Customer Behaviour Analysis

Ninos Kako

Capstone Project
Sprint 1

Customer Behaviour Analysis

- Business Questions
- Machine Learning Approach
- Dataset
- Next Steps

Business Questions

- Increase Sales
- Increase Customer Retention
- Improve Marketing
- Improve Pricing
- Improve Inventory Management and Assortment

Machine Learning Approach

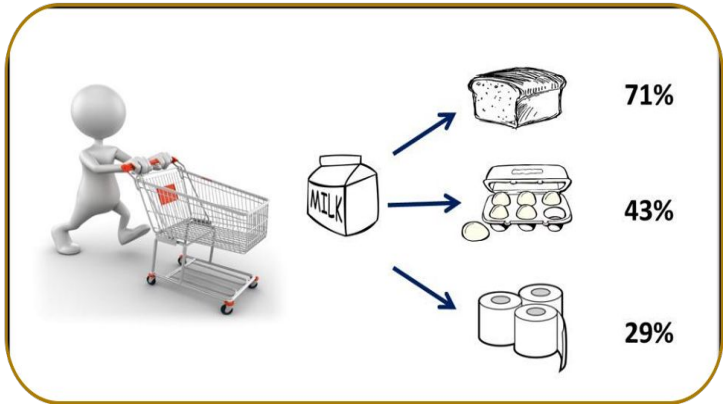
- **Recency, Frequency and Monetary RFM Analysis.**
- **K-means Clustering**
- **Market Basket Analysis**

RFM Analysis and K-means

- Customer Segmentation
  - *Recency*
  - *Frequency*
  - *Monetary*
- Main Benefits
  - Increase Sales and Revenue
  - Increase Customer Retention
  - Improve Marketing

Market Basket Analysis

- Understand customers' purchasing patterns
- Identify which items are frequently purchased together



- Main Benefits
  - Improve pricing and customer retention
  - Improve inventory management and assortment

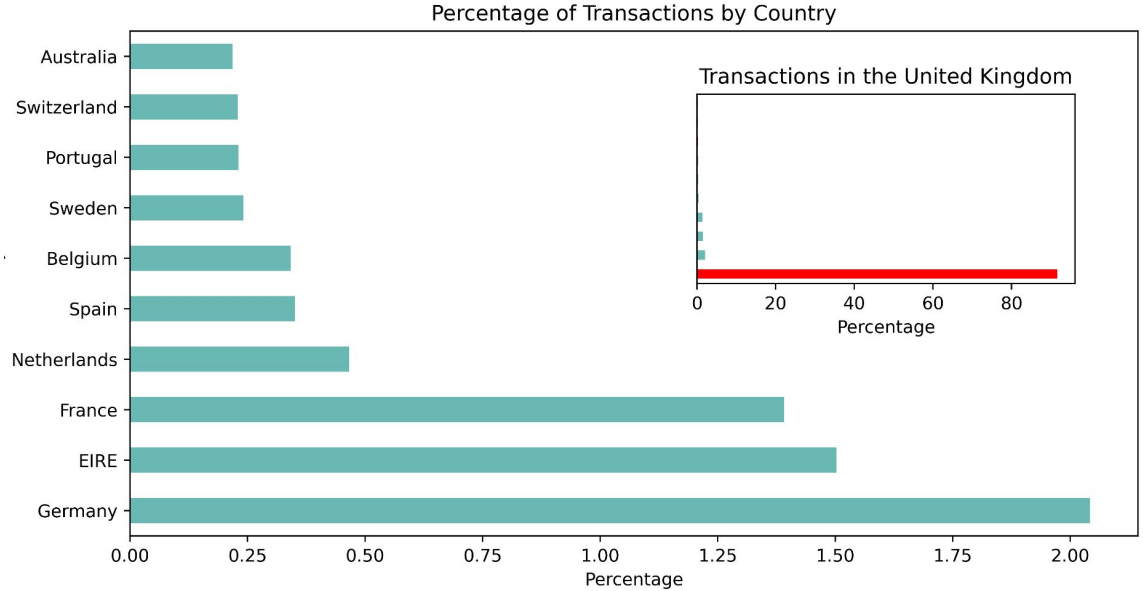
The Dataset: Online Retail

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
1	489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
2	489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom

- **6k** Customers
- From *DEC-2009* to *DEC-2011*
- Total number of transactions: **54k**
- **19.3m GBP** sales value

The Dataset: Transactions

- **49k** Transactions
- **91%** were made in the United Kingdom



The Dataset: Monthly Sales

- Average Sales **600k Jan-Aug**
- **Aug-Nov** increase to over **1.4m**

The Dataset: Issues

- Missing values:
  - Customer ID: **240k**
  - Description: **4k**
- Bad Documentation
  - **1.2k** items have more than one description.
  - **300** items have no description at all (missing values).
- Item price is not constant:
  - For example: a Car Flag price ranges between £0.42 - £1,000
- Other charges/expenses:
  - Bank Charges
  - Amazon Fees
  - Bad debt Adjustment

Next steps

1. Dealing with missing values:
    a. Description column.
    b. Customer ID column.
2. Feature engineering to calculate:
    a. Recency
    b. Frequency
    c. Monetary
3. K-means clustering.

Thank You!