CUSTOMER CHURN PREDICTION

MISHTY VERMA

INTRODUCTION AND OVERVIEW

A telecommunications company wanted to reduce customer churn. By analyzing customer behavior data using SQL and Excel, and applying statistical models, they identified key churn predictors. Power BI was used to create a dashboard that monitored churn risk in real-time. As a result, the company implemented targeted retention strategies, reducing churn by 12% within six months.

Customer churn is the phenomenon where customer stops doing business with a company or switches companies.

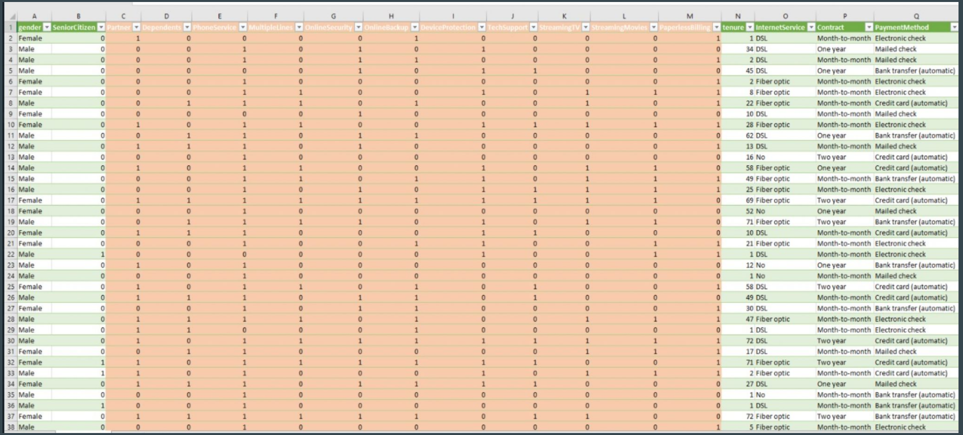
Here, the data was taken from kaggle telecom churn prediction data set.

The columns in the dataset are as follows:

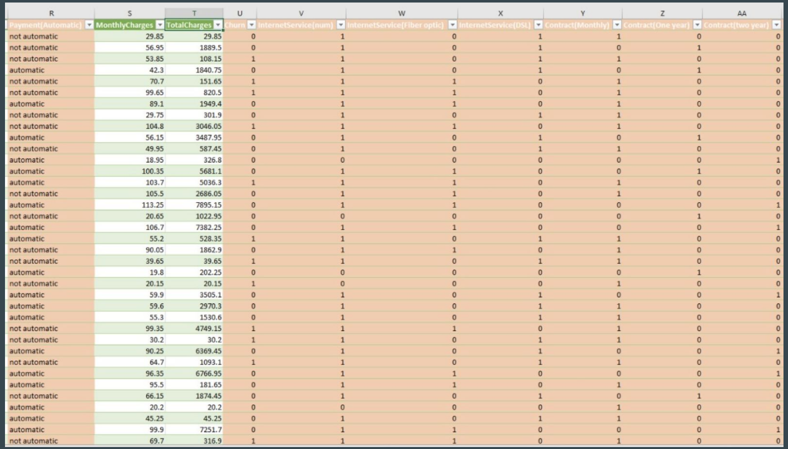
- customerID: unique value
- Gender: Male or Female
- SeniorCitizen: yes or no
- Partner: yes o no
- Dependents: yes or no
- Tenure: in months
- PhoneService: Yes or no
- MultipleLines: No phone service , yes or no
- InternetService: DSL FiberOptics or no service
- OnlineSecurity: Yes or no or no internet service
- OnlineBackup: Yes or no No internet
- DeviceProtection: Yes or no or no internets service
- TechSupport: Yes or no or no internet service
- StreamingTV: Yes or no or no internet service
- StreamingMovies: Yes or no or no internet service
- Contract: month-to-month, one year, two year
- PaperlessBilling: Yes or No
- PaymentMethod: electronic, mailcheck, bank transfer, credit card
- MonthlyCharges: Dollar
- TotalCharges: Dollar
- Churn: yes or no

EXCEL

First, the data was processed on Excel. Here, relevant groupings were done and categorical data was converted to numerical data for the purpose of applying machine learning models.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	gender	SeniorCitizen	Partner	Dependents	PhoneService	MultipleLines	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	PaperlessBilling	tenure	InternetService	Contract	PaymentMethod
2	Female	0	1	0	0	0	0	1	0	0	0	0	1	1	DSL	Month-to-month	Electronic check
3	Male	0	0	0	1	0	1	0	1	0	0	0	0	34	DSL	One year	Mailed check
4	Male	0	0	0	1	0	1	1	0	0	0	0	1	2	DSL	Month-to-month	Mailed check
5	Male	0	0	0	0	0	1	0	1	1	0	0	0	45	DSL	One year	Bank transfer (automatic)
6	Female	0	0	0	1	0	0	0	0	0	0	0	1	2	Fiber optic	Month-to-month	Electronic check
7	Female	0	0	0	1	1	0	0	1	0	1	1	1	8	Fiber optic	Month-to-month	Electronic check
8	Male	0	0	1	1	1	0	1	0	0	1	0	1	22	Fiber optic	Month-to-month	Credit card (automatic)
9	Female	0	0	0	0	0	1	0	0	0	0	0	0	10	DSL	Month-to-month	Mailed check
10	Female	0	1	0	1	1	0	0	1	1	1	1	1	28	Fiber optic	Month-to-month	Electronic check
11	Male	0	0	1	1	0	1	1	0	0	0	0	0	62	DSL	One year	Bank transfer (automatic)
12	Male	0	1	1	1	0	1	0	0	0	0	0	1	13	DSL	Month-to-month	Mailed check
13	Male	0	0	0	1	0	0	0	0	0	0	0	0	16	No	Two year	Credit card (automatic)
14	Male	0	1	0	1	1	0	0	1	0	1	1	0	58	Fiber optic	One year	Credit card (automatic)
15	Male	0	0	0	1	1	0	1	1	0	1	1	1	49	Fiber optic	Month-to-month	Bank transfer (automatic)
16	Male	0	0	0	1	0	1	0	1	1	1	1	1	25	Fiber optic	Month-to-month	Electronic check
17	Female	0	1	1	1	1	1	1	1	1	1	1	0	69	Fiber optic	Two year	Credit card (automatic)
18	Female	0	0	0	1	0	0	0	0	0	0	0	0	52	No	One year	Mailed check
19	Male	0	0	1	1	1	1	0	1	0	1	1	0	71	Fiber optic	Two year	Bank transfer (automatic)
20	Female	0	1	1	1	0	0	0	1	1	0	0	0	10	DSL	Month-to-month	Credit card (automatic)
21	Female	0	0	0	1	0	0	1	1	0	0	1	1	21	Fiber optic	Month-to-month	Electronic check
22	Male	1	0	0	0	0	0	0	1	0	0	1	1	1	DSL	Month-to-month	Electronic check
23	Male	0	1	0	1	0	0	0	0	0	0	0	0	12	No	One year	Bank transfer (automatic)
24	Male	0	0	0	1	0	0	0	0	0	0	0	0	1	No	Month-to-month	Mailed check
25	Female	0	1	0	1	1	0	1	0	1	0	0	1	58	DSL	Two year	Credit card (automatic)
26	Male	0	1	1	1	0	1	1	0	1	0	0	0	49	DSL	Month-to-month	Credit card (automatic)
27	Female	0	0	0	1	0	1	1	0	0	0	0	1	30	DSL	Month-to-month	Bank transfer (automatic)
28	Male	0	1	1	1	1	0	1	0	0	1	1	1	47	Fiber optic	Month-to-month	Electronic check
29	Male	0	1	1	0	0	0	1	0	0	0	0	0	1	DSL	Month-to-month	Electronic check
30	Male	0	1	0	1	1	1	1	1	1	1	1	1	72	DSL	Two year	Credit card (automatic)
31	Female	0	0	1	1	0	0	0	0	0	1	1	1	17	DSL	Month-to-month	Mailed check
32	Female	1	1	0	1	1	1	1	1	1	0	0	1	71	Fiber optic	Two year	Credit card (automatic)
33	Male	1	1	0	1	0	0	0	1	0	1	1	1	2	Fiber optic	Month-to-month	Credit card (automatic)
34	Female	0	1	1	1	0	1	1	1	1	0	0	0	27	DSL	One year	Mailed check
35	Male	0	0	0	1	0	0	0	0	0	0	0	0	1	No	Month-to-month	Bank transfer (automatic)
36	Male	1	0	0	1	0	0	0	0	0	0	0	0	1	DSL	Month-to-month	Bank transfer (automatic)
37	Female	0	1	1	1	1	1	1	0	1	1	0	0	72	Fiber optic	Two year	Bank transfer (automatic)
38	Male	0	0	0	1	0	0	0	0	0	0	0	1	5	Fiber optic	Month-to-month	Electronic check



R	S	T	U	V	W	X	Y	Z	AA
Payment(Automatic)	MonthlyCharges	TotalCharges	Churn	InternetService(num)	InternetService(Fiber optic)	InternetService(DSL)	Contract(Monthly)	Contract(One year)	Contract(two year)
not automatic	29.85	29.85	0	1	0	1	1	0	0
not automatic	56.95	1889.5	0	1	0	1	0	1	0
not automatic	53.85	108.15	1	1	0	1	1	0	0
automatic	42.3	1840.75	0	1	0	1	0	1	0
not automatic	70.7	151.65	1	1	1	0	1	0	0
not automatic	99.65	820.5	1	1	1	0	1	0	0
automatic	89.1	1949.4	0	1	1	0	1	0	0
not automatic	29.75	301.9	0	1	0	1	1	0	0
not automatic	104.8	3046.05	1	1	1	0	1	0	0
automatic	56.15	3487.95	0	1	0	1	0	1	0
not automatic	49.95	587.45	0	1	0	1	1	0	0
automatic	18.95	326.8	0	0	0	0	0	0	1
automatic	100.35	5681.1	0	1	1	0	0	1	0
automatic	103.7	5036.3	1	1	1	0	1	0	0
not automatic	105.5	2686.05	0	1	1	0	1	0	0
automatic	113.25	7895.15	0	1	1	0	0	0	1
not automatic	20.65	1022.95	0	0	0	0	0	1	0
automatic	106.7	7382.25	0	1	1	0	0	0	1
automatic	55.2	528.35	1	1	0	1	1	0	0
not automatic	90.05	1862.9	0	1	1	0	1	0	0
not automatic	39.65	39.65	1	1	0	1	1	0	0
automatic	19.8	202.25	0	0	0	0	0	1	0
not automatic	20.15	20.15	1	0	0	0	1	0	0
automatic	59.9	3505.1	0	1	0	1	0	0	1
automatic	59.6	2970.3	0	1	0	1	1	0	0
automatic	55.3	1530.6	0	1	0	1	1	0	0
not automatic	99.35	4749.15	1	1	1	0	1	0	0
not automatic	30.2	30.2	1	1	0	1	1	0	0
automatic	90.25	6369.45	0	1	0	1	0	0	1
not automatic	64.7	1093.1	1	1	0	1	1	0	0
automatic	96.35	6766.95	0	1	1	0	0	0	1
automatic	95.5	181.65	0	1	1	0	1	0	0
not automatic	66.15	1874.45	0	1	0	1	0	1	0
automatic	20.2	20.2	0	0	0	0	1	0	0
automatic	45.25	45.25	0	1	0	1	1	0	0
automatic	99.9	7251.7	0	1	1	0	0	0	1
not automatic	69.7	316.9	1	1	1	0	1	0	0

Irrelevant columns were also deleted and files were saved in two formats, csv(for machine learning model, and Excel file for power BI analysis.

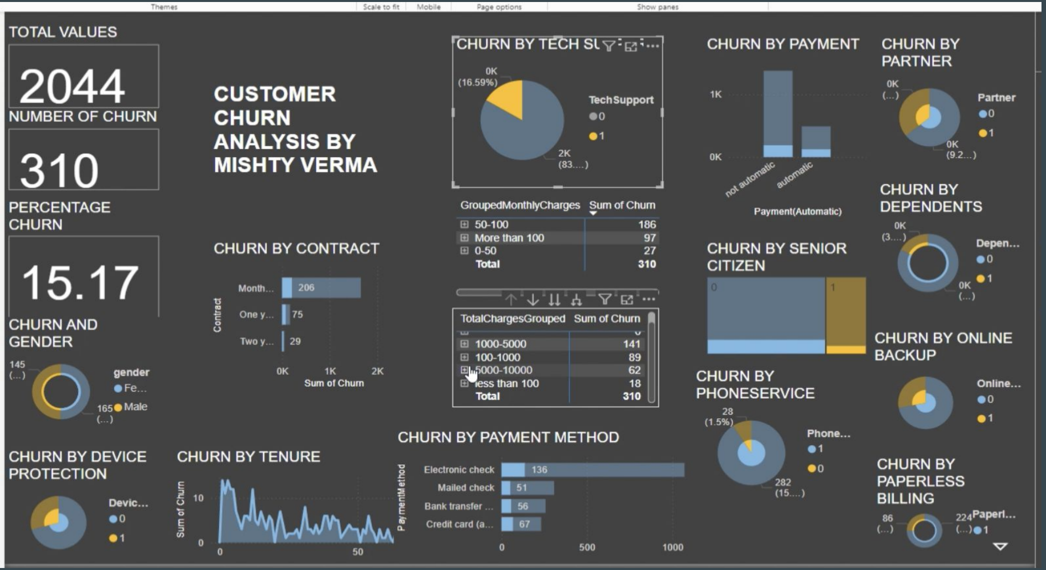
MonthlyCharges	TotalCharges	Churn	InternetService(num)	InternetService(Fiber optic)	InternetService(DSL)	Contract(Monthly)	Contract(One year)	Contract(two year)	Payment(Automatic_num)	Gender(num)	GroupedMonthlyCharges	TotalChargesGrouped
29.85	29.85	0	1	0	1	1	0	0	0	0	0-50	less than 100
56.95	1889.5	0	1	0	1	0	1	0	0	1	50-100	1000-5000
53.85	108.15	1	1	0	1	1	0	0	0	1	50-100	100-1000
42.3	1840.75	0	1	0	1	0	1	0	1	1	0-50	1000-5000
70.7	151.65	1	1	1	0	1	0	0	0	0	50-100	100-1000
99.65	820.5	1	1	1	0	1	0	0	0	0	50-100	100-1000
89.1	1949.4	0	1	1	0	1	0	0	1	1	50-100	1000-5000
29.75	301.9	0	1	0	1	1	0	0	0	0	0-50	100-1000
104.8	3046.05	1	1	1	0	1	0	0	0	0	More than 100	1000-5000
56.15	3487.95	0	1	0	1	0	1	0	1	1	50-100	1000-5000
49.95	587.45	0	1	0	1	1	0	0	0	1	0-50	100-1000
18.95	326.8	0	0	0	0	0	0	1	1	1	0-50	100-1000
100.35	5681.1	0	1	1	0	0	1	0	1	1	More than 100	5000-10000
103.7	5036.3	1	1	1	0	1	0	0	1	1	More than 100	5000-10000
105.5	2686.05	0	1	1	0	1	0	0	0	1	More than 100	1000-5000
113.25	7895.15	0	1	1	0	0	0	1	1	0	More than 100	5000-10000
20.65	1022.95	0	0	0	0	0	1	0	0	0	0-50	1000-5000
106.7	7382.25	0	1	1	0	0	0	1	1	1	More than 100	5000-10000
55.2	528.35	1	1	0	1	1	0	0	1	0	50-100	100-1000
90.05	1862.9	0	1	1	0	1	0	0	0	0	50-100	1000-5000
39.65	39.65	1	1	0	1	1	0	0	0	1	0-50	less than 100
19.8	202.25	0	0	0	0	0	1	0	1	1	0-50	100-1000
20.15	20.15	1	0	0	0	1	0	0	0	1	0-50	less than 100
59.9	3505.1	0	1	0	1	0	0	1	1	0	50-100	1000-5000
59.6	2970.3	0	1	0	1	1	0	0	1	1	50-100	1000-5000
55.3	1530.6	0	1	0	1	1	0	0	1	0	50-100	1000-5000
99.35	4749.15	1	1	1	0	1	0	0	0	1	50-100	1000-5000
30.2	30.2	1	1	0	1	1	0	0	0	1	0-50	less than 100
90.25	6369.45	0	1	0	1	0	0	1	1	1	50-100	5000-10000
64.7	1093.1	1	1	0	1	1	0	0	0	0	50-100	1000-5000
96.35	6766.95	0	1	1	0	0	0	1	1	0	50-100	5000-10000
95.5	181.65	0	1	1	0	1	0	0	1	1	50-100	100-1000
66.15	1874.45	0	1	0	1	0	1	0	0	0	50-100	1000-5000
20.2	20.2	0	0	0	0	1	0	0	1	1	0-50	less than 100
45.25	45.25	0	1	0	1	1	0	0	1	1	0-50	less than 100
99.9	7251.7	0	1	1	0	0	0	1	1	0	50-100	5000-10000
69.7	316.9	1	1	1	0	1	0	0	0	1	50-100	100-1000

CustChurnCsv Sheet1

POWER BI

A dynamic dashboard was created to find the key performance indicators. Relevant conclusions were made from the dashboard.

https://drive.google.com/file/d/1UJQLn-gL2cFJkEnfQneTjFPHeiLNGWW8/view?usp=sharing



Conclusions made from dashboard

Out of 7043 customers, 1869 or 26.5% customers churned in the past month.

Gender did not significantly affect the churn rate.

For no device protection, tech support, and online backup, churn was higher than that for customers with these services.

Customers with automatic mode of payment churned less than those without it.

People without dependents or partners churned more than those with dependents and partners.

People with paperless billing churned more.

Lower the tenure, higher was the churn and similarly people with month to month contract churned more and people with 2 year contract churned less.

Customers with monthly charges of 50-100 dollars churned more.

Logistic Regression Model

Logistic regression uses sigmoid function to map values to different classes. Unlike linear regression, which gives a continuous outcome, logistic regression predicts classes of values. The equation for logistic regression is

$$\text{logit}(p) = \ln(p1\text{-}p) = {}_0 + {}_1x_1 + \ldots\ldots + {}_nx_n$$

Where:

pis the probability of the outcome of interest(probability of churn), x_1,x_2,...x_n are the independent variables, ${}_0$ is the intercept,${}_1$,...${}_n$ are coefficients of the independent variables.

The logistic function (or sigmoid function) is used to transform the logit(linear combination of inputs) into a probability:

p = 11+e-(0 + 1x1 +........+nxn)

Logistic regression Model

Logistic regression was used to classify the value and predict churn(1) or no churn (0)

https://drive.google.com/file/d/1nUU8s4b2aGJFwjEmbjPK_wsXabN2vqbu/view?usp=drive_link

https://drive.google.com/file/d/1T4TWPrVMT7zMwBRoRxfNylwR8ZWDXPv-/view?usp=drive_link

Here are the links to my logistic regression models

Report:

Here the following steps can be implemented by the company to reduce customer churn:

Every customer should take device protection, tech support and online backup.

Customers should take up automatic payment mode.

Customers should take the service for 2 years or 1 year, customers with month to month service churned more.

If the price deal on the monthly plan is more than the overall deal on 2 year or 1 year plan, then the customer might prefer long term plan over short term plans.

The company van bring down the churn rate by implementing few of these steps.