

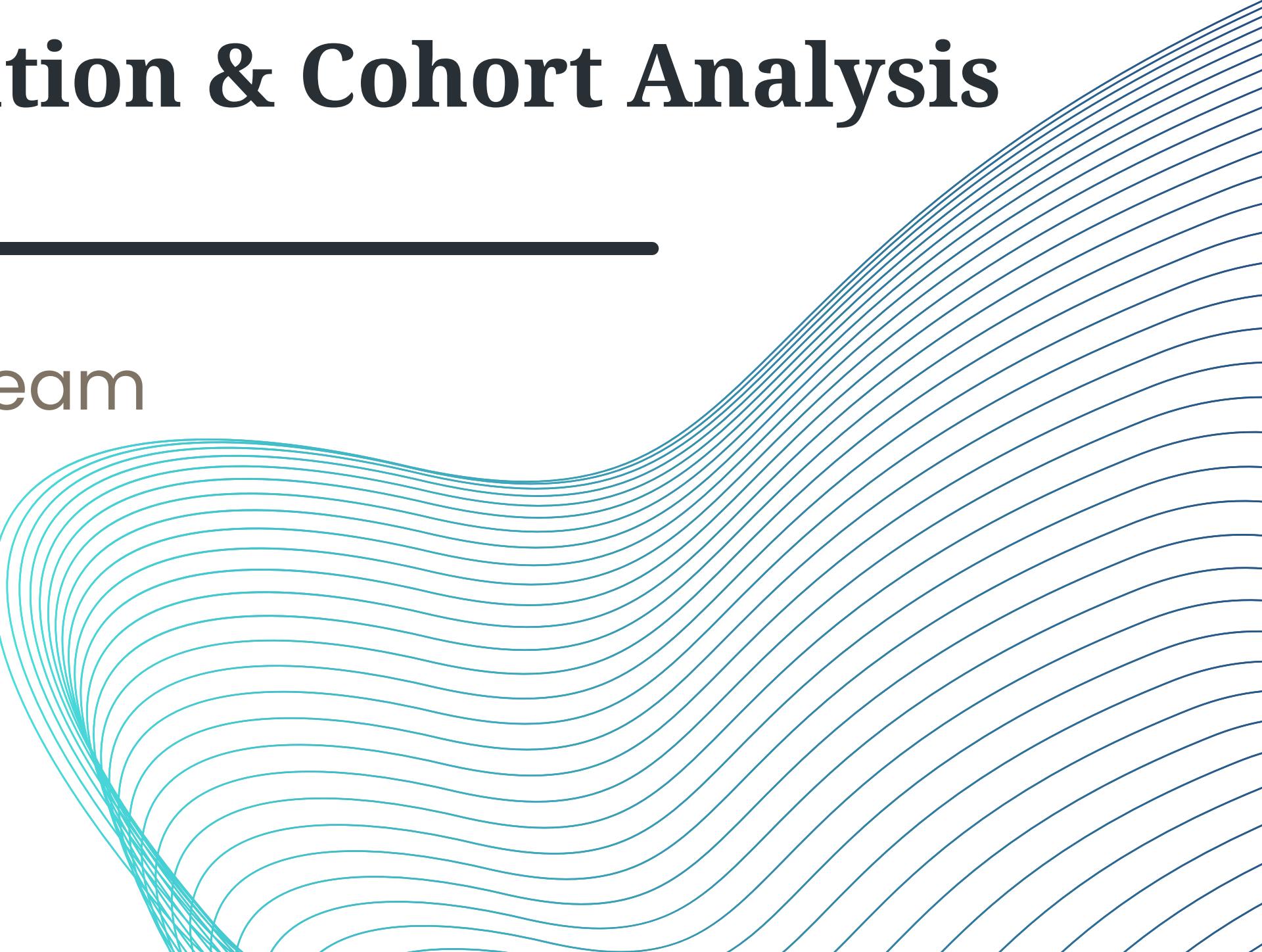
الأكاديمية السعودية الرقمية  
SAUDI DIGITAL ACADEMY

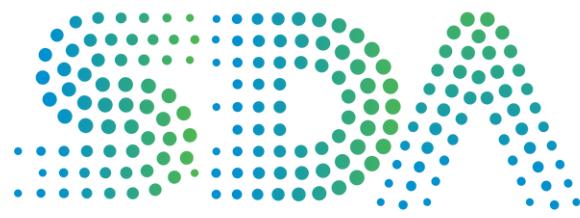


# RFM Customer Segmentation & Cohort Analysis

---

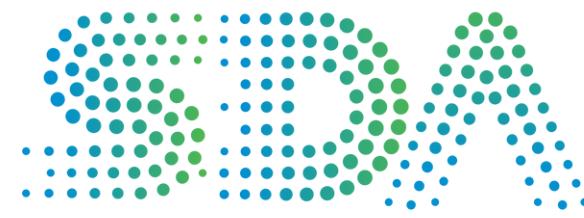
By: G-1 Team





## Outlines:

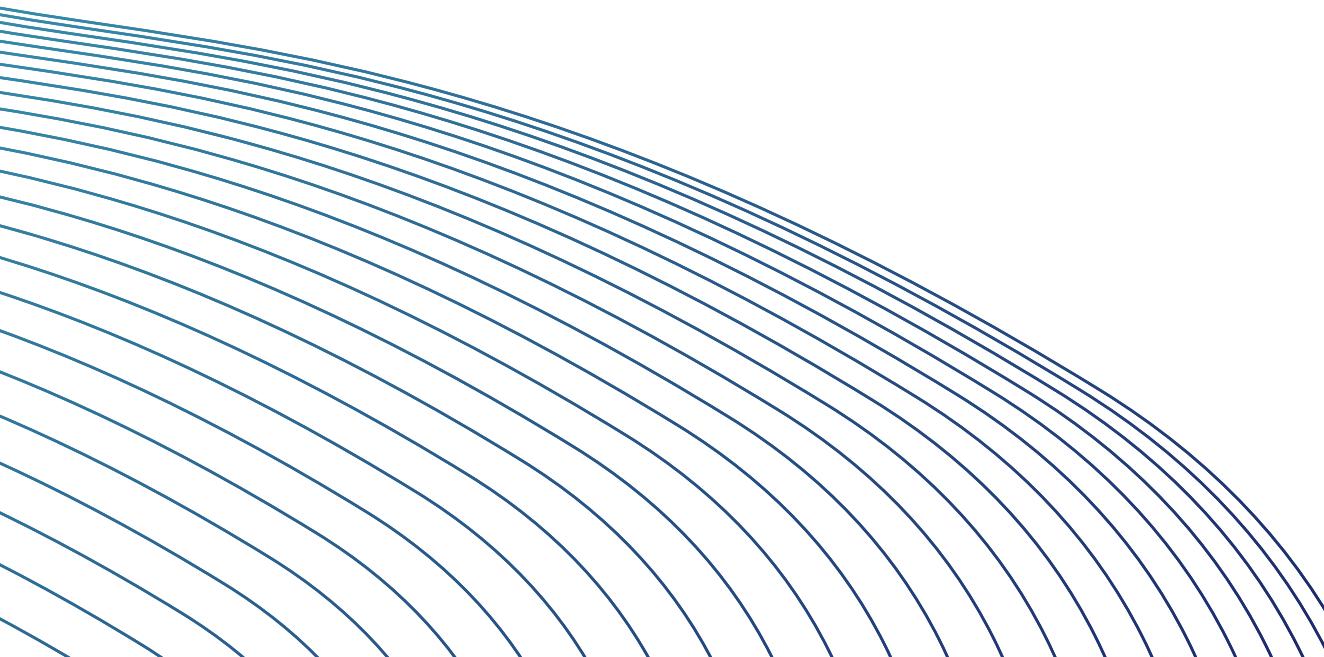
- Project Overview
- Data Cleaning & Exploratory Data Analysis
- RFM Analysis
- Customer Segmentation
- Applying Clustering
- Create Cohort and Conduct Cohort Analysis
- Conclusion

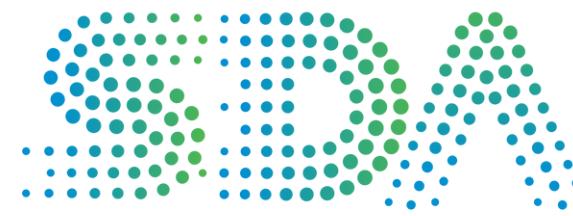


الأكاديمية السعودية الرقمية  
SAUDI DIGITAL ACADEMY



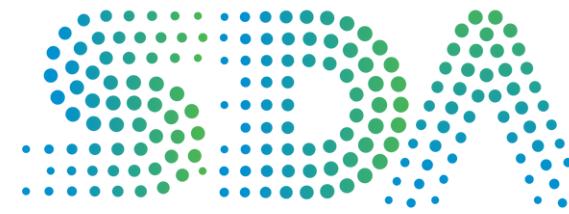
# Project Overview





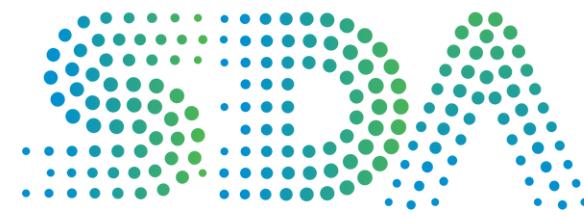
## Problem Statement

**In the retail industry, it is crucial for businesses to comprehend the value obtained from each customer.**



## Goal

The aim of this project is to analyze the online retail dataset, understand customer behavior, do customer segmentation in order to derive actionable insights for effective business strategies.

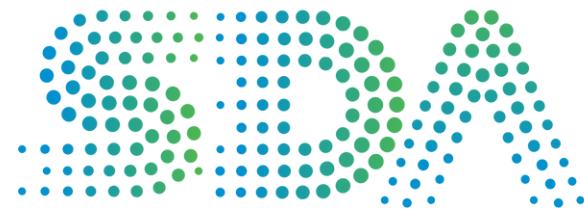


# Dataset

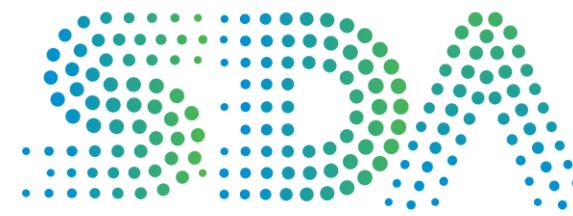
This is a transnational data set between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.

It consists of 8 columns & more than 541000 rows.

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
538349	22637	PIGGY BANK RETROSPOT	1	2010-12-10 14:59:00	5.06	NaN	United Kingdom
555488	23146	TRIPLE HOOK ANTIQUE IVORY ROSE	2	2011-06-03 13:43:00	3.29	16942.0	United Kingdom
546007	21078	SET/20 STRAWBERRY PAPER NAPKINS	6	2011-03-08 15:34:00	0.85	NaN	United Kingdom
551278	23254	CHILDRENS CUTLERY DOLLY GIRL	8	2011-04-27 13:13:00	4.15	12483.0	Sweden
579512	23437	50'S CHRISTMAS GIFT BAG LARGE	1	2011-11-29 16:47:00	2.46	NaN	United Kingdom



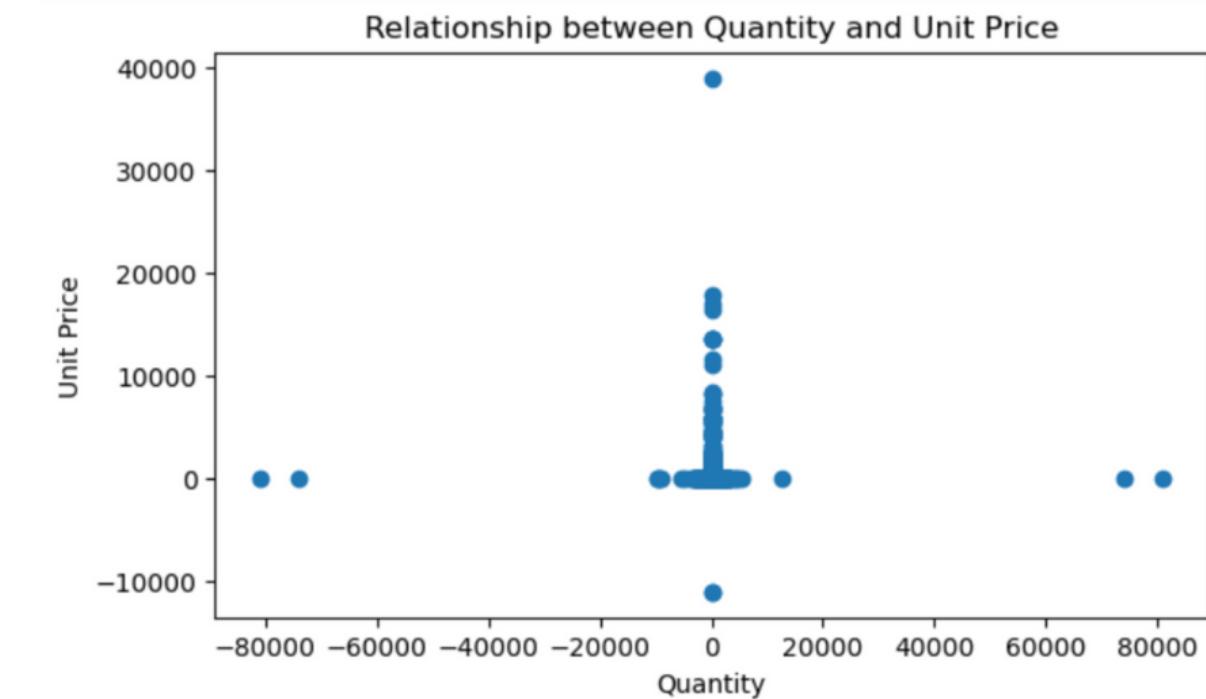
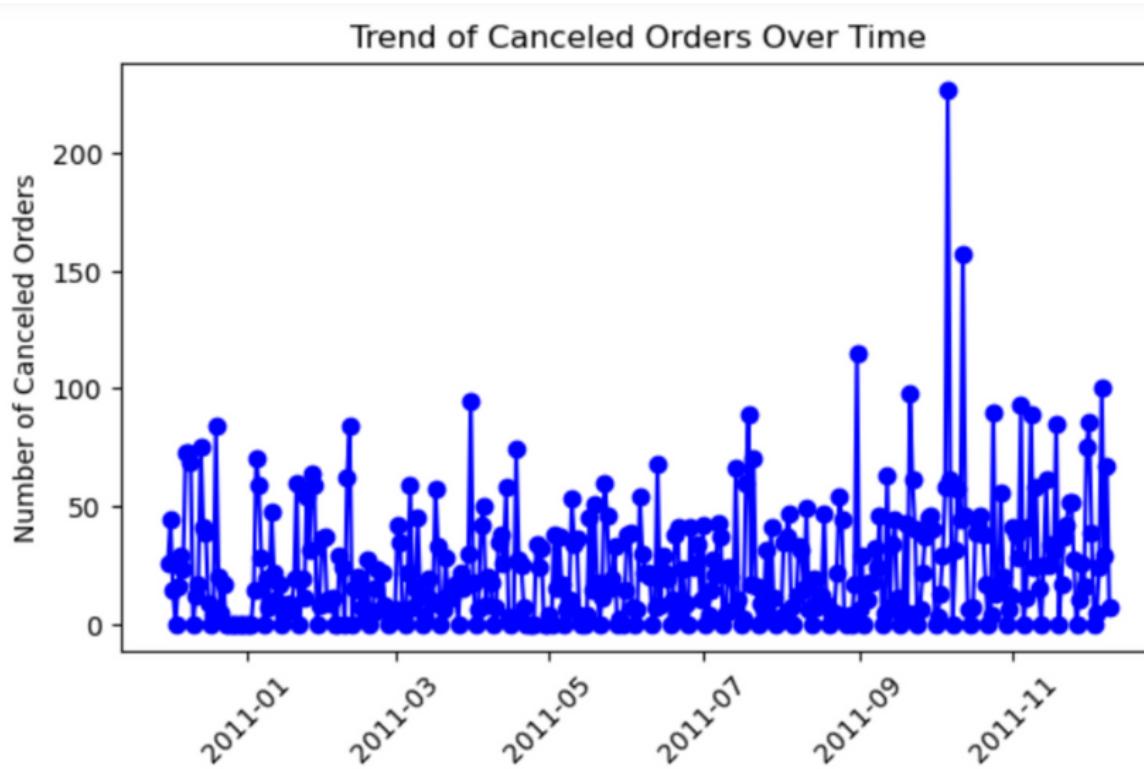
# Data Cleaning & Exploratory Data Analysis

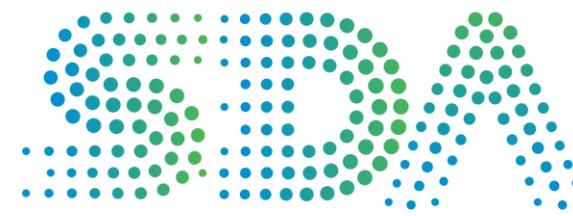


# Data Cleaning & EDA

## Dealing with noisy data:

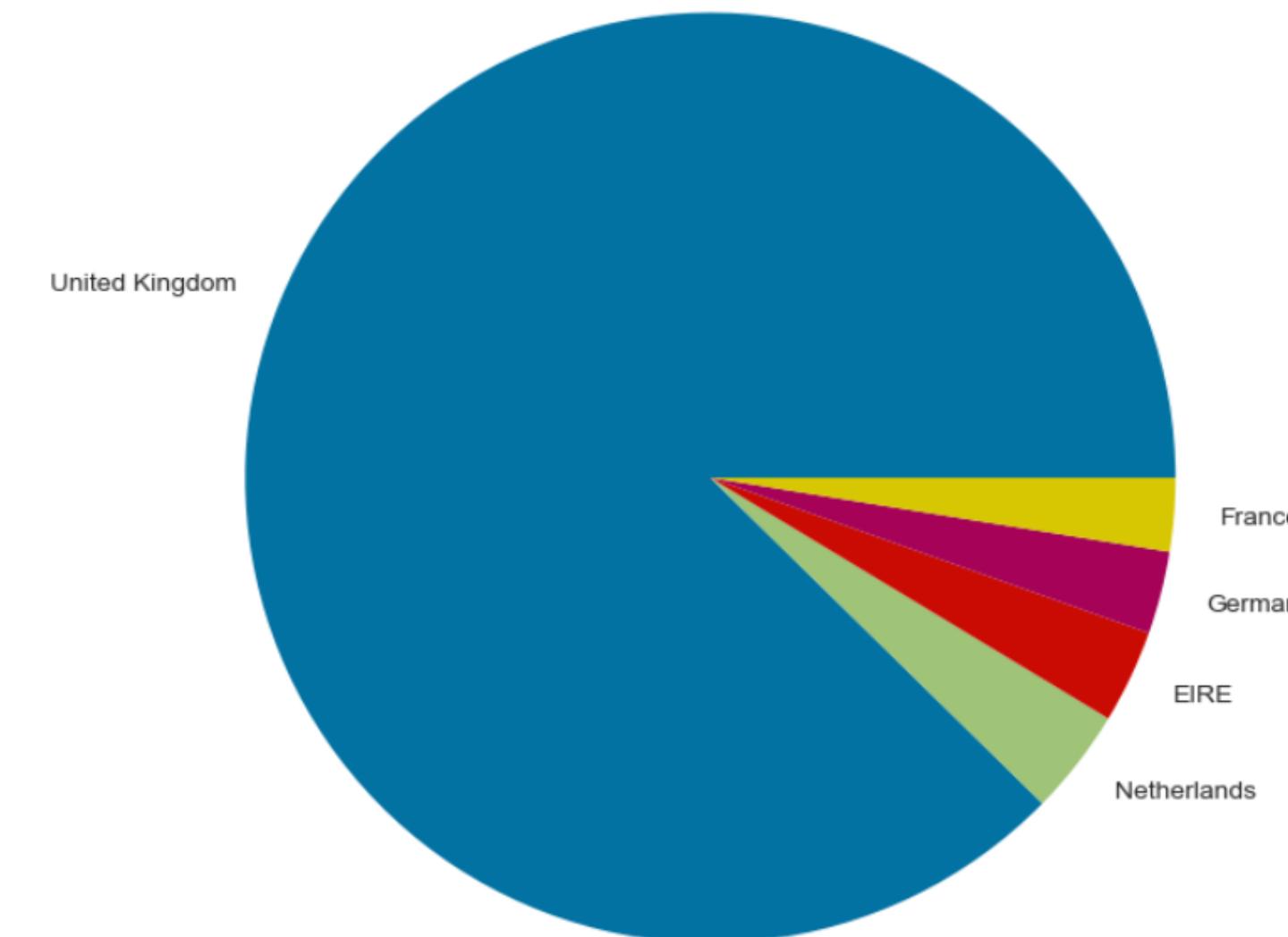
- Missing Values
- 0 values
- Duplicated
- Typos

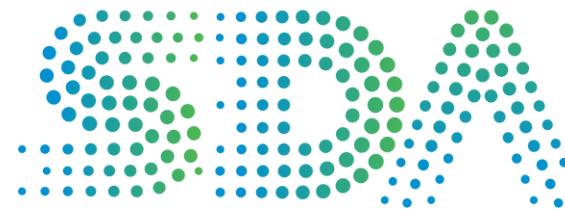




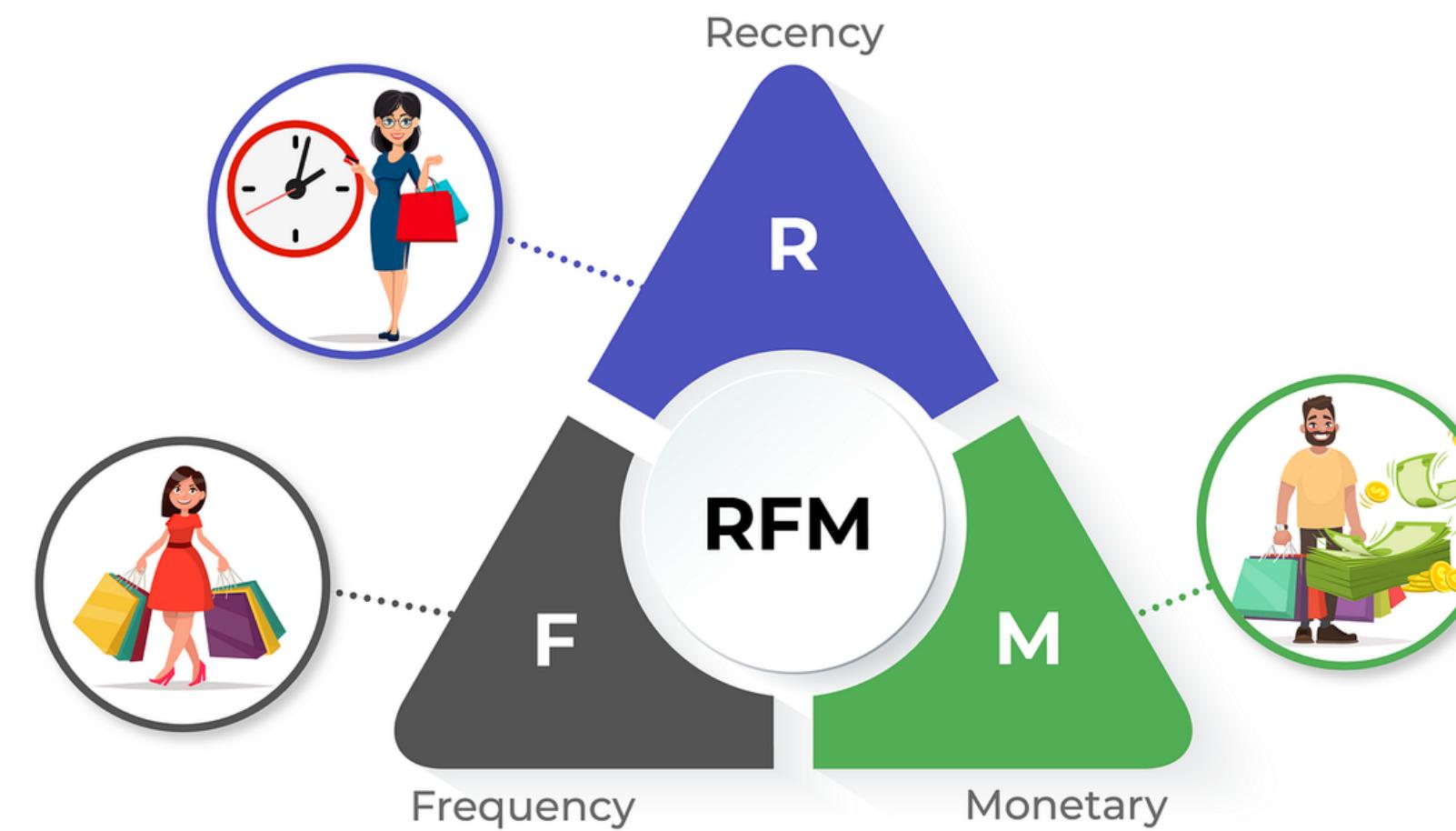
# Exploring Customers by Country

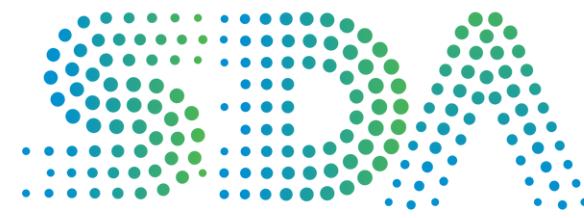
Total Price by Country (Top 5)





# RFM Analysis



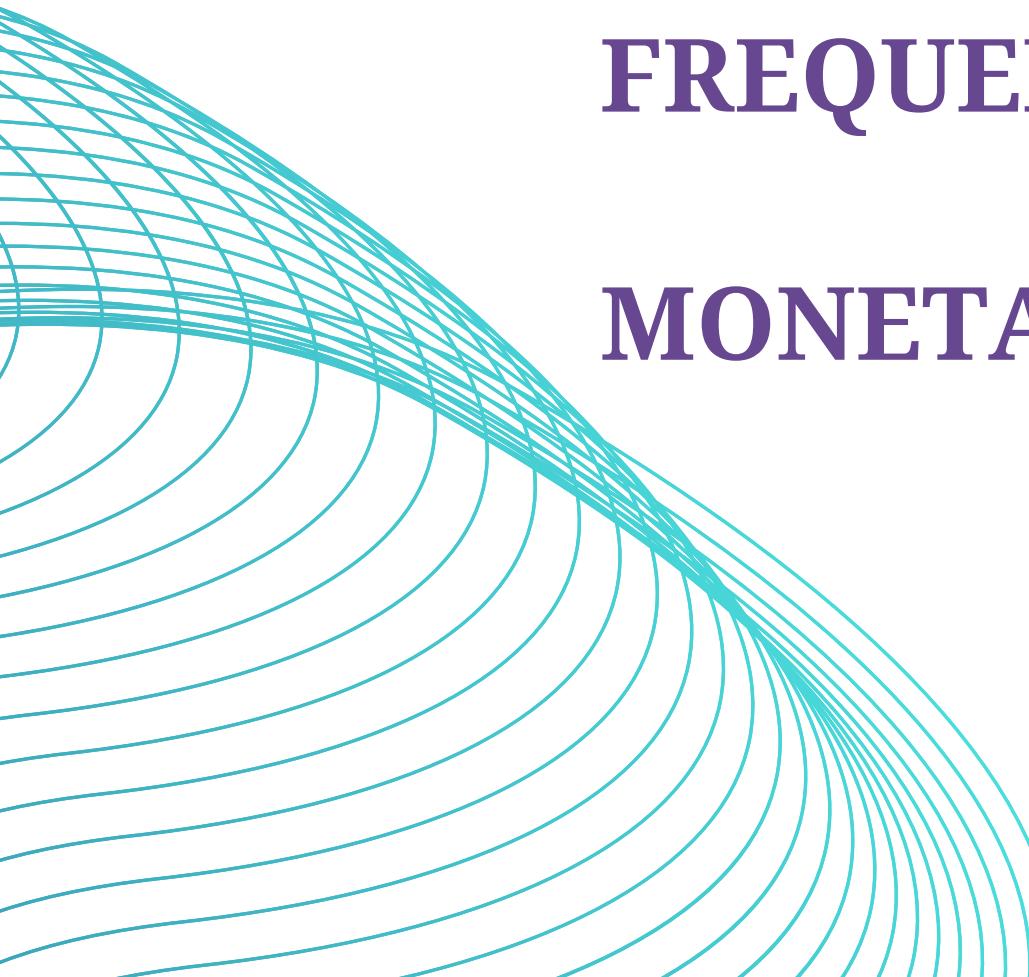


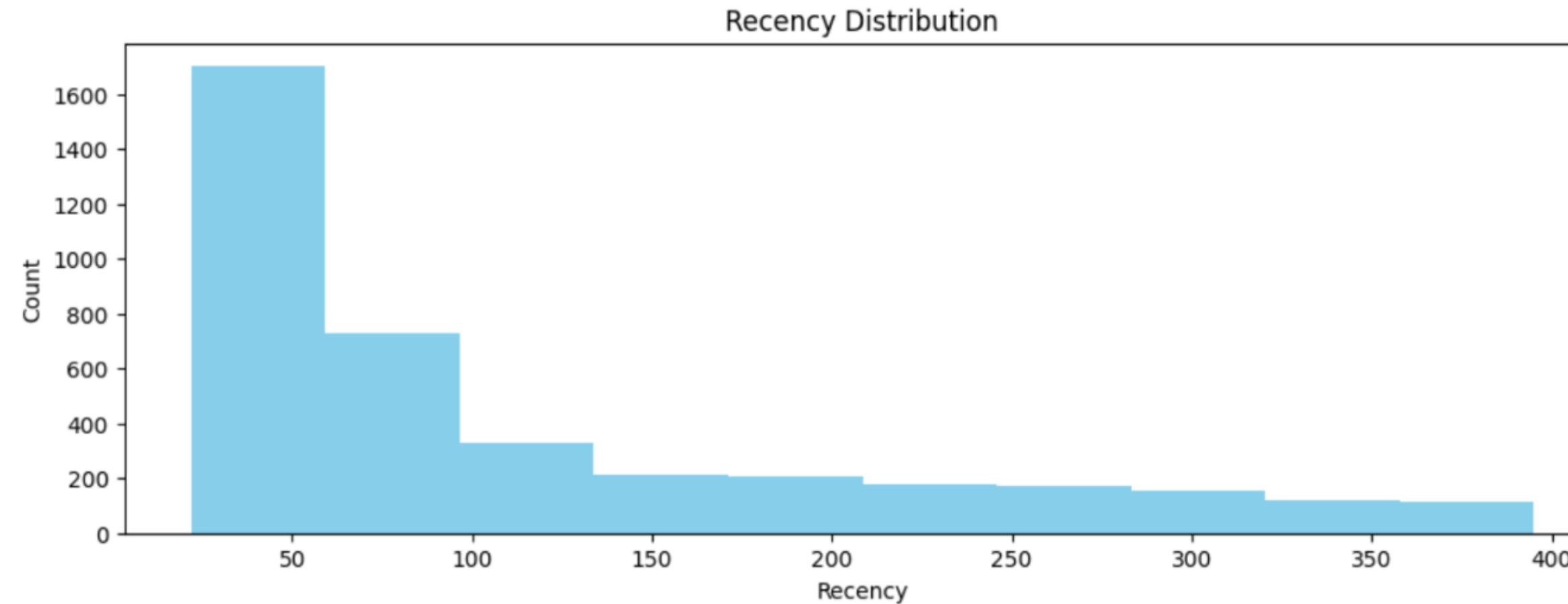
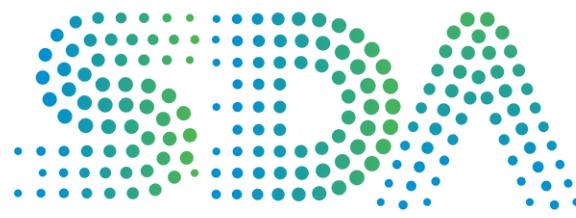
## What RFM means?

**RECENCY:** Time since last purchase.

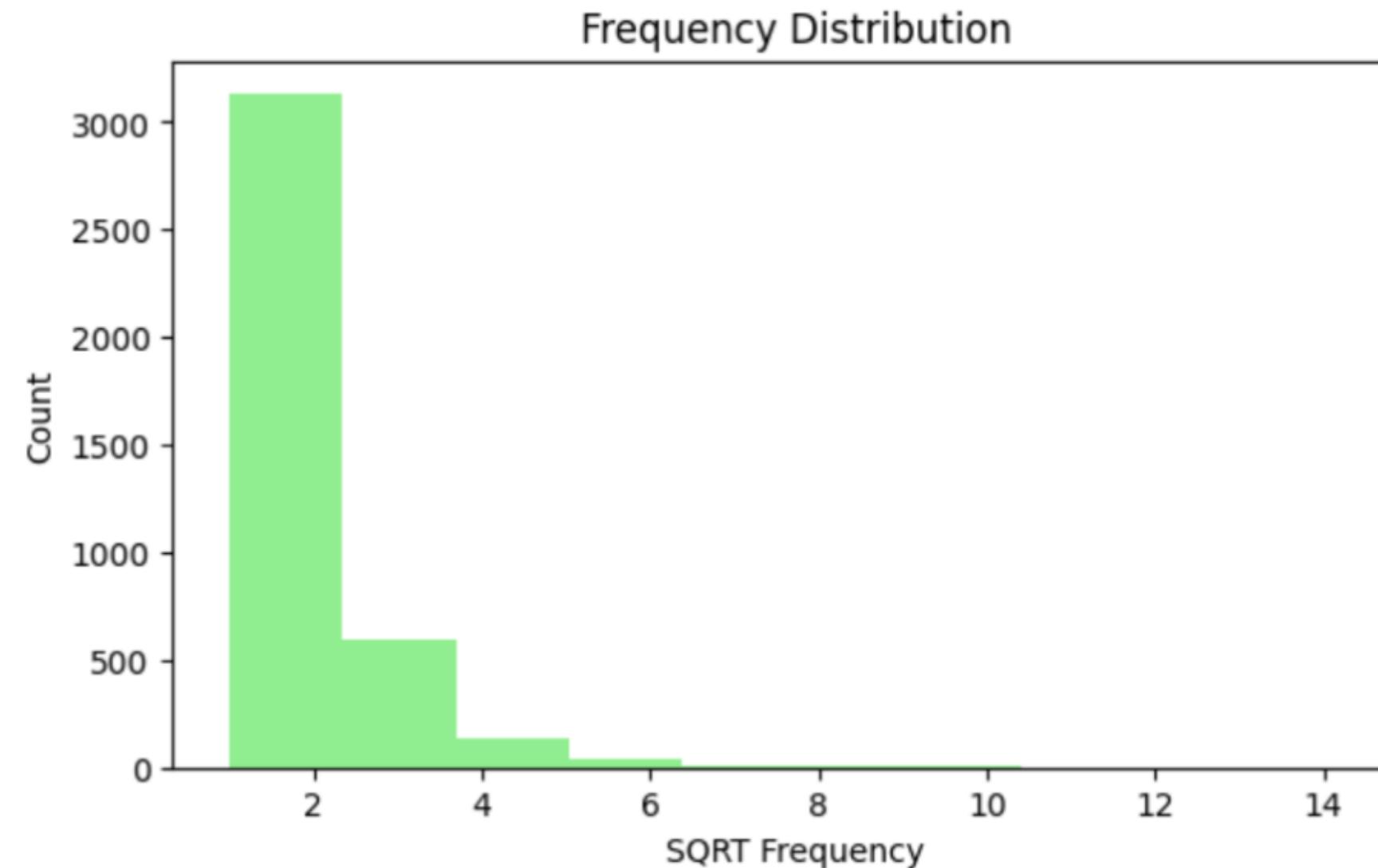
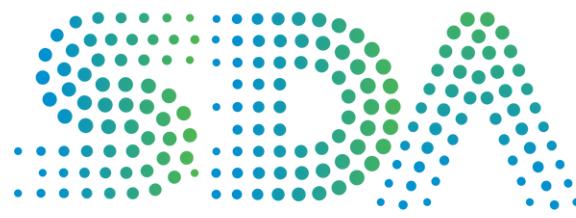
**FREQUENCY:** Total number of purchases.

**MONETARY VALUE:** Total monetary value.

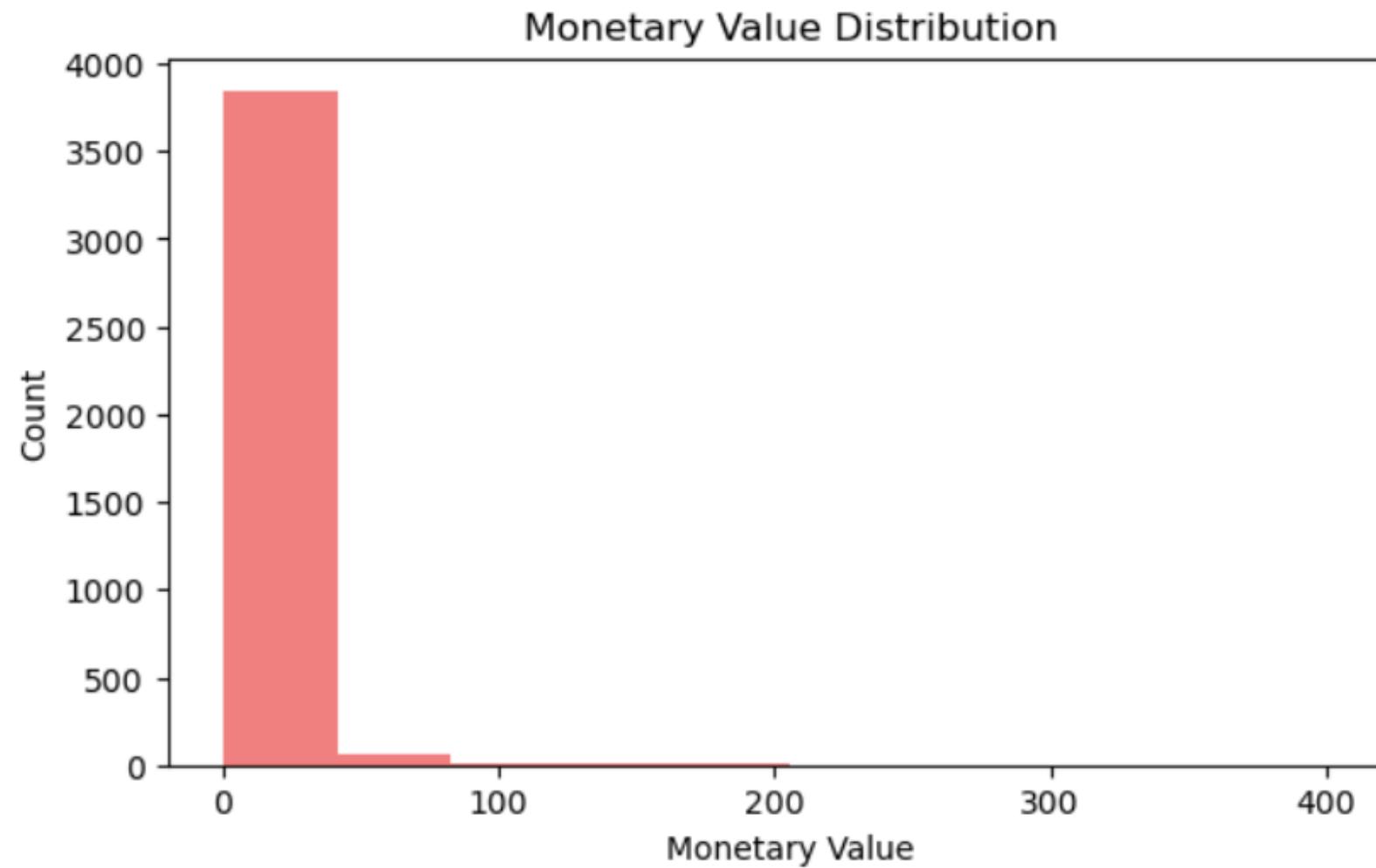
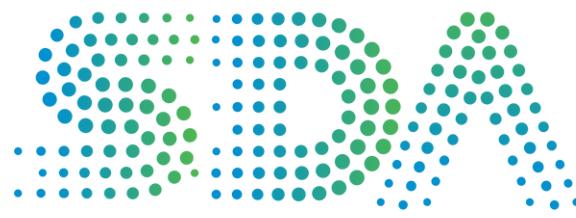




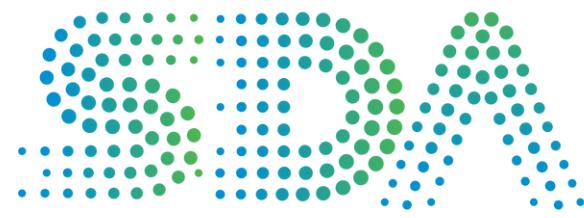
For each customer we calculated the number of days since the customer's last purchase by choosing the last day in the year as a point of reference.



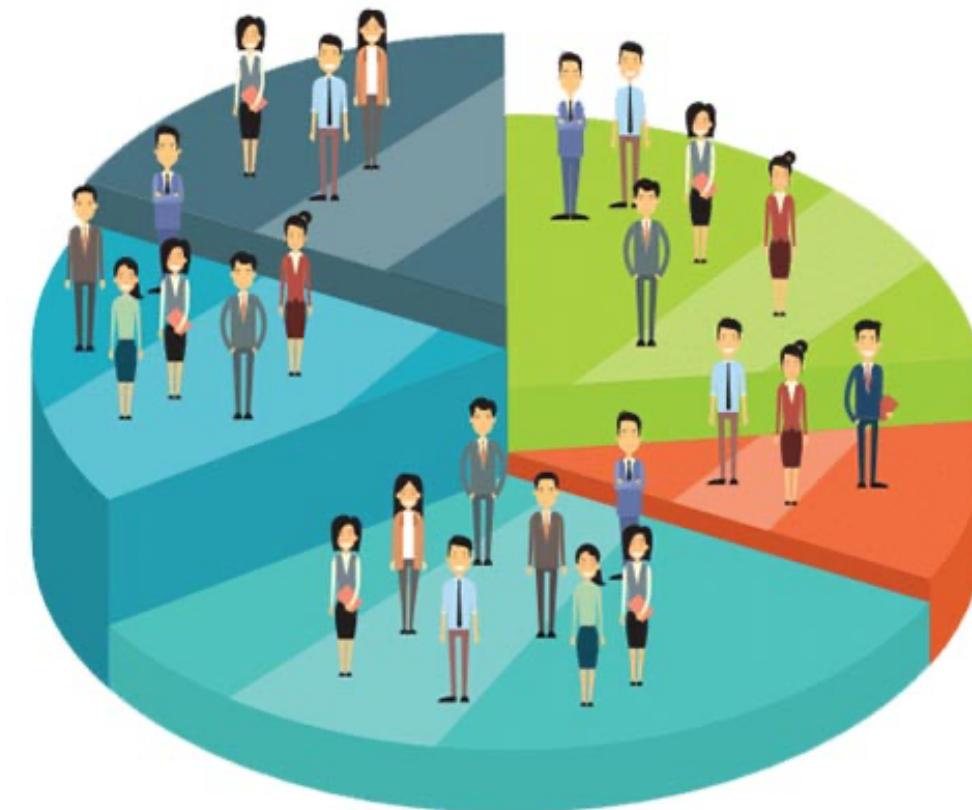
For each customer we calculated the total number of purchases, by counting the total number of invoices related to each customer.

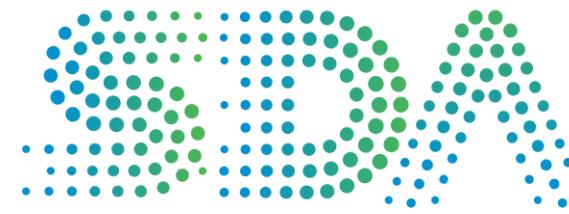


For each customer, we calculated the total amount of money spent by adding together the total price for each item the customer was purchased.



# Customer Segmentation

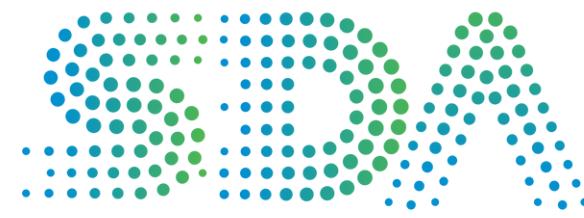




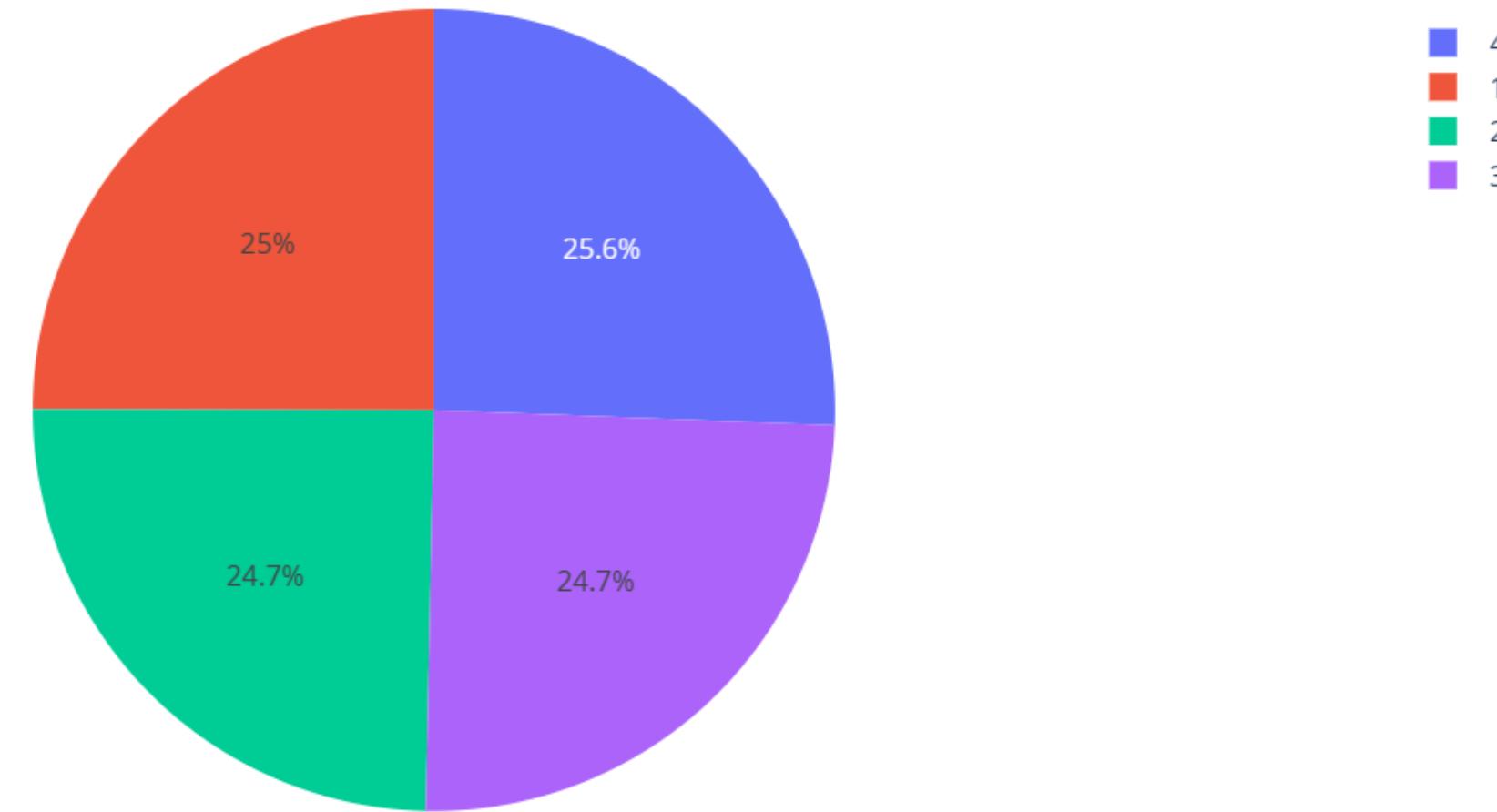
## What Customers Segmentation means?

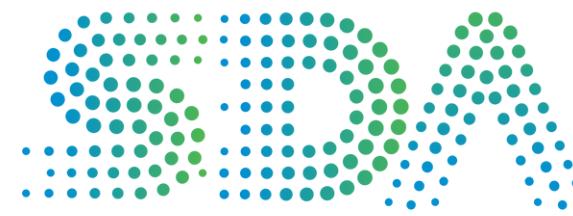
**Segmentation** is a marketing term that refers to aggregating prospective buyers into groups or segments with common needs and who respond similarly to a marketing action.

In our case we could segment customers according to their precise needs.

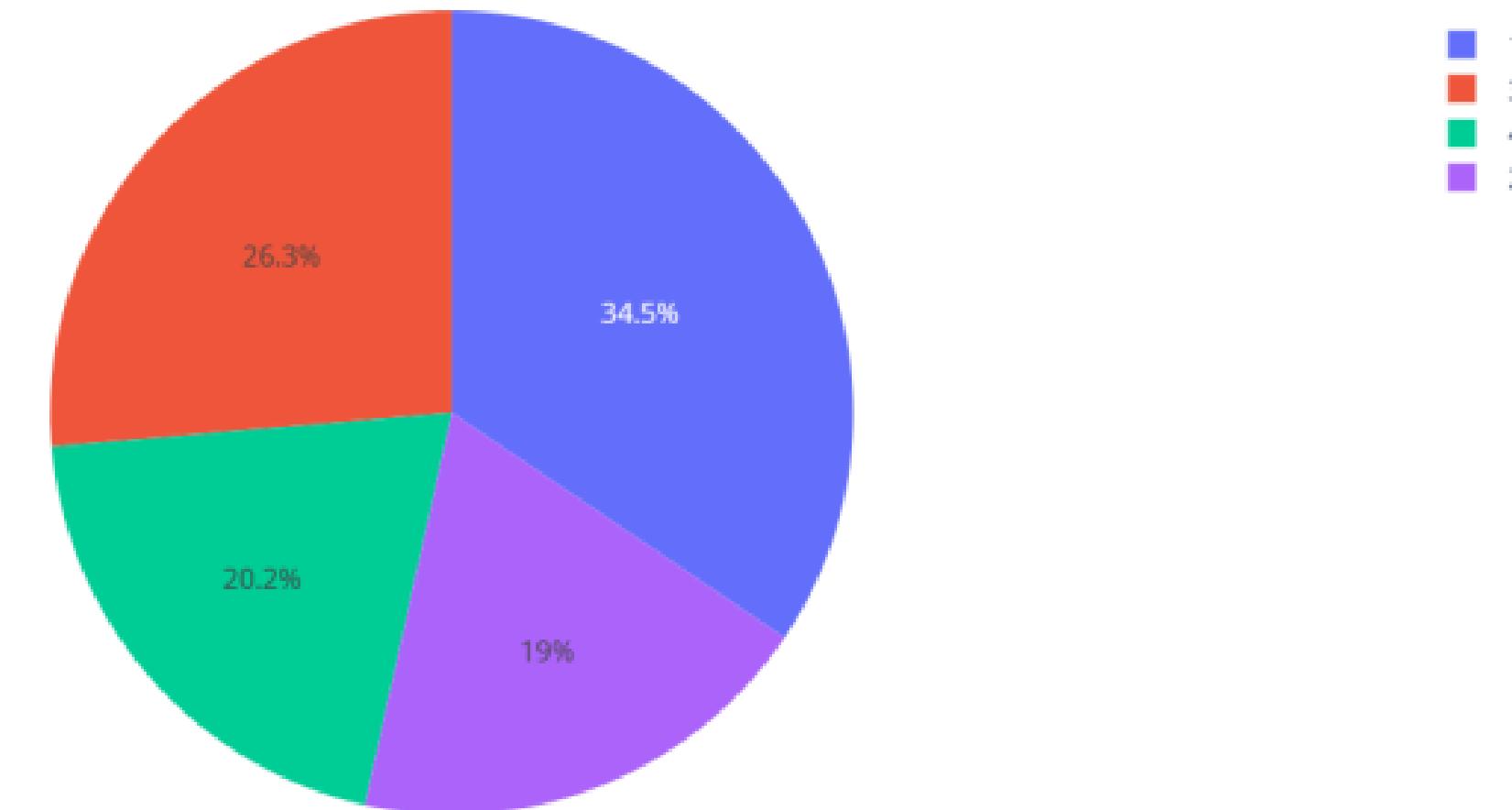


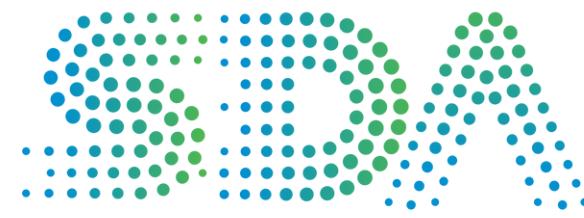
# Recency Score Distribution



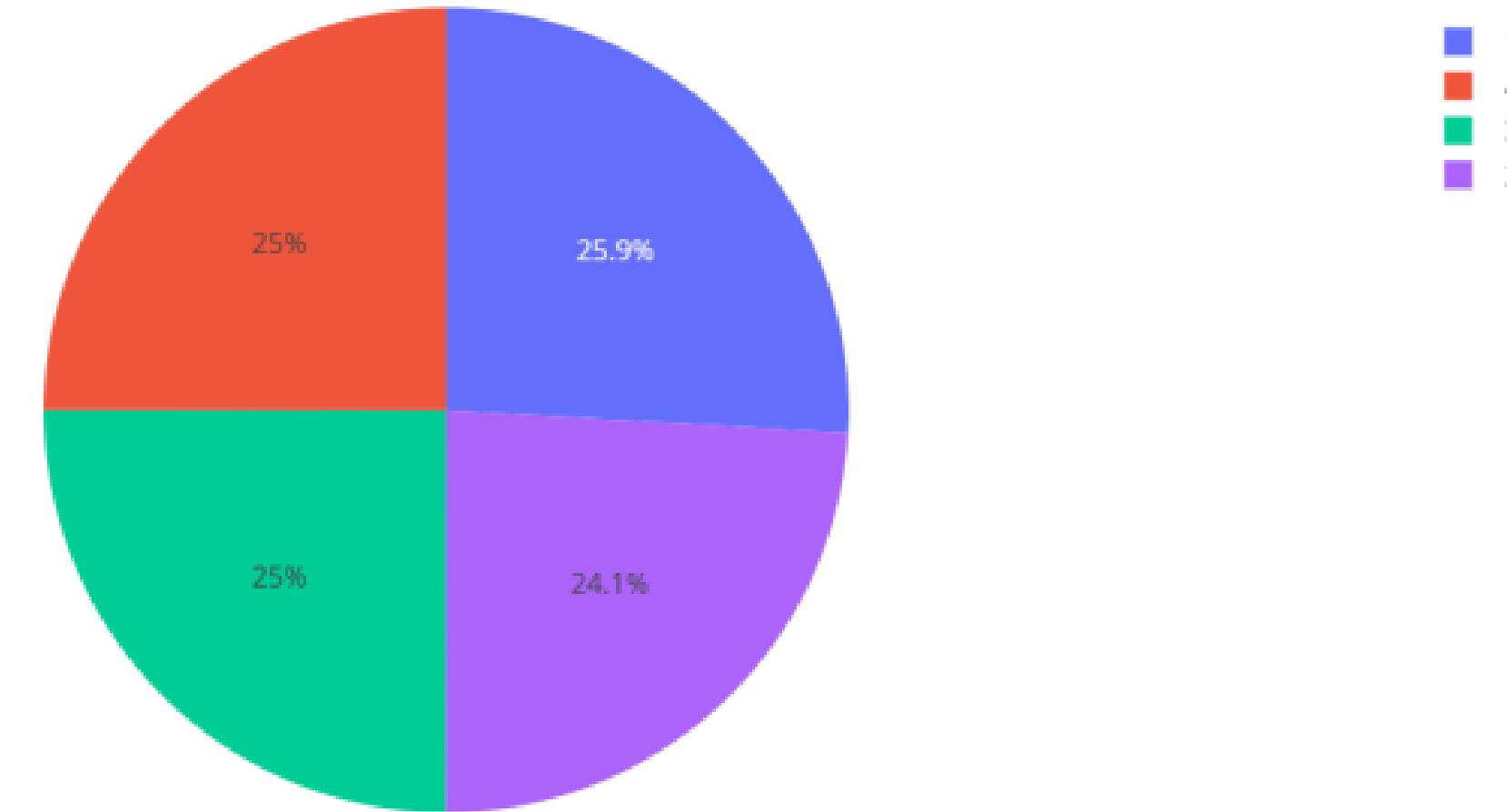


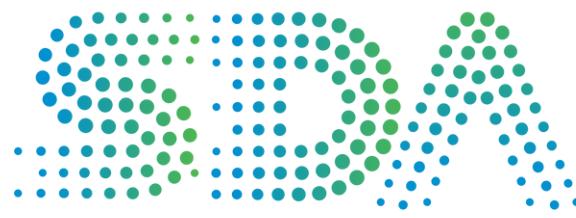
# Frequency Score Distribution



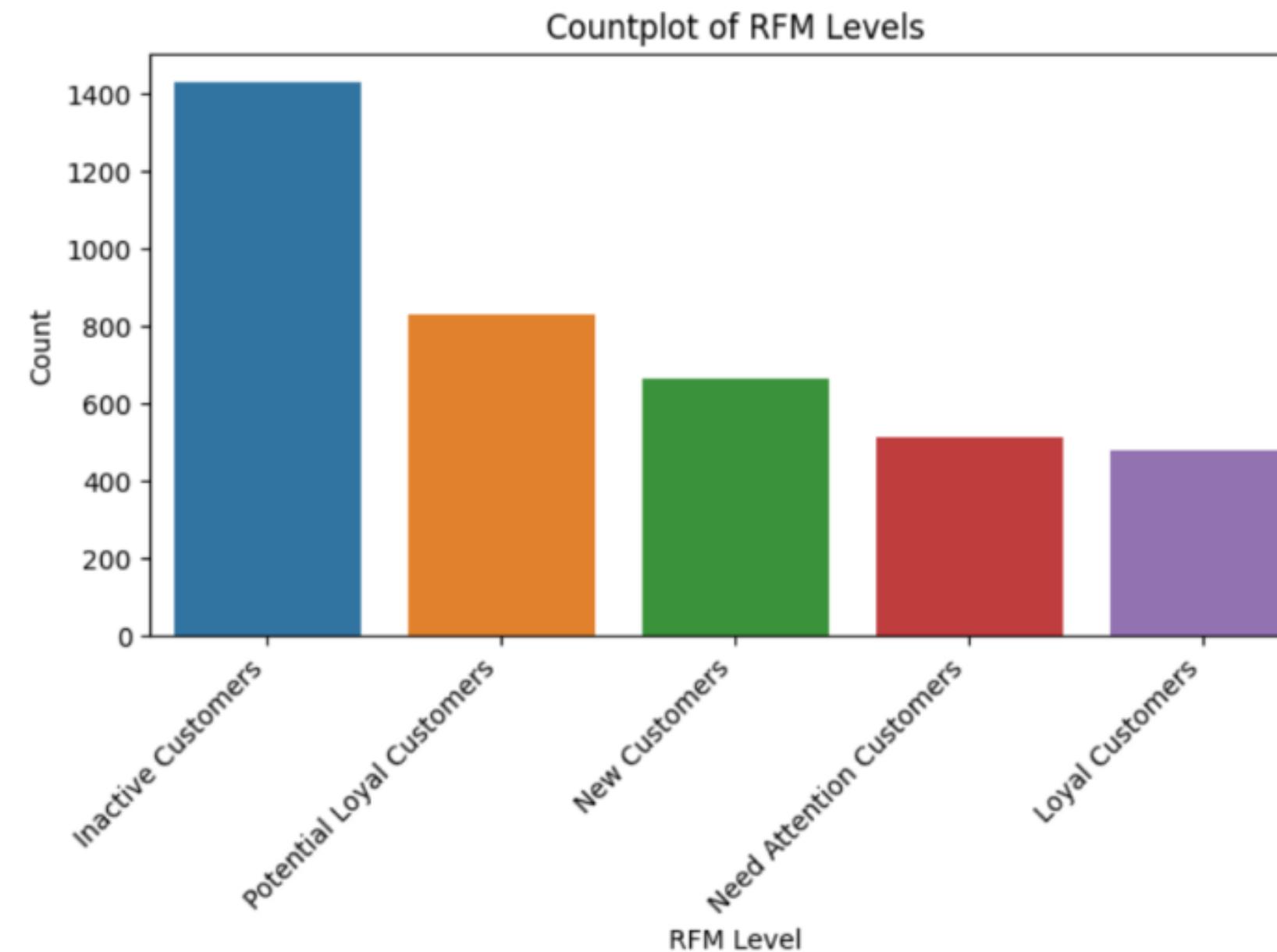


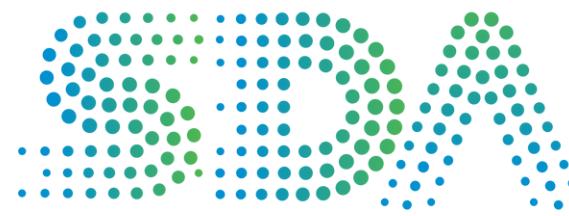
# Monetary Score Distribution



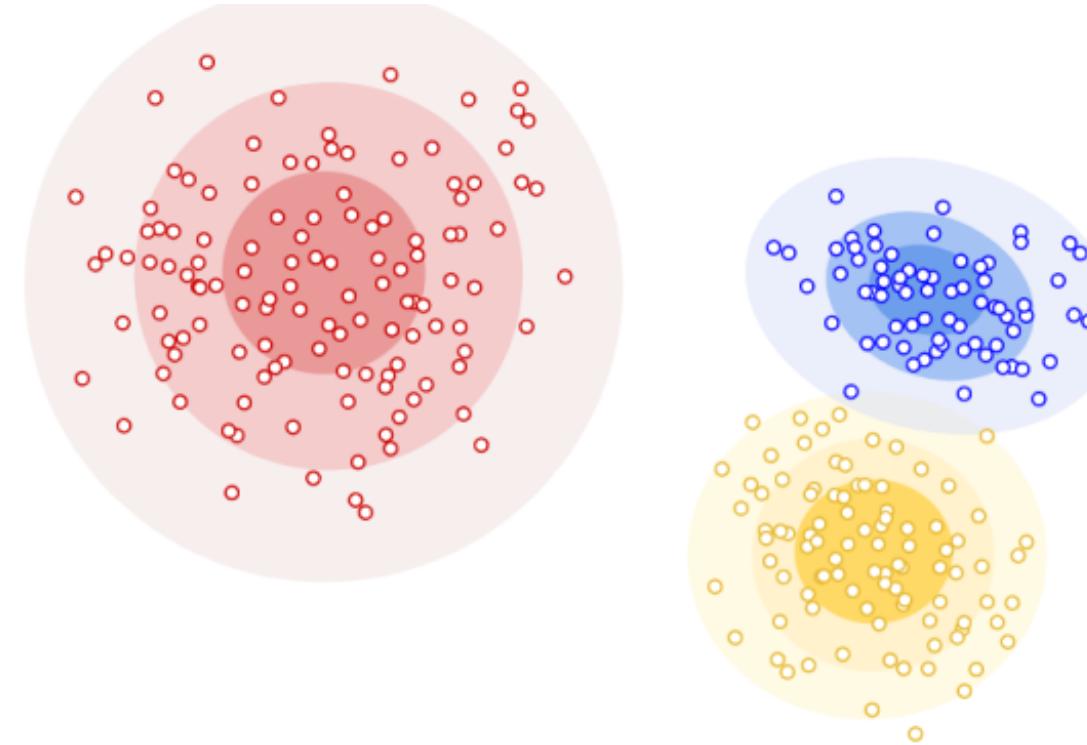
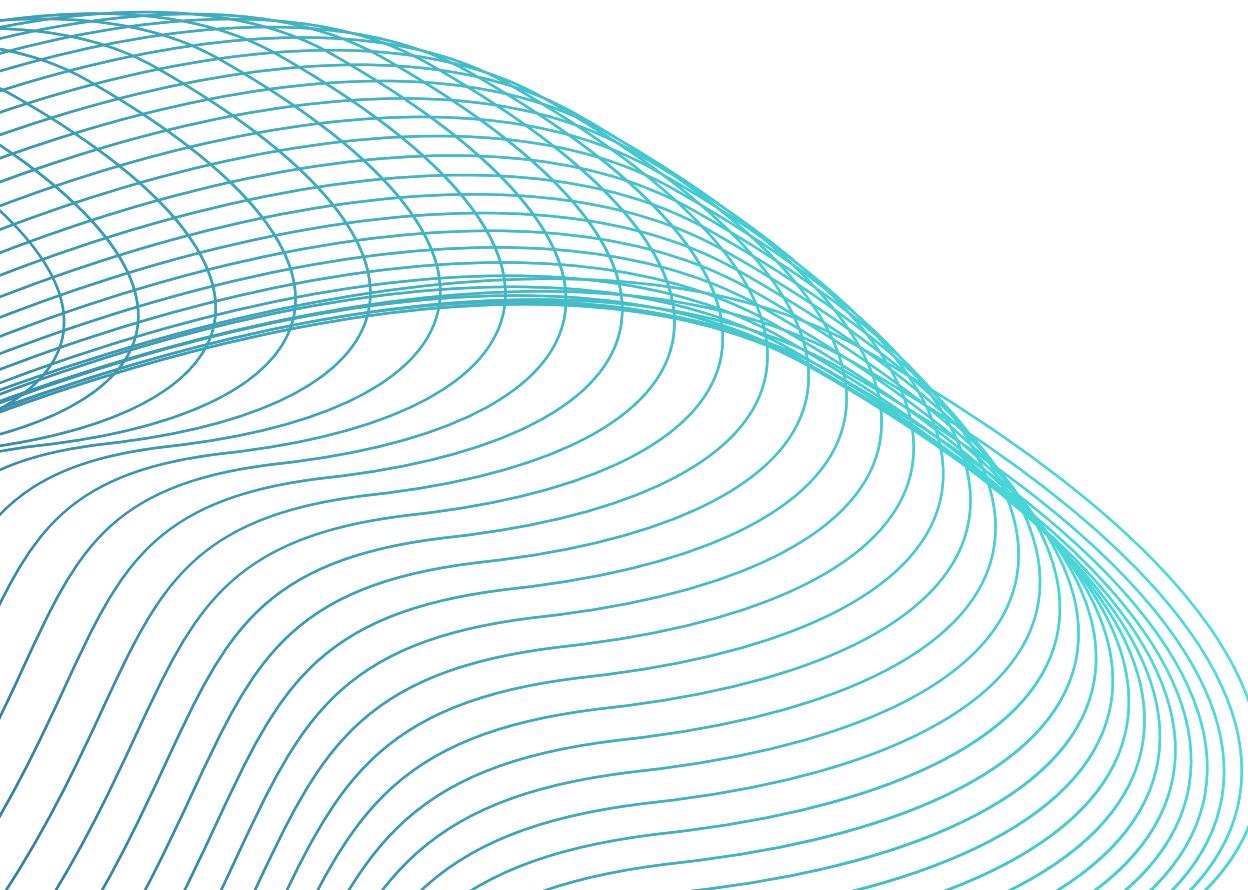


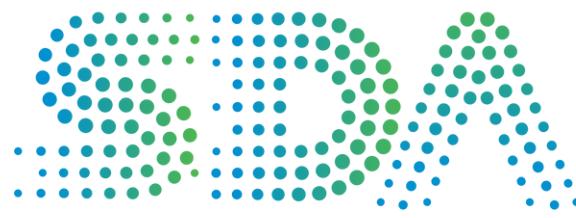
# RFM Segments plot





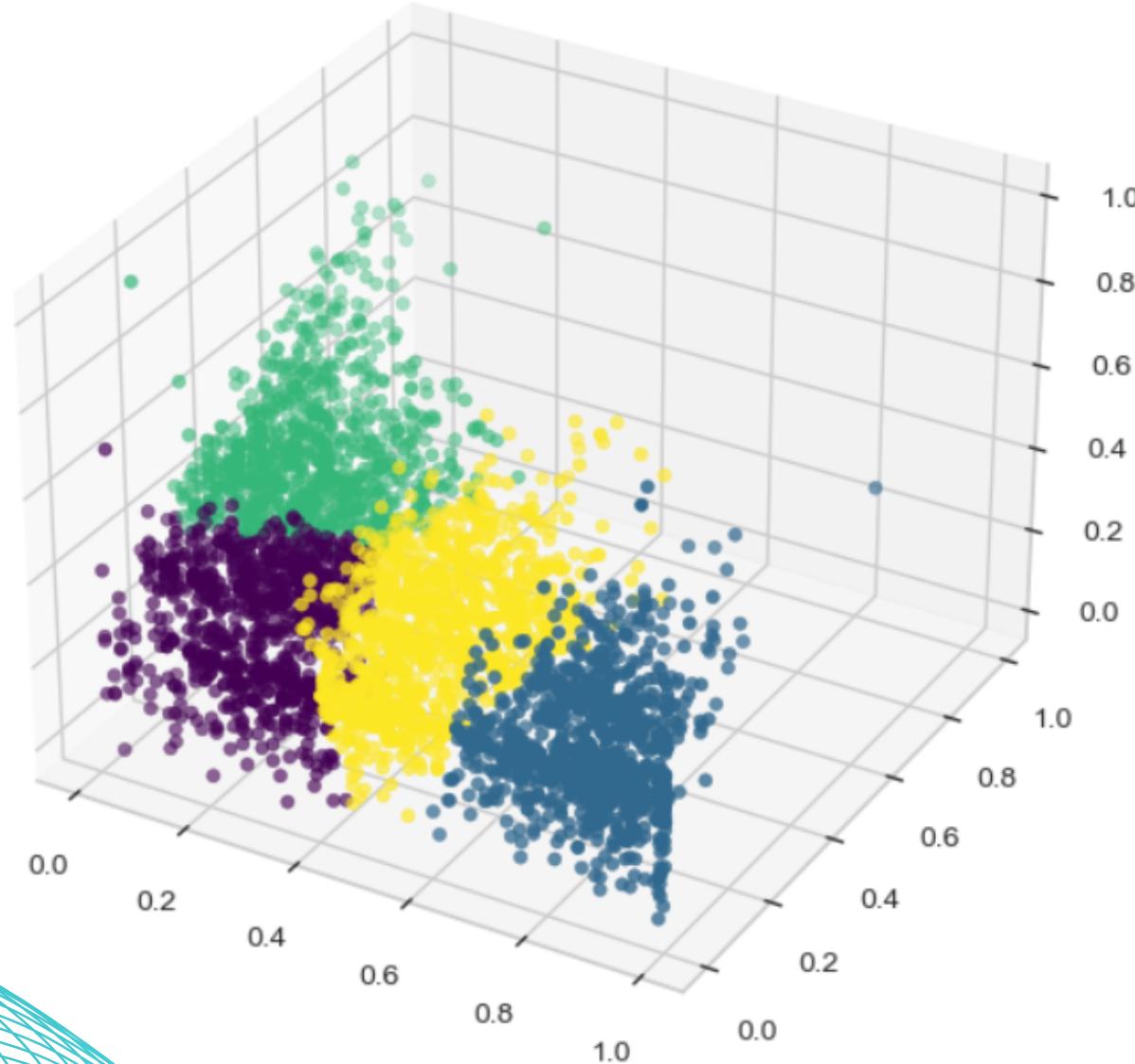
# Applying Clustering



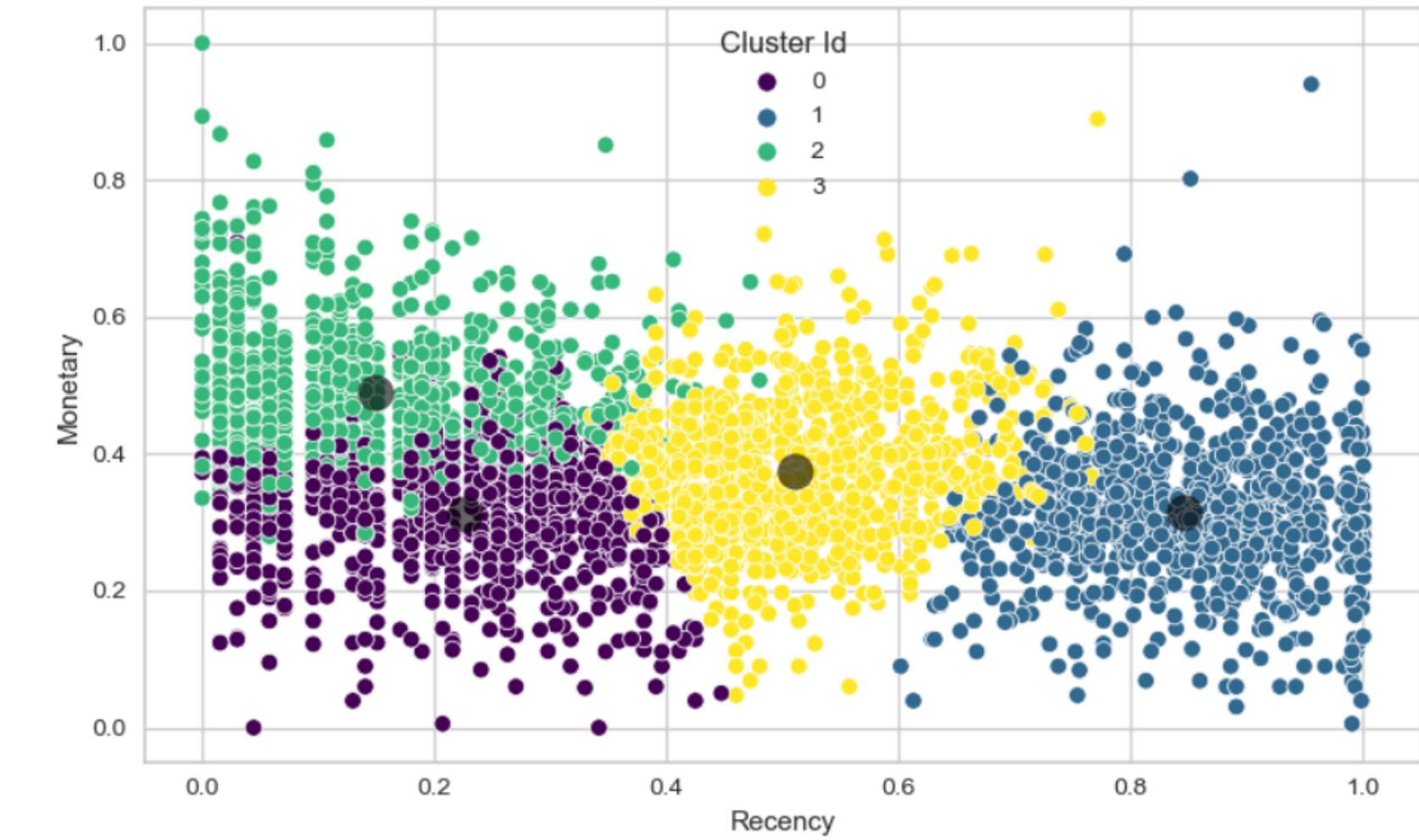


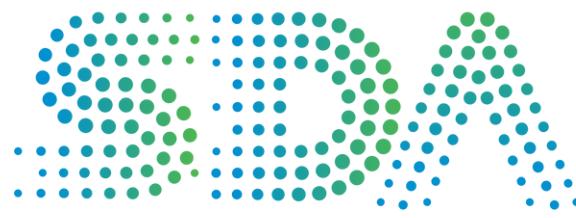
# Clustering Implementation

ClusterID-3

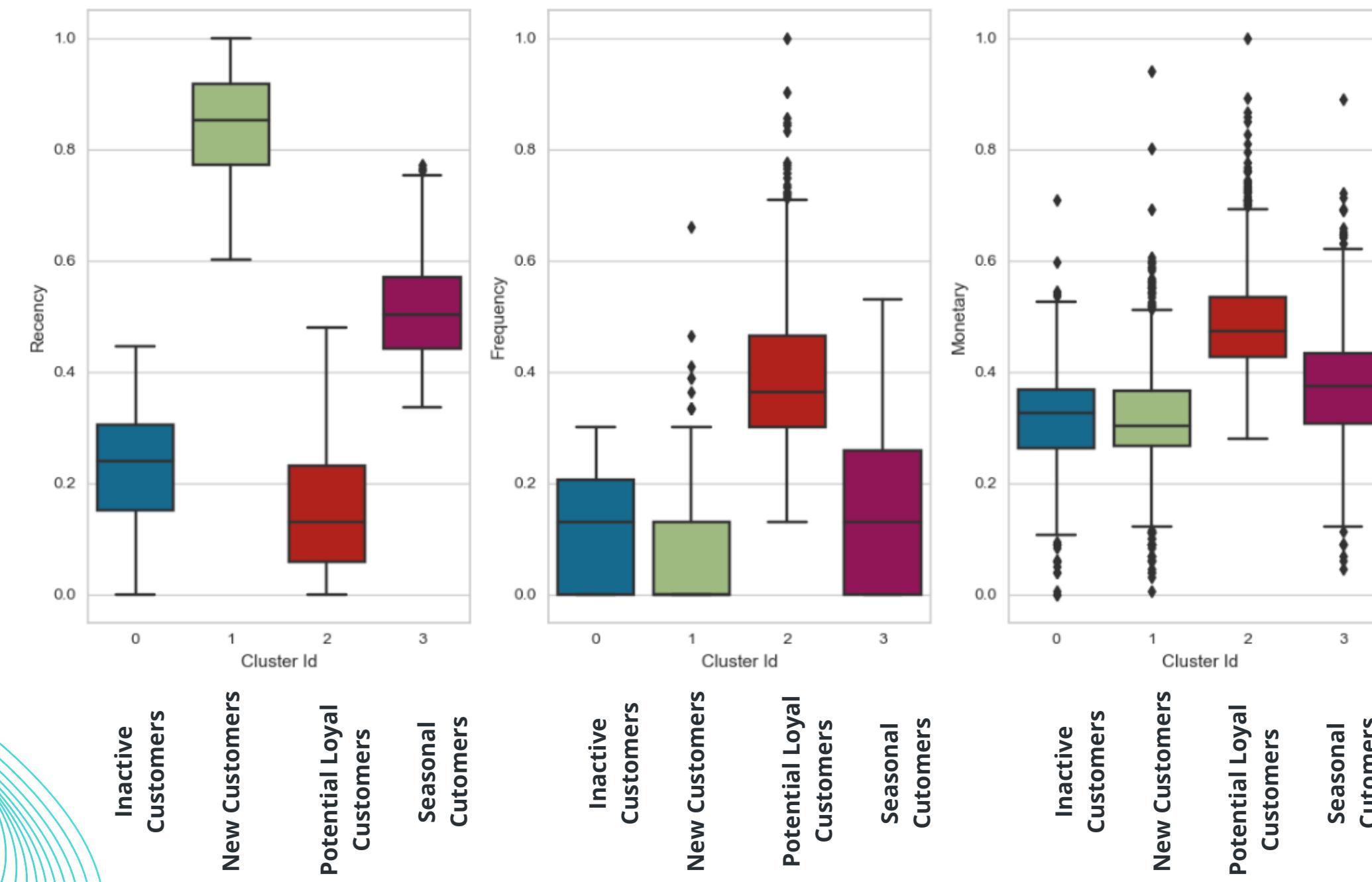


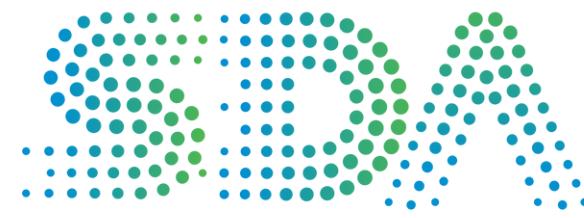
Recency-Monetary  
(4-CLUSTERS)





# Results

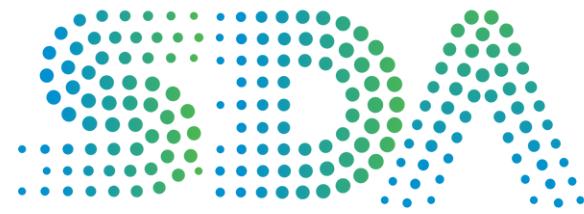




الأكاديمية السعودية الرقمية  
SAUDI DIGITAL ACADEMY



# Create Cohort and Conduct Cohort Analysis

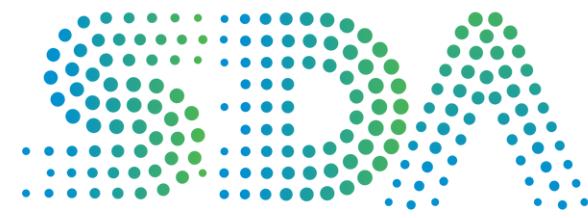


## What's Cohort Analysis?

Cohort analysis is a tool that groups customers into “cohorts” based on shared characteristics, especially the time they started using a product.

## Why we use Cohort Analysis?

We use cohort analysis for tracking and analyzing the behavior of groups of users over time.

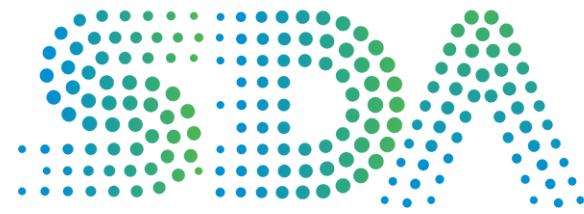


**We will conduct Cohort Analysis by creating three cohorts, which are:**

**1st Cohort: User number & Retention Rate**

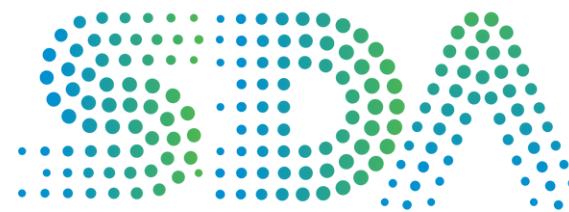
**2nd Cohort: Average Quantity Sold**

**3rd Cohort: Average Sales**



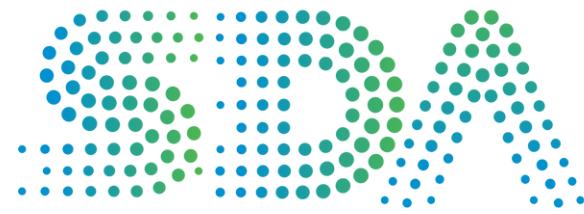
In order to create the three cohorts, we need to do the following steps:

- Month Extraction from InvoiceDate column
- Assigning Cohort to Each Transaction
- Assigning Cohort Index to each transaction



## Dataset after applying the steps :

CustomerID	Country	TotalPrice	date	invoice_date	cohort_date	cohortindex
12910.0	United Kingdom	15.6	2011-09-22	2011-09-01	2011-04-01	6
16942.0	United Kingdom	2.5	2011-06-16	2011-06-01	2011-04-01	3
16145.0	United Kingdom	10.5	2011-02-23	2011-02-01	2010-12-01	3

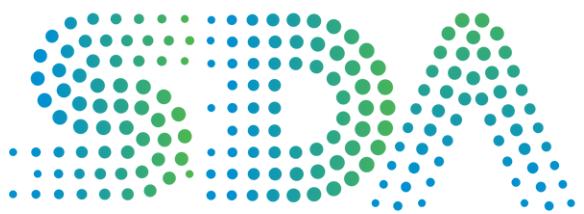


# 1st Cohort: User number & Retention Rate

## Retention Rate:

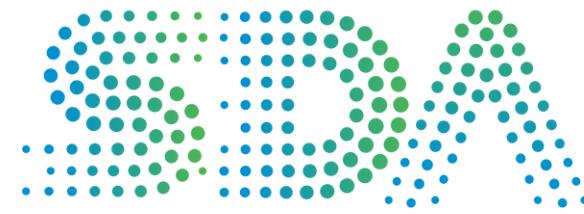
The retention rate gives you the percentage of active customers compared to the total number of customers

Therefore we will count how many customers per each cohort date & cohort index.



# 1st Cohort: User number & Retention Rate

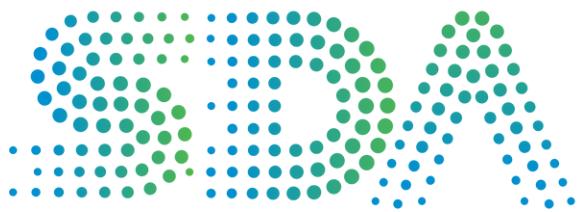




## 2nd Cohort: Average Quantity Sold

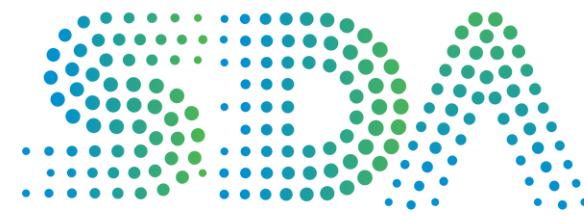
### Average Quantity Sold:

We will calculate the average quantity sold per each cohort date & cohort index.



# 2nd Cohort: Average Quantity Sold

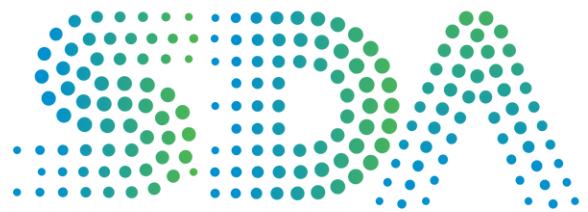




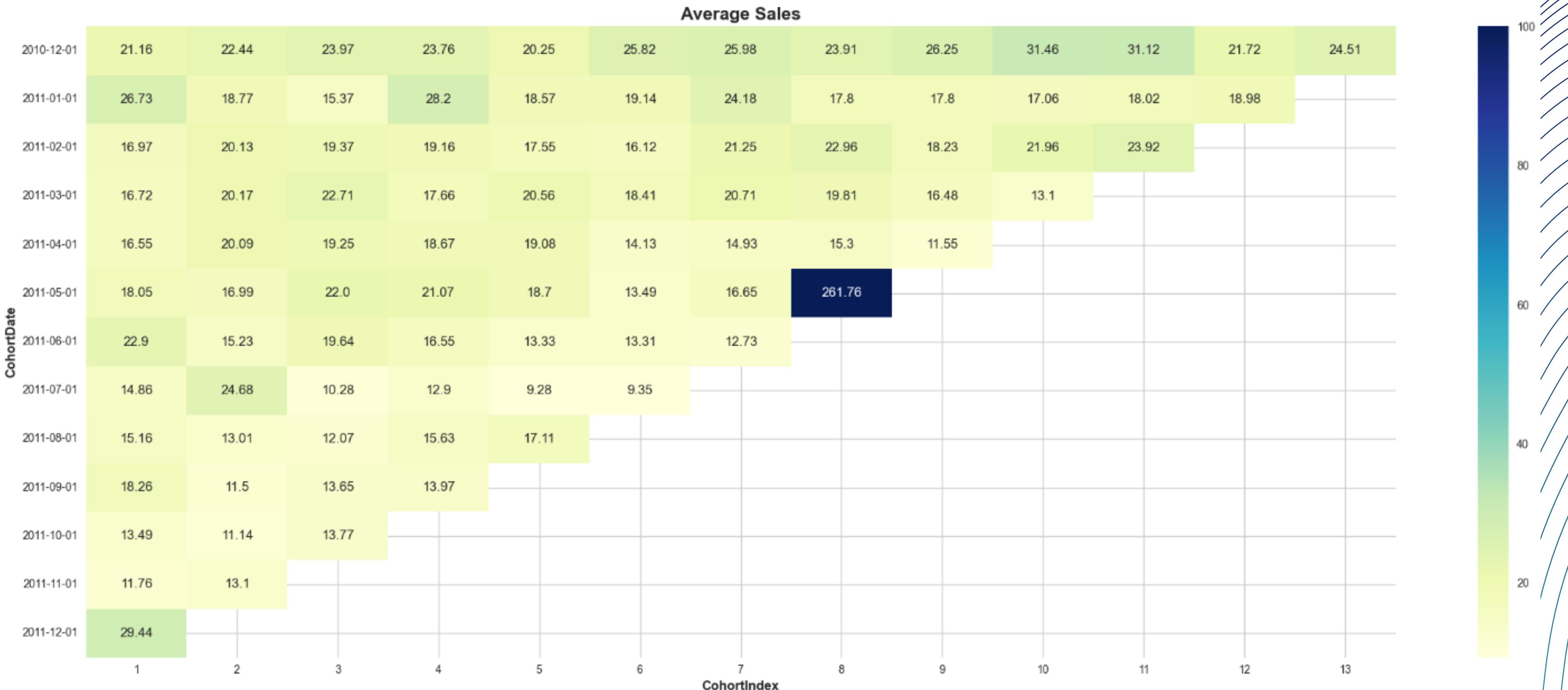
## 3rd Cohort: Average Sales

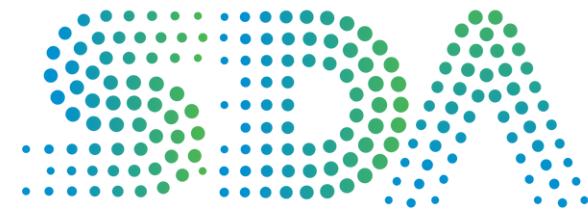
### Average Sales

We will calculate the average Total Price per each cohort date & cohort index.



## 3rd Cohort: Average Sales



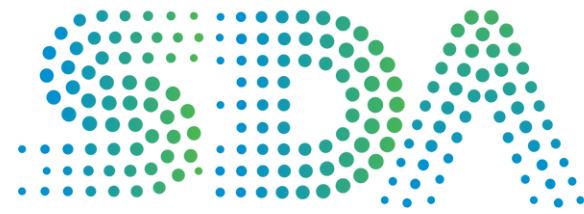


الأكاديمية السعودية الرقمية  
SAUDI DIGITAL ACADEMY

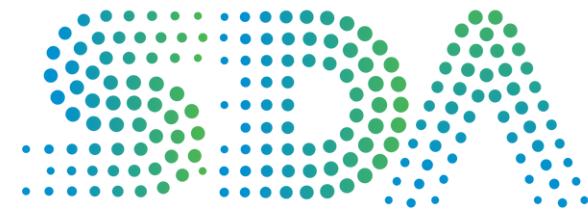


# Cohort Analysis Results

A large, abstract graphic element in the bottom right corner consists of numerous thin, light blue lines that curve upwards and outwards, creating a sense of motion and depth.



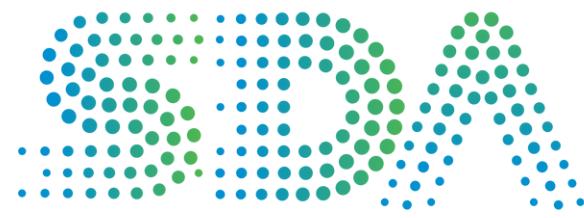
The cohorts analysis shows that people tend to buy more during the end-of-year season. If businesses communicate well and build relationships, customers who shop during specific seasons might become regular buyers. It's not just about getting new customers; keeping the ones you already have is just as important.



الأكاديمية السعودية الرقمية  
SAUDI DIGITAL ACADEMY



# Conclusion



**The project investigated customer behavior using RFM analysis, K-means clustering, and cohorts analysis. We found that the outcomes of K-means and RFM segmentation are similar.**

**The acquired insights offer practical strategies for focused marketing and enhanced customer engagement.**

The background features a minimalist design with abstract, flowing blue lines. These lines are thin and light blue, creating a sense of motion and depth. They form a large, sweeping wave that starts from the top left, moves across the upper portion of the frame, and then tapers off towards the bottom right. Another set of similar lines is visible on the far right side of the image.

**Thanks for Listening**