Capstone Project
Background and Objectives

TELECOM CUSTOMER ANALYTICS

DATA SCIENCE
INSTITUTE

Background

1. Background

Analyzing telecom customer churn data is crucial for providers to understand factors influencing customer departure. This data helps identify patterns in service quality, pricing, and customer satisfaction, allowing companies to proactively address issues, improve customer retention strategies, and enhance overall service delivery. Data analysis in this context enables telecom companies to stay competitive by adapting to evolving customer preferences and mitigating potential churn risks.

TA SCIENCE
INSTITUTE

Background

2. Primary Objectives

1. Analyze "Customer Churn" and understand the factors associated with it
2. Develop Churn Prediction Model
3. Implement Machine Learning Algorithms and select the best method for Churn Prediction

TA SCIENCE
INSTITUTE

Background

3. Data

The following datasets are available:

1. Customer Demographics
2. Services Availed
3. **Payment Details**
4. Churn Status

Data : Customer Demographics

Content

This dataset contains customer demographics like gender, senior citizen, partner, tenure and so on

CID	gender	SeniorCitizen	Partner	Dependents	tenure
CI-01	Female	0	No	No	18
CI-02	Male	0	No	No	60
CI-03	Male	0	Yes	Yes	44
CI-04	Female	0	Yes	No	72
CI-05	Female	0	No	No	69
CI-06	Male	0	No	Yes	27
CI-07	Female	0	Yes	Yes	72

Columns	Description	Type	Possible values
CID	Customer ID	Alpha numeric	
gender	Gender	Factor	
SeniorCitizen	Whether Customer is a senior citizen	Factor	0: No 1:Yes
Partner	If customer has a partner	Factor	
Dependents		Factor	
tenure	No. of months customer is using the services	Integer	

Data : Services Availed

Content

This dataset contains information of services availed by a customer

CID	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	numAdminTickets	numTechTickets
CI-01	Yes	No	DSL	No	No	Yes	No	No	No	0	0
CI-02	Yes	Yes	Fiber optic	Yes	Yes	Yes	Yes	Yes	Yes	0	0
CI-03	Yes	No	Fiber optic	No	No	Yes	No	Yes	Yes	0	0
CI-04	No	No phone service	DSL	No	Yes	Yes	No	Yes	Yes	0	7
CI-05	Yes	Yes	DSL	Yes	Yes	No	Yes	No	No	0	0

Columns	Description	Type	Possible values
CID	Customer ID	Alpha numeric	
PhoneService	If customer has availed phone service	Factor	Yes and No
MultipleLines	If customer has multiple lines	Factor	Yes, No, No phone service
InternetService	What Internet service a customer is using	Factor	DSL, Fibre optic, No
OnlineSecurity	If customer has availed online security	Factor	Yes, No, No internet service
OnlineBackup	If customer has availed online backup	Factor	Yes, No, No internet service
DeviceProtection	If customer has availed device protection	Factor	Yes, No, No internet service
TechSupport	If customer has availed Technical Support	Factor	Yes, No, No internet service
StreamingTV	If customer has availed streaming TV service	Factor	Yes, No, No internet service
StreamingMovies	If customer has availed streaming movie service	Factor	Yes, No, No internet service
numAdminTickets	No. of admin tickets raised by a customer	Integer	
numTechTickets	No. of technical tickets raised by a customer	Integer	

DATA SCIENCE INSTITUTE

Data : Payment Details

Content

This dataset contains payment details information of a customer

CID	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges
CI-01	Month-to-mont	Yes	Mailed check	52.19
CI-02	One year	Yes	Credit card (automatic)	118.79
CI-03	Month-to-mont	Yes	Bank transfer (automatic)	95.14
CI-04	Two year	Yes	Electronic check	56.39
CI-05	Two year	Yes	Mailed check	69.09
CI-06	Month-to-mont	No	Mailed check	21.34
CI-07	Two year	Yes	Electronic check	76.54

Columns	Description	Type	Possible values
CID	Customer ID	Alpha numeric	
Contract	Type of contract	Factor	1 year, 2 year, month to month
PaperlessBilling	If customer is using PaperlessBilling	Factor	Yes and No
PaymentMethod	Type of payment method used	Factor	Credit card (automatic), Electronic check, Mailed check, Bank transfer (automatic),
MonthlyCharges	Monthly Charges	Factor	Yes, No, No internet service

Data : Churn Status

Content

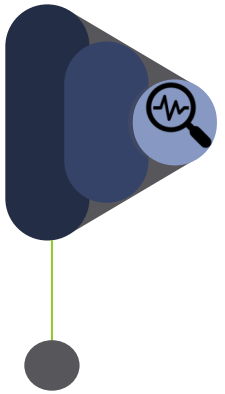
This dataset contains churn status of customers

CID	Churn
CI-01	No
CI-02	No
CI-03	No
CI-04	Yes
CI-05	No
CI-06	No

Columns	Description	Type	Possible values
CID	Customer ID	Alpha numeric	
Churn	Churn Status	Factor	Yes and No

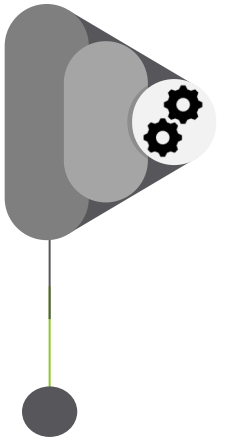
Next steps

Data management



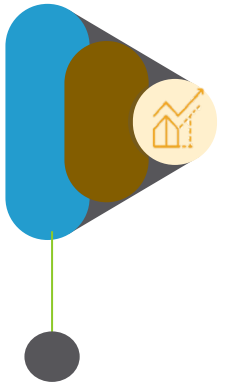
- Compile 4 datasets using Customer ID
- Data cleaning , Handling missing values and completing Basic Data checks
- Check if any variables needed to be feature coded i.e made into groups or want to be left as continuous variables

Descriptive Statistics & Data visualization



- Explore data for attrition rate, are any services playing a role in attrition, etc.
- How can this data be presented better visually ?
- Once again post Data visualization check if any variable needs to be feature coded

Predictive modelling



- Develop a model to predict customer churn
- Using different Predictive model techniques to find Significant variables
- Ensure you follow all steps like Train and test data , checking for Multicollinearity
- Check if any other ML technique fits better

DATA SCIENCE INSTITUTE

* You can go through the Questions and Hint ppt to understand each section better / cross check if you have done all the steps mentioned