

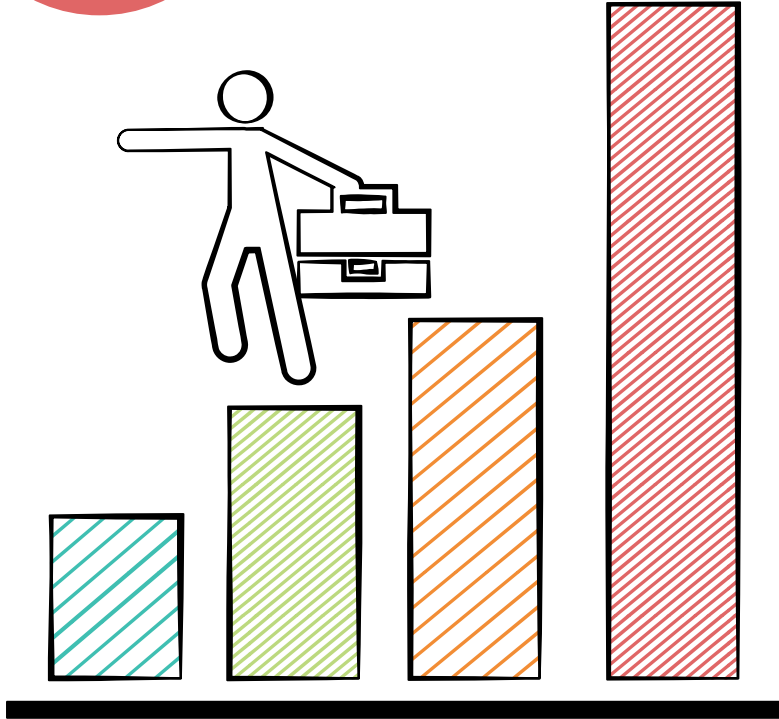


# Customer Churn Prediction

Team 18

Project Guide -  
Prof.Chiranjeevi

# 1 Introduction



Customer churn occurs when customers or subscribers stop doing business with a company or service, also known as customer attrition

One industry in which churn rates are particularly useful is the telecommunications industry, because most customers have multiple options from which to choose within a geographic location.

# 2 Agenda

## 1.Data Collection

Telco-Customer-Churn  
Dataset from Kaggle

## 5.Prediction using ML

Logistic Regression and  
Decision tree are used for  
prediction

## 4.Data visualization

Data visualization is done  
for model evaluation using  
highcharter.



## 2.Data Pre-processing

Data cleaning has done for  
inconsistency in data and so on

## 3.EDA

Exploratory data analysis -  
correlation between numeric  
columns

# Problem Statement



Predicting if a customer is likely to churn from the company or not using Logistic Regression and Decision Tree

# 3 Data Analysis

The data is loaded into RStudio:

```
churn <- read.csv("D:/Telco-Customer-Churn.csv")  
dim(churn)
```

```
View(churn)
```

Rows  
7043

Columns  
21

churn x												
Filter												
	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DevicePro
1	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	Yes	No
2	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes
3	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No
4	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	No	Yes
5	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No
6	9305-CDSKC	Female	0	No	No	8	Yes	Yes	Fiber optic	No	No	Yes
7	1452-KIOVK	Male	0	No	Yes	22	Yes	Yes	Fiber optic	No	Yes	No
8	6713-OKOMC	Female	0	No	No	10	No	No phone service	DSL	Yes	No	No
9	7892-POOKP	Female	0	Yes	No	28	Yes	Yes	Fiber optic	No	No	Yes
10	6388-TABGU	Male	0	No	Yes	62	Yes	No	DSL	Yes	Yes	No
11	9763-GRSKD	Male	0	Yes	Yes	13	Yes	No	DSL	Yes	No	No
12	7469-LKBCI	Male	0	No	No	16	Yes	No	No	No internet service	No internet service	No interr
13	8091-TTVAX	Male	0	Yes	No	58	Yes	Yes	Fiber optic	No	No	Yes
14	0280-XJGEX	Male	0	No	No	49	Yes	Yes	Fiber optic	No	Yes	Yes
15	5129-JLPIS	Male	0	No	No	25	Yes	No	Fiber optic	Yes	No	Yes
16	3655-SNQYZ	Female	0	Yes	Yes	69	Yes	Yes	Fiber optic	Yes	Yes	Yes
17	8191-XWSZG	Female	0	No	No	52	Yes	No	No	No internet service	No internet service	No internet service
18	9959-WOFKT	Male	0	No	Yes	71	Yes	Yes	Fiber optic	Yes	No	No
19	4190-MFLUW	Female	0	Yes	Yes	10	Yes	No	DSL	No	No	No

```
glimpse(churn)
```

```
> glimpse(churn)
```

Rows: 7,043

Columns: 21

\$ customerID	<chr>	"7590-VHVEG", "5575-GNVDE", "3668-QPYBK", "7795-CFOCW", "9237-HQITU", "9305-CDSKC", "1452-KIOVK", ...
\$ gender	<chr>	"Female", "Male", "Male", "Male", "Female", "Female", "Male", "Female", "Female", "Male", "Male", ...
\$ SeniorCitizen	<int>	0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0...
\$ Partner	<chr>	"Yes", "No", "No", "No", "No", "No", "No", "No", "Yes", "No", "Yes", "No", "Yes", "No", "No", "Ye...
\$ Dependents	<chr>	"No", "No", "No", "No", "No", "No", "Yes", "No", "No", "Yes", "Yes", "No", "No", "No", "No", "Yes...
\$ tenure	<int>	1, 34, 2, 45, 2, 8, 22, 10, 28, 62, 13, 16, 58, 49, 25, 69, 52, 71, 10, 21, 1, 12, 1, 58, 49, 30, ...
\$ PhoneService	<chr>	"No", "Yes", "Yes", "No", "Yes", "Yes", "Yes", "No", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Y...
\$ MultipleLines	<chr>	"No phone service", "No", "No", "No phone service", "No", "Yes", "Yes", "No phone service", "Yes"...
\$ InternetService	<chr>	"DSL", "DSL", "DSL", "DSL", "Fiber optic", "Fiber optic", "Fiber optic", "DSL", "Fiber optic", "D...
\$ OnlineSecurity	<chr>	"No", "Yes", "Yes", "Yes", "No", "No", "No", "Yes", "No", "Yes", "Yes", "No internet service", "N...
\$ OnlineBackup	<chr>	"Yes", "No", "Yes", "No", "No", "No", "Yes", "No", "No", "Yes", "No", "No internet service", "No"...
\$ DeviceProtection	<chr>	"No", "Yes", "No", "Yes", "No", "Yes", "No", "No", "Yes", "No", "No", "No internet service", "Yes...
\$ TechSupport	<chr>	"No", "No", "No", "Yes", "No", "No", "No", "No", "Yes", "No", "No", "No internet service", "No", ...
\$ StreamingTV	<chr>	"No", "No", "No", "No", "No", "Yes", "Yes", "No", "Yes", "No", "No", "No internet service", "Yes"...
\$ StreamingMovies	<chr>	"No", "No", "No", "No", "No", "Yes", "No", "No", "Yes", "No", "No", "No internet service", "Yes", ...
\$ Contract	<chr>	"Month-to-month", "One year", "Month-to-month", "One year", "Month-to-month", "Month-to-month", "...
\$ PaperlessBilling	<chr>	"Yes", "No", "Yes", "No", "Yes", "Yes", "Yes", "No", "Yes", "No", "Yes", "No", "No", "Yes", "Yes", "Yes"...
\$ PaymentMethod	<chr>	"Electronic check", "Mailed check", "Mailed check", "Bank transfer (automatic)", "Electronic chec...
\$ MonthlyCharges	<dbl>	29.85, 56.95, 53.85, 42.30, 70.70, 99.65, 89.10, 29.75, 104.80, 56.15, 49.95, 18.95, 100.35, 103...
\$ TotalCharges	<dbl>	29.85, 1889.50, 108.15, 1840.75, 151.65, 820.50, 1949.40, 301.90, 3046.05, 3487.95, 587.45, 326.8...
\$ Churn	<chr>	"No", "No", "Yes", "No", "Yes", "Yes", "No", "No", "Yes", "No", "No", "No", "No", "Yes", "No", "N...

```
library(skimr)
churn %>% skim()
```

# Statistics

1

```
— variable type: character —
skim_variable  n_missing complete_rate min max empty n_unique whitespace
1 customerID      0           1 10 10      0         7043           0
2 gender          0           1  4  6      0          2           0
3 Partner         0           1  2  3      0          2           0
4 Dependents      0           1  2  3      0          2           0
5 PhoneService    0           1  2  3      0          2           0
6 MultipleLines   0           1  2 16      0          3           0
7 InternetService 0           1  2 11      0          3           0
8 OnlineSecurity  0           1  2 19      0          3           0
9 OnlineBackup    0           1  2 19      0          3           0
10 DeviceProtection 0           1  2 19      0          3           0
11 TechSupport     0           1  2 19      0          3           0
12 StreamingTV     0           1  2 19      0          3           0
13 StreamingMovies 0           1  2 19      0          3           0
14 Contract        0           1  8 14      0          3           0
15 PaperlessBilling 0           1  2  3      0          2           0
16 PaymentMethod   0           1 12 25      0          4           0
17 churn           0           1  2  3      0          2           0
```

```
— variable type: numeric —
skim_variable  n_missing complete_rate mean sd p0 p25 p50 p75 p100 hist
1 SeniorCitizen 0           1      0.162 0.369 0  0  0  0  1  [bar chart]
2 tenure        0           1      32.4 24.6 0  9 29 55 72  [bar chart]
3 MonthlyCharges 0           1      64.8 30.1 18.2 35.5 70.4 89.8 119. [bar chart]
4 TotalCharges  11          0.998 2283. 2267. 18.8 401. 1397. 3795. 8685. [bar chart]
> |
```



2

# Data Preprocessing

## Handling Null Values:

```
as.data.frame(colSums(is.na(churn)))  
churn <- churn %>% drop_na()
```

customerID	0
gender	0
SeniorCitizen	0
Partner	0
Dependents	0
tenure	0
PhoneService	0
MultipleLines	0
InternetService	0
OnlineSecurity	0
OnlineBackup	0
DeviceProtection	0
TechSupport	0
StreamingTV	0
StreamingMovies	0
Contract	0
PaperlessBilling	0
PaymentMethod	0
MonthlyCharges	0
TotalCharges	11
Churn	0

> |

TechSupport	0
StreamingTV	0
StreamingMovies	0
Contract	0
PaperlessBilling	0
PaymentMethod	0
MonthlyCharges	0
TotalCharges	0
Churn	0

> |



# Inconsistency

```
graph TD; A[Inconsistency] --> B[The column Senior Citizen has values in 0's and 1's -> changing it to "No" and "Yes"]; A --> C[Mapping the column(10 : 15) values from "No Internet service" to only "No"]; A --> D[Changing the values of column "MultipleLines" from "No Phone Service" to only "No"]; A --> E[Grouping the column "tenure" by giving the following values -> "0-12" , "12-24" , "24-48" , "48-60" , " >60 "];
```

The column Senior Citizen has values in 0's and 1's -> changing it to "No" and "Yes"

Mapping the column(10 : 15) values from "No Internet service" to only "No"

Changing the values of column "MultipleLines" from "No Phone Service" to only "No"

Grouping the column "tenure" by giving the following values -> "0-12" , "12-24" , "24-48" , "48-60" , " >60 "

3

EDA 

### Correlation Plot :

```
numeric.var <- sapply(churn, is.numeric)
corr.matrix <- cor(churn[,numeric.var])
corrplot(corr.matrix, main="\n\nCorrelation Plot
for Numerical Variables", method="number")
```

