

BUSINESS PROCESS ANALYSIS FOR CUSTOMER RETENTION

ETL Process on SQL Server

- Database Creation
- Table Creation
- View Creation

Power BI - Churn Summary

- Data Transformation
- Custom Measures
- Basic Visualization
- Advanced Visualization

Data Analysis Approach

ML Model - Random Forest

- Data Preparation
- Processing
- Model Building & Evaluation
- Prediction on Joiner Data

Power BI - Churn Prediction

- Churner Profile
- Customer at Risk



Tech Stack: SQL Server , Power BI , Python (Random Forest) , ETL , KPI Analytics

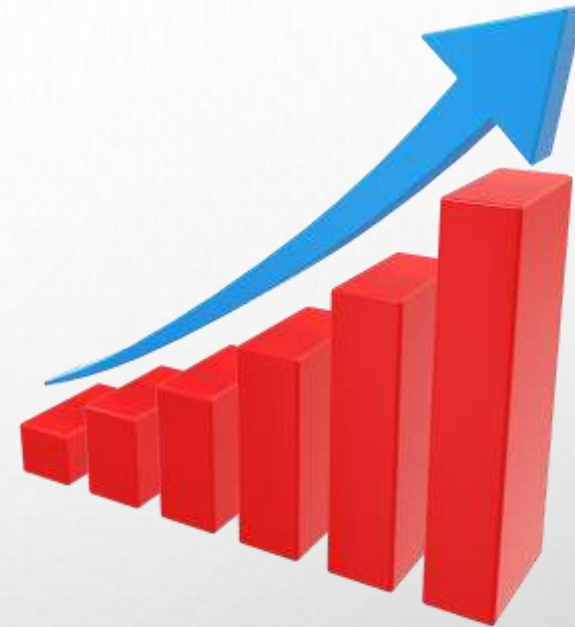
The Problem:

A telecom company is facing a high, undefined customer churn rate, leading to significant revenue loss and unstable growth.

Why?

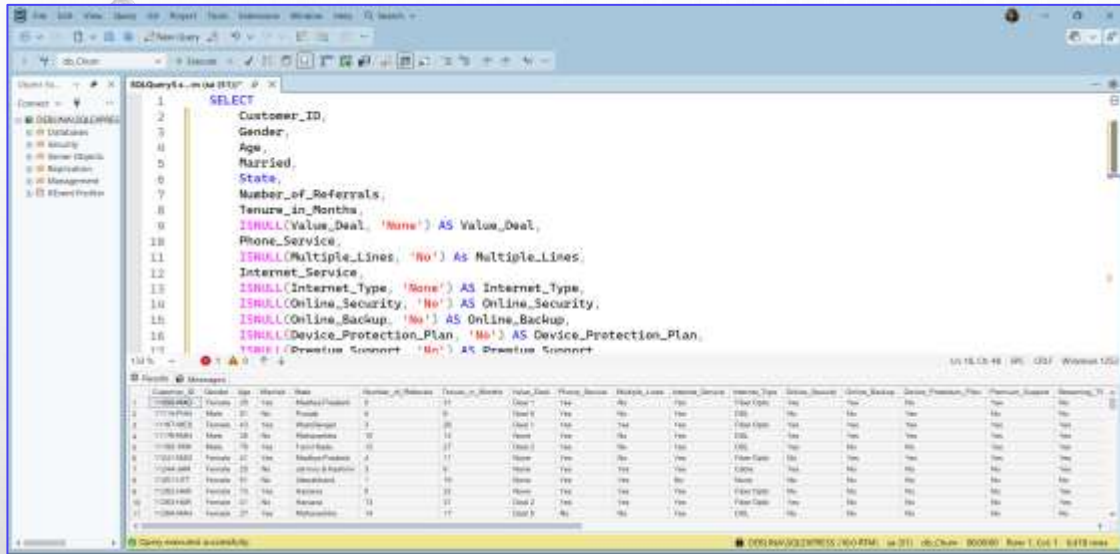
"The business had customer data, but no process to analyze it. This created a critical information gap:"

- **No Root Cause:** They didn't know the primary *drivers* of churn. (Was it price? Poor service? Contract terms?)
- **No Clear Profile:** They couldn't identify *who* was leaving. (Was it new customers? Old customers? Customers in a specific state?)
- **Reactive vs. Proactive:** All retention efforts were *reactive* (e.g., trying to win back someone who already left). There was no *proactive* system to stop it from happening.



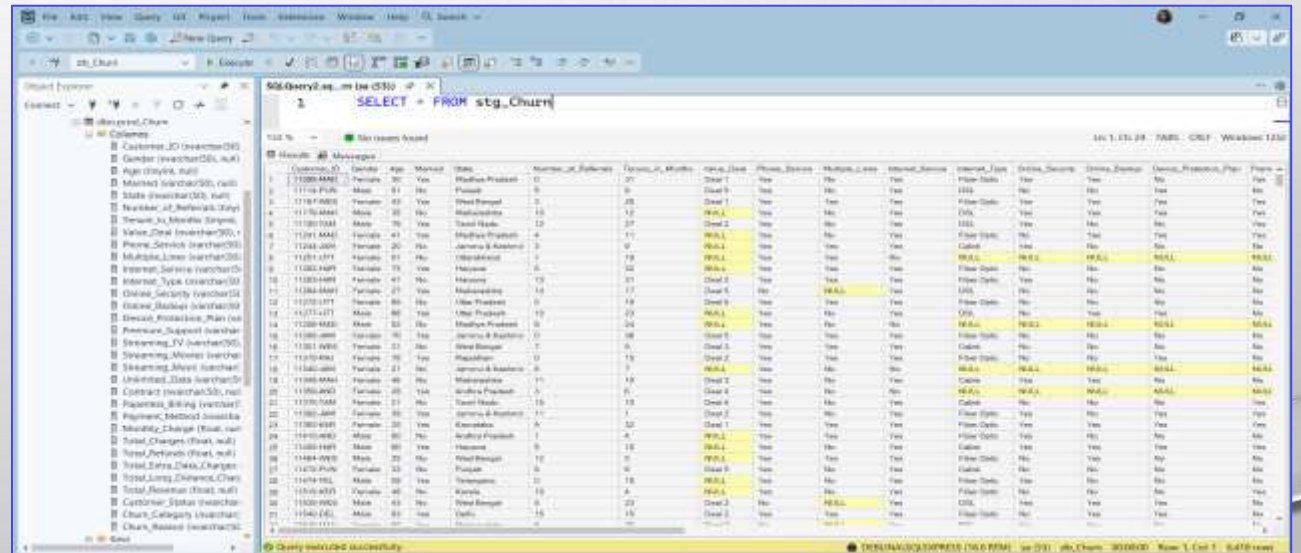
The Problem

To analyze the failing 'Customer Retention' process from end-to-end.



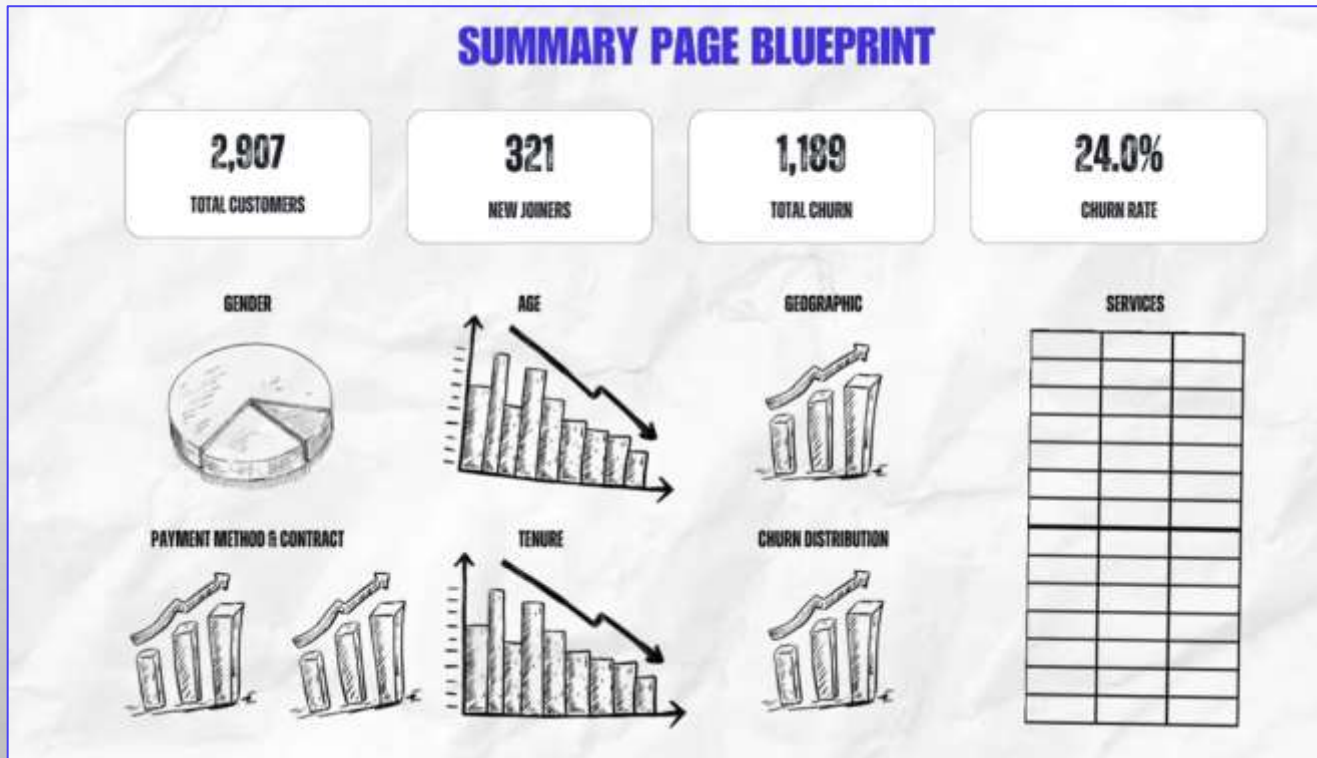
Remove null and insert the new data into Prod table

Checking to see we have all the data or not? – by select statement



POWER BI

To import the data of what we have done in the SQL Server and present it in Power BI perspective



Now, for the visualization, I have created a blueprint / agenda to create the first page. So, we have to cover a couple of metrics:

- I. Overall level of Numbers.
- II. Demographic
 - a. Gender related visuals
 - b. Account related visuals
 - c. Geographic
 - d. Churn Distribution
- III. Services used by the Customers

THE DASHBOARD

CHURN ANALYSIS - SUMMARY

Monthly Charge Range

All

Married

All

Churn Prediction

6,418

Total Customers

411

New Joiners

1,732

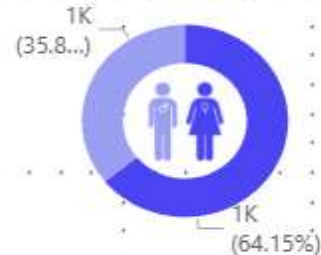
Total Churn

27.0%

Churn Rate



Total Churn by Gender



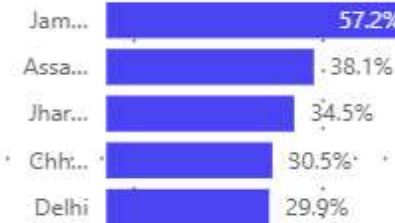
Gender
● Female
● Male

Total Customers and Churn Rate by Age Group



GEOGRAPHIC

Churn Rate by State

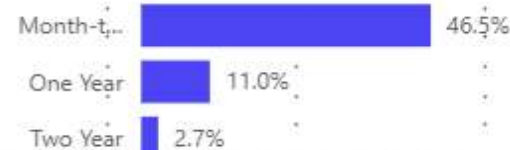


SERVICES USED

Churn Rate by Internet Type



Churn Rate by Contract



Total Customers and Churn Rate by Tenure Group



CHURN DISTRIBUTION

Total Churn by Churn Category



Churn Rate by Payment Method



Services	No	Yes
Device_Protection_Plan	71.02%	28.98%
Internet_Service	6.29%	93.71%
Multiple_Lines	54.79%	45.21%
Online_Backup	71.88%	28.12%
Online_Security	84.64%	15.36%
Paperless_Billing	25.40%	74.60%
Phone_Service	9.41%	90.59%
Premium_Support	83.49%	16.51%
Streaming_Movies	56.00%	44.00%
Streaming_Music	61.14%	38.86%
Streaming_TV	56.76%	43.24%
Unlimited_Data	19.92%	80.08%

CHURN ANALYSIS - SUMMARY

Monthly Charge Range

Married

All

All

Churn Prediction

6,418

Total Customers

411

New Joiners

1,732

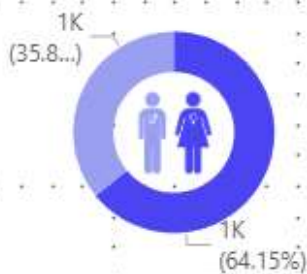
Total Churn

27.0%

Churn Rate



Total Churn by Gender

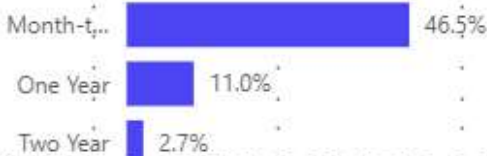


Gender

Female

Male

Churn Rate by Contract



Churn Rate by Payment Method



Total Customers

Total Customers

1.6K

23.5%

Total Customers

Total Customers

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

Total

761

DEMOGRAPHIC

State

57.2%

38.1%

34.5%

30.5%

29.9%

DISTRIBUTION

Churn Category

761

Attitude

301

Dissatisfaction

300

Price

196

Other

174

SERVICES USED

Churn Rate by Internet Type

Fiber Optic

41.1%

Cable

25.7%

DSL

19.4%

None

7.8%

Services

No

Yes

Device Protection Plan

71.02%

28.98%

Internet Service

6.29%

93.71%

Multiple Lines

54.79%

45.21%

Online Backup

71.88%

28.12%

Online Security

84.64%

15.36%

Paperless Billing

25.40%

74.60%

Phone Service

9.41%

90.59%

Premium Support

83.49%

16.51%

Streaming Movies

56.00%

44.00%

Streaming Music

61.14%

38.86%

Streaming TV

56.76%

43.24%

Unlimited Data

19.92%

80.08%

CHURN ANALYSIS - SUMMARY

Monthly Charge Range

All

Married

All

Churn Prediction

4,048

Total Customers

269

New Joiners

1,111

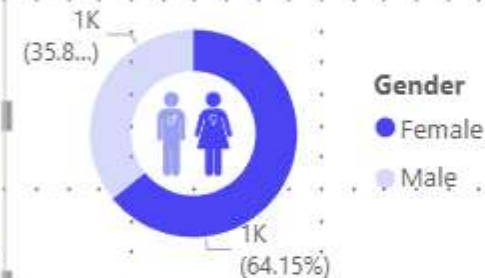
Total Churn

27.4%

Churn Rate



Total Churn by Gender

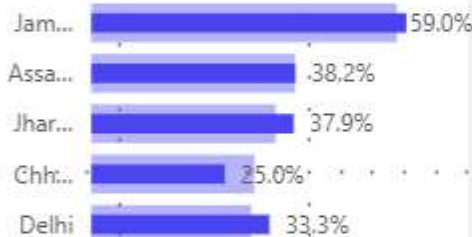


Total Customers and Churn Rate by Age Group



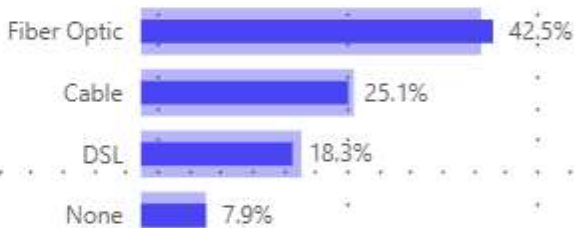
GEOGRAPHIC

Churn Rate by State

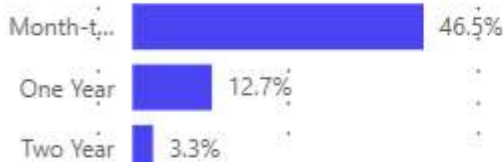


SERVICES USED

Churn Rate by Internet Type



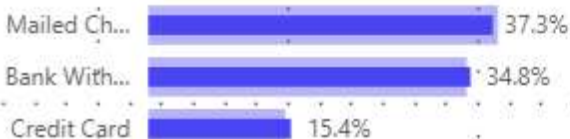
Churn Rate by Contract



Total Customers and Churn Rate by Tenure Group



Churn Rate by Payment Method



CHURN DISTRIBUTION

Total Churn by Churn Category



Services	No	Yes
Device_Protection_Plan	69.49%	30.51%
Internet_Service	6.39%	93.61%
Multiple_Lines	52.30%	47.70%
Online_Backup	70.66%	29.34%
Online_Security	83.26%	16.74%
Paperless_Billing	25.02%	74.98%
Phone_Service	8.19%	91.81%
Premium_Support	82.63%	17.37%
Streaming_Movies	54.46%	45.54%
Streaming_Music	59.05%	40.95%
Streaming_TV	55.63%	44.37%
Unlimited_Data	19.44%	80.56%

CHURN ANALYSIS - SUMMARY

Monthly Charge Range

All

Married

All

Churn Prediction

2,370

Total Customers

142

New Joiners

621

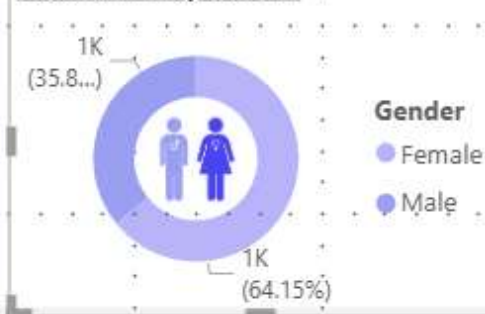
Total Churn

26.2%

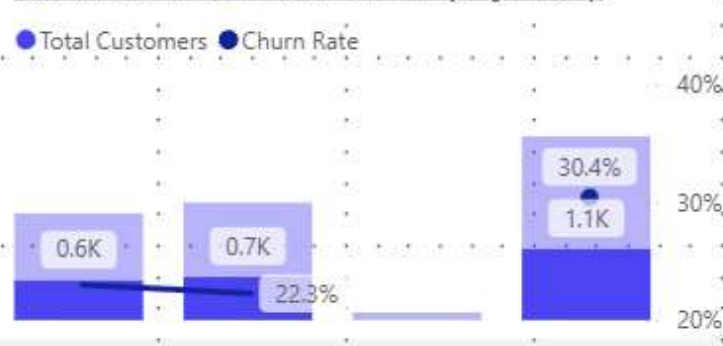
Churn Rate



Total Churn by Gender

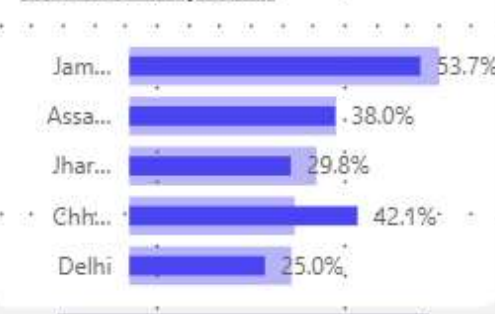


Total Customers and Churn Rate by Age Group



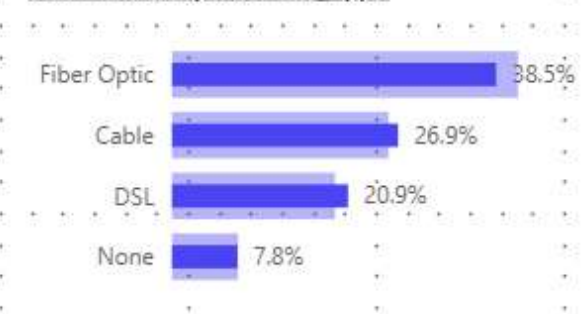
GEOGRAPHIC

Churn Rate by State

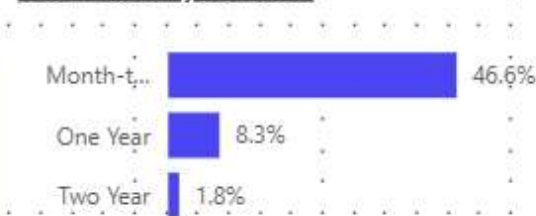


SERVICES USED

Churn Rate by Internet Type



Churn Rate by Contract



Total Customers and Churn Rate by Tenure Group

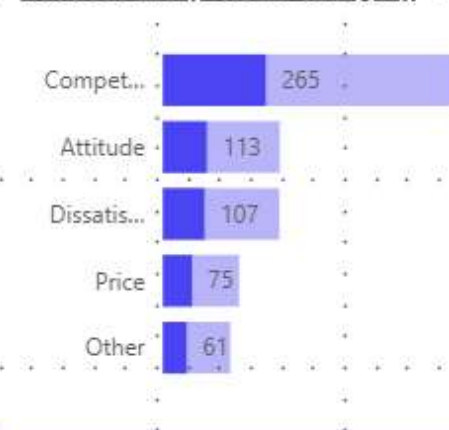


Churn Rate by Payment Method



CHURN DISTRIBUTION

Total Churn by Churn Category



Services	No	Yes
Device_Protection_Plan	73.75%	26.25%
Internet_Service	6.12%	93.88%
Multiple_Lines	59.26%	40.74%
Online_Backup	74.07%	25.93%
Online_Security	87.12%	12.88%
Paperless_Billing	26.09%	73.91%
Phone_Service	11.59%	88.41%
Premium_Support	85.02%	14.98%
Streaming_Movies	58.78%	41.22%
Streaming_Music	64.90%	35.10%
Streaming_TV	58.78%	41.22%
Unlimited_Data	20.77%	79.23%

CHURN ANALYSIS - SUMMARY

3,428

Total Customers

143

New Joiners

1,

Total

33.2%

Churn Rate

Churn Prediction

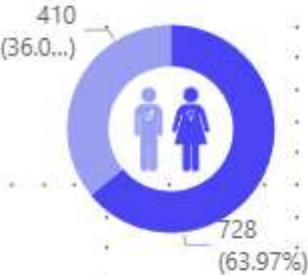


Monthly Charge Ran...
50-100
☐ < 20
☐ > 100
☐ 20-50
☒ 50-100

Married
All

DEMOGRAPHIC

Total Churn by Gender



Gender
● Female
● Male

Total Customers and Churn Rate by Age Group



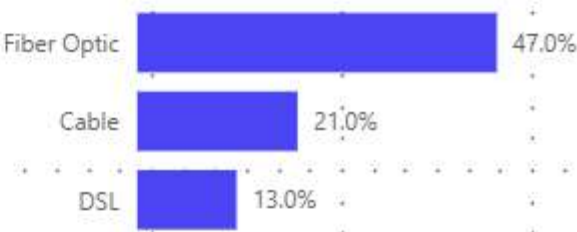
GEOGRAPHIC

Churn Rate by State



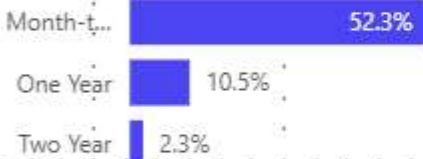
SERVICES USED

Churn Rate by Internet Type



ACCOUNT INFO

Churn Rate by Contract



Churn Rate by Payment Method



Total Customers and Churn Rate by Tenure Group



CHURN DISTRIBUTION

Total Churn by Churn Category



Services	No	Yes
Device_Protection_Plan	74.96%	25.04%
Internet_Service		100.00%
Multiple_Lines	51.41%	48.59%
Online_Backup	74.34%	25.66%
Online_Security	85.50%	14.50%
Paperless_Billing	22.14%	77.86%
Phone_Service	2.02%	97.98%
Premium_Support	85.15%	14.85%
Streaming_Movies	58.00%	42.00%
Streaming_Music	61.78%	38.22%
Streaming_TV	58.26%	41.74%
Unlimited_Data	14.32%	85.68%

CHURN ANALYSIS - SUMMARY

1,726

Total Customers

74

New Joiners

554

Total Churn

3

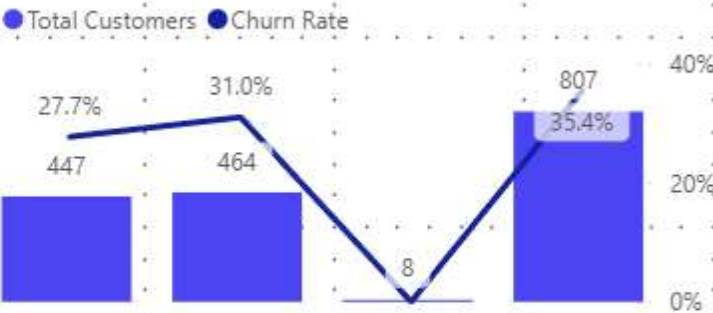
Monthly Charge Ran...
50-100

Married
Yes
No
Yes

Churn Prediction

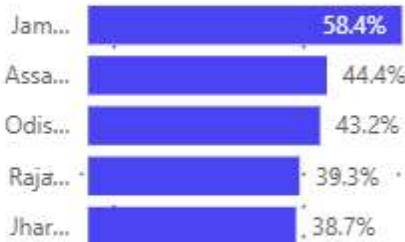


Total Customers and Churn Rate by Age Group



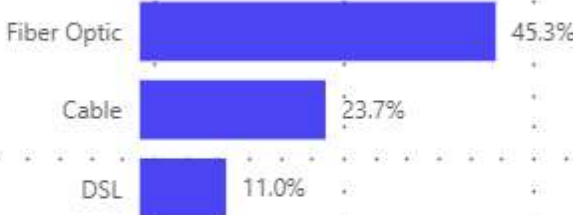
GEOGRAPHIC

Churn Rate by State

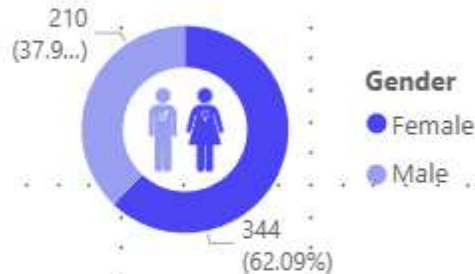


SERVICES USED

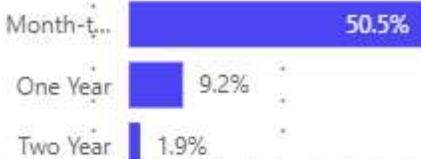
Churn Rate by Internet Type



Total Churn by Gender



Churn Rate by Contract



Churn Rate by Payment Method



Total Customers and Churn Rate by Tenure Group



CHURN DISTRIBUTION

Total Churn by Churn Category



Services	No	Yes
Device_Protection_Plan	74.55%	25.45%
Internet_Service		100.00%
Multiple_Lines	51.44%	48.56%
Online_Backup	72.38%	27.62%
Online_Security	84.66%	15.34%
Paperless_Billing	20.40%	79.60%
Phone_Service	2.53%	97.47%
Premium_Support	85.92%	14.08%
Streaming_Movies	57.22%	42.78%
Streaming_Music	60.47%	39.53%
Streaming_TV	61.19%	38.81%
Unlimited_Data	15.70%	84.30%

CHURN ANALYSIS - SUMMARY

Monthly Charge Ran...
50-100

Married

Yes

☐ No

☒ Yes

Churn Prediction

1,726

Total Customers

74

New Joiners

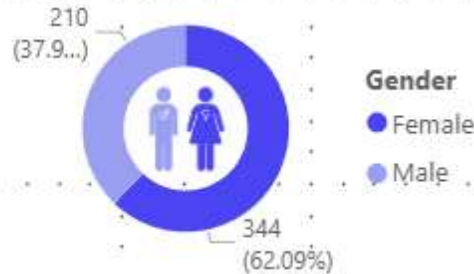
554

Total Churn

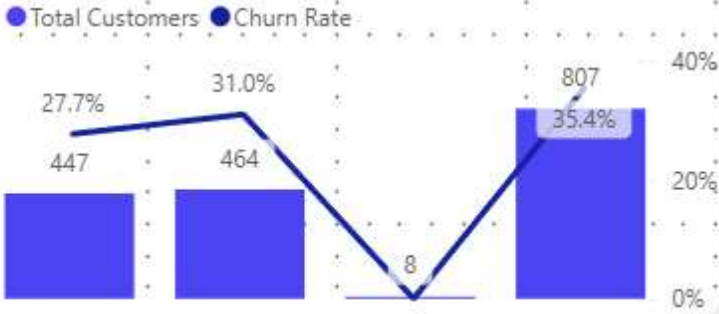
3



Total Churn by Gender



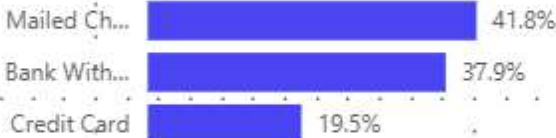
Total Customers and Churn Rate by Age Group



Churn Rate by Contract



Churn Rate by Payment Method



Total Customers and Churn Rate by Tenure Group

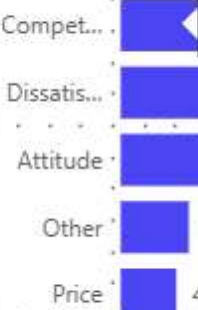


Churn Rate by State



CHURN DIS

Total Churn by Churn Reason



Churn_Reason	Total Churn
Competitor had better devices	102
Competitor made better offer	96
Competitor offered more data	34
Competitor offered higher download speeds	26
Total	258

Churn Reason	Percentage	Churn Rate
Online Backup	72.38%	27.62%
Online Security	84.66%	15.34%
Paperless Billing	20.40%	79.60%
Phone Service	2.53%	97.47%
Premium Support	85.92%	14.08%
Streaming Movies	57.22%	42.78%
Streaming Music	60.47%	39.53%
Streaming TV	61.19%	38.81%
Unlimited Data	15.70%	84.30%

WHAT'S DONE TILL NOW?

- I. ETL DESIGN
- II. DATA MODEL
- III. CUSTOM MEASURES AND METRICS
- IV. VISUALIZATION DASHBOARD
- V. EXPLORATORY DATA ANALYSIS
- VI. DESCRIPTIVE ANALYSIS

WHAT'S NEXT ?

MACHINE LEARNING ALGORITHM

Now, I will use **ML Algorithm** which we gonna use with this data and use that final machine learning output which will be a predictive output to predict future churners. I'm going to use **Random Forest** (one of the ML Algorithm)

**PLATFORM AND
PROGRAMMING
LANGUAGE
NEEDED**



THE EXCEL FILE WITH ALL THE CATEGORICAL DATA

HERE, WE HAVE IMPORTED BOTH THE CHURN_DATA AND THE JOIN DATA INTO 1 FILE. THEN WE WILL IMPORT THIS FILE FOR ML ALGORITHM.

Prediction.xlsx - Excel

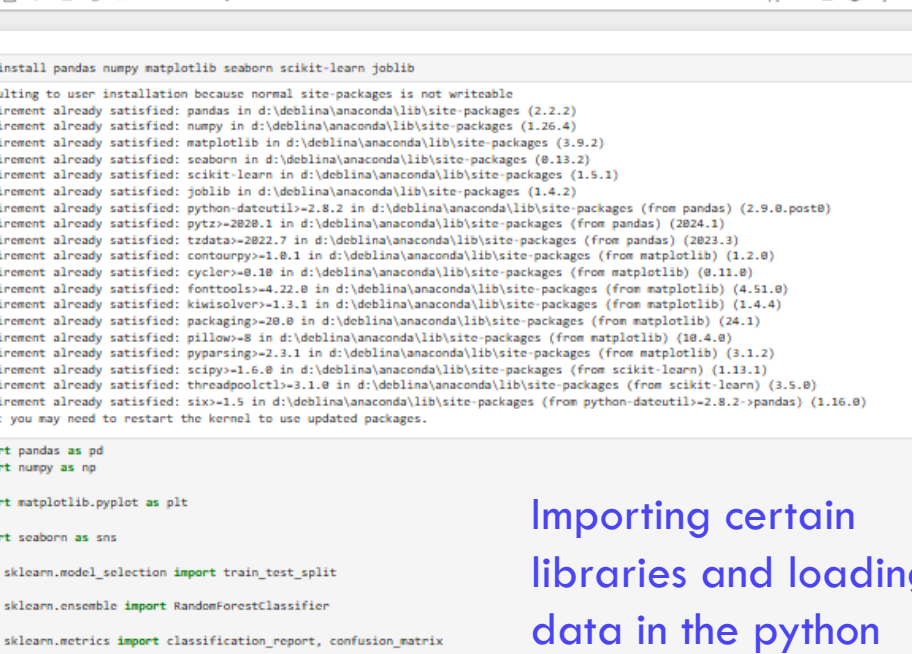
Deblina Mandal

FileHomeInsertPage LayoutFormulasDataReviewViewHelpPDFelementTell me what you want to do

CutCopyFormat Painter

Calibri11

<



The screenshot displays a JupyterLab environment. At the top, the 'Jupyter' logo and 'Customer Retention Last checkpoint 21 hours ago' are visible. The interface includes a menu bar (File, Edit, View, Run, Kernel, Settings, Help) and a toolbar with icons for file operations and execution. The main workspace is divided into two panes. The left pane shows a terminal window with the output of the command `pip install pandas numpy matplotlib seaborn scikit-learn joblib`. The output lists the installation of various packages and their dependencies, including pandas, numpy, matplotlib, seaborn, scikit-learn, joblib, python-dateutil, pytz, tzdata, contourpy, cycler, fonttools, kiwisolver, packaging, pillow, pyparsing, scipy, threadpoolctl, and six. The right pane shows a code cell with the following Python code:

```
[3]: import pandas as pd
import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.model_selection import train_test_split

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import classification_report, confusion_matrix

from sklearn.preprocessing import LabelEncoder

import joblib
```

Below the code cell, a comment indicates the next step: `# Defining the path to the Excel file`.

Overlaid on the right side of the image is a large blue text box containing the text: "Importing certain libraries and loading the data in the python environment".

2.

The screenshot shows a Jupyter Notebook interface with the following code and output:

```

# Define the path to the excel file
file_path = "D:\\data\\customer_data\\data\\new_checkout\\python\\project\\Product Recommendation\\python\\new_checkout\\customer_data\\ProductData.xlsx"

# Define the sheet name to read data from
sheet_name = 'new_checkout'

# Read the data from the specified sheet into a pandas dataframe
data = pd.read_excel(file_path, sheet_name=sheet_name)

# Display the first few rows of the fetched data
print(data.head(3))

```

The output displays three dataframes:

customer_data

customer_id	gender	Age	Married	State	Number_of_referrals	
0	11000-Male	Female	18	No	Madhya Pradesh	0
1	11114-Male	Male	31	No	Madhya Pradesh	0
2	11117-Female	Female	33	Yes	West Bengal	0
3	11179-Male	Male	43	No	Madhya Pradesh	10
4	11180-Female	Female	74	Yes	Madhya Pradesh	12

Insurance_data

Insurance_id	Months	value	Deal	Service	Multiple_Lines
0	13	Deal 1	Yes	No	---
1	0	Deal 1	Yes	No	---
2	26	Deal 1	Yes	No	---
3	12	Deal 1	Yes	No	---
4	21	Deal 2	Yes	No	---

Payment_data

Payment_Method	Monthly_Charge	Total_Charges	Total_Refunds	
0	Bank_Mithraoal	85.000000	3331.000000	0.00
1	Bank_Mithraoal	85.000000	100.000000	0.00
2	Bank_Mithraoal	110.000000	2277.000000	12.00
3	Credit_Late	85.000000	9999.000000	0.00
4	Credit_Late	72.000000	9999.000000	0.00

total_refunds_data

total_refunds_data	total_long_distance_charges	total_refunds
0	601.700000	1453.000000
1	122.000000	480.000000
2	1872.000000	1627.000000
3	239.000000	3333.000000
4	122.000000	1453.000000

customer_status_data

customer_status	status_category	status
0	Staged	Others
1	Staged	Others
2	Staged	Others
3	Staged	Others
4	Staged	Others

[n rows x 12 columns]

3.

Confusion Matrix:

```
[[ 984  161]
 [ 126 409]]
```

Classification Report:

	precision	recall	f1-score	support
0	0.88	0.82	0.85	977
1	0.78	0.64	0.71	454
accuracy			0.80	1382
macro avg	0.82	0.73	0.78	1382
weighted avg	0.80	0.69	0.75	1382

Feature Importance

Feature	Relative Importance
Content	0.24
Total Revenue	0.21
Recall_Changes	0.19
Monthly_Change	0.18
Total_Living_Months_Changes	0.17
Age	0.15
Months_in_Market	0.14
Scale	0.13
Months_of_Marketing	0.12
Market_Size	0.11
Market_Deal	0.10
Marketing_Method	0.09
Program_Support	0.08
Online_Security	0.07
Internet_Service	0.06
Response_Rating	0.05
Total_Fees_Charge_Changes	0.04
Market	0.03
Online_Marketing	0.02
Market_Protection_Risk	0.01
Multiple_Lines	0.01
Gender	0.01
Streaming_Music	0.01
Streaming_TV	0.01
Streaming_Games	0.01
Streaming_Movies	0.01
Total_Revenue	0.01
Phone_Service	0.01

4.

```

customer_id, gender, age, married, status, number_of_referrals
0 11731-105 female 28 No 12611 Status 0
1 12066-685 male 27 No next Referral 2
2 12166-051 female 25 Yes 666111110 2
3 12267-703 female 28 No 666110 0
4 12368-061 female 32 Yes 10611 0

feature_to_predict, value, deal, status, service, Multiple_Line, ...
0 0 Deal 0 No ...
1 20 30% Yes No ...
2 25 30% Yes No ...
3 3 30% Yes No ...
4 10 30% Yes No ...

Payment_Method, Monthly_Charge, Total_Charges, Initial_Months
0 Rollover Line 25.000000 26.000001 0.0
1 Basic Monthly 26.000001 26.000002 0.0
2 Basic Monthly 26.000000 26.000000 0.0
3 Credit Card 26.000000 26.000000 0.0
4 Credit Card 26.000000 26.000000 0.0

Total_Monthly_Charges, Total_Long-Distance_Charges, Total_Revenue
0 0 0.000000 26.000001
1 0 26.000002 26.000001
2 0 26.000000 26.000000
3 0 26.000000 26.000000
4 0 26.000000 26.000000

Customer_Status, Error_Category, Charge, Revenue
0 Inured Others Others
1 Inured Others Others
2 Inured Others Others
3 Inured Others Others
4 Inured Others Others

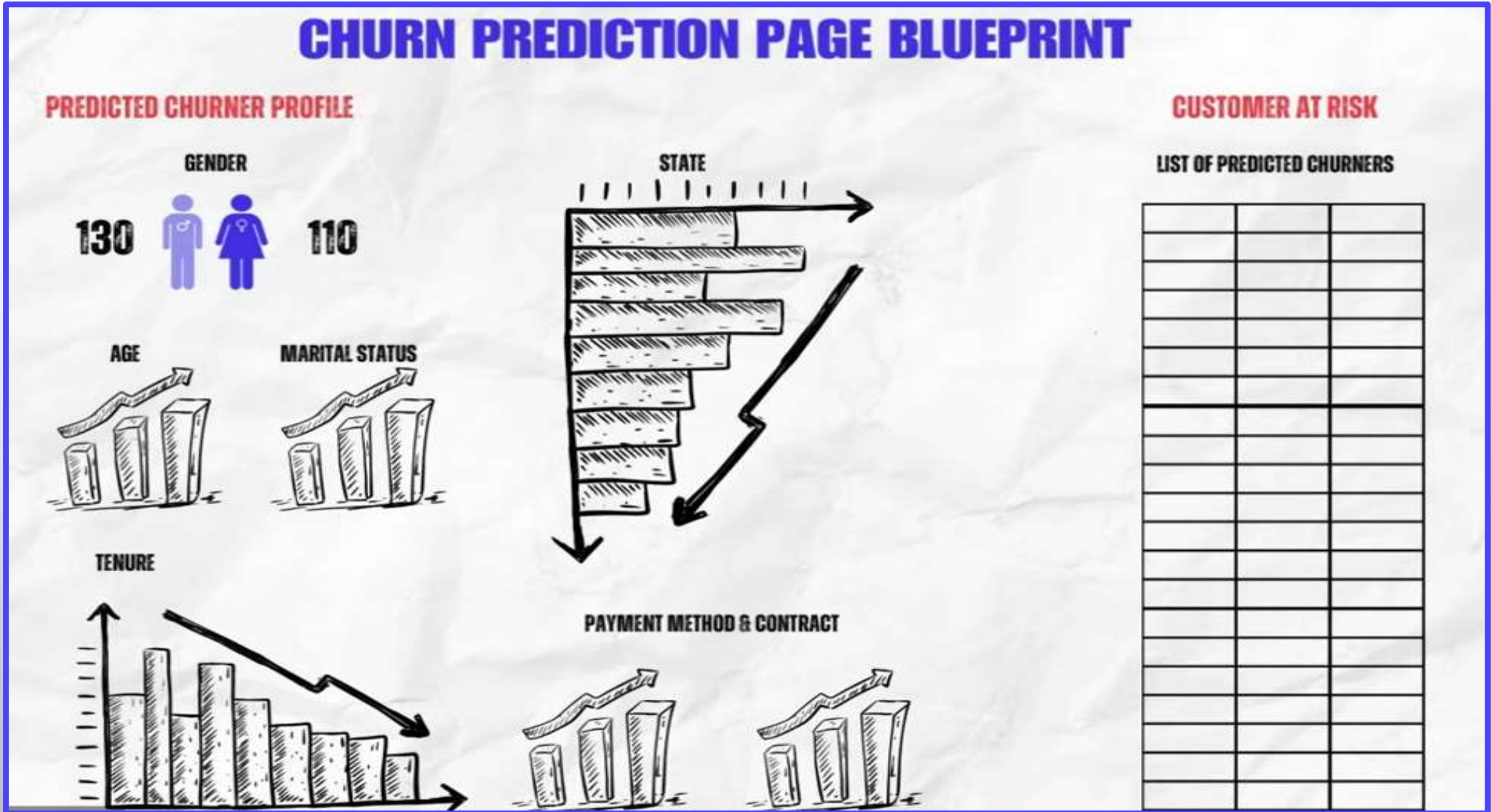
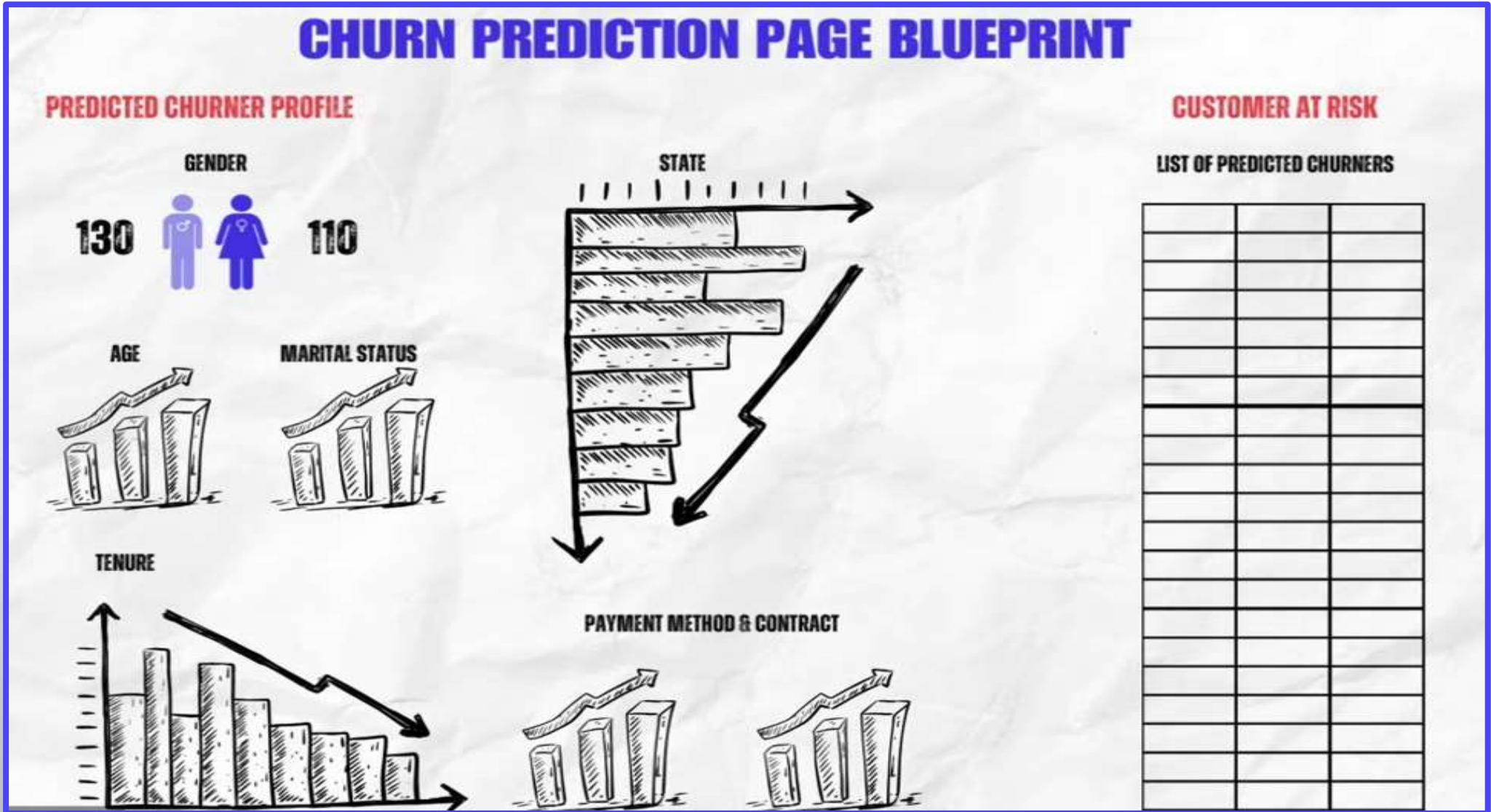
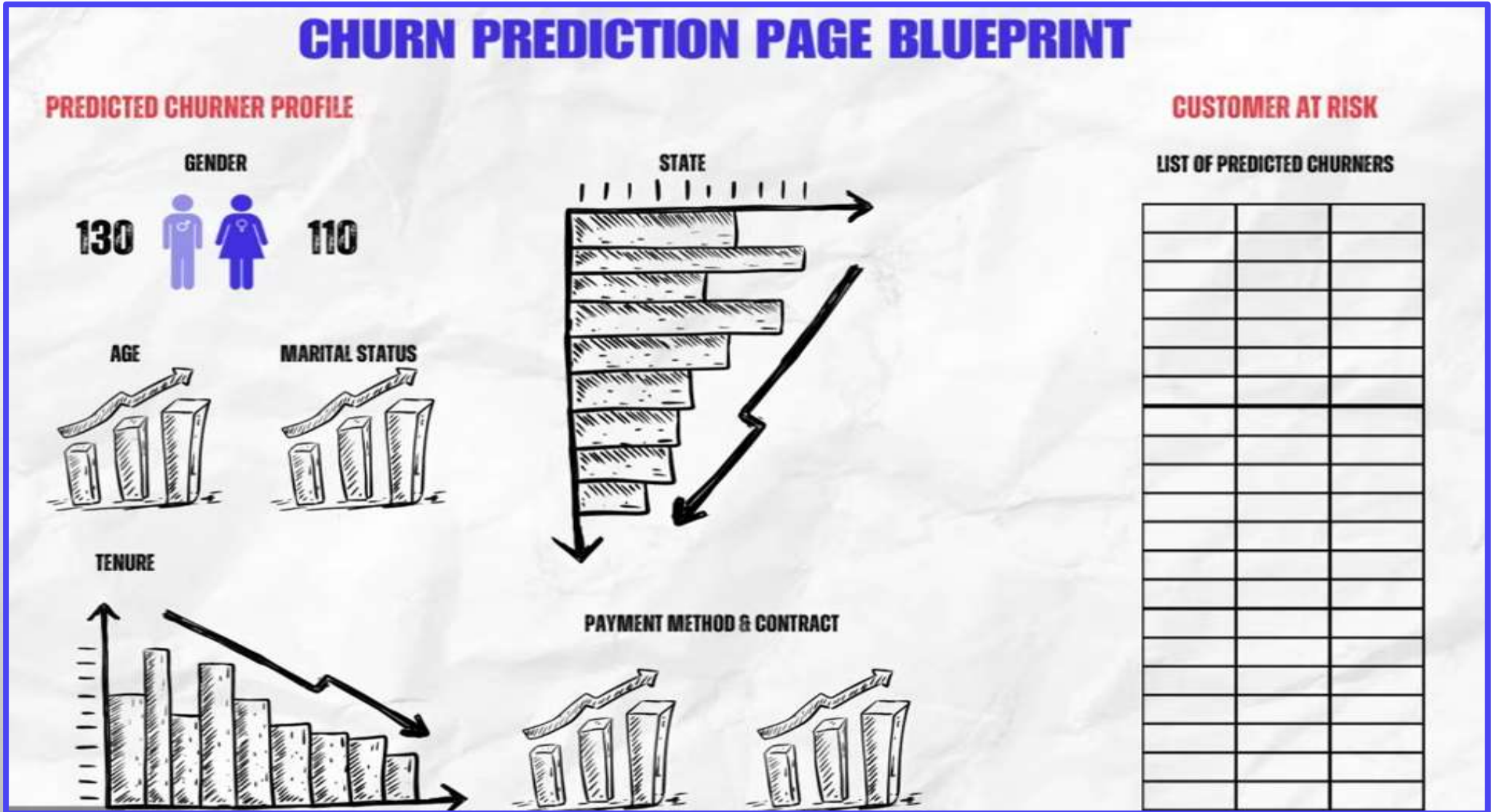
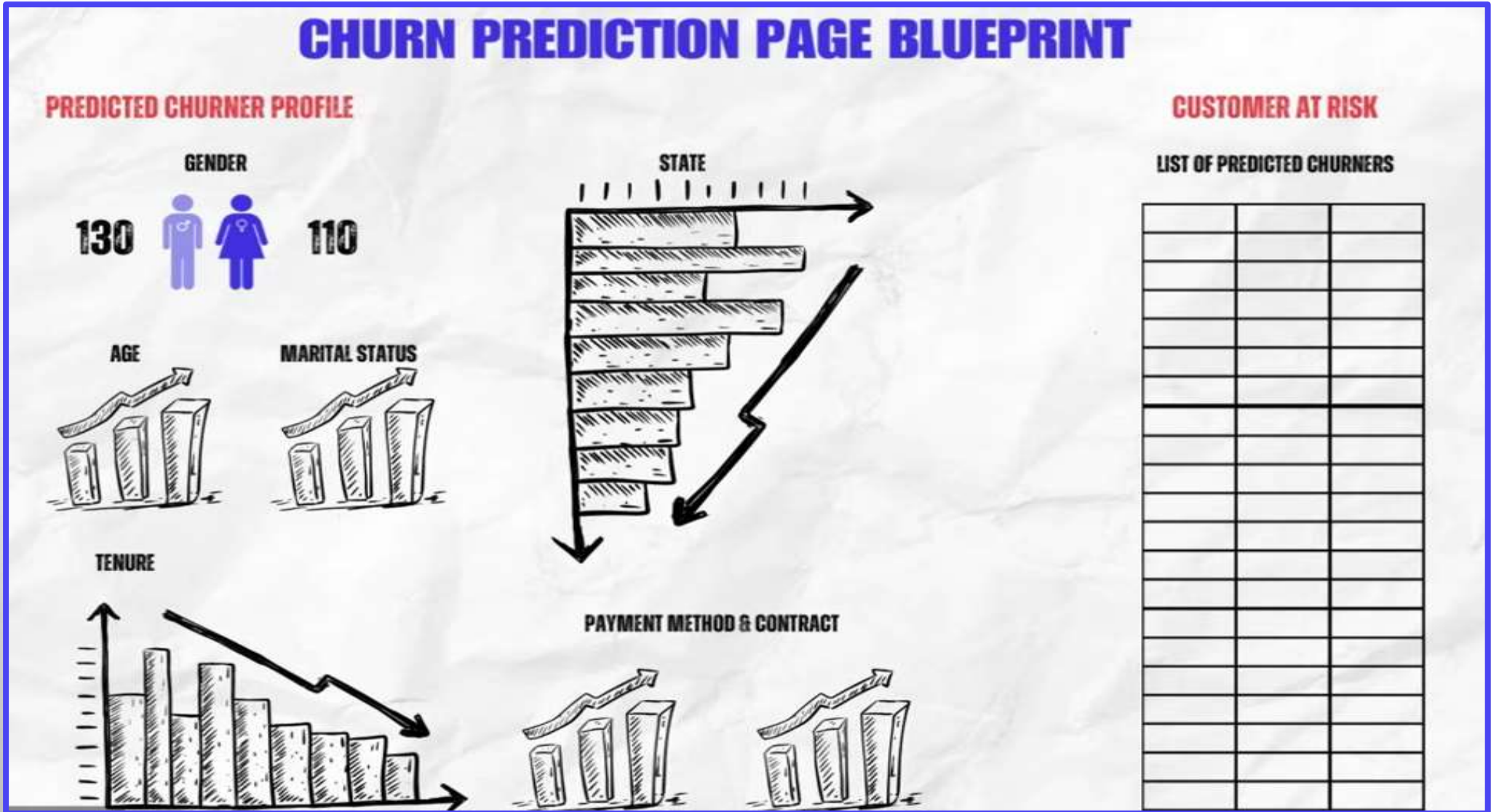
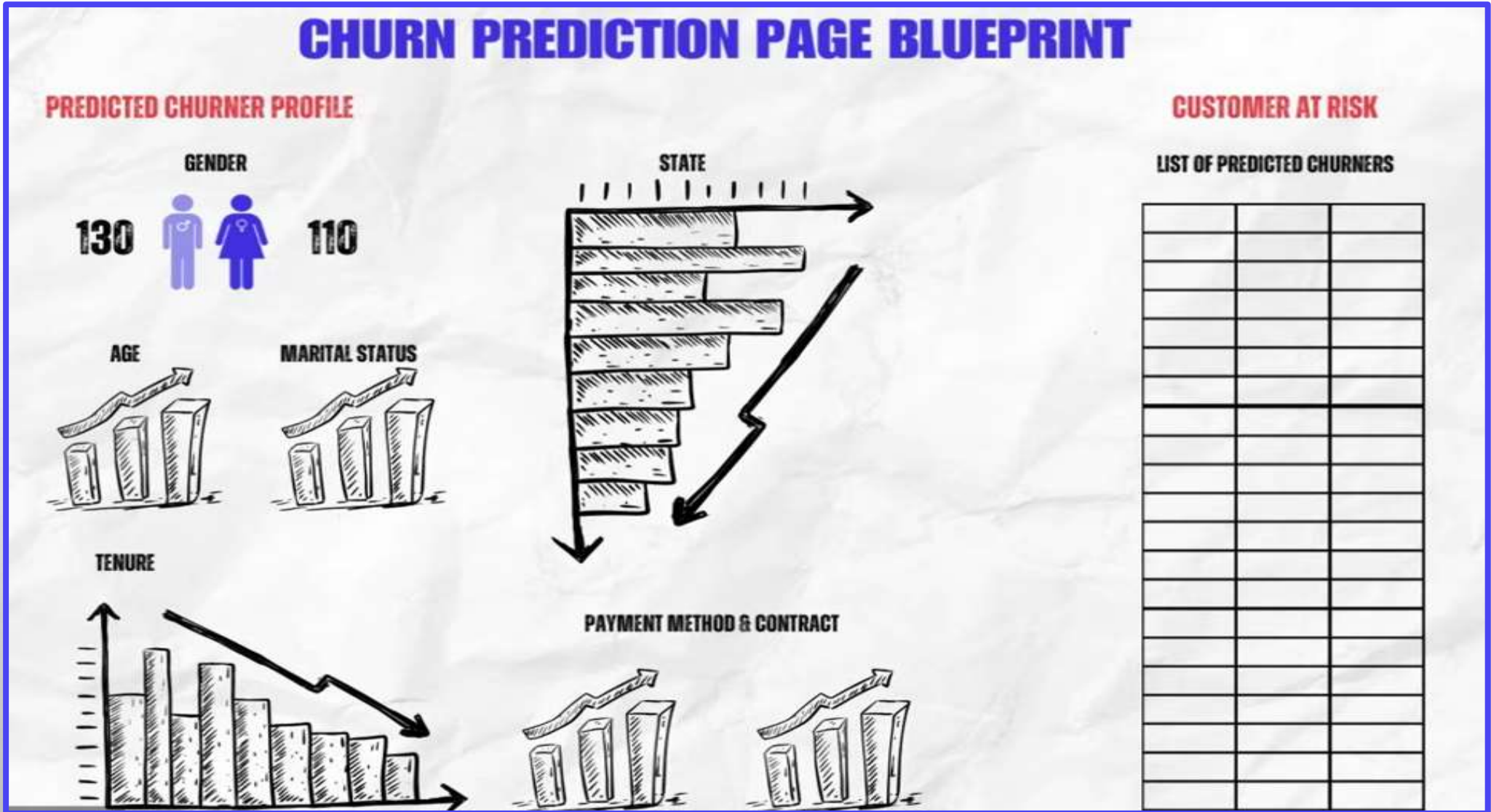
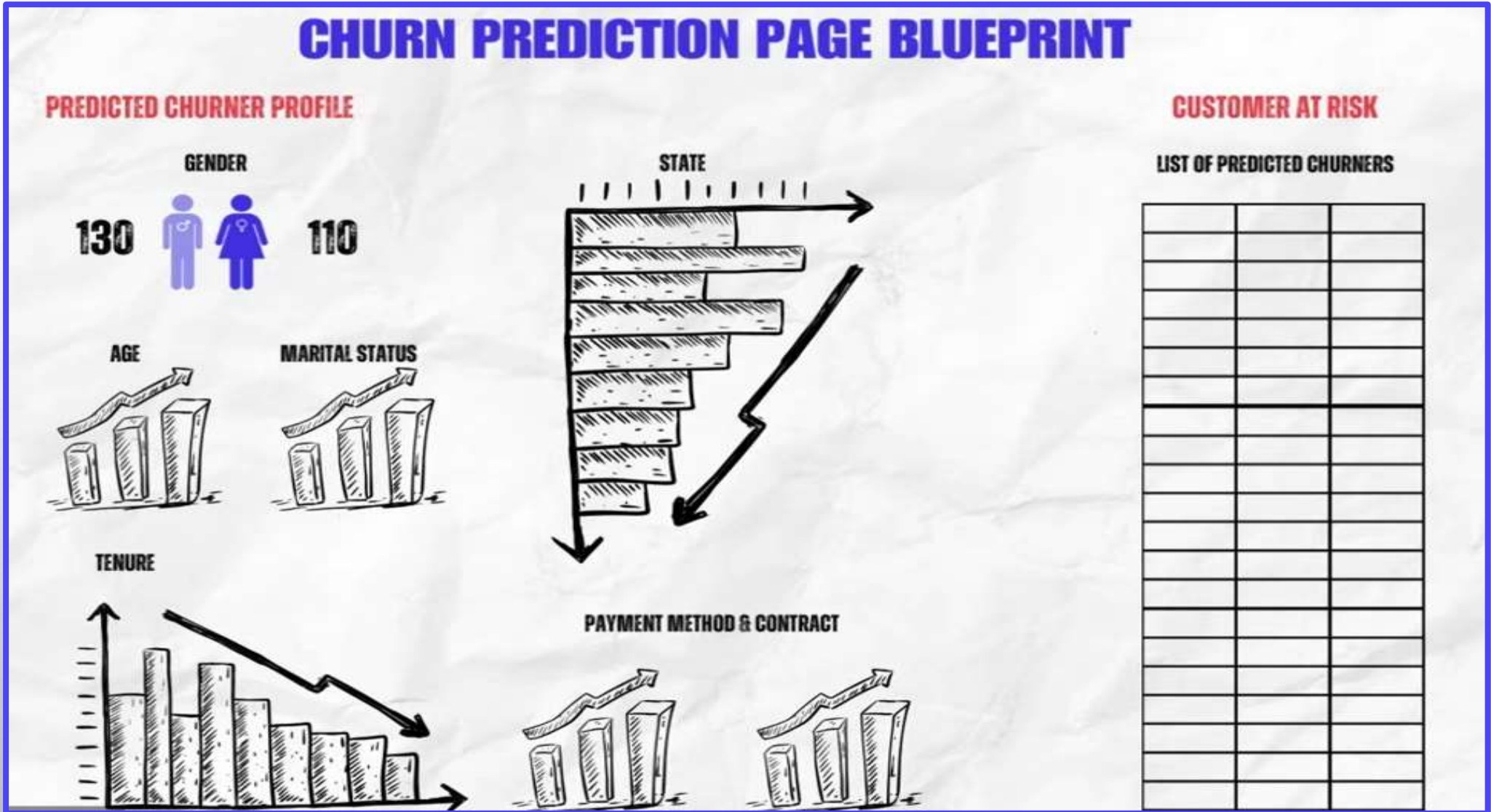
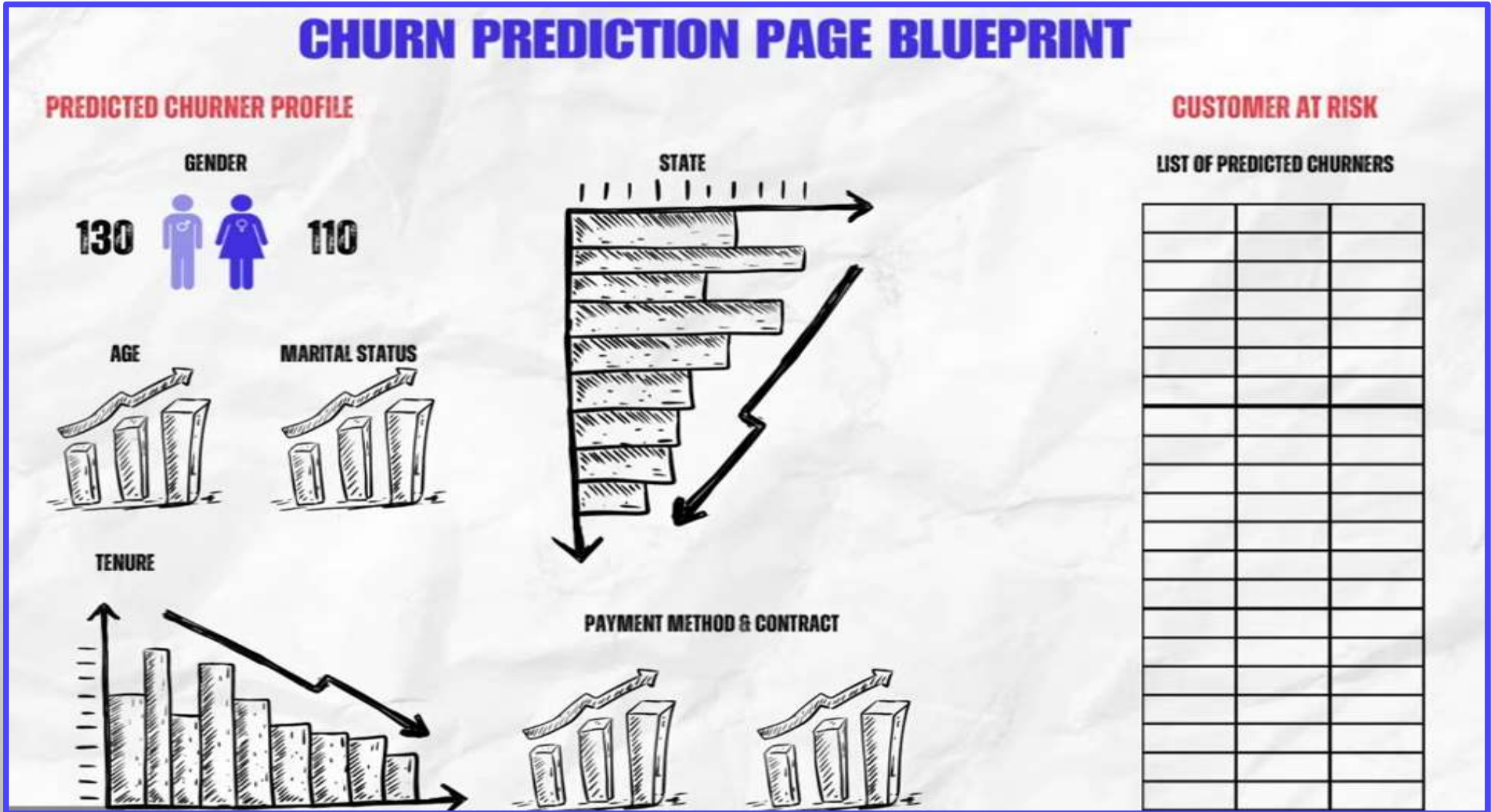
[V run a 42 columns]

```

Finally storing the new prediction model and saving the output with the csv file

The Predicted Column

Now, All the categorical data are there but at the end we have one more column that's says "Customer Status Predicted"

[illegible][illegible][illegible][illegible][illegible][illegible][illegible][illegible][illegible][illegible][illegible][illegible][illegible]

CHURN ANALYSIS - PREDICTION

Summary

PREDICTED CHURNER PROFILE

246

Female



132

Male

by Age Group



by Marital Status



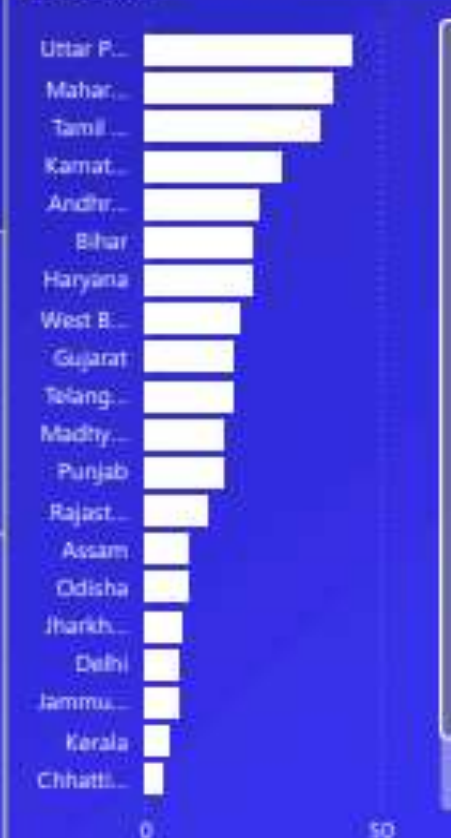
by Tenure Group



by Payment_Method



by State



by Contract



CUSTOMERS AT RISK

COUNT OF PREDICTED CHURNERS : 378

Customer_ID	Monthly_Charge	Number_of_Referrals	Total_Refunds	Total_Revenue
11751-TAM	24.30	5	0.00	38.49
12056-WES	90.40	2	0.00	362.88
12136-RAJ	19.90	2	0.00	31.78
12257-ASS	19.55	9	0.00	29.79
12340-DEL	62.80	0	0.00	104.95
12469-AND	55.30	11	0.00	91.98
12490-TEL	74.75	9	38.84	236.76
13058-MAD	46.10	13	0.00	138.13
13123-BIH	100.20	13	0.00	253.62
13666-UTT	95.40	15	0.00	344.16
13744-AND	19.65	8	0.00	33.50
13823-TEL	24.50	1	0.00	46.40
13946-HAR	19.65	1	0.00	43.32
14567-TAM	20.35	1	0.00	64.40
15349-UTT	50.15	9	0.00	90.03
15591-KAR	20.40	3	0.00	66.37
15803-UTT	19.15	6	0.00	41.52
16032-AND	46.60	8	0.00	91.64
16068-BIH	25.25	4	0.00	35.25
16244-UTT	19.55	5	0.00	24.68
16733-ODI	45.85	10	0.00	86.23
16764-WES	20.80	15	0.00	68.94
16931-BIH	19.30	8	0.00	27.96
17454-LAD	48.80	11	0.00	45.95

CHURN ANALYSIS - PREDICTION

Summary

PREDICTED CHURNER PROFILE

11

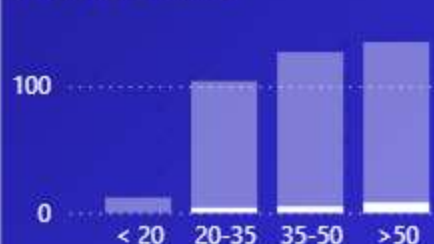
Female



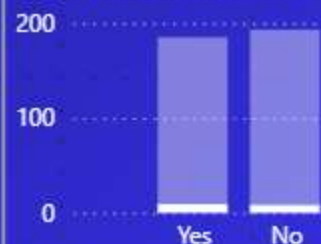
6

Male

by Age Group



by Marital Status



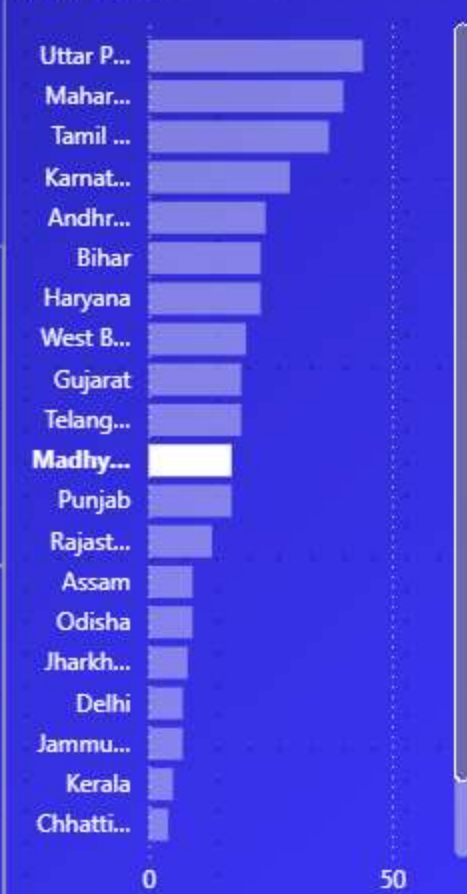
by Tenure Group



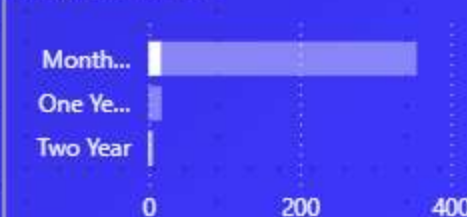
by Payment_Method



by State



by Contract



CUSTOMERS AT RISK

COUNT OF PREDICTED CHURNERS : 17

Customer_I D	Monthly_Char ge	Number_o f_Referrals	Total_Refunds	Total_Revenue
13058-MAD	46.10	13	0.00	138.13
27461-MAD	44.55	7	0.00	67.59
28378-MAD	19.45	12	0.00	66.04
33621-MAD	70.35	13	0.00	155.65
46792-MAD	35.90	6	0.00	35.90
51597-MAD	19.90	10	0.00	21.40
57457-MAD	20.00	14	0.00	106.76
58974-MAD	20.30	11	0.00	66.56
67994-MAD	46.30	12	0.00	57.81
72357-MAD	44.30	14	0.00	87.25
72469-MAD	55.70	9	0.00	73.38
74979-MAD	29.05	10	0.00	44.75
77311-MAD	20.25	9	0.00	60.71
81118-MAD	45.80	1	0.00	59.42

CHURN ANALYSIS - PREDICTION

Summary

PREDICTED CHURNER PROFILE

15

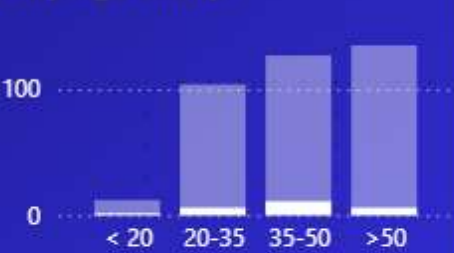
Female



9

Male

by Age Group



by Marital Status



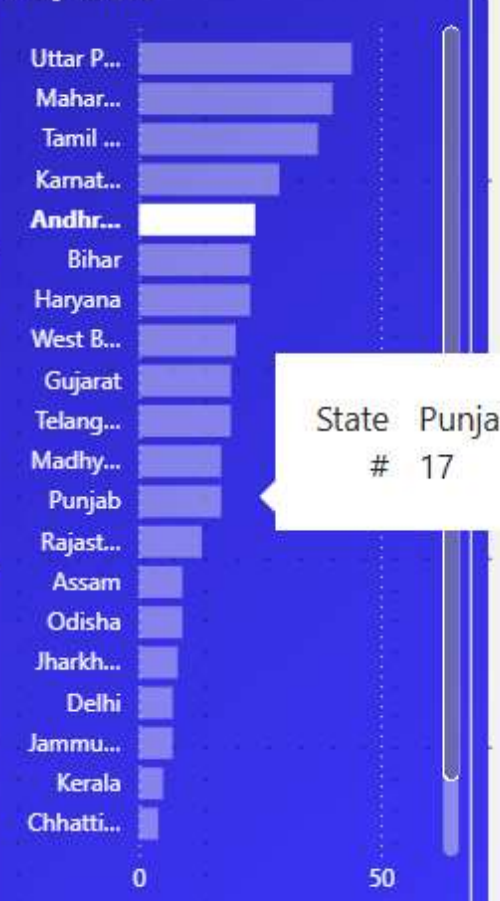
by Tenure Group



by Payment_Method

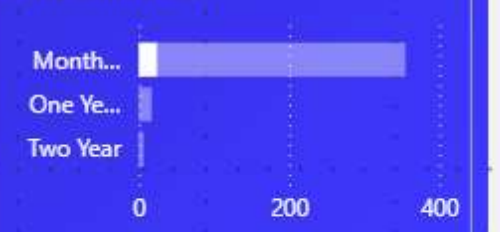


by State



State Punjab
17

by Contract



CUSTOMERS AT RISK

COUNT OF PREDICTED CHURNERS : 24

Customer_I D	Monthly_Char ge	Number_o f_Referrals	Total_Refunds	Total_Revenue
12469-AND	55.30	11	0.00	91.99
13744-AND	19.65	8	0.00	33.50
16032-AND	46.60	8	0.00	91.64
19041-AND	35.10	6	0.00	101.10
19998-AND	85.70	0	0.00	346.27
24754-AND	69.55	0	33.80	171.41
31129-AND	70.35	15	0.00	94.77
34024-AND	20.05	4	0.00	27.71
38748-AND	19.65	13	0.00	47.53
44208-AND	69.15	2	0.00	307.84
45213-AND	61.20	15	0.00	190.07
47492-AND	20.15	15	0.00	38.78
54564-AND	76.10	11	0.00	300.65
59750-AND	74.90	11	0.00	218.21
62359-AND	33.60	5	0.00	83.60
62535-AND	70.70	8	0.00	300.77
63464-AND	60.65	7	0.00	264.37
64327-AND	20.65	8	0.00	55.89
77658-AND	24.40	0	0.00	24.40
82473-AND	29.80	10	0.00	94.40
83939-AND	20.10	8	0.00	108.87
89303-AND	69.95	6	0.00	202.10
92398-AND	20.00	5	0.00	107.81
95822-AND	19.20	3	0.00	37.09

THANK YOU

