Nanyang Technological University

Nanyang Business School

BC2407 Analytics II: Advanced Predictive Techniques
AY 2021/22, Semester 2

Group Project: Seminar Group 6, Team 7

Proof-of-Concept for Customer Retention for AT&T

AT&T Customer Churn reduction through service aggregation and Increasing Revenue by predicting customer importance through Customer Lifetime Value

Members:

Name	**Matriculation Number**
Clifford Choo Sing Link	U2011513B
David Alexander Yosal	U2010520B
Lim Qing Rui	U2010816G
Low Yong Zhuo	U1911161B
Shao Yakun	U1920578C

Prepared for: Professor Neumann Chew C. H.

Table of Contents

Executive Summary

High churn rate is one of the top recurring issues faced by many subscription-based and contract-based firms. Especially so in the highly competitive and dynamic telecommunications industry, many companies are greatly concerned with customer retention and face problems attracting existing customers to come back. AT&T, one of the top telecommunications firms in the United States, is facing record-level customer attrition despite putting in place aggressive promotions. Most notably, their premium streaming services have lost over 4 million subscribers in 2019 (Szalai G., 2020) and its subscribers are still falling at a distressing rate of 14.6% annually (Munson B, 2021).

Hence, this report will focus on the use of Machine Learning algorithms to better understand the underlying root causes behind customer churn in AT&T, generate insights on improve customer retention and discover any trends to better explain AT&T's customer purchasing behaviour. With these insights obtained, appropriate recommendations will be proposed to minimise churn in AT&T.

The datasets used are procured from Kaggle, which consist of customer and telecommunications service-related information that can help us better understand customer churn. Before conducting any analysis, we combined all the datasets and cleaned them to ensure no missing values, no duplicates, and retained only relevant variables. Data exploration and visualisation, in conjunction with secondary research, was conducted on the factors to provide a preliminary analysis.

Using the combined and cleaned dataset, we explored the different variables and visualized them using multiple libraries. Several of the visualizations that we produced are correlation heatmaps, geospatial analysis, distribution and density plots, etc. After ensuring that there were no outliers, imbalanced data and skewed distributions, we plucked in our data into multiple types of supervised models to predict Churn and CLTV and an unsupervised model to derive associative services that AT&T can provide to customers.

To find the best model to predict Churn and CLTV, we ran many different types of models and tried several ways to improve their accuracy. We did hyperparameter optimisation, data normalisation, data balancing and feature selection. Then, we compared the scoring metrics of the different model variations.

For Churn, we found out that XGBoost on unprocessed data yields the best results for all the scoring metrics, with an accuracy of 93.1 percent. From the model, we learned that perhaps the biggest customer reason to leave AT&T is due to their customer service not being able to solve their problems or low satisfactory response from service requests. Also, we found some inverse correlation between customer's age and retention probability and with price-sensitivity in consumer pricing.

Using random forest, we can identify which variables had the highest impact on CLTV. We found some obvious variables such as tenure and charges that customers were currently paying to be highly significant. However, age displayed interesting correlative properties with other variables in determination of CLTV.

Lastly, after analysing our association rules output, we found a couple of services that often showed up together, such as PhoneService and InternetService. Also, we tried to observe the difference in rules between those who stayed and those who did not but were unable to find significant differences. In addition, we will derive insights leading from XGBoost and Random Forest which manifested as outliers or showed as model limitations.

These insights can be implemented inside AT&T's internal and external strategies, such as staff training and marketing campaigns. Internally, having a more trained fleet of customer service operators would greatly help customer retention, as total customer service request is the most significant variable affecting churn. Externally, we recommend AT&T to introduce some service bundlings and setting some options as the default. We believe that these will improve sales and provide a better customer experience with AT&T. Also, appealing to the younger audience would be helpful in improving customer retention.

1. Overview

With an increase in emergence of subscription-based businesses in the telecommunications industry, greater pressure has been placed on existing businesses looking to bolster customer retention and increase acquisition efforts. As high acquisition costs are often involved in attracting new customers which may in turn lower profitability, firms tend to look at improving their retention efforts instead. On the flip side, a high churn rate lowers revenue as higher profitability customers, such as those with higher Customer Lifetime Value (CLTV) may be lost, stifling business growth and expansion. Therefore, prompt identification of factors leading to customer churn and the uncovering is required for Software as a Service Companies (SaaS) to strategize and increase product suitability for customers.

1.1. Company Overview

In recent years, companies are looking to retain existing customers on top of acquiring newer customers at a cheaper cost to help them stay competitive in this ever-changing climate. AT&T, the third most popular Telecommunication Service Provider (Telco) in the United States of America (USA) with a market share of about 27% is no stranger to this *(Dano M., 2020)*. The Telco has been building up its 5G infrastructure since 2017 and is the first company to launch its 5G network in 12 major cities across the USA in 2018. With the introduction of 5G, this provides tailwinds for AT&T to adapt and gain greater market share of the serviceable market through providing internet and mobile-based services to grow its business and potentially improve profit margins.

1.2. Business Opportunity

One of the most prevalent problems affecting subscription-based models, and a significant factor affecting revenue levels in the Telco industry, is customer churn. Customer churn can be defined as the number of users who terminate their subscriptions, choose not to renew their contracts, or cease consumption of products or services offered by a company. Churn can be categorised as (1) involuntary – an unintentional opt out by customers such as credit card payment not being captured, or (2) voluntary – when customers are dissatisfied with the product or service provided and discontinued their subscription *(Unscrambl., 2021)*.

Today, churn is a hot topic for Telcos in the USA due to intense competition in a highly saturated market. With 5G being the talk of the town of late, more companies are striving to provide the best 5G experience and service at the most affordable price. Customers who are more price sensitive will easily switch to competitors that offer a better package or provide a higher level of service. In 2020, the average churn rate in the Telco industry in USA is 21% *(Statista, 2022)*. Thus, AT&T should aim to achieve a churn rate lower than the industry average. Based on preliminary findings conducted, a primary cause of churn in the Telco market ties closely to customer service issues (e.g., waiting too long, incompetent agents) *(Techsee, 2019)*. On top of service-related problems, there may be other causes surrounding the high churn rates for case of AT&T.

Therefore, this report will examine the reasons behind the cause of churn for AT&T, specifically for its internet, mobile, and streaming services sector, the ways to resolve these issues, and how it can provide improved customer experience and a better mix of products/services to reduce churn and raise profitability in this cut-throat industry.

1.3. Reasons for Proposal

1.3.1 Shift in focus towards a Customer-centric model and Cost of churn

In a consumer research report, 76% of customers expect businesses to understand their needs, and companies concurred that being customer-centric has helped them achieve an additional 60% profitability as compared to those who are not *(Fontella C., n.d.)*. Hence, companies have begun to pivot towards differentiating themselves from their counterparts in this area. In addition, by understanding the needs of customers through the use of data analytics and market segmentation of customers can mitigate an annual loss amounting to US$1.6 trillion resulting from dissatisfactory customer service *(Fontella C., n.d.)*.

1.3.2 Rise in investors' concern about Profitability

With over 80% of AT&T's Operating Income (USD$18.64 billion) and 73% of AT&T Revenue (USD$124.41 billion) derived from their telecommunication service in FY2021, AT&T is struggling to maintain growth in this area despite the increasing penetration for 5G in the USA market. A sluggish growth in the mobile phone subscription business and services that AT&T can provide to their customers would indirectly lead to a decreasing forecasted profitability in the upcoming years and lower shareholder value *(Moritz S., 2022)*. Hence, AT&T should optimise their services provided to customers to overcome normalisation of earnings and possible erosion of tailwinds from telecommunication services rollouts.

1.3.3. Customer Churn and Services Mix Purchased as business performance indicators

The average annual customer churn rate in the telecommunication industry is 21% where any churn rate greater than the average is indicative of the company's product or service failing to meet customers' expectations or goals. Hence, subscription-based businesses such as AT&T needs to take note of internal factors (telecommunication services that they provide) or external factors (customer age, partners, dependents) that can influence attrition in customer base. As AT&T is a telecommunication provider, they may provide different services, where some may face greater demand than others *(Smith O., n.d.)*. As such, these services may be bundled together for greater attractiveness to the target audiences of AT&T. Through identification of such services, AT&T can choose to discontinue them or provide as a free service to entice customers to employ their usage. These suggestions will be provided with the analytics that we obtained from our Machine Learning Models.

2. Project Objective and Feasibility

2.1. Analytical Problem

The report aims to evaluate current predictive Machine Learning models to identify trends and key contributing factors leading to variances in Customer Churn rates where Internet, Phone services and Streaming services may be provided. As stated in the proposal, the key objective is to enhance and value-add to current services provided by AT&T based on upcoming trends in 5G and create new product varieties for the localised (USA) telecommunication service outlooks, through a more predictable and accurate data analysis model (by analysing the impact of machine learning) to expand the economic and business product variety of AT&T. Keeping this in mind, the predictive model accuracy chosen should reflect an appropriate reduction of cost, Statistical Significance and Confidence Interval. The chosen reduction of cost (e.g., 5%) or confidence interval (e.g., 95%) will give the predictive model the lower bound range of the accuracy. Predictive accuracy derived below the lower bound will be considered an ineffective model of prediction in this report, with an evaluation given for each effective and ineffective model and mitigations to overcome any hurdles in business limitations from the models.

2.2. Business Objectives

In this project, we would be conducting predictive analysis on the churn rate and the different services provided by AT&T as variables to find out their correlations. Afterwards, the team would be constructing various predictive models using machine learning methodologies to provide AT&T with a model for predicting Churn rates and service bundles to reduce churn metrics. At the same time, the team would be doing a comparison analysis for all methodologies in the project to evaluate the best model that is the most accurate for answering each question. From the predictive models, AT&T would be able to perform and derive marketing plans to better value add to the quality of the services and leverage on business intelligence, coupled with data-driven models to provide value for stakeholders and investors, while focusing on a customer-driven business model. However, internal-related variables derived from the company's sales and revenue may be more restrictive and are usually only available for the company's analytics. Nevertheless, the team was still able to obtain the sufficient churn and telecommunication information from online sources to determine churn reduction and optimisation of services to decrease churn rate across various consumer categories closely within AT&T's business activity, to target the following project feasibility needs:

With a predictive model for customer-centric variable prediction on business services and macro-economic factors, it will allow companies such as AT&T to be better able to identify and analyse churn metrics, Customer Lifetime Value (CLTV), and attractive services that may be rolled out/implemented as business opportunities to scale in the industry.

1. **Identify salient variables for Services offered:** Data inputs and variables are explored during predictive and descriptive analysis to evaluate the weightage of each association and their statistical significance. Companies can turn towards compatibility of services to reduce churn across various Churn Outcomes or enhance their current policies to further leverage on the models to create greater business values to consumers and improve customer retention.
2. **Predicting and monitoring variables related to Churn Score:** The ability to assess risk and re-evaluate current policies or services to attract and retain customers allows companies to reduce risk of customer loss and hindsight in customer retention strategies. The company can identify intervention points of their current service packages. An example would be customers with increasing service requests being linked to a higher probability of Churn, and hence AT&T can implement strategies for a seamless customer journey or experience.
3. **Identify factors and customers leading to increasing CLTV:** The models describe variable importance or a coefficient which has a weight on determining Customer Churn or the customer lifetime value. Acquiring a new customer is said to be anywhere from 5 to 25 times more costly than just retaining an existing customer. Existing customers also typically spend 67% more than new customers *(Karnes, n.d)*. Thus, it is of utmost importance for AT&T to understand how they can increase CLTV, to earn more revenue from them and to minimize potential customer acquisition costs.

2.3. Desired Outcome

The desired objective is to decrease overall churn and increase average CLTV, while retaining customers who have high CLTV as they would contribute to the AT&T's profitability. Thus, we aim to reduce churn by accurately predicting variables and recommend insights generated from the models to help the business achieve quantitative revenue targets.

1. Actionable data-driven insights to answer key business questions derived from good models

   To derive valid insights from our data, we need to have a good machine learning model. To evaluate the models, we are going to use a few statistical metrics such as accuracy, precision, recall, ROC-AUC and F1 score. As these metrics measures different things and there may be no single model that excels in all of these metrics, finding an all-rounded model with good scores in all of them would be optimal.

2. Trends to explain AT&T customers' behaviors and find domains which we can explore further

   Through data visualizations and unsupervised models, we hope to discover many interesting trends and associations between the variables. Through these discoveries, we would attempt to identify which opportunities to improve which we are missing out. This will allow us to better understand AT&T customers' purchasing behaviours and telco service preferences to better personalise their experiences.

3. Data Preparation

3.1. Data Sources

Constructing the predictive models require the team to explore the Customer Churn related variables as well as Customer Churn scores. As data from AT&T is confidential and limited, we will be using and merging data of similar telco companies. The primary datasets used are “Telco customer churn: IBM dataset.csv” and “JB Link Telco customer churn.csv”. All of which are sourced from Kaggle, a reliable open-source community site.

The IBM dataset *(Kaggle., 2021)* contains customer-related information (e.g., their geographical region, services subscribed, service usage, monthly charges & other personal information) as well as the churn scores of individual customers. The other datasets are chosen because of the sufficiency of the data provided by respective organisations which helps to complement the main IBM dataset. Kaggle is an online data science community which allows anyone, from beginners to professional data scientists, to publish and distribute datasets, programme codes, and data models, to allow everyone to achieve their data science objectives. Since everyone is free to share data, there may be concerns regarding the quality and reliability of the data. However, we have assessed all relevant datasets to ensure that they are dependable and provide enough relevant information to work with. The main IBM dataset is especially trustworthy as it had been used to build analytics models in a data science publication *(Utterback C., 2021)*.

The second dataset *(Kaggle., 2022)* used was based on a small size telecom company (JB Link) providing phone and internet services to customers in more than 1,000 cities. In addition, customer attrition rate turned out to be 27% in the most recent quarter and decreased in almost 12% of the total number of customers. With the realism of the data employed from an actual scenario, we found it attractive to use it in this case study.

3.2. Data Cleaning and Manipulation

a. Merging of datasets

The 2 datasets mentioned in Section 3.1 have identical number of observations and customerID column. Therefore, we merge the 2 datasets using customerID as the merging key into a combined dataframe for easier manipulation.

b. Dealing with missing values

Upon checking, there are 3 columns with more than 80% of missing values, and the rest of the columns contains no missing value. As the proportion of missing values for those 3 columns are too high, and it is likely to lead to high inaccuracy if we try to fill in the values, we dropped those columns for our analysis.

c. Encoding categorical variables

One-Hot encoding is applied to multi-categorical variables to prepare for machine learning models later on.

d. Classification of variables

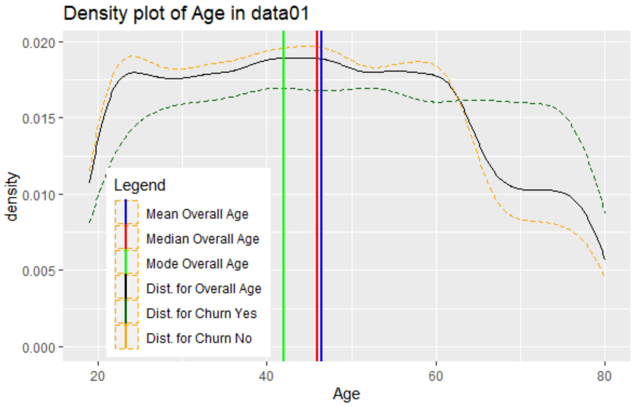
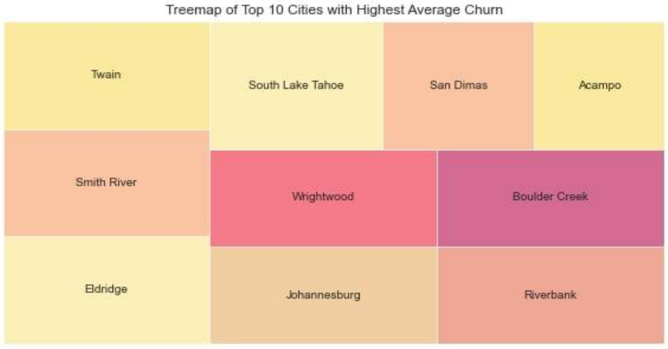
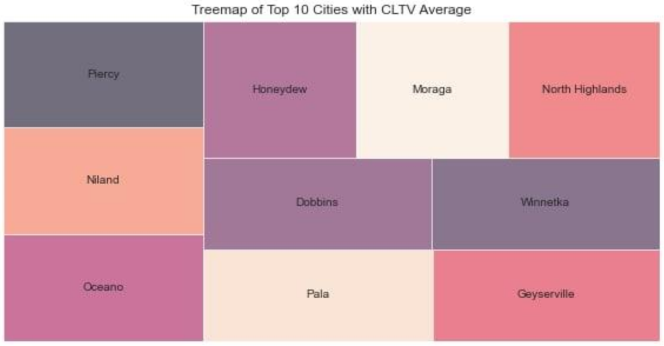
As the number of features is large in our data, we categorized all variables into 4 categories so that we can do: customer demographic-related, service/products-related, customer-billing-related, and churn-related. The details of each category are available in **Appendix I**. Those categories will be useful for our problem analysis and generating business insights.

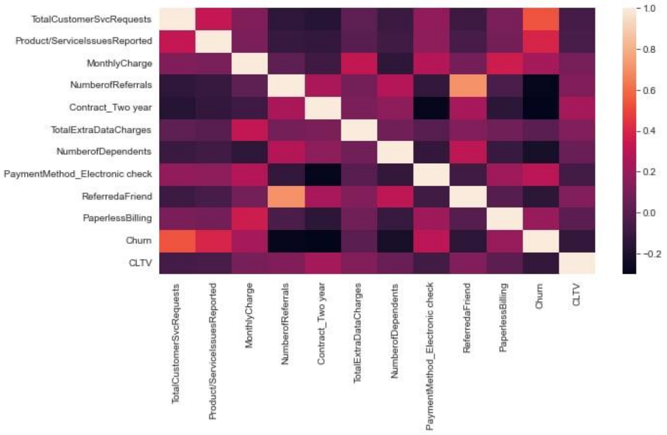
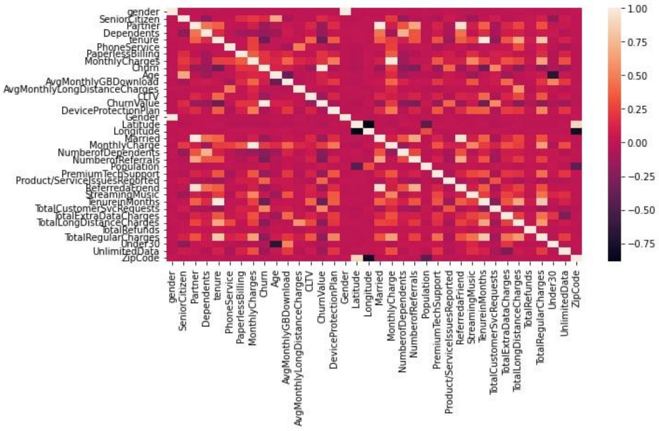
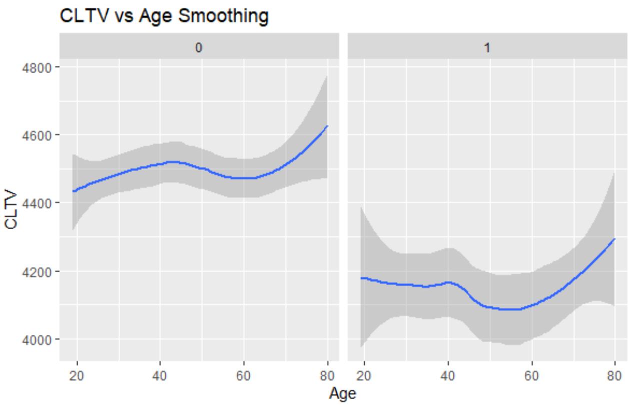
e***. Other feature engineering***

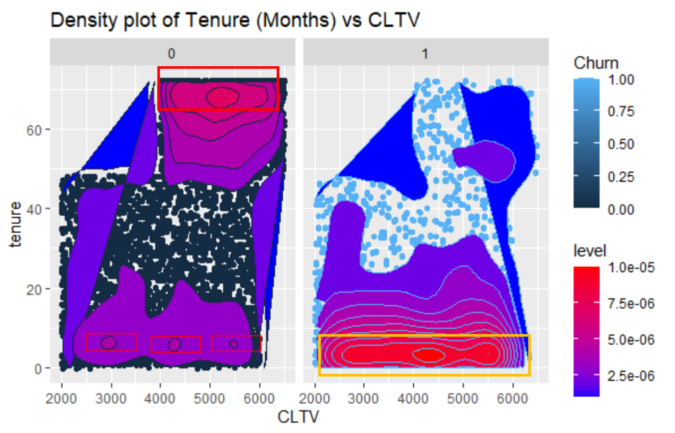
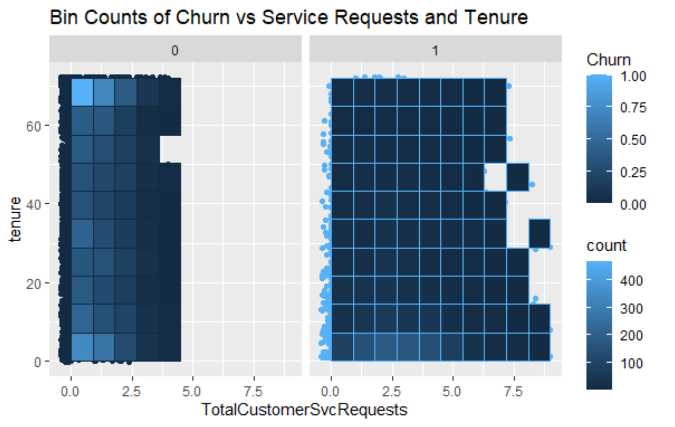
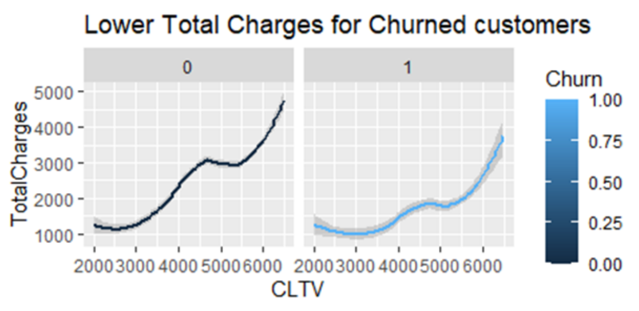
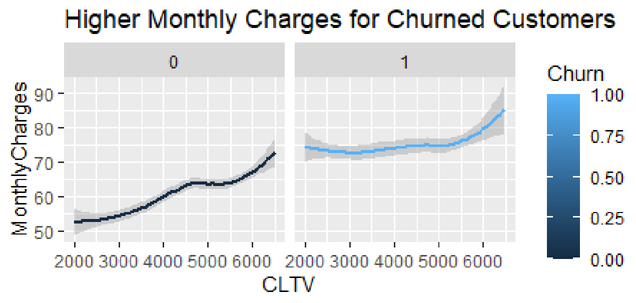
Other feature engineering includes: replacing “yes” and “no” to 1 and 0 for categorical columns for easier processing; created a new column “InternetService” to indicate whether or not the customer currently subscripted to internet service, and replace the “no internet service” values in other categorical columns with “0” (so that we can do filtering directly from the created column, and transform those other columns to binary categorical variables for easier processing).

4. Data Exploration

Through preliminary data exploration, it was found that there is a significant class imbalance of the Y variables, Churn and Customer Lifetime Value (CLTV), where Churn is categorical and CLTV is continuous. From a macro-overview, certain data variables exhibit high correlative properties with these Y variables (Appendix B). Taking into account these variables, the graphs are scaled to obtain a more accurate outlook on the relationship between the X and Y variables and to discover any dependent relationships that would otherwise have been overlooked due to the discrepancies which may affect the models and variable selection leading to biased results.

Variable	Data Visualisation	Elaboration
Age		From the data, we can see that there is an even distribution of age. However, there is a greater composition of people in their late 60s who exhibited higher tendencies to leave the current Telco Company rather than to stay with the company. Individuals with an older age could have higher attrition due to unsuitability in the products of the telco as they could be unfamiliar with the usage and deployment.
Top 10 Cities with Highest Average Churn	 Refer to Appendix B image b for an enlarged image	Geospatial data analysis revealed that customers displayed a higher rate of churn in these cities. This could be due to presence of other competing companies such as T-Mobile and Verizon.
Top 10 Cities with Highest Average CLTV	 Refer to Appendix B image c for an enlarged image	Geospatial data exploration showed that customers possess a higher value to the company as they would increase purchases of services with AT&T. hence, AT&T can utilise marketing techniques from these areas to areas with higher churn for customer retention and choose to focus on developing strategies to increase expenditures from residents in these areas.

Correlation plot of macro-factors influencing attrition and CLTV		Macro factors are factors that are out of the business control, such as customer requests and issues, number of friends, preference for paperless billing. Such variables may have a positive correlation on CLTV or Churn. For example, the 2-year contract has the highest contribution to CLTV, where customers who are with the company for a longer time tend to spend more on services with AT&T. We can also observe that the total number of service requests is highly related to Churn, and hence AT&T should position towards better customer service as a differentiating factor from competitors to avoid attrition.
Correlation plot of services provided that customers purchase related to Churn and CLTV		In this correlation plot, we focus on internal factors relating to services that AT&T can provide, such as phone services, billing methods, device protection plans in relation to CLTV and Churn value. Similar to macro factors, we can see those services such as issues reported by customers and monthly charges are highly linked to Churn. For CLTV, the top 3 variables as shown by the correlation plot are tenure, total long-distance charges, and total regular charges. Such charges are direct influences of the revenue obtainable by AT&T.
Smoothed plot of CLTV vs Age		A general trend across CLTV compared with age in the Churn categories is that CLTV is higher for customers who do not leave, while customers who possess an overall lower value to the business. Amongst both churn categories, we can see that churn value decreases after age of 40 years old. This is a good point of inquisition for the business as they can choose to focus less on marketing for customers within the age group specified. However, correlation does not signify causation and there could be external factors involved in the lower CLTV for those age ranges which will be explored by our models.

Tenure (Months) vs CLTV		The 2D density estimation map provides a bird's eye view of Tenure in relation to CLTV, where redder areas are indicative of high CLTV and are optimal from a visualisation aspect for AT&T. The heatmaps have been split into 2 categories: Churn = 0 and Churn = 1. We can observe that for customers without churn (Left Graph), AT&T can aim to retain customers for tenures at areas denoted by the red boxes, while for customers with churn (Right graph), they would be in the orange box.
Service requests and Tenure vs Churn		Through binning the data in tenure variable into 10 different bins, a distinct trait is that for customers who have a positive churn, they possess an equally distributed churn count across different tenure durations and service requests. In addition, service requests for customers who did not stay with the company have a greater number of service requests than customers who stayed. Measures can be put in place to mitigate the additional service requests to reduce the churn rate.
Lower Total Charges for Churn Customers		There is a general increasing trend for total charges for churn customers. However, the Total Charges at higher CLTV levels are lower and this could be due to the preference of low expenditure that customers would have if they were more likely to leave the company. This may also be due to the sunk cost fallacy – customers who spent more in the company are more likely to stay committed, and thus less likely to churn.
Higher Monthly Charges for Churn Customers		Churn customers experience higher levels of monthly charges across all CLTV levels. The pricing of services for customers in this area is not justifiable given the lower total charges for churn customers. The higher price could lead to price sensitivity in customers and will be investigated in our models for churn outcome.

5. Analytics Solutions to our Business Problem

Under this section, we will explain how we have utilized data analytics tools to solve the 3 key business problem identified in **Section 1.2 Business Objective**: namely, 1) identify consumer's preference for our service products, 2) better predict and monitor customer churn, and 3) determine what are the group of customers that are most valuable to us.

For each of the business problems, we include 2 parts:

1) **Solution Methodology & Analytic Tools**
   i.e., our approach to the problem, analytics tools involved, and how we try to achieve our desired outcome during the process of developing the solutions
2) **Findings & Business Insights**
   i.e., Key relevant findings we obtained, how our solution tackle the business problem, and any other valuable insights related to the business problem for AT&T to consider

5.1. Problem 1 – What are the more popular combinations of telco services/products?

a. Solution Methodology & Analytics Tools Involved

We are using the apriori algorithm from the mlxtend's python library to create the association rules. Since we are focusing on finding out the associations between the services, we sliced the dataset based on 3 general services provided: internet services, phone services and other services. The exhaustive list for all the columns categorized inside those 3 general services will be available in the **Appendix C.**

To find out which products/services are often found together, we ran the apriori algorithm several times on datasets containing different mix of services, which are: All services, internet services, internet and other services and phone and other services.

Aside from answering the problem statement, we also tried running the apriori algorithms on the same datasets, sliced by Churn == 0 and Churn == 1. From this, we tried to spot any behaviour differences between customers who stayed or left the company.

b. Findings & Business Insights

In this part, we will first describe some noticeable associations found from the results of the apriori algorithm (for all 12 variations), and then provide the business insights and recommendations

In this part, we will first describe the noticeable associations found from the results of the apriori algorithm (only the noticeable ones, the full list of results can be found in **Appendix D**, and then provide the business insights and recommendations derived from the results.

Table of Noticeable Findings (overall from 4 main categories):

1) Noticeable associations from rules from all services:

1. InternetService, PaperlessBilling and PhoneService are heavily connected with each other. As the top 6 rules with the highest lift are permutations of these services UNLESS InternetService → PhoneService or vice versa.
2. If a customer is subscribed to InternetService he/she is less likely to go for PhoneService and vice versa, although the lift difference from the balance point (1.0) of this rule is extremely small.
3. When PaperlessBilling and InternetService are picked together, customers are somewhat less likely to get PhoneService and vice versa (very small difference between the balance point)
   a. Although this rule's confidence is high, it is likely not statistically significant as the lift is very small.
4. For customers that have both PhoneService and InternetService they are likely to opt for Contract_Month-to-month and vice versa
   a. This is likely majorly influenced by the InternetService part of the product, as customers with only InternetService are also likely to opt for the monthly contract.
5. As for those that subscribed to the PhoneService, they are likely to opt for MultipleLines and vice versa.

2) Noticeable associations from rules only from internet-related services:

1. StreamingMusic, StreamingMovies and InternetServices are highly correlated with each other, as the top 5 rules are permutation of those services.

Comparing association rules between customers who stayed and left the company
(In terms of rules from all services)

1. In the dataset where customers churned, there are no rules pertaining to the UnlimitedData service, where in the dataset where customers stayed, getting UnlimitedData will get InternetService and vice versa is the best rule, ranked by lift. In fact, the rules are 100% confident that if a customer has UnlimitedData, then he/she will get InternetService (which intuitively make sense).
2. In both datasets, the trend of having either InternetService or PhoneService means it is less likely to have the other staying true with also very small significance.

With the above findings from our model, below are the business recommendations for product/service bundling combinations that could potentially help to increase revenue.

Product/service bundling is a great way for companies to boost sales and persuade customers to shop more. Finding the most optimal bundling is very important. From the association rules, we found some services that customers tend to have together and thus, we can create a bundling promotion for it. Overall, we have 5 recommendations:

1) InternetService and PhoneService

Although there is a rule that states getting an InternetService means it is less likely to get the other one and vice versa, we believe that since the lift difference from the balance point (1.0) is so small, it is not statistically significant enough. However, there is a significant rule that customers will get InternetService if he gets PhoneService together with PaperlessBilling. Hence, there is an indirect correlation between the two.

2) InternetService, StreamingMovies and StreamingMusic

These 3 services are heavily intertwined with one another on multiple rules. This implies that customers like picking these 3 or combinations or these 3 products together.

3) PhoneService, MultipleLines

Similar to the other bundles, PhoneService and MultipleLines shows up as a rule with high lift and passable confidence.

4) InternetService, UnlimitedData

When zooming deep between rules created from datasets that left and stayed with the company, we found that the most noticeable difference is that people who stayed within the company would get UnlimitedData given that they have InternetService. Although we write a direct correlation between UnlimitedData and staying with the company, it is something worth exploring.

5) Set PaperlessBilling as the default

Taking a quick glance at the rules, the PaperlessBilling showed up on a lot of the rules with high significance, especially those that contains InternetService and/or PhoneService. Aside from providing a better experience, AT&T can save on printing and mailing bills’ costs.

5.2. Problem 2 – How to improve prediction and monitoring of customer churn?

For the second business problem, our goal for our solution is to

1) Develop an analytics tool to help AT&T predict customer churn accurately (with the least cost of wrong prediction), and
2) Gain insights on what factors have been contributing to customer churn so that operational adjustments, if necessary, can be made.

a. Solution Methodology & Analytics Tools involved

As an overview, the following were the steps taken to develop our analytics tool solution:

1) Firstly, we **tested multiple popular Machine learning models** on the given dataset to predict customer churn, and obtain the performances of each model
   a. 4 models are tested – Logistic Regression, CART, Random Forest, XGBoost
   b. 5 metrics used to evaluate the performance – Accuracy, Precision, Recall, ROC-AUC, F1 Score
2) **Compare the performance** of all models using the 5 metrics, the best performer is chosen as the final model option as our prediction tool
3) **Optimize** the chosen model (hyperparameter tuning & feature selection)
4) **Final prediction model** will be the model with the optimal hyperparameters, using the most preferred group of features (from feature selection) as model input

With the above steps, the final model chosen is:
Optimized XGBoost Classification Model with Top25 important features as model input

Detailed explanations are as below:

Step 1&2: Obtain the performance of multiple models, pick the best one

Considering our business context (i.e., to predict customer churn), we have chosen 5 relevant performance metrics to for model evaluation, our rationale is as below:

1) **Accuracy**
It is used to ensure that the predicted results when compared to the overall results are close to the testing data.

2) **Precision Score**
It can be used to detect the errors that may result from the model, as it can detect the false positive rates as false positive rates may be costly for the business where additional resources may be required but wasted to implement a customer retention strategy for a customer who is not undergoing churn.

3) **Recall Score**
It helps to compare the actual positives to the overall positives, which can be translated to helping AT&T achieve their target conversion rates or predict the targeted Customer Churn *(Exsillio., 2016).*

4) **ROC-AUC Score**
It provides a visualization of the correlative probability of the predictions and target variables, showing the suitability of the models in ranking predictions.

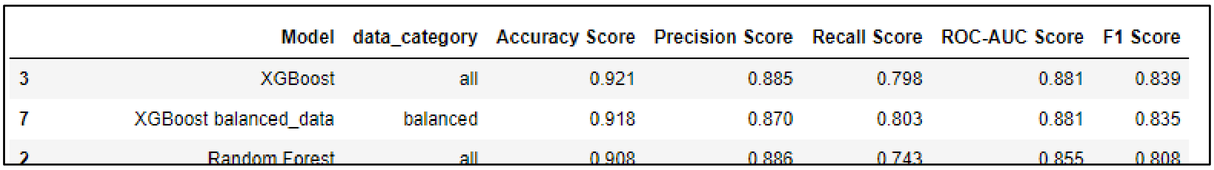
5) **F1 Score**
It combines precision and recall into one metric while being able to measure imbalanced data. This helps us to select the model as stable models would display consistent F1 scores whether it is balanced or not *(Czakon J., 2021).*

On the technical aspect for model testing, the details are

- Seed = 2022 is used for all random states; 80-20 train-test split were used
- All models are run with 2 trainset data variations
   a. Original Trainset data

b. Balanced Trainset data (balanced via oversampling, using the Synthetic Minority Oversampling Technique (SMOTE) tool)

The full performance results for all models' variations can be found in **Appendix E**. Among all model and variation testing, XGBoost with original trainset data had the best performance (highest score across all 5 metrics).



	Model	data_category	Accuracy Score	Precision Score	Recall Score	ROC-AUC Score	F1 Score
3	XGBoost	all	0.921	0.885	0.798	0.881	0.839
7	XGBoost balanced_data	balanced	0.918	0.870	0.803	0.881	0.835
2	Random Forest	all	0.908	0.886	0.743	0.855	0.808

Figure 3.2.1. XGBoost with original trainset data rank the 1st in terms of all metrics

Therefore, we chose XGBoost as the model to use for our customer churn prediction tool.

Step 3: Model Optimization

Our model optimization is done in 2 aspects: hyperparameter tuning, and optimal feature selection (for model input).

Hyperparameter tuning

Hyperparameter tuning is needed here to improve the overall performance of our model and also to mitigate technical issues such as overfitting.

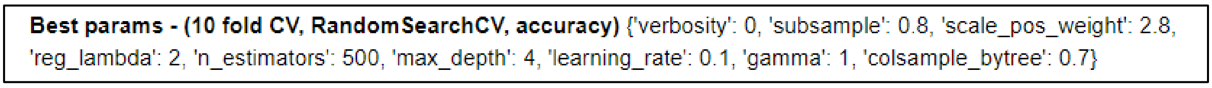
As **our business goal** is to develop a tool that is **both accurate** and also **produces the least cost for a false prediction**, we chose 2 key scoring metrics to tune the model: "accuracy", and "recall_score".

"recall_score" was chosen as one of the scoring metrics because we recognized that a False-Negative (FN) prediction is more costly than a False-Positive (FP) prediction. This is so because a FN prediction means that we predict the customer will NOT leave (churn=0) while in reality he DOES (churn=1); but if our customer management team operate based on the prediction (will not leave), they will be unlikely to direct extra attention to this customer, result in a direct loss in customer count. Such consequence is much more serious compared to that of FP prediction (where the worst case will be paying more attention to the customer, but unlikely to result in loss of customer count).

Overall, the following hyperparameters were chosen for tuning:

"learning_rate","reg_lambda", "scale_pos_weight","max_depth","gamma","sub_sample","colsample_bytree","n_estimators",and "verbosity"

Operational wise, we further split the trainset data to achieve overall 60-20-20 train-validation-test set split and performed a 10-fold cross-validation on the train and validation set for the 2 scoring metrics, using the RandomizedSearchCV to obtain the optimized set of parameters (as a traditional grid search is too computationally expensive). We then run our model with the 2 sets of parameters on the testset data and chose the one with the better performance. Here, the better one is the set optimized with "accuracy":



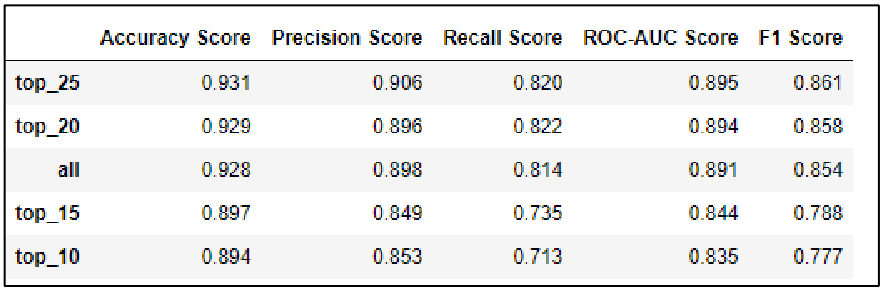
```
Best params - (10 fold CV, RandomSearchCV, accuracy) {'verbosity': 0, 'subsample': 0.8, 'scale_pos_weight': 2.8,
'reg_lambda': 2, 'n_estimators': 500, 'max_depth': 4, 'learning_rate': 0.1, 'gamma': 1, 'colsample_bytree': 0.7}
```

Refer to Appendix F for the comparison between the 2 sets of parameters.

Feature selection for model inputs

Feature selection is also needed here, as we want to reduce the complexity of our model (so that it can works faster) and improve the convenience of use (as a smaller number of data input will be required).

This step is fairly simple – we ranked all features in terms of their respective feature importance (by Gain Score) in the model and selected the top ones as the final required inputs for our model. Here we tried a few variations: top10, top15, top20, top 25, versus all, and we found that with top 25 variables as model input, the model performs the best. (Refer to Appendix G for feature importance ranking)



	Accuracy Score	Precision Score	Recall Score	ROC-AUC Score	F1 Score
top_25	0.931	0.906	0.820	0.895	0.861
top_20	0.929	0.896	0.822	0.894	0.858
all	0.928	0.898	0.814	0.891	0.854
top_15	0.897	0.849	0.735	0.844	0.788
top_10	0.894	0.853	0.713	0.835	0.777

Figures 3.2.2. Top 25 features as model inputs generates the best performance

Step 4: Final Model (i.e., our analytics solution model)

With the above steps, we were able to develop an analytics model which can predict the customer churn most accurately & with the least cost of any false prediction – i.e., *Optimized XGBoost Classification Model with Top25 features (ranked by gain score) as model inputs.*

A sample performance test for our model is as below, where it achieves 93.12% test accuracy:

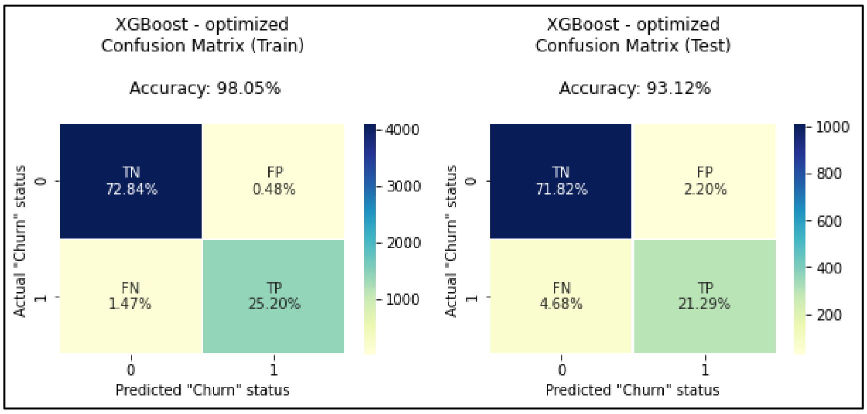


Figure 3.2.3. Prediction Model Performance

This satisfied both our business objectives as well as our desired outcomes for business problem 2.

b. Findings & Business Insights

Recall that we have mentioned that there is another aspect we are looking into for business problem 2 (customer churn) other than just making predictions – we also want to derive some insights on what the factors have been contributing to customer churn, so that AT&T could make necessary operational adjustments if required.

This could be done by investigating further into our prediction model and finding out how are the input features contributing differently to the churn (i.e., whether they are contributing positively or negatively, and what's the degree of contribution).

We used the SHapley Additive exPlanations (SHAP) to tackle this aspect of the problem. SHAP is a game theoretic approach to explain the output of a model and it is useful in interpreting Blackbox models such as XGBoost. Here, we did a credit allocation analysis on our model (with different categories of features), and some interesting findings & valuable business insights are as below (refer to Appendix H for all relevant credit allocation plots):

1. Among all variables, we found that the **number of total customer service requests** has the highest contribution to the churn value from the plot below. It can be observed that having a total customer service request of 5 (and above) could significantly increase the probability of a churn (i.e., churn = 1).

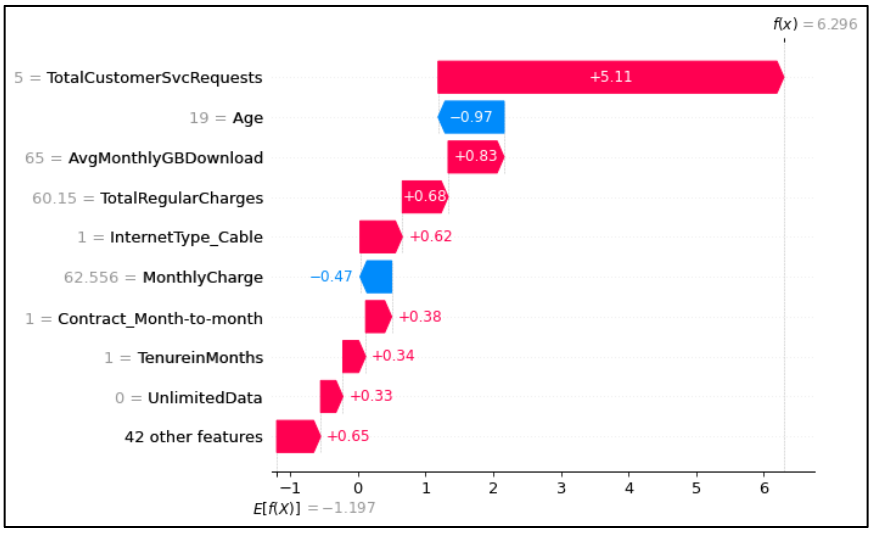


Figure 3.2.4. Credit allocation analysis on all variables

This makes sense and matches well with human intuition. According to a study by Forum Corporation, poor customer service (70%) is the leading cause of churn *(Plaksjj Z., 2021)*. The more service requests customers had to make, the more unsatisfied they felt about the service experience, leading them to end their contracts eventually.

Business insight: AT&T **should try to resolve issues faced by customers in less than 5 calls/requests**, so as to increase customer retention. Hence, better-trained customer service staff who can resolve customer issues will be crucial for AT&T.

Another worth noting interesting finding from the plot above is on how higher total regular charges and lower monthly charges influence churn (as seen from the plot above, *TotalRegularCharge* is in red while the latter is in blue). This contradiction may seem illogical and so a closer examination could be done in the future. For now, a probable explanation for this discrepancy is that customers would be more price sensitive in the short term, due to negative price elasticity when results from demand of the telco's service as price increases, while regular customers would not be as concerned with paying higher charges. This provides AT&T with pricing ability for their services. Hence, AT&T can **leverage on pricing economics to attract newer customers in the short term with lower charges**.

2. Among all demographic variables, we found that customer age has the highest contribution to customer churn as see from below. Here we see, having an age of 19 (and above) could significantly reduce the chances of a churn:

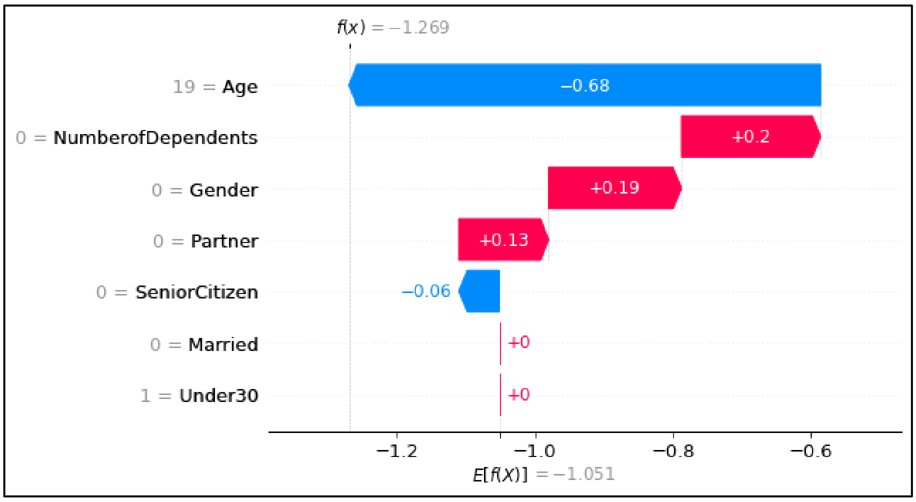


Figure 3.2.5. Credit allocation analysis on demographical variables

Again, this finding matches well with real-world situations – a research report from Deloitte stated that the GenZ population (i.e., younger age) tend to switch their streaming service subscription providers more often than the older age groups *(Weprin A., 2022)*. One reason is that the younger generations are digital natives, who do not face much trouble adding and cancelling a service unlike the older generations. In addition, the study also mentioned that almost half of millennials and 34% of Gen Zs who had unsubscribed, have resubscribed to the same service in the past year. This is because many younger consumers are just subscribing to watch a particular show that is released on the service.

Business insight: AT&T can **potentially relook into their current programmes and services**, to evaluate what type of new and relevant content can be brought in to **persuade younger customers to stay**.

3. Among all billing-related variables, we found that **total regular charges** and **a month-to-month contract type** have the highest contribution for customer churn. Typically, a total regular charge of 60.15 (and above) and a month-to-month contract type can significantly increase the chances of churn:

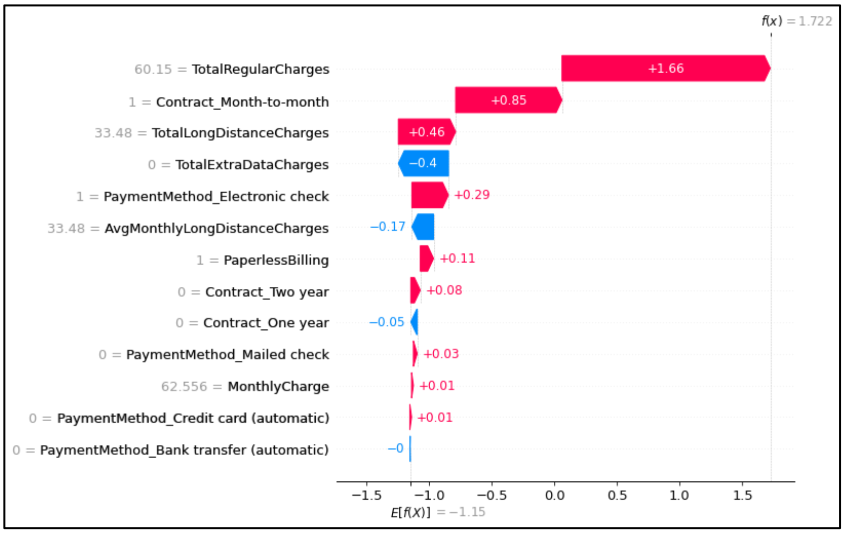


Figure 3.2.6. Credit allocation analysis on billing-related variables

This finding makes sense intuitively too – if the total regular charge is too high, the customer will be more likely to leave AT&T and switch to a cheaper option, therefore lead to a higher churn; in addition, when the contract duration is shorter (M-on-M is the shortest option available), the customer will have more chance to switch service to other options (compared to if the contract period is longer, the customer is more likely to stay as they have been used to AT&T's services and have higher loyalty).

Business insight: With this, AT&T can potentially look into setting thresholds of Total Regular Charges not more than $60 and devising contracts on a longer-term scale (3 Months minimal contracts without cancellation).

4. Among all service-related variables, we see that the use of cable internet also has a greater impact on causing churn than DSL or fibre optic, as seen from the below plot:

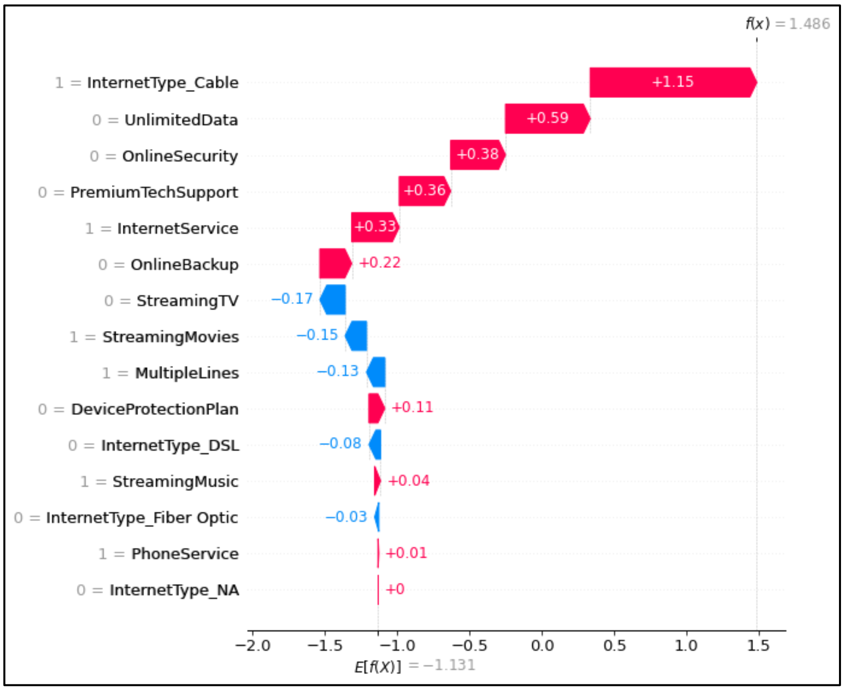


Figure 3.2.7. Credit allocation analysis on billing-related variables

This is surprising because compared to DSL and fibre optic, the price and speed of cable internet is in the middle of those two.

Business insight: AT&T could delve **deeper into their cable services and investigate if customers face greater problems in the cable services** compared to the others, and make any adjustments or improvements, if possible, to **reduce the customer churn caused by the use of internet cable**.

5.3. Problem 3 – Which customers are more valuable to us?

a. Solution Methodology & Analytics Tools Involved

Linear Regression was used to determine variables that are important to CLTV values and to obtain a best fit linear relationship between the factors in Figure 3.3.2 and CLTV. Based on the imputed values into the Linear model, the variables will be ranked according to highest positive to lowest negative correlation in comparison to the target variable CLTV. This will be shown in 3.3.2.

b. Findings and Business Insights

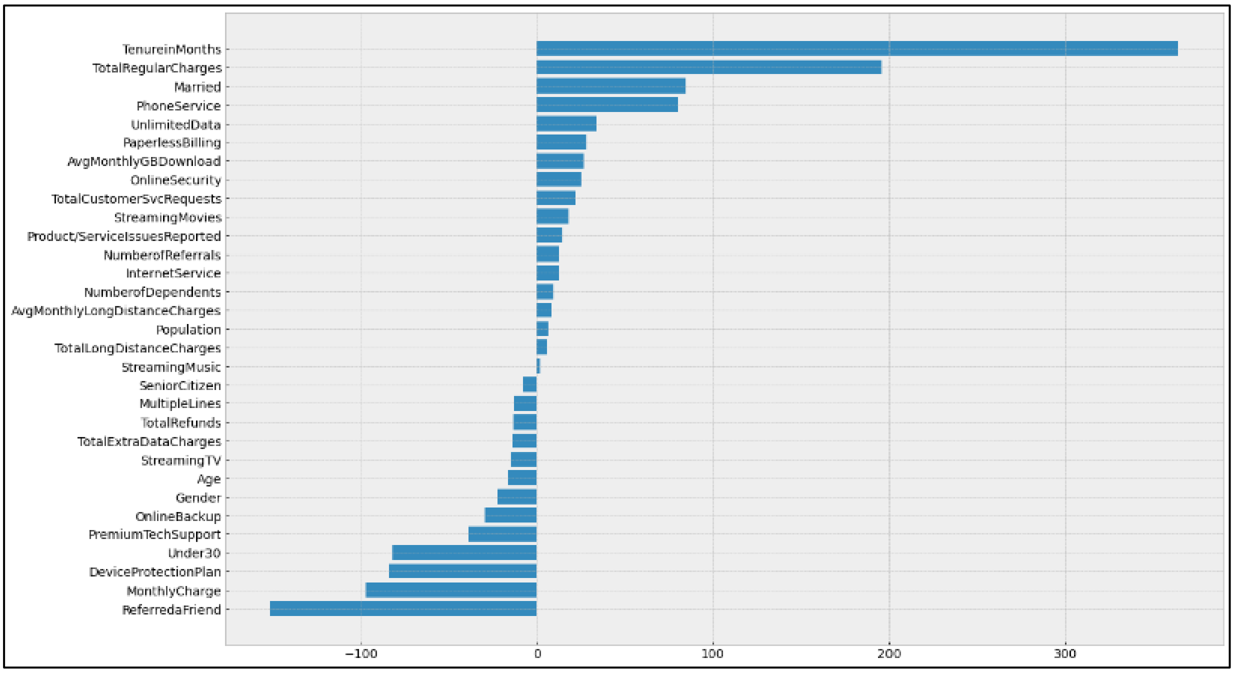


Figure 3.3.2. Feature importance in Linear Regression

1. Tenure, which is defined as the length of time between the customer's first and last order date, can be used to measure the existing recent expenditure in the current period. Customer Lifetime Value (CLTV) can be highly linked to the Tenure that the customer stays with the company. The importance of tenure in the relationship with CLTV is that it is less costly to retain existing customers than to conduct newer customer acquisition, so the value of existing customers through a longer tenure helps to increase CLTV (Qualtrics., n.d.). The decreased cost of newer acquisition increases the Customer Lifetime Value (CLTV) and is better for AT&T as lower expenditure will be needed for marketing and internal relations.

Business insight: AT&T can aim to **extend tenures on customers** or to pivot towards customer centric services by **conducting surveys with current customers** and garnering feedback (to better understand their needs), instead of running newer customers acquisitions campaign.

2. The second to fourth most important features in CLTV are the 'Monthly Charges' and 'Total Regular Charges'. This is congruent to CLTV being a revenue metric, and as charges appear as the top few variables, it shows a relationship between the charges that the business can put in place for pricing in the short run (Monthly pricing/charges) and long run (Total Charges). Lower monthly charges may be enticing customers to stay, leading to higher lifetime value. This leads to higher total charges which results in an increased customer lifetime value, where AT&T can use the predictive results to identify customers with similar CLTVs who raise the maximum revenue for the company and focus on retaining them. AT&T may also leverage on acquisition of new customers with similar expenditures, interests, and behaviours as the current highest CLTV customers through lookalike modelling for acquisition *(Srivastava D., n.d.)*. However, one important perspective to take note of is that customer with a higher monthly churn display higher Monthly Charge on average (as shown in data exploration), and pricing for services/products to customers is important too, while customers may be indifferent on a longer timeframe.

Business insight: Low-cost leadership strategy for customer acquisition whereby customers will be enticed to sign on with AT&T due to initial low costs per month, and once they are with AT&T for a longer term, AT&T may rise prices when they have market pricing power.

3. Age also plays an important role in predicting customer shopping behaviour and where personalization of individual actions and preferences of telco providers/services are interlinked *(Mckinsey., 2021).* Different individuals across age groups may possess preferences for different services which have various profits margins attached to them. In our insights, age is negatively linked to CLTV. For example, the younger generation may prefer faster speed Internet Services which can be costlier than the preferences for legacy architecture internet systems/services in the older population.

Business insight: AT&T can **segment customers by age**. They can choose to look towards improving the CLTV of the elderly for higher profitability, attracting them with what they want – stability, informational resources, and strong customer support. They can continue their focus on the already strong CLTV of the young population segment through bundling of services such as phone service and unlimited data which can negate the downside risks (of churn) from sales to younger generations.

6. Limitations of our business solutions from the model

6.1. Limitations of solution for problem 1

One main limitation is defining what a "good" association rule is. We need to be able to decide the optimal parameters to use in the algorithm which will change the quantity and quality of association rules. This will impact our analysis and subsequent recommendations. As of now, there is no widely accepted cut-off values for these metrics, which boosts the model's subjectivity and making it hard to get the most optimal results possible.

6.2. Limitations of solutions for problem 2

To predict the Churn outcome, we used different categories to test the Model. One of the categories showed that in demographic, when Presence of Dependents was 0, and when Number of Dependents (Number of family members) was 0 (Figure 3.2.5), both variables showed different correlative directions with CLTV, despite the intended outcome of absence of dependents. Hence, one of the variables for dependents could have been removed to provide greater analysis under demographic categories to predict Churn Outcome.

Tenure of customers (Figure 5.2.2) was classified under the service category, where customers would be more likely to leave after one month. However, when our team reconciliates the data, we can derive that a limitation of XGBoost is its inability to perform well on sparse data. As the 2D density estimation map showed that the data was uneven and scattered, this could lead to a biased selection of Tenure = 1 month leading to a higher probability of Churn. Hence, tenure as a variable estimator could be an unreliable indicator, as the model is sensitive to outliers which creates low scalability in the model outputs *(Hachcham A., 2021).*

6.3. Limitations of solution for problem 3

AT&T needs to understand that just because a variable is a strong predictor in linear regression, it does not mean that the business should focus on modifying that variable. For example, higher total regular charges have a strong impact on increasing CLTV, only because they are correlated revenue metrics. Therefore, this does not mean AT&T should just increase its regular charges to improve its CLTV. Doing that can actually result in the opposite, as higher prices (especially if it is unreasonable) may anger customers, making them more likely to churn, and thus cutting their CLTV.

Another factor that is anomalous is that Friend referrals are negatively linked to CLTV. This may not indicate causation as in theory, friend referrals are supposed to bring in higher revenue/profitability.

One more possible limitation is that higher customer service requests are associated with higher CLTV. Again, this may not indicate causation because more service requests would indicate that customers are having some issues with the service. In theory, with more negative experiences with the company, they are less likely to spend more and more likely churn, which should lead to lower CLTV instead.

7. Business Applications

In order to allow managers from AT&T to fully and more conveniently utilize our solution tools, in addition to generating insights from our models, our team aims to value add through the following implementations:

- Allow real-time dynamic update of prediction results & business insights from our models, based on a constantly changing customer database
- Enable easy and convenient access to the key information from the current database as well as the results & recommendations from our model.

Therefore, we also developed an interactive dashboard using PowerBI as an integrated solution platform for the managers.

We have recognized a few important aspects of the data information and solution insights, and decided to include the following 5 sections in our dashboard:

1) Current customer characteristics

   This section will include useful plots about customer distribution, such as geographical distribution, demographical overview (age, gender etc.), to help the manager gain an understanding of their current customer pool (so that they can potentially do demographically targeted business strategies)

2) Current service/products distribution

   This section includes the distribution of the current service/products provision, such as % of customers who subscribed to a certain service, and what are the most popular ones. Those information will be helpful to allow managers to have an overview of the performance of their existing products and services.

3) Recommendation for potential bundle provision (solution for problem 1)

   Next, the outputs from our association rule model will be cleaned and displayed as one section in the dashboard, where the managers will be able to filter by confidence/support/lift level and see what the most popular combinations of service/products are, such that they can provide bundling services to increase sales.

4) Churn prediction results & critical factors to pay attention to (solution for problem 2)

   The churn rate from the model will also be shown in the dashboard, and the results from our SHAP analysis will also be displayed – this is to help the managers to directly access information such as what are the most important factors, what they should look into now to increase customer retention, as well as what is the projected churn rate given their current customer pool.

   Moreover, they will also be able to see the list of customers with high probability of churn (based on our model analysis), so that they can do targeted monitoring to make those customers stay with AT&T.

5) Potential list of customers with high value to the company (solution for problem 3)

   Lastly, the managers are also able to see the list of customers with high value to the company (this is done by filtering the customers with higher values for the positively correlated factors, and lower values for the negatively correlated factors). Those will help them know the customers that are of high importance to the company, and they can potentially provide some VIP or special service/promotion to those customers, so that they will continue to stay with us.

In the ideal situation, we would want to have all of the above sections for the dashboard, and also connect both the dashboard and our models to AT&T's real time cloud database to achieve real time updates. However, due to technical constraints and time limits, we are not able to develop the dashboard with full functionality yet. A sample screenshot of our dashboard can be found in Appendix J.

8. Synthesizing our Findings – Feasible Improvement Plans for AT&T to Adopt

With the solutions to the three business problems identified and outlined (and how we can achieve the business outcome), this report dives deep to extract and synthesize some of the most prominent findings we had found and provide a few feasible suggestions for improvement and immediate plans that AT&T could take on operationally to increase customer retention.

8.1. Product/Service bundling and changing defaults (Short-term Plan)

Products/services bundling is a common marketing technique that is used across all industries around the globe. Through selling multiple services together, we can simplify our customer's purchasing journey and increase sales efficiency *(Zix, 2020)*. Furthermore, instead of giving customers the whole ala carte list of our services, a more simplify and filtered down options can be presented to them, increasing the chance of customers subscribing to these additional services offered *(Murphy D., 2020)*.

Hence, we have recommended the following list of product bundles:

1. Internet service and Phone service
2. Internet service, Streaming movies and Streaming music
3. Phone service and Multiple lines
4. Internet service and Unlimited data

All these services are found to be highly related to one another statistically, where a customer that subscribes to one service would be likely to get the other(s). Furthermore, the act of bundling makes sense intuitively. For instance, the first package – Internet and Phone services are essential for connectivity in our daily lives in this day and age. The second package is also logical, as the music and video streaming has been gaining traction especially with the hit by the pandemic as seen by the rise of streaming services like Netflix, Disney+ and Spotify *(Kennedy, 2021)*. For the third package, it was found that customers who purchase a phone service are likely to get multiple lines on top of that. As for the fourth package, customers who stay subscribed to the company were observed to purchase Internet service and Unlimited data. Although we were unable to make a direct link between the two, this relationship can be further explored.

On top of product/service bundling, we would also recommend that AT&T set paperless billing to be the default option. People given limited options tend to make faster decisions and are more satisfied with their choices as compared to those given a more extensive array of options *(Schwartz B., 2005)*. This suggests that having a more streamlined process would simplify sign-ups and subscriptions, aid customers in decision making and improve overall satisfaction levels. As Paperless billing was found to be statistically most related to many AT&T services provided, it was often found together in multiple association rules, especially those containing Internet service.

8.2. Appealing to the younger audience (Long-term Plan)

As mentioned under Section 5.2.2, the younger populace (e.g., millennials and Gen Zs) are most likely to churn but they are also much more likely to return after a churn *(Weprin A., 2022)*. Thus, it is important to come up with strategies to retain them or incentivise those who have recently churned.

With regards to streaming services, AT&T can redesign some of their current programmes by introducing more tie-ups with trending shows that younger people are interested in watching. AT&T can also collaborate with production companies to produce exclusive shows that are only available on their platform, along with limited edition merchandises inspired by those shows. According to Y Pulse (2020), the primary reason younger consumers stick to streaming services is to escape advertisements. Thus, removing advertisements may be the key to pulling back their customers.

As for their telecommunications division, a survey by InMobi showed that younger customers desire improved internet and 5G connectivity *(Kaplan M., 2021)*. AT&T can focus on bringing in these improvements to the young. Besides that, AT&T can focus more on digital marketing towards younger customers, provide mass-personalisation, and also provide more engaging user experiences.

8.3. Improve customer service quality (Long-term Plan)

AT&T can provide better service standards to reduce total customer service requests. This can be implemented through internal processes such as service development curriculums and service trainings for employees. Despite the notion that the implementation of preventive strategies may be costlier to the company, it has been shown that with the support at key organisation levels such as higher levels of leadership, global organizations are able to pivot towards executing a seamless and consistent customer service experience. In addition, this may reduce the occurrence of repeat calls and misdirects. Customer service may be provided along the value chain too, as upstream and downstream partners in telcos such as the physical retailers may leverage on insights and services that are preferred by customers to deliver a differentiated experience and leverage on data and trends from previous customer touch points to tailor to the various needs of their customers and preferences *(Northridge., n.d.)*.

9. Conclusion

From this paper, we have successfully utilised advanced analytics methods to predict and explain customer churn and CLTV. We have explored the data and applied a few pre-processing techniques to statistically ensure that the data we fed into our models is optimal. Then, we derived which factors are important in determining customer churn and which have a significant impact on customers' CLTV. Moreover, we have explored popular combinations of services/products offered by AT&T through utilising an unsupervised machine learning model. Then from the insights derived from those models, we have proposed 3 recommendations that we believe would increase customer retention and improve sales.

Limitations of the report

However, we would like to recognize the limitations of our recommendations. Mainly, there are two limiting factors for our report, namely data quality and model limitations.

A biased data would train a biased model and yield biased results. In our dataset, we only have data for people who lived in California. This could result in our models being overfitted to the Californian market and not being generalised or correlative enough for the entire US market.

As mentioned beforehand in part 6 of our report, there are certain limitations inside our results due to the nature of the machine learning models. For example, XGBoost's tendency to perform poorly on sparse data may affect our results, as the data for the variable tenure data distribution is quite spread apart, which may not be indicative of actual results.

Statistical significance and variable importance should be used in conjunction with secondary research and thus the variables importance as reflected in the models selected should indicate to AT& the direction of policies that they should focus on.

Value of the report

Although limited, we believe this report has provided a data-driven recommendation for AT&T. Through implementing those recommendations, we believe that AT&T can observe an increase in customer retention, services sales and ultimately an improvement on their bottom line and increase in market share.

Future Prospects

Moving forward, AT&T can scale up our project into something they can use daily through larger data counts across different geographical locations. They can upload our model into the cloud and/or dashboards and have real-time updates from the models. Some use cases of those are: a real-time churn probability for their customers, so that their customer service team know which customer to prioritize, a real-time update of association rules for each customer, helping salespersons to pitch products the customer would like, and so on. The employment of time-series analysis with the various models will also spur the model for greater outputs and analysis with trends predictions and newer services that AT&T can provide to stand out from their competition.

10. References

Czakon J. (31 December 2021). F1 Score vs ROC AUC vs Accuracy vs PR AUC: Which Evaluation Metric Should you choose? Retrieved 27 March 2022, from: https://neptune.ai/blog/f1-score-accuracy-roc-auc-pr-auc

Dano M. (16 October 2020) US Wireless snapshot: Subscribers, market share and Q3 estimates. Retrieved 21 March 2022, from: https://www.lightreading.com/4g3gwifi/us-wireless-snapshot-subscribers-market-share-and-q3-estimates/d/d-id/764688

Exsillio. (9 September 2016). Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures. Retrieved 27 March 2022, from https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/

Fontella C. (n.d.) 16 customer experience trends and stats that'll define the next year. Retrieved 21 March 2022, from: https://blog.hubspot.com/service/customer-experience-trends

Hachcham A. (12 August 2021). XGBoost: Everything you need to know. Retrieved 1 April 2022, from: https://neptune.ai/blog/xgboost-everything-you-need-to-know#:~:text=Disadvantages,overall%20method%20is%20hardly%20scalable.

Kaggle. (2021). Telco customer churn: IBM dataset. Retrieved 12 February 2022, from: https://www.kaggle.com/yeanzc/telco-customer-churn-ibm-dataset

Kaggle. (2022). JB Link Telco Customer Churn. Retrieved 12 February 2022, from: https://www.kaggle.com/datasets/johnflag/jb-link-telco-customer-churn

Kaplan M. (1 November 2021) How Telcos can Shape the U.S. Mobile User Experience. Retrieved 2 April 2022, from: https://www.inmobi.com/blog/2021/11/01/how-telcos-can-shape-the-us-mobile-user-experience

Karnes. (n.d.). Customer Lifetime Value: What is it and how to calculate. Retrieved 1 April 2022, from: https://clevertap.com/blog/customer-lifetime-value/

Kennedy, G. (2021, June 17). The rise of streaming platforms: More shows, more money, more problems. The Crimson White. https://cw.ua.edu/81877/culture/the-rise-of-streaming-platforms-more-shows-more-money-more-problems/

Mckinsey. (27 October 2021). Customer Lifetime value: The customer compass. Retrieved 1 April 2022, from: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/customer-lifetime-value-the-customer-compass

Moritz S. (26 January 2022). AT&T Drops on concerns about slowdown for mobile business. Retrieved 21 March 2022, from: https://www.bloomberg.com/news/articles/2022-01-26/at-t-profit-tops-estimates-as-it-braces-for-a-wireless-slowdown

Munson B. (22 April 2021). AT&T loses 620,000 video subs in Q1 as DirecTV deal moves ahead Retrieved 1 April 2022, from: https://www.fiercevideo.com/video/at-t-loses-620-000-video-subs-q1-as-directv-deal-moves-ahead

Murphy D. (22 December 2020). Service Bundling Strategy-7 benefits for your small business. Retrieved 2 April 2022, from: https://www.business2community.com/small-business/service-bundling-strategy-7-benefits-for-your-small-business-02372696

Northridge. (n.d.) 7 tips to improve the customer experience and reduce costs. Retrieved 2 April 2022, from: https://www.northridgegroup.com/blog/seven-tips-dramatically-improving-customer-experience-reducing-costs/

Plaksij Z., 2021. (4 May 2021). Customer Churn: 12 ways to stop churn immediately. Retrieved 1 April 2022, from: https://www.superoffice.com/blog/reduce-customer-churn/

Qualtrics. (n.d.). What is customer lifetime value and how to measure it?. Retrieved 1 April 2022, from: https://www.qualtrics.com/au/experience-management/customer/customer-lifetime-value/#:~:text=Customer%20lifetime%20value%20is%20the,great%20way%20to%20drive%20growth.

Schwartz B. (July 2005). The paradox of Choice. Retrieved 2 April 2022, from: https://www.ted.com/talks/barry_schwartz_the_paradox_of_choice

Smith O. (n.d.). 10 Customer Retention Metrics and How to measure them. Retrieved 21 March 2022, from: https://blog.hubspot.com/service/customer-retention-metrics#:~:text=of%20the%20Period-,2.,stop%20doing%20business%20with%20you

Srivastava D. (n.d.) CLTV: Why is it an important metric and how to calculate it. Retrieved 1 April 2022, from: https://easyinsights.ai/blog/cltv-why-is-it-an-important-metric-and-how-to-calculate-it/

Statista. (11 January 2022). Customer Churn rate by industry U.S. 2020. Retrieved 21 March 2022, from: https://www.statista.com/statistics/816735/customer-churn-rate-by-industry-us/

Szalai G. (29 January 2020). AT&T Loses 1.16M Streaming, Pay TV Subs, HBO Max Hits WarnerMedia Earnings. Retrieved 1 April 2022, from: https://www.hollywoodreporter.com/business/business-news/at-t-loses-streaming-pay-tv-subscribers-warnermedia-earnings-drop-1274002/

TechSee. (2019) Reasons for Customer Churn in the telecom industry: 2019 Survey Results. Retrieved 21 March 2022, from: https://techsee.me/resources/surveys/2019-telecom-churn-survey/

Unscrambl. (24 February 2021) Reducing Customer Churn for Telcos – A Data-Driven Approach Powered by Business Intelligence. Retrieved 21 March 2022, from: https://unscrambl.com/blog/reduce-customer-churn-for-telcos-data-driven-approach/

Utterback C. (27 April 2021). Predict Customer Churn with Precision. Retrieved 21 March 2022, from: https://towardsdatascience.com/predict-customer-churn-with-precision-56932ae0e5e3

Weprin A. (29 March 2022). Streaming subscriber churn "is here to stay", Deloitte Survey Forecasts. Retrieved 1 April 2022, from: https://www.hollywoodreporter.com/business/digital/churn-deloitte-2022-digital-media-trends-1235121006/

YPulse. (13 February 2020). How brands are reaching Netflix Massive Youth Audience without ads. Retrieved 2 April 2022, from: https://www.ypulse.com/article/2020/02/13/how-brands-are-reaching-netflixs-massive-youth-audience-without-ads/

Zix. (4 December 2020). 6 Benefits of Bunding Services. Retrieved 2 April 2022, from: https://www.msspalert.com/cybersecurity-guests/6-benefits-bundling/

11. Appendices

11.1. Appendix A: Data Dictionary

This data dictionary is for the combined dataset used in the report for analysis, which have been cleaned and merged using 2 raw datasets (*"telco_churn_data.csv" & "Telco_customer_churn.csv"*).

Variables	**Data Type**	**Description**
Age	Numerical (Discrete)	Current age of customer
Avg Monthly GB Download	Numerical (Discrete)	Average download volume in gigabytes by the customer
Avg Monthly Long Distance Charges	Numerical (Continuous)	Average charges for long distance of the customer
Churn Category	Categorical (Nominal)	High-level category for the customer's reason for churning ["Attitude", "Competitor", "Dissatisfaction", "Other", "Price"]
Churn Reason	Categorical (Nominal)	Specific reason for the customer leaving the company
Churn	Categorical (Nominal)	Indication of whether the customer churned [0: the customer remained with the company, 1: the customer left the company]
City	Categorical (Nominal)	City of the customer's primary residence
CLTV	Numerical (Discrete)	Predicted Customer Lifetime Value, calculated using corporate formulas
Contract	Categorical (Nominal)	Current contract type of the customer ["Month-to-Month", "One Year", "Two Year"]
Customer ID	Categorical (Nominal)	Unique identifier of the customer
Customer Satisfaction	Categorical (Ordinal)	Overall satisfaction rating by the customer [1: Very Unsatisfied, 2: Unsatisfied, 3: Neutral, 4: Satisfied, 5: Very Satisfied]

Dependents	Categorical (Nominal)	Indication of whether the customer lives with any dependents (E.g. of dependents: children, parents) [“Yes”, “No”]
Device Protection Plan	Categorical (Nominal)	Indicates if the customer subscribes to an additional device protection plan for their Internet equipment provided by the company [“Yes”, “No”]
Gender	Categorical (Nominal)	Sex of customer [“Female”, “Male”]
Internet Service	Categorical (Nominal)	Indication of whether the customer subscribes to Internet service with the company [“Yes”, “No”]
Internet Type	Categorical (Nominal)	Type of Internet service the customer subscribes to with the company [“None”, “Cable”, “DSL”, “Fiber Optic”]
Latitude	Numerical (Continuous)	Latitude of the customer’s primary residence
Longitude	Numerical (Continuous)	Longitude of the customer’s primary residence
Married	Categorical (Nominal)	Indication of whether the customer is married [“Yes”, “No”]
Monthly Charge	Numerical (Continuous)	Current total monthly charge for all their services subscribed to by the customer
Multiple Lines	Categorical (Nominal)	Indication of whether the customer subscribes to multiple telephone lines with the company [“Yes”, “No”]
Number of Dependents	Numerical (Discrete)	Number of dependents that the customer lives with
Number of Referrals	Numerical (Discrete)	Number of referrals made by the customer
Offer	Categorical (Nominal)	Last marketing offer that the customer accepted, if applicable

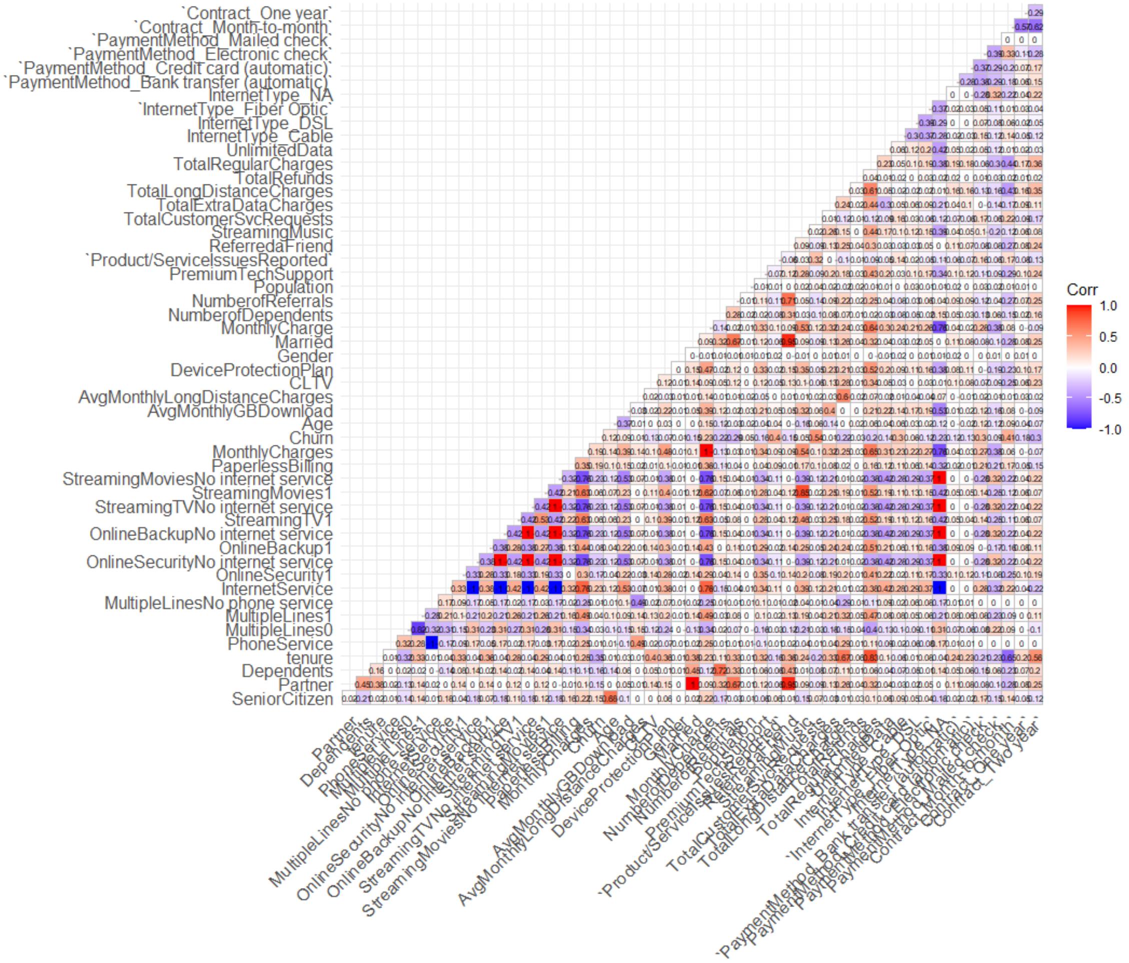
		[“None”, “Offer A”, “Offer B”, “Offer C”, “Offer D”, “Offer E”]
Online Backup	Categorical (Nominal)	Indication of whether the customer subscribes to an additional online backup service provided by the company [“Yes”, “No”]
Online Security	Categorical (Nominal)	Indication of whether the customer subscribes to an additional online security service provided by the company [“Yes”, “No”]
Paperless Billing	Categorical (Nominal)	Indication of whether the customer has chosen paperless billing [“Yes”, “No”]
Payment Method	Categorical (Nominal)	Billing payment mode of the customer [“Bank Withdrawal”, “Credit Card”, “Mailed Check”]
Phone Service	Categorical (Nominal)	Indication of whether the customer subscribes to home phone service with the company [“Yes”, “No”]
Population	Numerical (Discrete)	Current population estimate for the entire Zip Code area
Premium Tech Support	Categorical (Nominal)	Indication of whether the customer subscribes to an additional premium technical support plan provided by the company [“Yes”, “No”]
Product/Service Issues Reported	Numerical (Discrete)	Number of times the customer reported an issue with a product or service
Referred a Friend	Categorical (Nominal)	Indication of whether the customer has ever referred a friend or family member to the company [“Yes”, “No”]

Senior Citizen	Categorical (Nominal)	Indication of whether customer is 65 or older [“Yes”, “No”]
Streaming Movies	Categorical (Nominal)	Indication of whether the customer uses their Internet service to stream movies from a third-party provider [“Yes”, “No”]
Streaming Music	Categorical (Nominal)	Indication of whether the customer uses their Internet service to stream music from a third-party provider [“Yes”, “No”]
Streaming TV	Categorical (Nominal)	Indication of whether the customer uses their Internet service to stream television programming from a third-party provider [“Yes”, “No”]
Tenure in Months	Numerical (Discrete)	Number of months the customer has been with the company
Total Customer Svc Requests	Numerical (Discrete)	Number of times the customer contacted customer service
Total Extra Data Charges	Numerical (Continuous)	Total charges for extra data downloads above those specified in their plan of the customer
Total Long Distance Charges	Numerical (Continuous)	Total charges for long distance above those specified in their plan of the customer
Total Refunds	Numerical (Continuous)	Total refunds of the customer
Total Regular Charges	Numerical (Continuous)	Total regular charges, excluding additional charges of the customer
Under 30	Categorical (Nominal)	Indication of whether customer is under 30 years old [“Yes”, “No”]

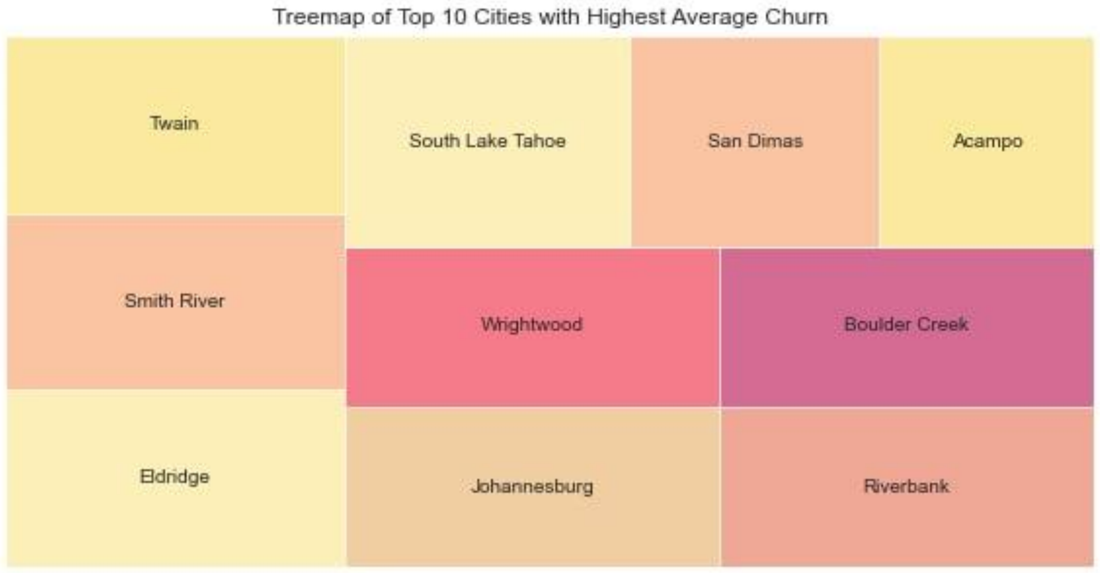
Unlimited Data	Categorical (Nominal)	Indication of whether the customer has paid an additional monthly fee to have unlimited data downloads/uploads [“Yes”, “No”]
Zip Code	Numerical (Discrete)	Zip code of the customer’s primary residence

11.2. Appendix B: Data Exploration Graphs

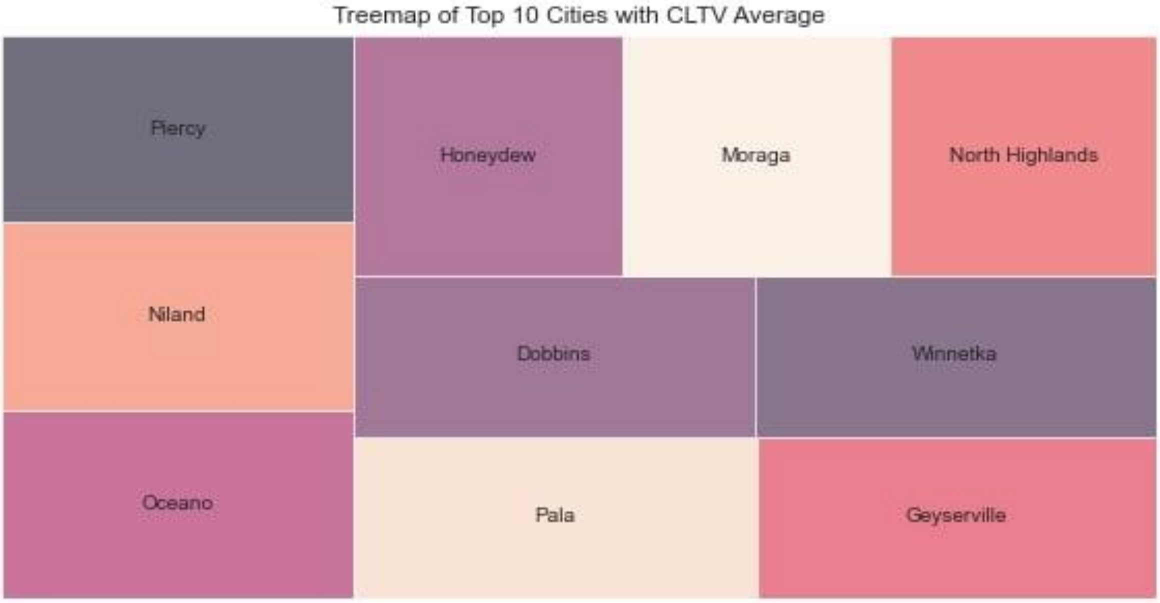
a. Correlation plot of all variables to determine multi-collinearity



b. Treemap of Top 10 cities with Highest average churn



c. Treemap of Top 10 cities with Highest average churn



11.4 Appendix C: Complete list of variables inside the 3 generalized services

Variable version

Internet services	InternetService, InternetType, UnlimitedData, StreamingTV, StreamingMovies, StreamingMusic, OnlineSecurity, OnlineBackup, DeviceProtectionPlan, PremiumTechSupport
Phone services	PhoneService, MultipleLines
Other services	PaperlessBilling, Contract

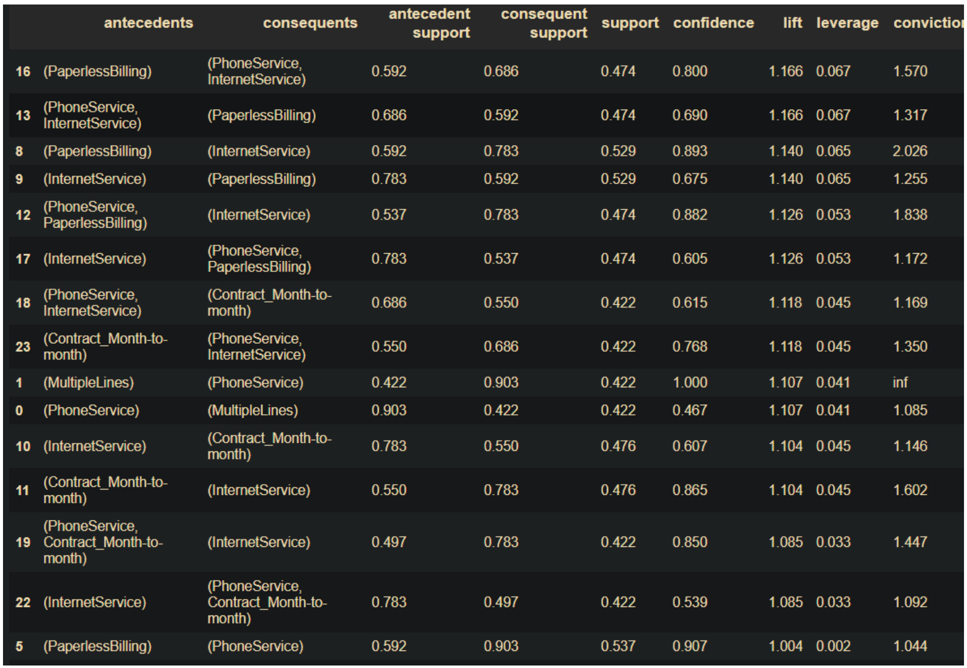
After encoded

Internet services	InternetService, InternetType, UnlimitedData, InternetType_NA, InternetType_DSL, InternetType_Cable, InternetType_Fiber Optic, StreamingTV, StreamingMovies, StreamingMusic, OnlineSecurity, OnlineBackup, DeviceProtectionPlan, PremiumTechSupport
Phone services	PhoneService, MultipleLines
Other services	PaperlessBilling, Contract_Two year, Contract_One year, Contract_Month-to-month

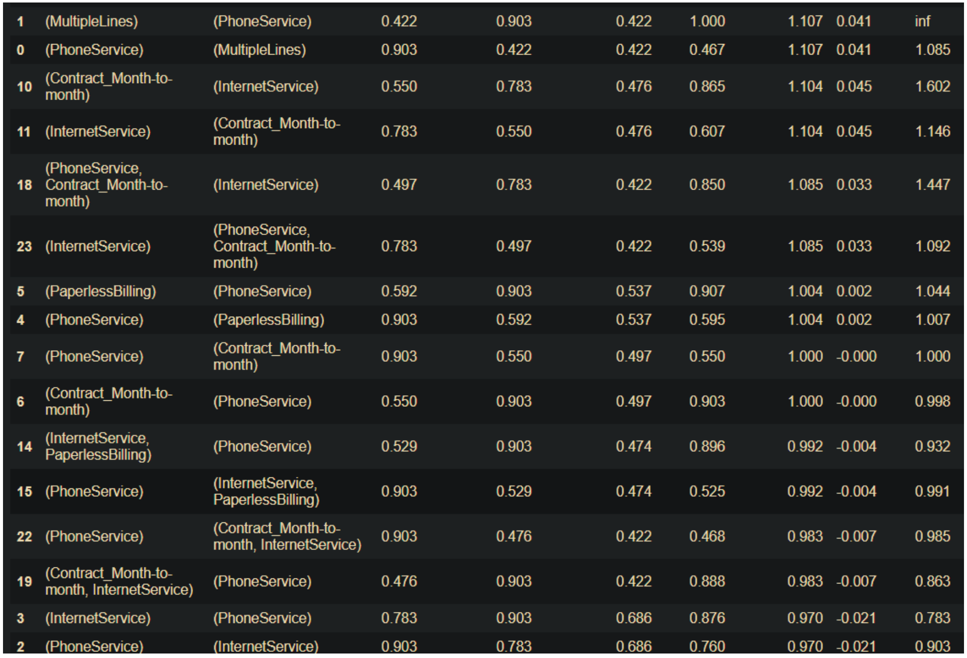
11.6 Appendix D: (Solution 1) Association rules outputs

All the rules are sorted by highest lift

a. Association rules, dataset all, all services

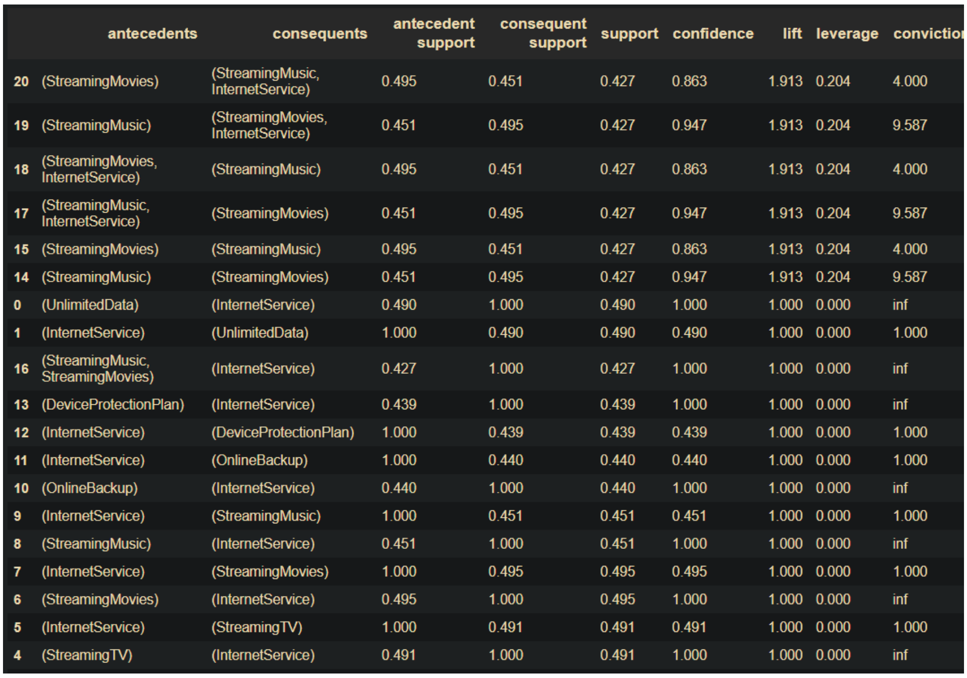


	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	convictio
16	(PaperlessBilling)	(PhoneService, InternetService)	0.592	0.686	0.474	0.800	1.166	0.067	1.570
13	(PhoneService, InternetService)	(PaperlessBilling)	0.686	0.592	0.474	0.690	1.166	0.067	1.317
8	(PaperlessBilling)	(InternetService)	0.592	0.783	0.529	0.893	1.140	0.065	2.026
9	(InternetService)	(PaperlessBilling)	0.783	0.592	0.529	0.675	1.140	0.065	1.255
12	(PhoneService, PaperlessBilling)	(InternetService)	0.537	0.783	0.474	0.882	1.126	0.053	1.838
17	(InternetService)	(PhoneService, PaperlessBilling)	0.783	0.537	0.474	0.605	1.126	0.053	1.172
18	(PhoneService, InternetService)	(Contract_Month-to-month)	0.686	0.550	0.422	0.615	1.118	0.045	1.169
23	(Contract_Month-to-month)	(PhoneService, InternetService)	0.550	0.686	0.422	0.768	1.118	0.045	1.350
1	(MultipleLines)	(PhoneService)	0.422	0.903	0.422	1.000	1.107	0.041	inf
0	(PhoneService)	(MultipleLines)	0.903	0.422	0.422	0.467	1.107	0.041	1.085
10	(InternetService)	(Contract_Month-to-month)	0.783	0.550	0.476	0.607	1.104	0.045	1.146
11	(Contract_Month-to-month)	(InternetService)	0.550	0.783	0.476	0.865	1.104	0.045	1.602
19	(PhoneService, Contract_Month-to-month)	(InternetService)	0.497	0.783	0.422	0.850	1.085	0.033	1.447
22	(InternetService)	(PhoneService, Contract_Month-to-month)	0.783	0.497	0.422	0.539	1.085	0.033	1.092
5	(PaperlessBilling)	(PhoneService)	0.592	0.903	0.537	0.907	1.004	0.002	1.044



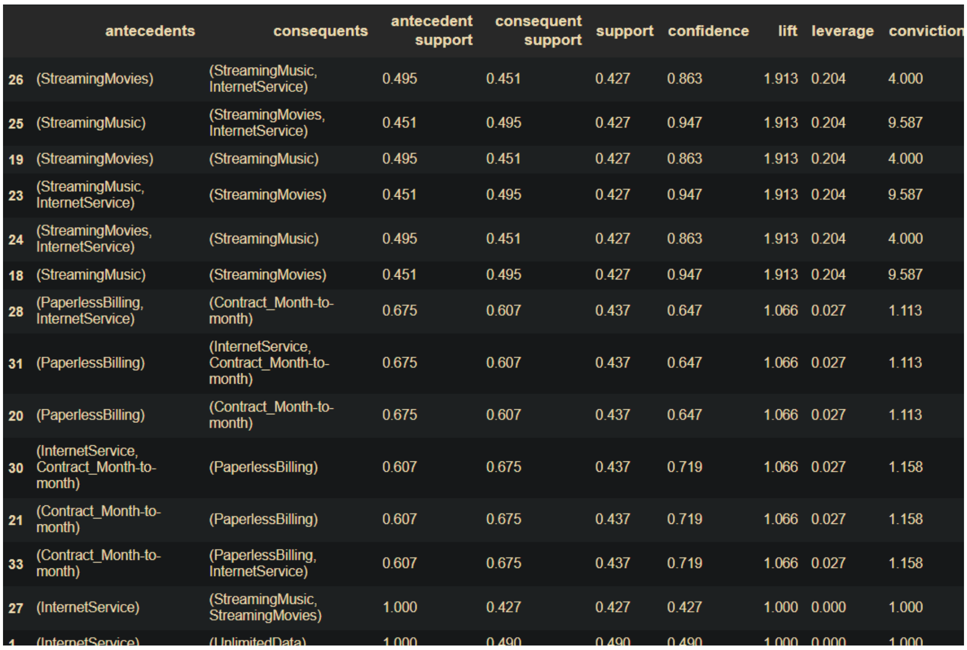
1	(MultipleLines)	(PhoneService)	0.422	0.903	0.422	1.000	1.107	0.041	inf
0	(PhoneService)	(MultipleLines)	0.903	0.422	0.422	0.467	1.107	0.041	1.085
10	(Contract_Month-to-month)	(InternetService)	0.550	0.783	0.476	0.865	1.104	0.045	1.602
11	(InternetService)	(Contract_Month-to-month)	0.783	0.550	0.476	0.607	1.104	0.045	1.146
18	(PhoneService, Contract_Month-to-month)	(InternetService)	0.497	0.783	0.422	0.850	1.085	0.033	1.447
23	(InternetService)	(PhoneService, Contract_Month-to-month)	0.783	0.497	0.422	0.539	1.085	0.033	1.092
5	(PaperlessBilling)	(PhoneService)	0.592	0.903	0.537	0.907	1.004	0.002	1.044
4	(PhoneService)	(PaperlessBilling)	0.903	0.592	0.537	0.595	1.004	0.002	1.007
7	(PhoneService)	(Contract_Month-to-month)	0.903	0.550	0.497	0.550	1.000	-0.000	1.000
6	(Contract_Month-to-month)	(PhoneService)	0.550	0.903	0.497	0.903	1.000	-0.000	0.998
14	(InternetService, PaperlessBilling)	(PhoneService)	0.529	0.903	0.474	0.896	0.992	-0.004	0.932
15	(PhoneService)	(InternetService, PaperlessBilling)	0.903	0.529	0.474	0.525	0.992	-0.004	0.991
22	(PhoneService)	(Contract_Month-to-month, InternetService)	0.903	0.476	0.422	0.468	0.983	-0.007	0.985
19	(Contract_Month-to-month, InternetService)	(PhoneService)	0.476	0.903	0.422	0.888	0.983	-0.007	0.863
3	(InternetService)	(PhoneService)	0.783	0.903	0.686	0.876	0.970	-0.021	0.783
2	(PhoneService)	(InternetService)	0.903	0.783	0.686	0.760	0.970	-0.021	0.903

b. Association rules, dataset all, internet services



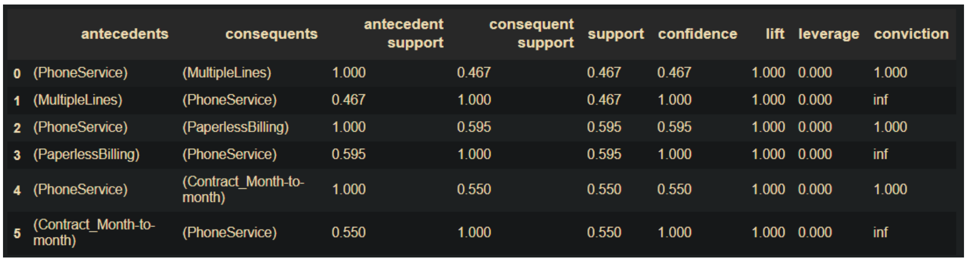
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	convictio
20	(StreamingMovies)	(StreamingMusic, InternetService)	0.495	0.451	0.427	0.863	1.913	0.204	4.000
19	(StreamingMusic)	(StreamingMovies, InternetService)	0.451	0.495	0.427	0.947	1.913	0.204	9.587
18	(StreamingMovies, InternetService)	(StreamingMusic)	0.495	0.451	0.427	0.863	1.913	0.204	4.000
17	(StreamingMusic, InternetService)	(StreamingMovies)	0.451	0.495	0.427	0.947	1.913	0.204	9.587
15	(StreamingMovies)	(StreamingMusic)	0.495	0.451	0.427	0.863	1.913	0.204	4.000
14	(StreamingMusic)	(StreamingMovies)	0.451	0.495	0.427	0.947	1.913	0.204	9.587
0	(UnlimitedData)	(InternetService)	0.490	1.000	0.490	1.000	1.000	0.000	inf
1	(InternetService)	(UnlimitedData)	1.000	0.490	0.490	0.490	1.000	0.000	1.000
16	(StreamingMusic, StreamingMovies)	(InternetService)	0.427	1.000	0.427	1.000	1.000	0.000	inf
13	(DeviceProtectionPlan)	(InternetService)	0.439	1.000	0.439	1.000	1.000	0.000	inf
12	(InternetService)	(DeviceProtectionPlan)	1.000	0.439	0.439	0.439	1.000	0.000	1.000
11	(InternetService)	(OnlineBackup)	1.000	0.440	0.440	0.440	1.000	0.000	1.000
10	(OnlineBackup)	(InternetService)	0.440	1.000	0.440	1.000	1.000	0.000	inf
9	(InternetService)	(StreamingMusic)	1.000	0.451	0.451	0.451	1.000	0.000	1.000
8	(StreamingMusic)	(InternetService)	0.451	1.000	0.451	1.000	1.000	0.000	inf
7	(InternetService)	(StreamingMovies)	1.000	0.495	0.495	0.495	1.000	0.000	1.000
6	(StreamingMovies)	(InternetService)	0.495	1.000	0.495	1.000	1.000	0.000	inf
5	(InternetService)	(StreamingTV)	1.000	0.491	0.491	0.491	1.000	0.000	1.000
4	(StreamingTV)	(InternetService)	0.491	1.000	0.491	1.000	1.000	0.000	inf

c. Association rules, dataset all, internet and other services



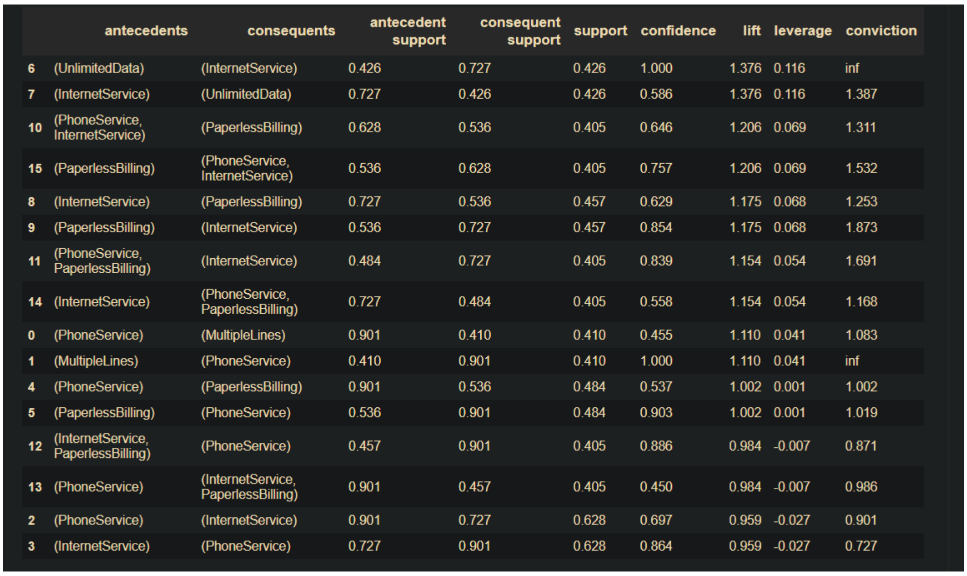
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
26	(StreamingMovies)	(StreamingMusic, InternetService)	0.495	0.451	0.427	0.863	1.913	0.204	4.000
25	(StreamingMusic)	(StreamingMovies, InternetService)	0.451	0.495	0.427	0.947	1.913	0.204	9.587
19	(StreamingMovies)	(StreamingMusic)	0.495	0.451	0.427	0.863	1.913	0.204	4.000
23	(StreamingMusic, InternetService)	(StreamingMovies)	0.451	0.495	0.427	0.947	1.913	0.204	9.587
24	(StreamingMovies, InternetService)	(StreamingMusic)	0.495	0.451	0.427	0.863	1.913	0.204	4.000
18	(StreamingMusic)	(StreamingMovies)	0.451	0.495	0.427	0.947	1.913	0.204	9.587
28	(PaperlessBilling, InternetService)	(Contract_Month-to-month)	0.675	0.607	0.437	0.647	1.066	0.027	1.113
31	(PaperlessBilling)	(InternetService, Contract_Month-to-month)	0.675	0.607	0.437	0.647	1.066	0.027	1.113
20	(PaperlessBilling)	(Contract_Month-to-month)	0.675	0.607	0.437	0.647	1.066	0.027	1.113
30	(InternetService, Contract_Month-to-month)	(PaperlessBilling)	0.607	0.675	0.437	0.719	1.066	0.027	1.158
21	(Contract_Month-to-month)	(PaperlessBilling)	0.607	0.675	0.437	0.719	1.066	0.027	1.158
33	(Contract_Month-to-month)	(PaperlessBilling, InternetService)	0.607	0.675	0.437	0.719	1.066	0.027	1.158
27	(InternetService)	(StreamingMusic, StreamingMovies)	1.000	0.427	0.427	0.427	1.000	0.000	1.000
1	(InternetService)	(UnlimitedData)	1.000	0.490	0.490	0.490	1.000	0.000	1.000

d. Association rules, dataset all, phone and other services



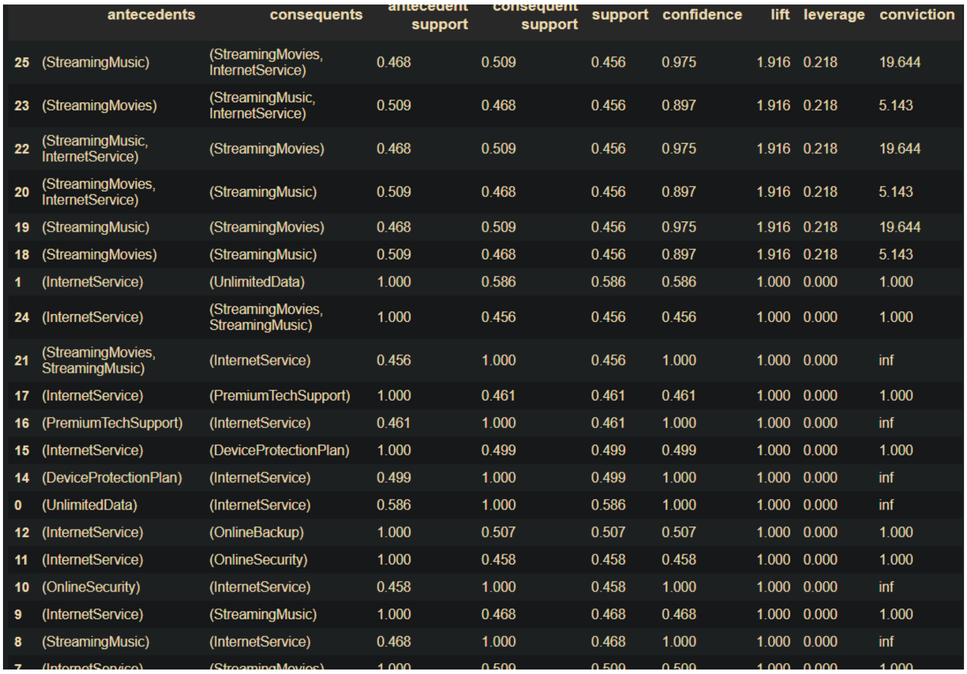
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(PhoneService)	(MultipleLines)	1.000	0.467	0.467	0.467	1.000	0.000	1.000
1	(MultipleLines)	(PhoneService)	0.467	1.000	0.467	1.000	1.000	0.000	inf
2	(PhoneService)	(PaperlessBilling)	1.000	0.595	0.595	0.595	1.000	0.000	1.000
3	(PaperlessBilling)	(PhoneService)	0.595	1.000	0.595	1.000	1.000	0.000	inf
4	(PhoneService)	(Contract_Month-to-month)	1.000	0.550	0.550	0.550	1.000	0.000	1.000
5	(Contract_Month-to-month)	(PhoneService)	0.550	1.000	0.550	1.000	1.000	0.000	inf

e. Association rules, dataset where churn == 0, all services



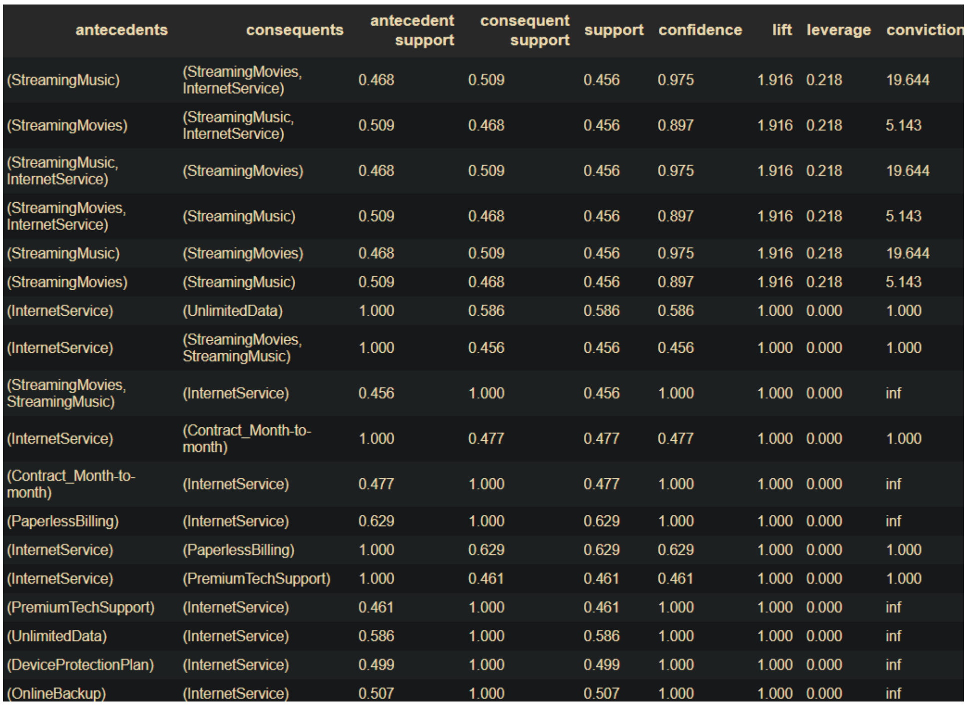
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
6	(UnlimitedData)	(InternetService)	0.426	0.727	0.426	1.000	1.376	0.116	inf
7	(InternetService)	(UnlimitedData)	0.727	0.426	0.426	0.586	1.376	0.116	1.387
10	(PhoneService, InternetService)	(PaperlessBilling)	0.628	0.536	0.405	0.646	1.206	0.069	1.311
15	(PaperlessBilling)	(PhoneService, InternetService)	0.536	0.628	0.405	0.757	1.206	0.069	1.532
8	(InternetService)	(PaperlessBilling)	0.727	0.536	0.457	0.629	1.175	0.068	1.253
9	(PaperlessBilling)	(InternetService)	0.536	0.727	0.457	0.854	1.175	0.068	1.873
11	(PhoneService, PaperlessBilling)	(InternetService)	0.484	0.727	0.405	0.839	1.154	0.054	1.691
14	(InternetService)	(PhoneService, PaperlessBilling)	0.727	0.484	0.405	0.558	1.154	0.054	1.168
0	(PhoneService)	(MultipleLines)	0.901	0.410	0.410	0.455	1.110	0.041	1.083
1	(MultipleLines)	(PhoneService)	0.410	0.901	0.410	1.000	1.110	0.041	inf
4	(PhoneService)	(PaperlessBilling)	0.901	0.536	0.484	0.537	1.002	0.001	1.002
5	(PaperlessBilling)	(PhoneService)	0.536	0.901	0.484	0.903	1.002	0.001	1.019
12	(InternetService, PaperlessBilling)	(PhoneService)	0.457	0.901	0.405	0.886	0.984	-0.007	0.871
13	(PhoneService)	(InternetService, PaperlessBilling)	0.901	0.457	0.405	0.450	0.984	-0.007	0.986
2	(PhoneService)	(InternetService)	0.901	0.727	0.628	0.697	0.959	-0.027	0.901
3	(InternetService)	(PhoneService)	0.727	0.901	0.628	0.864	0.959	-0.027	0.727

f. Association rules, dataset where churn == 0, internet services



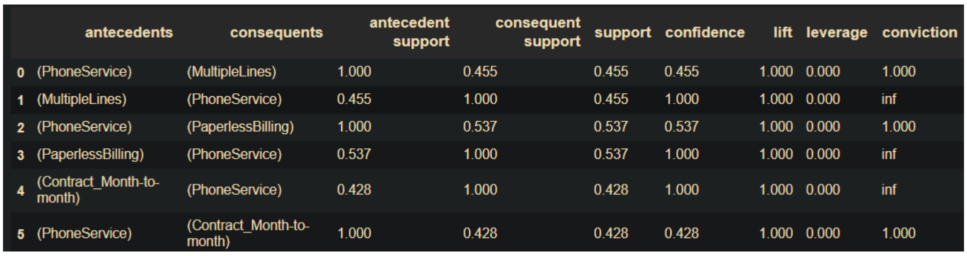
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
25	(StreamingMusic)	(StreamingMovies, InternetService)	0.468	0.509	0.456	0.975	1.916	0.218	19.644
23	(StreamingMovies)	(StreamingMusic, InternetService)	0.509	0.468	0.456	0.897	1.916	0.218	5.143
22	(StreamingMusic, InternetService)	(StreamingMovies)	0.468	0.509	0.456	0.975	1.916	0.218	19.644
20	(StreamingMovies, InternetService)	(StreamingMusic)	0.509	0.468	0.456	0.897	1.916	0.218	5.143
19	(StreamingMusic)	(StreamingMovies)	0.468	0.509	0.456	0.975	1.916	0.218	19.644
18	(StreamingMovies)	(StreamingMusic)	0.509	0.468	0.456	0.897	1.916	0.218	5.143
1	(InternetService)	(UnlimitedData)	1.000	0.586	0.586	0.586	1.000	0.000	1.000
24	(InternetService)	(StreamingMovies, StreamingMusic)	1.000	0.456	0.456	0.456	1.000	0.000	1.000
21	(StreamingMovies, StreamingMusic)	(InternetService)	0.456	1.000	0.456	1.000	1.000	0.000	inf
17	(InternetService)	(PremiumTechSupport)	1.000	0.461	0.461	0.461	1.000	0.000	1.000
16	(PremiumTechSupport)	(InternetService)	0.461	1.000	0.461	1.000	1.000	0.000	inf
15	(InternetService)	(DeviceProtectionPlan)	1.000	0.499	0.499	0.499	1.000	0.000	1.000
14	(DeviceProtectionPlan)	(InternetService)	0.499	1.000	0.499	1.000	1.000	0.000	inf
0	(UnlimitedData)	(InternetService)	0.586	1.000	0.586	1.000	1.000	0.000	inf
12	(InternetService)	(OnlineBackup)	1.000	0.507	0.507	0.507	1.000	0.000	1.000
11	(InternetService)	(OnlineSecurity)	1.000	0.458	0.458	0.458	1.000	0.000	1.000
10	(OnlineSecurity)	(InternetService)	0.458	1.000	0.458	1.000	1.000	0.000	inf
9	(InternetService)	(StreamingMusic)	1.000	0.468	0.468	0.468	1.000	0.000	1.000
8	(StreamingMusic)	(InternetService)	0.468	1.000	0.468	1.000	1.000	0.000	inf
7	(InternetService)	(StreamingMovies)	1.000	0.509	0.509	0.509	1.000	0.000	1.000

g. Association rules, dataset where churn == 0, internet and other services



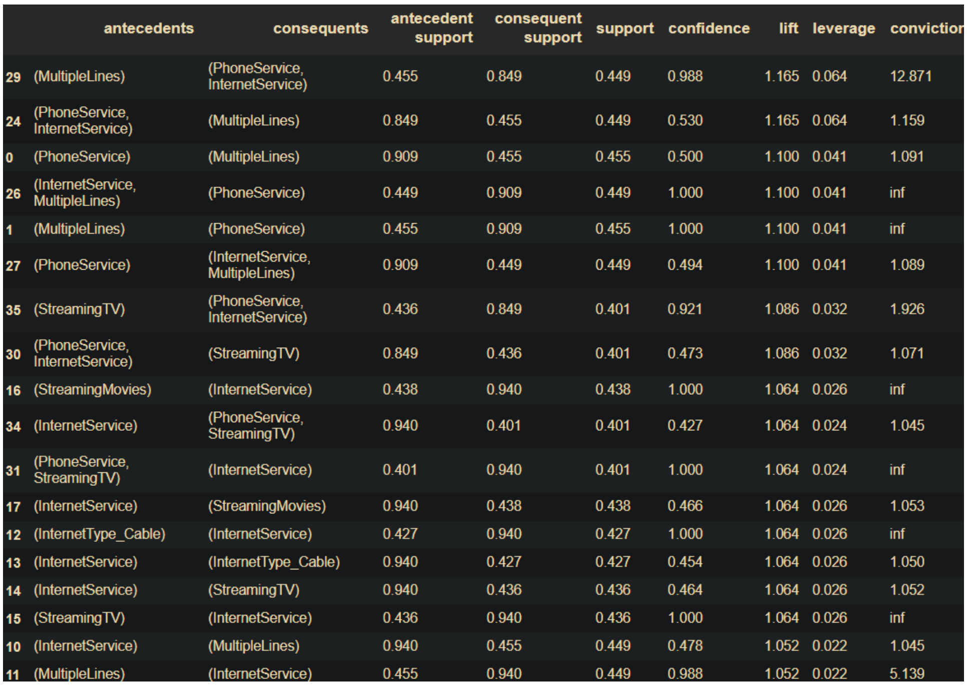
antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
(StreamingMusic)	(StreamingMovies, InternetService)	0.468	0.509	0.456	0.975	1.916	0.218	19.644
(StreamingMovies)	(StreamingMusic, InternetService)	0.509	0.468	0.456	0.897	1.916	0.218	5.143
(StreamingMusic, InternetService)	(StreamingMovies)	0.468	0.509	0.456	0.975	1.916	0.218	19.644
(StreamingMovies, InternetService)	(StreamingMusic)	0.509	0.468	0.456	0.897	1.916	0.218	5.143
(StreamingMusic)	(StreamingMovies)	0.468	0.509	0.456	0.975	1.916	0.218	19.644
(StreamingMovies)	(StreamingMusic)	0.509	0.468	0.456	0.897	1.916	0.218	5.143
(InternetService)	(UnlimitedData)	1.000	0.586	0.586	0.586	1.000	0.000	1.000
(InternetService)	(StreamingMovies, StreamingMusic)	1.000	0.456	0.456	0.456	1.000	0.000	1.000
(StreamingMovies, StreamingMusic)	(InternetService)	0.456	1.000	0.456	1.000	1.000	0.000	inf
(InternetService)	(Contract_Month-to-month)	1.000	0.477	0.477	0.477	1.000	0.000	1.000
(Contract_Month-to-month)	(InternetService)	0.477	1.000	0.477	1.000	1.000	0.000	inf
(PaperlessBilling)	(InternetService)	0.629	1.000	0.629	1.000	1.000	0.000	inf
(InternetService)	(PaperlessBilling)	1.000	0.629	0.629	0.629	1.000	0.000	1.000
(InternetService)	(PremiumTechSupport)	1.000	0.461	0.461	0.461	1.000	0.000	1.000
(PremiumTechSupport)	(InternetService)	0.461	1.000	0.461	1.000	1.000	0.000	inf
(UnlimitedData)	(InternetService)	0.586	1.000	0.586	1.000	1.000	0.000	inf
(DeviceProtectionPlan)	(InternetService)	0.499	1.000	0.499	1.000	1.000	0.000	inf
(OnlineBackup)	(InternetService)	0.507	1.000	0.507	1.000	1.000	0.000	inf

h. Association rules, dataset where churn == 0, phone and other services

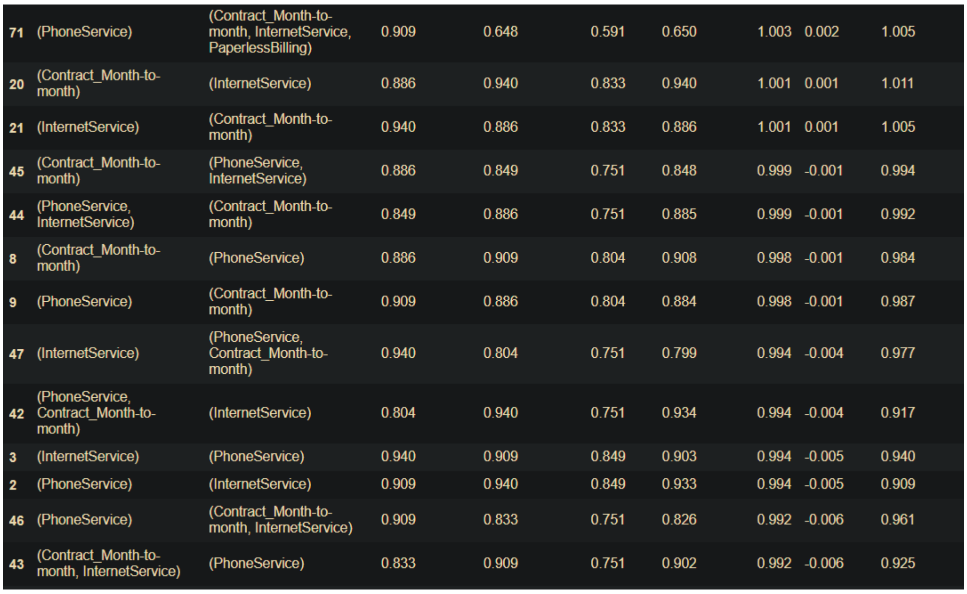


	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(PhoneService)	(MultipleLines)	1.000	0.455	0.455	0.455	1.000	0.000	1.000
1	(MultipleLines)	(PhoneService)	0.455	1.000	0.455	1.000	1.000	0.000	inf
2	(PhoneService)	(PaperlessBilling)	1.000	0.537	0.537	0.537	1.000	0.000	1.000
3	(PaperlessBilling)	(PhoneService)	0.537	1.000	0.537	1.000	1.000	0.000	inf
4	(Contract_Month-to-month)	(PhoneService)	0.428	1.000	0.428	1.000	1.000	0.000	inf
5	(PhoneService)	(Contract_Month-to-month)	1.000	0.428	0.428	0.428	1.000	0.000	1.000

i. Association rules, dataset where churn == 1, all services

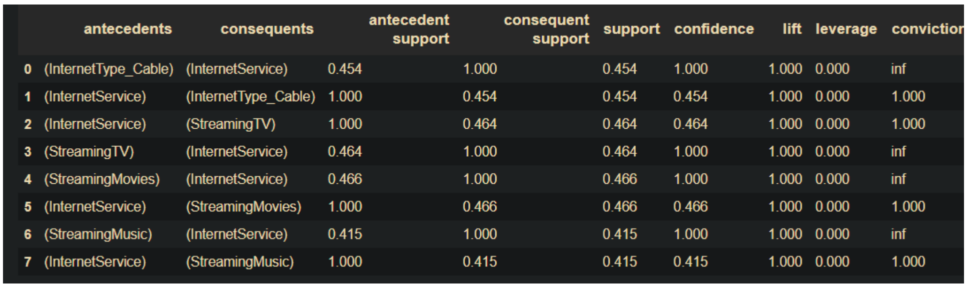


	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
29	(MultipleLines)	(PhoneService, InternetService)	0.455	0.849	0.449	0.988	1.165	0.064	12.871
24	(PhoneService, InternetService)	(MultipleLines)	0.849	0.455	0.449	0.530	1.165	0.064	1.159
0	(PhoneService)	(MultipleLines)	0.909	0.455	0.455	0.500	1.100	0.041	1.091
26	(InternetService, MultipleLines)	(PhoneService)	0.449	0.909	0.449	1.000	1.100	0.041	inf
1	(MultipleLines)	(PhoneService)	0.455	0.909	0.455	1.000	1.100	0.041	inf
27	(PhoneService)	(InternetService, MultipleLines)	0.909	0.449	0.449	0.494	1.100	0.041	1.089
35	(StreamingTV)	(PhoneService, InternetService)	0.436	0.849	0.401	0.921	1.086	0.032	1.926
30	(PhoneService, InternetService)	(StreamingTV)	0.849	0.436	0.401	0.473	1.086	0.032	1.071
16	(StreamingMovies)	(InternetService)	0.438	0.940	0.438	1.000	1.064	0.026	inf
34	(InternetService)	(PhoneService, StreamingTV)	0.940	0.401	0.401	0.427	1.064	0.024	1.045
31	(PhoneService, StreamingTV)	(InternetService)	0.401	0.940	0.401	1.000	1.064	0.024	inf
17	(InternetService)	(StreamingMovies)	0.940	0.438	0.438	0.466	1.064	0.026	1.053
12	(InternetType_Cable)	(InternetService)	0.427	0.940	0.427	1.000	1.064	0.026	inf
13	(InternetService)	(InternetType_Cable)	0.940	0.427	0.427	0.454	1.064	0.026	1.050
14	(InternetService)	(StreamingTV)	0.940	0.436	0.436	0.464	1.064	0.026	1.052
15	(StreamingTV)	(InternetService)	0.436	0.940	0.436	1.000	1.064	0.026	inf
10	(InternetService)	(MultipleLines)	0.940	0.455	0.449	0.478	1.052	0.022	1.045
11	(MultipleLines)	(InternetService)	0.455	0.940	0.449	0.988	1.052	0.022	5.139



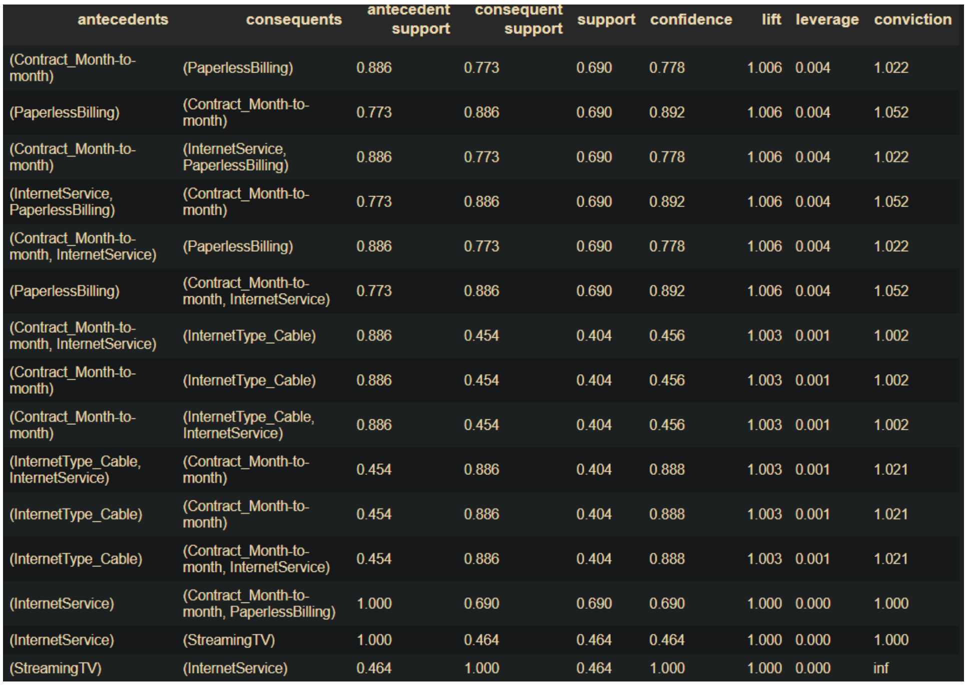
71	(PhoneService)	(Contract_Month-to-month, InternetService, PaperlessBilling)	0.909	0.648	0.591	0.650	1.003	0.002	1.005
20	(Contract_Month-to-month)	(InternetService)	0.886	0.940	0.833	0.940	1.001	0.001	1.011
21	(InternetService)	(Contract_Month-to-month)	0.940	0.886	0.833	0.886	1.001	0.001	1.005
45	(Contract_Month-to-month)	(PhoneService, InternetService)	0.886	0.849	0.751	0.848	0.999	-0.001	0.994
44	(PhoneService, InternetService)	(Contract_Month-to-month)	0.849	0.886	0.751	0.885	0.999	-0.001	0.992
8	(Contract_Month-to-month)	(PhoneService)	0.886	0.909	0.804	0.908	0.998	-0.001	0.984
9	(PhoneService)	(Contract_Month-to-month)	0.909	0.886	0.804	0.884	0.998	-0.001	0.987
47	(InternetService)	(PhoneService, Contract_Month-to-month)	0.940	0.804	0.751	0.799	0.994	-0.004	0.977
42	(PhoneService, Contract_Month-to-month)	(InternetService)	0.804	0.940	0.751	0.934	0.994	-0.004	0.917
3	(InternetService)	(PhoneService)	0.940	0.909	0.849	0.903	0.994	-0.005	0.940
2	(PhoneService)	(InternetService)	0.909	0.940	0.849	0.933	0.994	-0.005	0.909
46	(PhoneService)	(Contract_Month-to-month, InternetService)	0.909	0.833	0.751	0.826	0.992	-0.006	0.961
43	(Contract_Month-to-month, InternetService)	(PhoneService)	0.833	0.909	0.751	0.902	0.992	-0.006	0.925

j. Association rules, dataset where churn == 1, internet services



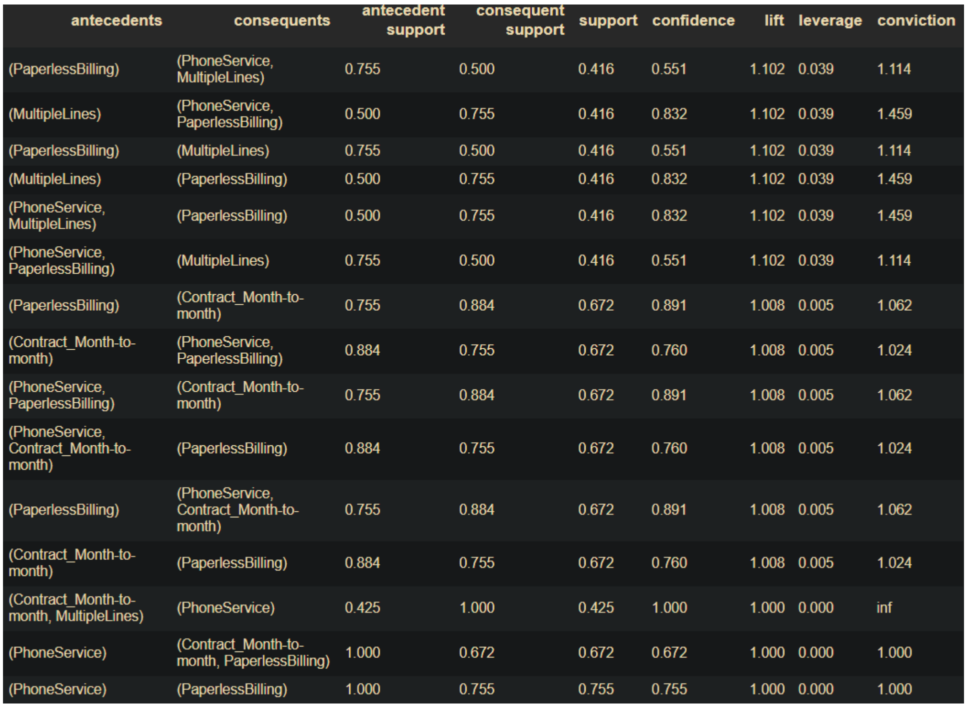
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(InternetType_Cable)	(InternetService)	0.454	1.000	0.454	1.000	1.000	0.000	inf
1	(InternetService)	(InternetType_Cable)	1.000	0.454	0.454	0.454	1.000	0.000	1.000
2	(InternetService)	(StreamingTV)	1.000	0.464	0.464	0.464	1.000	0.000	1.000
3	(StreamingTV)	(InternetService)	0.464	1.000	0.464	1.000	1.000	0.000	inf
4	(StreamingMovies)	(InternetService)	0.466	1.000	0.466	1.000	1.000	0.000	inf
5	(InternetService)	(StreamingMovies)	1.000	0.466	0.466	0.466	1.000	0.000	1.000
6	(StreamingMusic)	(InternetService)	0.415	1.000	0.415	1.000	1.000	0.000	inf
7	(InternetService)	(StreamingMusic)	1.000	0.415	0.415	0.415	1.000	0.000	1.000

k. Association rules, dataset where churn == 1, internet and other services



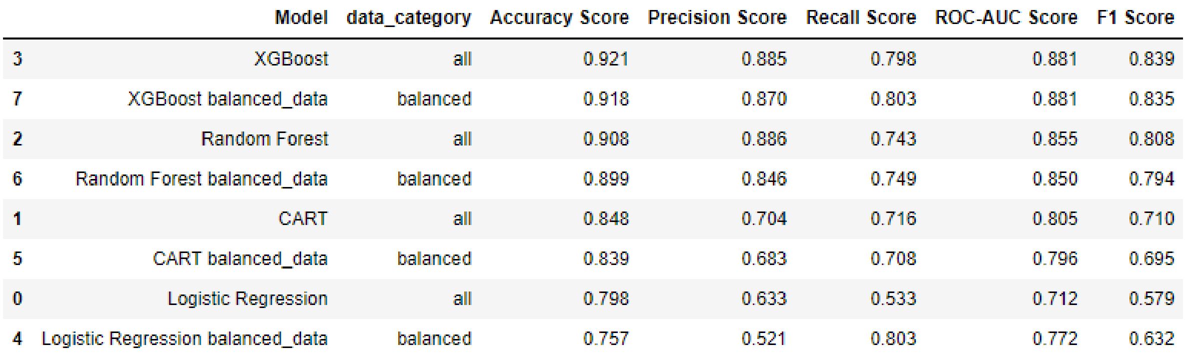
antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
(Contract_Month-to-month)	(PaperlessBilling)	0.886	0.773	0.690	0.778	1.006	0.004	1.022
(PaperlessBilling)	(Contract_Month-to-month)	0.773	0.886	0.690	0.892	1.006	0.004	1.052
(Contract_Month-to-month)	(InternetService, PaperlessBilling)	0.886	0.773	0.690	0.778	1.006	0.004	1.022
(InternetService, PaperlessBilling)	(Contract_Month-to-month)	0.773	0.886	0.690	0.892	1.006	0.004	1.052
(Contract_Month-to-month, InternetService)	(PaperlessBilling)	0.886	0.773	0.690	0.778	1.006	0.004	1.022
(PaperlessBilling)	(Contract_Month-to-month, InternetService)	0.773	0.886	0.690	0.892	1.006	0.004	1.052
(Contract_Month-to-month, InternetService)	(InternetType_Cable)	0.886	0.454	0.404	0.456	1.003	0.001	1.002
(Contract_Month-to-month)	(InternetType_Cable)	0.886	0.454	0.404	0.456	1.003	0.001	1.002
(Contract_Month-to-month)	(InternetType_Cable, InternetService)	0.886	0.454	0.404	0.456	1.003	0.001	1.002
(InternetType_Cable, InternetService)	(Contract_Month-to-month)	0.454	0.886	0.404	0.888	1.003	0.001	1.021
(InternetType_Cable)	(Contract_Month-to-month)	0.454	0.886	0.404	0.888	1.003	0.001	1.021
(InternetType_Cable)	(Contract_Month-to-month, InternetService)	0.454	0.886	0.404	0.888	1.003	0.001	1.021
(InternetService)	(Contract_Month-to-month, PaperlessBilling)	1.000	0.690	0.690	0.690	1.000	0.000	1.000
(InternetService)	(StreamingTV)	1.000	0.464	0.464	0.464	1.000	0.000	1.000
(StreamingTV)	(InternetService)	0.464	1.000	0.464	1.000	1.000	0.000	inf

l. Association rules, dataset where churn == 1, phone and other services

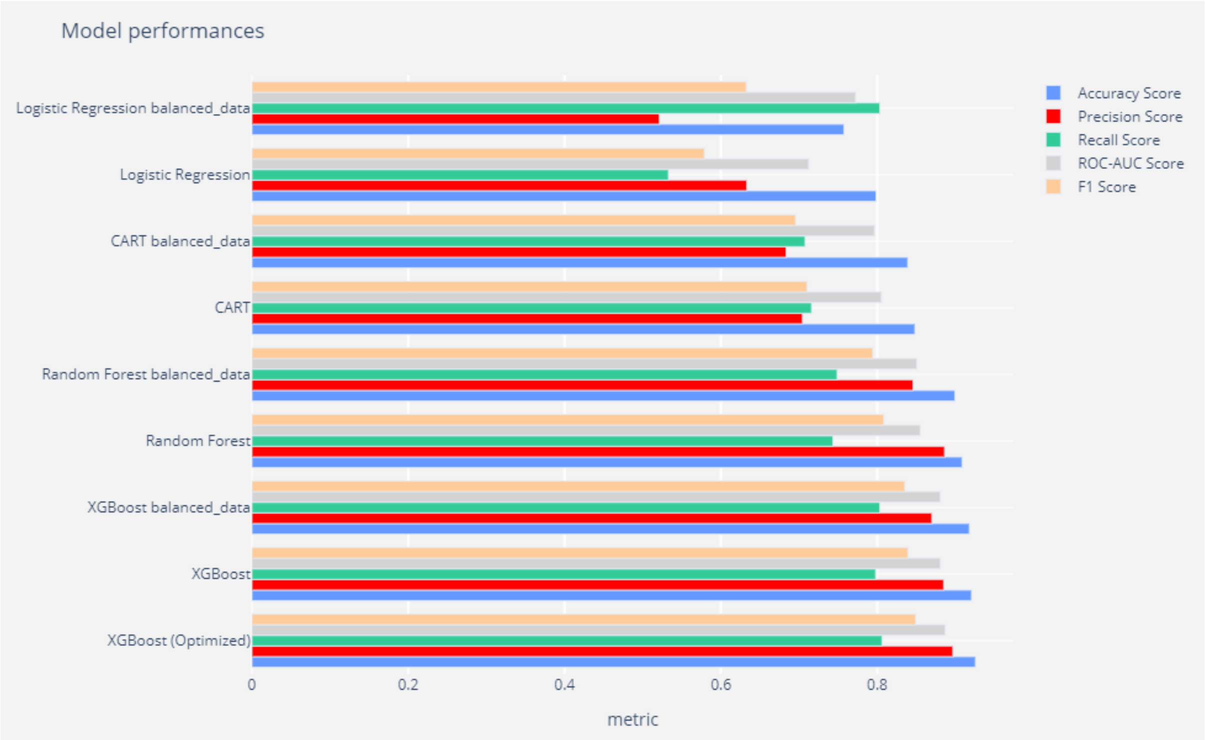


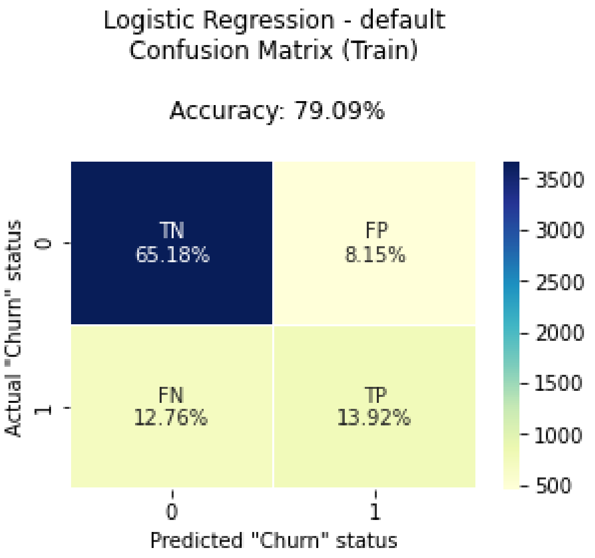
antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
(PaperlessBilling)	(PhoneService, MultipleLines)	0.755	0.500	0.416	0.551	1.102	0.039	1.114
(MultipleLines)	(PhoneService, PaperlessBilling)	0.500	0.755	0.416	0.832	1.102	0.039	1.459
(PaperlessBilling)	(MultipleLines)	0.755	0.500	0.416	0.551	1.102	0.039	1.114
(MultipleLines)	(PaperlessBilling)	0.500	0.755	0.416	0.832	1.102	0.039	1.459
(PhoneService, MultipleLines)	(PaperlessBilling)	0.500	0.755	0.416	0.832	1.102	0.039	1.459
(PhoneService, PaperlessBilling)	(MultipleLines)	0.755	0.500	0.416	0.551	1.102	0.039	1.114
(PaperlessBilling)	(Contract_Month-to-month)	0.755	0.884	0.672	0.891	1.008	0.005	1.062
(Contract_Month-to-month)	(PhoneService, PaperlessBilling)	0.884	0.755	0.672	0.760	1.008	0.005	1.024
(PhoneService, PaperlessBilling)	(Contract_Month-to-month)	0.755	0.884	0.672	0.891	1.008	0.005	1.062
(PhoneService, Contract_Month-to-month)	(PaperlessBilling)	0.884	0.755	0.672	0.760	1.008	0.005	1.024
(PaperlessBilling)	(PhoneService, Contract_Month-to-month)	0.755	0.884	0.672	0.891	1.008	0.005	1.062
(Contract_Month-to-month)	(PaperlessBilling)	0.884	0.755	0.672	0.760	1.008	0.005	1.024
(Contract_Month-to-month, MultipleLines)	(PhoneService)	0.425	1.000	0.425	1.000	1.000	0.000	inf
(PhoneService)	(Contract_Month-to-month, PaperlessBilling)	1.000	0.672	0.672	0.672	1.000	0.000	1.000
(PhoneService)	(PaperlessBilling)	1.000	0.755	0.755	0.755	1.000	0.000	1.000

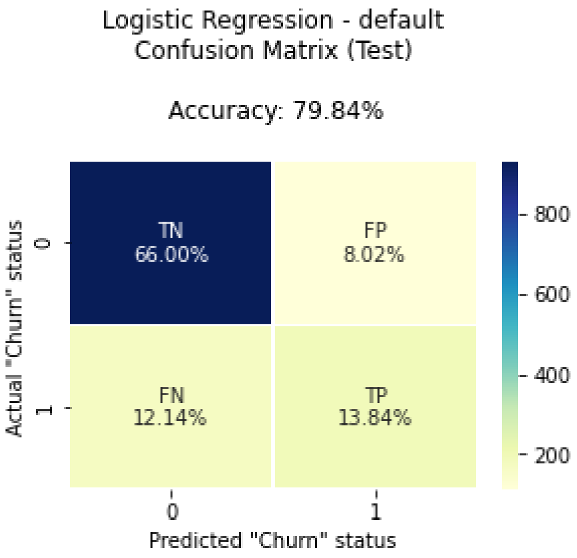
11.5 Appendix E: (Solution 2) All Models Variations Performance Output Comparison

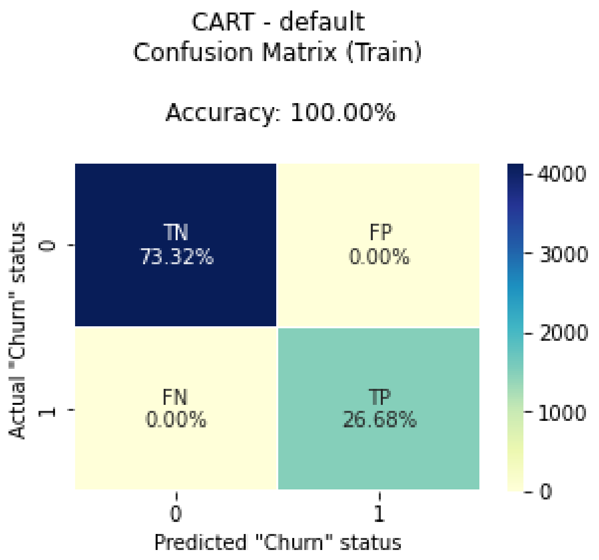


	Model	data_category	Accuracy Score	Precision Score	Recall Score	ROC-AUC Score	F1 Score
3	XGBoost	all	0.921	0.885	0.798	0.881	0.839
7	XGBoost balanced_data	balanced	0.918	0.870	0.803	0.881	0.835
2	Random Forest	all	0.908	0.886	0.743	0.855	0.808
6	Random Forest balanced_data	balanced	0.899	0.846	0.749	0.850	0.794
1	CART	all	0.848	0.704	0.716	0.805	0.710
5	CART balanced_data	balanced	0.839	0.683	0.708	0.796	0.695
0	Logistic Regression	all	0.798	0.633	0.533	0.712	0.579
4	Logistic Regression balanced_data	balanced	0.757	0.521	0.803	0.772	0.632



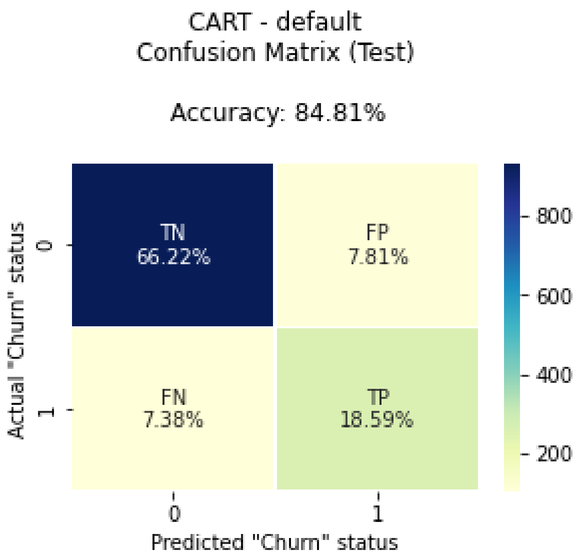






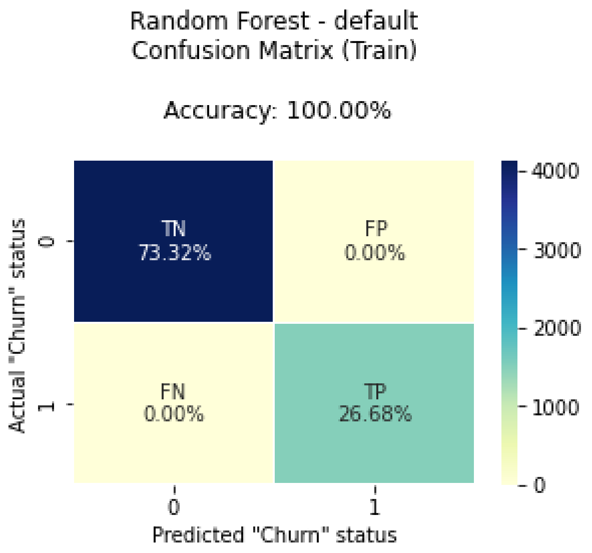
CART - default
Confusion Matrix (Train)

Accuracy: 100.00%



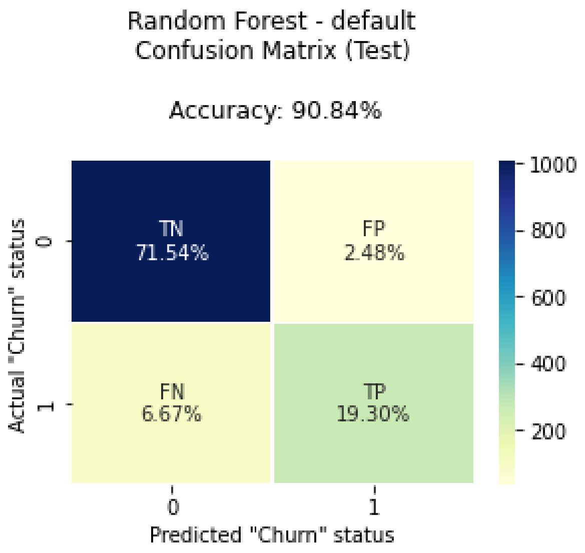
CART - default
Confusion Matrix (Test)

Accuracy: 84.81%



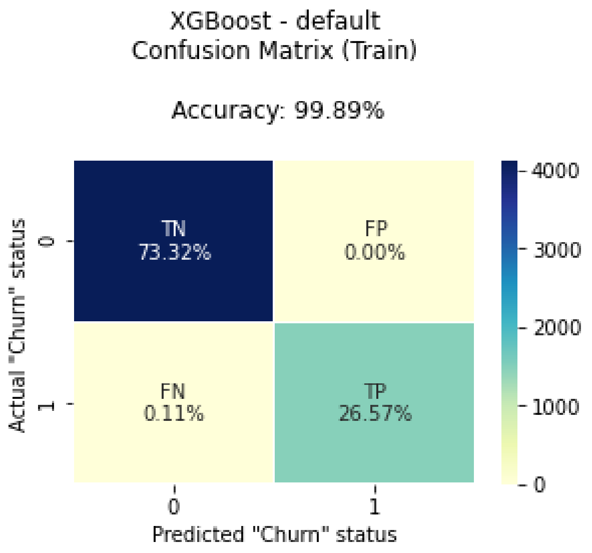
Random Forest - default
Confusion Matrix (Train)

Accuracy: 100.00%



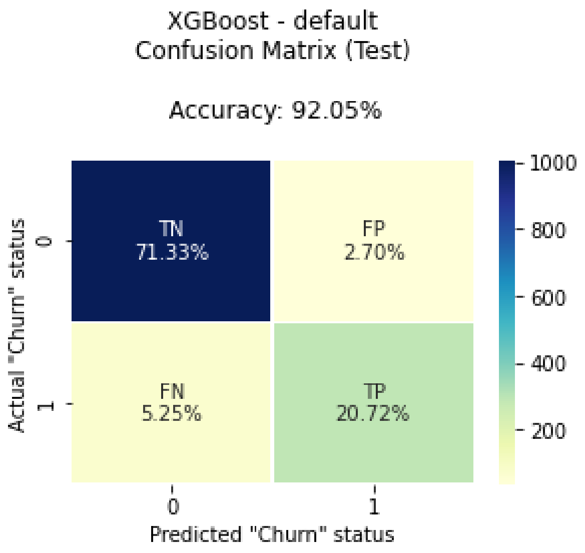
Random Forest - default
Confusion Matrix (Test)

Accuracy: 90.84%



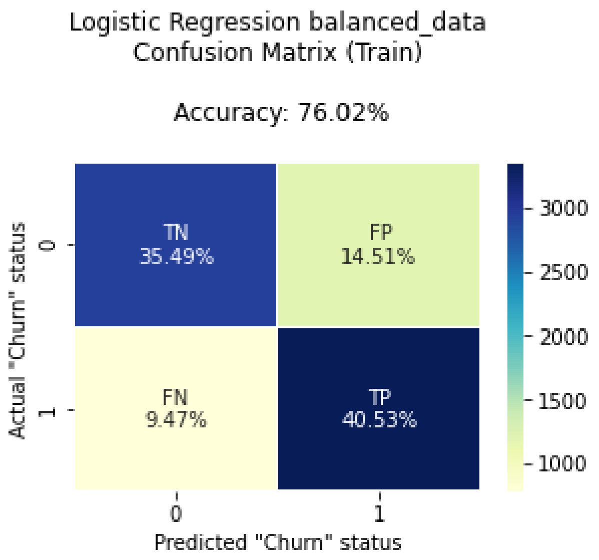
XGBoost - default
Confusion Matrix (Train)

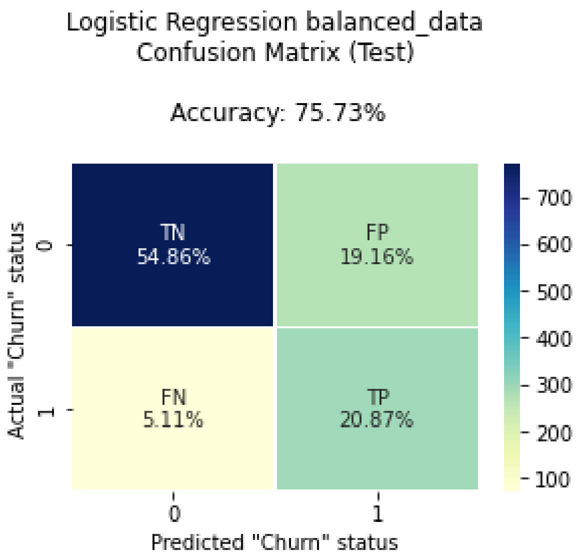
Accuracy: 99.89%

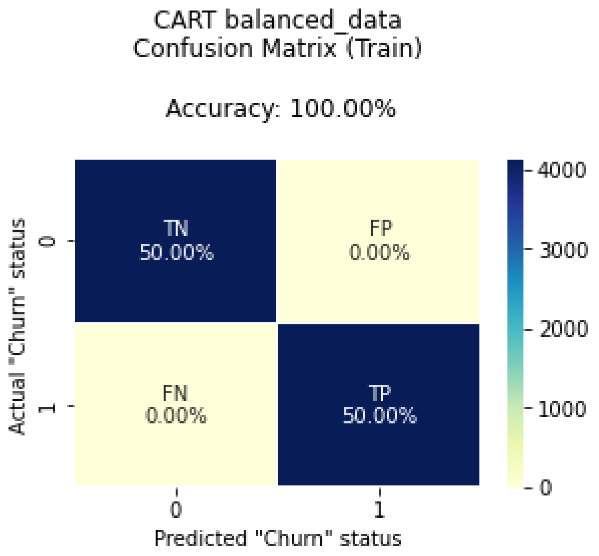


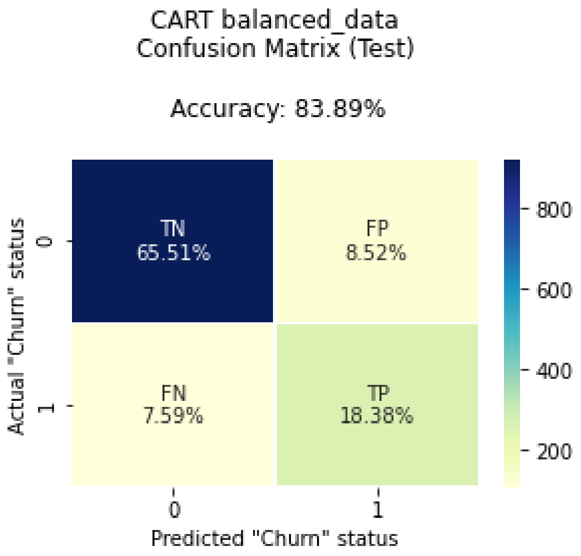
XGBoost - default
Confusion Matrix (Test)

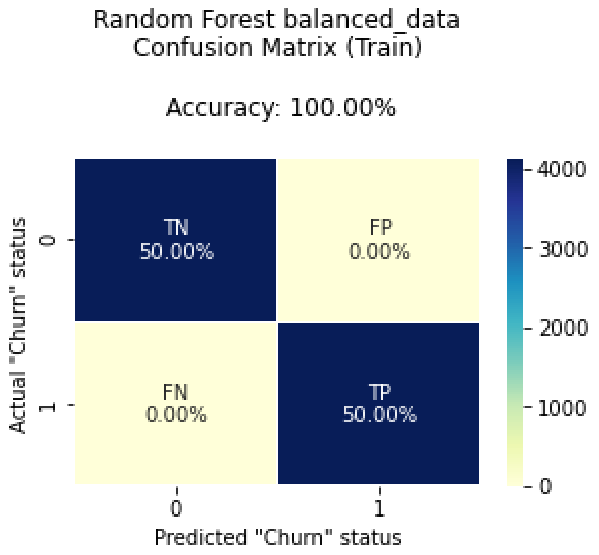
Accuracy: 92.05%

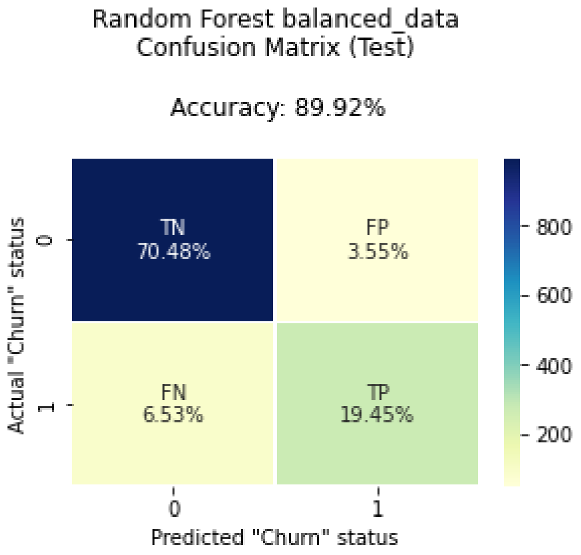

Logistic Regression balanced_data
Confusion Matrix (Train)
Accuracy: 76.02%
TN
35.49%
FP
14.51%
FN
9.47%
TP
40.53%
Actual "Churn" status
0
1
Predicted "Churn" status
0
1
3000
2500
2000
1500
1000

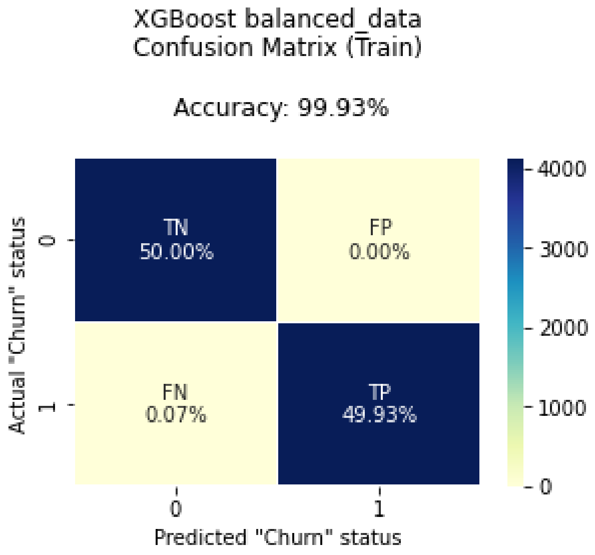

Logistic Regression balanced_data
Confusion Matrix (Test)
Accuracy: 75.73%
TN
54.86%
FP
19.16%
FN
5.11%
TP
20.87%
Actual "Churn" status
0
1
Predicted "Churn" status
0
1
700
600
500
400
300
200
100

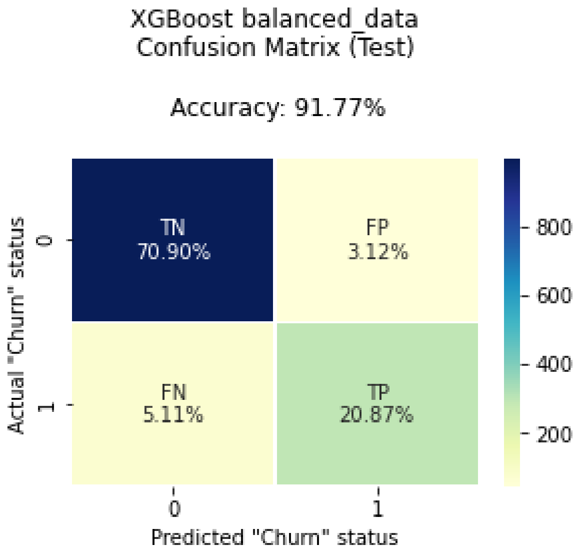

CART balanced_data
Confusion Matrix (Train)
Accuracy: 100.00%
TN
50.00%
FP
0.00%
FN
0.00%
TP
50.00%
Actual "Churn" status
0
1
Predicted "Churn" status
0
1
4000
3000
2000
1000
0


CART balanced_data
Confusion Matrix (Test)
Accuracy: 83.89%
TN
65.51%
FP
8.52%
FN
7.59%
TP
18.38%
Actual "Churn" status
0
1
Predicted "Churn" status
0
1
800
600
400
200

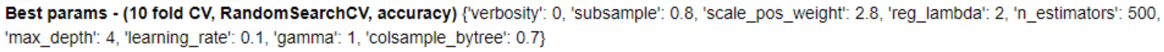

Random Forest balanced_data
Confusion Matrix (Train)
Accuracy: 100.00%
TN
50.00%
FP
0.00%
FN
0.00%
TP
50.00%
Actual "Churn" status
0
1
Predicted "Churn" status
0
1
4000
3000
2000
1000
0


Random Forest balanced_data
Confusion Matrix (Test)
Accuracy: 89.92%
TN
70.48%
FP
3.55%
FN
6.53%
TP
19.45%
Actual "Churn" status
0
1
Predicted "Churn" status
0
1
800
600
400
200

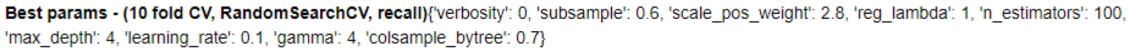

XGBoost balanced_data
Confusion Matrix (Train)
Accuracy: 99.93%
TN
50.00%
FP
0.00%
FN
0.07%
TP
49.93%
Actual "Churn" status
Predicted "Churn" status
0
1
0
1
4000
3000
2000
1000
0


XGBoost balanced_data
Confusion Matrix (Test)
Accuracy: 91.77%
TN
70.90%
FP
3.12%
FN
5.11%
TP
20.87%
Actual "Churn" status
Predicted "Churn" status
0
1
0
1
800
600
400
200

11.5 Appendix F: (Solution 2) XGBoost Hyperparameter tuning

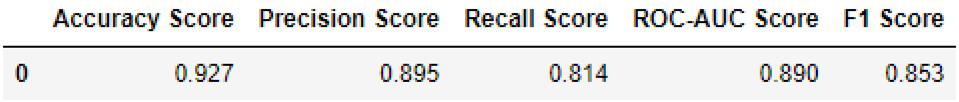


Best params - (10 fold CV, RandomSearchCV, accuracy) {'verbosity': 0, 'subsample': 0.8, 'scale_pos_weight': 2.8, 'reg_lambda': 2, 'n_estimators': 500, 'max_depth': 4, 'learning_rate': 0.1, 'gamma': 1, 'colsample_bytree': 0.7}



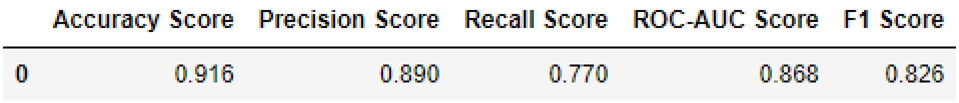
Best params - (10 fold CV, RandomSearchCV, recall){'verbosity': 0, 'subsample': 0.6, 'scale_pos_weight': 2.8, 'reg_lambda': 1, 'n_estimators': 100, 'max_depth': 4, 'learning_rate': 0.1, 'gamma': 4, 'colsample_bytree': 0.7}

Performance for 1st set:



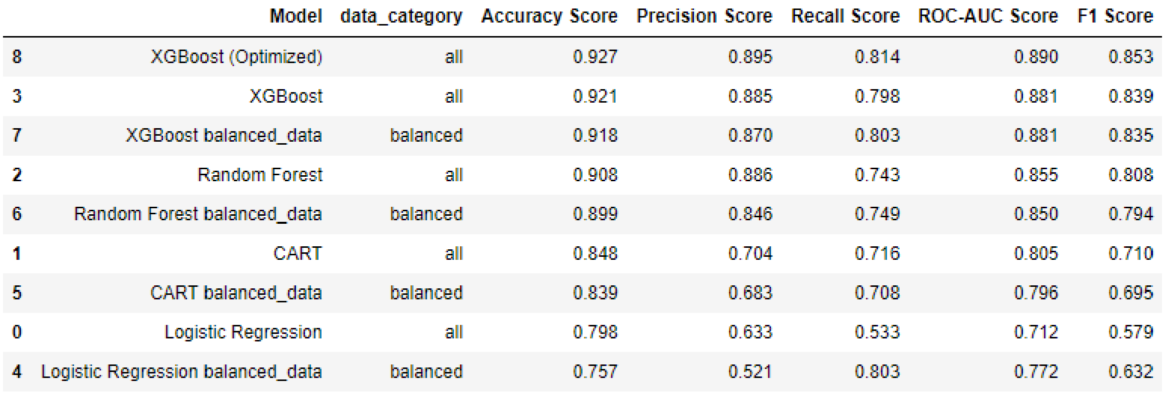
	Accuracy Score	Precision Score	Recall Score	ROC-AUC Score	F1 Score
0	0.927	0.895	0.814	0.890	0.853

Performance for 2nd set

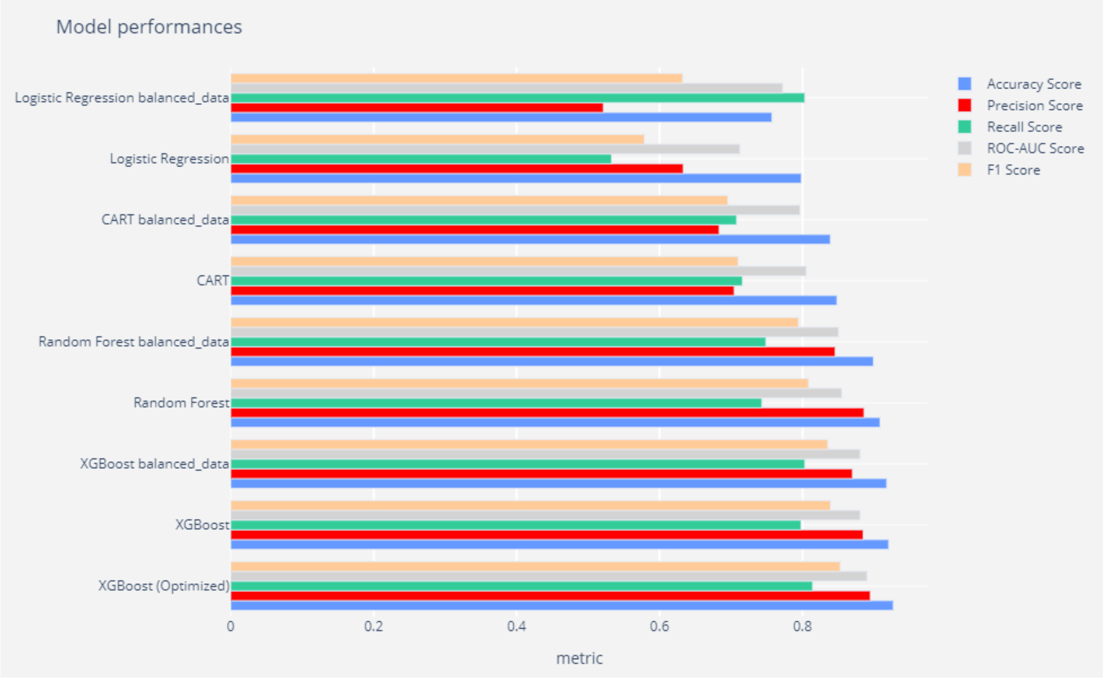


	Accuracy Score	Precision Score	Recall Score	ROC-AUC Score	F1 Score
0	0.916	0.890	0.770	0.868	0.826

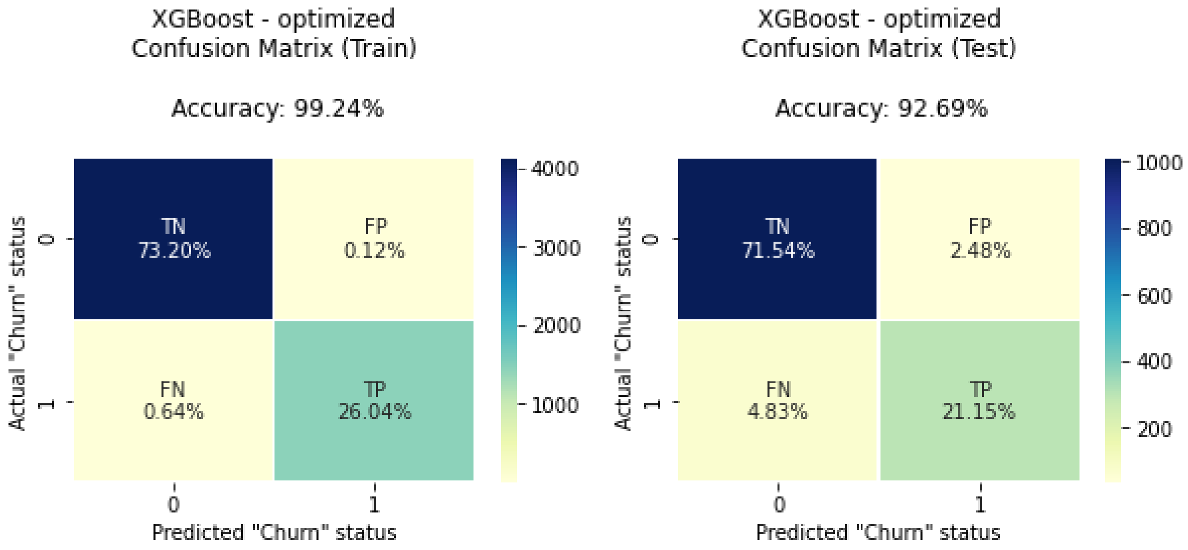
The first set of parameters is chosen, the performance is the best



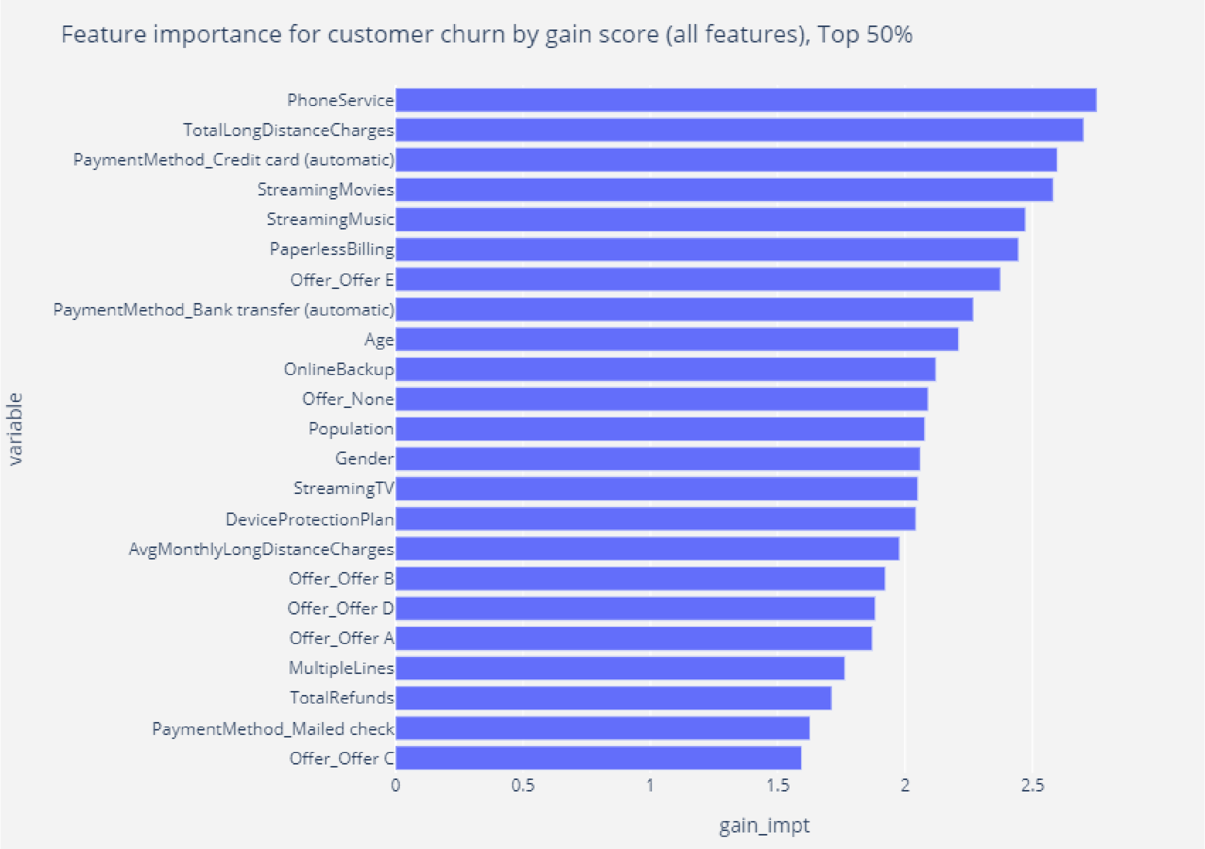
	Model	data_category	Accuracy Score	Precision Score	Recall Score	ROC-AUC Score	F1 Score
8	XGBoost (Optimized)	all	0.927	0.895	0.814	0.890	0.853
3	XGBoost	all	0.921	0.885	0.798	0.881	0.839
7	XGBoost balanced_data	balanced	0.918	0.870	0.803	0.881	0.835
2	Random Forest	all	0.908	0.886	0.743	0.855	0.808
6	Random Forest balanced_data	balanced	0.899	0.846	0.749	0.850	0.794
1	CART	all	0.848	0.704	0.716	0.805	0.710
5	CART balanced_data	balanced	0.839	0.683	0.708	0.796	0.695
0	Logistic Regression	all	0.798	0.633	0.533	0.712	0.579
4	Logistic Regression balanced_data	balanced	0.757	0.521	0.803	0.772	0.632

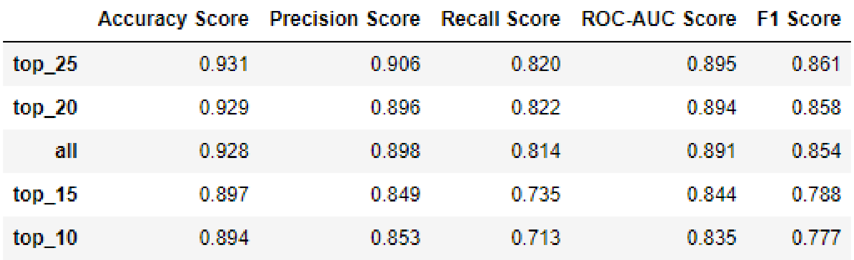


Confusion matrix for the optimal model

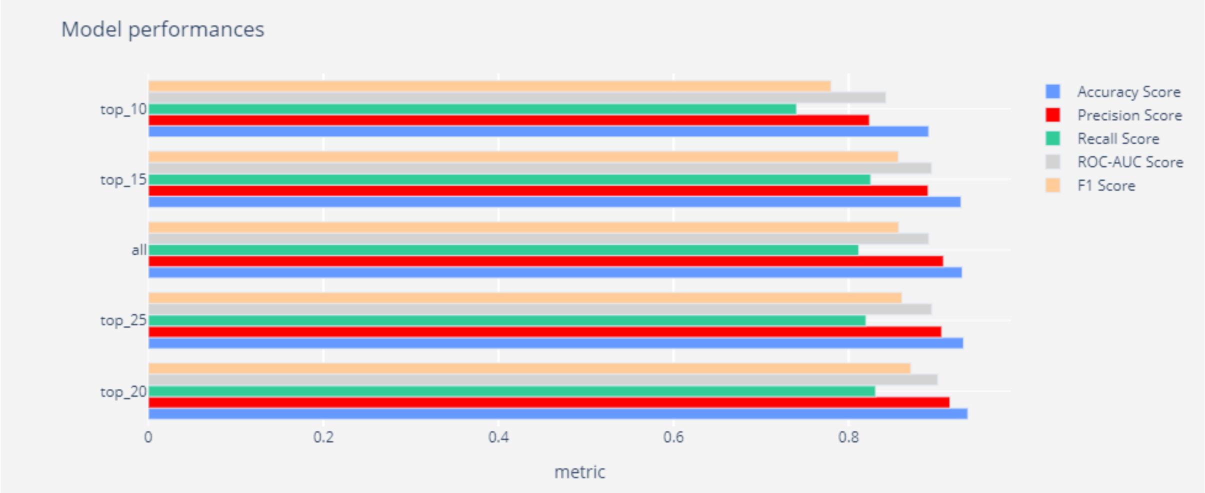


11.5 Appendix G: (Solution 2) XGBoost Feature Selection Variations

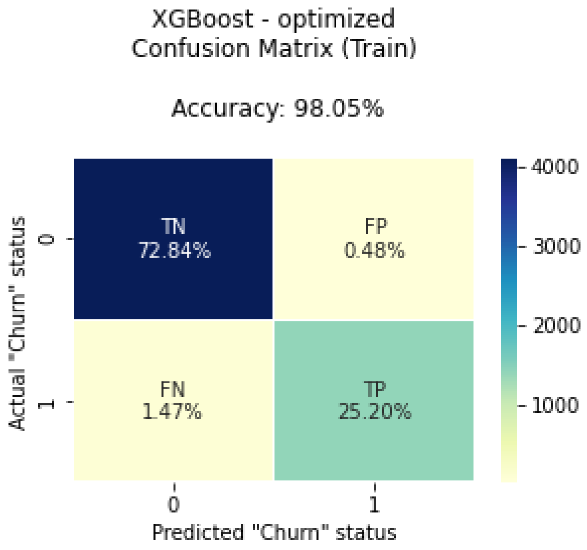


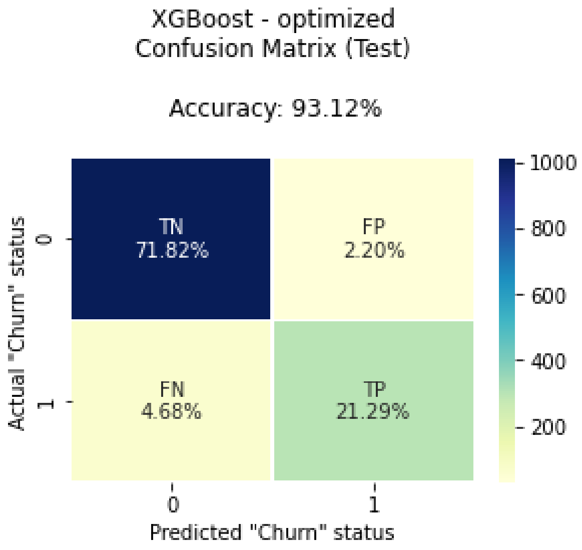


	Accuracy Score	Precision Score	Recall Score	ROC-AUC Score	F1 Score
top_25	0.931	0.906	0.820	0.895	0.861
top_20	0.929	0.896	0.822	0.894	0.858
all	0.928	0.898	0.814	0.891	0.854
top_15	0.897	0.849	0.735	0.844	0.788
top_10	0.894	0.853	0.713	0.835	0.777



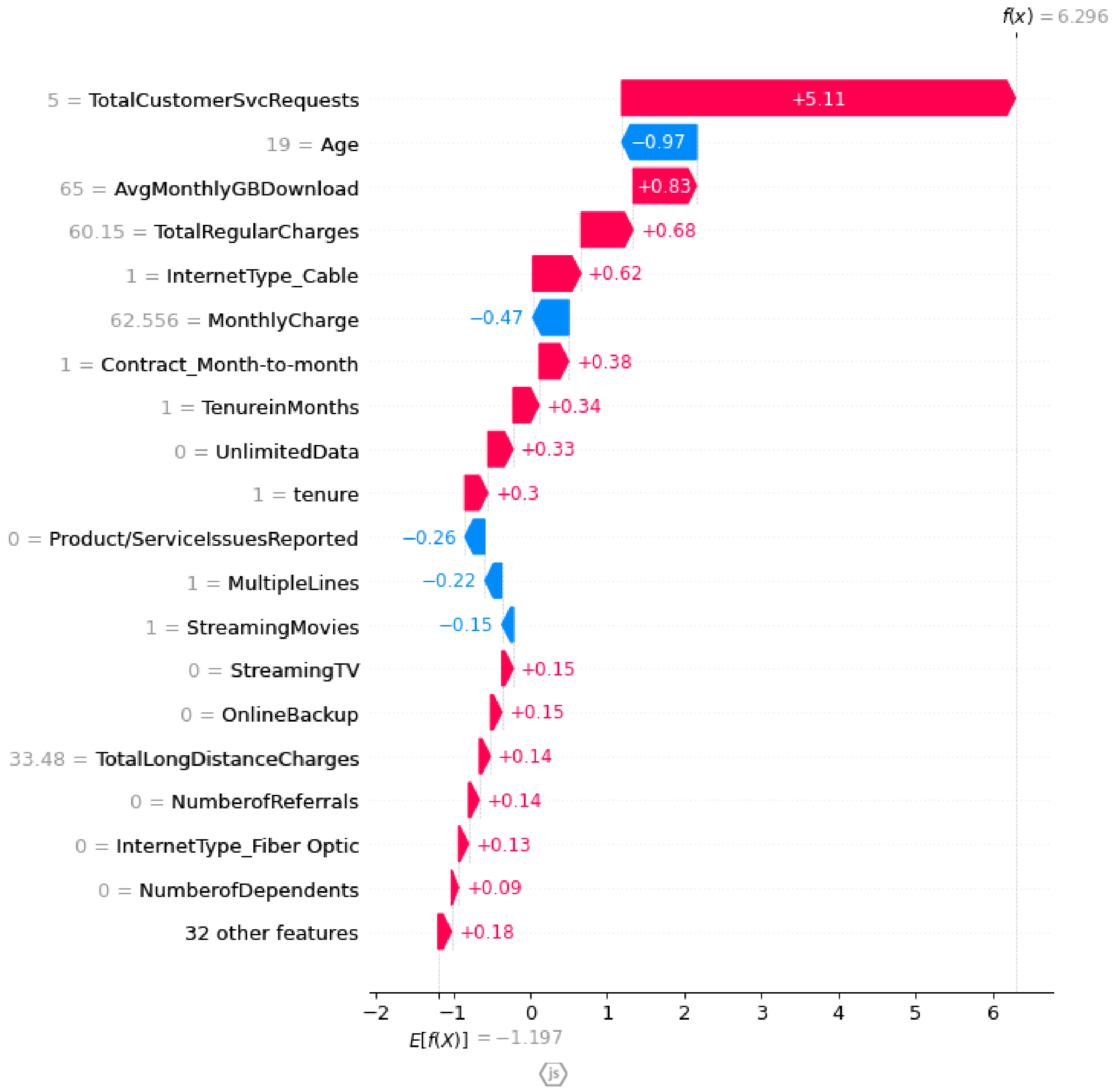
Optimized XGBoost with top 20 selected features:

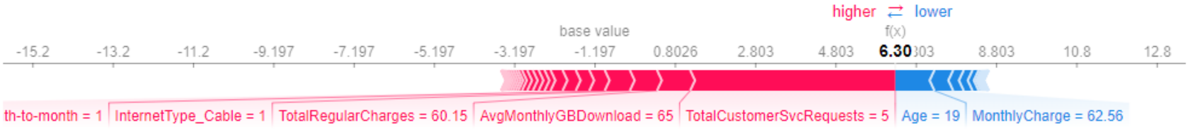


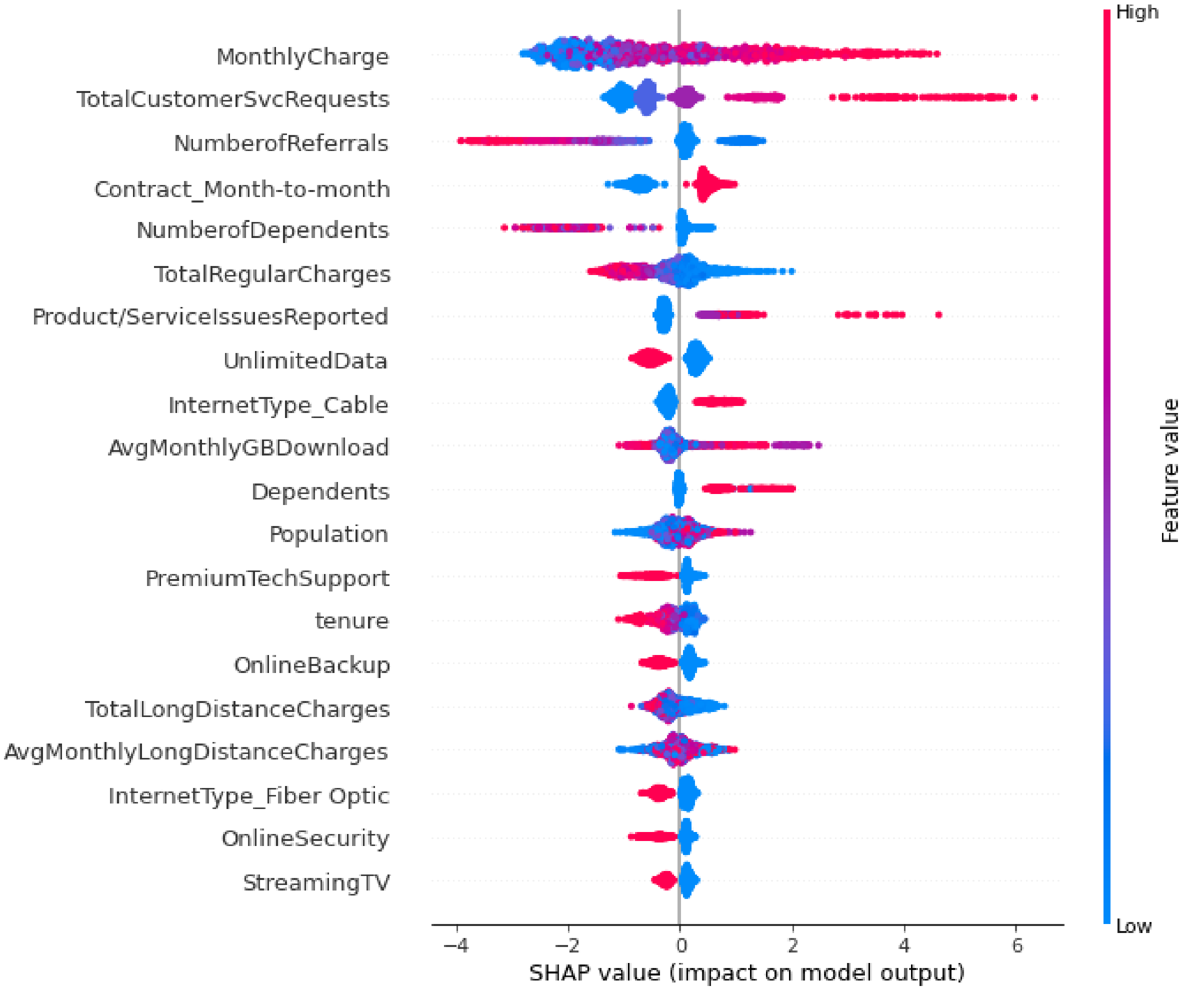


11.6 Appendix H: (Solution 2) Features Credit Allocation Insights Full

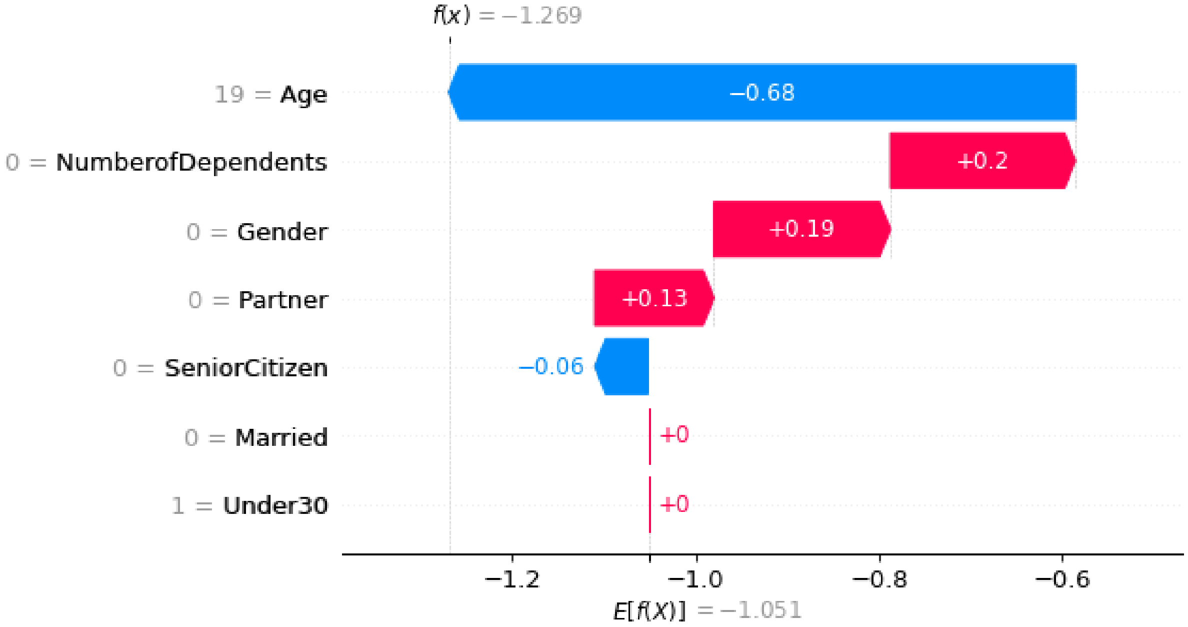
For all variables



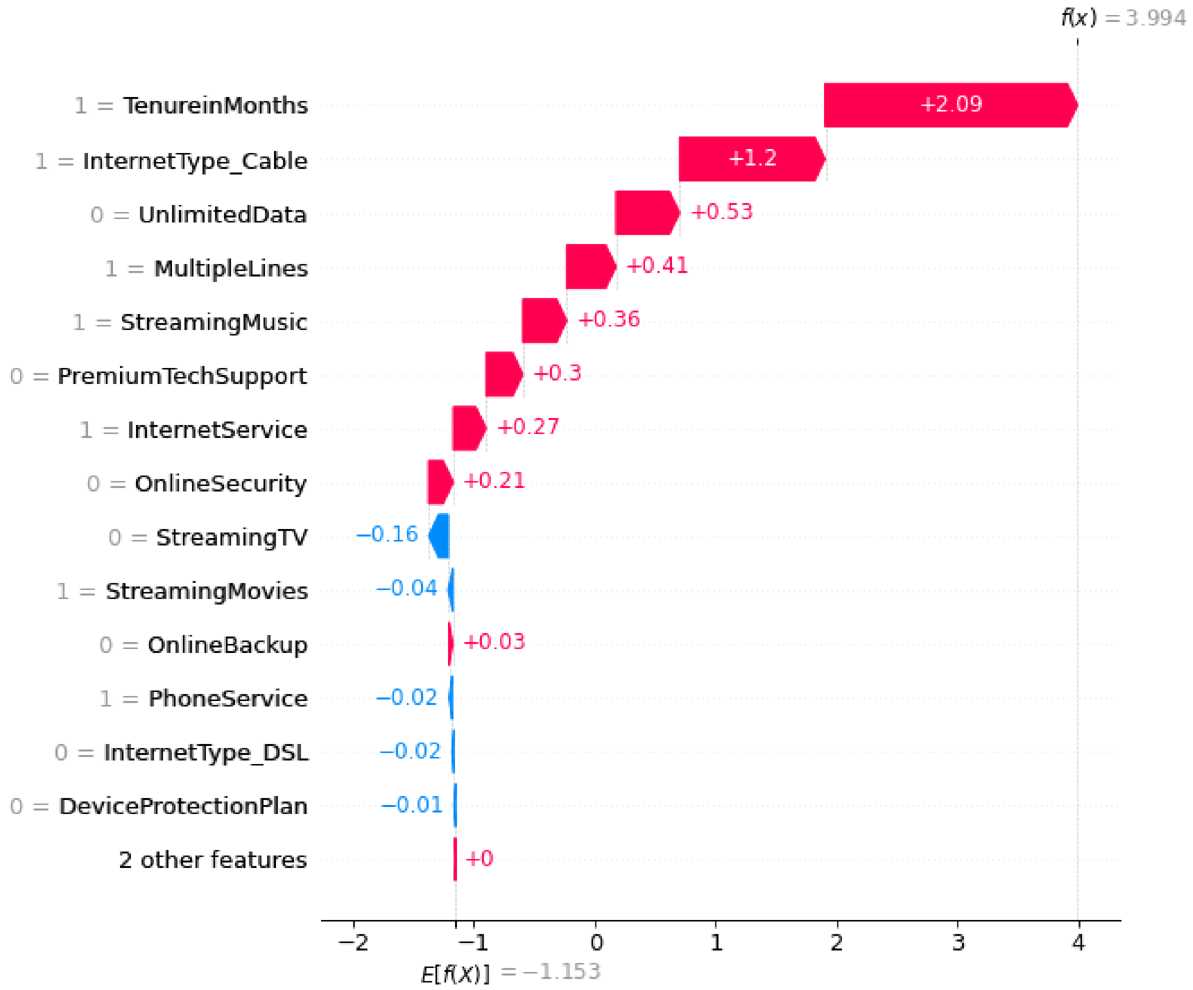




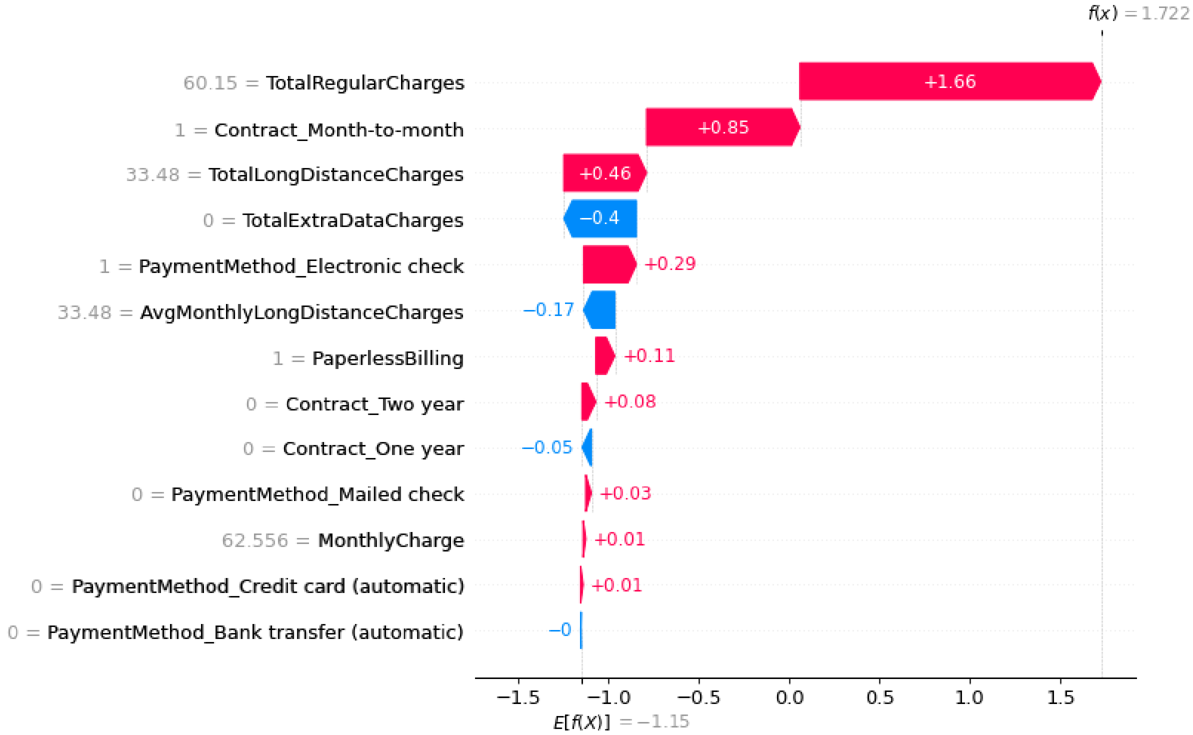
Demographical columns



Service features



Billing features

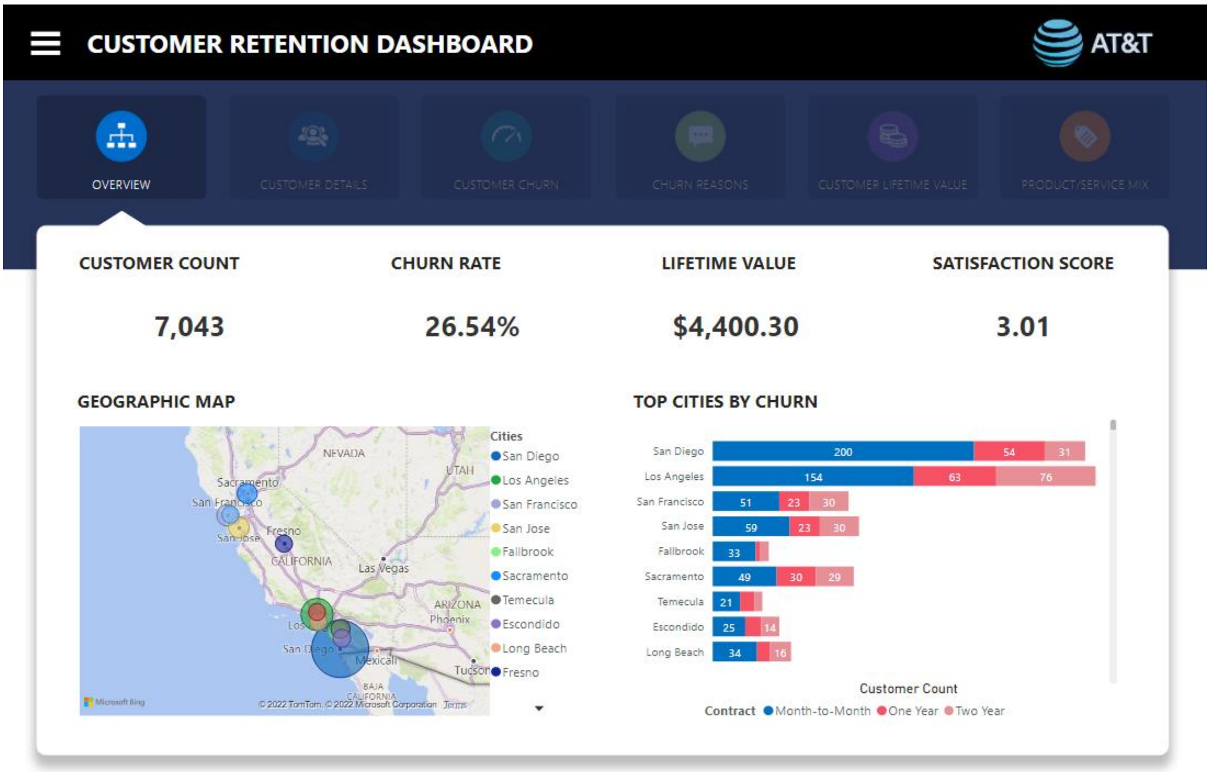

f(x) = 1.722
60.15 = TotalRegularCharges
+1.66
1 = Contract_Month-to-month
+0.85
33.48 = TotalLongDistanceCharges
+0.46
0 = TotalExtraDataCharges
−0.4
1 = PaymentMethod_Electronic check
+0.29
33.48 = AvgMonthlyLongDistanceCharges
−0.17
1 = PaperlessBilling
+0.11
0 = Contract_Two year
+0.08
0 = Contract_One year
−0.05
0 = PaymentMethod_Mailed check
+0.03
62.556 = MonthlyCharge
+0.01
0 = PaymentMethod_Credit card (automatic)
+0.01
0 = PaymentMethod_Bank transfer (automatic)
−0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
E[f(X)] = −1.15

11.7. Appendix I: Variables grouped into categories for analysis

Category name	**Variables**
Demographic	Senior Citizen, Partner, Age, Gender, Married, Number of Dependents**,** Under 30
Services	Phone Service, Multiple Lines, Internet Service, Online Security, Online backup, Streaming TV, Streaming Movies, Device Protection Plan, premium tech Support, Streaming Music, tenure in months, Unlimited Data, Internet Type_Cable, Internet Type_DSL, 'Internet Type_Fiber Optic, Internet Type_NA
Billing Column	Paperless Billing, Avg Monthly Long-Distance Charges Monthly Charge, Payment Method_Bank transfer (automatic), payment Method_Credit card (automatic), Payment Method_Electronic check, Payment Method_Mailed check, Total Extra Data Charges, Total Long Distance Charges, Total Regular Charges, Contract_Month-to-month, Contract_One year, Contract_Two year

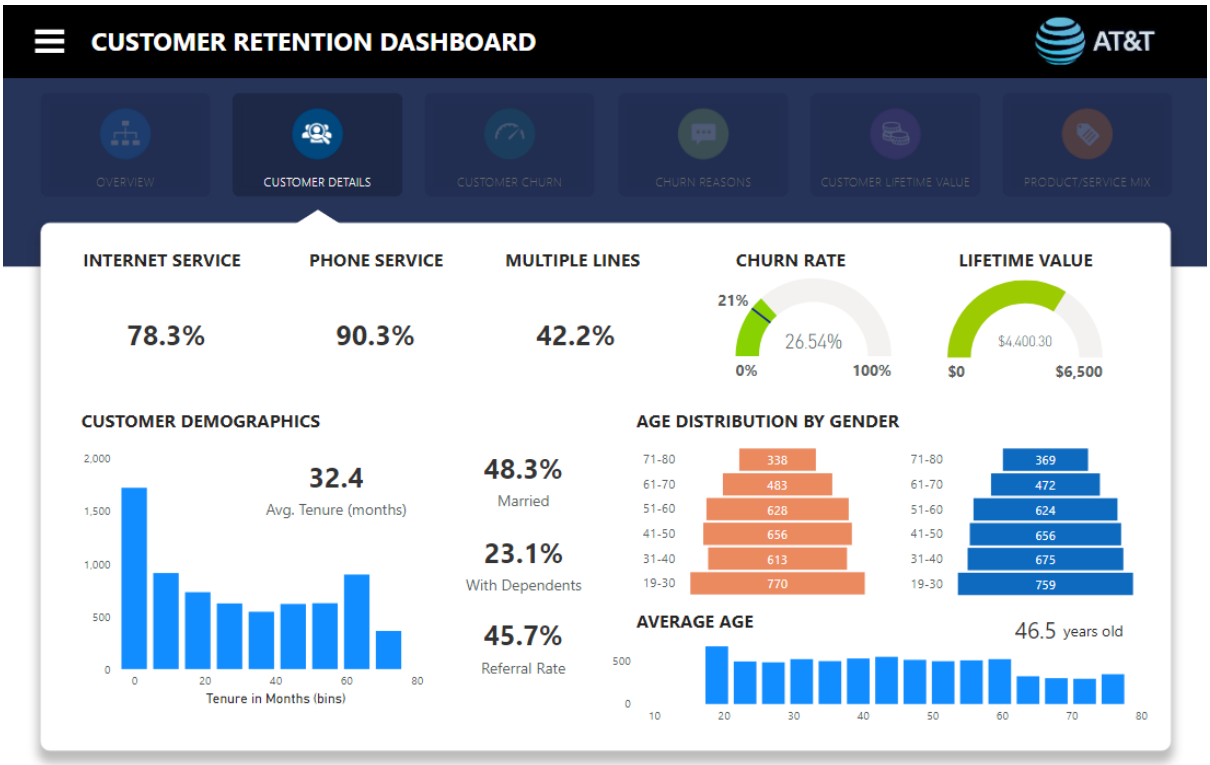
11.8. Appendix J: Screenshot of Dashboard

a. Dashboard overview



CUSTOMER COUNT	CHURN RATE	LIFETIME VALUE	SATISFACTION SCORE
7,043	26.54%	$4,400.30	3.01

b. Customer Details in Dashboard



INTERNET SERVICE	PHONE SERVICE	MULTIPLE LINES	CHURN RATE	LIFETIME VALUE
78.3%	90.3%	42.2%	26.54%	$4,400.30

32.4 Avg. Tenure (months)

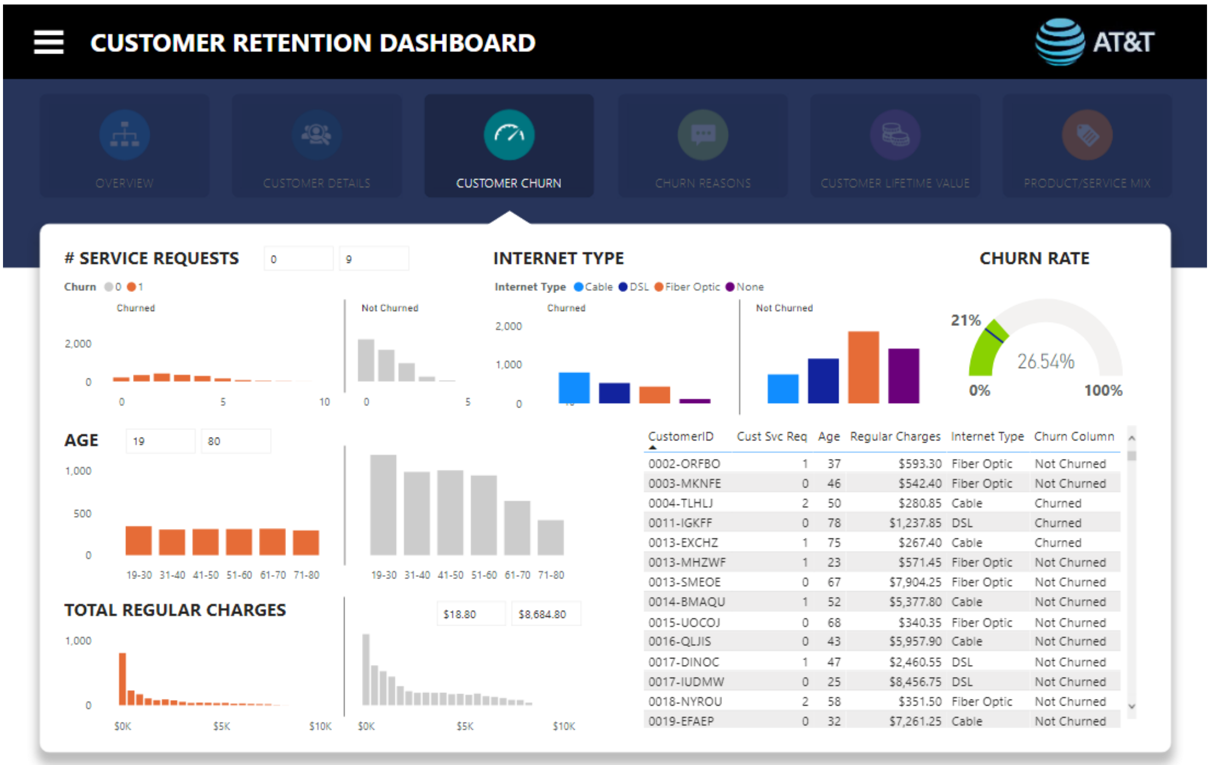
48.3% Married

23.1% With Dependents

45.7% Referral Rate

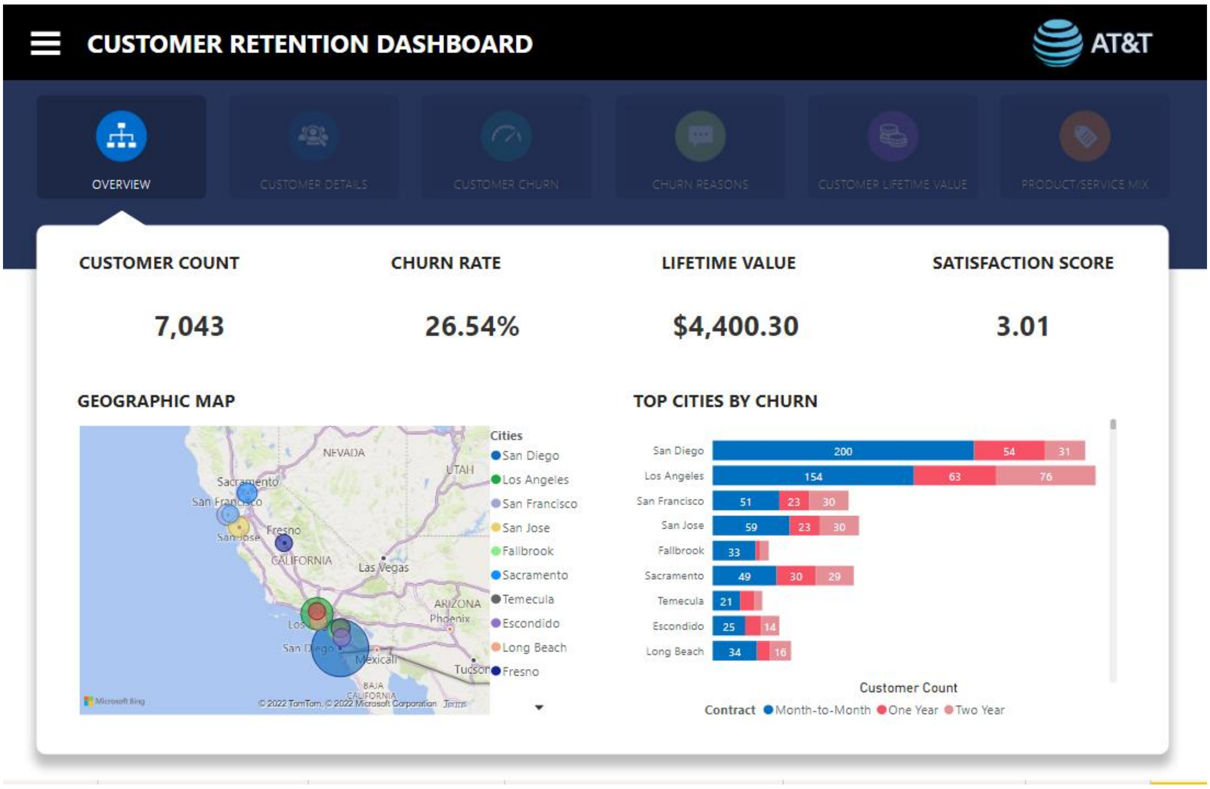
46.5 years old

c. Customer Churn by Telco Services



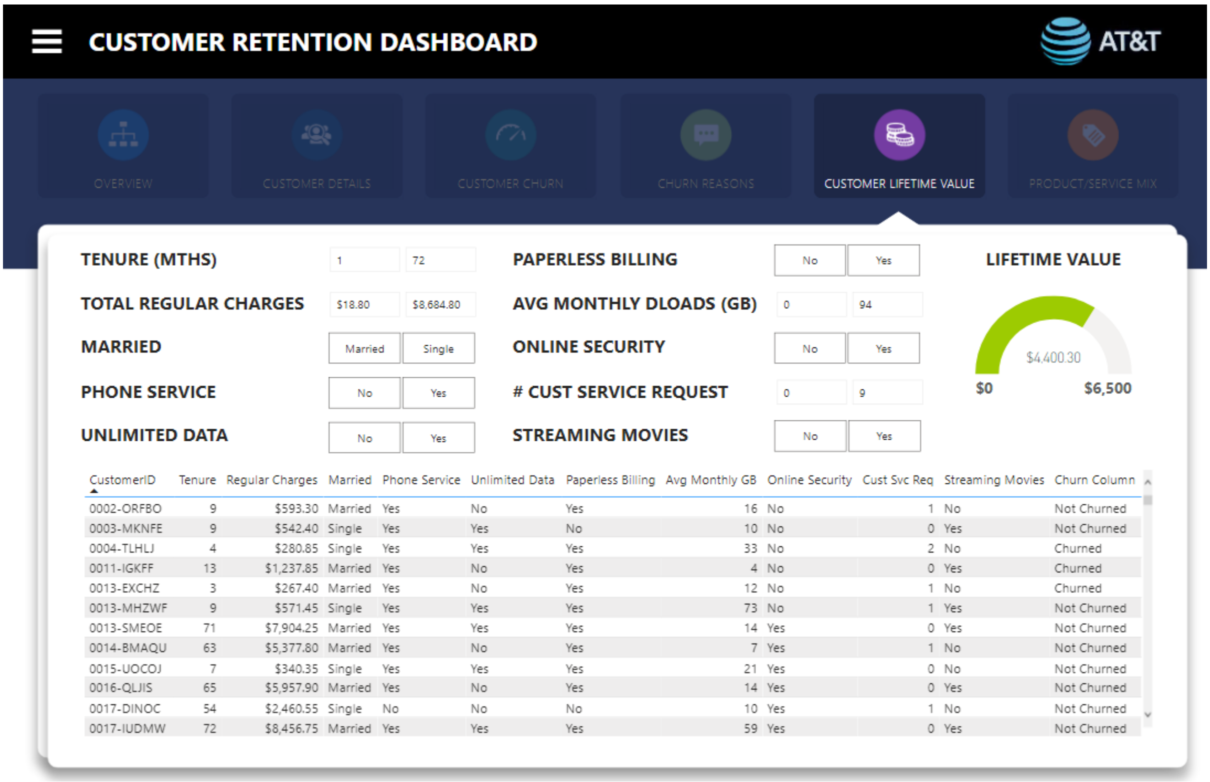
CustomerID	Cust Svc Req	Age	Regular Charges	Internet Type	Churn Column
0002-ORFBO	1	37	$593.30	Fiber Optic	Not Churned
0003-MKNFE	0	46	$542.40	Fiber Optic	Not Churned
0004-TLHLJ	2	50	$280.85	Cable	Churned
0011-IGKFF	0	78	$1,237.85	DSL	Churned
0013-EXCHZ	1	75	$267.40	Cable	Churned
0013-MHZWF	1	23	$571.45	Fiber Optic	Not Churned
0013-SMEOE	0	67	$7,904.25	Fiber Optic	Not Churned
0014-BMAQU	1	52	$5,377.80	Cable	Not Churned
0015-UOCOJ	0	68	$340.35	Fiber Optic	Not Churned
0016-QLJIS	0	43	$5,957.90	Cable	Not Churned
0017-DINOC	1	47	$2,460.55	DSL	Not Churned
0017-IUDMW	0	25	$8,456.75	DSL	Not Churned
0018-NYROU	2	58	$351.50	Fiber Optic	Not Churned
0019-EFAEP	0	32	$7,261.25	Cable	Not Churned

d. Customer Churn by Geography



CUSTOMER COUNT	CHURN RATE	LIFETIME VALUE	SATISFACTION SCORE
7,043	26.54%	$4,400.30	3.01

e. Customer Lifetime Value in Dashboard



TENURE (MTHS)	1	72	PAPERLESS BILLING	No	Yes
TOTAL REGULAR CHARGES	$18.80	$8,684.80	AVG MONTHLY DLOADS (GB)	0	94
MARRIED	Married	Single	ONLINE SECURITY	No	Yes
PHONE SERVICE	No	Yes	# CUST SERVICE REQUEST	0	9
UNLIMITED DATA	No	Yes	STREAMING MOVIES	No	Yes

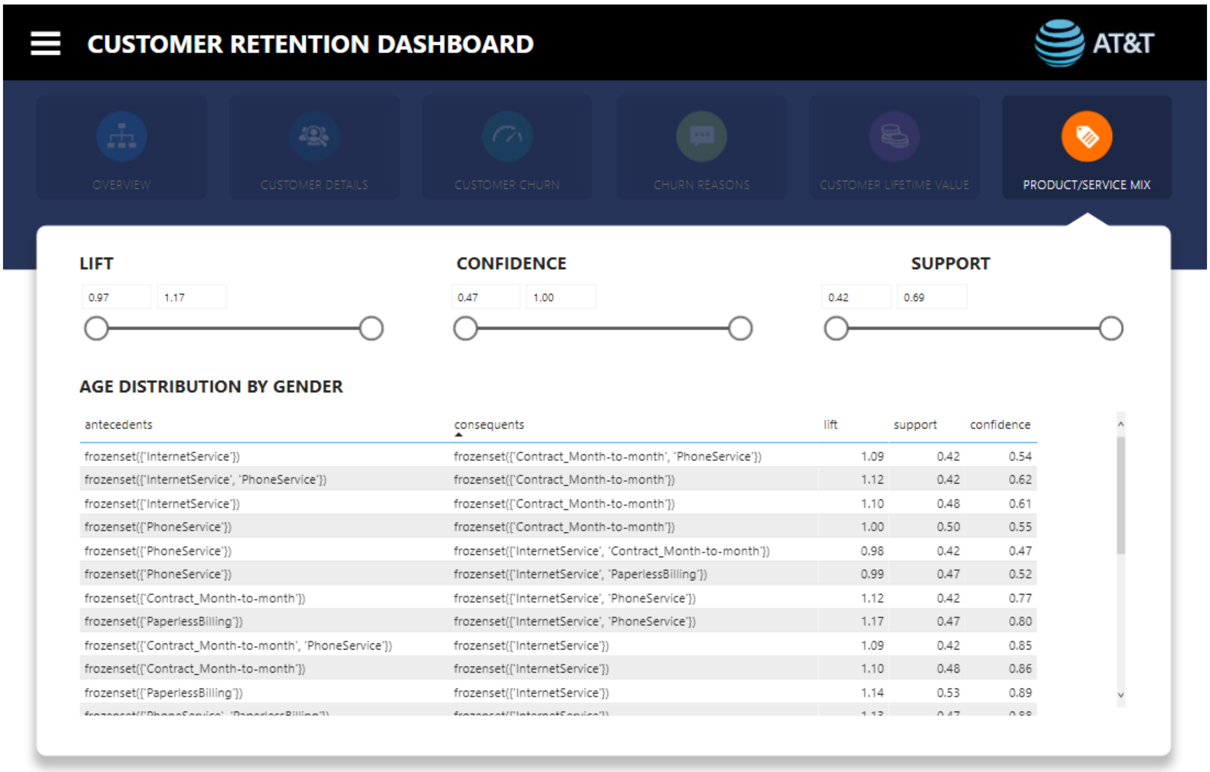
LIFETIME VALUE

$4,400.30

$0 $6,500

CustomerID	Tenure	Regular Charges	Married	Phone Service	Unlimited Data	Paperless Billing	Avg Monthly GB	Online Security	Cust Svc Req	Streaming Movies	Churn Column
0002-ORFBO	9	$593.30	Married	Yes	No	Yes	16	No	1	No	Not Churned
0003-MKNFE	9	$542.40	Single	Yes	Yes	No	10	No	0	Yes	Not Churned
0004-TLHLJ	4	$280.85	Single	Yes	Yes	Yes	33	No	2	No	Churned
0011-IGKFF	13	$1,237.85	Married	Yes	No	Yes	4	No	0	Yes	Churned
0013-EXCHZ	3	$267.40	Married	Yes	No	Yes	12	No	1	No	Churned
0013-MHZWF	9	$571.45	Single	Yes	Yes	Yes	73	No	1	Yes	Not Churned
0013-SMEOE	71	$7,904.25	Married	Yes	Yes	Yes	14	Yes	0	Yes	Not Churned
0014-BMAQU	63	$5,377.80	Married	Yes	No	Yes	7	Yes	1	No	Not Churned
0015-UOCOJ	7	$340.35	Single	Yes	Yes	Yes	21	Yes	0	No	Not Churned
0016-QLJIS	65	$5,957.90	Married	Yes	No	Yes	14	Yes	0	Yes	Not Churned
0017-DINOC	54	$2,460.55	Single	No	No	No	10	Yes	1	No	Not Churned
0017-IUDMW	72	$8,456.75	Married	Yes	Yes	Yes	59	Yes	0	Yes	Not Churned

f. Product/Service Bundling using Association rules



LIFT		CONFIDENCE		SUPPORT	
0.97	1.17	0.47	1.00	0.42	0.69

AGE DISTRIBUTION BY GENDER

antecedents	consequents	lift	support	confidence
frozenset({'InternetService'})	frozenset({'Contract_Month-to-month', 'PhoneService'})	1.09	0.42	0.54
frozenset({'InternetService', 'PhoneService'})	frozenset({'Contract_Month-to-month'})	1.12	0.42	0.62
frozenset({'InternetService'})	frozenset({'Contract_Month-to-month'})	1.10	0.48	0.61
frozenset({'PhoneService'})	frozenset({'Contract_Month-to-month'})	1.00	0.50	0.55
frozenset({'PhoneService'})	frozenset({'InternetService', 'Contract_Month-to-month'})	0.98	0.42	0.47
frozenset({'PhoneService'})	frozenset({'InternetService', 'PaperlessBilling'})	0.99	0.47	0.52
frozenset({'Contract_Month-to-month'})	frozenset({'InternetService', 'PhoneService'})	1.12	0.42	0.77
frozenset({'PaperlessBilling'})	frozenset({'InternetService', 'PhoneService'})	1.17	0.47	0.80
frozenset({'Contract_Month-to-month', 'PhoneService'})	frozenset({'InternetService'})	1.09	0.42	0.85
frozenset({'Contract_Month-to-month'})	frozenset({'InternetService'})	1.10	0.48	0.86
frozenset({'PaperlessBilling'})	frozenset({'InternetService'})	1.14	0.53	0.89
frozenset({'PhoneService', 'PaperlessBilling'})	frozenset({'InternetService'})	1.13	0.47	0.88