PRESENTATION

CUSTOMER CHURN PREDICTION ANALYSIS

TELECOM CUSTOMER RETENTION

INSIGHTS AND STRATEGY

OBJECTIVE:

Predict which customers are likely to churn based on behavior, service usage, and demographics.

Deliverables:

- Churn prediction model
- Key churn drivers
- Retention strategy recommendations

DATA SOURCE

TELECOM CUSTOMER DATASET (N=7,043)

FOCUS

PREDICTIVE MODELING

RETENTION STRATEGY

DATA SET OVERVIEW

COMPREHENSIVE ANALYSIS OF CUSTOMER BEHAVIOR, AND SERVICE USAGE

7,043

TOTAL CUSTOMERS

Unique customer records

22

ATTRIBUTES

Demographics Services Billing

99.8%

DATA COMPLETENESS

Only 11 rows missing (Total Charges)

CHURN STATUS

Binary Classification

Values: Yes/ NO

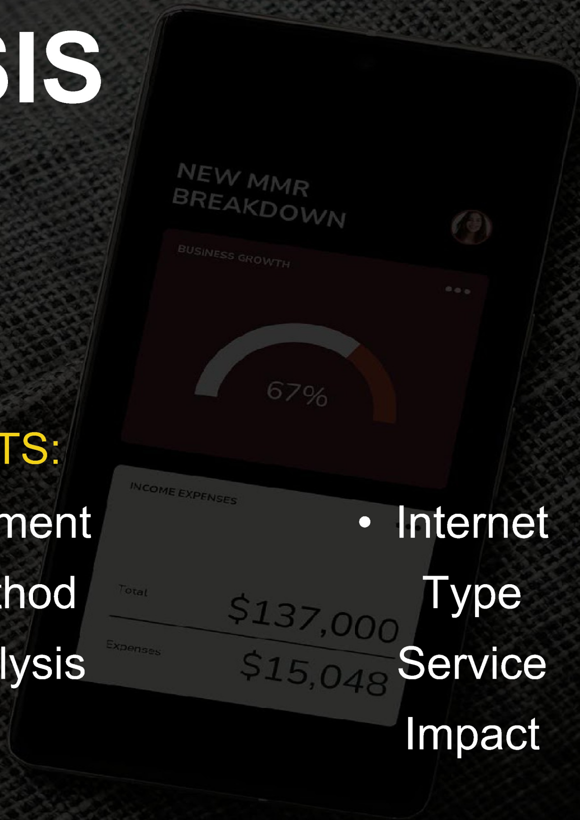
KEY FEATURES

- Customer Demographics (gender, senior citizen, dependents)
- Service Usage (phone, internet, security, backup, streaming)
- Billing Information (MonthlyCharges, TotalCharges)
- Subscription Details (contract type, payment method)

BEFORE MOVING FORWARD A QUICK ANALYSIS OF THE DATA SET

ANALYSIS USING THE GIVEN POINTS:

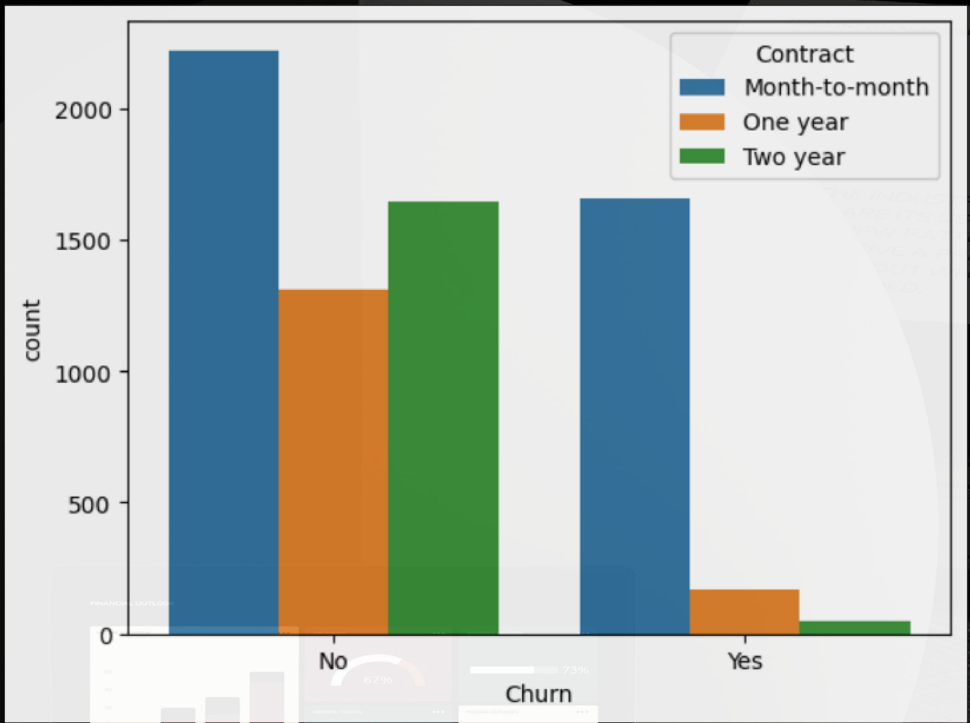
- Contract Type Analysis
- Payment Method Analysis
- Internet Type Service Impact
- Demographics & Billing Insights



CONTRACT TYPE ANALYSIS

Contract

Month-to-month	3875
Two year	1695
One year	1473



CHURN RATE :

MONTHLY - High Risk

ONE YEAR - Medium Risk

TWO YEAR - Low Risk

Strategic Action

Incentive migration from Month-to-Month plans.

PAYMENT METHOD ANALYSIS

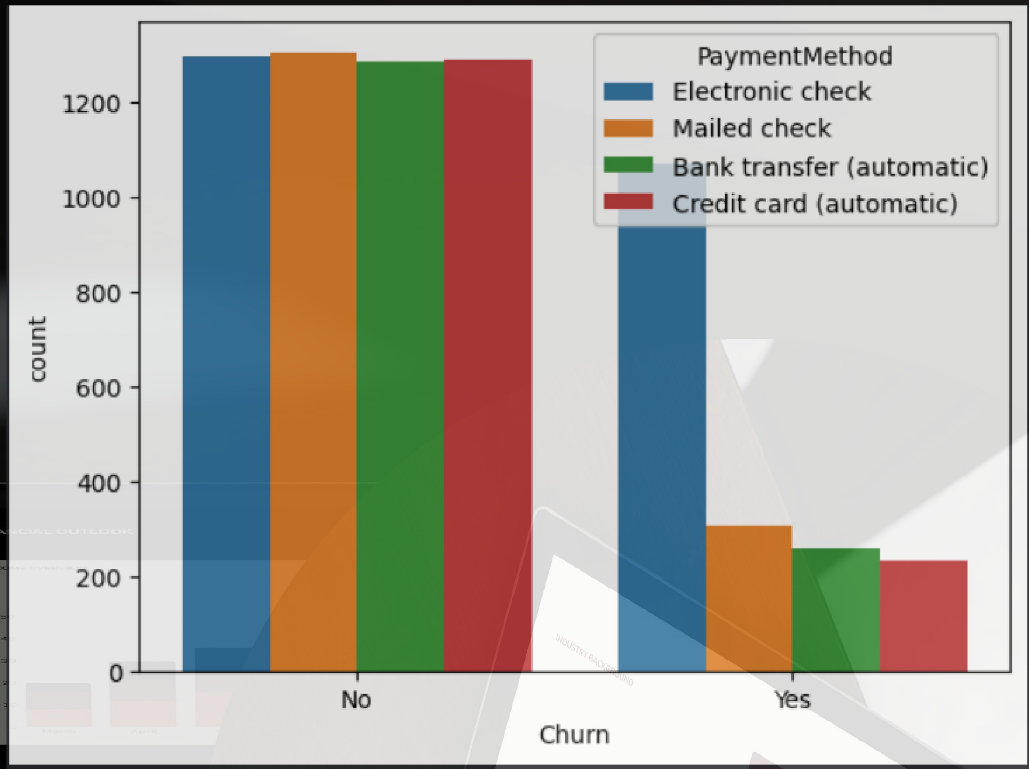
HIGH RISK CHANNEL

Electronic Check users represent the highest churn risk group. Nearly half of these customers leave, significantly higher than automatic payment methods.

Electronic Check	:	High Risk
Mailed Check	:	Medium
Credit Card (Auto)	:	Low Risk

Strategic Action

Migrate Electronic Check users to Autopay.



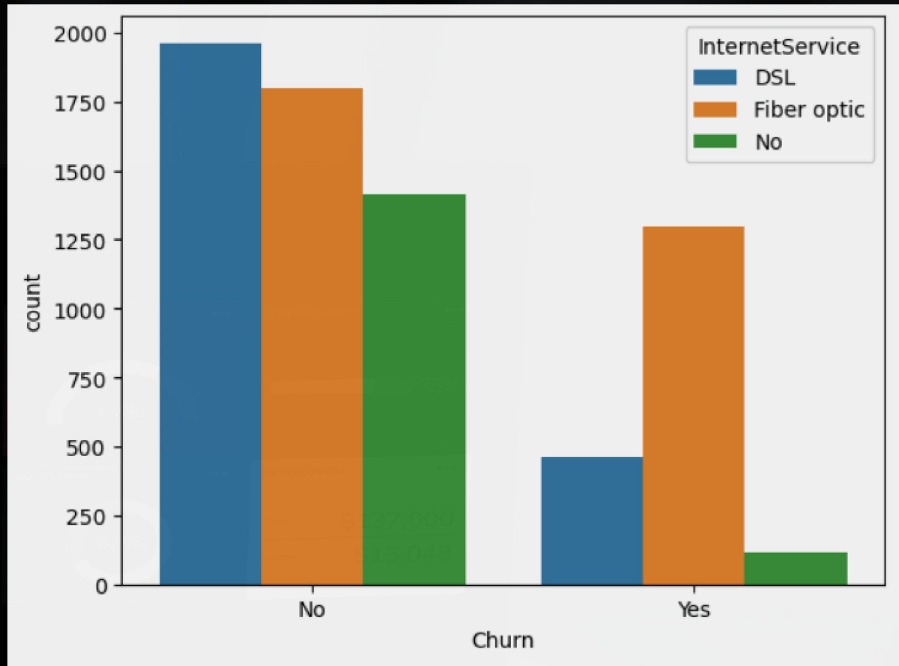
INTERNET TYPE SERVICE IMPACT

Fiber Optic customers show significantly higher churn rates despite being a premium service. This indicates potential price-value mismatch or technical reliability issues.

Fiber Optic	:	High Risk
DSL	:	Moderate
No Internet	:	Low Risk

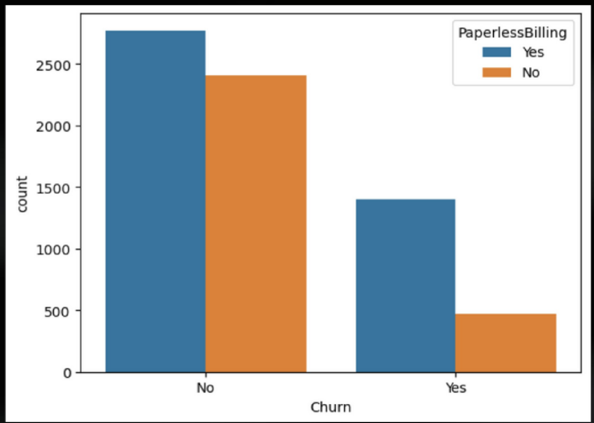
Strategic Action

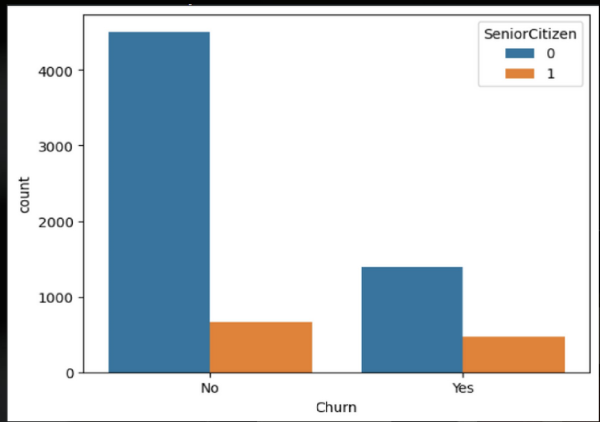
Investigate fiber quality & support experience.

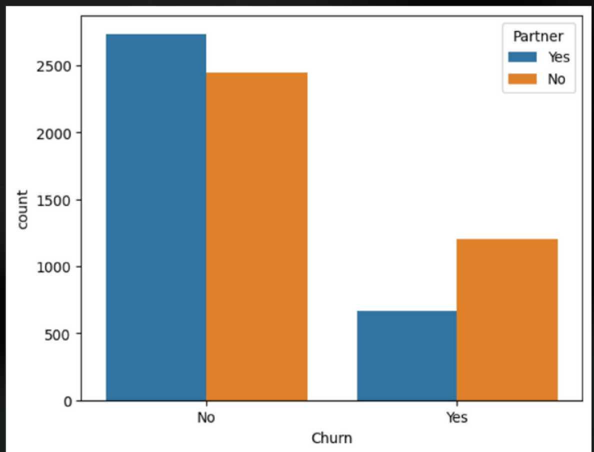


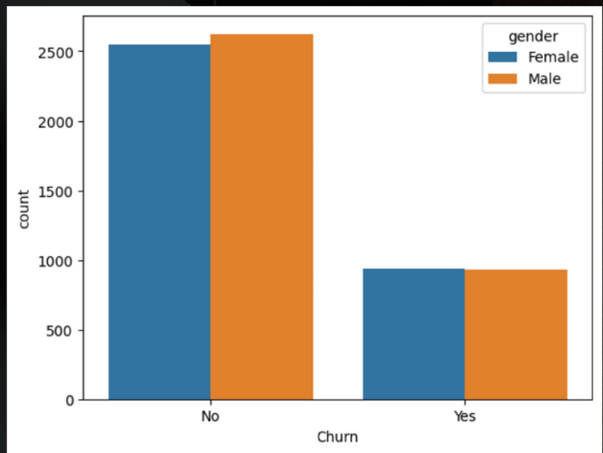
INTERNET SERVICE

DEMOGRAPHICS & BILLING INSIGHTS









Strategic Implication

- **Solo customers** (no partner/dependents) and **Senior Citizens** represent the highest demographic risk segments.
- **Family** ties create natural retention barriers.

GENDER NEUTRALITY

Minimal impact on churn behavior Not a key differentiator.

CHURN PREDICTION MODEL

- Model Used: Logistic Regression
- Inputs: Tenure, Monthly Charges, Services, Contract, etc.
- Output: Probability of churn (0 to 1)
- Goal: Identify high-risk customers early

```
# Split the dataset into features (X) and target (y)
X = df.drop('Churn', axis=1)
y = df['Churn']
```

```
# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Convert 'TotalCharges' column to float, and handle errors='coerce' to replace non-numeric values with NaN
X_train['TotalCharges'] = pd.to_numeric(X_train['TotalCharges'], errors='coerce')
X_test['TotalCharges'] = pd.to_numeric(X_test['TotalCharges'], errors='coerce')
```

```
# Replace missing values with the mean
X_train = X_train.fillna(X_train.mean(numeric_only=True))
X_test = X_test.fillna(X_train.mean(numeric_only=True))
```

```
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

Logistic Regression

```
lg = LogisticRegression()                #initialize
lg.fit(X_train,y_train)                  #train model
y_pred = lg.predict(X_test)
y_pred
print(classification_report(y_test,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.81      0.91      0.86      1036
           1       0.62      0.39      0.48       373

    accuracy                           0.78      1409
   macro avg       0.71      0.65      0.67      1409
weighted avg       0.76      0.78      0.76      1409
```

Accuracy score

```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

```
0.7757274662881476
```

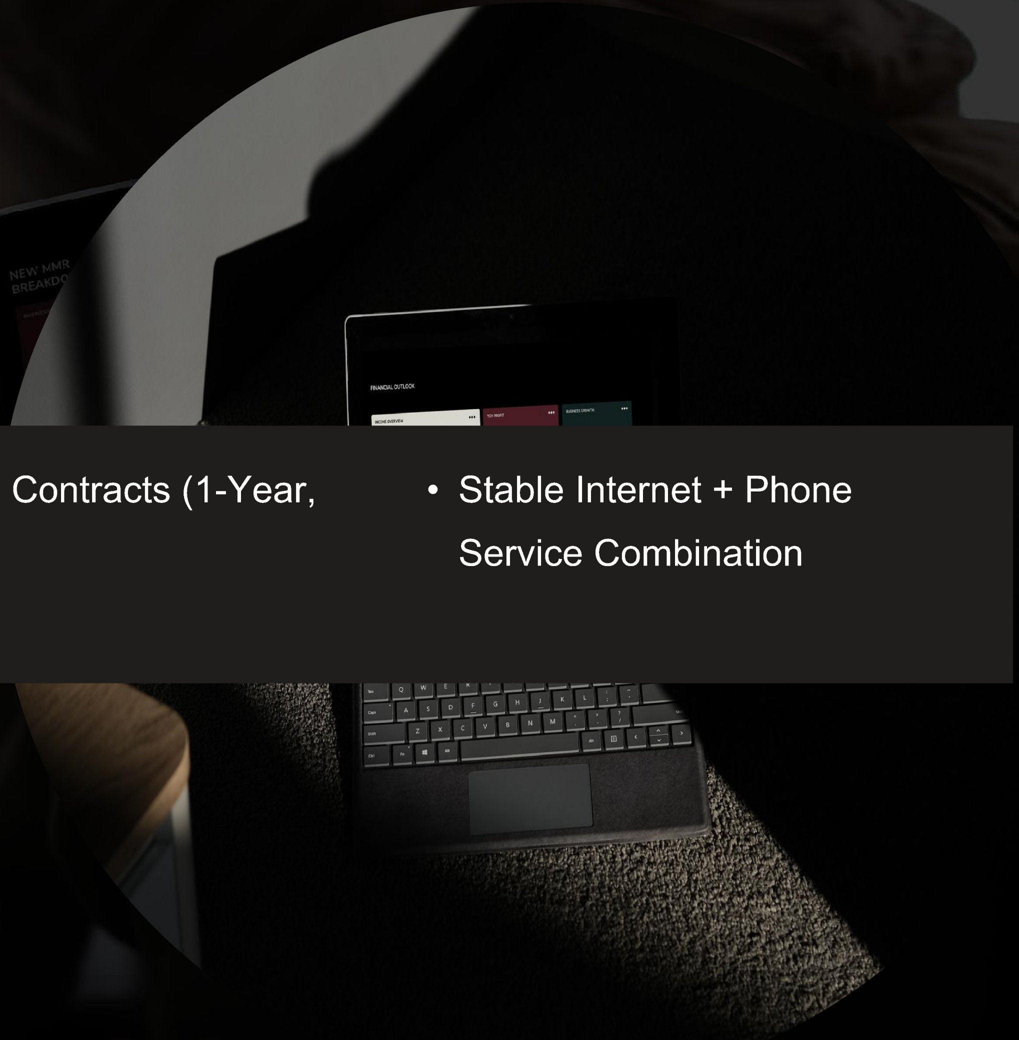
click to see the source code

KEY CHURN DRIVERS

- MONTH-TO-MONTH CONTRACTS → HIGHEST CHURN RISK
- High Monthly Charges → Increased churn probability
- Low Tenure → New customers at higher risk
- Fiber Optic Internet Users → More likely to churn

DRIVERS THAT REDUCE CHURN

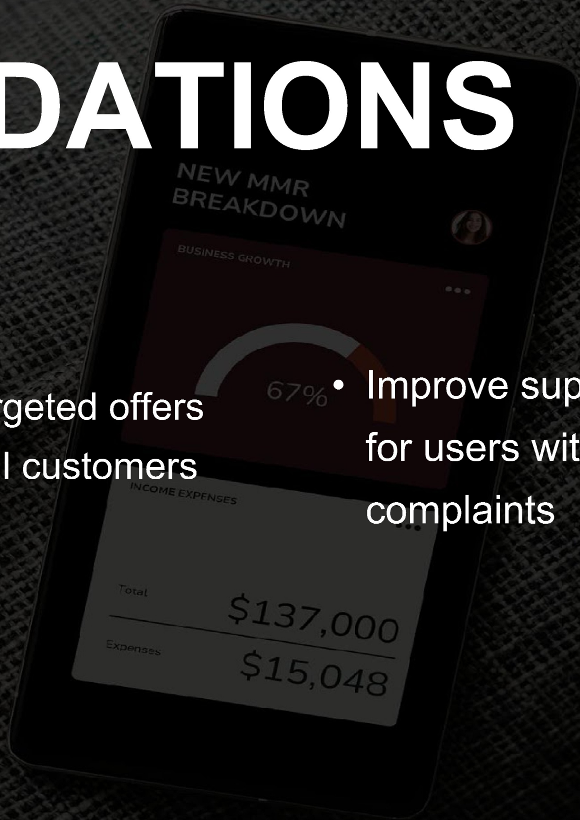
- Auto-Pay Billing Methods
- Long-term Contracts (1-Year, 2-Year)
- Stable Internet + Phone Service Combination



RETENTION STRATEGY RECOMMENDATIONS

RECOMMENDATIONS:

- Convert Month-to-Month users to annual plans using discounts
- Provide targeted offers for high-bill customers
- Improve support quality for users with frequent complaints
- Promote auto-pay & paperless billing enrollment



CONCLUSIONS

```
    # Create a DataFrame from the dictionary
    df = pd.DataFrame(data)


    # Encode the categorical columns
    categorical_columns = ['gender', 'SeniorCitizen', 'Partner', 'Dependents', 'PhoneService', 'MultipleLines', 'Contract']
    for column in categorical_columns:
        df[column] = label_encoder.fit_transform(df[column])
    df = scaler.fit_transform(df)

    result = lg.predict(df).reshape(1,-1)
    return result[0]
gender = "Female"
Seniorcitizen = "No"
Partner = "Yes"
Dependents = "No"
tenure = 1
Phoneservice="No"
multiline = "No phone service"
contact="Month-to-month"
totalcharge = 29.85
result = prediction(gender,Seniorcitizen,Partner,Dependents,tenure,Phoneservice,multiline,contact,totalcharge)

if result==1:
    print('churn')
else:
    print('not churn')
```

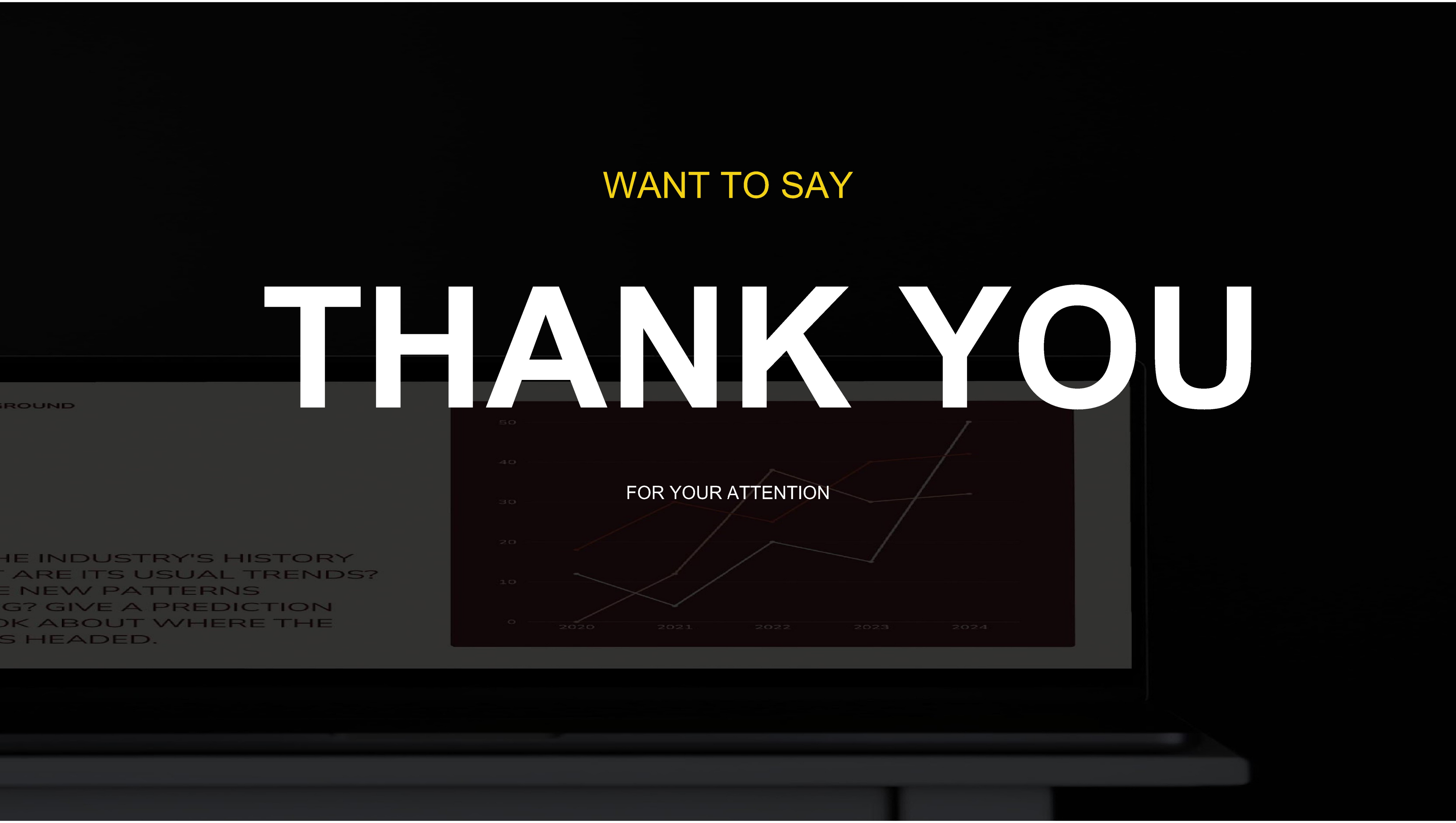
```
not churn
```

Accuracy score

```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

```
0.7757274662881476
```

- Logistic Regression model successfully predicts churn
- Key churn drivers identified clearly
- Practical retention strategies recommended


WANT TO SAY
THANK YOU
FOR YOUR ATTENTION