

# CUSTOMER CHURN ANALYSIS FROM DATA TO STRATEGY

Summer Projects - 2024

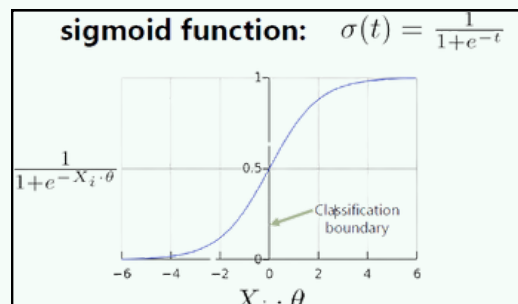
## Data Science

## Consulting

### Machine Learning Concept

#### Logistic Regression

Logistic regression is a supervised machine learning algorithm widely used for binary classification tasks. It uses a sigmoid function for classifying data.



### Exploratory Data Analysis

We used the following libraries to analyse the data



- EDA helps in knowing more about the data
1. The type of data - categorical or numerical.
  2. Number of Missing Values in the data.
  3. Finding Correlation between the different features.

### Data Encoding

Data Encoding is the process of converting categorical data into numerical data that can be fed into the model

- 1. One - Hot Encoding:** It is a binary encoding for mapping unrelated data

id	color
1	red
2	blue
3	green
4	blue

One Hot Encoding

id	color_red	color_blue	color_green
1	1	0	0
2	0	1	0
3	0	0	1
4	0	1	0

- 2. Ordinal Encoding:** It is a hierarchical encoding method, used for related data categories

Original Encoding	Ordinal Encoding
Poor	1
Good	2
Very Good	3
Excellent	4

### Data Imputation

Imputation fills in the missing values with some number. For instance, we can fill in the mean value along each column. The simplest method is to drop the datapoint. According to the data type and correlation, the missing data points are filled. We have also used **KNN Imputer** and **MICE Algorithm** which are powerful statistical tools to fill the missing values.

	col1	col2	col3	col4	col5		col1	col2	col3	col4	col5
0	2	5.0	3.0	6	NaN	mean()	0	2.0	5.0	3.0	6.0
1	9	NaN	9.0	0	7.0		1	9.0	11.0	9.0	0.0
2	19	17.0	NaN	9	NaN		2	19.0	17.0	6.0	9.0

### Feature Engineering

#### Feature Transformation

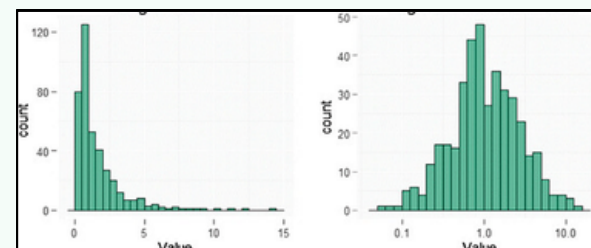
This includes handling missing values, encoding categorical variables, detecting and treating outliers, and applying feature scaling techniques like standardization and normalization.

#### Feature Scaling

1. Standardization:  $(X_i - X_{\text{mean}})/\sigma$
2. Min - Max Scaling:  $(X_i - X_{\text{min}})/(X_{\text{max}} - X_{\text{min}})$
3. Mean Normalization
4. Robust Scaling
5. Max Absolute Scaling

### Transformations

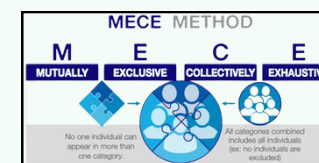
Many machine learning algorithms assume that the data features are normally distributed, this is why handling skewed distribution becomes an essential task in data transformation process, as the skewed data might lead to biased or inaccurate model. They are of 2 types - Power and Function Transformations.



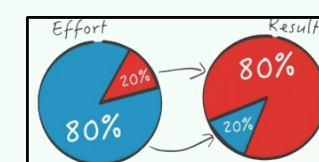
#### Logarithmic Transformation

### Structured Thinking

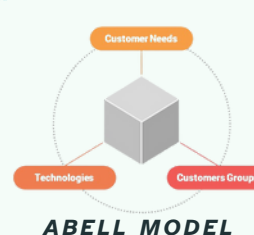
#### MECE PRINCIPLE



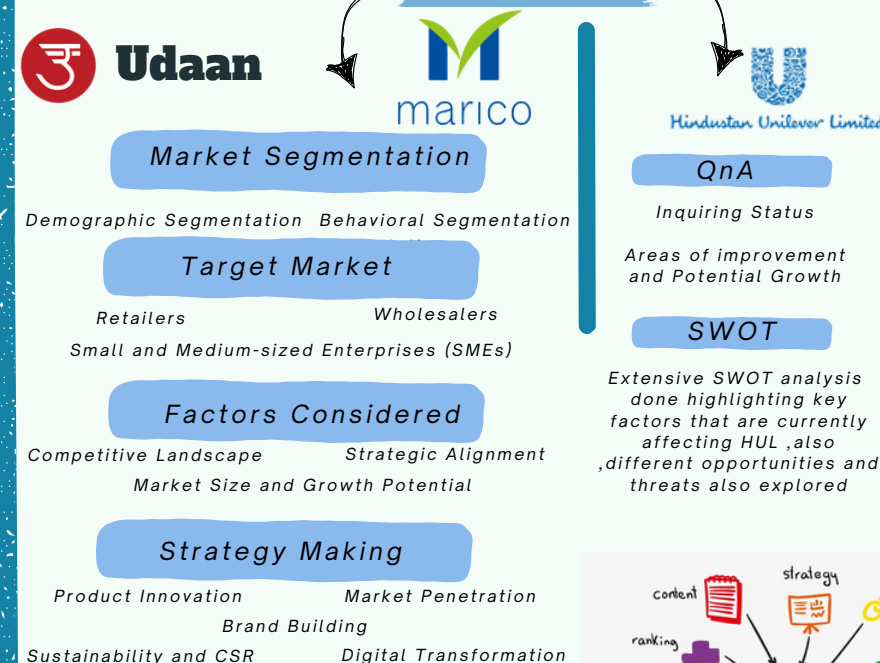
#### PARETO PRINCIPLE



### Frameworks

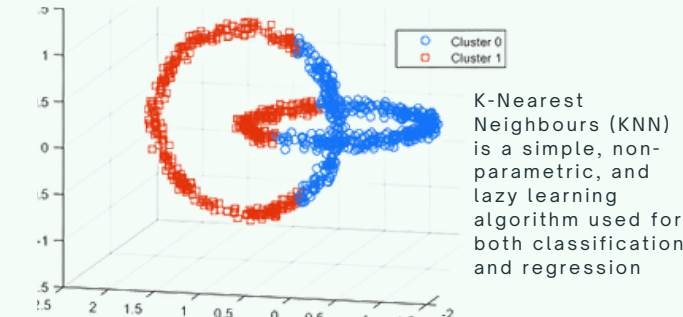


### Case Studies



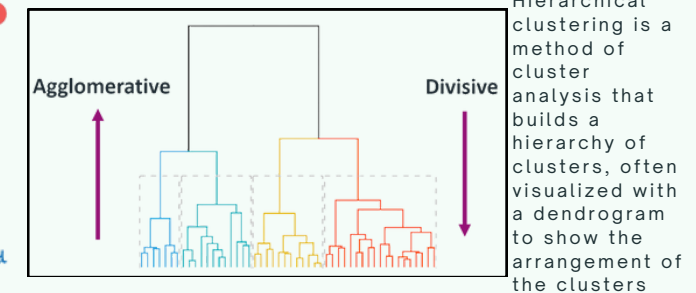
### Clustering

#### K NEIGHBOURS CLUSTERING



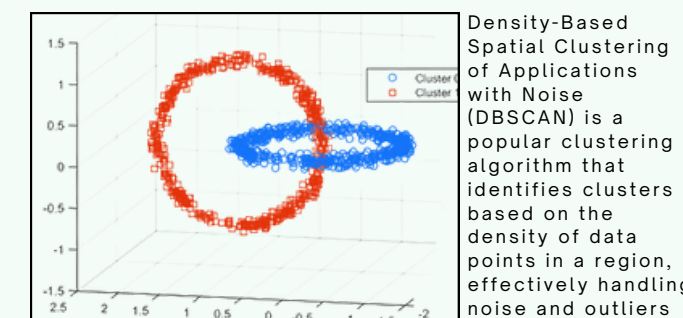
K-Nearest Neighbours (KNN) is a simple, non-parametric, and lazy learning algorithm used for both classification and regression

#### HIERARCHICAL CLUSTERING



Hierarchical clustering is a method of cluster analysis that builds a hierarchy of clusters, often visualized with a dendrogram to show the arrangement of the clusters

#### DBSCAN CLUSTERING



Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a popular clustering algorithm that identifies clusters based on the density of data points in a region, effectively handling noise and outliers

### Types of Marketing Strategies



## Assignments



1 DS  
Basics of libraries



2 DS  
Linear regression using OLS and Gradient Descent



3 DS  
Bank Customer Churn Analysis



4 Consulting  
Marketing Strategy Product Innovation



5 Consulting  
SWOT Analysis



6 Consulting+DS  
Clustering on Bank Customer DB