**Author: Eduardo Matos**

**Subject: Machine Learning for Business**

**CCT College Dublin**

CA1 – Machine Learning

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | Machine Learning for Business |
| **Assessment Title:** | CA1 |
| **Student Full Name:** | **Eduardo Matos** |
| **Student Number:** | **Sba23298 - Eduardo Matos** |
|  |  |
|  |  |
|  |  |
|  |  |
| **Assessment Due Date:** | Sunday 14th April 2024 |
| **Date of Submission:** | Sunday 14th April 2024 |

Index

[Introduction 1](#_Toc164025884)

[Objectives 1](#_Toc164025885)

[Problem definition 1](#_Toc164025886)

[Exploratory Data Analysis 1](#_Toc164025887)

[Percent of products returned. 1](#_Toc164025888)

[15 Items Purchased Less 2](#_Toc164025889)

[Number of Purchases by Country 3](#_Toc164025890)

[Revenue/Growth of The Company Through the Year 3](#_Toc164025891)

[Clustering Analysis 4](#_Toc164025892)

[Clustering Revenue vs Frequency - K-Means 4](#_Toc164025893)

[Clustering Revenue vs Recency – DBSCAN 5](#_Toc164025894)

[Clustering Revenue vs Frequency - K-Means 7](#_Toc164025895)

[References 7](#_Toc164025896)

## Introduction

In the retail industry, organizations can gain valuable insights to attract and retain customers, improve business efficiency, and ultimately increase sales and customer engagement. By utilizing advanced analysis, retail organizations can better understand their customers, enhance forecasting, and achieve superior results. It is crucial to identify and target customers to secure their loyalty, improve business efficiency, and ultimately enhance performance.

This analysis focuses on clustering regularly purchased products to identify patterns and gain insights into consumer buying behaviour. This approach facilitates a better understanding of the dynamics of everyday purchases, enabling more targeted marketing strategies.

## Objectives

This analysis will uncover customer patterns and buying habits with the intention of providing the company with insight to consider in order to help define future marketing and business strategies by:

* 1. Exploratory Data Analysis.
  2. Clustering Analysis.

## Problem definition

The analysed company has a non-positive steady monthly growth and turnover rate and needs to conduct an analysis to determine strategies to improve business performance.

* To achieve a more precise clustering of products, we excluded products that cost over 7.5 € from our datasets. Our analysis aims to cluster regularly purchased products, and those over this price point are not considered in this case.
* To perform an analysis with less bias, the clustering analysis will solely focus on the United Kingdom, as it accounts for the majority of the company's purchases. This is due to the potential variation in customer behaviour between countries.

## Exploratory Data Analysis

### Percent of products returned.

The percent of invoices returned is 1.8%. What corresponds to 6.038 invoices out of 328.139.

* - It is not considered necessary to carry out further analysis with regard to returns, as 1.8% could be considered a reasonable level.

Esquemático

Descripción generada automáticamente

### 15 Items Purchased Less

The items purchased less are:

Imagen que contiene Patrón de fondo

Descripción generada automáticamente

|  |
| --- |
| BAKING MOULD EASTER EGG MILK CHOC |
| WOODEN ROUNDERS GARDEN SET |
| PEN, 10 COLOURS - SPACEBOY |
| MAGIC DRAWING SLATE FIDO |
| TRAY FOR TV DINNER ,SPACEBOY |
| BEECH WOOD P/FRAME |
| HANGING CHICK CREAM DECORATION |
| NUMBER TILE VINTAGE FONT No |
| ACRYLIC JEWEL ICICLE, BLUE |
| SET OF 6 SPICE TINS PANTRY DESIGN |
| HEN PARTY BANNER TAPE |
| PINK BITTY LIGHT CHAIN |
| MIDNIGHT BLUE CRYSTAL DROP EARRINGS |
| CRACKED GLAZE NECKLACE RED |

Further analysis could be carried out to determine whether it is worth keeping minimum spend items in the catalogue.

### Number of Purchases by Country

* United Kingdom is the country with the highest number of purchases with 92.08% of sales, 14.593.
* Australia, Austria, Bahrain, Belgium, Brazil, Canada, Channel Islands, Cyprus, Denmark, Finland are 10 of the countries with less than 1% of the sales.

Forma

Descripción generada automáticamente con confianza baja

Further analysis could be undertaken to consider whether it is necessary to cease offering products in some countries.

### Revenue/Growth of The Company Through the Year

Over 2010, the company had a steady turnover until August, since then and until November it experienced an exponential turnover, finally in December the turnover dropped drastically.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Over the year as a whole, the company did not experience significant growth, although there was some growth in the first and last quarters of the year.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

As can be seen over the year as a whole, the company did not experience significant growth, although there was some growth in the first and last quarters of the year.

## Clustering Analysis

### Clustering Revenue vs Frequency - K-Means

Gráfico

Descripción generada automáticamente

* As can be seen from this graph, most of the customers appear to be occasional customers who do not generate a great deal of revenue for the company, but it looks like there is a second and third group that appear to be very interesting and although their frequency of purchase is not significant, the revenue they generate for the company is revealing.

Further analysis could be made to consider how to attract more the attention of these groups.

### Clustering Revenue vs Recency – DBSCAN

* Gráfico, Gráfico de dispersión

  Descripción generada automáticamenteAs can be seen in the graph most of the customers are inactive and occasionally spender. Additionally, there seems to be a few more groups that could be interesting to consider for future analysis.
  + Frequent and big spenders - customers who have not spent more than 150 days since their last purchase and sometimes generate more than €5,000 in revenues.

Further analysis could be defined to incentivize this public, frequent and big spenders, as the number of purchases each year seems possible to increase considering how often they buy and the benefits they make for the company in each one.

### Clustering Revenue vs Frequency - K-Means

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

* As can be seen in the graph above, 4 groups are considered, in this analysis. Customers are quite spread being a minimum majority concentrated as a group of low activity and defined by not purchasing often but generating most of the revenues.
* A second group emerges that although seems to buy more often than the previous one with low activity, this tome this one seems to generate a third of what is generated by the previous one.
* In addition, a third group seems to be the most interesting for further analysis, where high activity customers buy more frequently than the previous two groups and generate slightly more revenue for the company than the previous moderate activity group.

Further analysis of this third group, classified as high activity, could be carried out to encourage more activity from this group. As they are studied as a group that purchases relatively often, marketing strategies could be promoted to engage them more.

## References

Scikit-learn.org. (2017). sklearn.cluster.DBSCAN — scikit-learn 0.22 documentation. [online] Available at: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html.

Sharma, A. (2020). How Does DBSCAN Clustering Work? | DBSCAN Clustering for ML. [online] Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2020/09/how-dbscan-clustering-works/.

GeeksforGeeks. (2010). Print nodes at k distance from root. [online] Available at: https://www.geeksforgeeks.org/print-nodes-at-k-distance-from-root/ [Accessed 10 Apr. 2024].