Capstone Project
Background and Objectives

Retail Analtyics

Background

1. Background

The Online Retail Data Project for the UK in 2023 aims to analyze and derive insights from the growing e-commerce sector. With the increasing trend of online shopping, the project focuses on understanding customer behavior, identifying key market trends, and optimizing business strategies for retailers. Leveraging data science techniques, such as customer segmentation, anomaly detection and campaign response model the project seeks to enhance decision-making processes, improve customer experiences, and drive overall efficiency in the dynamic landscape of the UK's online retail industry.

TA SCIENCE
INSTITUTE

Background

2. Primary Objectives

1. **Analyze sales performance-Overall and By Product**s
2. Study customer buying behavior and create customer segments
3. **Develop Campaign Response Model –Use traditional and ML algorithms**

TA SCIENCE
INSTITUTE

Background

3. Data

The following datasets are available:

1. Online Retail Sales Data
2. Campaign Response

Data : Online Retail Sales Data

Content

This dataset contains sales information of each customer like customer id, invoice no, invoice data, product description, quantity, unit price and so on

CustomerID	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	Country
13313	539993	22386	JUMBO BAG PINK POLKADOT	10	04-01-2023 10:00	1.95	United Kingdom
13313	539993	21499	BLUE POLKADOT WRAP	25	04-01-2023 10:00	0.42	United Kingdom
13313	539993	21498	RED RETROSPOT WRAP	25	04-01-2023 10:00	0.42	United Kingdom
13313	539993	22379	RECYCLING BAG RETROSPOT	5	04-01-2023 10:00	2.1	United Kingdom
13313	539993	20718	RED RETROSPOT SHOPPER BAG	10	04-01-2023 10:00	1.25	United Kingdom
13313	539993	85099B	JUMBO BAG RED RETROSPOT	10	04-01-2023 10:00	1.95	United Kingdom
13313	539993	20682	RED RETROSPOT CHILDRENS UMBRELLA	6	04-01-2023 10:00	3.25	United Kingdom

Columns	Description	Type	Possible values
CustomerID	Customer ID	numeric	
InvoiceNo	Invoice no	numeric	
StockCode	Product code	Alpha numeric	
Description	Product description	Factor	
Quantity	Quantity of product purchased	numeric	
InvoiceDate	Invoice Date	Timestamp	
UnitPrice	Price of product	Float	
Country	Country	Factor	

Data : Campaign Response Data

Content

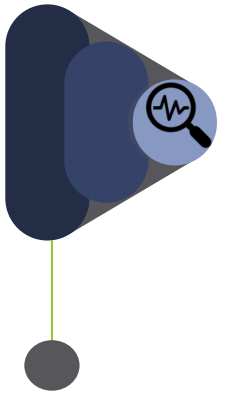
This dataset contains campaign response data for each customer

CustomerID	response	n_comp	loyalty	nps	n_communications
12346	1	2	1	7	8
12747	0	1	1	3	3
12748	1	0	1	9	6
12749	0	4	1	2	5
12820	0	4	1	2	2
12821	1	2	1	5	7

Columns	Description	Type	Possible values
CustomerID	Customer ID	numeric	
response	Response to the email marketing campaign	Factor	1: responded positively 0: no response
n_comp	No. of complaints 3 months prior to the campaign	numeric	
loyalty	Member of loyalty program	Factor	1: Yes 0: No
nps	Net promoter score	Ordinal	0 to 10
n_communications	No. of communications 3 months prior to the campaign	numeric	

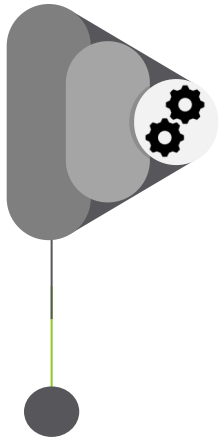
Next steps

EDA



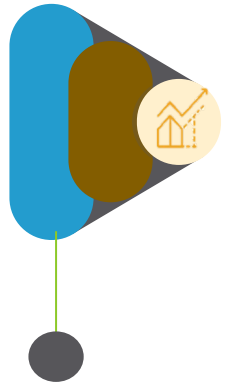
- Explore data to find top/bottom products
- Derive new features at customer level
- Analyze customer buying behavior
- Merge customer level features with campaign response data

Customer Segmentation



- Perform cluster analysis
- Interpret cluster solution

Campaign response modelling



- Develop a model to predict customer response
- 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