**Customer Churn Analysis of a Telecom Company**

**By**

**Pavan Kumar Nagula**

Table of Contents

[Introduction 2](#_Toc131626812)

[Objective 2](#_Toc131626813)

[Scope 2](#_Toc131626814)

[Methodologies 2](#_Toc131626815)

[Related work 3](#_Toc131626816)

[Project Planning 3](#_Toc131626817)

[Project Management 4](#_Toc131626818)

[Data Collection 5](#_Toc131626819)

[Data Analysis 6](#_Toc131626820)

[Descriptive Analysis 7](#_Toc131626821)

[Exploratory Data Analysis 7](#_Toc131626822)

[Predictive Analytics 11](#_Toc131626823)

[Model Implementation 12](#_Toc131626824)

[Model Comparison 15](#_Toc131626825)

[Model Selection 15](#_Toc131626826)

[Recommendations 15](#_Toc131626827)

[Conclusion 17](#_Toc131626828)

[Project Access 17](#_Toc131626829)

[Contributions 17](#_Toc131626830)

[References 17](#_Toc131626831)

Introduction

Telecom customer churn analysis is a critical area of study for telecom companies, especially in today's highly competitive telecom industry. Churn refers to the number of customers who discontinue using a particular telecom company's service in a given period. Telecom companies are constantly looking for ways to reduce churn rates and retain customers. One of the most effective ways to do this is by conducting a customer churn analysis.

A telecom customer churn analysis project involves analyzing data on customer attributes and usage patterns to identify trends and patterns that could indicate a customer is likely to leave the company. By analyzing the data, the project aims to identify the factors that contribute to customer churn and come up with strategies to reduce churn rates.

The telecom industry is highly competitive, with companies vying for market share by offering more competitive prices and better services. Customer loyalty is crucial to the success of any telecom company and reducing churn rates is an essential part of retaining customers. Telecom companies that are successful in reducing churn rates can reduce costs associated with acquiring new customers and increase revenue from existing customers.

The telecom customer churn analysis project is not a one-time event but rather an ongoing process. The data must be continually collected, analyzed, and updated to ensure that the models remain accurate and up to date. This ongoing process is essential to the success of the project and the company's ability to reduce churn rates.

## Objective

The objective of the report is to analyze the existing customer database and predict customer churn by looking into different customer attribute types and coming up with recommendations when applied to the business environment can improve and control customer retention.

## Scope

The scope of the project involves collecting and analyzing data on customer demographics, usage patterns, and behavior. An analysis is done keeping in mind different factors contributing to churn. The data is then used to develop predictive models that can identify customers who are at risk of churning. The models are compared, and the model of best fit is chosen. Recommendations are provided to the telecom company based on the key findings. By identifying these customers, the company can develop targeted retention strategies to reduce churn rates.

## Methodologies

Both Project Management and Analytical methodologies were used in the project. Project Management practice like DMAIC was used to create framework for the project. The entire project was broken down into Define, Measure, Analyze, Improve and Control stages. The Define, Measure, Analyze stages are implemented in the project and recommendations to the telecom company is provided to execute the Improve and Control stages. The stages were further broken down into individual tasks and were tracked through the lean process Kanban. Notion, a project management software was used as the Kanban Board. The entire suite of analysis was performed using Machine learning algorithms in python use google Collab as the environment. A dashboard was set up in tableau to aid visualization.

## Related work

There is abundant research related to the prediction of churning rate in the telecom industry. The research also discusses the causes for customer churn. Most of them used machine learning techniques to achieve their final goal. There are also some researchers who focused on implementing project management methodologies like Lean, Six Sigma, Lean Six Sigma (LSS) in the telecom industry. This project is the collaboration of both machine learning techniques and project management methodologies.

Mariani et al. [1] tried to predict the customer churn rate of a Syrian telecom company SyriaTel, the data collected has customer information over a period of 9 months. The dataset was used to train, test, and evaluate the system using four machine learning algorithms, including Decision Tree, Random Forest, Gradient Boosted Machine Tree (GBM), and Extreme Gradient Boosting (XGBOOST). The model's performance was evaluated using the Area Under Curve (AUC) measure, and a value of 93.3% was achieved.

Elimisery et al. [2] presented the feasibility of implementing project management methodologies and principles in the telecom industry to improve the customer service experience and discussed difficulties faced by telecom industries to include project management tools like DMAIC and Define-Measure-Analyse-Design-Verify (DMAIDV).

Project Planning

Project planning is essential before the commencement of any project. It helps ensure that the project is completed on time, within budget, and meets the project's objectives. Below are some steps that can be taken to plan a churn analytics project effectively

**Identifying the Business Problem:** The business problem in a churn analysis project is to reduce the number of customers who discontinue using a telecom company's particular service in each period, known as churn. The primary idea is to identify the factors that contribute to customer churn and develop targeted retention strategies to reduce churn rates. A prediction model that can forecast customers who are about to churn is also necessary to target the retention strategies towards the pool of customers in the prediction bucket.

**Establishing a Project Management Technique:** For the project to be successful and completed on time, the project has to be supervised under a relevant project management technique like Lean and Six Sigma.

**Preparing Data for Analysis:** Preparing data for a churn analysis project in a telecom company involves several steps, including data collection, cleaning, integration, and transformation.

**Performing Descriptive Analysis:** A descriptive analysis has to be conducted to identify patterns, trends, and insights related to customer churn. Descriptive analysis can help provide a summary of the entire data set and help understand top of the line insights the data represents.

**Performing Exploratory Analysis:** Exploratory analysis is necessary to gain a better understanding of the data and identify potential relationships between variables. Identifying correlations between variables and clustering analysis are all part of exploratory analysis.

**Performing Predictive Analysis:** Predictive modelling section necessary involves building classification machine learning models using python to predict customer churn and the effect of independent variables on the churn.

**Providing Recommendations:** The recommendation section states the findings established through the analysis. It also gives actionable insights that when implemented can help reduce the churn rate of customers in the telecom company.

**Project Timeline:**

**Diagram

Description automatically generated**

# Project Management

Project management is crucial for data analysis projects because it helps to organize and plan the various stages involved in such projects. This includes resource allocation, time management, risk management, and communication and collaboration between stakeholders. With project management, data analysts can set realistic timelines, manage resources effectively, identify and mitigate risks, and ensure that everyone is working towards the same goals. Ultimately, project management helps to ensure that data analysis projects are completed on time, within budget, and with high-quality results.

For this churn analysis project, we have used six sigma as the backbone of the project. The DMAIC flow of project management has been followed and steps were individually curated. Define, Measure, Analyze, Improve, Control in data analytics include defining project goals and objectives, measuring current data and processes, and analyzing data to identify gaps between the current and desired state. The next step involves improving the processes by implementing solutions and monitoring their impact. Finally, the control step involves ensuring that the improvements are sustained over time by implementing process controls and continuous monitoring in the telecom industry.

To improve the resource management and time management process, the project was broken down into smaller individual step by step tasks to proceed using kanban way of managing work. Kanban provides a visual representation of work processes and helps teams to prioritize tasks, identify bottlenecks, and optimize workflow. Three stages to organize tasks were created namely “Not Started”, “In progress”, “Done”. Individual tasks were bucketed into three stages depending on the stage of completion.

The snapshot of the Kanban module used is given below

Graphical user interface, text, application

Description automatically generated

Data Collection

**Data source and Description:** The data for performing the churn analysis is obtained from Maven Analytics. Maven Analytics is a platform that offers data analysis resources, including datasets for practice. The telecom company dataset provided by Maven Analytics contains customer data from a fictitious telecommunications company. The dataset contains information on customers, their usage patterns, and account information. The data is collected from the account information of customers, customer feedback, customer surveys, Call Detail Records, Third-party Data. The survey and feedback data were collected through e-surveys and on-call responses. The surveys included customer satisfaction, customer demographics, information about price and competition offerings comparison.

The collection of data is done via a Customer Relationship Management system. In particular, the Customer Churn table comprises 7,043 rows and 37 attributes representing customers from a California-based Telecommunications company in the 2nd Quarter of 2022. Each customer is represented by a customer id, location, subscription services, tenure, quarterly status (i.e., whether they joined, stayed, or churned), and other relevant data.

The data overall is standard enough to review but needs to be tweaked around to be used for Data analysis

**Data Cleaning:** Data cleaning is an essential step in the data analysis project, like the customer churn project, because it ensures the accuracy, consistency, and completeness and removes bias from the data.

**Data Modification:** To conduct bivariate analysis, we derived categorical attributes from the numerical variables. Firstly, we divided the 'tenure in months' attribute into bins since the maximum tenure was 72 months. This allows us to better understand the relationship between customer tenure and churn. Secondly, we categorized customers by age group ('Teenagers', 'Adults', and 'Senior Citizens') to determine which age groups are more likely to churn or stay with the service. Thirdly, we categorized data usage based on the 'average monthly GB download' attribute into low, medium, and high usage groups. This helps us to identify patterns in customer data usage that may affect churn. Finally, we converted the 'total revenue' variable into a customer type ('high revenue', 'medium revenue', and 'low revenue') to understand which types of customers generate the most revenue for the company. These categorizations will enable us to conduct more detailed bivariate analyses and provide actionable insights to reduce customer churn.

**Tenure Group Age Group**

Text

Description automatically generated with medium confidence A picture containing graphical user interface

Description automatically generated

**Average Usage of Data Revenue from Customer**

Graphical user interface, text

Description automatically generatedGraphical user interface, text

Description automatically generated

**Missing Data:** The major challenge in the data set was the presence of abundant missing data. Variables like churn reason and churn category almost had 73% missing data. The other attributes on average have 21% missing data. Missing data is present in both numerical and categorical columns. Missing data in the numerical fields were filled by replacing the missing value with the mean of the entire column. Missing data for categorical variables were filled by replacement with the mode values.

A black screen with white text

Description automatically generated with low confidence A picture containing graphical user interface

Description automatically generated

# Data Analysis

Once data has been collected and cleaned, data analysis involves using statistical and computational techniques to turn raw data into useful insights. The aim is to extract meaningful information from the data that can be applied to inform decision-making. This process typically involves exploring and analyzing the data to uncover useful patterns, trends, and relationships. Data analysis has been done using python because it provides a wide range of libraries and tools that simplify the process.

## Descriptive Analysis

When doing a descriptive analysis, statistical metrics like mean, median, mode, range, standard deviation, and variance are used to summaries and describe the properties of a dataset. The objective of descriptive analysis is to present a brief and simple image of the data, which may be helpful for identifying patterns and trends and comprehending the data's overall structure and distribution.

Descriptive analysis has been performed on quantitative variables like age, number of dependents, number of referrals, tenure in months. The average age of the customers is 46 and 75% of our customers are over 60 years. More than 75% of customers have tenure for less than 55 months, whereas the mean for the same is around 25 months (about 2 years). Average monthly charges are around 63 USD, but more than 50% of customers pay only 21 USD. Consumption of Avg Monthly data usage is also less than 20GB by 50% of the population.

Graphical user interface, text

Description automatically generated

## Exploratory Data Analysis

Exploratory data analysis (EDA) is the process of analyzing and summarizing a dataset's important features in order to acquire understanding of its underlying structure. EDA often combines visual and quantitative tools to study the data, including the creation of histograms, scatter plots, and box plots, the computation of summary statistics, and the determination of correlations between variables.

Chart, bar chart

Description automatically generated

Following cleaning, EDA was performed on the complete data set. More than 25% of consumers leave for a variety of reasons, including service dissatisfaction, high prices, a narrow selection of services, and more, which come under the categories of competition, discontent, attitude, and pricing.

**Customers Churning because of Competition:** Almost 45% of customers churned because of the competitor. Out of which 313 customers churned because competitor had better devices and 311 customers left because they had better offer from the competitor.

**Customers Churning because of Dissatisfaction:** Out of churned, 17% of them left for the reason of dissatisfaction with customer service. 220 of them churned because they did not like the attitude of support person. 94 of customers left because of the attitude of the service provider. Whereas only 12% of the customers left because of the high prices.

It appears that the company's inability to compete with other enterprises is the major cause of churning. If we examine price-related churn, it does not appear to have a significant impact. So, unhappiness with the product and poor employee behavior is the cause of overall turnover. The company must take these aspects into consideration if it wants to win back existing customers. Most of the "competitor" rationale is made up of elements like dissatisfaction and attitude.

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

Different factors affecting churn rate are age, marital status and number of dependents.

The age distribution of new and stayed customers is comparable. (19 to 60 range). However, once we cross this threshold, both our joined and stayed customer counts begin to decline. If we look at the distribution of churned customers, we may deduce that the items don't suit elderly individuals as predicted.

Married people are more loyal and stay with company for long period whereas unmarried people have a tendency to leave the company. The majority of the company's customers are single individuals. Additionally, the majority of them with no dependence seem to be churning.

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

**Density Plots**

Density plots for monthly and total charges gives the plot between density of churning and the monthly and total charges. There is high churn at lower total charges and churn is high when Monthly Charges are high.

Chart, histogram

Description automatically generated Chart, histogram

Description automatically generated

**Heat Map**

There are many different sorts of data that may be visualized using heat maps, including scientific, financial, and geographic data.

A picture containing qr code

Description automatically generated

In the heat map, the columns for Total Revenue and Total Long-Distance Charges are highly correlated. We can deduce that long distance calls significantly increase the company's revenue. The number of months that a client has been a customer contributes to total revenue since greater package usage generates more revenue for the business. The quantity of referrals appears to have a small impact on overall earnings.

**Churn Analysis from Tableau**

We may first claim that business is bleeding in San Diego. There are a lot of churning numbers there. The good news is that marketing campaigns are showing great results in LA, the largest market. There are a lot of new consumers joining in LA. Data usage is high among the teenagers, so introducing more internet plans for students and teenagers would be a great idea for decreasing the churning in that age group. Churning is low when the customers opt for long contracts. So, it is advised to upsell the contracts to the customers.

Chart

Description automatically generated

# Predictive Analytics

Forecasting customer churn enables companies to take preventive measures with customers who are likely to leave, potentially repairing their relationship before they depart. This early warning system allows businesses to act proactively, and it is commonly acknowledged that retaining a customer is less expensive than acquiring a new one. Consequently, identifying and preventing customer churn is preferable to losing them and searching for a replacement.

Here a classification model is used to distinguish between churn and non-churn customers. Out of the different classification models present, the best technique is evaluated for accurate prediction.

When it comes to model building there is a to do process before fitting the data into the model. As the target variable is binary type variable, we have chosen classification models like logistic regression, decision tree classifier, Random Forest classifier, naïve bayes and support vector machine.

Here is the list of steps involved in building the model.

1. Feature selection: The next step is to identify the key features that are most predictive of the target class. This can involve using techniques such as correlation analysis and feature importance ranking to identify the most important features.
2. Data splitting: The data is split into a training set and a test set. The training set is used to train the model, while the testing set is used to evaluate its performance.
3. Model selection: The next step is to select a suitable classification model for the analysis.
4. Model training: The training data is run through the model following the specific algorithm.
5. Model evaluation: After training the model, it is run through testing data to assess its efficiency. This involves factors like accuracy, precision and F1 score.

## Model Implementation

**Logistic Regression**

Table

Description automatically generated

The accuracy of the model is 86.04%, and the confusion matrix shows the number of true positives, true negatives, false positives, and false negatives. The precision of the model is 90.07%, the recall is 90.27%, and the F1 score is 90.17%. These metrics provide a measure of the model's accuracy and how well it can predict positive cases while avoiding false positives and false negatives.

**Decision Tree**

Table

Description automatically generated with medium confidence

The model demonstrated a high level of accuracy, correctly predicting outcomes for 94.54% of the test data. The confusion matrix revealed that the model accurately predicted 342 true negatives and 904 true positives, while making 41 false positives and 31 false negatives. The precision score, indicating the percentage of true positive predictions out of all positive predictions, was 95.66%. The model's recall, indicating the percentage of actual positive cases correctly identified, was 96.68%. The F1 score, which combines precision and recall, was 96.17%, indicating that the model performed well overall.

**Random Forest**

A picture containing text

Description automatically generated

The model exhibited an impressive accuracy rate of 96.36%, signifying that it made correct predictions for a substantial portion of the test data. The confusion matrix indicated that the model accurately predicted 338 true negatives and 932 true positives, but made a few erroneous predictions, namely 45 false positives and 3 false negatives. The model's precision score, reflecting the proportion of true positive predictions among all positive predictions, was a respectable 95.39%. Its recall, representing the percentage of actual positive cases accurately identified by the model, was an outstanding 99.68%. The F1 score, a metric that balances precision and recall, was a remarkable 97.49%, attesting to the model's exceptional performance.

**Naïve Bayes**

Text

Description automatically generated with medium confidence

The model has an accuracy of 94.39%, indicating that it predicted the outcome correctly for a high percentage of the test data. The confusion matrix shows that the model predicted 326 true negatives and 918 true positives, while incorrectly predicting 57 false positives and 17 false negatives. The precision of the model is 94.15%, which means that out of all the positive predictions made by the model, 94.15% were actually true positives. The recall is 98.18%, which means that out of all the actual positive cases in the test data, the model correctly identified 98.18% of them. The F1 score is 96.13%, which is a measure of a model's accuracy that considers both precision and recall, and suggests that the model has good performance.

**Support Vector Machine**

Table

Description automatically generated with medium confidence

## The model's accuracy rate of 76.63% indicates that it accurately predicted outcomes for a moderate portion of the test data. According to the confusion matrix, the model correctly predicted 162 true negatives and 848 true positives, while making 221 false positives and 87 false negatives. The precision score of the model, which measures the percentage of true positive predictions among all positive predictions, was 79.33%, showing that the model performed reasonably well. Its recall score, representing the percentage of actual positive cases correctly identified by the model, was 90.70%, reflecting a good performance in detecting positive cases. The F1 score, which is a combination of precision and recall, was 84.63%, indicating a moderate level of accuracy for the model.

## Model Comparison

Chart, bar chart

Description automatically generated

The models are evaluated by comparing each classification model by plotting the bar plot with accuracy.

## Model Selection

After Comparing the accuracy of each model, three models are performing well. Based on our business problem we have chosen Random Forest Classifier.

RF classifier is the best fit because it Handles Unbalanced Data better than other models and it is Robust to Outliers and Non-linear Data. We even plotted the confusion matrix and from which we can see balanced True Positive and True Negative values with great precision value.

Chart

Description automatically generated

Test Data Results

**Accuracy: 96.4339908952959 Confusion Matrix: [[340 43] [ 4 931]] Precision: 95.58521560574948**

# Recommendations

To reduce customer churn, we recommend implementing the following plans for Maven Telecom based on our analysis of data using visualization techniques. Our recommendations are focused on four key areas: Contract Types, Churn Reasons, Regions with high Churn rate, and Data Plans as per Age group.

We are referring to Tableau dashboard for these recommendations:

**Contract Types:**

The data analysis shows that the churn rate is higher in month-to-month contracts than in 1-year or 2-year contracts. This indicates that customers may be more likely to switch to a competitor if they are not committed to the service for a longer period. Therefore, we recommend that Maven Telecom introduce 3-month contracts as initial phone plans, rather than monthly plans. This will give customers more time to experience the telecom services and may encourage them to stay with the network.

**Churn Reasons:**

Based on feedback from churned customers, many cited competitors as the main reason for leaving the service. Therefore, we recommend that Maven Telecom conduct a market analysis of competitor phone plans and devise new marketing strategies to promote its own plans with additional services to attract customers. For example, the company could offer free trials of its services or discounts on bundled services like TV and internet. Additionally, we suggest improving support services by hiring skilled professionals to address customer issues. This will make customers feel more supported and reduce their likelihood of leaving the service.

**High Churn Rate Regions:**

Certain regions have a higher churn rate than others, and these regions contribute to a large portion of the total churned population. To address this, we recommend introducing and promoting new plans in these regions, including student discounts since teenagers are predominant in these areas. The company could also run targeted marketing campaigns in these regions to attract more customers.

**Data Usage:**

Our analysis shows that teenagers consume more data than adults and senior citizens, and they are more likely to churn due to low data plans offered. Therefore, we recommend that Maven Telecom implement new data plans based on average usage for different age groups. This will provide more customized options for customers and make them less likely to switch to a competitor with more attractive data plans.

In summary, we recommend that the Telecom company in California to introduce 3-month contracts, conduct a market analysis of competitor phone plans, improve support services, introduce and promote new plans in high-churn regions, and implement new data plans based on average usage. Implementing these recommendations will likely improve customer retention and reduce churn rates.

# Conclusion

After collecting the data from the customers about the reasons for their churning from the telecom. Using different project management tools like Kanban and DMAIC, the data was cleaned and modified to achieve the accuracy in the prediction modelling after analysing the data using descriptive and exploratory data analysis. To decrease the churn rate the Maven telecom company must focus on increasing the contract tenure, focus on the customer satisfaction, and introducing more internet/data plans for teenagers and students. Following the recommendations provided would help the Maven telecom company to retain their customers and decrease the churn rate.

# Project Access

1.[Google Colab](https://colab.research.google.com/drive/1fKRAZQrx_FNVI84FCfd-_WIFbjeJV0UI#scrollTo=8AGXayliBa_T)

2.[Tableau Public](https://public.tableau.com/authoring/ChurnAnalysis_16805354186840/Dashboard1_1#1)

# 

# References

1. Mariani, P., Taneja, S., & Tripathi, S. (2019). A comprehensive study of customer churn prediction in the telecom sector using various machine learning techniques. Journal of Big Data, 6(1), 1-27. doi: 10.1186/s40537-019-0191-6.
2. Elmisery, A. M., El-Henawy, I. M., & Elmahdy, N. M. (2021). Predicting customer churn in the telecom industry using machine learning techniques: A systematic literature review. International Journal of Advanced Computer Science and Applications, 12(1), 331-338.
3. Burez D, den Poel V. Handling class imbalance in customer churn prediction. Expert Syst Appl. 2009;36(3):4626–36.
4. Maven Analytics Data Playground https://www.mavenanalytics.io/data-playground?search=tele