



الأكاديمية السعودية الرقمية  
SAUDI DIGITAL ACADEMY

# CUSTOMER SEGEMENTATION





# **TABLE OF CONTENT**

- 1- Introduction**
- 2- Cleaning Dataset**
- 3- Exploratory Data Analysis**
- 4- RFM Analysis**
- 5- K-means Clustering**
- 6- Cohort Analysis**
- 7- Conclusion**

# ONLINE RETAIL INTRODUCTION

**The dataset we are working on is a transnational data set which the company mainly sells unique all-occasion gifts. Many customers of the company are whole salers.**

---

**The transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.**

# ONLINE RETAIL INTRODUCTION

**we looked at the data  
and we did exploratory data analysis  
and cleaning data set then preprocess it  
and used machine learning**



# ONLINE RETAIL INTRODUCTION

Variable Name	Role	Type	Demographic	Description
InvoiceNo	ID	Categorical		a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation
StockCode	ID	Categorical		a 5-digit integral number uniquely assigned to each distinct product
Description	Feature	Categorical		product name
Quantity	Feature	Integer		the quantities of each product (item) per transaction
InvoiceDate	Feature	Date		the day and time when each transaction was generated
UnitPrice	Feature	Continuous		product price per unit
CustomerID	Feature	Categorical		a 5-digit integral number uniquely assigned to each customer
Country	Feature	Categorical		the name of the country where each customer resides



# DATA CLEANING

**Missing values:** Description: 1454  
CustomerID: 135080

**Duplicated values:** 5268

We dropped the duplicated and  
missing values

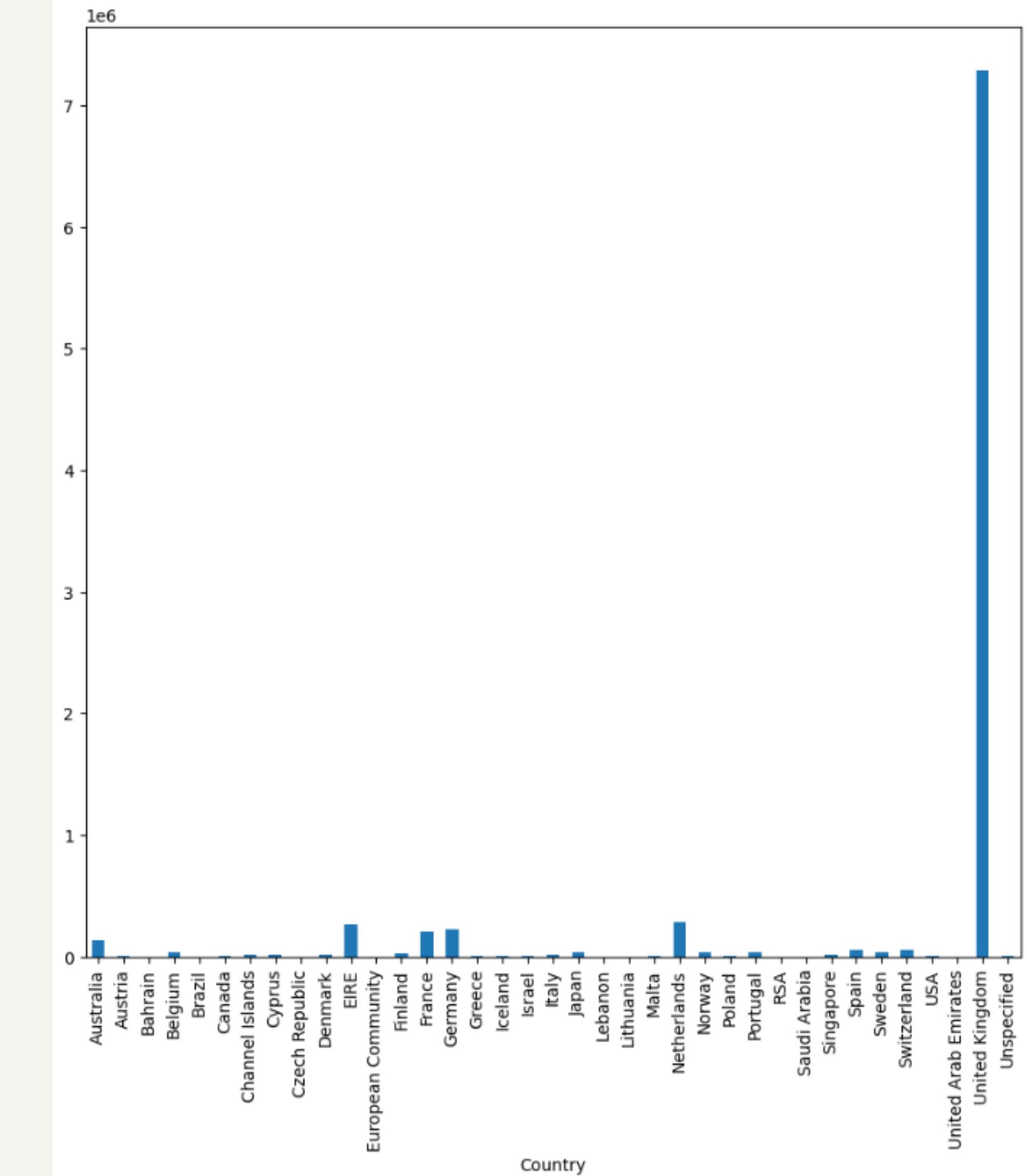




# EXPLORATORY DATA ANALYSIS

**Number of customer per country**

**UK is dominant**

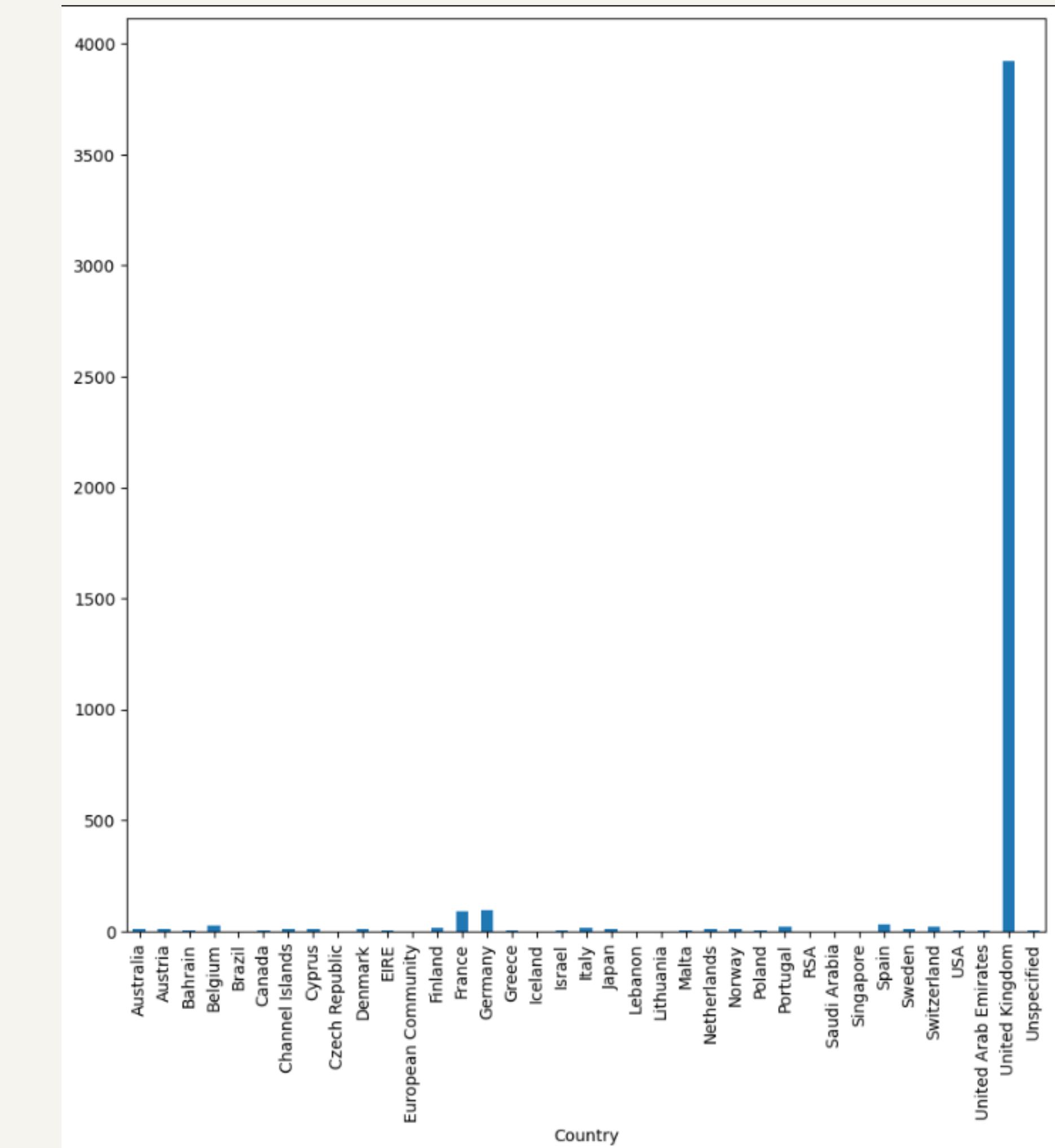




# EXPLORATORY DATA ANALYSIS

Total “revenu” per country

UK is dominant

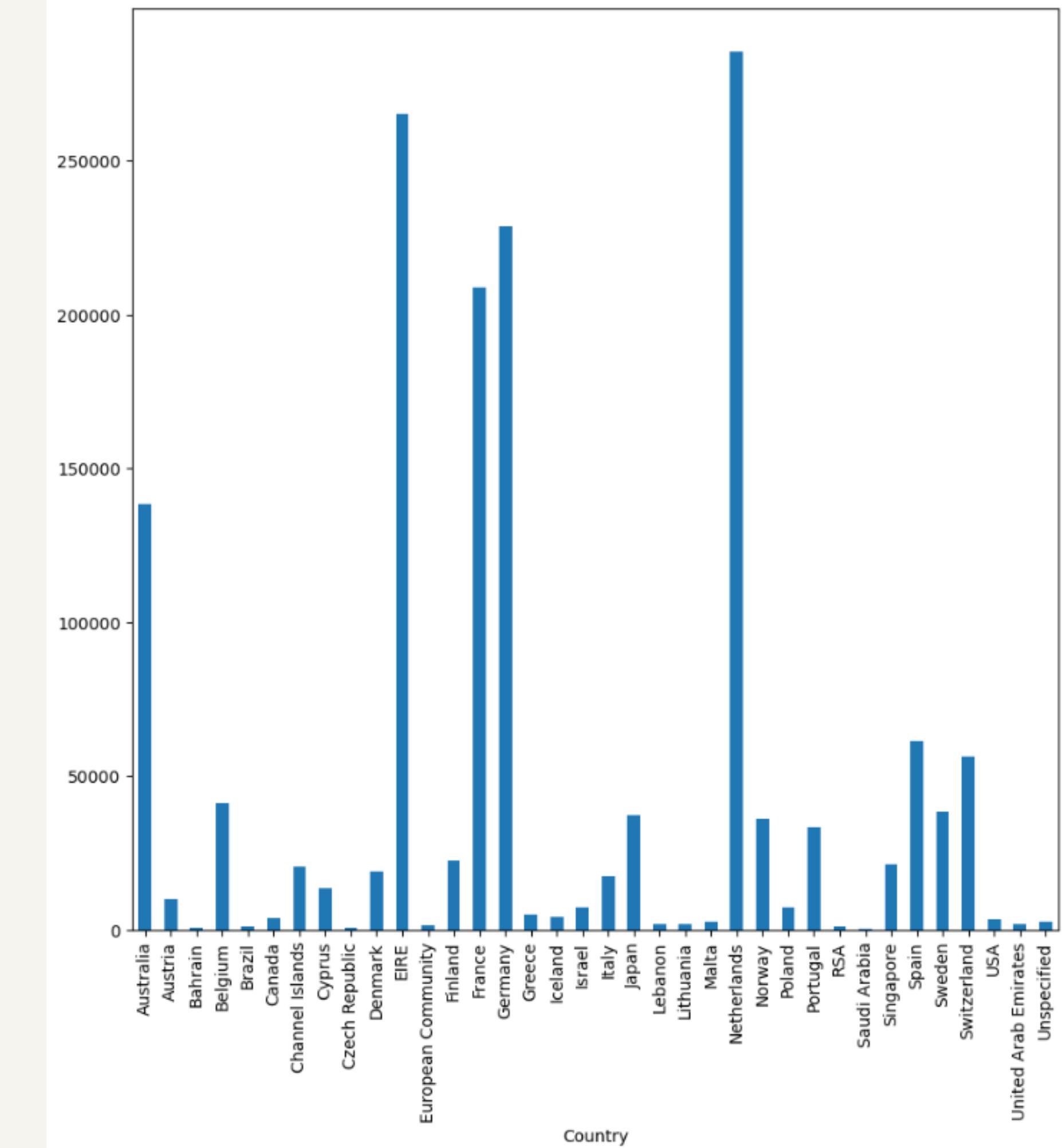




# EXPLORATORY DATA ANALYSIS

Total “revenu” per country

Without UK



# EXPLORATORY DATA ANALYSIS



The **UK** not only has the most sales revenue, but also the most customers.

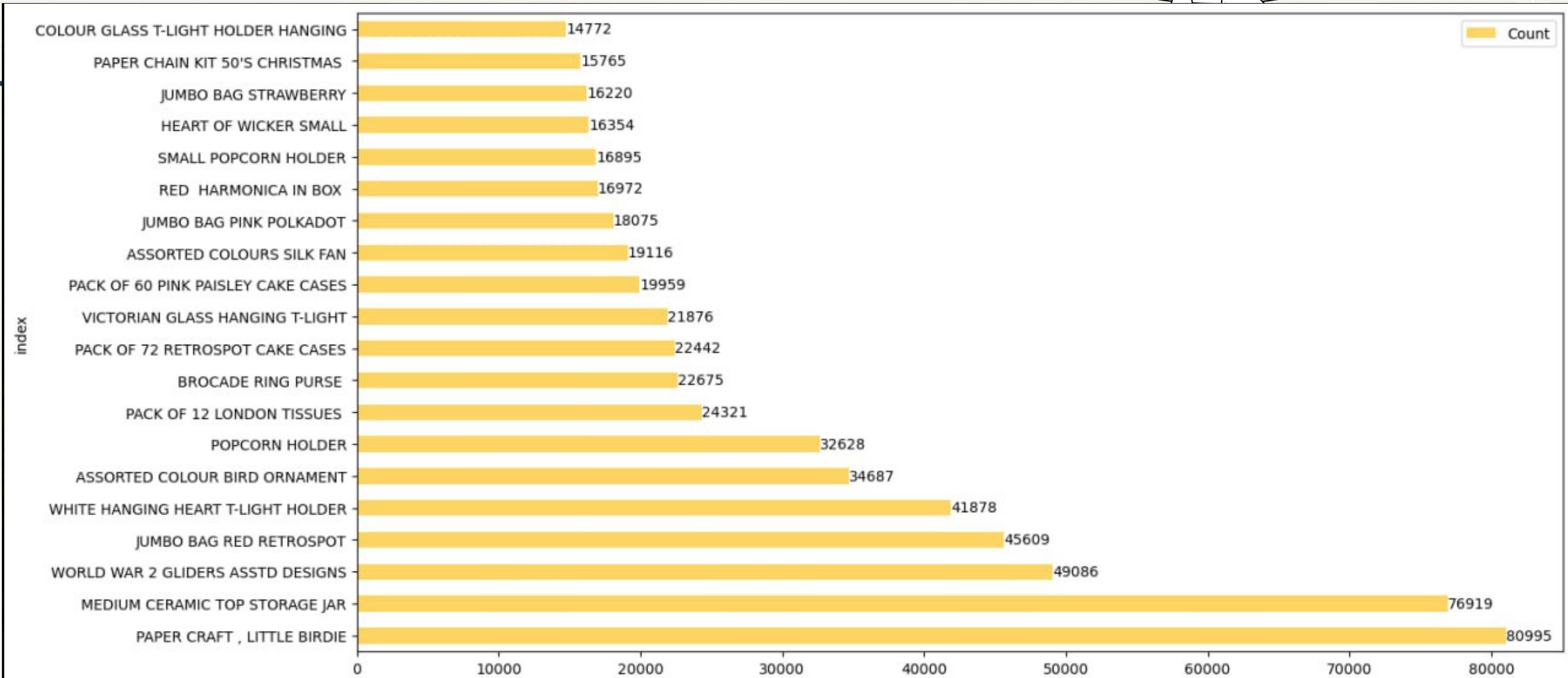
Since the majority of this data set contains orders from the UK, we can explore the UK market further

---

# EXPLORATORY DATA ANALYSIS

most popular products that are bought in the UK

With Quantity

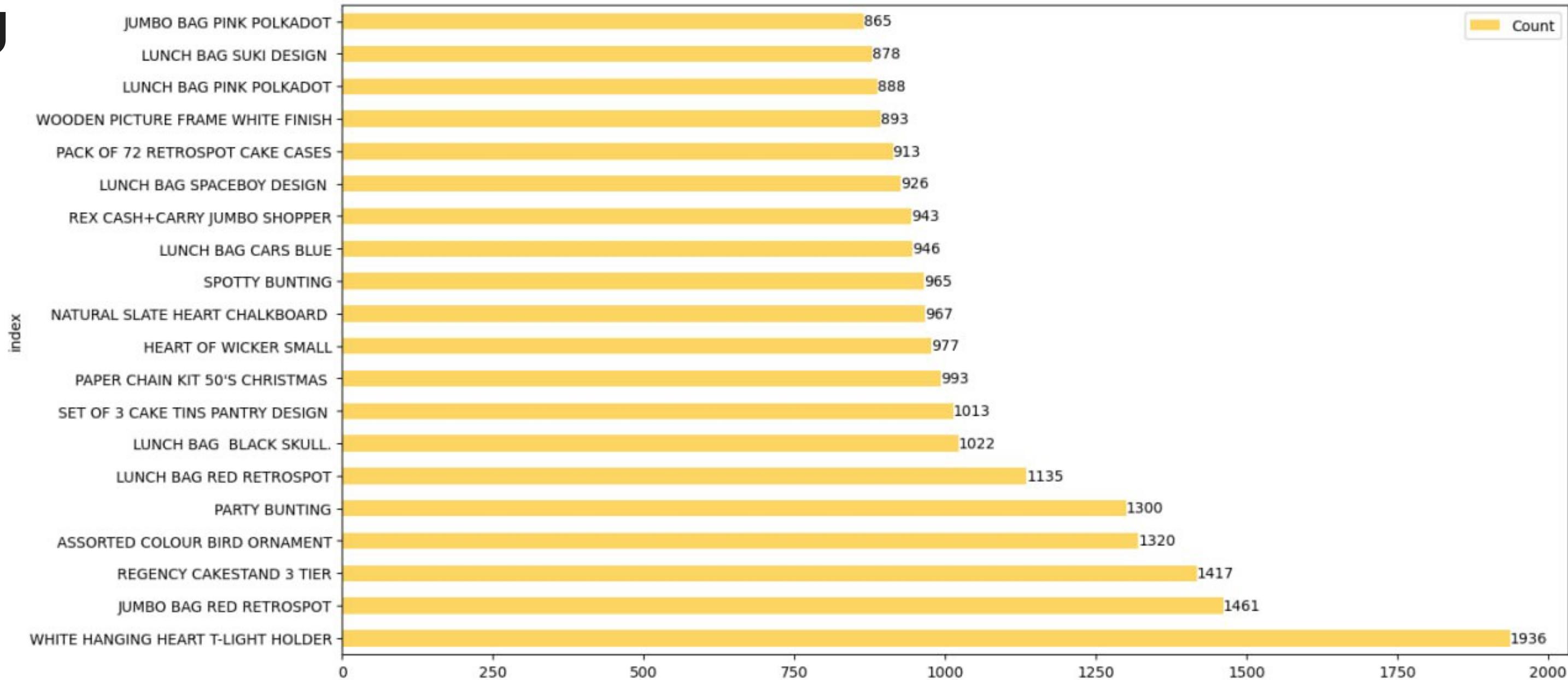


# EXPLORATORY DATA ANALYSIS

most popular products that are bought in the UK



## Without Quantity



# RFM ANALYSIS

**RFM (Recency, Frequency, Monetary) Analysis is a customer segmentation technique for analyzing customer value based on past buying behavior.**

**RECENCY (R): Time since last purchase**

**FREQUENCY (F): Total number of purchases**

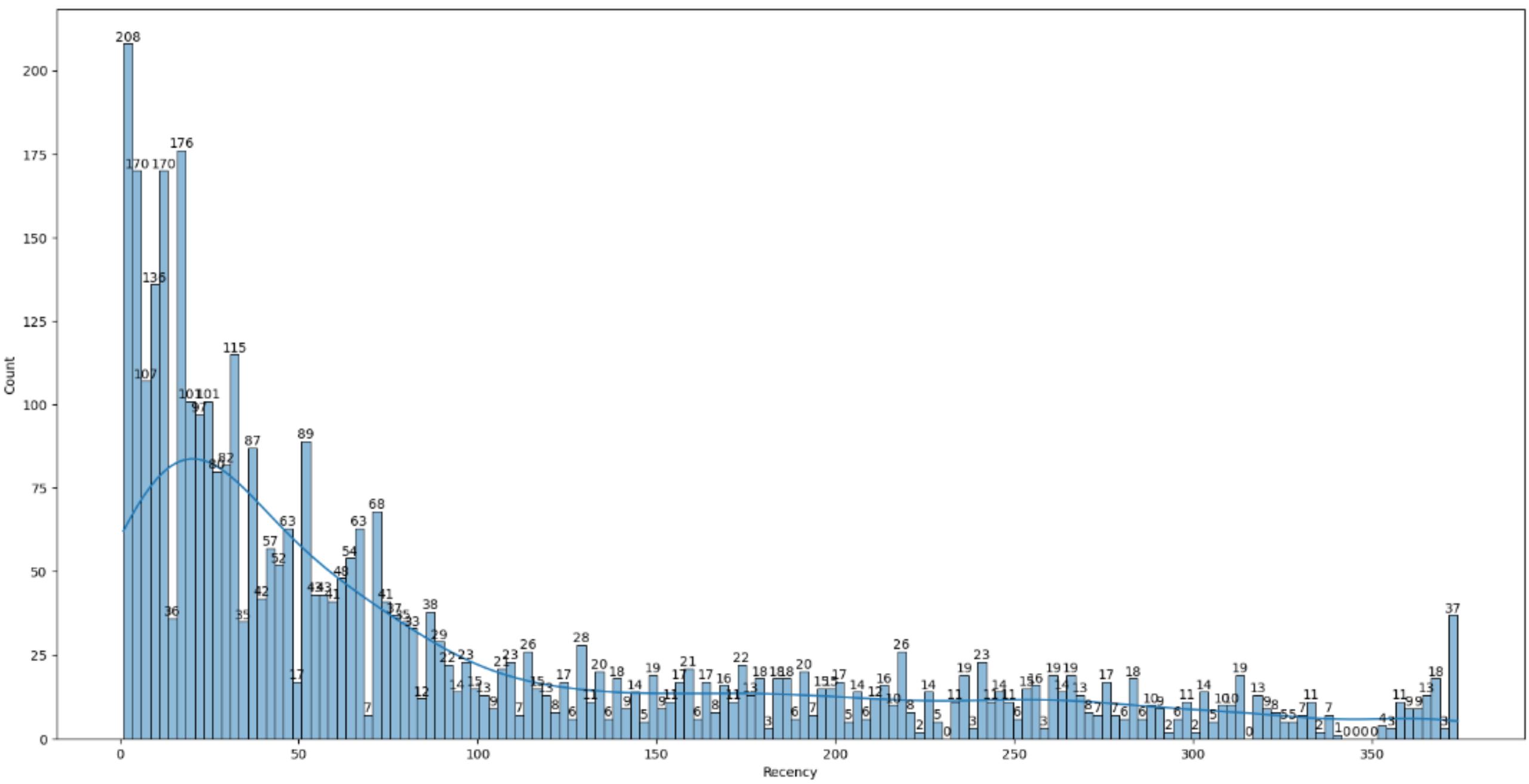
**MONETARY (M): Total amount of money spent**





# RFM ANALYSIS

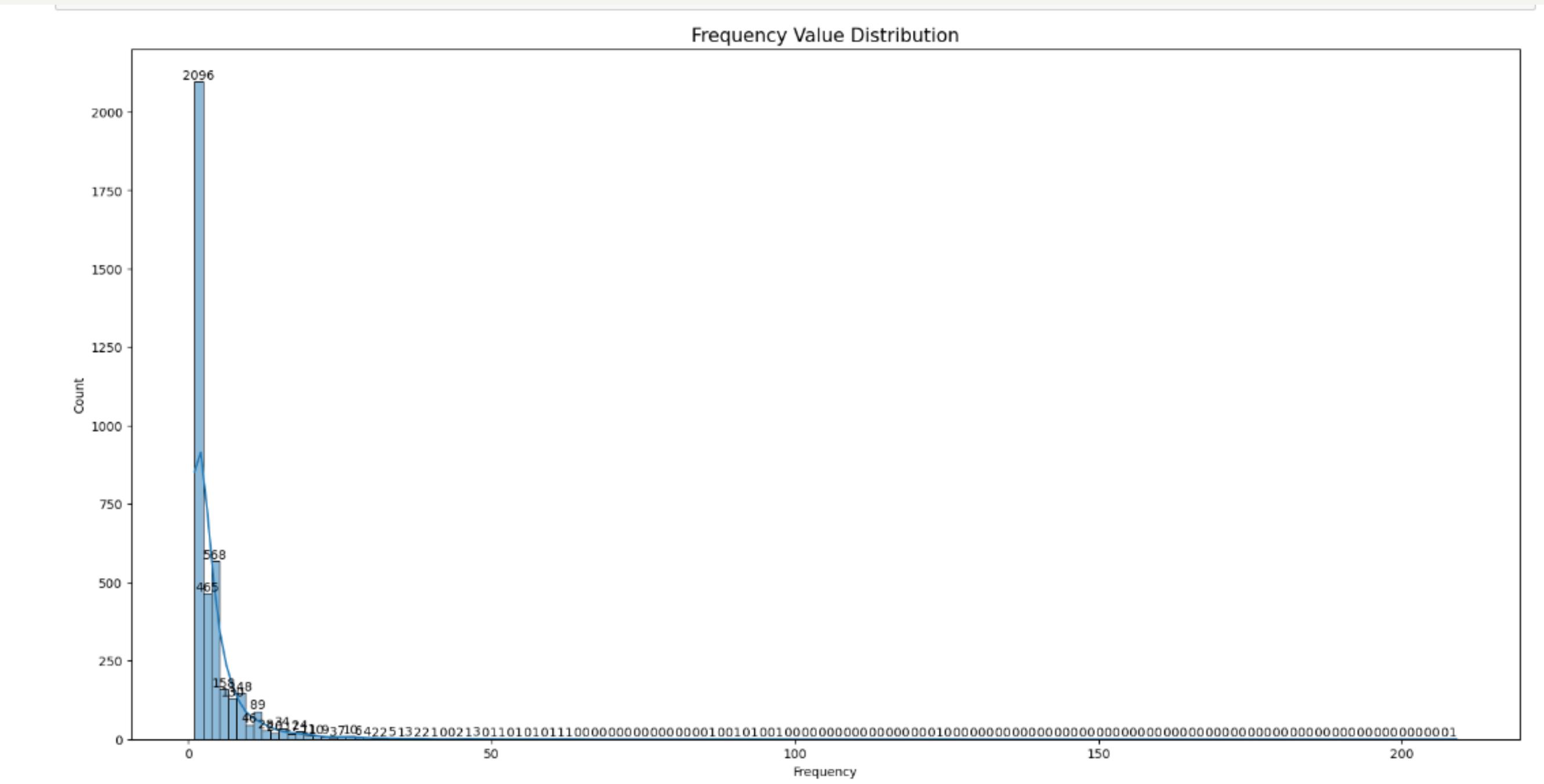
## Plot the Recency





# RFM ANALYSIS

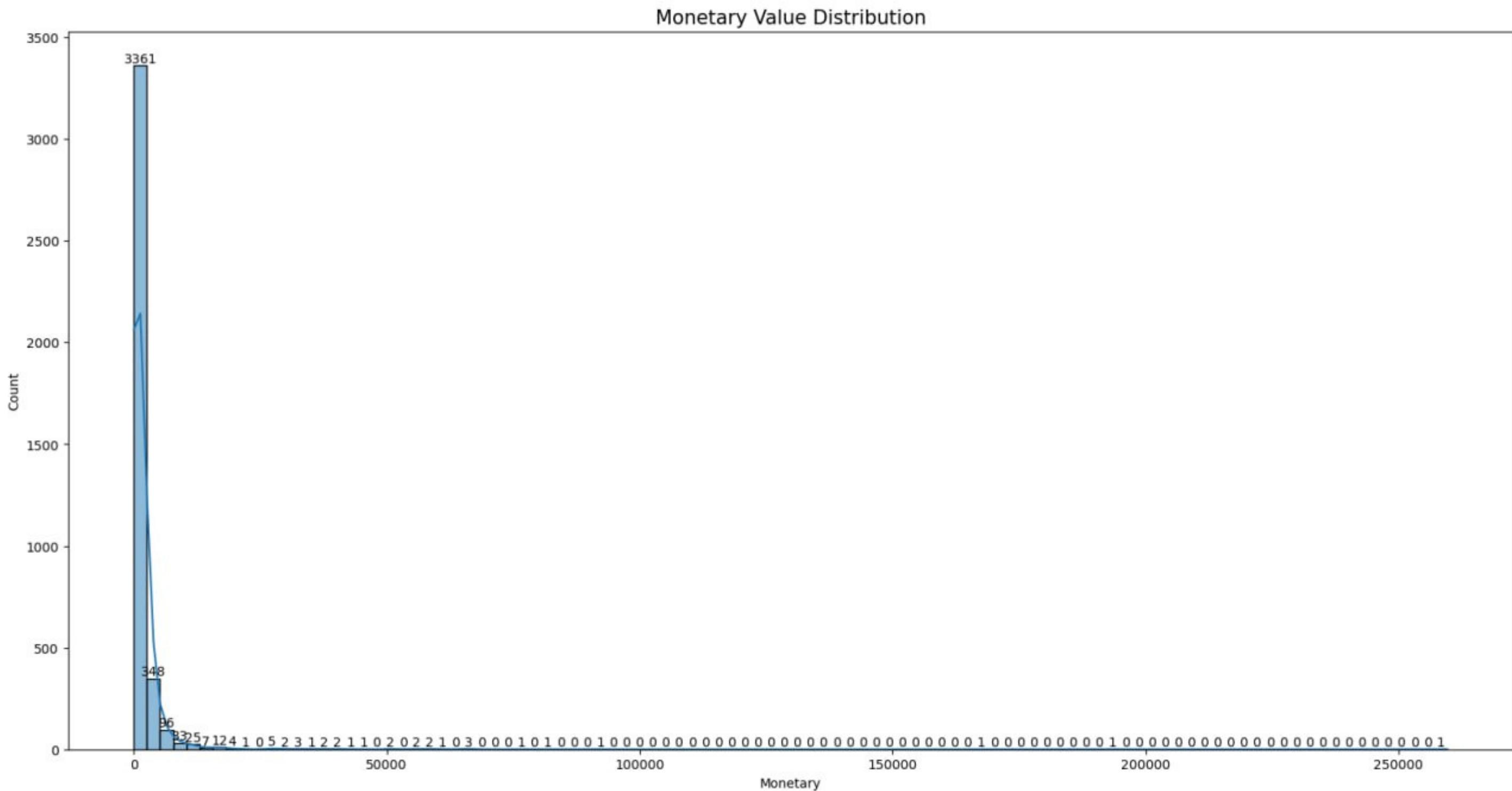
## Plot the Frequency





# RFM ANALYSIS

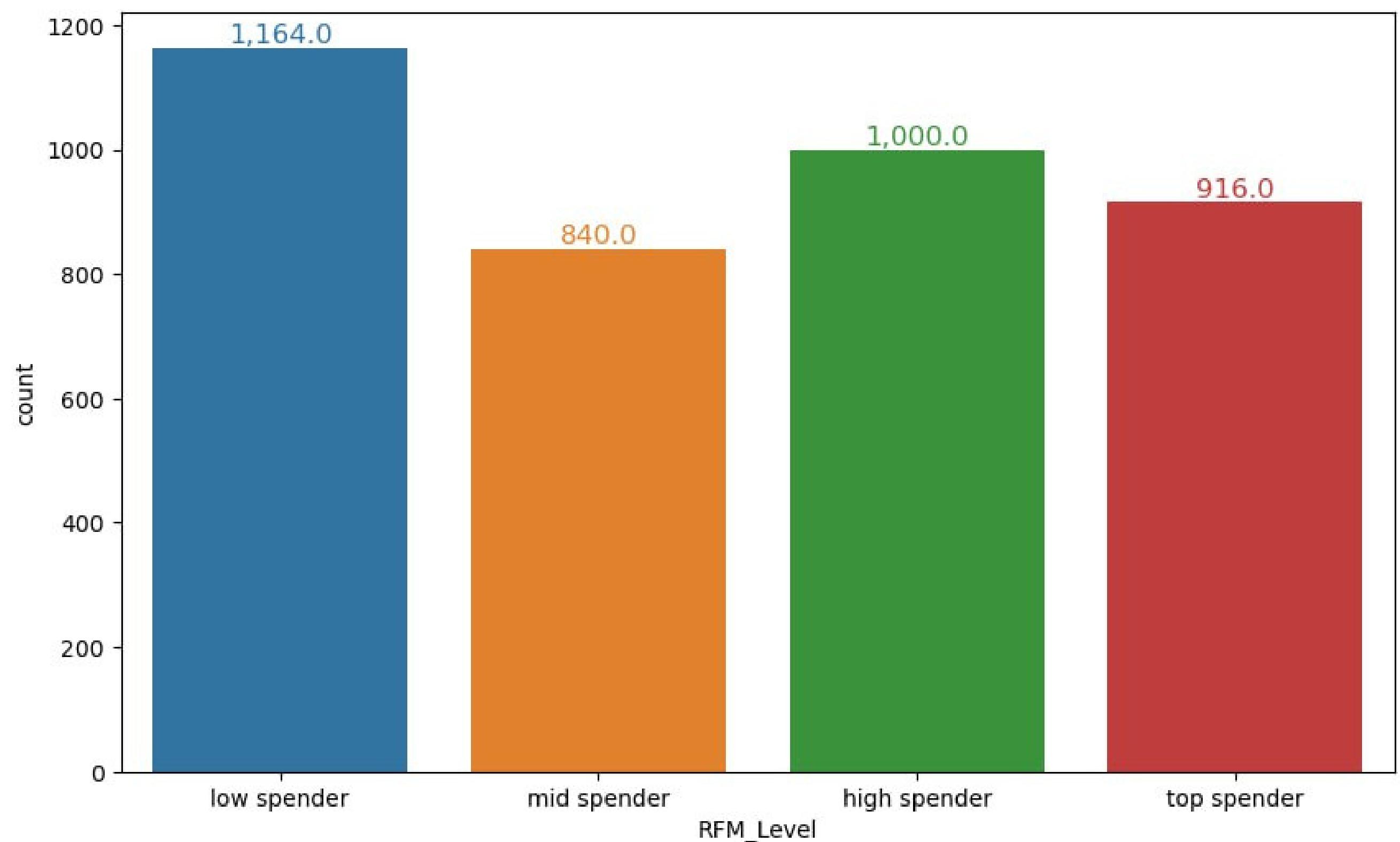
## Plot the Momentry





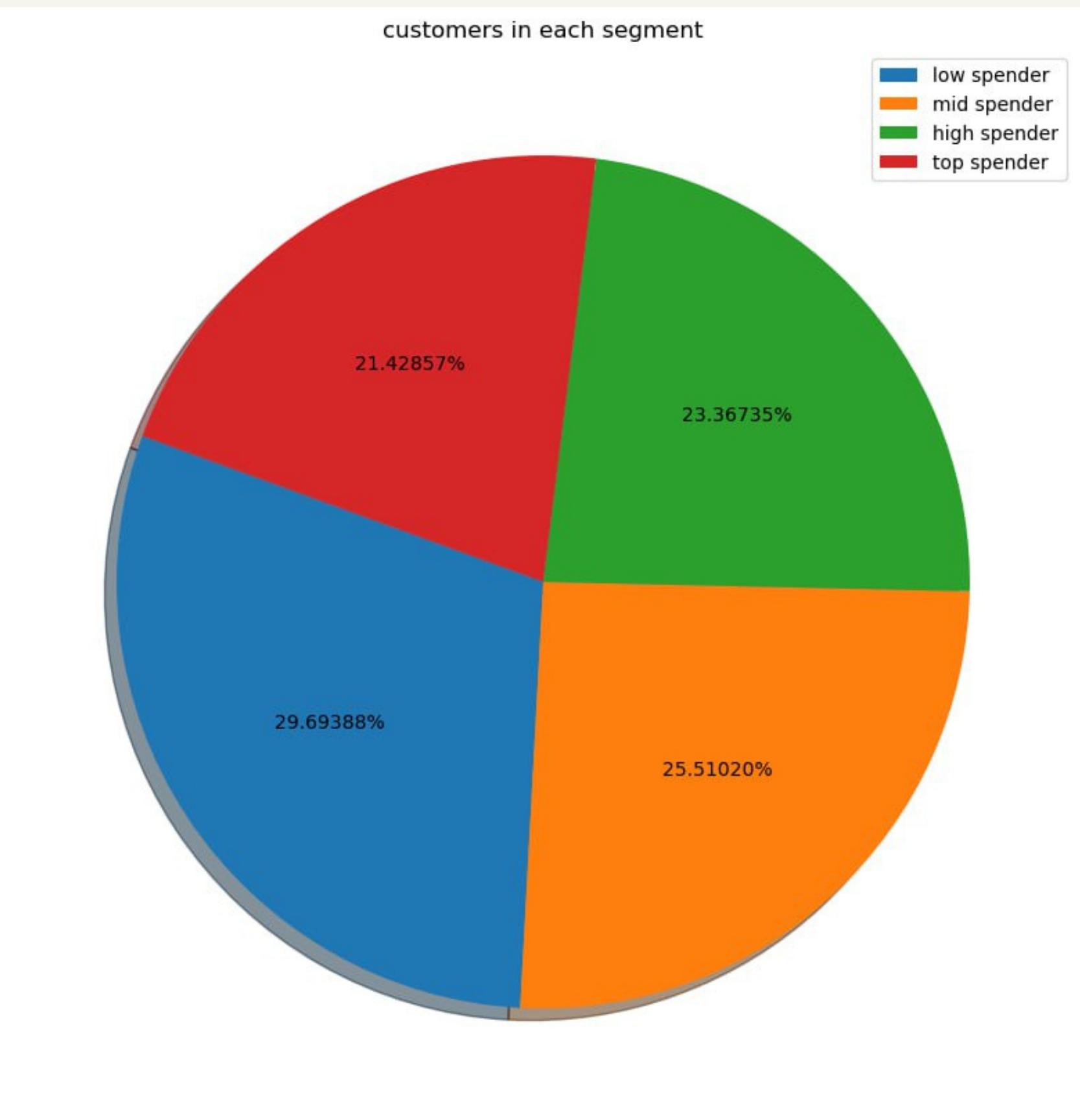
# RFM ANALYSIS

## Plot the RFM Level



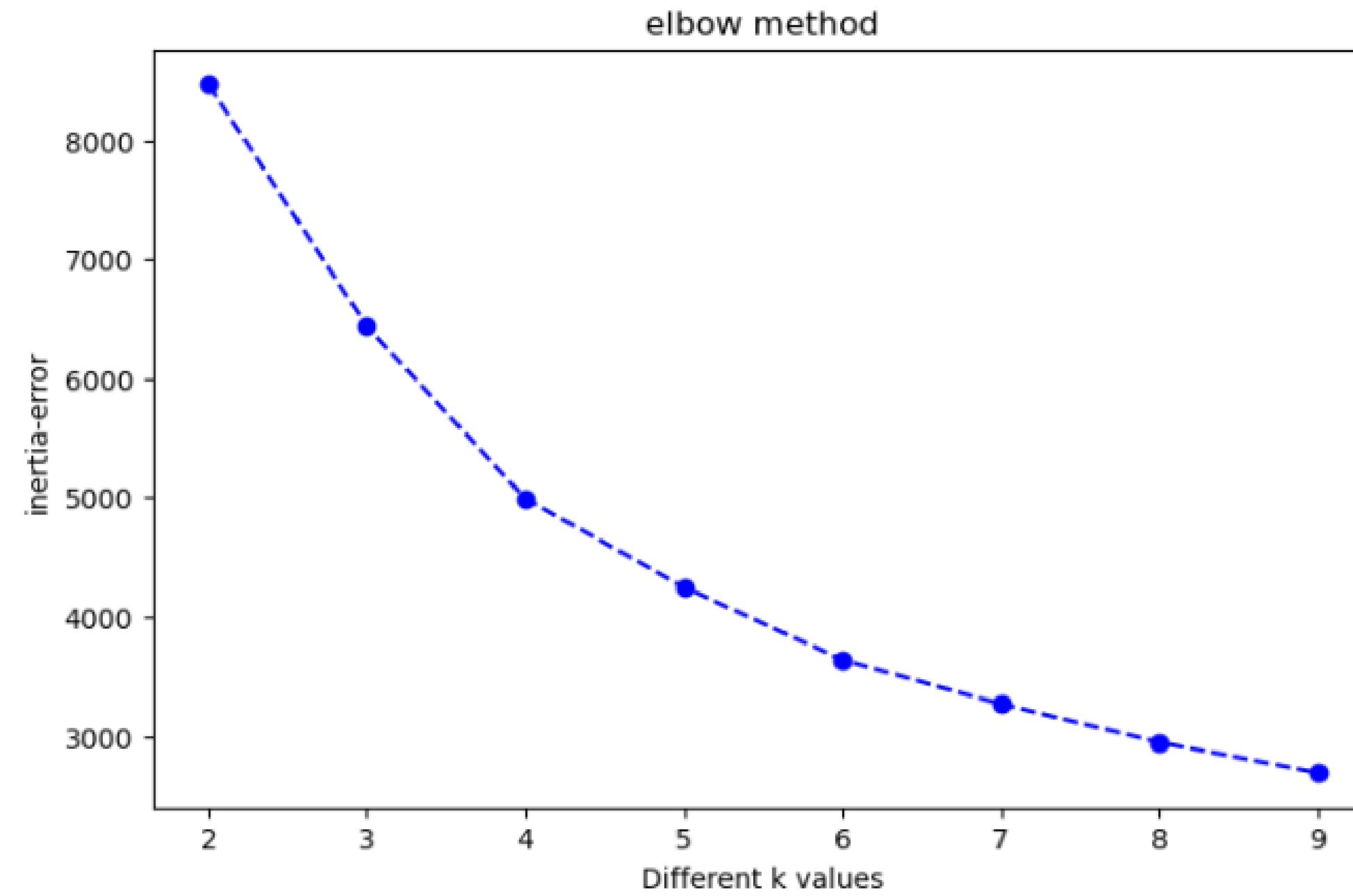


# RFM ANALYSIS



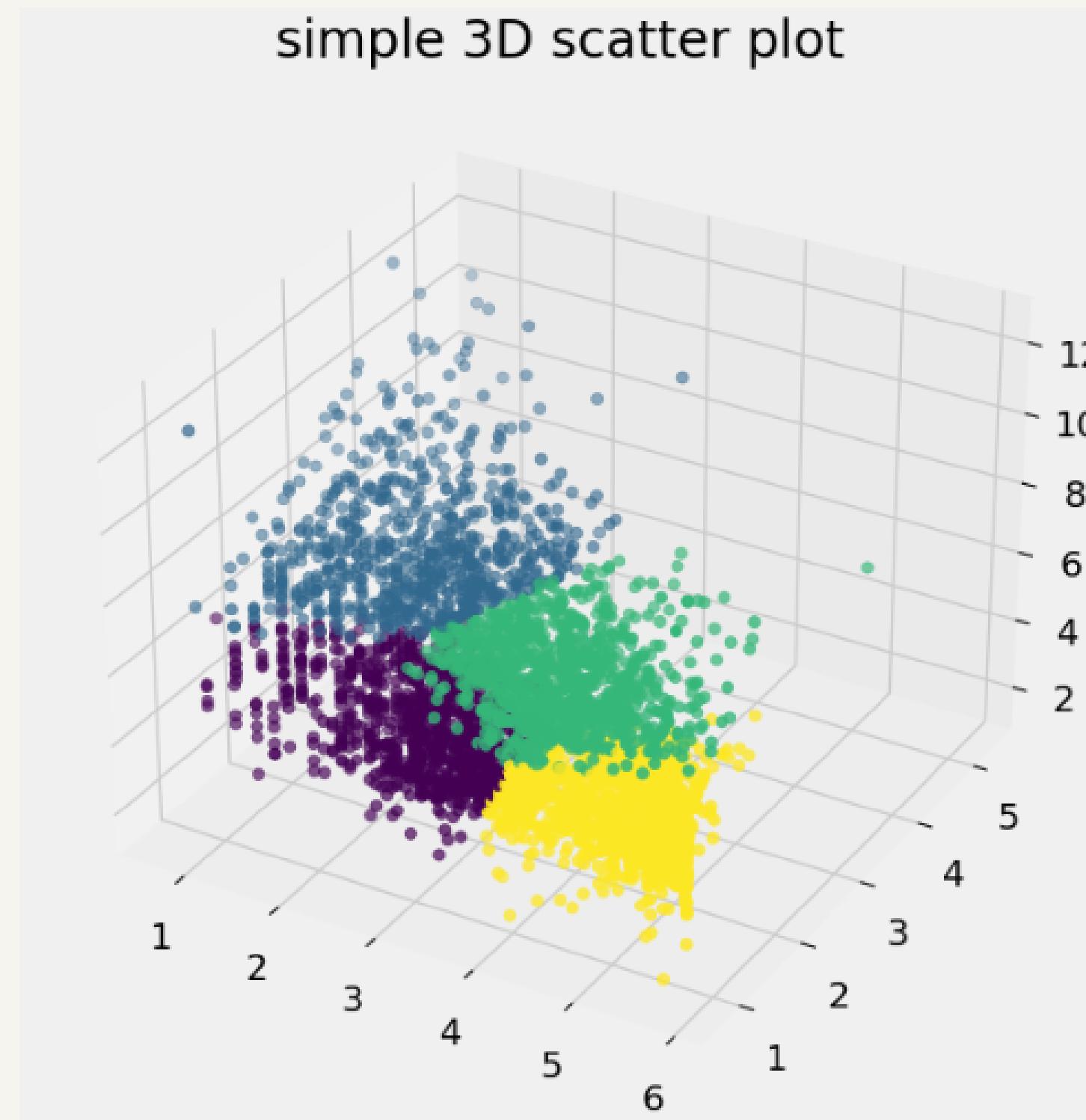


# K-means clustering

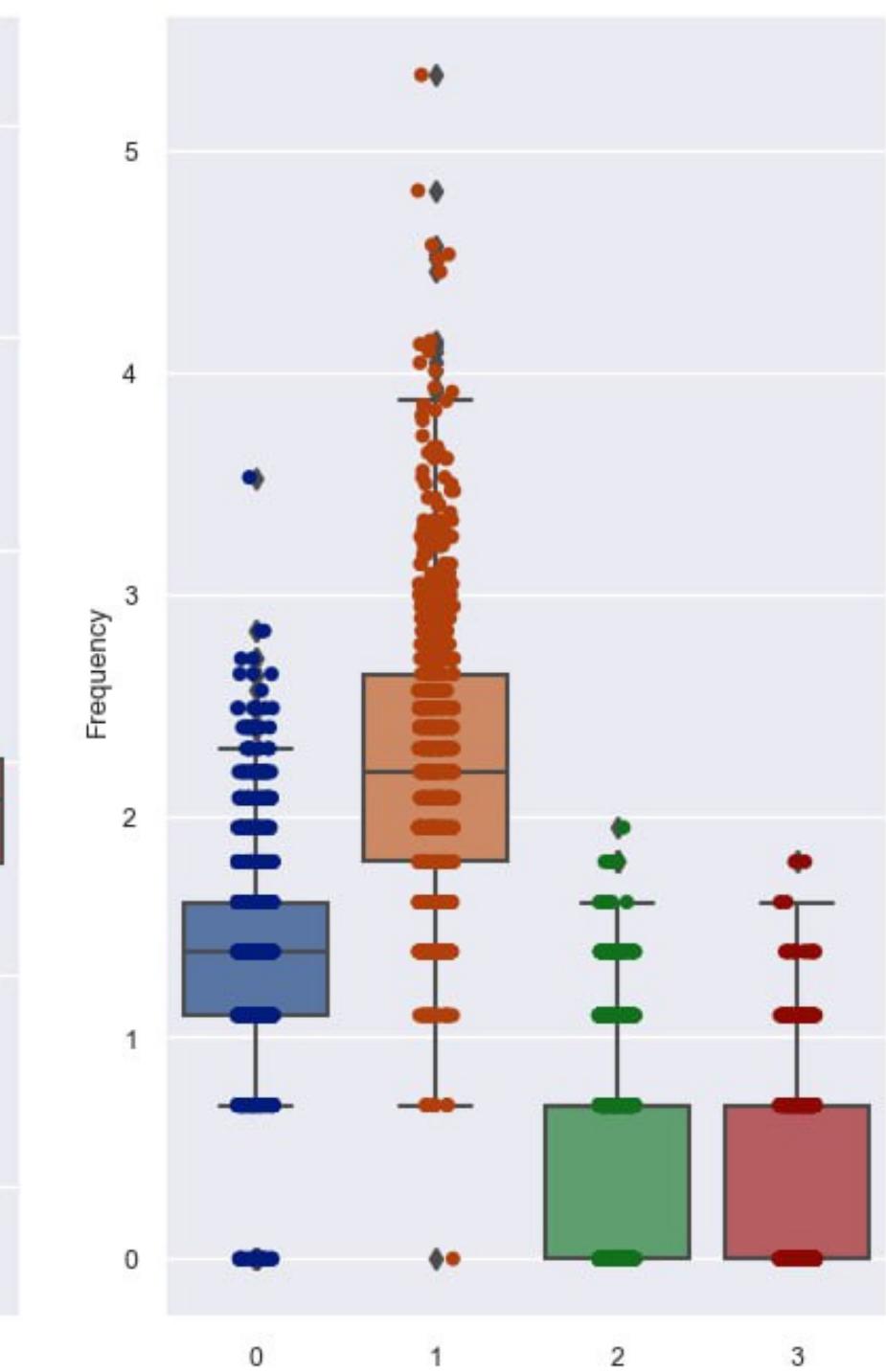
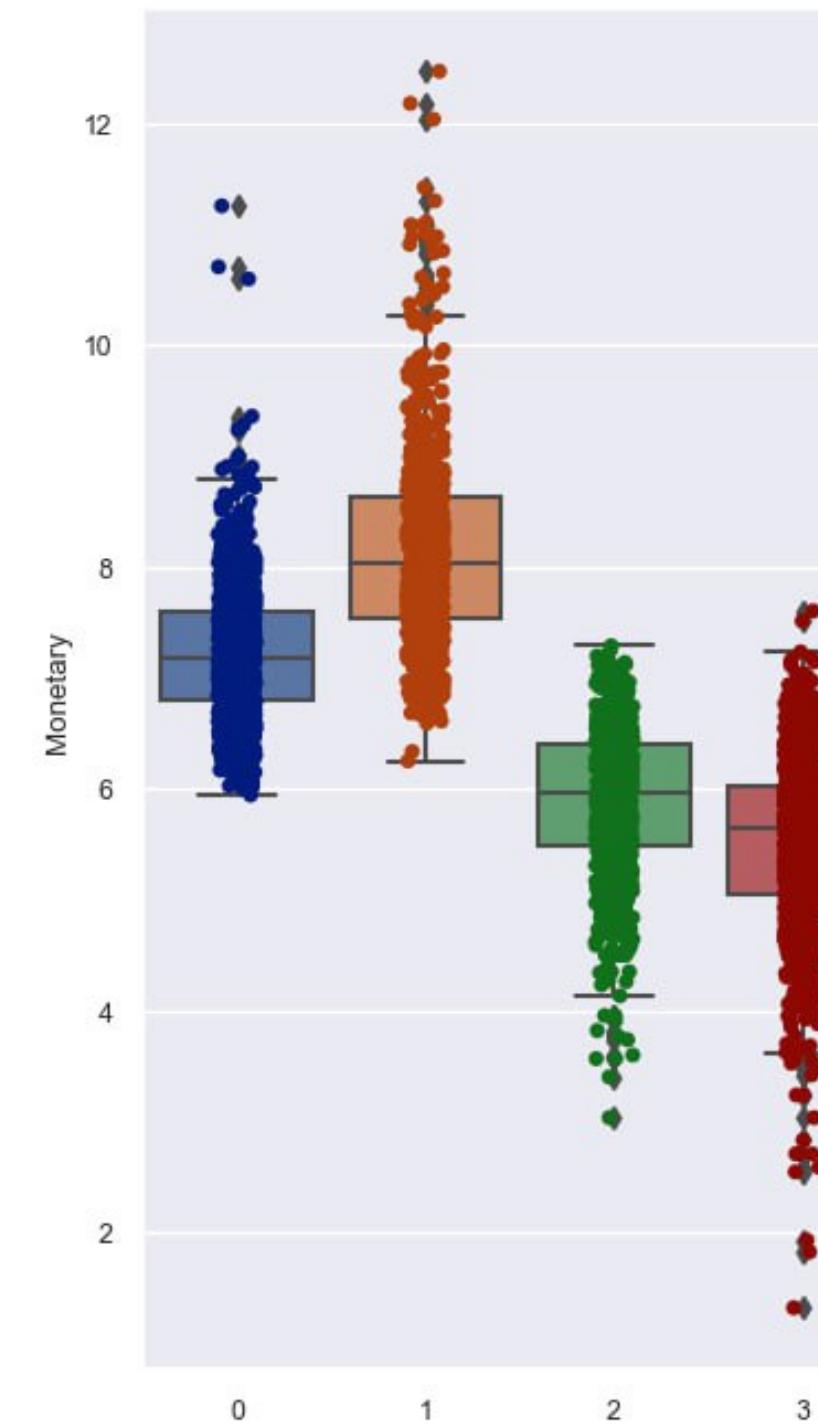
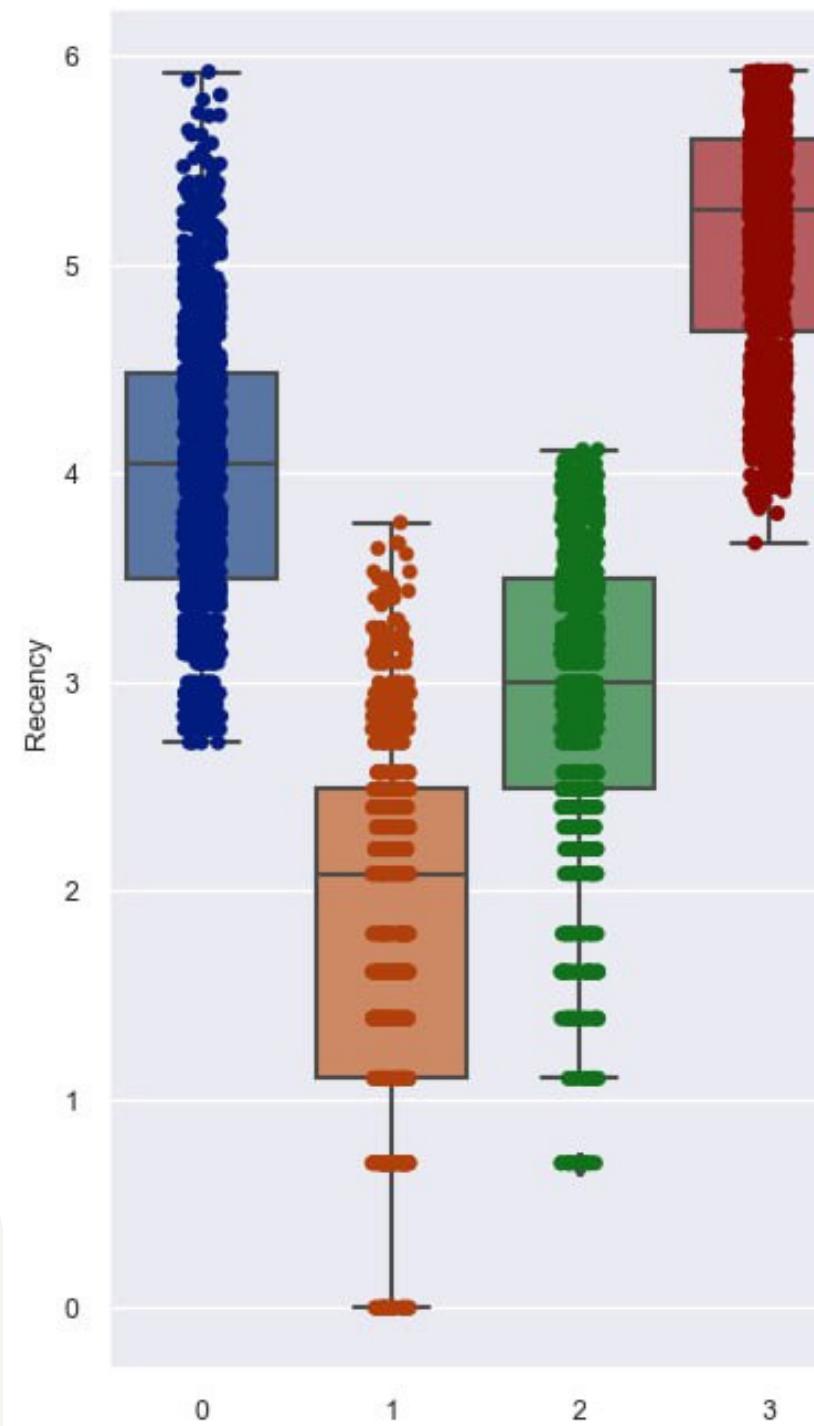




# K-means clustering

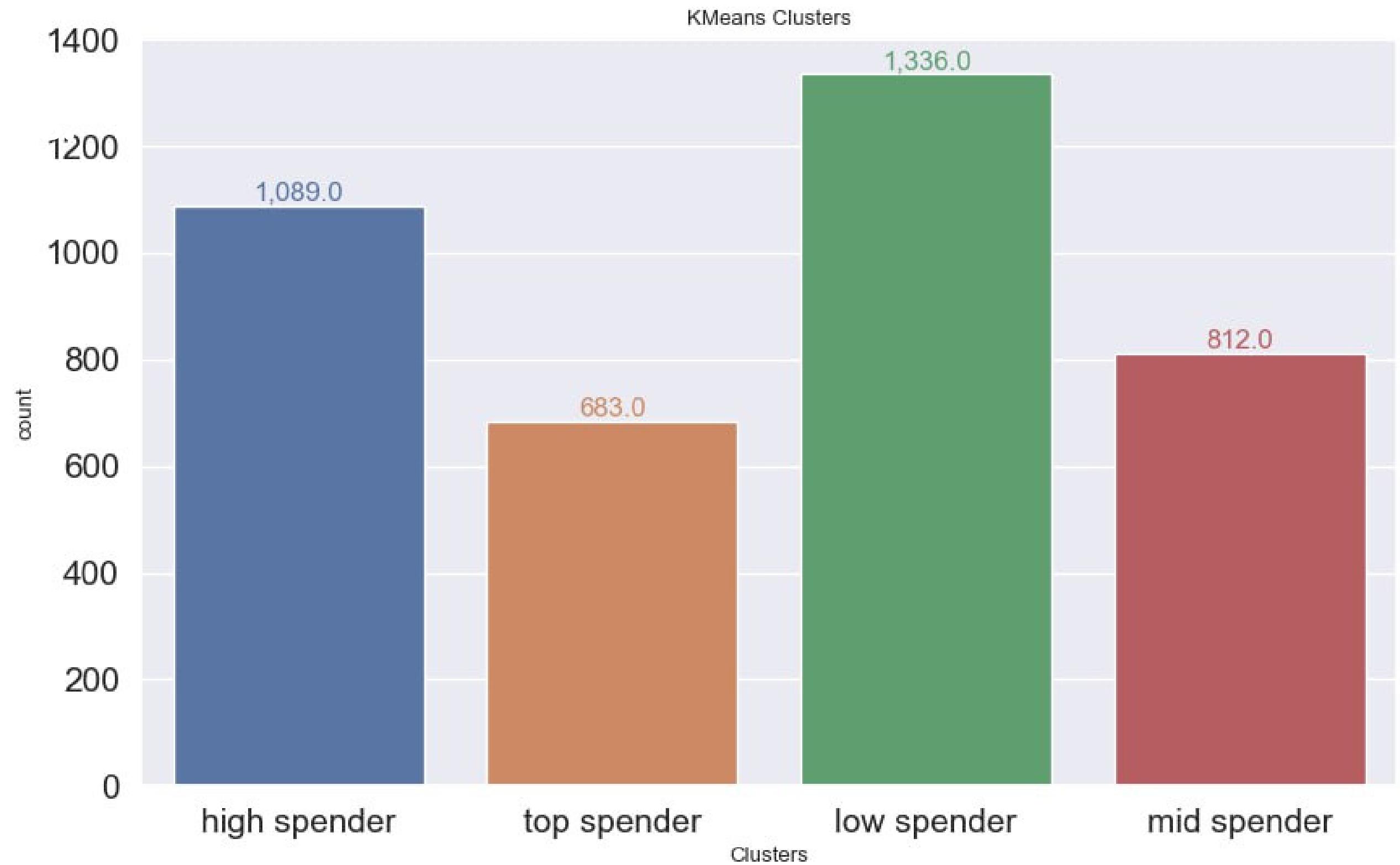


# K-MEANS CLUSTERING



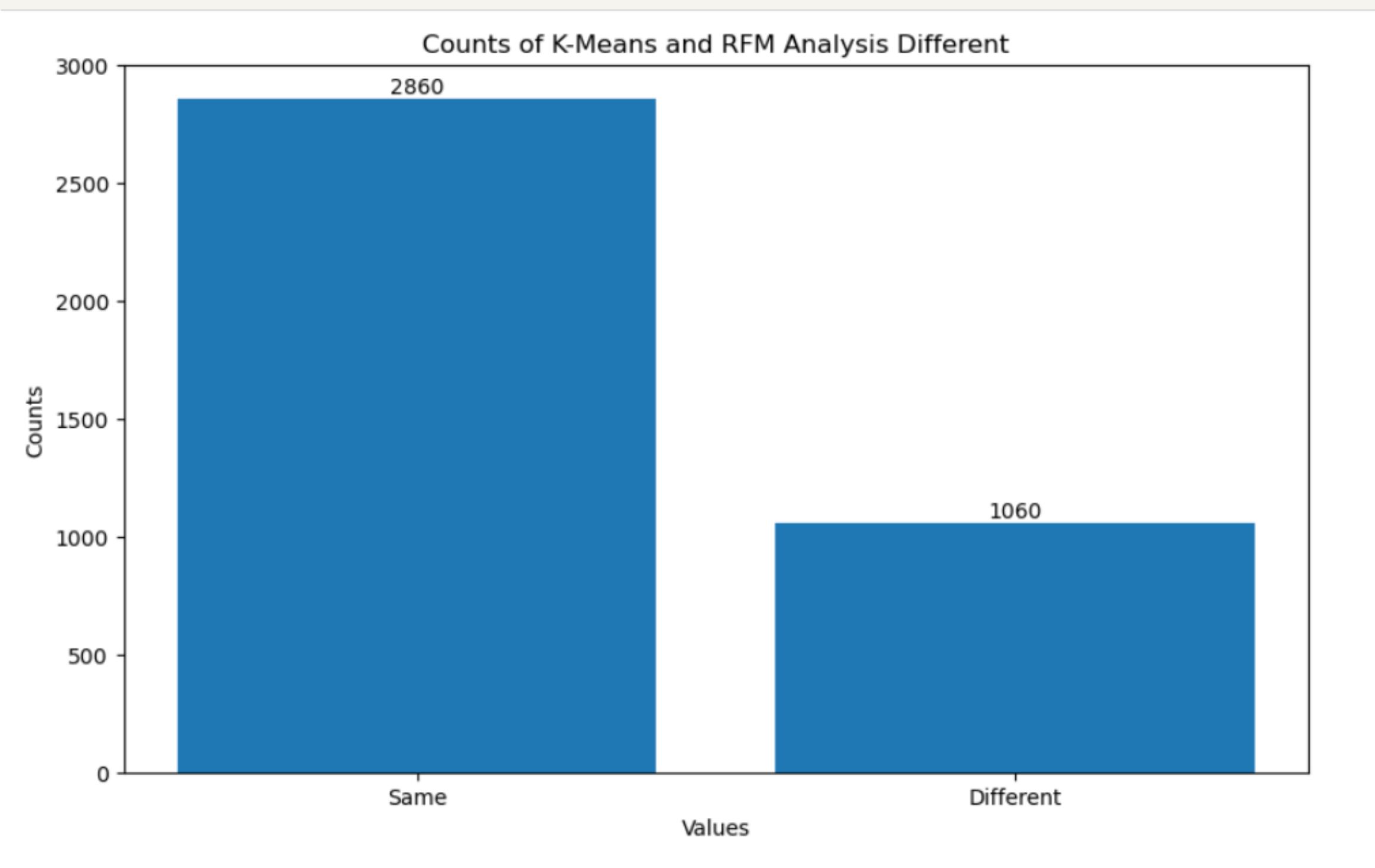


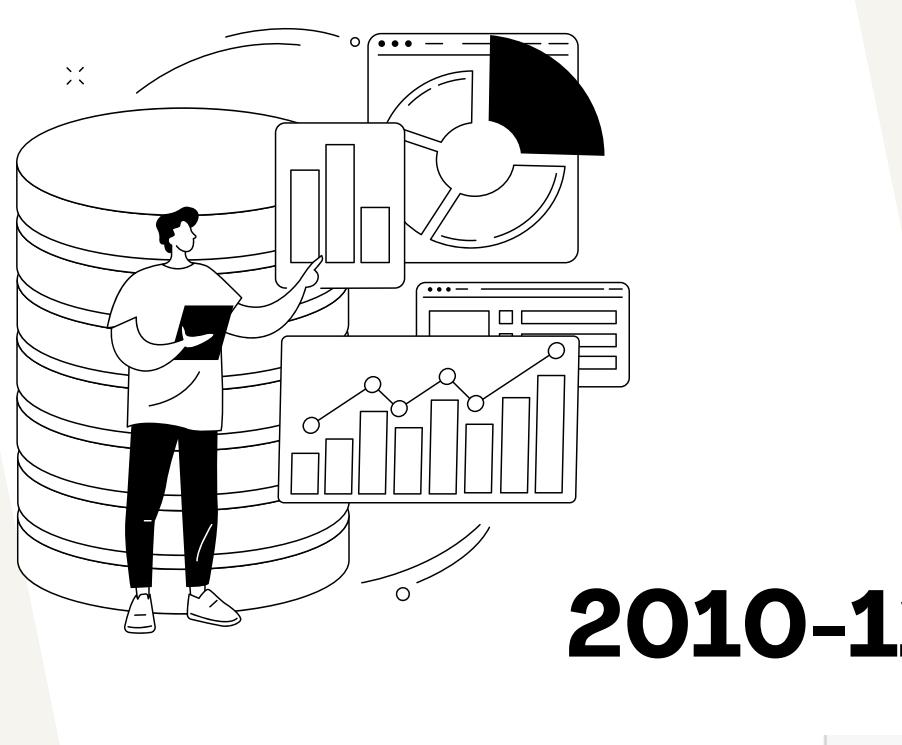
# K-means clustering





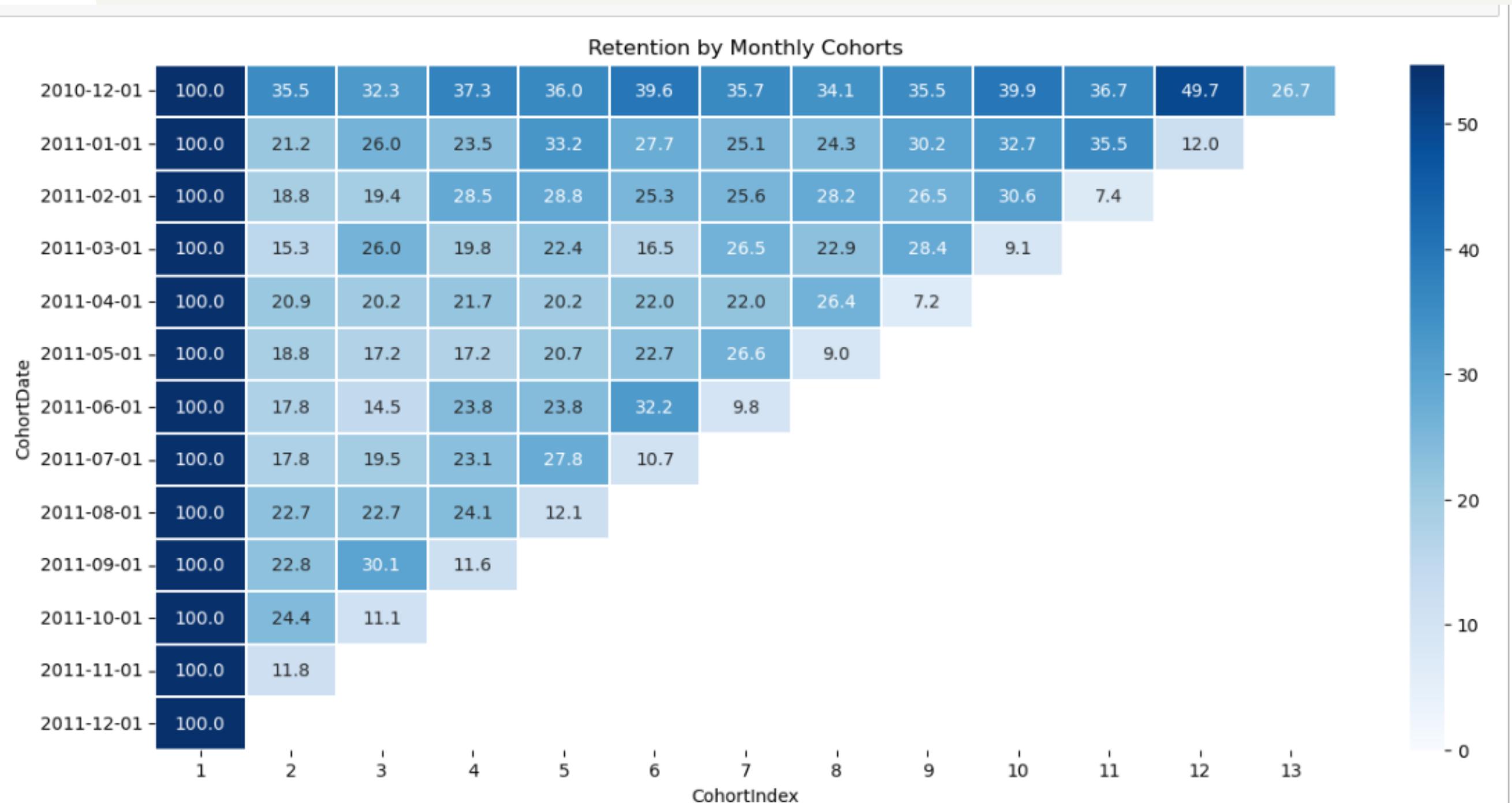
# K-means clustering

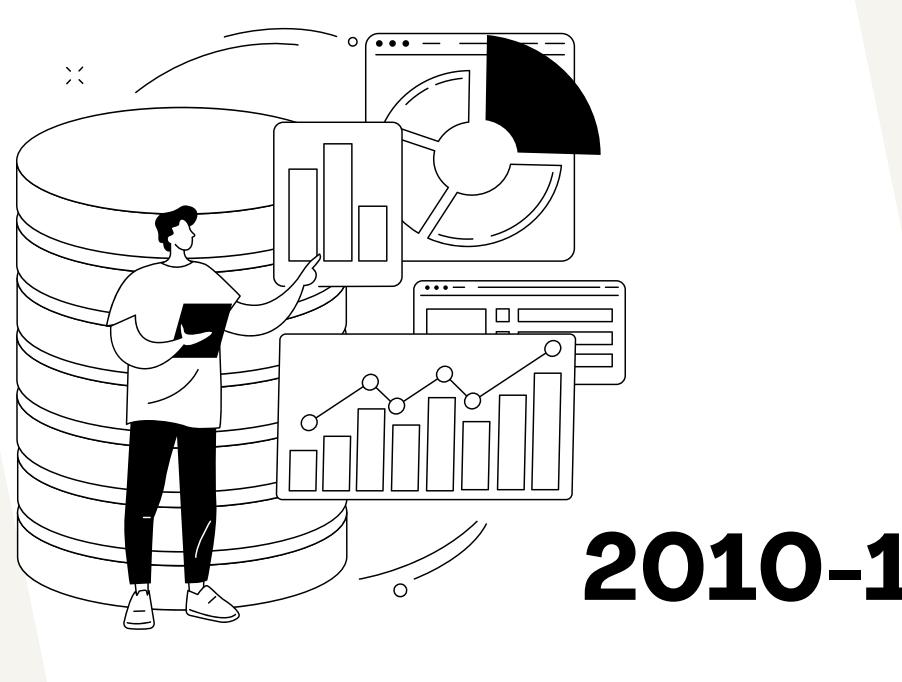




# K-means clustering

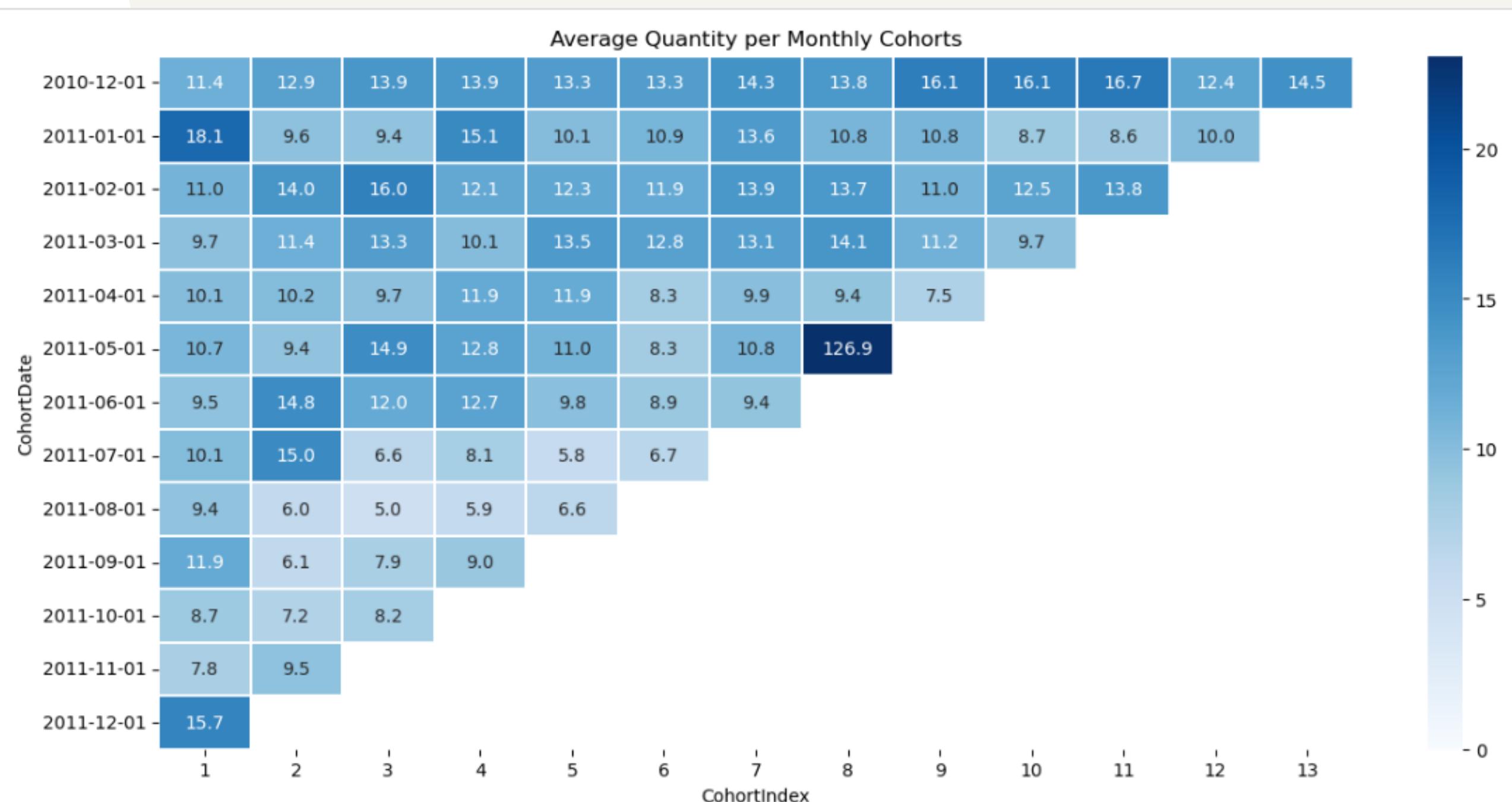
**2010-11 have the highest customer retention for each CohortDate**





# K-means clustering

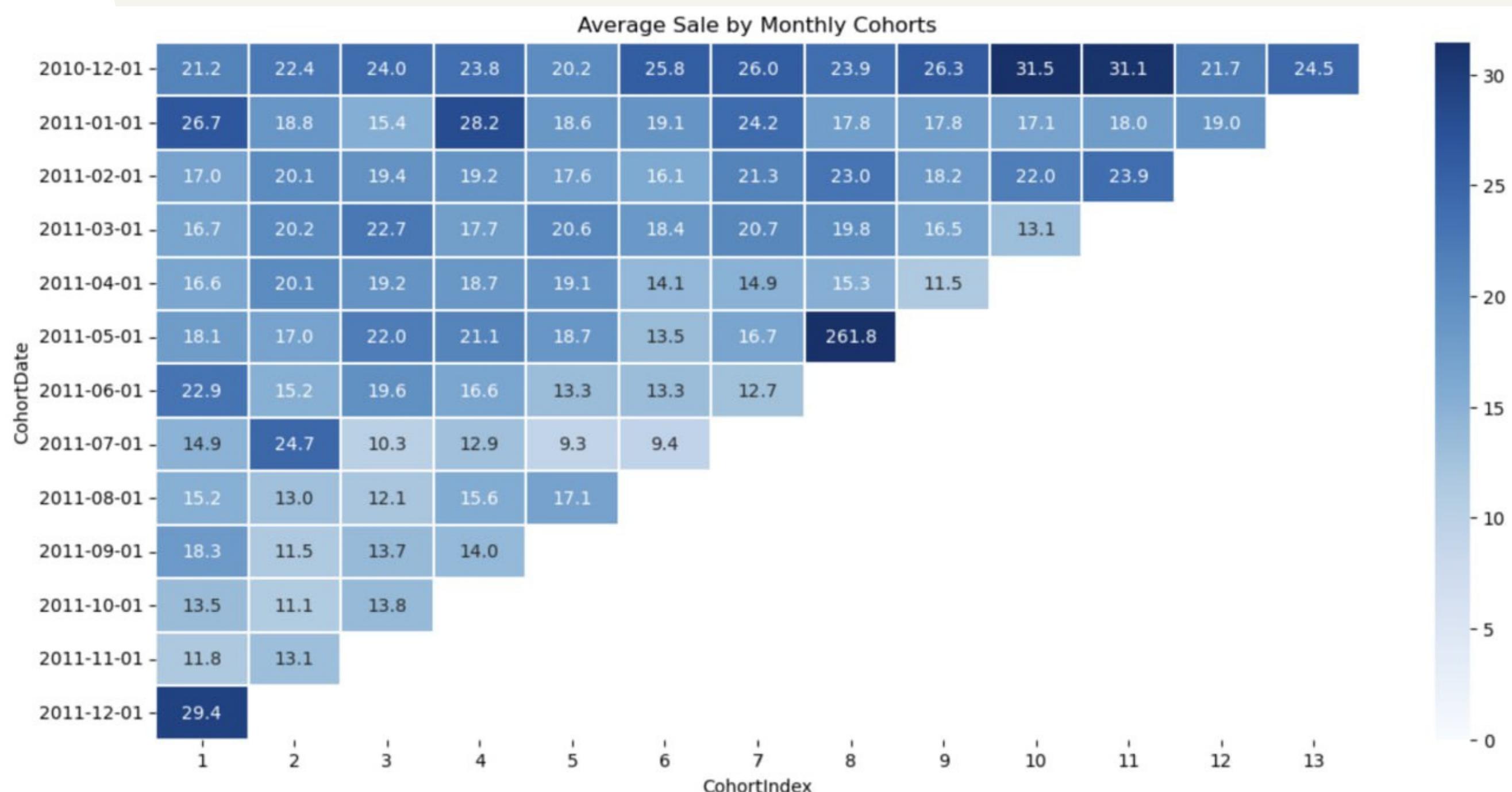
**2010-12 the average quantity sold was increasing until cohort 12**





# K-means clustering

**high spender for 2011-8 and their spends are increasing**



# Conclusion

- Data Cleaning & Exploratory Data Analysis
- RFM Analysis
- Customer Segmentation
- Applying Clustering
- Create a Cohort and Conduct Cohort Analysis





**Thank you, We hope you Enjoyed Our  
Presentation  
Do you have Questions?**