Customer Churn Analysis,
Clustering, and Prediction Model Development

June 2024

Content

Project Overview

The goal of this project is to effectively reduce customer churn and increase revenue. The project will present an end-to-end data analysis process, including pinpointing business challenges, conducting data analytics, developing machine learning models to increase effectiveness, providing actionable recommendations, and creating a financial forecasting model.

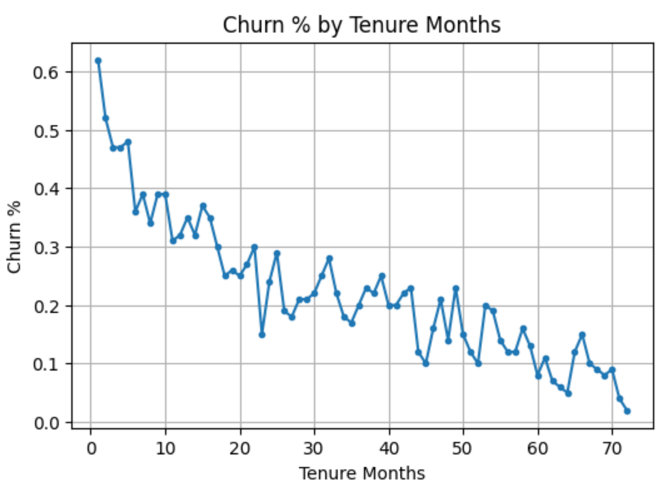
The project began with a customer churn analysis, uncovering insights that customers with short-term contracts and those using fiber optic internet tend to churn at higher rates. Based on these insights, I used K-means clustering to build a customer segmentation model, effectively dividing the existing customers into four groups with distinct characteristics that align with the analysis results. Additionally, I developed a machine learning prediction model that can identify churn customers with a 77% recall rate. Combining these analysis insights and machine learning labels, I created short-term and long-term recommendations for improving customer retention and increasing revenue.

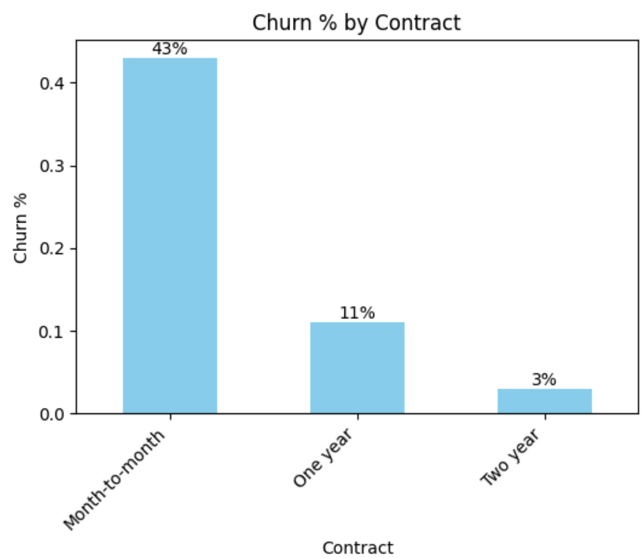
Customer Churn Analysis

The total churn rate is 26.54%, which is slightly lower than the average churn rate in the U.S. telecom industry (30% to 35%). Given this, there's a significant opportunity to address the churn issue early on through strategic planning and proactive measures.

Who Churns?

1. **Customers on short-term contracts tend to churn at higher rates.**
   Customers with shorter-term relationships with the company tend to churn more compared to those with longer-term relationships. As the line chart on the left, customers with less than two months of tenure have a churn rate exceeding 50%. Furthermore, the bar chart on the right shows that customers on a month-to-month contract have a 43% churn rate, which is much higher than that of customers with yearly contracts.





2. **Monthly contract customers used fiber optic internet or lacked add-on services churned the most**
   55% of our customers are on month-to-month contracts, and 43% of them have churned.
   Monthly contract customers who churn exhibit some traits compared to those who didn't churn:
   - 70% of them used fiber optic internet service, compared to 44% of the customers who didn't churn.
   - 70% to 80% of them didn't have add-on services, compared to 40% to 60% of the customers who didn't churn.
     - Add-On Services include Tech Support, Online Security, Online Backup, or Device Protection.

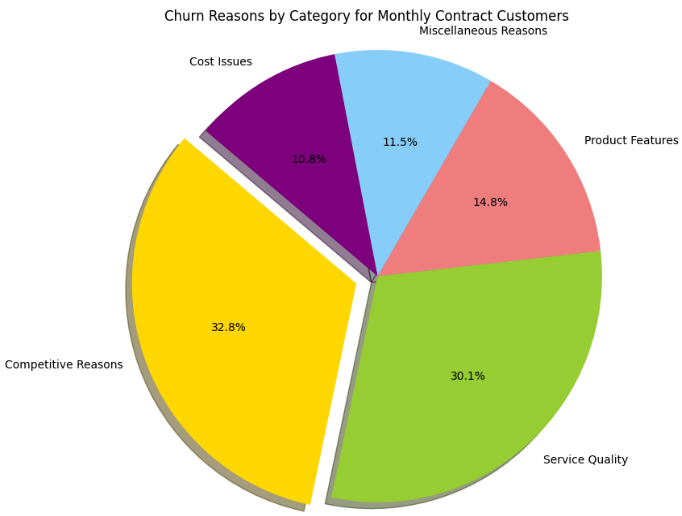
Monthly Contract Customers Trait	Churned Customers	Non-Churned Customers
Used Fiber Optic Internet	70%	44%
Lacked Add-On Services*	70%-80%	40% - 60%

Why Churn?

Competition and service quality dissatisfaction are top reasons for customer churn

32.8% of monthly contract customers churn due to switching to competitor services, and 30.1% churn due to dissatisfaction with service quality.

(The pie chart on the left-hand side categorizes churn reasons, while the table on the right-hand side details these reasons.)



index	Churn Reason	Count	Percentage
0	Attitude of support person	174	10.51
1	Competitor offered higher download speeds	161	9.73
2	Don't know	141	8.52
3	Competitor offered more data	140	8.46
4	Competitor made better offer	126	7.61
5	Competitor had better devices	116	7.01
6	Attitude of service provider	115	6.95
7	Network reliability	92	5.56
8	Price too high	89	5.38
9	Product dissatisfaction	85	5.14
10	Lack of self-service on Website	84	5.08
11	Service dissatisfaction	83	5.02
12	Extra data charges	53	3.2
13	Moved	43	2.6
14	Limited range of services	40	2.42
15	Long distance charges	37	2.24
16	Lack of affordable download/upload speed	36	2.18
17	Poor expertise of online support	18	1.09
18	Poor expertise of phone support	16	0.97
19	Deceased	6	0.36

For detailed coding of the churn analysis, please refer to my GitHub repository here.

Customer Segmentation

Now, let's leverage the insights that we gained from the above analysis with machine learning techniques to segment our customers for a deeper understanding of customer churn.

Customer Segmentation using K-means Clustering

Step 1: Data Cleaning & Preprocessing

- Retain only the important factors and remove rows with null values.
- Combine the columns Online Security, Online Backup, Device Protection, and Tech Support into one column named 'add-on service,' where having any of these services will be marked as 1, and none as 0.
- Encode categorical variables into dummy variables.

Step 2: Determine the Number of Clusters

From the Elbow Method, the optimal number of clusters is 4, which is the "elbow" representing the point where adding another cluster does not significantly improve the fit.

Step 3: Interpreting the Clusters

From the K-means clustering, we identified 4 clusters with distinct traits. Clusters 1 and 2 represent short-tenure customers, while Clusters 3 and 4 are long-tenure customers. We also notice that Clusters 2 and 3 have lower add-on services and no fiber optic internet service, leading to lower monthly charges for these customers. Clusters

1 and 4 comprise nearly 60% of our customer base and contribute nearly 80% of our monthly revenue. However, Cluster 1 has the highest churn rate, about 49%.

Cluster	Traits	Churn Rate	Tenure Months	Monthly Charges	Add-on service	Fiber optic	Number of Customers	Percentage of Customers	Proportion of Monthly Rev
1	Short Tenure, High Charges	49%	14.77	81.09	64%	78%	2,186	31%	39%
2	Short Tenure, Low Charges	25%	10.59	32.64	30%	0%	1,745	25%	12%
3	Long Tenure, Low Charges	5%	54.17	33.96	40%	0%	1,153	16%	9%
4	Long Tenure, High Charges	16%	58.58	93.28	96%	71%	1,959	28%	40%

* Monthly Rev = Monthly Charges * Number of Customers

Table: Cluster Summary of Mean Values for Each Cluster

For detailed coding of the K-means clustering algorithm, please refer to my GitHub repository here.

Analysis Insight & Conclusion

Customer Churn Prevention Insight

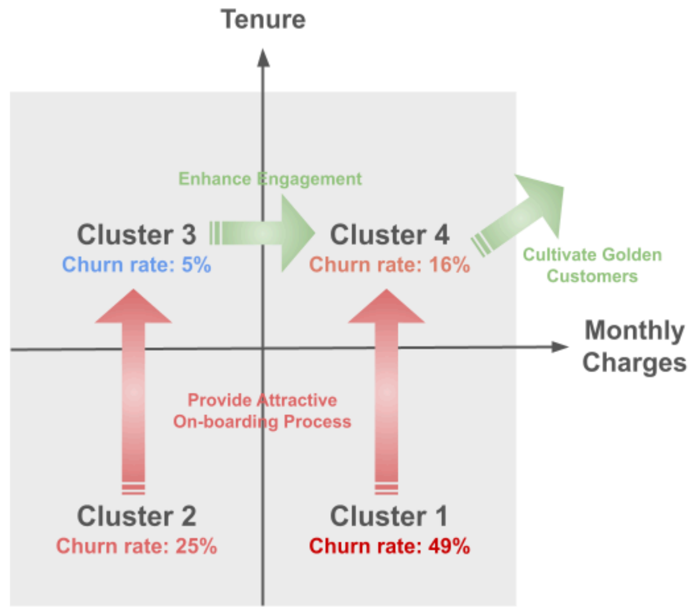
From churn analysis and clustering, two insights emerge:

1. Short-tenure customers churn more than long-tenure customers.
2. Customers with fiber optic internet services (high charges) churn more than those without.

Therefore, it is crucial to **implement tailored strategies for different customer clusters:**

- **Clusters 1 and 2:** Optimize the onboarding process for short-tenure customers and prioritize their transition to long-tenure status. For those customers who are about to churn, offer them retain incentive. For example, provide discounts on future service upgrades or exclusive access to new features for a limited period. Additionally, personalize the offer based on their usage patterns or feedback to increase its effectiveness in retaining their subscription.
- **Cluster 3:** Despite having the lowest churn rate, they contribute only 9% of the monthly revenue. To enhance engagement, consider offering incentives like a 3-month trial of optic internet exclusively for long-tenure customers, potentially moving them towards cluster 4 (higher charges).
- **Cluster 4:** These "golden" customers exhibit lower churn rates and higher charges. Provide them with personalized customer support and services, including tailored recommendations based on their optic fiber internet usage to optimize their plans for better cost-effectiveness.

Implement Tailored Strategies for Different Customer Clusters



Moreover, **enhancing fiber optic internet service quality and customer services** are essential to prevent churn. For short-tenure customers, introduce a program for them to trial add-on services, shifting their focus beyond fees and internet quality to include better service support and other value-added services. These strategies collectively aim to effectively prevent churn across diverse customer segments.

Financial Forecasting

Potential Revenue Gains & CLV Incremental from Churn Prevention

Clusters	Monthly Revenue ($)			Customer LifeTime Value (CLV) ($)		
	Original	10% Reduction in Churn Rate	Optimized Churn Reduction & Customer Segmentation*	Original	10% Reduction in Churn Rate	Optimized Churn Reduction & Customer Segmentation*
1	90,404	99,090	89,181	1,335,267	1,538,967	1,460,479
2	42,718	44,142	39,727	452,379	479,445	443,486
3	37,198	37,394	39,314	2,015,020	2,035,805	2,073,484
4	153,498	156,422	183,083	8,991,903	9,312,172	10,362,803
Total	323,818	337,047	351,305	12,794,570	13,366,389	14,340,253
Incremental Percentage**	-	**4.09%**	**8.49%**	-	**4.47%**	**12.08%**

* Optimized Churn Reduction & Customer Segmentation: This strategy involves a 10% reduction in the churn rate and moving 10% of active customers to their ideal clusters (Cluster 1 to 4, Cluster 2 to 3, Cluster 3 to 4).

** Incremental Percentage: : Calculated as (New Value - Original) / Original

Potential Revenue Gains from Churn Prevention

Based on our forecasting calculations, if we successfully reduce churn by 10% in each cluster and move 10% of customers in each cluster to the next engagement level, we could **potentially increase monthly revenue by 4% to 8% and Customer Lifetime Value by approximately 4% to 12%.** Before fully implementing these strategies, conducting A/B testing to gauge the potential churn reduction rates in each cluster and gathering additional information—such as effective marketing content, channels, and optimal timing—will better prepare us.

Action Plan Recommendations

Short-term

I recommend developing a machine learning prediction model to effectively predict churn customers. By integrating this prediction with our existing customer segmentation, we can devise more personalized retention strategies. Conduct A/B testing to refine these strategies and prevent customer churn. Evaluate the A/B testing results to enhance forecasting accuracy for potential ROI and adjust future plans based on insights gained from the testing phase.

ML Prediction Model Development

I've developed a model with a 76% overall prediction accuracy rate and a 77% recall rate. Given the primary goal of identifying customers at risk of churn, I prioritized recall as the main metric. This approach aims to capture as many actual churners as possible, despite potentially higher false positive rates(customers who do not churn but are predicted to churn).

For detailed coding of the machine learning prediction model, please refer to my GitHub repository here.

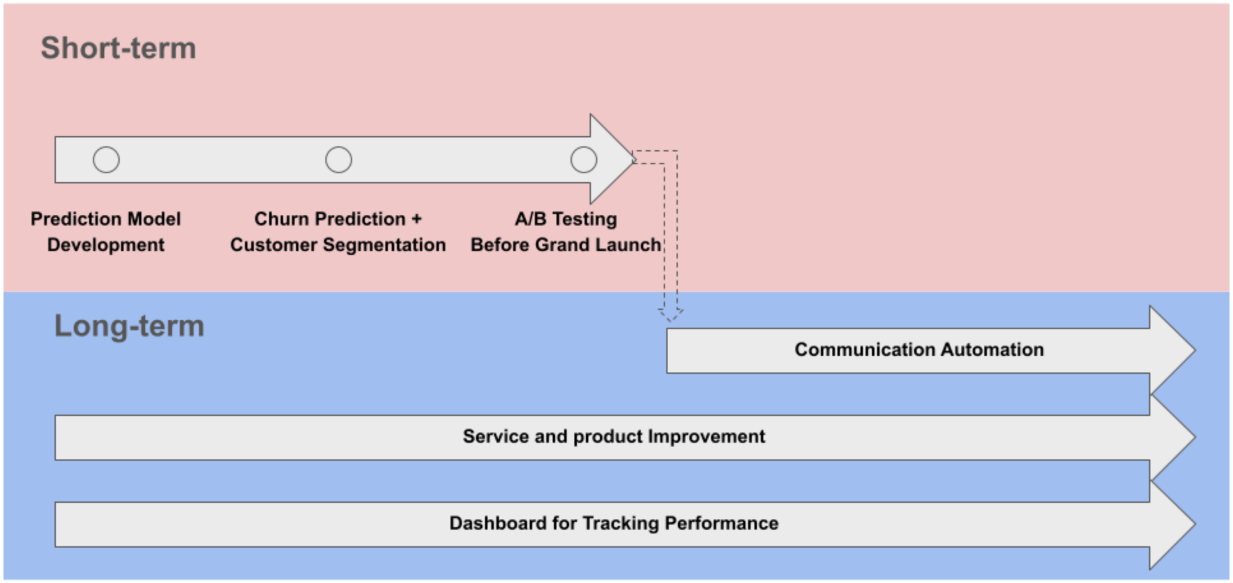
Churn Prediction + Customer Segmentation

With these two powerful machine learning labels, we can effectively identify customers about to churn and use personalized marketing content based on their segment's behavior. This not only allows for better use of our marketing budget and resources but also provides customers with more attractive incentives.

A/B Testing Before Grand Launch

To ensure wiser investments and future automated communication with our customers, we can conduct A/B testing on marketing content, timing, and offers for different customer segments before launching the long-term customer churn project.

Action Plan Recommendations



Long-term

From our churn analysis results, around 65% of the churn reasons are due to competition and service quality dissatisfaction, indicating that improving customer satisfaction is essential. This should be considered a long-term optimization plan that requires continuous improvement.

At the same time, we can utilize A/B testing results to refine our customer churn prevention plan, integrating it into the marketing automation process to make our resources more effective. Additionally, we should build a dashboard to monitor the performance of these automated processes. The dashboard will enable stakeholders to identify problems and take immediate action when a decline is detected.

Project Conclusion

This project provides an end-to-end framework for reducing customer churn and enhancing revenue by combining data analysis with machine learning tools and developing tailored strategies for different customer segments. Implementing these strategies with a real-time tracking dashboard will ensure continued effectiveness and adaptability to changing customer behaviors. Ultimately, this project lays a solid foundation for improving customer retention and driving revenue growth through data-driven insights and proactive customer management.