CUSTOMERS CHURN ANALYSIS

GREETINGS

Good Day, It's Pernel.

We will be discussing customer churn within Databel's telecom customer base, a topic crucial for understanding customer retention strategies.

I hope that each reader will enjoy reading this report as much as I enjoyed creating it.

STAY HYSINGHTS !

PLAN

About the Project
Project Workflow
Recommendations

Mars 2024

ABOUT THE PROJECT

Churn refers to the point at which a customer quits using or breaks up with a company. It extends to a large range of sectors from Telecoms, Cable TV or SaaS. When a company can anticipate churn, it is in the position to proactively take action to prevent the event and keep those customers that are most profitable.

This is the case for **Databel**, a data services company that serves over six thousand customers state-wide. Databel has operated in the market for five years and has seen considerable growth. Nevertheless, the rise in competition and changing consumer preferences have led to worries regarding the churn rate, which reflects the proportion of subscribers who terminate their memberships on a monthly basis.

Our analysis was not strictly to uncover those factors, but also to understand whatever we can do to avoid churn and drive higher customer loyalty so we can ensure that in the future we are able to easily predict churn. We can then use these insights based on data to make the necessary changes in-house to keep more of our customer base.

PROJECT WORKFLOW

ABOUT THE DATASET

We are working with a dataset of **6,687 entries**, which includes **29 features** encompassing both numerical and categorical data. These features offer deep insights into customer behavior, with key variables such as age, region, contract type, and churn reasons. Additionally, we have granular details on customer service calls, data plans, and additional charges.

From a quick analysis, we've observed that **26.86%** of the customer base has churned. This represents a significant proportion, and it's essential that we explore the underlying factors contributing to this churn.

We have found no duplicate entries; however, it is important to highlight that the Churn Category and Churn Reason contain some missing data. This absence of information constitutes a significant portion of the data pertaining to individuals who are not classified as churners, which is understandable.

Please Take a look at the metadata file to explore all the features

ANALYSIS & VISUALIZATION

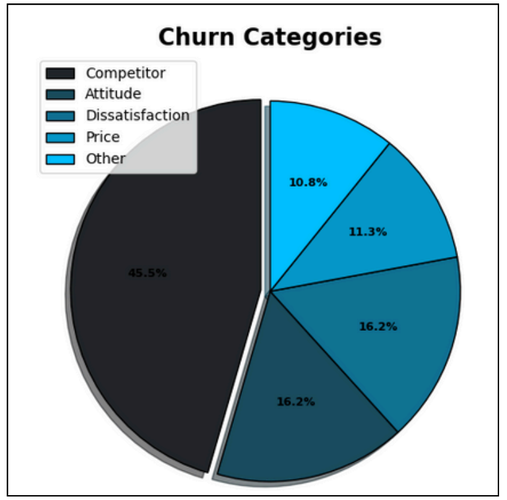
Understanding the factors driving customer churn is essential for enhancing both satisfaction and retention.

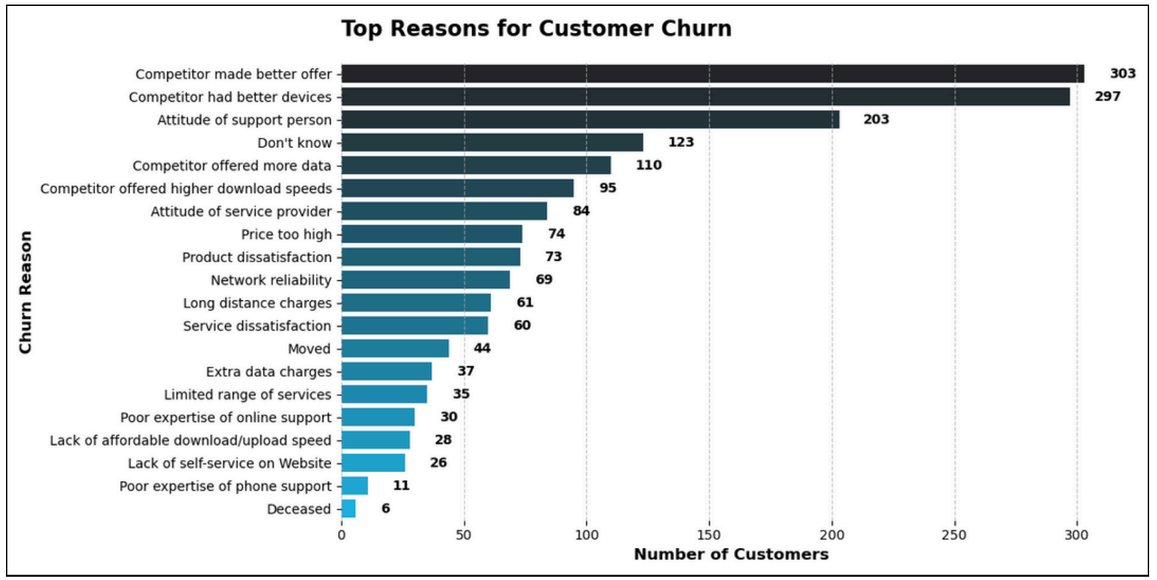
The 3 top reasons of customers churn are :

- **Competitor made better offer**
- **Competitor has better devices**
- **Attitude of support person**

The pie chart showcases the primary reasons behind customer churn, with competitor-related factors emerging as the most significant (45.5%).

Additionally, attitude and dissatisfaction are responsible for 16.2% of churn cases, signaling that improvements in customer service and experience are crucial.

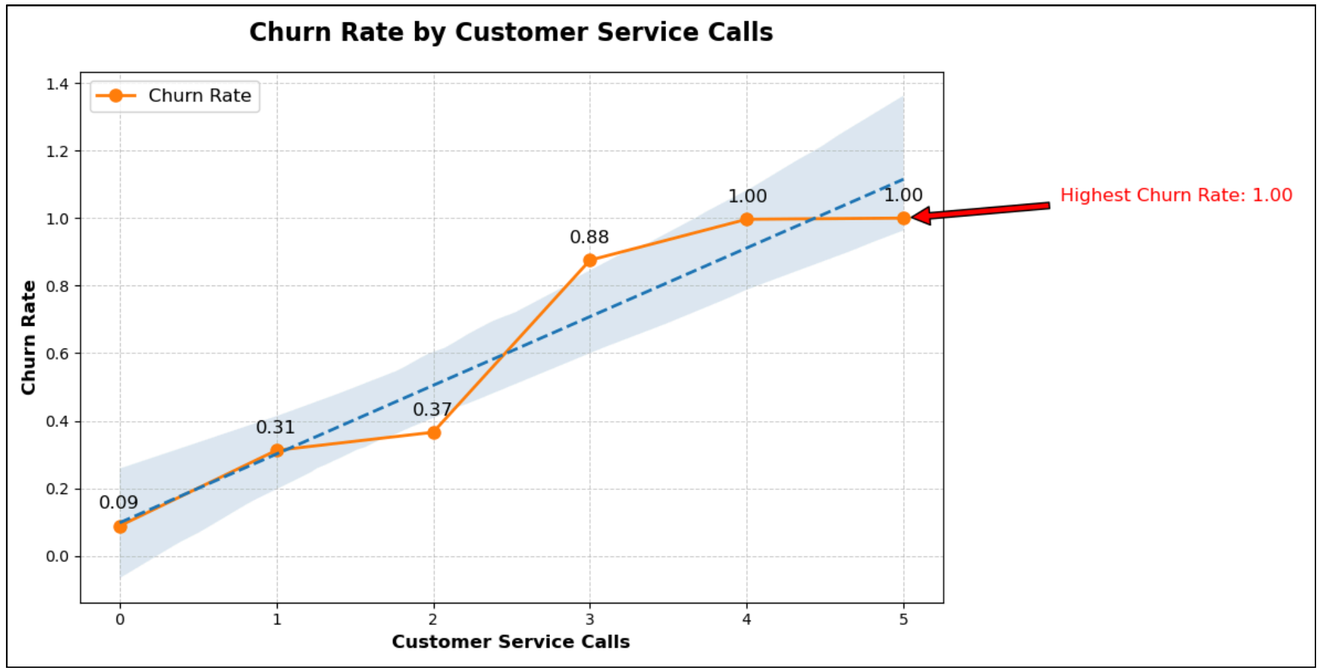




ANALYSIS & VISUALIZATION

The chart highlights a **strong positive correlation** between the number of customer service calls and the churn rate, indicating that customers who frequently seek support are more likely to churn.

Notably, the churn rate increases sharply after just two calls, suggesting that customer service interactions play a critical role in driving churn.

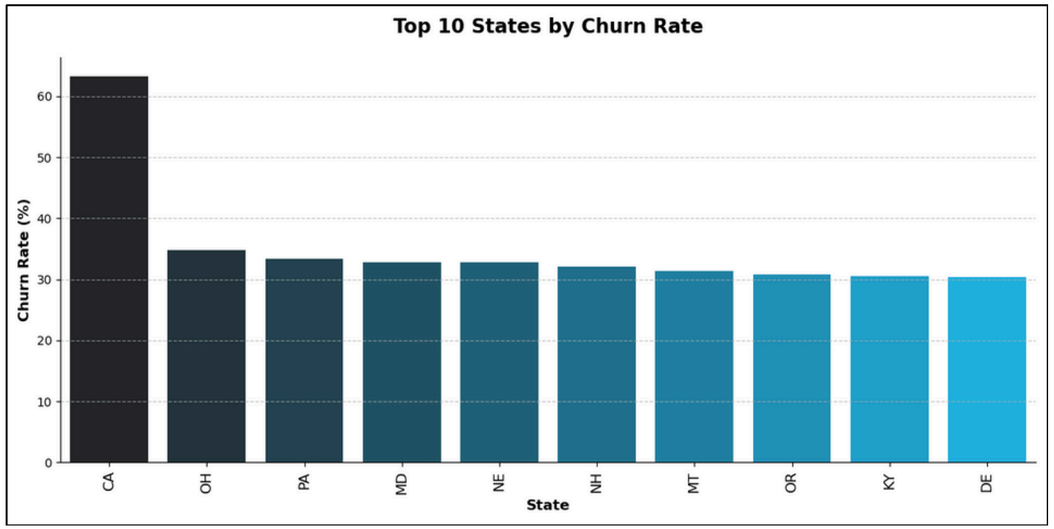


This trend underscores the potential impact of customer service quality on retention, with frequent calls possibly reflecting underlying dissatisfaction that leads to higher churn rates. Improving the customer service experience could therefore be key to reducing churn.

ANALYSIS & VISUALIZATION

Understanding regional variations in churn rates is essential for optimizing operations and enhancing customer satisfaction.

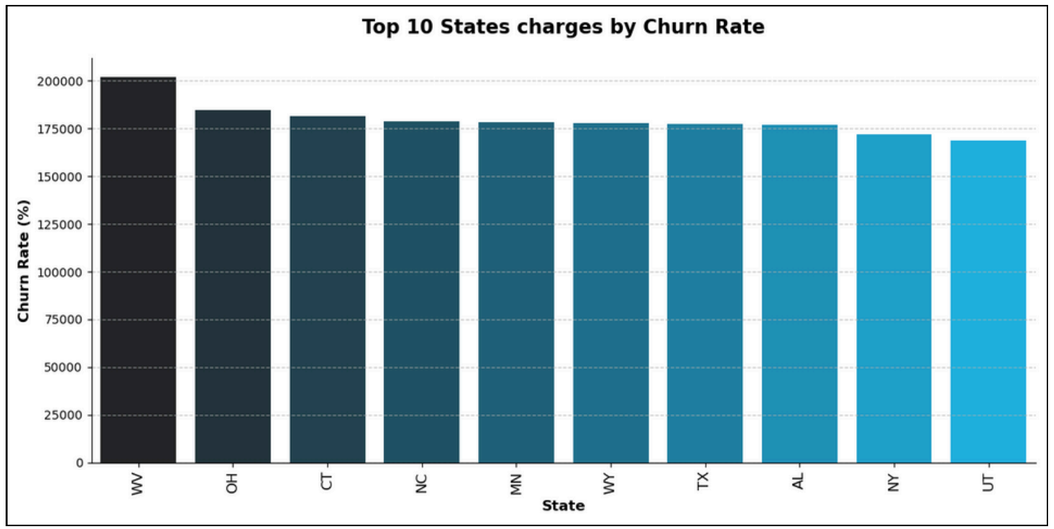
This bar chart highlights the top 10 states with the highest churn rates, with **California** standing out as the leader, significantly surpassing other states. This suggests an urgent need for a deeper investigation into the specific factors contributing to churn in California.



Additionally, most of the states with elevated churn rates are concentrated in the Eastern U.S., pointing to potential regional trends influenced by factors such as economic conditions, demographic differences, or unique market dynamics that may require targeted strategies.

ANALYSIS & VISUALIZATION

This bar chart showcases the top 10 states with the highest charges and churn rates. Interestingly, California is neither first nor on the list, indicating that high churn in this region may not be driven by pricing factors.



Most states with elevated charges and churn rates are concentrated in the **Eastern U.S.**, suggesting potential regional patterns. States such as **Ohio**, **Connecticut**, **North Carolina**, and **Michigan** show a relatively consistent relationship between charges and churn rates, pointing to a more balanced dynamic in these areas.

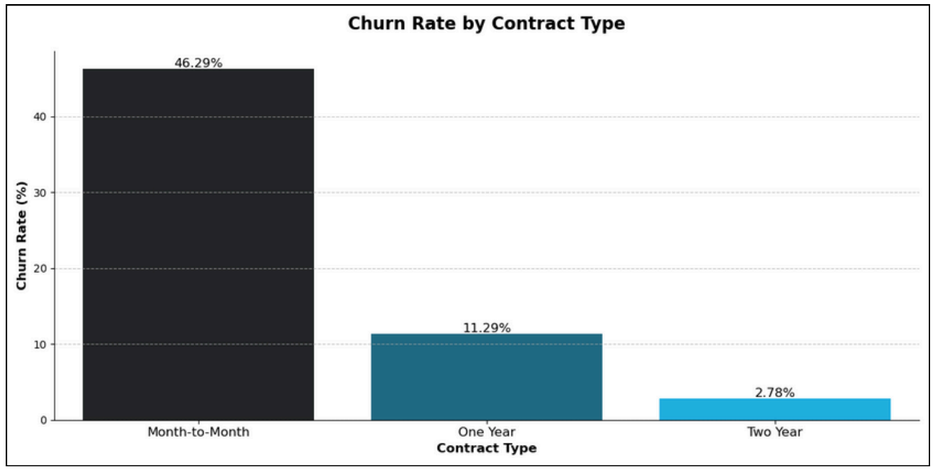
In contrast, **Western states** like Wyoming, Texas, Alabama, New York, and Utah exhibit lower charges and churn rates compared to their Eastern counterparts, hinting at differences in pricing strategies, customer satisfaction, or service quality across regions.

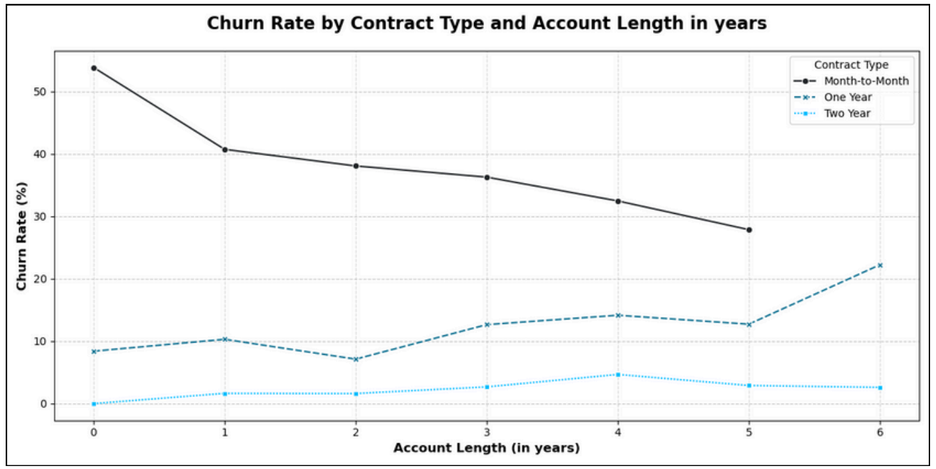
ANALYSIS & VISUALIZATION

Understanding the relationship between contract type, account length, and churn is important for improving customer retention strategies. **Contract type** plays a major role in churn rates, with **Month-to-Month** customers experiencing the highest churn, followed by **One-Year** customers, while **Two-Year** customers have the lowest churn rates.

Moreover, as account length increases, churn rates tend to decrease across all contract types, indicating that customer loyalty strengthens over time. **Two-Year contracts** are especially effective in minimizing churn, even in the early stages of the contract, showing significantly lower churn rates.

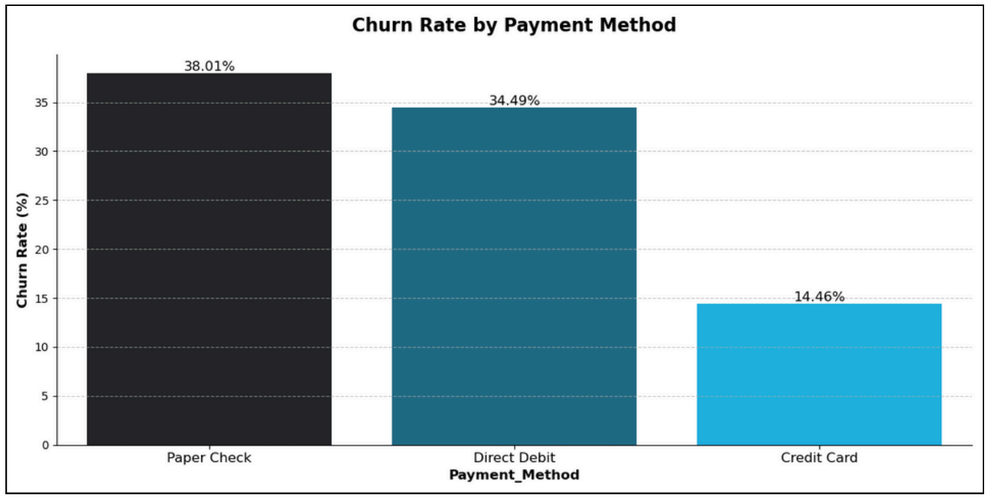
In contrast, we need to review our retention strategies **Month-to-Month customers** particularly in the initial phase of their relationship with the company.





ANALYSIS & VISUALIZATION

Grasping the relationship between payment methods and churn is essential for optimizing billing practices and improving customer satisfaction. The bar chart highlights critical insights into the effects of payment methods on churn rates.



Customers who choose **Paper Check** as their payment method demonstrate the highest churn rate at 38.01%, which may indicate issues like processing delays, inconvenience, or higher fees. Additionally, **Direct Debit** is also a notable factor in customer churn.

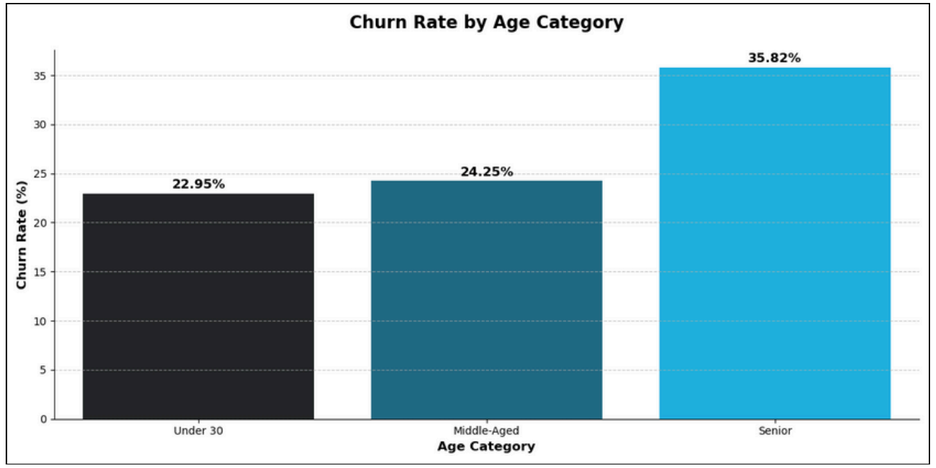
On the other hand, **Credit Card** payments are likely to offer superior convenience, making them a preferred option for customers and potentially leading to lower churn rates.

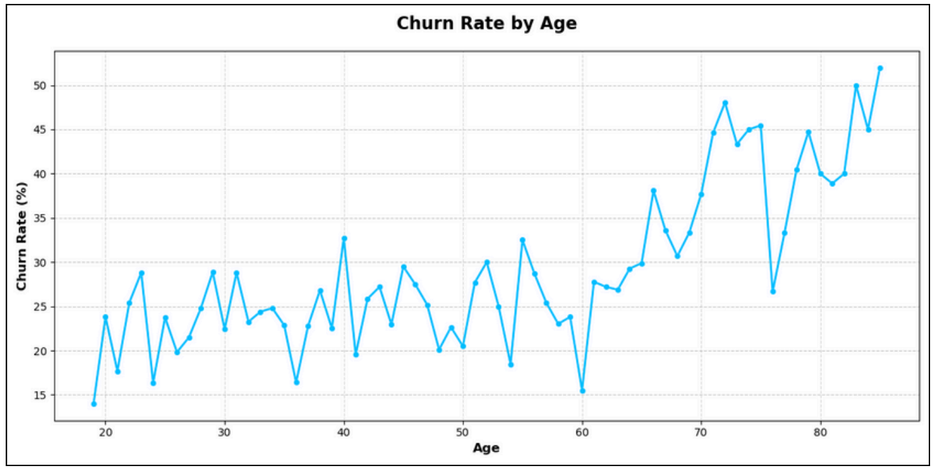
ANALYSIS & VISUALIZATION

To understand the impact of age on churn rate, we categorized customers into three groups: **Under 30**, **Middle-Aged (30-60)**, and **Senior (60 and older)**.

The chart shows that age is a significant factor influencing churn. Senior customers have the highest churn rate at 35.82%, followed by Middle-Aged customers at 24.25%.

Younger customers are less likely to churn, possibly due to greater engagement with technology or differing expectations from service providers.

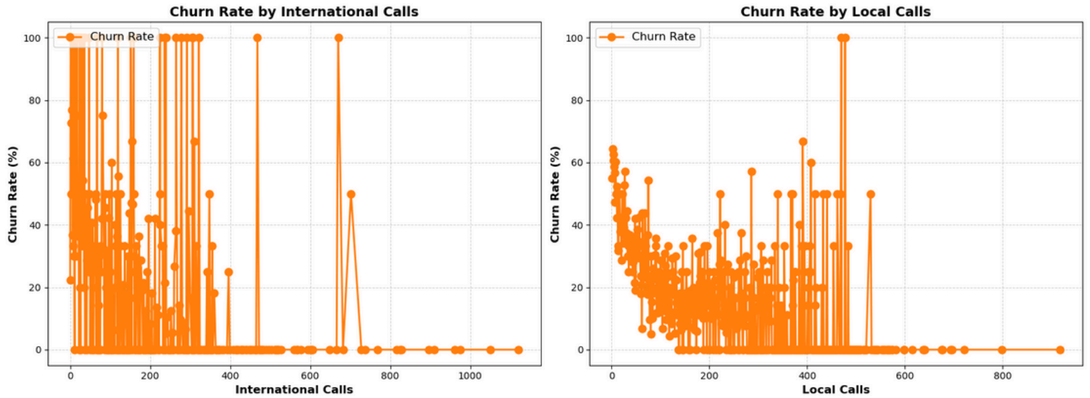


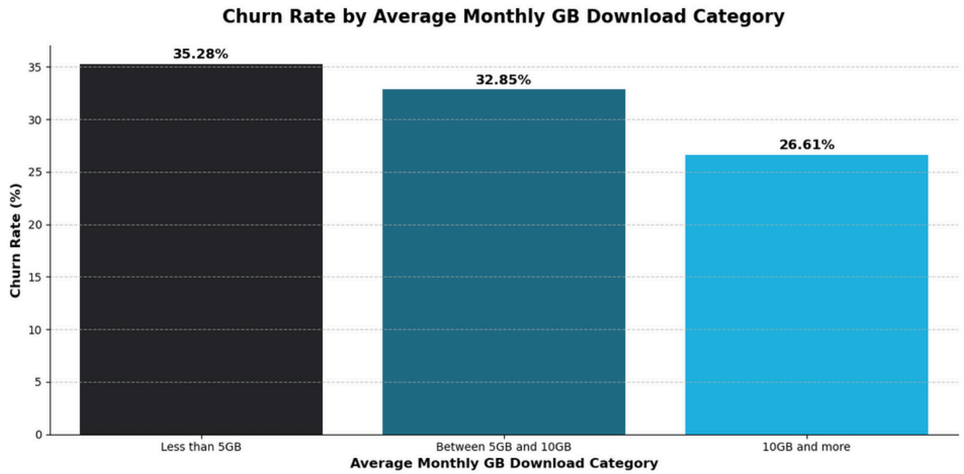


Analysis & Visualization

The impact of call volume in churn rate reveals no strong correlation between the number of international calls and churn rate.

There is a slight inverse correlation between the number of local calls and churn rate. This suggests that customers who make more local calls may be less likely to churn.



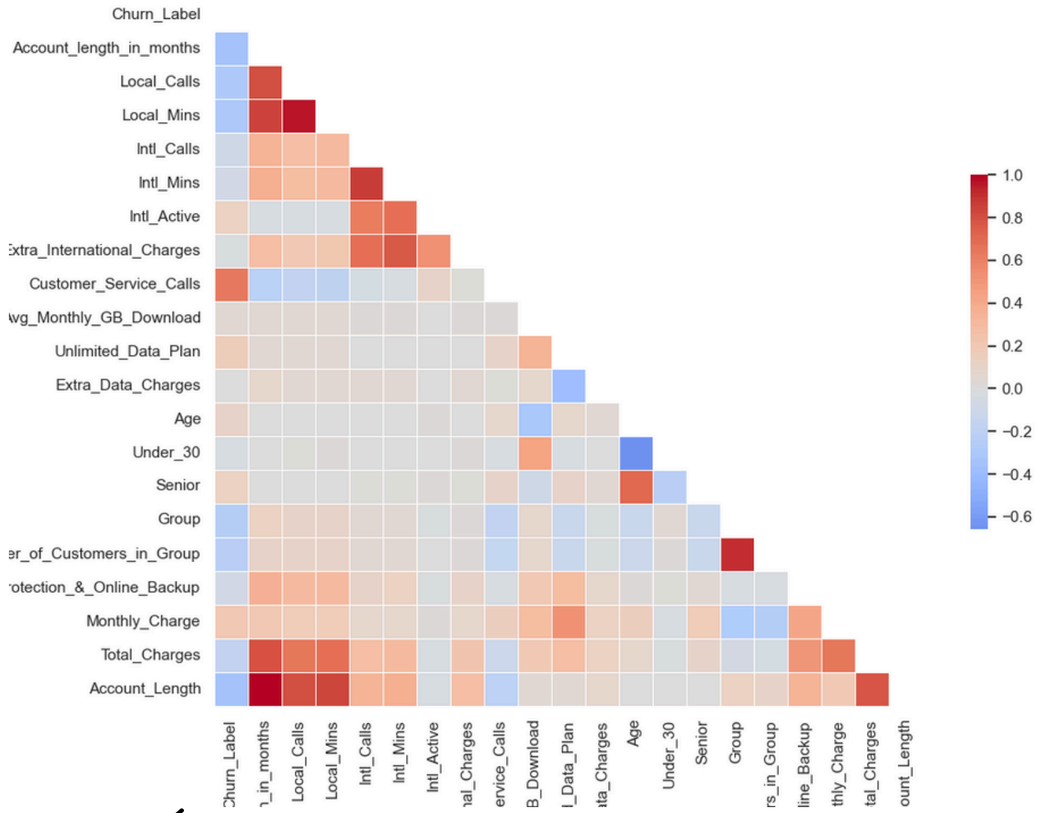


We categorized Average Monthly GB Data usage into three groups: **Less than 5GB**, **5GB to 10GB**, and **10GB or more**.

The bar chart shows that customers who use less than 5GB per month have the highest churn rate at 35.28%, suggesting they may be less satisfied with the service or have found more suitable alternatives. Customers who use 10GB or more per month have the lowest churn rate at 26.61%, which implies greater satisfaction or stronger reliance on the service, making them less likely to churn.

ANALYSIS & VISUALIZATION

The correlation matrix is an important visualization that offers valuable insights into churn and other various customers attributes interdependencies.



Keys points

- A **moderate negative correlation** between account length and churn, indicating that customer loyalty grows over time
- A **moderate negative correlation** between unlimited data plans and churn indicates that offering unlimited data may improve customer retention
- A **moderate negative correlation** between contract type and churn shows that longer-term contracts help reduce churn
- A **weak positive correlation** between total charges and churn suggests that pricing may play a role in churn decisions
- A **weak positive correlation** between age and churn, suggesting older customers might be slightly more likely to churn, warranting further investigation
- A **moderate positive correlation** between customer service calls and churn highlights the importance of promptly addressing service issues

Recommandations

REFINE SERVICE VALUE PROPOSITION

Customer Feedback: Gather detailed feedback from churned customers, especially those using less than 5GB of data and those who prefer Paper Check as a payment method. Investigate pain points and explore ways to offer better-suited packages or services.
Differentiation Strategy: Analyze competitors, particularly in regions like California and Eastern U.S. states, and adjust the service offering to stand out.

REEVALUATE PRICING & REGION

Region-Specific Pricing: Consider region-based pricing strategies. Since high churn is prominent in certain states, like those in the Eastern U.S., adjusting prices to match local economic conditions may attract and retain more customers.
Offer Discounted Long-Term Contracts: Two-Year contracts were shown to reduce churn, so offering attractive discounts or incentives for signing long-term contracts could help decrease churn, especially in high-risk areas.

Recommandations

OPTIMIZE CUSTOMER SERVICE

Train Customer Support Teams: The strong correlation between customer service calls and churn highlights a need for better training. Focus on improving response time, empathy, and problem-solving to reduce frustration and enhance customer satisfaction.
Proactive Issue Resolution: Implement proactive customer service for customers who make more than two calls. Follow up with them and solve their problems quickly before frustration leads to churn.

ENHANCE PAYMENT METHODS AND BILLING EXPERIENCE

Improve Paper Check Experience: With Paper Check users showing the highest churn rate (38.01%), investigate issues such as processing delays, fees, or lack of convenience. Offer incentives for customers to switch to faster, digital payment methods like Direct Debit or Credit Card.
Offer Flexible Payment Plans: Introduce more payment flexibility, especially for older customers, to ensure that billing processes are convenient and easy to use.

Recommandations

TAILOR PLANS TO DATA USAGE PATTERNS

Adapt Data Plans: Customers who use less than 5GB per month have the highest churn rate (35.28%). Consider offering low-data packages that are better suited to their needs at competitive prices.
Promote Unlimited Data Plans: Since unlimited data plans are associated with lower churn, promote these plans more actively to customers as a way to retain them, highlighting the benefits of unlimited usage without overage concerns.

TARGET AND SUPPORT OLDER CUSTOMER SEGMENTS

Simplify Tech and Offer Specialized Plans: The data suggests older customers (Senior category) have higher churn rates. Offer simplified technology, additional support, and senior-friendly service plans to cater to their needs. Ensure that they feel comfortable using the services and aren't overwhelmed by complexity.
Dedicated Customer Support for Seniors: Provide tailored customer service for older customers who may have different expectations and needs from younger customers, ensuring they receive personalized assistance.

Recommandations

IMPROVE RETENTION FOR MONTH-TO-MONTH CUSTOMERS

Promote Loyalty: Month-to-Month customers show the highest churn rates, especially in the early stages. Offering loyalty programs, exclusive discounts, or free services after a few months could reduce the likelihood of churn during this critical period.

Upgrade Offers for Contract Extension: Encourage Month-to-Month customers to switch to One-Year or Two-Year contracts by offering special upgrade promotions that make the transition attractive and affordable.

Modelisation

Classifications metrics are the metrics that explains the efficiency of the model.

	Model	Accuracy	Precision	Recall	F1 Score	CROSS
0	Random Forest Classifier	90.73	0.868	0.773	0.817	89.55
2	Support Vector Machines	89.83	0.872	0.728	0.794	89.19
3	KNeighborsClassifier	87.08	0.830	0.653	0.731	86.16
1	Decision Tree Classifier	85.41	0.723	0.739	0.731	84.13
4	Naive Bayes	77.93	0.559	0.849	0.674	77.99

- **Accuracy:** The overall proportion of correct predictions
- **Precision:** The proportion of positive predictions that were actually correct
- **Recall:** The proportion of actual positive cases that were correctly predicted
- **F1 Score:** The harmonic mean of precision and recall, providing a balanced measure of both
- **CROSS:** This metric is likely the cross-validation score, indicating the model's performance on multiple training-testing splits

MODELISATION

MODEL PERFORMANCE ANALYSIS

Random Forest Classifier

- **Accuracy (90.73%):** This model has the highest accuracy, meaning it correctly predicts churn and non-churn cases 90.73% of the time
- **Precision (0.868):** Out of all the customers predicted to churn, 86.8% truly churned, indicating strong predictive power with fewer false positives
- **Recall (0.773):** It captures 77.3% of actual churners, meaning it misses some churners but balances it with precision
- **F1 Score (0.817):** The harmonic mean of precision and recall shows the model is well-balanced between catching actual churners and avoiding false alarms
- **CROSS (89.55%):** The cross-validation score is close to the accuracy, indicating the model's stability across different datasets

Conclusion: This model is the best overall, with a high ability to both identify churners and avoid incorrect predictions of churn.

MODELISATION

MODEL PERFORMANCE ANALYSIS

Support Vector Machines (SVM)

- **Accuracy (89.83%):** Slightly lower than Random Forest but still performs very well in churn prediction
- **Precision (0.872):** It performs marginally better than Random Forest in terms of precision, meaning it makes slightly fewer false-positive errors
- **Recall (0.728):** It catches 72.8% of actual churners, which is lower than Random Forest, meaning it misses more churners
- **F1 Score (0.794):** A lower F1 score compared to Random Forest indicates a slight trade-off between precision and recall
- **CROSS (89.19%):** A close match to the accuracy suggests this model is also stable during validation

Conclusion: SVM performs well, especially with high precision, but sacrifices recall, meaning it may miss more actual churners than Random Forest.

MODELISATION

MODEL PERFORMANCE ANALYSIS

KNeighborsClassifier

- **Accuracy (87.08%):** Lower than the top two models, indicating it's less effective at making correct predictions
- **Precision (0.830):** A good level of precision, meaning that out of the predicted churners, 83% are correctly identified
- **Recall (0.653):** It captures only 65.3% of actual churners, meaning it misses many real churners
- **F1 Score (0.731):** The balance between precision and recall is acceptable, but much lower than Random Forest and SVM
- **CROSS (86.16%):** Lower cross-validation score indicates that it may not generalize as well as the top models

Conclusion: This model is less effective than Random Forest and SVM, especially with a lower recall and F1 score, meaning it struggles to identify a large portion of churners.

MODELISATION

MODEL PERFORMANCE ANALYSIS

Decision Tree Classifier

- **Accuracy (85.41%):** Lower than the previous models, indicating it doesn't perform as well in overall churn prediction
- **Precision (0.723):** Lower precision means more false positives, predicting churn where it doesn't exist
- **Recall (0.739):** Slightly better than KNeighbors but still misses a significant portion of churners
- **F1 Score (0.731):** Similar to KNeighbors, showing a balance between recall and precision, but less reliable than top models.
- **CROSS (84.13%):** A lower score for cross-validation indicates it might not perform as consistently across different datasets

Conclusion: Decision Tree is moderately effective but lags behind the top-performing models in both accuracy and precision.

MODELISATION

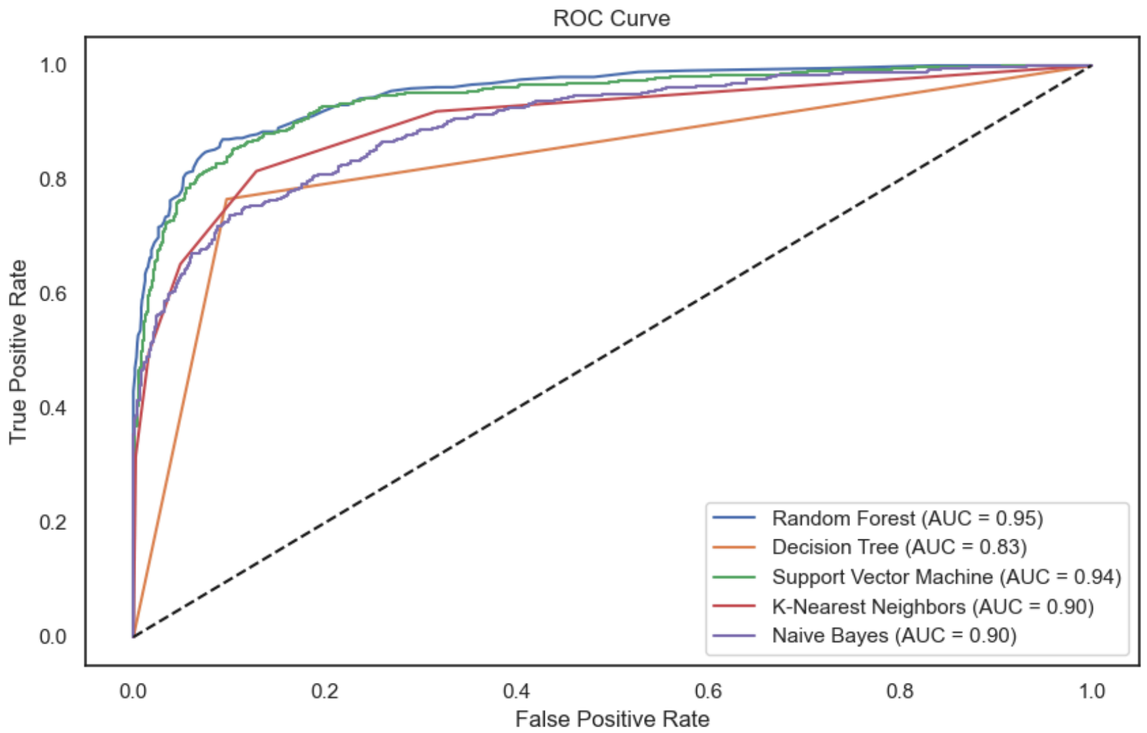
MODEL PERFORMANCE ANALYSIS

Naive Bayes

- **Accuracy (77.93%):** The lowest accuracy among all models, meaning it makes more incorrect predictions.
- **Precision (0.559):** The lowest precision, meaning it frequently predicts churn incorrectly (high false positives).
- **Recall (0.849):** Surprisingly, it has the highest recall, capturing 84.9% of actual churners, meaning it detects most churners but at the cost of many false positives.
- **F1 Score (0.674):** The low F1 score shows an imbalance between recall and precision.
- **CROSS (77.99%):** The lowest cross-validation score confirms that it does not perform consistently across different samples

Conclusion: Naive Bayes identifies many churners (high recall), but its high rate of false positives (low precision) and low accuracy make it less effective overall.

Modelisation



Random Forest Classifier is the best model, offering a strong balance between accuracy, precision, recall, and overall model stability.Naive Bayes, while strong in recall, suffers from poor precision and accuracy, making it unreliable for churn prediction. Other models like SVM and KNeighbors perform well but do not surpass Random Forest in terms of overall predictive capability.

The choice of the best model depends on your specific business objectives. If precision is crucial, SVM might be preferred. If recall is important, Random Forest might be a better choice.

To effectively decrease churn and improve customer retention, the company should concentrate on elevating customer service standards, adjust pricing and service offerings to meet the diverse needs of different regions and demographics, encourage the adoption of long-term contracts, and resolve particular challenges associated with payment methods and data consumption. It is imperative to tailor services to various customer segments, notably older adults and low-data users, as this will play a significant role in improving customer satisfaction and loyalty.

Pernel Hyldras AVOUGNASSOU
pernelavougnassou@gmail.com

Thank you.