📘 Customer Churn Analysis Report

1. Executive Summary

The organization currently serves **6,418 customers**, with **1,732 churned**, resulting in a **27% churn rate**. This level of churn poses a significant risk to revenue stability and long-term growth.

Analysis across all Power BI pages reveals that churn is driven primarily by:

- **Competitive pressure**
- **Service dissatisfaction**
- **Month-to-month contracts**
- **Lack of service add-ons**
- **Fiber Optic service issues**
- **Manual payment methods**

The **Churn Prediction** model identifies **374 customers** at high risk of leaving, enabling targeted retention strategies.

The **Churn Reason** page shows that **competitor-related reasons** and **support/service attitude issues** are the top drivers of churn.

Churn_Reason	Total Churn
Attitude of service provider	93
Attitude of support person	208
Competitor had better devices	289
Competitor made better offer	274
Competitor offered higher download speeds	92
Competitor offered more data	106
Deceased	5

Churn_Reason	Total Churn
Don't know	124
Extra data charges	34
Lack of affordable download/upload speed	28
Lack of self-service on Website	27
Limited range of services	33
Long distance charges	62
Moved	45

Network reliability	66
Others	0
Poor expertise of online support	30
Poor expertise of phone support	12
Price too high	72
Product dissatisfaction	71
Service dissatisfaction	61
Total	**1,732**

2. Introduction

The purpose of this report is to analyze customer churn patterns using the organization's Power BI dashboards and identify:

- Why customers are leaving
- Which customer segments are most at risk
- What factors most strongly predict churn
- How to reduce churn through targeted interventions

This report integrates insights from three Power BI pages:

1. **Summary** – Overall churn pattern
2. **Churn Reason** – Customer-reported reasons for leaving
3. **Churn Prediction** – Machine Learning Model-based future churn risk

The goal is to provide a data-driven foundation for strategic decision-making and customer retention planning.

3. Findings

3.1 Overall Churn Profile (Summary Page)

Demographics

- Female customers account for **64.15%** of churn.
- Highest churn among **<20** and **50+** age groups (31% each).
- Tenure does not significantly reduce churn; all groups show 26–28% churn rate (26.88% in average).

Geography

Highest churn rates (Top 5):

- Pennsylvania (38.2%)
- Iowa (36.1%)
- Ohio (32.8%)
- Utah (31.6%)
- Florida (31.5%)

Internet Service Type Churn Rates

- Fiber Optic - **41.1%** (highest).
- Cable - **25.7%**
- DSL - **19.4%**
- None (Only phone service) - **7.8%** (lowest).

Service Add-Ons

Customers **without** add-ons show extremely high churn:

- Online Security: **84.6%**
- Premium Support: **83.5%**
- Online Backup: **71.9%**
- Device Protection: **71.0%**

Contracts

- Month-to-month contracts: **46.5% churn**
- One-year contracts: **11% churn**
- Two-year contracts: **2.7% churn**

Billing

- Mailed check users: **37.8% churn**
- Bank Withdrawal: **34.4% churn**
- Credit card users: **14.8% churn**

3.2 Churn Reasons (Churn Reason Page)

The Churn Reason page highlights the **specific reasons customers gave for leaving**:

Top Competitor-Related Reasons

Churn_Reason	**Total Churn**
Competitor had better devices	**289**
Competitor made better offer	**274**
Competitor offered more data	**106**
Competitor offered higher download speeds	**92**

Insight:

Competitors outperform in pricing, devices, and data packages — a major churn driver.

Service & Support Issues

Churn_Reason	**Total Churn**
Attitude of support person	**208**
Attitude of service provider	**93**
Poor expertise of online support	**30**
Poor expertise of phone	**12**

Insight:

Customer service quality is a significant pain point.

Product & Pricing Issues

Churn_Reason	Total Churn
Price too high	**72**
Product dissatisfaction	**71**
Extra data charges	**34**
Limited range of services	**33**

Network & Reliability

- Network reliability issues: **66**
- Lack of affordable speeds: **28**

Other Reasons

- Don't know: **124**
- Moved: **45**
- Deceased: **5**

Overall Insight:

The majority of churn reasons fall into **competitor advantage**, **service dissatisfaction**, and **pricing concerns**.

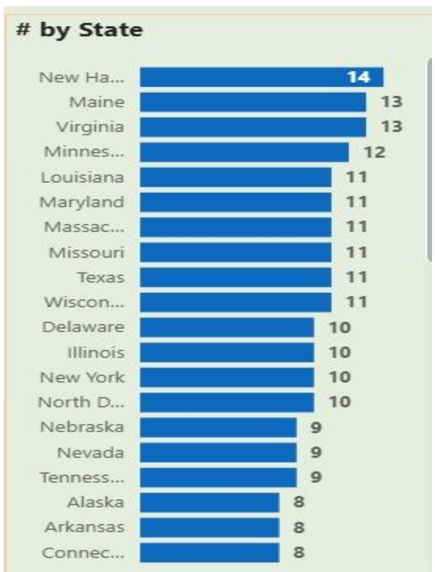
3.3 Predicted Churn (Churn Prediction Page)

Predicted Churn Volume

- **374 customers** predicted to churn.

High-Risk Segments

- **Month-to-month contracts:** 356 predicted churners
- **Older customers (50+):** 135 predicted churners
- **Credit card users:** 189 predicted churners
- **States and the prediction churn**



Customer-Level Risk

The prediction table identifies individual customers with:

- Monthly charges
- Total revenue
- Refund history
- Referral count

This enables **targeted retention outreach**.

4. Conclusion

The analysis shows that churn is driven by a combination of:

- **Competitive disadvantages** in pricing, devices, and data offerings
- **Customer service and support issues**
- **High churn among month-to-month contract holders**
- **Low adoption of service add-ons**
- **Fiber Optic service dissatisfaction**
- **Geographic hotspots with elevated churn**

The prediction model confirms these patterns and highlights **374 customers** who are likely to churn soon.

Addressing these issues will significantly reduce churn and improve customer lifetime value.

5. Recommendations

5.1 Strengthen Competitive Position

- Review competitor pricing and device offerings in high-churn states.
- Introduce competitive bundles (higher speeds + more data).
- Offer loyalty discounts to at-risk customers.

5.2 Improve Customer Service Quality

- Retrain support teams to address attitude-related complaints.
- Implement customer service quality monitoring.
- Introduce faster escalation paths for technical issues.

5.3 Increase Add-On Adoption

- Bundle Online Security, Backup, and Device Protection at a discount.
- Auto-enroll new customers with opt-out options.
- Promote add-ons during customer onboarding.

5.4 Reduce Month-to-Month Churn

- Offer incentives for 12- or 24-month contracts.
- Provide device upgrades for contract renewals.
- Target predicted churners with personalized offers.

5.5 Improve Fiber Customer Experience

- Conduct service quality audits in Fiber regions.
- Offer Fiber-specific loyalty rewards.
- Address network reliability issues.

5.6 Target High-Risk Customers

- Prioritize outreach to the **374 predicted churners**.
- Use customer-level data (revenue, refunds, referrals) to personalize offers.
- Focus on high-value customers first.

6. Methodology

6.1 SQL QUERY DOCUMENTATION

a. Gender percentage

```
SELECT Gender, COUNT(Gender) as TotalNumGender,
(COUNT(Gender) * 100.0/ (SELECT COUNT(*) FROM dbo.stage_churn)) as Percentage
FROM dbo.stage_churn
Group by Gender
```

	Gender	TotalNumGender	Percentage
1	Male	2370	36.927391710813
2	Female	4048	63.072608289186

b. Contract percentage

```
SELECT Contract, COUNT(Contract) as TotalContract,
COUNT(Contract) * 100.0/(SELECT COUNT(*) FROM dbo.stage_churn) as Percentage
FROM dbo.stage_churn
GROUP by Contract
```

	Contract	TotalContract	Percentage
1	Month-to-Month	3286	51.199750701153
2	One Year	1413	22.016204425054
3	Two Year	1719	26.784044873792

C. Customer_status revenue percentage

```
SELECT Customer_Status, COUNT(Customer_Status) as TotalCustomer_Status,
SUM(Total_Revenue) as Total_Revenue,
SUM(Total_Revenue) * 100.0/ (SELECT SUM(Total_Revenue) FROM dbo.stage_churn) AS
Percentage
FROM dbo.stage_churn
GROUP by Customer_Status
```

	Customer_Status	TotalCustomer_Status	Total_Revenue	Percentage
1	Joined	411	49281.5598697662	0.253097281975677
2	Churned	1732	3411960.5796299	17.5229426827105
3	Stayed	4275	16010148.2622757	82.2239600353138

D. Highest-lowest Customer contribution by state

```
SELECT State, COUNT(State) as TotalNumState,
COUNT(State) * 100.0/(SELECT COUNT(*) FROM dbo.stage_churn) as Percentage
FROM dbo.stage_churn
GROUP by State
ORDER BY Percentage DESC;
```

	State	TotalNumState	Percentage
1	Texas	242	3.770645060766
2	Kansas	146	2.274851978809
3	Massachusetts	144	2.243689622935
4	New York	141	2.196946089124
5	Virginia	139	2.165783733250
6	Louisiana	137	2.134621377376
7	Missouri	137	2.134621377376
8	Utah	136	2.119040199439
9	California	135	2.103459021502
10	Wyoming	135	2.103459021502
11	Arizona	134	2.087877843564
12	Illinois	134	2.087877843564
13	Delaware	134	2.087877843564
14	South Carolina	134	2.087877843564
15	Oregon	133	2.072296665627
16	Hawaii	133	2.072296665627
17	Maine	132	2.056715487690
18	Maryland	131	2.041134309753
19	Minnesota	131	2.041134309753
20	Alabama	131	2.041134309753
21	Rhode Island	131	2.041134309753
22	Nevada	130	2.025553131816
23	Florida	130	2.025553131816
24	Georgia	128	1.994390775942
25	Colorado	128	1.994390775942
26	Nebraska	127	1.978809598005
27	Arkansas	126	1.963228420068
28	Tennessee	126	1.963228420068
29	North Carolina	124	1.932066064194
30	North Dakota	124	1.932066064194
31	Indiana	123	1.916484886257
32	Iowa	122	1.900903708320
33	South Dakota	122	1.900903708320
34	Ohio	122	1.900903708320
35	Wisconsin	120	1.869741352446
36	Kentucky	120	1.869741352446
37	New Mexico	119	1.854160174509
38	Montana	118	1.838578996572
39	Michigan	118	1.838578996572
40	Washington	118	1.838578996572
41	Oklahoma	116	1.807416640698
42	New Hamps...	115	1.791835462760
43	Vermont	115	1.791835462760
44	Mississippi	113	1.760673106886
45	New Jersey	113	1.760673106886
46	Connecticut	112	1.745091928949
47	Idaho	111	1.729510751012
48	Pennsylvania	110	1.713929573075
49	West Virginia	110	1.713929573075
50	Alaska	108	1.682767217201

E. Moving database

Once the raw dataset is cleaned, move the churn database to the production database. The database is now ready to be used in PowerBI as well as for prediction purposes (Machine Learning).

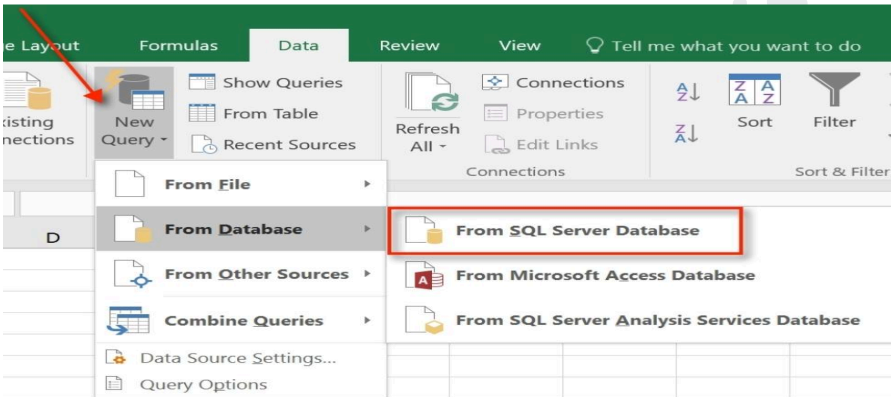
```sql
SELECT
	Customer_ID,
	Gender,
	Age,
	Married,
	State,
	Number_of_Referrals,
	Tenure_in_Months,
	ISNULL(Value_Deal, 'None') AS Value_Deal,
	Phone_Service,
	ISNULL(Multiple_Lines, 'No') AS Multiple_Lines,
	Internet_Service,
	ISNULL(Internet_Type, 'None') AS Internet_Type,
	ISNULL(Online_Security, 'No') AS Online_Security,
	ISNULL(Online_Backup, 'No') AS Online_Backup,
	ISNULL(Device_Protection_Plan, 'No') AS Device_Protection_Plan,
	ISNULL(Premium_Support, 'No') AS Premium_Support,
	ISNULL(Streaming_TV, 'No') AS Streaming_TV,
	ISNULL(Streaming_Movies, 'No') AS Streaming_Movies,
	ISNULL(Streaming_Music, 'No') AS Streaming_Music,
	ISNULL(Unlimited_Data, 'No') AS Unlimited_Data,
	Contract,
	Paperless_Billing,
	Payment_Method,
	Monthly_Charge,
	Total_Charges,
	Total_Refunds,
	Total_Extra_Data_Charges,
	Total_Long_Distance_Charges,
	Total_Revenue,
	Customer_Status,
	ISNULL(Churn_Category, 'Others') AS Churn_Category,
	ISNULL(Churn_Reason, 'Others') AS Churn_Reason

INTO [churn_db].[dbo].[production_churn]
FROM [churn_db].[dbo].[stage_churn]
```

F. View Syntax

View is used to pull data from Churned and Stayed vs Joined. Joined customers' dataset are not labeled data yet, as it does not have historical outcomes or do not have target labels for models to learn and train_test. Pulling Churned and Stayed in the same csv file and Joined data separately is crucial at this point. Where, Churned : 1 and Stayed: 0 encoded for training. Joined customers are the prediction dataset, not the training dataset.

```sql
CREATE VIEW vw_churnData as
	SELECT * FROM dbo.production_churn
	WHERE Customer_Status IN ('Churned', 'Stayed')


CREATE VIEW vw_JoinData as
	SELECT * FROM dbo.production_churn
	WHERE Customer_Status = 'Joined'
```

View data can be connected into excel and then converted into csv for Machine Learning and PowerBI reports



6.2 MACHINE LEARNING PREDICTION DOCUMENTATION

LINK: [CLICK] to check the full code. The only important part is covered below.

a. Data Preparation

First load the " vw_churnData.csv " in google collab

```python
from google.colab import files

# Upload the CSV file
 uploaded = files.upload()
```

b. Collecting list of columns to be label encoded (converting category to numerical values)

```
columns_to_encode = ['Gender', 'Married', 'State', 'Value_Deal',
'Phone_Service','Multiple_Lines','Internet_Service', 'Internet_Type',
'Online_Security','Online_Backup','Device_Protection_Plan',
'Premium_Support', 'Streaming_TV', 'Streaming_Movies','Streaming_Music',
'Unlimited_Data', 'Contract', 'Paperless_Billing','Payment_Method']
```

c. Encoding categorical variables but not target variables i.e Customer_Status.

```
label_encoders = {}
for column in columns_to_encode:
    label_encoders[column] = LabelEncoder()
    data[column] = label_encoders[column].fit_transform(data[column])
```

d. Manually encode the target variable 'Customer_Status'

```
data['Customer_Status'] = data['Customer_Status'].map({'Stayed': 0,
'Churned': 1})
```

e. Model Development

```
# Split data into features and target
X = data.drop('Customer_Status', axis=1)
y = data['Customer_Status']

# 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)
```

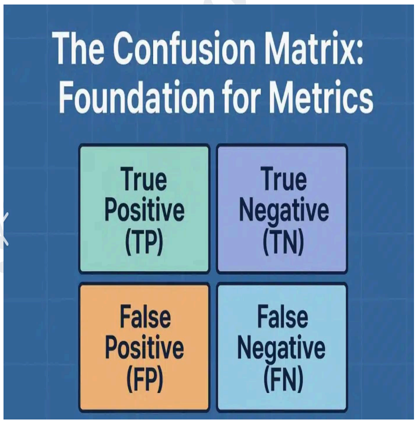
```
# Training Random Forest Model
# Initializing the Random Forest Classifier
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)

# Training the model
rf_model.fit(X_train, y_train)
```

f. Model Evaluation

```python
# Evaluate Model
# Make predictions
y_pred = rf_model.predict(X_test)

# Confustion Matrix
print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

# Classification Report
print("Classification Report:")
print(classification_report(y_test, y_pred))
```

What does this evaluation indicate?

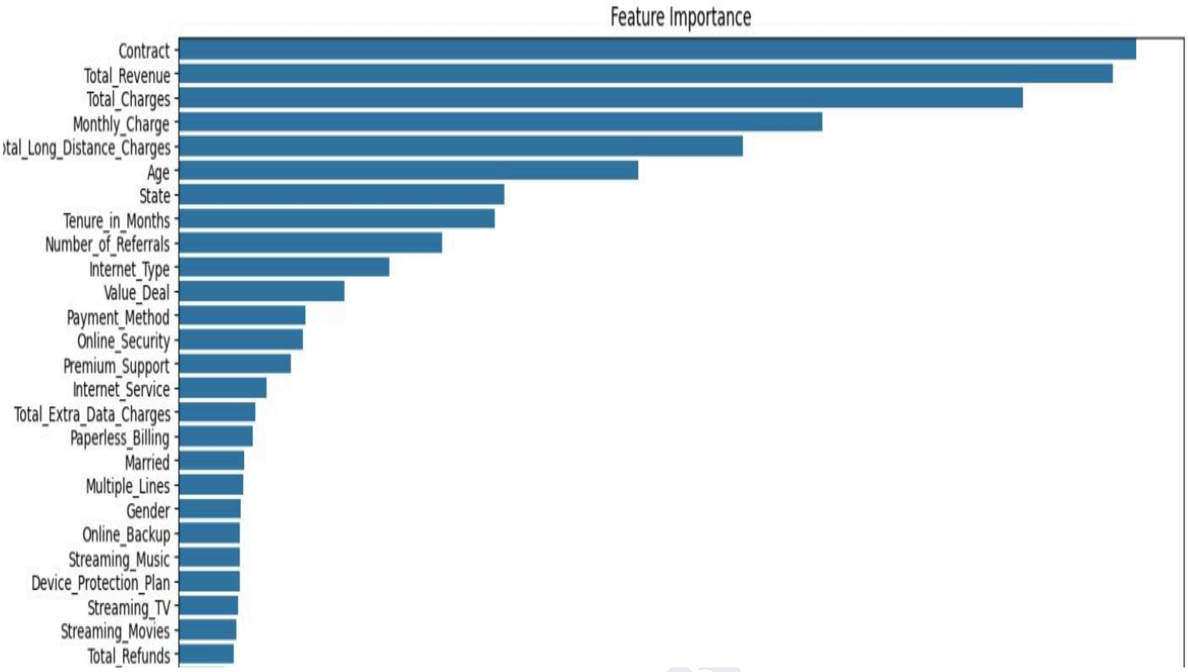
```
Confusion Matrix:
[[801  52]
 [122 227]]
Classification Report:
              precision    recall  f1-score   support

           0       0.87      0.94      0.90       853
           1       0.81      0.65      0.72       349

    accuracy                           0.86      1202
   macro avg       0.84      0.79      0.81      1202
weighted avg       0.85      0.86      0.85      1202
```



Confusion Matrix:
The confusion matrix is a table that is used to evaluate the performance of the classification model. The confusion matrix report indicates that 801, no churn customers are correctly identified and 227, yes churn customers are correctly identified. Whereas, 52 no churn customers are incorrectly predicted as yes churn and 122 yes churn customers are incorrectly predicted as no churn.

Classification Report:
It caught 65% of the actual churners (1), as the recall is 0.65. Precision 0.81 indicates when the model predicts someone would churn , it will be correct 81% of the time. The remaining 19% are the False Positive (FP)

g. Plotting the important features in descending order



h. load the " vw_JoinData.csv " in google collab

```
from google.colab import files

    # Upload the CSV file
    uploaded = files.upload()
```

i. Copy and store vw_joinData (df_data) in original_df before dropping unnecessary columns

```
# Copying the original Dataframe to prevent from getting encoded columns(0 and 1)
original_df = df_data.copy()
```

```
# Dropping unnecessary columns for prediction
df_data = df_data.drop(['Customer_ID',
'Churn_Category','Customer_Status','Churn_Reason'], axis=1)
```

Customer_ID is not a predictive feature. `'Churn_Category'` is empty, `'Customer_Status'` not valid for prediction and `'Churn_Reason'`, Keeping them would leak information into the model.

j. Encoding categorical variable to numerical values

```
# Encoding catergorical variables
for column in df_data.select_dtypes(include=['object']).columns:
    df_data[column] = label_encoders[column].transform(df_data[column])
```

k. These customers have no churn label yet, so run predictions.

```
# Predictions
new_predictions = rf_model.predict(df_data)

# Adding Predictions to the original Dataframe
original_df['Customer_Status_Predicted'] = new_predictions

# Filtering the dataframe to include only records predicted as "Churned"
original_df = original_df[original_df['Customer_Status_Predicted'] == 1]

# Saving the results and downloading file
original_df.to_csv('predicted_churned_customers.csv', index=False)
files.download('predicted_churned_customers.csv')
```

Although model evaluation and comparing a few model's performance would have given an idea to pick one best model, Random forest machine learning model is used because of its ensemble nature , robustness against overfitting, actionable insights through feature importance and many more. Churned and Stayed data are used, where their behavioral patterns and churn reasons plays an important part in predicting whether the newly joined customers are likely to churn or stay in the future.