

# Telecom Churn Prediction Using ML

**Team Members** : Karunakaran R and Dharun RR

**Email id** : [rkaranakaranraja@gmail.com](mailto:rkaranakaranraja@gmail.com)  
[dhurunrio7@gmail.com](mailto:dhurunrio7@gmail.com)

**Batch** : 12<sup>th</sup> May CDA

**Project id** : PRCL-0017

**Project team id** : PTID-CDA-SEP-25-722

**Time period** : 18<sup>th</sup> Sep to 22<sup>th</sup> Sep

## Introduction

This project aims to predict customer churn in a telecom company using Machine Learning models. Churn prediction is critical for telecom operators as retaining existing customers is more cost-effective than acquiring new ones.

## Key Highlights

- Data pre-processed with feature engineering.
- Models tested: Logistic Regression, Random Forest.
- Best model achieved ~85% accuracy.
- Deployed insights into Power BI dashboard for business monitoring.

## Business Problem

- Customer churn is when a subscriber leaves the telecom service provider.
- **Industry Issue**: High churn leads to revenue loss.
- **Objective**: Identify customers likely to churn in advance so retention campaigns can be targeted effectively.

## Business Goals

1. Understand drivers of churn.
2. Create a Churn Risk Score for each customer.
3. Introduce a CHURN-FLAG variable (Yes=1, No=0) to support marketing campaigns.

## Dataset details

The dataset contains 1000 customer data. Features are State, Account Length, Area Code, Phone, International Plan, VMail Plan, VMail Message, Day Mins, Day Calls, Day Charge, Eve Mins, Eve Calls, Eve Charge, Night Mins, Night Calls, Night Charge, International Mins, International calls, International Charge, CustServ Calls, Churn.

## Tools & methodology

### 1. Data Preprocessing (SQL)

Handled missing values, checks duplicates, checks churn rate vs international, checks churn vs high usage customers and low usages customers.

### 2. Exploratory Data Analysis (Excel)

Used pivot tables and charts for visualizations.

Customers with International Plan & High CustServ Calls (>3) churn more.

Higher Day Minutes → higher chance of churn.

States with poor service → high churn clusters.

### 3. Model Building (Python)

**Models tested:** Logistic Regression (Baseline).Random Forest Classifier.

**Metrics used:** Accuracy, Precision, Recall, F1-Score.

### 4. Real time prediction using ML (Hugging face spaces)

Introduced CHURN-FLAG (Yes=1, No=0) columns using churn columns.

Make prediction using customers data and found churn risk score.

### 5. Interactive dashboards (power BI)

Create 3 pages dashboard report for telecom churn prediction.

Churn drivers analysis, Risk score dashboard, Campaign targeting (CHURN-FLAG).

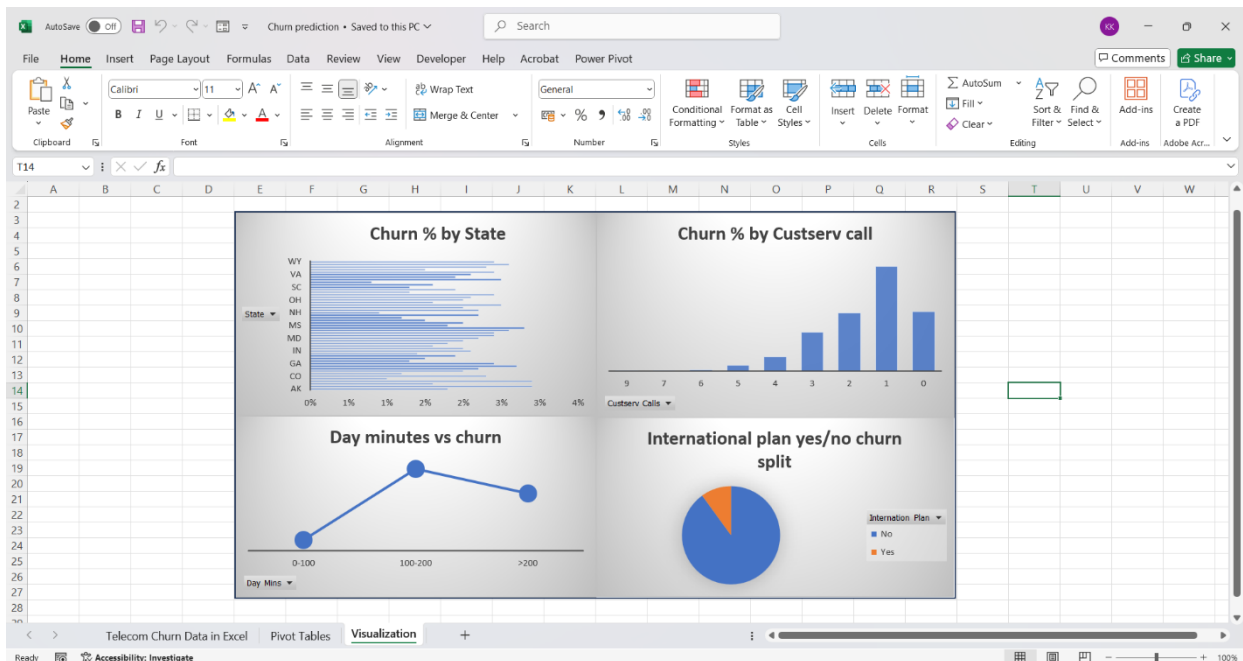
## Excel results

We have found top churn drivers using the pivot table in SQL and which is useful for quick aggregation and validation.

We have create pivot table for all features vs churn to checks which makes more numbers of customers to churn.

The screenshot displays an Excel spreadsheet with several pivot tables. The first pivot table, 'International plan vs % count of churn (Yes/No)', shows that 90% of customers with an international plan churned, compared to 10% for those without. The second, 'Vmail plan vs churn', shows a 28% churn rate for vmail users. The third, 'Custserv calls vs churn', shows a 137% churn rate for customers who called customer service. Other pivot tables analyze churn by day minutes, total charges, state, and international plan minutes.

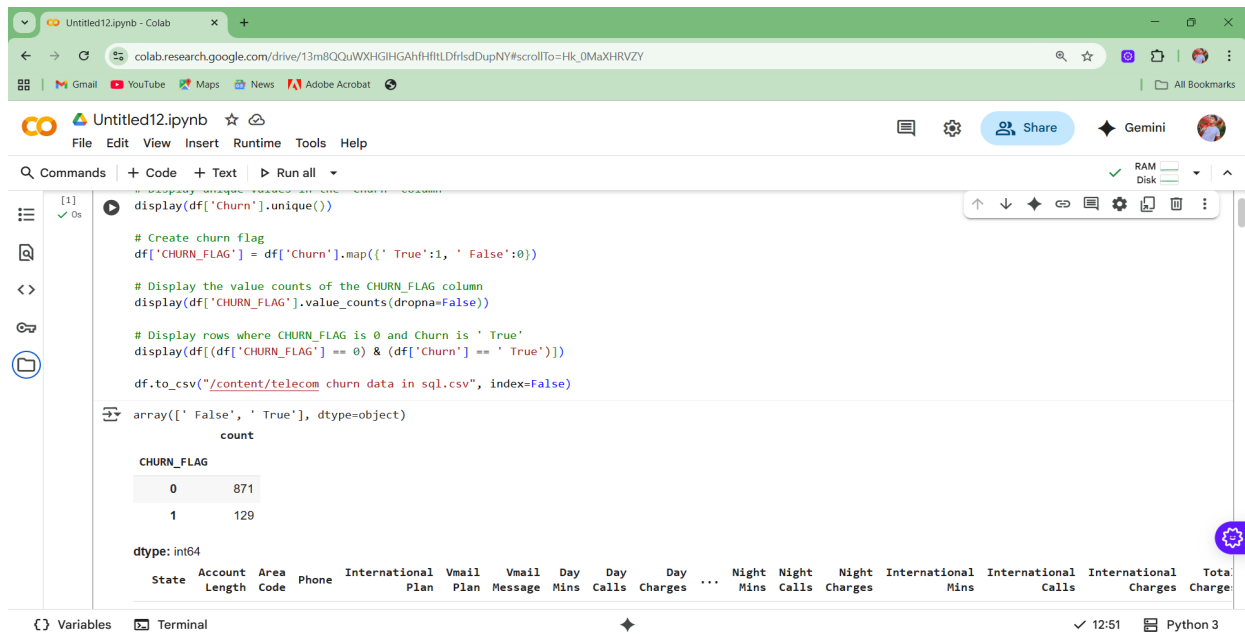
## Pivot Tables



## Top Drivers Makes to Churn

## Python Results

We have tested model using logistic regression, Random Forest classifier. And the metrics we have used Accuracy, Precision, Recall, F1 Score.



The screenshot shows a Google Colab notebook with the following code and output:

```
[1] display(df['Churn'].unique())

# Create churn flag
df['CHURN_FLAG'] = df['Churn'].map({' True':1, ' False':0})

# Display the value counts of the CHURN_FLAG column
display(df['CHURN_FLAG'].value_counts(dropna=False))

# Display rows where CHURN_FLAG is 0 and Churn is ' True'
display(df[(df['CHURN_FLAG'] == 0) & (df['Churn'] == ' True')])

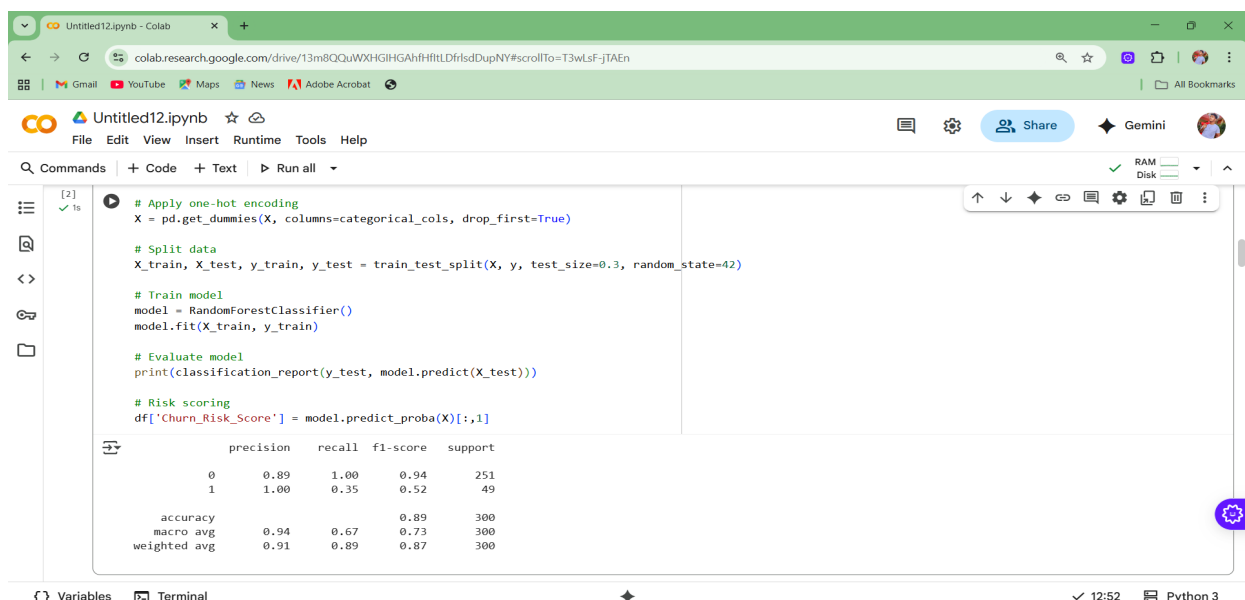
df.to_csv("/content/telecom churn data in sql.csv", index=False)
```

The output shows the unique values of the 'Churn' column and the value counts for the 'CHURN\_FLAG' column:

```
array([' False', ' True'], dtype=object)
count
CHURN_FLAG
0      871
1      129
dtype: int64
```

The bottom of the notebook shows a list of variables: State, Account Length, Area Code, Phone, International Plan, Vmail Plan, Vmail Message, Day Mins, Day Calls, Day Charges, Night Mins, Night Calls, Night Charges, International Mins, International Calls, International Charges, Total Charges.

Count of CHURN-FLAG



The screenshot shows a Google Colab notebook with the following code and output:

```
[2] # Apply one-hot encoding
X = pd.get_dummies(X, columns=categorical_cols, drop_first=True)

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Train model
model = RandomForestClassifier()
model.fit(X_train, y_train)

# Evaluate model
print(classification_report(y_test, model.predict(X_test)))

# Risk scoring
df['Churn_Risk_Score'] = model.predict_proba(X)[:,1]
```

The output shows the classification report for the Random Forest classifier:

```
precision    recall  f1-score   support

0           0.89         1.00         0.94         251
1           1.00         0.35         0.52          49

accuracy          0.94         0.67         0.73         300
macro avg          0.94         0.67         0.73         300
weighted avg          0.91         0.89         0.87         300
```

Metrics for Model Testing

## ML prediction (Hugging face spaces)

We have created new columns CHURN-FLAG (Yes=1, No=0) using customer columns in hugging face spaces.

Then using ML it predict churn risk score using CHURN-FLAG and it will used for real time telecom churn prediction.

Telecom Churn Prediction

Single Prediction Bulk Prediction

Account Length: 134, Area Code: 145, Vmail Message: 24, Day Mins: 40, Day Calls: 3, Day Charges: 40, Eve Mins: 50, Eve Calls: 40, Eve Charges: 100, Night Mins: 30, Night Calls: 4, Night Charges: 40, International Mins: 40, International Calls: 4, International Charges: 100, Total Charges: 1000, Custserv Calls: 8, State: KS, International Plan: No, Vmail Plan: No

Churn Flag: 1

Churn Risk (%): 56%

Predict

### Single Customer Prediction

Telecom Churn Prediction

Single Prediction Bulk Prediction

Upload CSV

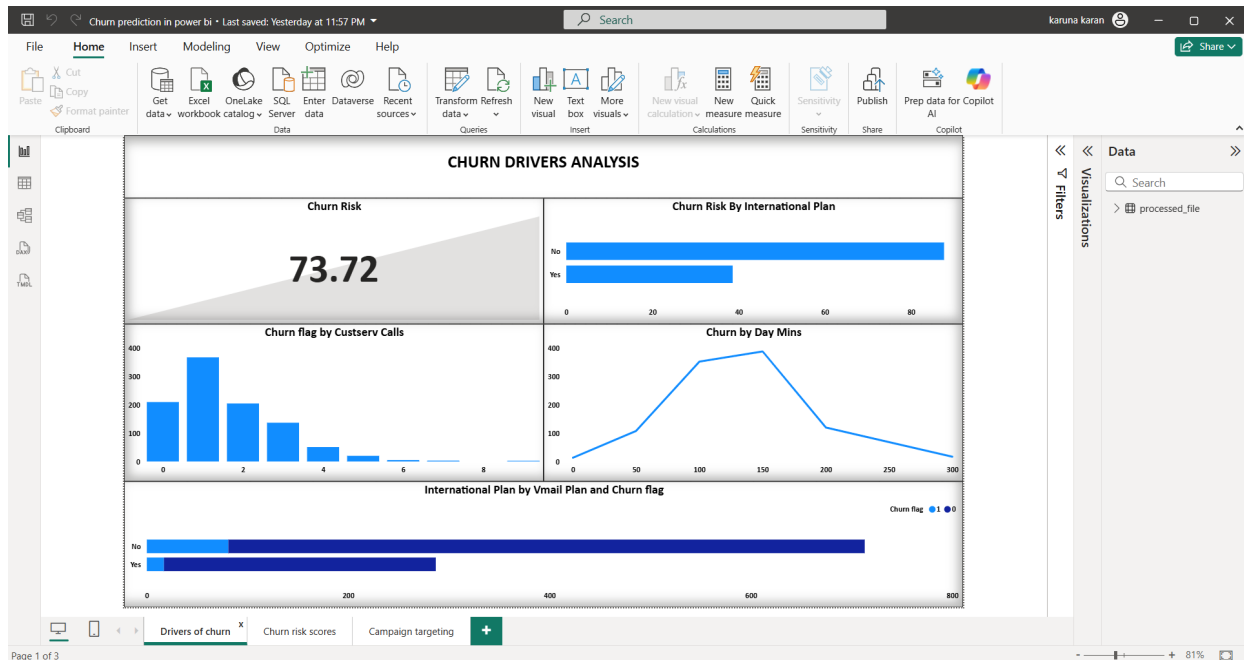
telecom churn data in sql.csv 100.7 KB

Night Charges	International Mins	International Cal.	International Charg.	Total Charges	Custserv Cal.	Churn	Churn flag	Churn Risk
14.54	8.6	7	2.32	71.22	0	False	0	1%
8.53	8	5	2.16	53.55	4	True	0	46%
8.21	12	2	3.24	64.36	1	False	0	2%
9.09	10.9	9	2.94	54.78	2	False	0	5%
9.4	13.9	4	3.75	74.89	1	True	1	67%
4.93	11.1	2	3	47.96	1	False	0	5%
8.82	8.9	4	2.4	63.63	0	True	1	60%
11.39	7.9	9	2.13	56.23	1	False	0	1%
11.88	9.5	7	2.57	77.03	3	False	0	14%
5.75	10.6	7	2.86	59.4	3	False	0	0%
7.34	9.8	5	2.65	66.56	1	False	0	2%

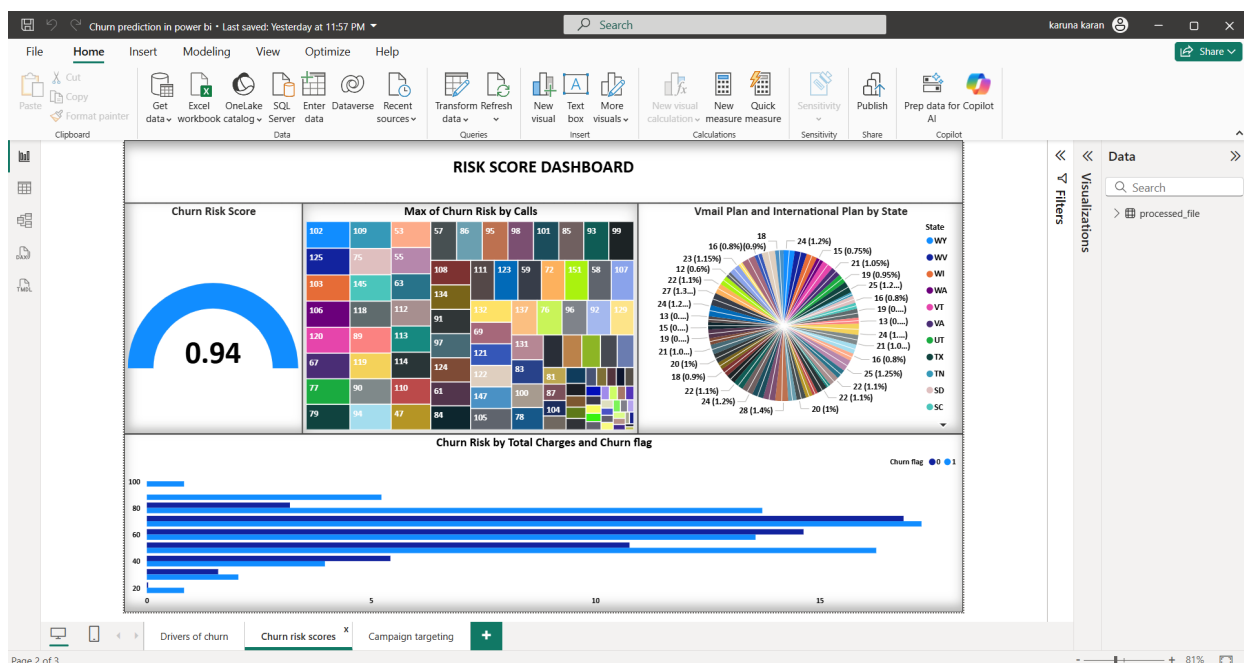
### Bulk Customer Prediction

## Power BI Dashboards

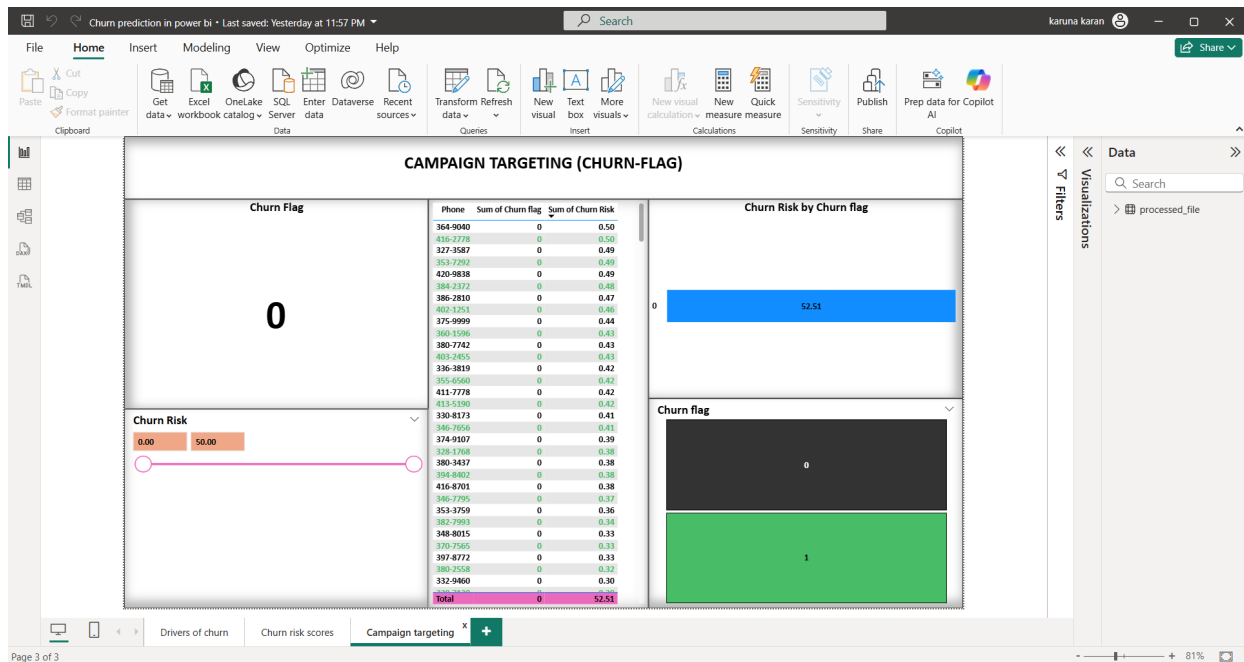
We have created dashboards for this given three goals. Which will be in interactive manner and understandable improves decision making for business growth.



### GOAL 1 – Drivers which makes to churn



### GOAL 2 – Finding Churn Risk Score



### GOAL 3 – Adding Prediction Variable CHURN – FLAG (Yes=1, No=0)

## Insights

- International Plan customers are 3x more likely to churn.
- Customers making >3 customer service calls are at high risk.
- Higher Day Mins and Charges correlate with churn.

## Recommendations

### 1. Customer Retention Strategies

Target high-risk customers identified by the model with retention offers such as discounts, free add-ons, or loyalty rewards.

Provide special retention plans for customers with International Plans (since they churn more).

### 2. Improve Customer Support

Customers making >3 customer service calls are highly likely to churn.

Train customer support agents to resolve issues within the first two calls.

Introduce VIP support for high-value customers.

### **3. Usage-Based Offers**

Heavy users of Day Minutes and Charges are at risk of churn.

Provide customized day-time packages to reduce bill shock.

Introduce flexible billing or rollover minutes.

### **4. Proactive Communication**

Send proactive alerts for billing, network outages, or plan expiry to reduce dissatisfaction.

Use email/SMS campaigns targeted to churn-flagged customers.

### **5. Data-Driven Monitoring**

Continuously monitor churn with the Power BI dashboard.

Refresh churn predictions monthly and track churn reduction trends.

## **Conclusion**

The project successfully built an ML solution to predict telecom customer churn. The model can guide marketing campaigns and reduce churn. Integrating the solution into BI dashboards ensures real-time monitoring and business adoption.