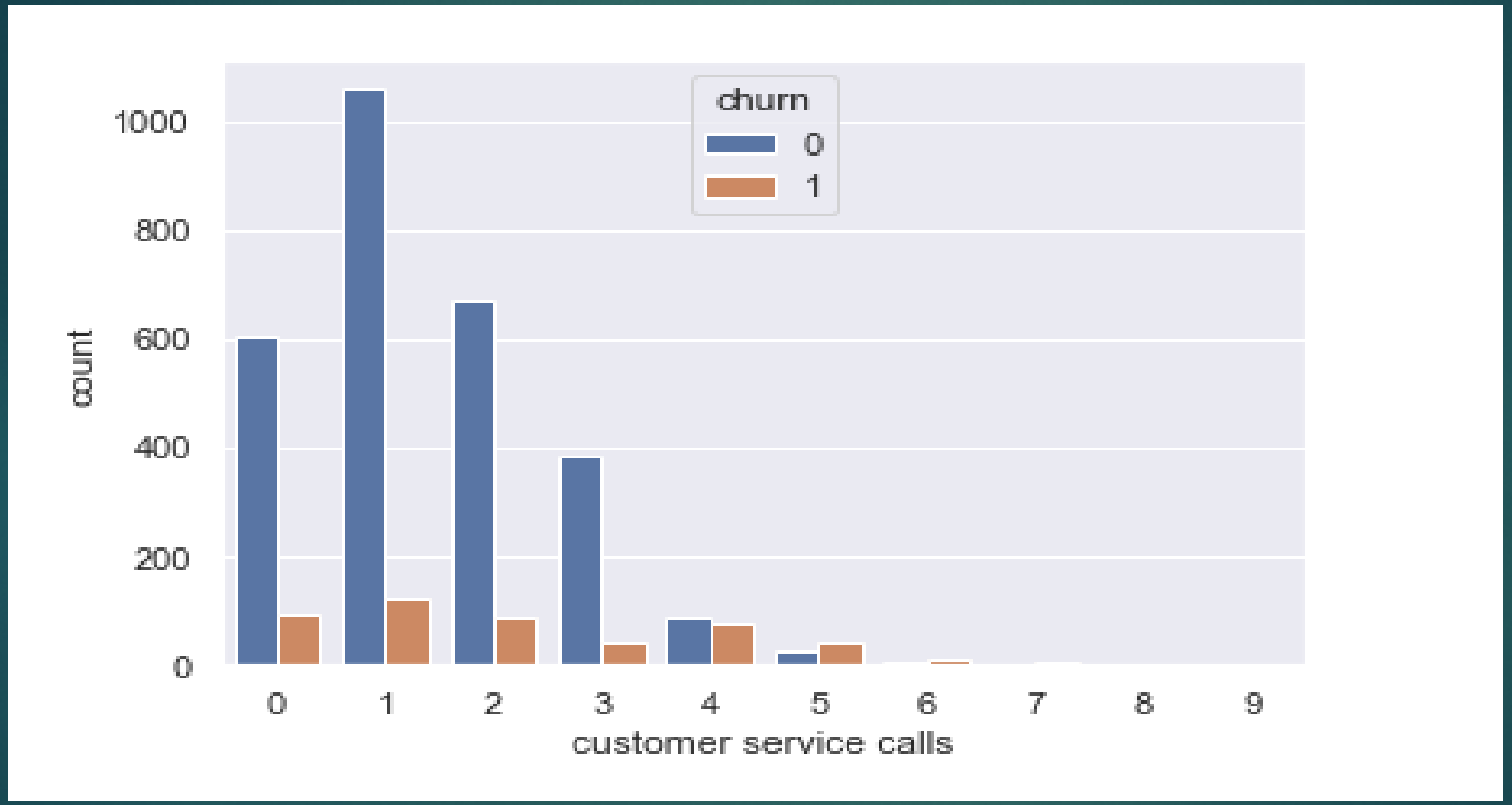


Visualization of the numerical features

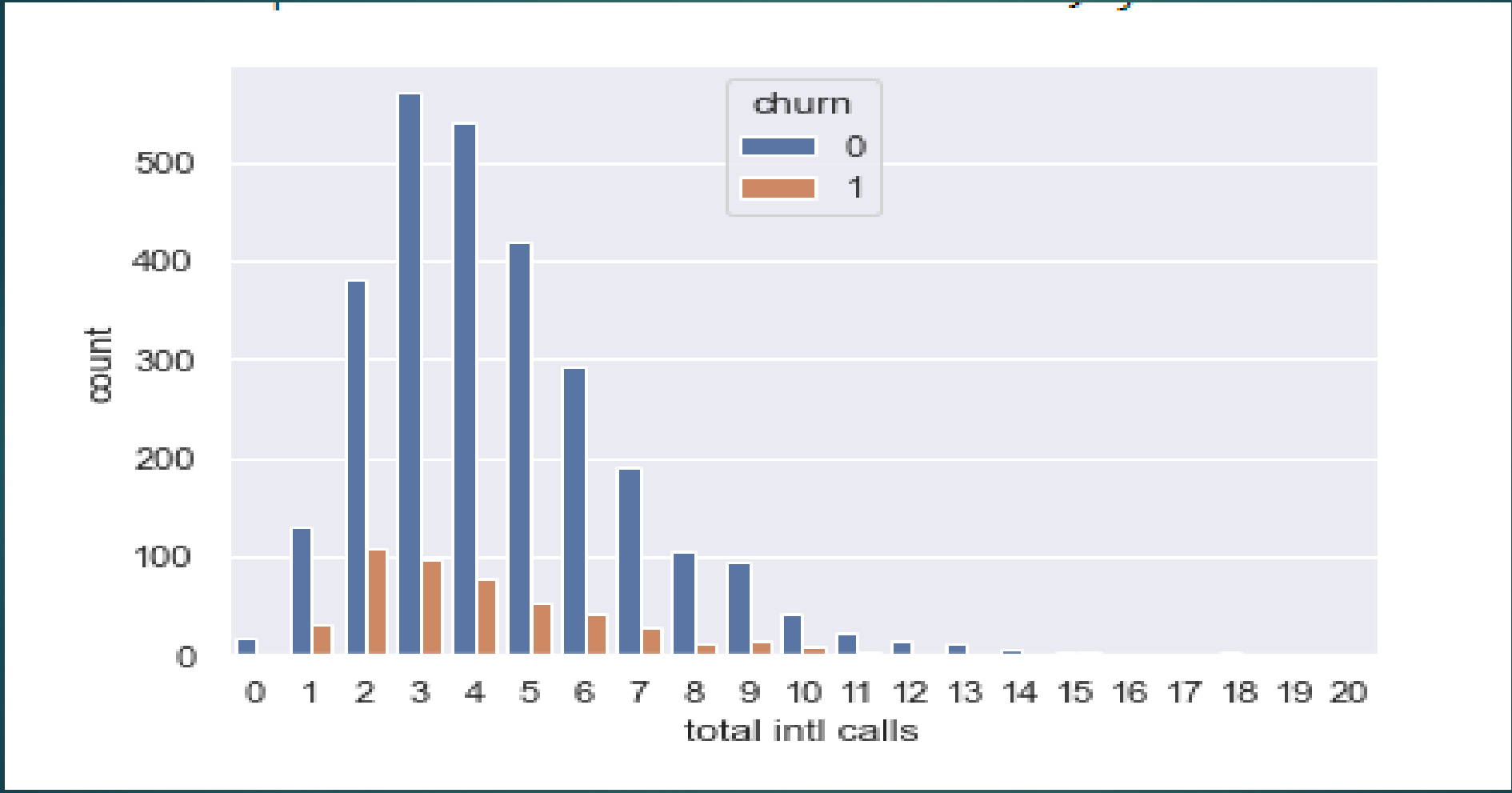
Visualization of the numerical features (2)



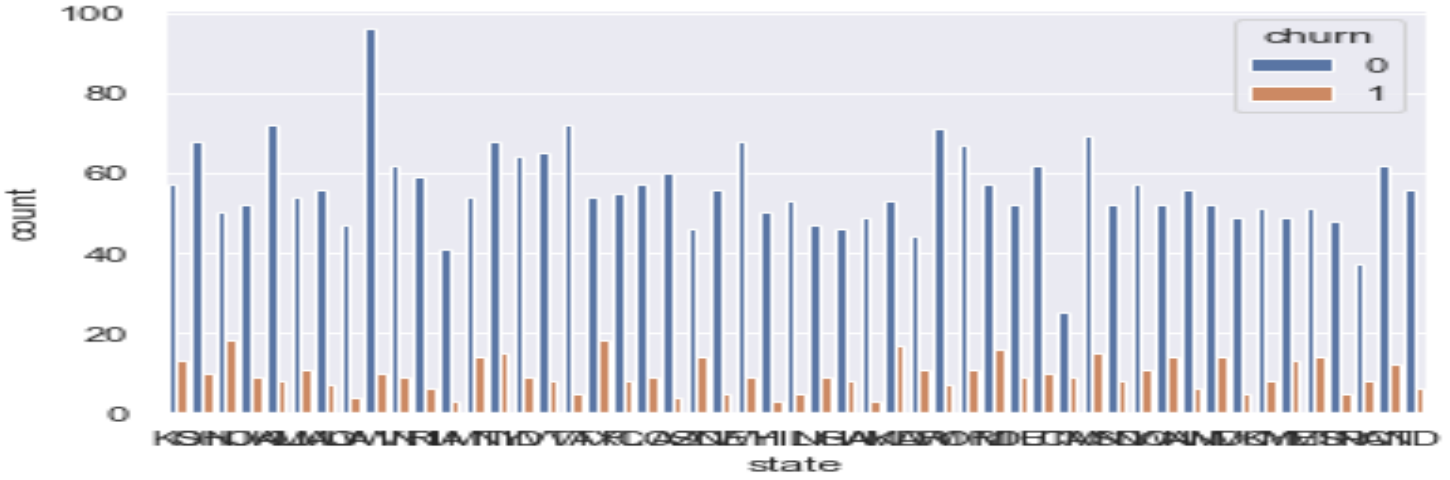
Analysis of customer service calls against churn

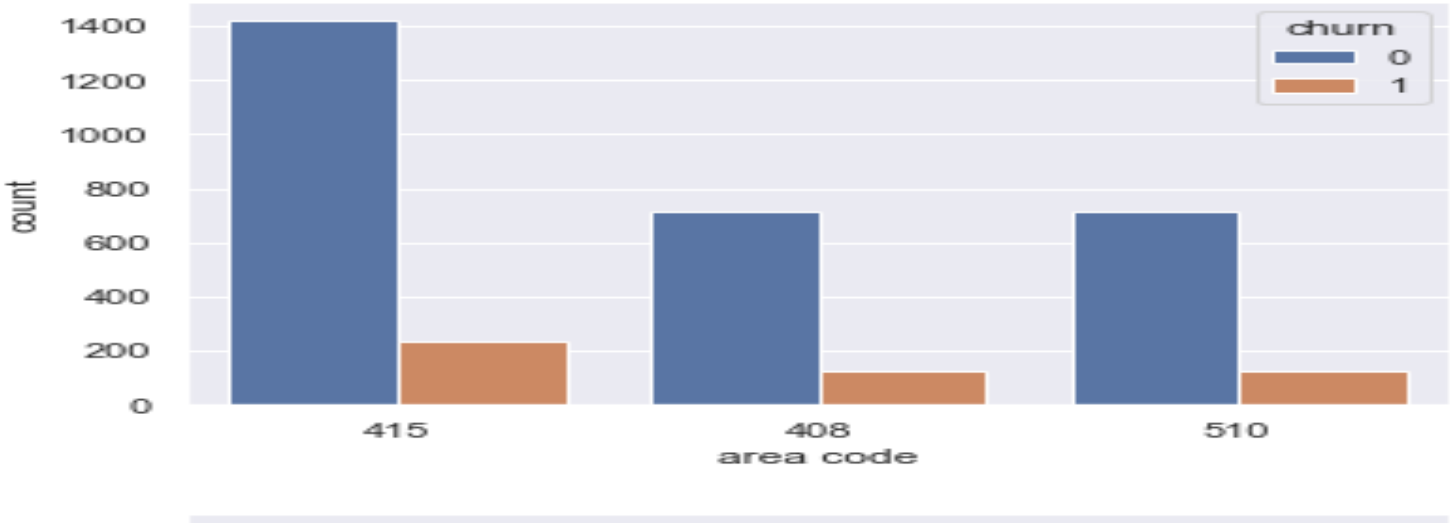


Analysis of total intl calls against churn

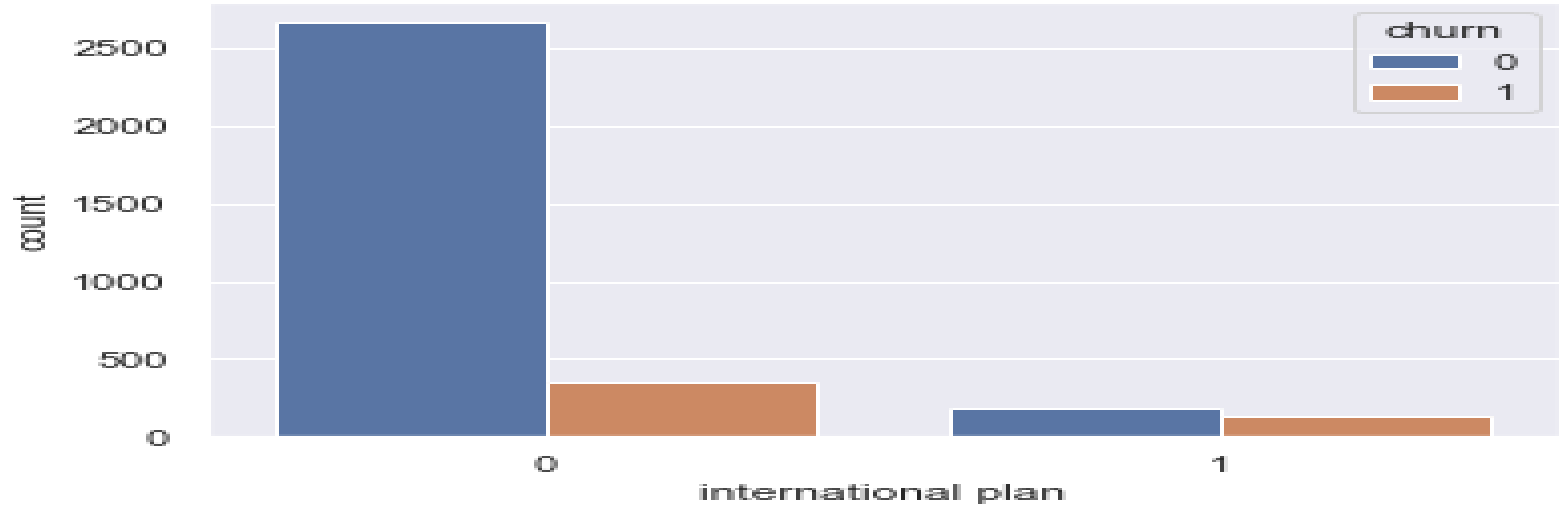


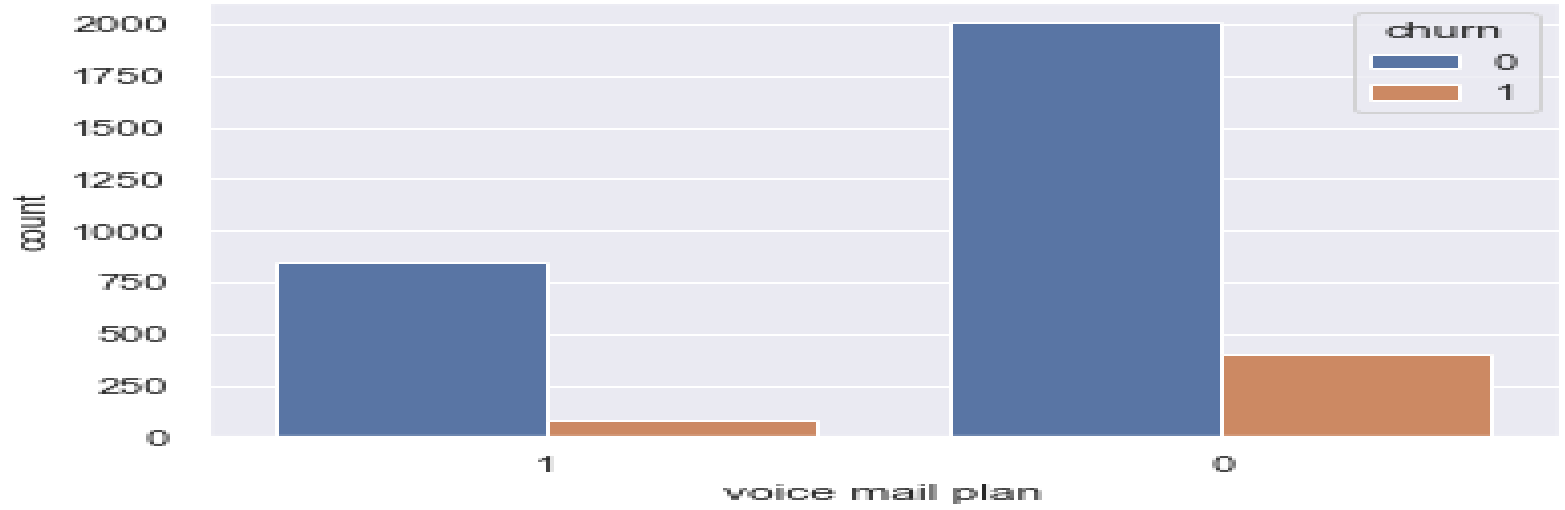
Analysis of categorical variables against churn (2)



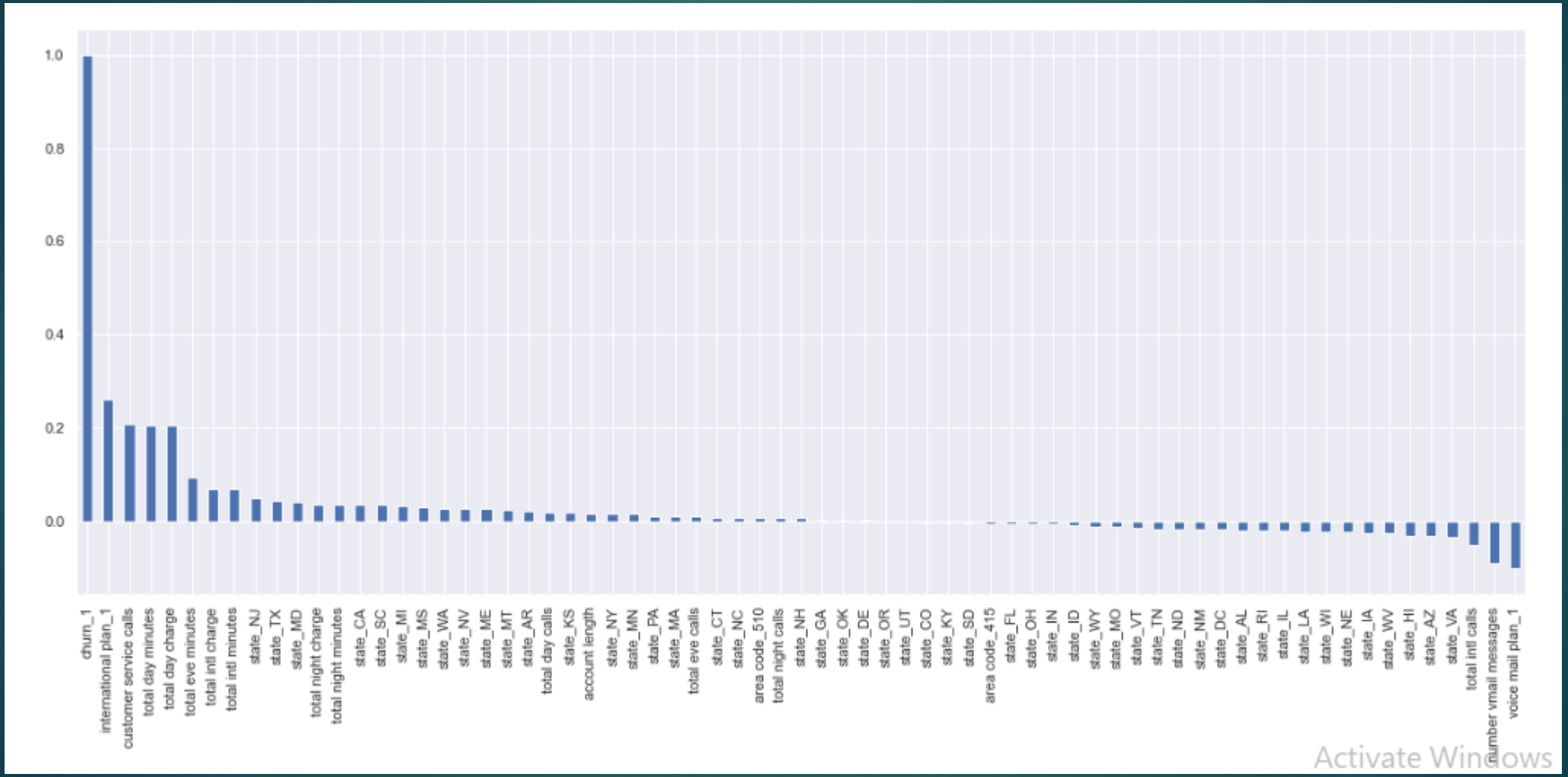


Analysis of categorical variables against churn





Summary of the highest and lowest churn in the dataset



Conclusion

- Churn is high for those with international plans, those who make several customer service calls especially more than 4 the churn rate seems to be pretty high.

- Clients who make day calls also have a high churn rate than clients who make night calls this could probably be the charge rate for day calls being higher.

- Churn is low for clients with a voice mail plan, with more number of voice mail messages, and those clients with a high number of international calls.

- From the plots above we can say that clients from area code 415 seem to be the ones who churn the most compared to area code 408 and 510 and having a customer service Centre could maybe help understand as to why clients in that area code seem to churn the most.

- Clients without an international plan also seem to churn more than clients with an international plan.
- Clients without a voice mail plan also churn more than clients with a voice mail plan

- Precision was used as our error metric for our algorithm, which is true positives / (false positives + true positives),which ensured that we minimize how much money we lose with false positives. We'd rather minimize our potential losses than maximize our potential gains. For this reason ,SVM is the preferred model because it maximizes on the true positives.