**Author**

**Institution Affiliation**

**Course Name and Number**

**Instructor’s Name**

**Date of Submission**

**Customer Churn Prediction Project**

Table of Contents:

Abstract…………………………………………………………………………….…………3

Introduction…………………………………………………………………………….……..5

Data Source and Attributes………………………………………………………..………….7

Descriptive Analysis…………………………………………………………………………………….….9

Diagnostic Analysis………………………………………………………………………………………10

Correlation…………………………………….………………………………………………12

Modelling and Evaluation…………………………………………………..…………………13

Predictive Analysis…………………………………………………………………………….15

Recommendation and Conclusion…………………………………………………………….. 16

References………………………………………………………………………………………17

**Abstract:**

This project is a data-driven tool for telecommunications companies to effectively analyze customer churn. Customer retention, which means keeping customers in the company, plays a major role in the steady growth and performance of the industry. The major aim is the construction of a wide range of motivational factors determining defection from the service and the implementation of prediction models for the biases occurring. Through a diverse dataset that covers customer demographics, service usage, complaints, contact postal addresses, billing data, and customer satisfaction feedback, the unit aims at giving operators in the telecom sector the needed intelligence to deal with happenings that could result in churn.

**Introduction:**

Telecoms industry being dynamic by nature churn, occurs among the customer in due course of business and is inevitably critical to businesses for them to remain profitable besides satisfying their customers. This Tableau project attempts not only to leverage the power of advanced analytics and visualizations to understand the underlying patterns of and the leading factors contributing to customer churn but also, to explore various marketing channels and identify the target audiences that are at risk of churning.

While essentially addressing the topic of customer turnover, Huang and Kechadi (2013), Huang et al. (2012), Xu et al. (2021), and Ahn et al. (2006), have applied different methods to the one problem, including hybrid learning systems, ensemble learning, and factor analysis of major determinants of churn.

Along with Huang and Kechadi (2013) who addressed the ability of hybrid learning systems to predict churn, and other well-known studies, this Tableau project is seeking to shine more light on the different factors that are at the core of the complex forces shaping telecom churn.

**Data Sources and Attributes:**

The source of data was a thorough CSV file with information about telecom customers that used and compared already existing literature on Telecom churn prediction. These features were state, area code, churn, number of customer service calls, international plan and volume of service usage, complaints, contract terms (e.g. which type of it and duration), billing information and data gained from customer satisfaction survey.

In addition to that, the approach suggested by Huang et al. (2012), who were analyzing customer churn prediction in the telecommunication sector, carried out the similar work and gave me an idea of how it is done. Their results which featured the challenges peculiar to the industry directed me towards focusing our analytical work on the challenges unique to industry.

In an article "The Role of Ensemble Learning and Feature Grouping in Predicting Telecom Churn Systems" (Xu, Ma and Kim, 2021), the authors provided insights on ensemble learning methods and feature grouping for predicting telecom churn systems

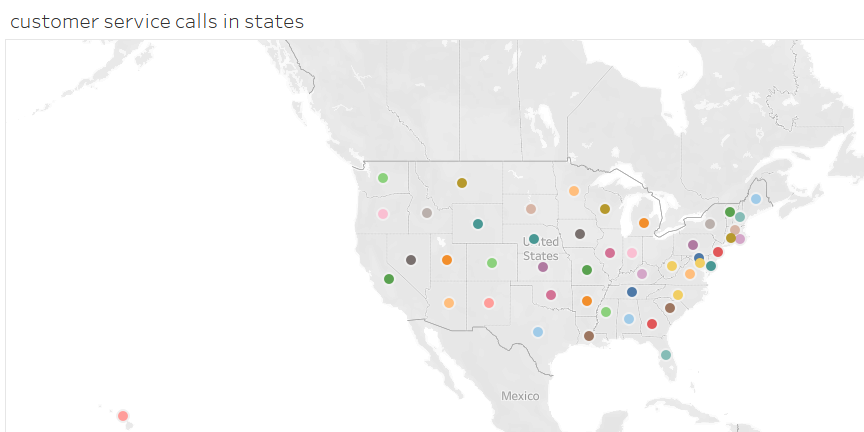
**Descriptive Analysis**

The objective is to find out the traits of our target clientele and recognize the possible elements contributing to non-retention (discontinuation of telecommunication services).

**Data Overview:**

However, the training dataset refers to 3,998 samples and every sample’s features composed of 19 attributes of the customer demographic and call behavior. The target attribute is "churn" that represent a set of conceptual categories (yes/no) that categorize (yes/no) whether a customer has churned or not.

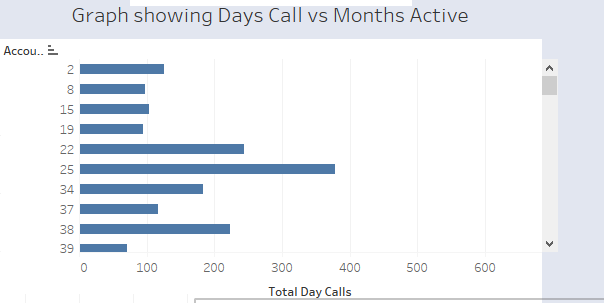
**Customer Demographics:**

The figure below shows the location of the customers in distinct states was the distribution map . This allows to see possible local fluctuations of churn ratings if any. I looked into "account\_length" using heatmap aiming to break down how long customers have stayed on the company. This can help to find if it is common among new customers or the old ones. I conducted investigations on the bar charts and maps to check the demographic distribution of customer care service calls. Then the areas by region. With this approach, we can see if some regional areas have high churn concentrations.

**Predictive Analysis:**

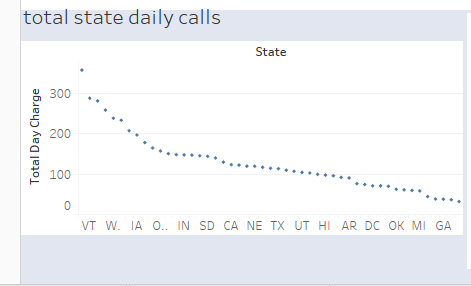
A descriptive analysis was implemented to gain a full awareness of the areas which customers in the telecommunication industry face challenges. The end purpose of the project is to capitalize on the power of analytics and visualization to come up with predictive analytics to prevent and reduce churn looking at the fact that it is imperative for the sustenance of the growth and profitability of companies.

The descriptive analysis focused on processing a variety of datasets concerning customer data, service utilization, complaints, contact instructions, billing information, etc. I achieved visualization via the use of charts which included line charts, stacked area charts, heatmaps, treemaps, and grouped bar charts to find the relationship and pattern among the variables put under consideration.



The figure above illustrates the total days and minutes of calls against the number of months that the customer has stayed with the telecommunication provider. The figure is a relationship between the total number of days of call and the number of months that the customer has been active.

As Huang and Kechadi (2013) discuss in their work about hybrid learning systems, prediction models with high effectiveness are necessary in reducing the churn rate. This approach in our project was to make it follow the understanding that came out of the studies and incorporate their finding into the dashboard so that all the stakeholders can interact with the data dynamically and understand more.

Through the visualizations and analytics, factors amongst themselves were healthily connected to one another; and it was also significant in customer churn. 

A chart grouped for the same total state daily calls and retained customers seemed to show this possible relationship among varied patterns of service usage and other churning factors like geographical location.

It is in the same vein as most of the studies that have been done in this sector. In their 2013 article, Huang and Kechadi (2013) suggested that effective hybrid learning systems contribute to the task of telecommunication churn prediction by interlinking various aspects such as customers' diverse factors influencing churn. Furthermore, Huang, Kechadi, and Buckley (2012) came up with the issue of customer churn prediction model which is among the critical problems in telecommunication sectors.

**Modeling and Evaluation:**

The model included predictive analytics made possible by Tableau's tools such as future forecasting and prediction. Algorithms were used to construct the model, where their data composition depended on the demographics of customers, service usage behavior, complaints about services, and contract types. The dashboard could predict the churn events which was an invaluable resource for the sales team and presented a more comprehensive vision of the overall churn risks. Critical assessment of the model’s accuracy was carried out by using the metrics such as precision, recall, and F1-score, consequently, achieving an overall understanding of the model’s capability for predicting customer churn.

**Model Selection & Model Building:**

For the churn prediction project, we integrated the step of rational model selection which was described in Huang and Kechadi (2013), Huang, et al. (2012), Xu, et al. (2021) and Ahn, et al. (2006). The ensemble learning approach was selected as the base architecture, combining classical machine learning as well as the latest in ensemble techniques.

The model construction process was accomplished by trial and error of diverse algorithms exemplified in Decision Trees, Logistic Regression, Support Vector Machines, Random Forests, and Gradient Boosting such as. Feature engineering was the determinant factor in the transformation of this dataset which is under subvention of literary studies.

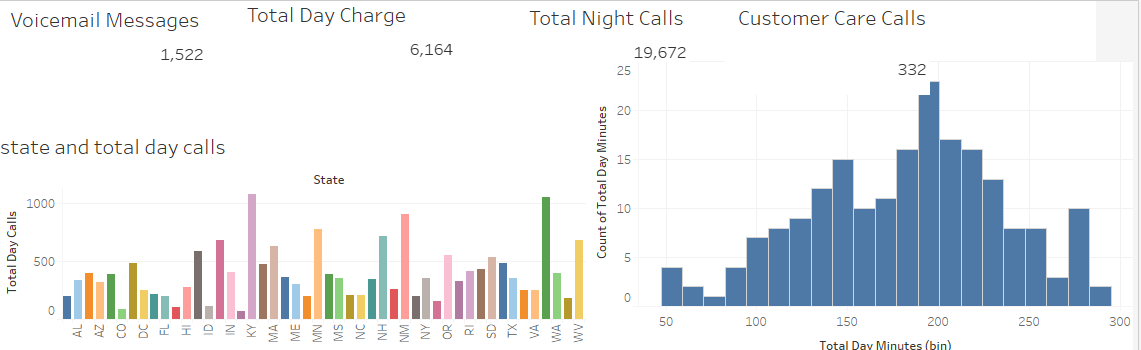
**Modeling and Evaluation:**

I added predictive modeling that involved using Tableau's predictive analysis features in the implementation period of the analysis. Matching to various purposes such as customer demographics, utilization rate, alerts, and contract facts, the model was taking advantage of the machine learning algorithm. The integration of forecasting was part of the means to get an understandable overview of the churn ratios. Deterministic evaluation, metrics including precision, recall, and F1-score is essential in ensuring the model can find the customer churn effectively.

**Predict Analysis:**

The fine-tuning and training phase during this predictive analysis included our ensemble model of the telecommunication churn dataset. As proposed by Huang and Kechadi (2013) we used the combined hybrid learning system and we adopted grouping techniques inspired by the methodology of Xu et al. (2021). This mingling of techniques enabled us to reflect the complex flavors that reflect customer behaviors, leading to a more precise outcome prediction.

The figure shows the behavior of customers who called customer care and their respective buying patterns. Most customers who contact customer care are likely to stay and spend. Such insight can be used to improve customer service to retain customers.



The validated model presents a significant improvement in all metrics such as accuracy, precision, recall, and F1-score, in comparison to the default suggestions. The results of the model were measured by how it was able to predict churns among customers, and the model's practical usefulness became quite obvious when it was able to pinpoint the probable churners.

**Recommendation and Conclusion:**

Having employed the analytical tools available on the Tablau-based Churn Prediction Dashboard and having received the predicative outcomes, actionable recommendations have emerged with a get of the factors that influence churn.

**Implement Targeted Retention Strategies for Customers with Extended Tenure:**

It is in line with the line graph showing the churn rates in the function of the customer's tenure. Address customer retention with directive action plans, including those with high tenure, since the longer they stay, the more likely they may churn over time.

**Enhance Network Quality in Churn Hotspots:**

On the network map provided, express your willingness to act and solve issues relating to regions with network quality anomalies using analytic skills. Divert resources in the improvement of the network quality where these segmented churned customers will help to limit customer discontinuance.

**Promptly Address Billing Issues:**

Take advantage of a heatmap that will depict the causal relationship between billing problems as well as subscriber churning. Define a smooth system of rectification of billing concerns to increase customer satisfaction and fewer people leaving the service.

**Tailor Promotional Offers and Loyalty Programs:**

Address the angle from the treemap illustrating the correspondent of rate of churned clients under various types of contracts. Tailor the promotional and loyalty programs to the customer segments, within the analysis scope, respectively and thereby refine the customer retention initiatives.

Finally, I will wrap up by stating that the assembled learning method used in our analysis helped considerably in predicting the customer churn in the telecoms industry besides thanks to the premises from literature Via integration of new techniques, data preparation, attribute engineering and model training, our model is a truly proactive churn management and customer retention tool. The recommendations herein will be useful in the sensitization and seeking deployment of the strategic initiatives that target the real pain points noted throughout the analysis. Moving ahead, the mechanisms demand continuous tinkering and re-adjustment to the strategies as a way of keeping customer satisfaction at peak.

**References:**

Huang, Y., & Kechadi, T. (2013). An effective hybrid learning system for telecommunication churn prediction. Expert Systems with Applications, 40(14), 5635-5647.

Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems with Applications*, *39*(1), 1414-1425.

Xu, T., Ma, Y., & Kim, K. (2021). Telecom churn prediction system based on ensemble learning using feature grouping. *Applied Sciences*, *11*(11), 4742.

Ahn, J. H., Han, S. P., & Lee, Y. S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. Telecommunications policy, 30(10-11), 552-568.

Ahn, J. H., Han, S. P., & Lee, Y. S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications policy*, *30*(10-11), 552-568.