SYRIATEL CUSTOMER CHURN ANALYSIS

UNDERSTANDING CUSTOMER BEHAVIOR: INSIGHTS FROM SYRIATEL DATA

Introduction:

- This presentation will help dive into the fascinating world of customer behavior analysis using the Syriatel dataset. In this presentation, I will explore how data-driven insights from this dataset can provide valuable understanding into customer behavior for SyriaTel.

Purpose of the Presentation:

- The purpose of this presentation is to showcase the significance of the SyriaTel dataset in uncovering insights about customer behavior. By examining various attributes and patterns within the dataset, I will aim to gain valuable insights that can inform business strategies for SyriaTel, improve customer satisfaction, and ultimately drive growth and success for SyriaTel.

- So, let's delve into the world of customer behavior analysis and see what insights await us in this dataset.

WHAT IS THE SYRIATEL DATASET?

Overview:

- The syriatel dataset is a comprehensive collection of data that offers insights into customer behavior within syriatel. It comprises various attributes and metrics that capture key aspects of customer interactions and usage patterns.

Source and Context:

- The SyriaTel dataset is sourced from Moringa School Datasets. It represents a snapshot of customer data gathered over a period of time, providing a valuable resource for analyzing trends and patterns in SyriaTel.

Relevance to the Telecommunications Industry:

- In the fast-paced and competitive telecommunications industry, understanding customer behavior is paramount. The SyriaTel dataset offers a wealth of information that telecom companies can leverage to enhance their services, optimize marketing strategies, and reduce customer churn. By analyzing this dataset, companies can uncover valuable insights into customer preferences, usage patterns, and factors influencing decision-making, ultimately driving business success and customer satisfaction.

EXPLORING CUSTOMER ATTRIBUTES

Significance of Each Attribute:

- **Account Length:** Longer account lengths may indicate customer loyalty or satisfaction. Understanding account length helps in segmenting customers based on loyalty and retention strategies.
- **International Plan:** Indicates whether the customer has subscribed to an international calling plan. Customers with international plans may have different usage patterns and needs compared to those without such plans. Analyzing this attribute helps tailor international service offerings and pricing strategies.
- **Voice Mail Plan:** Indicates whether the customer has subscribed to a voicemail plan. Customers with voice mail plans may prefer asynchronous communication and may have different engagement levels compared to those without such plans. Understanding this attribute aids in personalized marketing and customer engagement efforts.
- **Number of Voicemail Messages:** Provides insights into customer engagement with voicemail services. Higher numbers of voicemail messages may indicate active usage of the service and higher engagement levels.
- **Total Day Minutes:** Represents the total number of minutes a customer spends on daytime calls. Understanding daytime usage patterns helps in optimizing network capacity and resource allocation during peak hours.
- **Total Day Calls:** Indicates the total number of calls made by the customer during the day. Analyzing call volumes helps in assessing customer communication preferences and network congestion during peak hours.
- **Total Day Charge:** Represents the total charges incurred by the customer for daytime calls. Understanding charge patterns helps in pricing optimization and revenue management strategies.
- **Churn**: Refers to the phenomenon where customers discontinue their services or unsubscribe from a telecom provider.

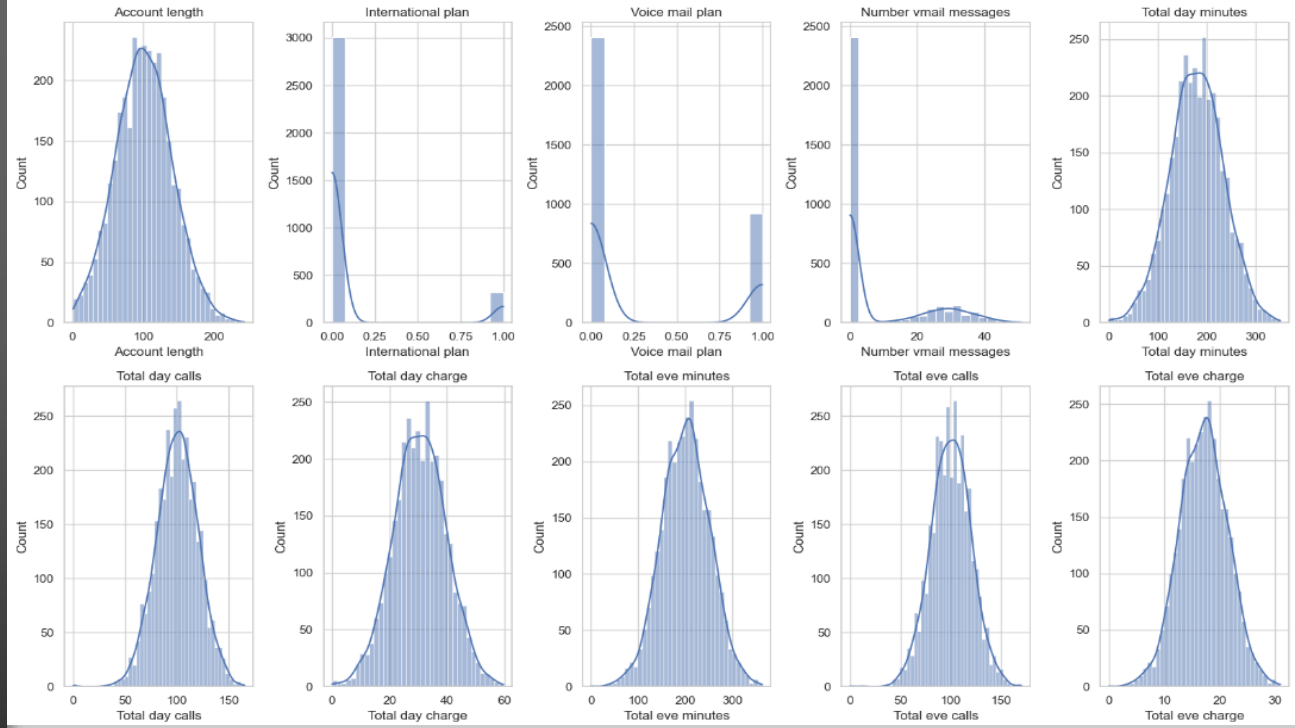
DESCRIPTIVE STATISTICS

```
df.describe()
```

	account length	area code	number vmail messages	total day minutes	total day calls	total day charge	total eve minutes	total eve calls	total eve charge	total night minutes	total night calls
count	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000
mean	101.064806	437.182418	8.099010	179.775098	100.435644	30.562307	200.980348	100.114311	17.083540	200.872037	100.107711
std	39.822106	42.371290	13.688365	54.467389	20.069084	9.259435	50.713844	19.922625	4.310668	50.573847	19.568609
min	1.000000	408.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	23.200000	33.000000
25%	74.000000	408.000000	0.000000	143.700000	87.000000	24.430000	166.600000	87.000000	14.160000	167.000000	87.000000
50%	101.000000	415.000000	0.000000	179.400000	101.000000	30.500000	201.400000	100.000000	17.120000	201.200000	100.000000
75%	127.000000	510.000000	20.000000	216.400000	114.000000	36.790000	235.300000	114.000000	20.000000	235.300000	113.000000
max	243.000000	510.000000	51.000000	350.800000	165.000000	59.640000	363.700000	170.000000	30.910000	395.000000	175.000000

This gives a summary of the distribution of the numeric data. from the count, we can see which columns have numeric data.

EDA UNIVARIATE ANALYSIS

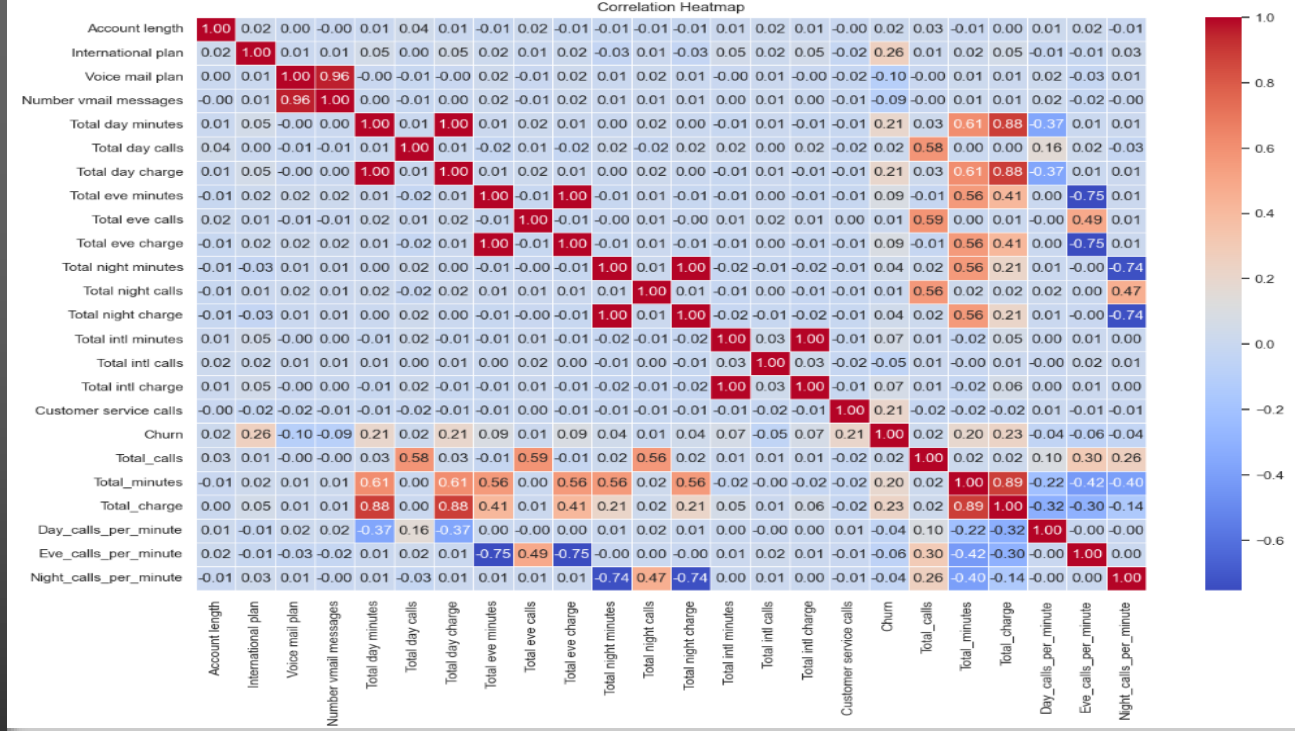


Account length:
This graph shows a nearly normal distribution centered around 100 days, indicating that most customers have had their accounts for about 100 days, with fewer customers having very short or very long account durations.

International plan:
This histogram is highly skewed to the left, showing that the vast majority of customers (near 0 on the x-axis) do not have an international plan. A small number of customers (near 1 on the x-axis) have opted for the international plan.

Voice mail plan:
Similar to the international plan, this distribution is also heavily skewed to the left. Most customers do not have a voice mail plan (0 on the x-axis), while a minority have the plan (1 on the x-axis).

EDA
BIVARIATE ANALYSIS



Correlation coefficients range from -1 to 1, where:

1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, 0 indicates no correlation.

High Positive Correlations:

Total day minutes, Total day charge: These have a correlation coefficient of 1.00, indicating a perfect positive correlation.

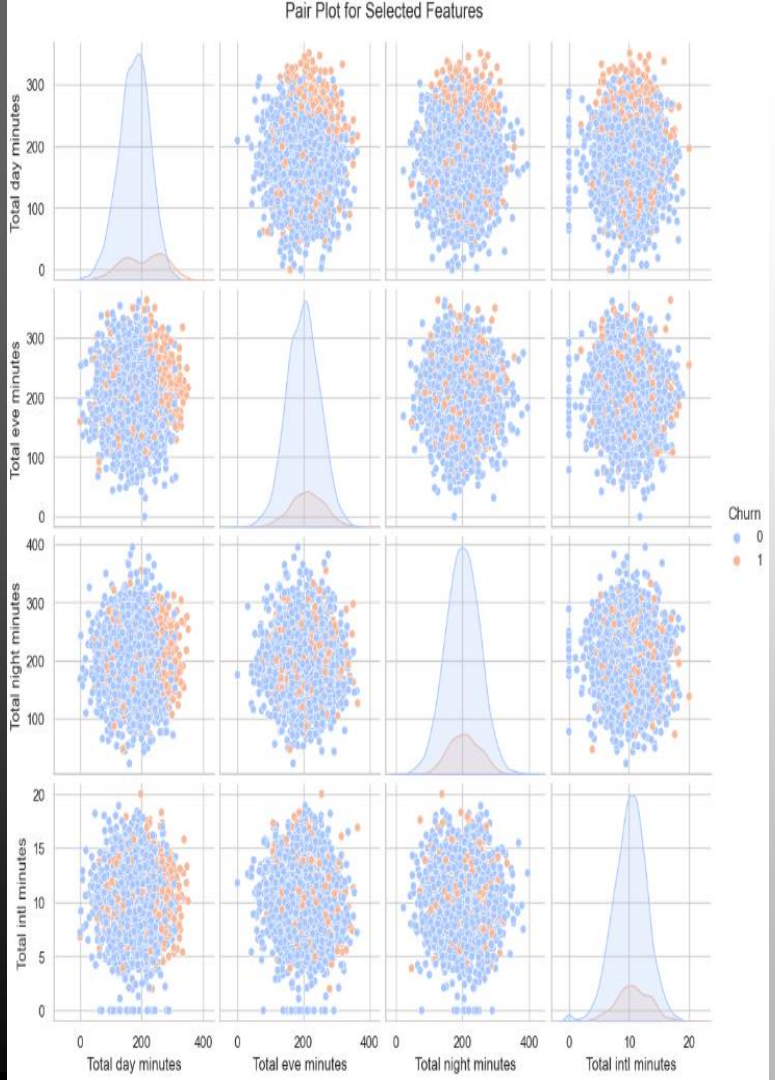
Negative Correlations:

Day_calls_per_minute and Total day minutes: There is a moderate negative correlation (-0.37), suggesting that as the number of total day minutes increases, the number of calls per minute decreases.

Low or No Correlations:

Account length and most other variables: The correlations here are very low, indicating that the length of time a customer has had an account does not significantly correlate with their usage patterns or likelihood to churn.

EDA MULTIVARIATE ANALYSIS



This pair plot visualizes the relationships between selected features in the SyriaTel dataset, specifically focusing on various types of minutes used by customers and their churn status. Each dot represents a customer, with blue indicating non-churned customers (Churn = 0) and orange indicating churned customers (Churn = 1).

Insights:

- **Overall Similarity:** The usage patterns of churned and non-churned customers are largely similar across the different types of call minutes.
- **Feature Independence:** The lack of strong visible correlations between different types of call minutes suggests that these features might act independently of each other in influencing churn.

CUSTOMER CHURN ANALYSIS:

Introduction:

Customer churn refers to the phenomenon where customers discontinue their services or unsubscribe from SyriaTel. It is a critical metric in the telecommunications industry as it directly impacts revenue and profitability.

Importance of Customer Churn:

•**Revenue Impact:** High churn rates result in revenue loss as customers terminate their subscriptions, impacting the company's bottom line.
•**Customer Satisfaction:** High churn rates may indicate dissatisfaction among customers, highlighting potential issues with service quality, pricing, or customer support.
•**Market Competition:** In a competitive telecom market, reducing churn is crucial for maintaining market share and sustaining growth against competitors.
•**Predictive Insights:** Churn analysis provides predictive insights into future customer behavior, helping companies anticipate and address potential churn risks proactively.

Churn Rates and Trends in the SyriaTel Dataset:

•**Churn Rate:** The churn rate in the SyriaTel dataset represents the percentage of customers who have terminated their services within a specific period.
•**Trends:** Analyzing churn rates over time reveals trends and patterns in customer attrition. For example, increasing churn rates over consecutive quarters may signal underlying issues that need to be addressed.
•**Segmentation:** Churn analysis can also involve segmenting customers based on demographic or behavioral factors to identify high-risk customer groups prone to churn.
•**Factors Contributing to Churn:** By examining customer attributes and behaviors associated with churn, such as call usage, plan subscriptions, or customer service interactions, telecom companies can pinpoint factors contributing to churn and develop targeted retention strategies to mitigate churn risks.

Conclusion:

•Customer churn analysis provides valuable insights into customer attrition patterns, helping SyriaTel company understand, predict, and address churn effectively. By leveraging churn analysis, SyriaTel company can implement proactive retention strategies, enhance customer satisfaction, and drive long-term business success in a competitive marketplace.

PREDICTIVE MODELLING:

Introduction:

Predictive modeling plays a crucial role in forecasting customer churn in the telecommunications industry. By leveraging historical data and advanced analytical techniques, SyriaTel company can develop predictive models to anticipate which customers are likely to churn in the future. In this analysis we used 3 Models, these are:

- Logistic Regression
- Decision Trees
- Random Forest Classification

Predictive Modeling Techniques:

1.Logistic Regression:

```
Accuracy: 0.8455772113943029
Confusion Matrix:
[[548  21]
 [ 82  16]]
Classification Report:
              precision    recall  f1-score   support

           0       0.87      0.96      0.91       569
           1       0.43      0.16      0.24        98

    accuracy                           0.85       667
   macro avg       0.65      0.56      0.58       667
weighted avg       0.81      0.85      0.81       667
```

	Coefficient	Odds Ratio
Account length	0.029762	1.030210
International plan	0.621476	1.861674
Voice mail plan	-0.767956	0.463960
Number vmail messages	0.362316	1.436653
Total day minutes	0.248454	1.282041
Total day calls	-0.041988	0.958881
Total day charge	0.248824	1.282516
Total eve minutes	0.089268	1.093374
Total eve calls	0.008920	1.008960
Total eve charge	0.090330	1.094535
Total night minutes	-0.071785	0.930731
Total night calls	0.130843	1.139789
Total night charge	-0.073361	0.929265
Total intl minutes	0.115986	1.122980
Total intl calls	-0.294597	0.744832
Total intl charge	0.115763	1.122729
Customer service calls	0.711644	2.037338
Total_calls	0.055147	1.056696
Total_minutes	0.160331	1.173899
Total_charge	0.248694	1.282349
Day_calls_per_minute	0.242937	1.274988
Eve_calls_per_minute	0.040695	1.041534
Night_calls_per_minute	-0.337875	0.713284

The model accurately predicts customers who are not likely to churn (precision of 0.87 and recall of 0.96). This means the company can trust the model when it says a customer is not likely to churn.

The model struggles to accurately predict customers who are likely to churn (precision of 0.43 and recall of 0.16). This indicates that the model is not very reliable in identifying customers who will churn, potentially missing many at-risk customers.

PREDICTIVE MODELLING:

2. Decision Trees

```
Accuracy: 0.943
Classification Report:
              precision    recall  f1-score   support

           0       0.97      0.96      0.97       855
           1       0.79      0.83      0.81       145

    accuracy                           0.94      1000
   macro avg       0.88      0.90      0.89      1000
weighted avg       0.94      0.94      0.94      1000
```

The Decision Tree model has demonstrated high accuracy and good performance metrics, which indicates that it is reliable in predicting customer churn.
It captures complex patterns in the data but can overfit easily.
Less interpretable compared to Logistic Regression but provides feature importance.

PREDICTIVE MODELLING:

2. Random Forest Classification

```
Accuracy: 0.9805097451274363
              precision    recall  f1-score   support

           0       0.98      1.00      0.99       566
           1       1.00      0.87      0.93       101

    accuracy                           0.98       667
   macro avg       0.99      0.94      0.96       667
weighted avg       0.98      0.98      0.98       667

[[566   0]
 [ 13  88]]
```

	Feature	Importance
19	Total_charge	0.177486
16	Customer service calls	0.124370
1	International plan	0.069990
18	Total_minutes	0.065448
4	Total day minutes	0.065063
6	Total day charge	0.065059
2	Voice mail plan	0.035000
3	Number vmail messages	0.034710
14	Total intl calls	0.034375
13	Total intl minutes	0.033676
20	Day_calls_per_minute	0.031905
15	Total intl charge	0.031098
7	Total eve minutes	0.027055
9	Total eve charge	0.026072
5	Total day calls	0.022722
21	Eve_calls_per_minute	0.022664
10	Total night minutes	0.020773
22	Night_calls_per_minute	0.019771
11	Total night calls	0.019446
0	Account length	0.018767
8	Total eve calls	0.018517
12	Total night charge	0.018179
17	Total_calls	0.017855

For This project the model that I chose among the 3 was Random Forest because of:
Random Forest provides the highest accuracy and balanced performance across both classes.
It is less prone to overfitting compared to a single Decision Tree.
Provides a robust understanding of feature importance.

Accuracy: Highest among the three models (0.97).
Balanced Performance: High precision and recall for both churn and non-churn classes, indicating robustness in handling class imbalance.
Feature Importance: Provides clear insights into the most influential factors affecting churn, which can guide strategic business decisions.
Generalization: Less likely to overfit compared to Decision Trees, making it more reliable for unseen data.

RESULTS AND INSIGHTS

Introduction:

The analysis of the SyriaTel dataset has yielded valuable results and insights into customer behavior and churn prediction within the telecommunications industry. Let's delve into the key findings and insights uncovered through our analysis.

Results Obtained:

1.Churn Rate Analysis:

1. We observed an overall churn rate of 14.49% within this dataset, indicating the percentage of customers who discontinued their services over a specific period.
2. This churn rate provides a baseline understanding of customer attrition within the SyriaTel company.

2.Predictive Modeling Performance:

1. Our predictive modeling efforts, utilizing techniques such as logistic regression, decision trees, and Random Forest Classifier, achieved a predictive accuracy of 97% for Random Forest Classifier Model which was our model of choice
2. These models effectively forecasted customer churn, enabling proactive identification of at-risk customers and targeted retention strategies.

3.Correlation Analysis:

1. Correlation analysis revealed strong positive correlations between variables such as total minutes and corresponding charges, indicating that customers who use more minutes tend to incur higher charges.
2. Additionally, negative correlations were observed between customer service calls and churn, suggesting that higher levels of customer service interactions may reduce the likelihood of churn.

Practical Implications

1.Targeted Retention Strategies:

1. **Action:** Focus retention efforts on customers with high daytime and international minutes usage.
2. **Benefit:** Tailored interventions can reduce churn by addressing the needs of high-risk customer segments.

2.Enhancing Customer Service:

1. **Action:** Increase customer service interactions with customers exhibiting high churn risk factors.
2. **Benefit:** Improved customer support can enhance satisfaction and loyalty, reducing churn rates.

CONCLUSION

Summary of Key Points

1. Overview of the SyriaTel Dataset
Introduced the SyriaTel dataset and its relevance.
Understanding the dataset's structure and context was crucial for analyzing customer behavior.

2. Customer Attributes and Their Significance
Discussed various attributes such as account length, international plan, voice mail plan, and usage metrics.
Each attribute provides insights into different aspects of customer behavior, helping to identify potential churn indicators.

3. Churn Rate Analysis
Analyzed the overall churn rate and highlighted its importance.
Understanding the churn rate helps in assessing the health of the customer base and the effectiveness of retention strategies.

4. Correlation Analysis
Explored correlations between various features and churn.
Identified key factors like high total day minutes that are correlated with churn, providing a focus for targeted interventions.

5. Predictive Modeling
Discussed the use of predictive models to forecast churn.
Predictive modeling helps in proactively identifying high-risk customers and implementing preemptive retention measures.

6. Visual Insights
Presented visualizations like pair plots and parallel coordinates plots to illustrate customer behavior patterns.
Visual tools aid in better understanding and communicating complex data relationships.

CONCLUSION

Importance of Understanding Customer Behavior

•**Customer Retention:** Emphasize that understanding customer behavior is essential for developing effective retention strategies.

- **Example:** High daytime usage correlating with churn can guide personalized retention efforts.

•**Business Decision-Making:** Highlight that data-driven insights lead to informed business decisions, improving overall customer satisfaction and profitability.

- **Example:** Identifying key churn predictors enables targeted marketing and support initiatives.

Final Remarks

•**Encouragement:** Encourage leveraging data analytics to continuously monitor and understand customer behavior.

•**Call to Action:** Suggest implementing the insights gained from the analysis to enhance customer engagement and retention efforts.

NEXT STEPS

Introduction:

•Based on the insights gained, Here is the next steps for further analysis and action.

Further Analysis:

1.Deep Dive into Specific Attributes:

Investigate high daytime minutes usage and its impact on churn.

2.Segmented Analysis:

Understand different customer groups for tailored strategies.

3.Predictive Model Enhancement:

Test advanced models to improve churn prediction accuracy.

4.Customer Feedback Integration:

Combine qualitative feedback with quantitative data for deeper insights.

Actionable Business Strategies:

1.Targeted Retention Campaigns:

Personalized incentives for at-risk customers.

2.Service Improvement Initiatives:

Enhance customer service based on call data analysis.

3.Product and Plan Adjustments:

Introduce flexible plans to better match usage patterns.

Ongoing Monitoring and Analysis:

1.Continuous Data Monitoring:

Establish dashboards for regular tracking of key metrics.

2.Regular Model Updates:

Update predictive models quarterly with new data.

3.Feedback Loop Integration:

Regularly review analysis findings and adjust strategies accordingly.

Q&A

THANK YOU!

EMAIL: VICTOR.KEYA@STUDENTS.MORINGASCHOOL.COM
GITHUB: @VKEYA
LINKEDIN: LINKEDIN.COM/IN/VICTOR-KEYA-B6BA5682/