SyriaTel Customer Churn Prediction

Presented by - Allan ngeiywa

Content

- **Business Problem Understanding**
- **Data Summary**
- **Exploratory Data Analysis**
- **Machine Learning Models**
- **Model Comparison**
- **Conclusion**
- **Q & A**

Business Problem Understanding

- SyriaTel, a prominent telecommunications company, is facing challenges related to customer churn. A significant portion of their customer base is discontinuing their services, resulting in revenue loss and reduced customer loyalty. The company seeks a data-driven solution to predict customer churn and implement strategies to reduce it.

- The primary business objective is to build a predictive model that can identify customers at risk of churning in advance. The goal is to leverage this predictive model to reduce churn rates and increase customer retention.

- Churn Prediction: Build a classifier that can accurately identify customers who are likely to discontinue their services ("churners") and those who are likely to continue as loyal customers ("non-churners").

- Predictable Patterns: Explore the data to identify predictable patterns and key factors influencing churn. This analysis should uncover insights into why customers are leaving SyriaTel's services.

- Communication: Present the findings and proposed strategies to SyriaTel's senior management and relevant teams to gain their support for implementing the recommendations.

Data Summary

Variables Description

- State :All 51 states
- Account Length : How long account has been active
- Area Code : Code Number of Area
- Intl Plan : International plan activated (yes, no)
- VMail Plan : Voice Mail plan activated (yes ,no)
- VMail Message : No.of voice mail messages
- Day Mins : Total day minutes used
- Day calls : Total day calls made
- Day Charge : Total day charge
- Eve Mins : Total evening minutes
- Eve Calls : Total evening calls
- Eve Charge : Total evening charge
- Night Mins : Total night minutes
- Night Calls : Total night calls
- Night Charge : Total night charge
- Intl Mins : Total International minutes used
- Intl Calls : Total International calls made
- Intl Charge : Total International charge
- CustServ calls : Number of customer service calls made
- Churn : Customer churn (Target Variable True=1, False=0)

Descion Variable - Churn

Categorical Variables- State, **international plan, voice mail plan**

Data Summary

SyriaTel Churn Dataset

The below table shows Syria table top 5 and bottom 5 rows

	State	Account length	Area code	International plan	Voice mail plan	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	Total night charge	Total intl minutes	Total intl calls	Total intl charge	Customer service calls	Churn
0	KS	128	415	No	Yes	25	265.1	110	45.07	197.4	99	16.78	244.7	91	11.01	10.0	3	2.70	1	False
1	OH	107	415	No	Yes	26	161.6	123	27.47	195.5	103	16.62	254.4	103	11.45	13.7	3	3.70	1	False
2	NJ	137	415	No	No	0	243.4	114	41.38	121.2	110	10.30	162.6	104	7.32	12.2	5	3.29	0	False
3	OH	84	408	Yes	No	0	299.4	71	50.90	61.9	88	5.26	196.9	89	8.86	6.6	7	1.78	2	False
4	OK	75	415	Yes	No	0	166.7	113	28.34	148.3	122	12.61	186.9	121	8.41	10.1	3	2.73	3	False

|	State	Account length	Area code	International plan	Voice mail plan	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	Total night charge	Total intl minutes	Total intl calls	Total intl charge	Customer service calls	Churn
3328	AZ	192	415	No	Yes	36	156.2	77	26.55	215.5	126	18.32	279.1	83	12.56	9.9	6	2.67	2	False
3329	WV	68	415	No	No	0	231.1	57	39.29	153.4	55	13.04	191.3	123	8.61	9.6	4	2.59	3	False
3330	RI	28	510	No	No	0	180.8	109	30.74	288.8	58	24.55	191.9	91	8.64	14.1	6	3.81	2	False
3331	CT	184	510	Yes	No	0	213.8	105	36.35	159.6	84	13.57	139.2	137	6.26	5.0	10	1.35	2	False
3332	TN	74	415	No	Yes	25	234.4	113	39.85	265.9	82	22.60	241.4	77	10.86	13.7	4	3.70	0	False

Churn Analysis

- There were False – 2,850 and True 483

- The chart reveals that out of the total dataset, 85.5% (2,850 customers) have not churned, while 14.5% (483 customers) have churned
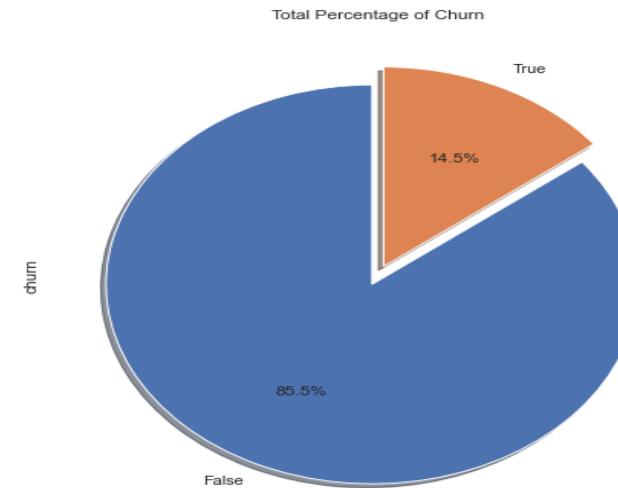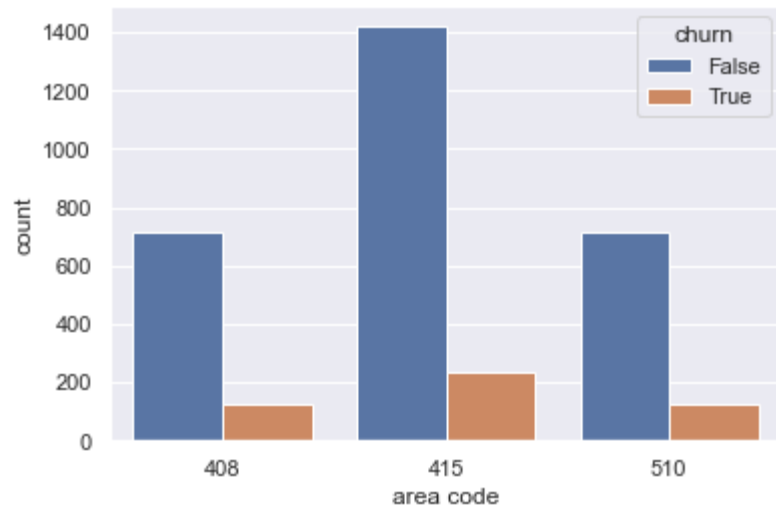

In the above data, we notice that there is only 3 unique value are there i.e 408,415,51

Analysis of State Column

Top 10 States with Highest Churn Rates
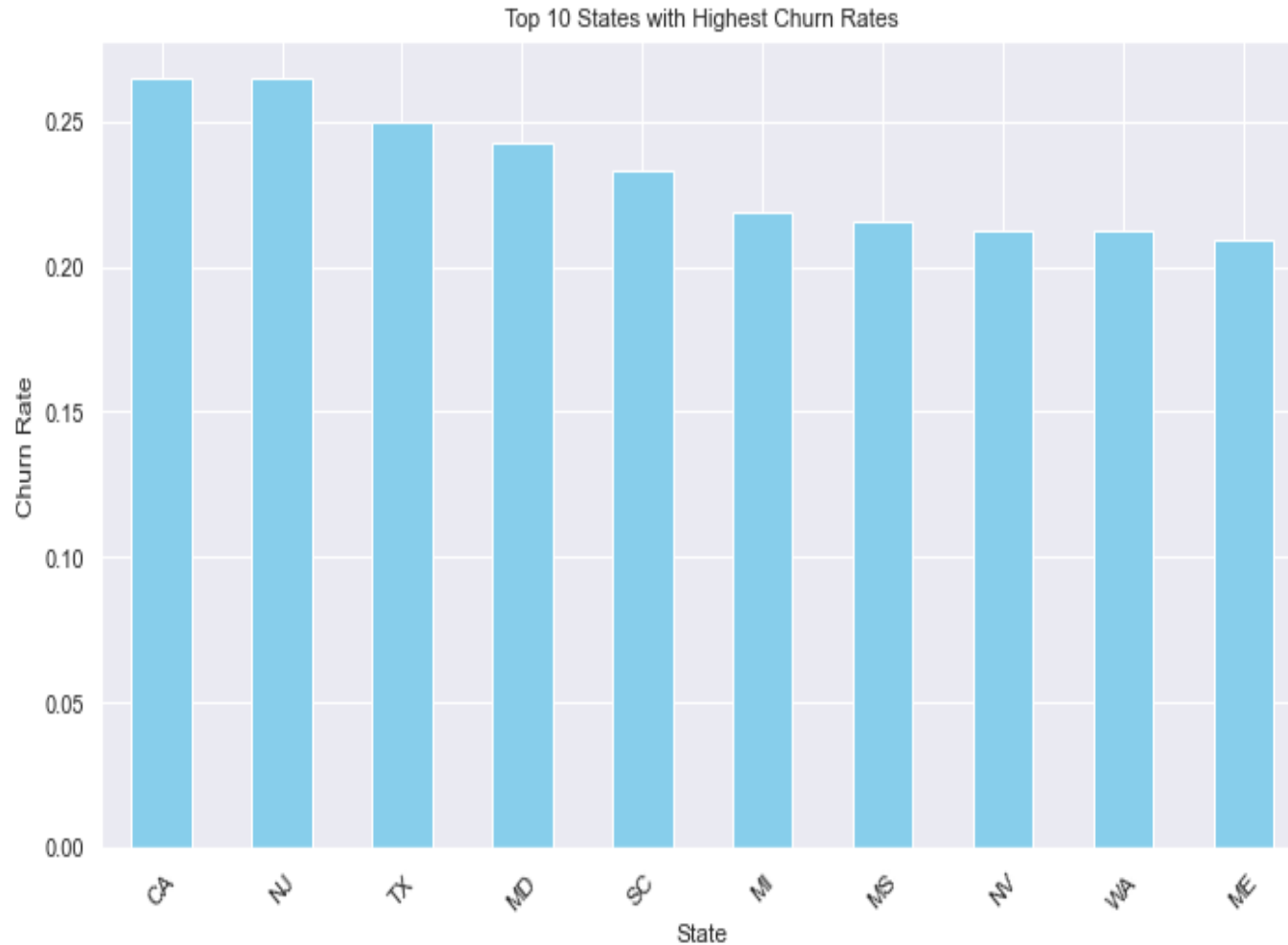
There are 51 unique state present who have different churn rate.
From the above analysis CA, NJ, TX, MD, SC, MI are the ones who have a higher churn rate of more than 22%
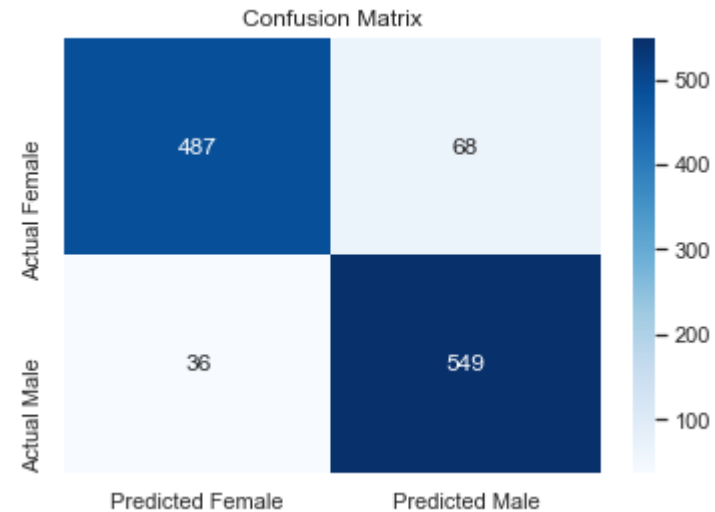
Analysis of Churn

Donut Plot for International plan

No — 90.3%

Yes — 9.7%
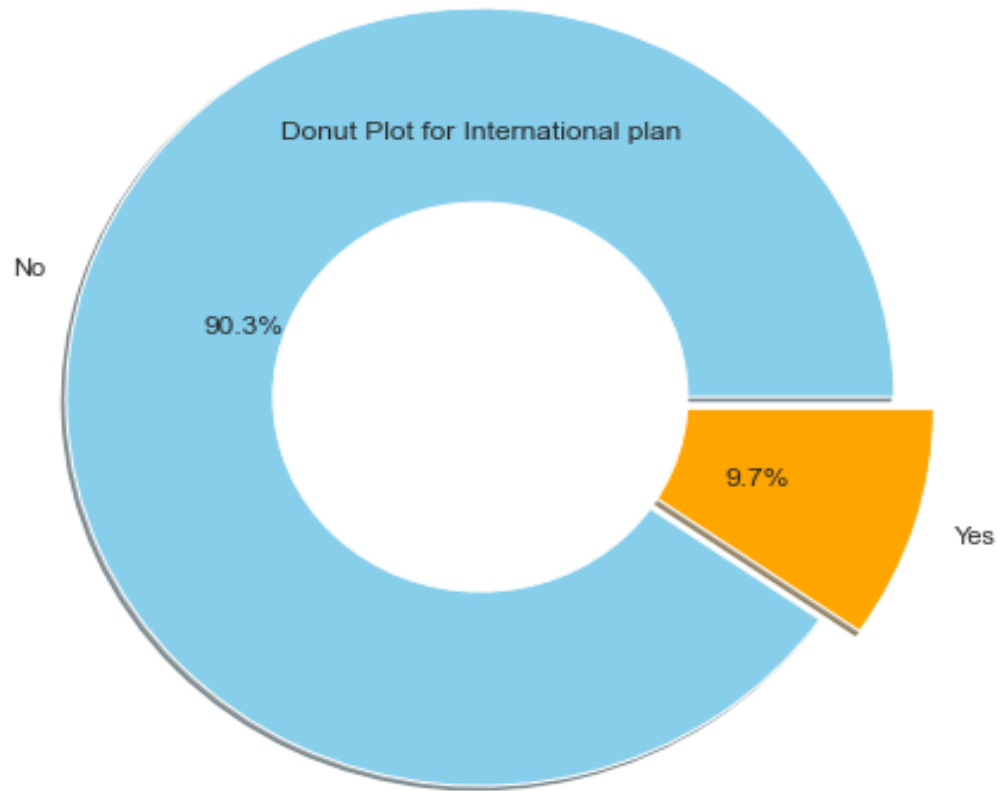
In this analysis, 3010 dont have an international plan, 323 have an international plan.

Machine Learning Models

▶ Logistic Regression


```
              precision    recall  f1-score   support

           0       0.93      0.88      0.90       555
           1       0.89      0.94      0.91       585

    accuracy                           0.91      1140
   macro avg       0.91      0.91      0.91      1140
weighted avg       0.91      0.91      0.91      1140
```
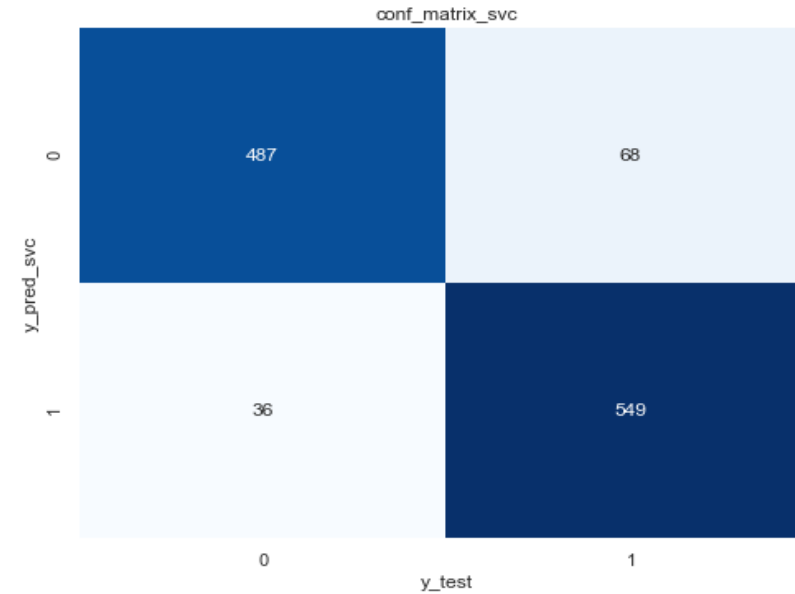
Model accuracy is 91%, which isn't bad. F1 score is 91.3%

Support Vector Classifier

	precision	recall	f1-score	support
0	0.96	0.82	0.89	555
1	0.85	0.97	0.91	585
accuracy			0.90	1140
macro avg	0.91	0.90	0.90	1140
weighted avg	0.91	0.90	0.90	1140

conf_matrix_svc

y_pred_svc	0	1
0	487	68
1	36	549

y_test

Model accuracy is 91%. F1 score is 91%

KNeighbors Classifier

	precision	recall	f1-score	support
0	0.96	0.85	0.90	555
1	0.87	0.97	0.92	585
accuracy			0.91	1140
macro avg	0.92	0.91	0.91	1140
weighted avg	0.92	0.91	0.91	1140

confusion_matrix_knn
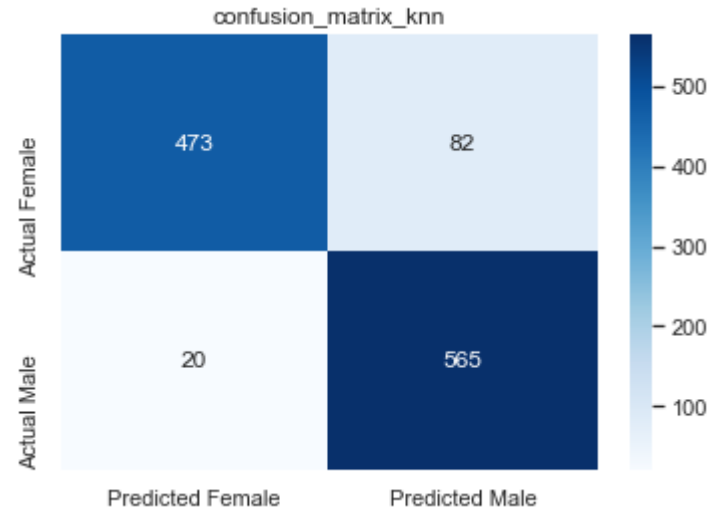
Model accuracy is 92%. F1 score is 92%

Decision Tree Classifier

```
           precision    recall  f1-score   support

        0       0.86      0.92      0.89       555
        1       0.92      0.86      0.89       585

 accuracy                          0.89      1140
macro avg       0.89      0.89      0.89      1140
weighted avg    0.89      0.89      0.89      1140
```
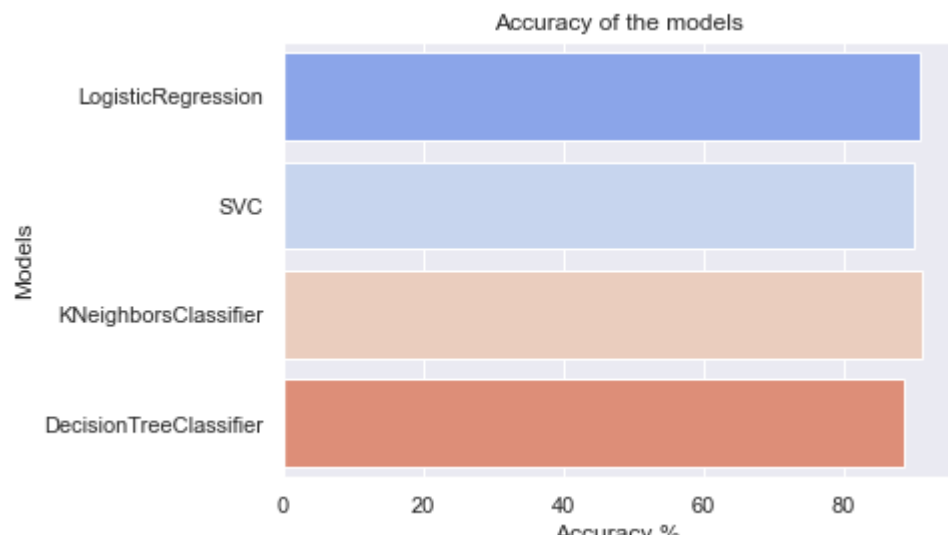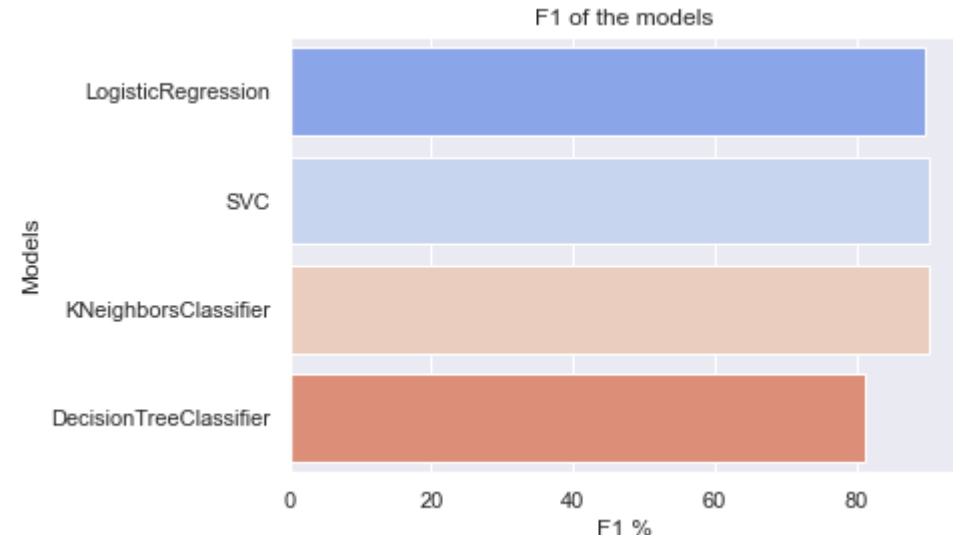

Confusion Matrix
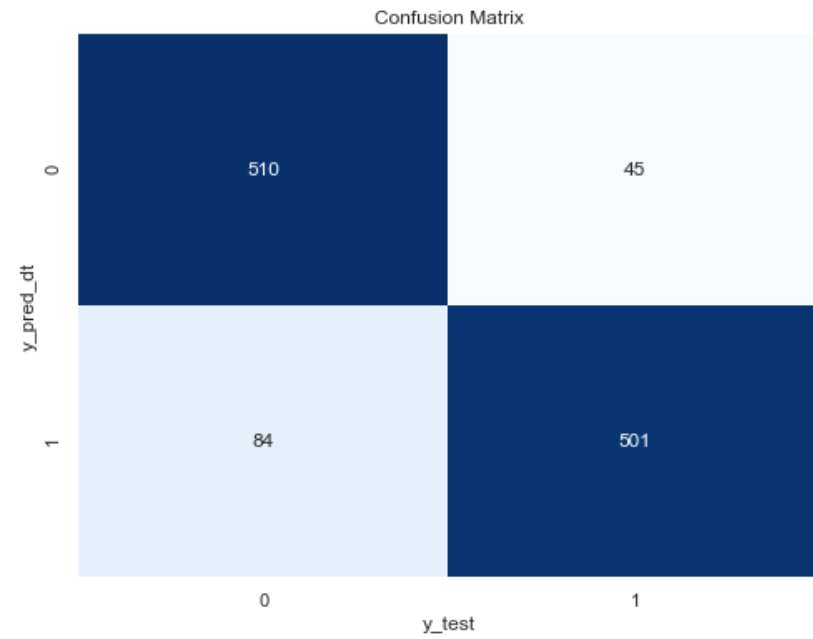
Model accuracy is 89%. F1 score is 89%

Model Comparisons

Accuracy of the models

	Models	Accuracy
0	KNeighborsClassifier	91.052632
0	LogisticRegression	90.877193
0	SVC	89.912281
0	DecisionTreeClassifier	88.684211

F1 of the models

	Models	F1
0	KNeighborsClassifier	90.092238
0	SVC	90.079953
0	LogisticRegression	89.543552
0	DecisionTreeClassifier	81.279757

The best performing model will have the highest accuracy. Of the four models tested, KNeighborsClassifier classifier has the highest accuracy.

KNeighborsClassifier classifier had the highest F1 score

Conclusion

▶ The company should improve on customer retention and reduce customer churn. This project analyzed a churn dataset to identify the main factors contributing to churn and gain valuable insights.

▶ Through exploratory data analysis, we were able to gain insight into the Syrialtel churn dataset.

▶ The churn pie chart reveals that out of the total dataset, 85.5% (2,850 customers) have not churned, while 14.5% (483 customers) have churned. This indicates a notable increase in customer churn and necessitates prompt attention.

▶ CA, NJ, TX, MD, SC, MI are the states who have a higher churn rate of more than 22%.

▶ The area code may not be relevant and can be excluded

Q & A

THANK YOU