PREDICTING SYRIATEL CUSTOMER CHURN USING MACHINE LEARNING

A DATA-DRIVEN RETENTION STRATEGY

BUSINESS UNDERSTANDING

What is telecom Business?

- It involves operations of telecommunication systems such as mobile communication, fixed-line services, and internet provision to a wide range of customers.

- Highly competitive and rapidly evolving

- The business is often faced by challenges such as customer churn.

- Decision making based on data-driven insights are essential for survival

BUSINESS OBJECTIVE

Main objective:

To help SyriaTel, a telecommunication company, reduce revenue loss caused by customer churn by developing a predictive model that identifies customers who are most likely to leave.

Specific objectives:

1. Explore customer data to identify patterns that influence churn.
2. Preprocess and engineer useful features.
3. Build and train classification models.
4. Evaluate model performance.
5. To provide recommendations and retention strategies.

PROBLEM STATEMENT

- Syria is losing customers to competitors.
- It lack an effective way of predicting customer churn.
- A predictive model is needed to identify customer’s risk of leaving.
- Use data to reduce revenue loss and improve customer retention.

RESEARCH QUESTIONS

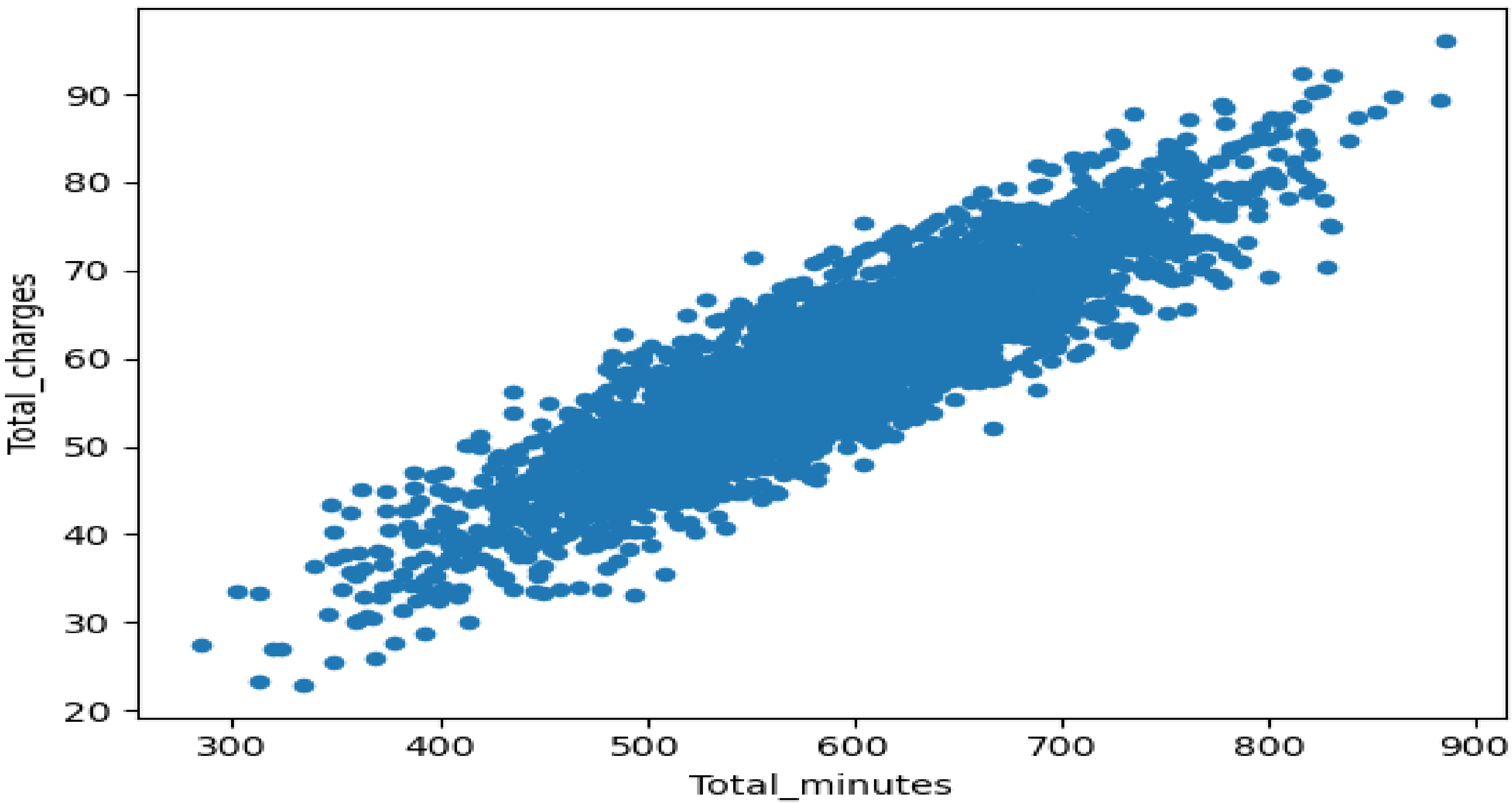
1. What key factors influence churn at Syriatel?
2. Which classification model performs best to predict customer churn?
3. How do customer’s behavior differ between the churners and non-churner?
4. What strategies can reduce churn based on data insights?

DATASET OVERVIEW

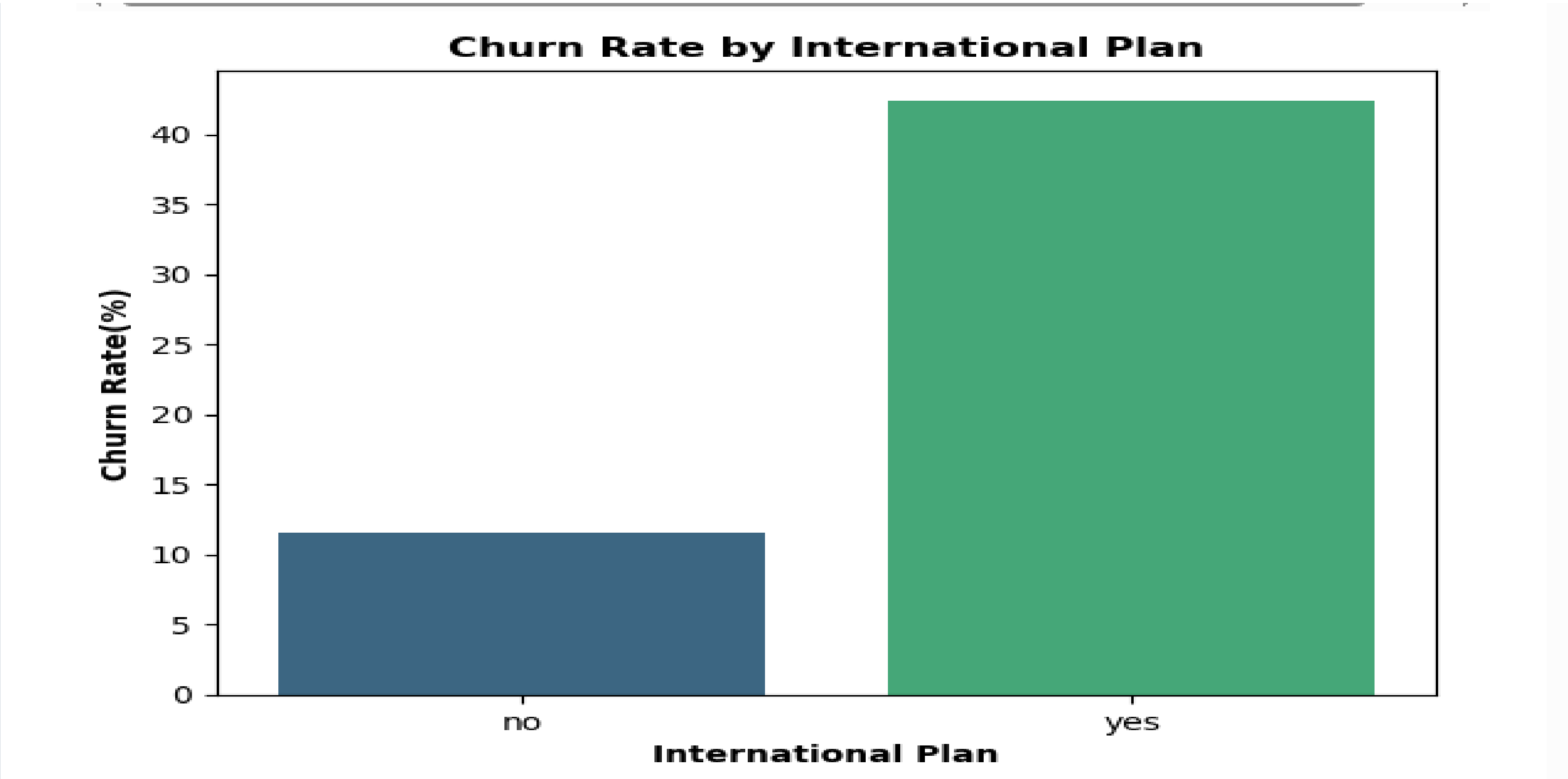
- Source: Kaggle, telecom churn data set
- Have 3333 records and 21 features
- They include:

✓ Customer usage(calls, minutes, charges)
✓ Demographics(area code, state)
✓ Plans(international, voicemail)
✓ Churn(target variable)

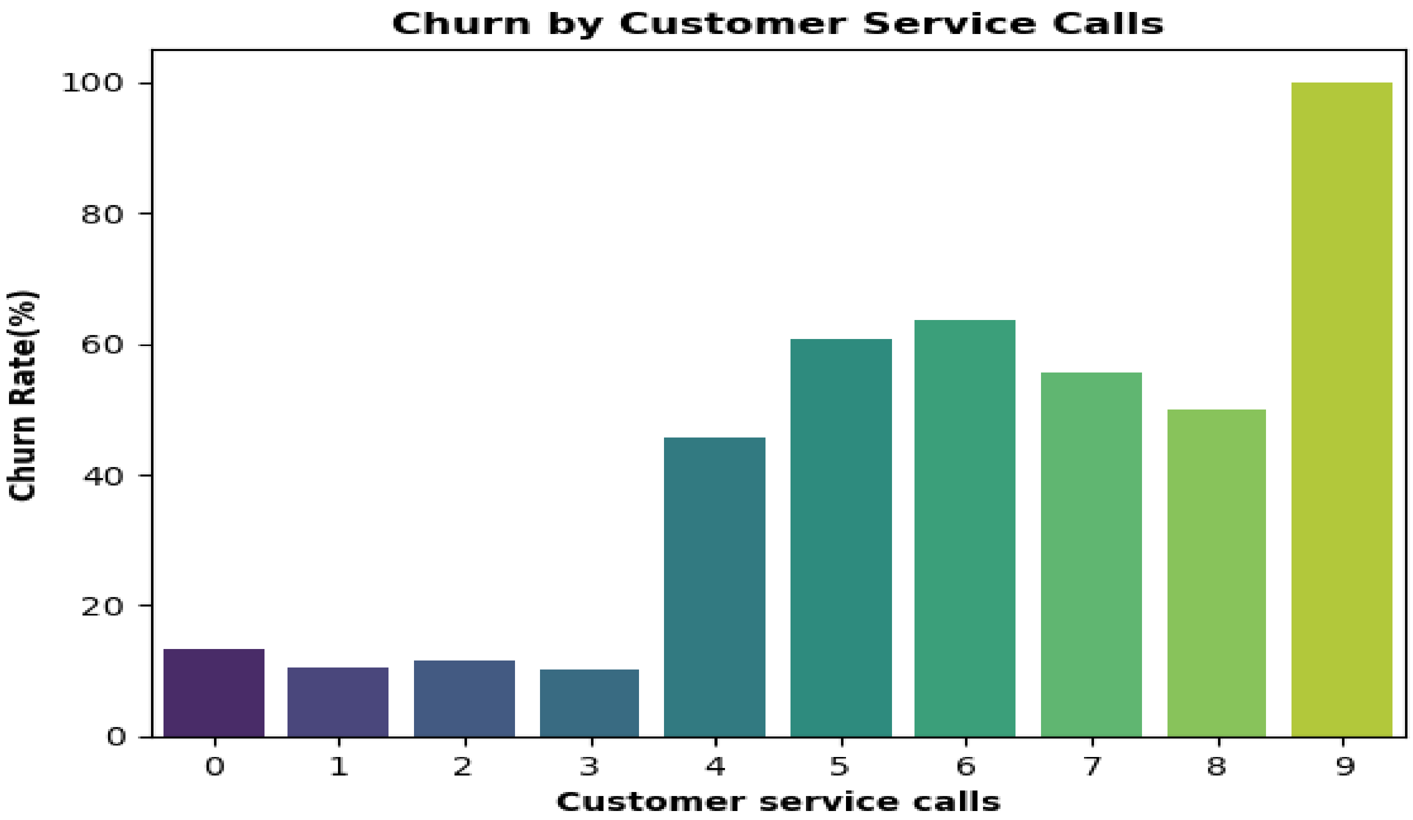
EXPLORATORY DATA ANALYSIS



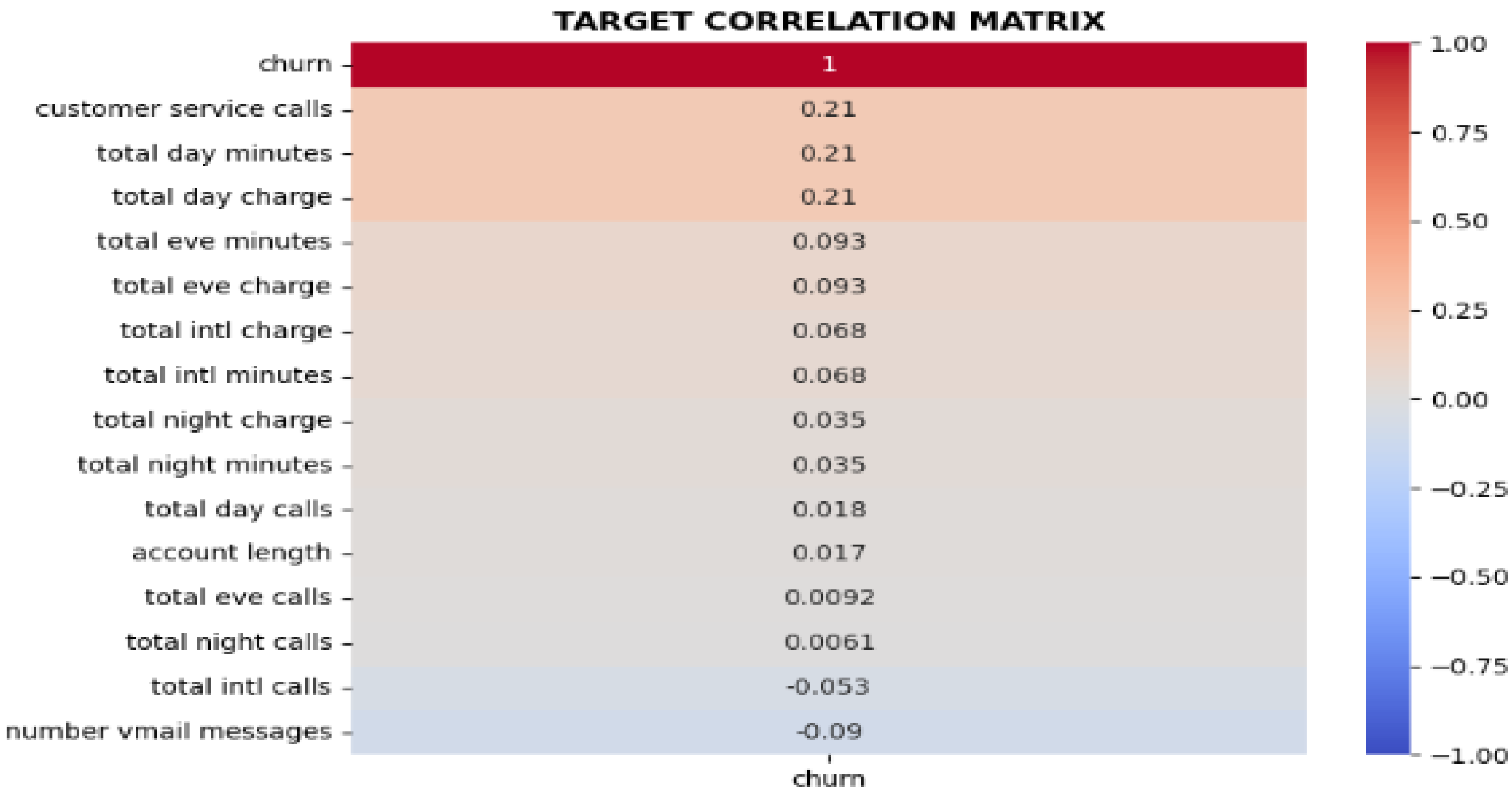
- **The scatter plot shows the linear relationship between total-charges and the total-minutes**



- **Customers with international plan are more likely to churn.**



Customers that contacted customer servives churned more, indicating a possible dissatisfaction in the services, such as unresolved issues, or long delays in resolving the issues



- **From the correlation matrix above, it is clear that charges and minutes columns are redundant because of similar correlation with each other. Churn interprets them as similar information.
  From this inference, we drop the minutes columns and analyse user data based on billing rather than usage time.**

MODELING

Modeling overview:

For the churn prediction the Model used are:

- Logistic Regression.
- Decision Tree Classifier.

Why we chose them?

- Logistic Regression-   for interpretability.
- Decision Tree for-   non-linear patterns.

- The SMOTE technique was used to handle imbalance problem

Models Results Before Balancing Summary.

Metric	Logistic regression	Decision Tree
Accuracy	0.87	0.97
Recall	0.97	1.0
Precision	0.89	0.96
F1-Score	0.97	0.98

Models Results after Balance Summary

Metric	Logistic Regression	Decision Tree
Accuracy	0.76	0.91
AUC	0.82	0.88
Recall	0.77	0.93
Precision	0.95	0.97
F1- score	0.82	0.94

EVALUATION

- Has better generalization on test data,(Accuracy: 86.6%) before balancing. This indicates Logistic Regression handles unseen data more reliably compared to the decision trees.

- The AUC of 0.88 is significantly high meaning the Decision Tree has strong ability to separate churners from non-churners. If you randomly select one churner and one non-churner, there's an 88% chance the tree will correctly assign a higher churn probability to the churner.

- Compared to Logistic Regression (AUC = 0.82), this tree performs better in terms of separating churners vs non-churners, but LR may generalize more reliably.

CONCLUSION

- Customers with frequent customer service are more likely to churn compared to those with few contact.
- Customer having many services calls and high daytime charges will probably leave due to stronger pattern of frustration.
- A customer without voice mail plan and high daytime charges are more likely to churn. On other hand, those with high voice mail plan and low daytime usage are more likely to stay.
- Customers who have an international plan and high daytime charges are also at a higher risk of churning, possibly due to high costs or unmet expectations with international services.

RECOMMENDATION

- **The company should improve customer services** as customers who call support often may be unhappy. The company should review the customer service process to solve problems faster and reduce repeated calls.
- **Company should come up with a retention plan for international users** as customers with international plans and high daytime usage are more likely to leave. The company can create special offers, discounts, or loyalty rewards for these customers to keep them satisfied.
- **Promotion of voicemail plans**: Regular daytime callers without voicemail plans are more likely to churn. The company should encourage these customers to take voicemail plans, maybe through bundled promotions.

- **Company should have proactive outreach** where they contact high-usage customers before they complain by offering check-ins or small perks to make them feel valued and reduce the chance of churn.
- **Develop targeted communication** such as sending personalized messages to at-risk groups like frequent service callers, heavy international users while explaining benefits of staying and giving tailored offers.
- **Analyze call costs** as high daytime and international charges may frustrate customers. Reviewing pricing structures or offering flexible packages could reduce dissatisfaction.
- **Customer education** as some customers may not fully understand plan benefits. Educating them in different ways such as explaining how voicemail or bundled offers save money, can improve satisfaction.

THANK YOU