

# **CUSTOMER CLUSTERING IN A RETAIL MARKET: DATA DRIVEN APPROACHES**

**- VIPUL CHALOTRA (280269)**

## **THESIS ADVISORS:**

- Fabrizio Rossi
- Andrea Manno
- Stefano Smriglio

**December 16, 2023**



**Università degli  
Studi dell'Aquila**



# Table of Contents

- **Introduction**
- **Exploratory Data Analysis**
- **Customer Clustering Insights**
- **Conclusion**



# Study Objectives

- **Retail Industry Trends.**
- **Exploratory Analysis to Uncover Customer Characteristics.**
- **Effective Data-Driven Customer Clustering Methods.**
- **Customer Segmentation Patterns and Characteristics.**
- **Impact of Customer Clustering on Retail Performance Metrics.**
- **Practical Recommendations for Retail Strategy Enhancement.**



# Table of Contents

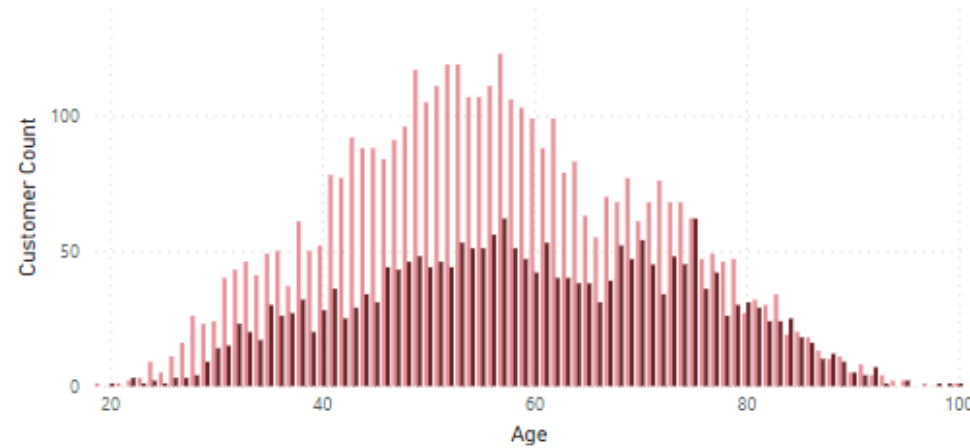
- Introduction
- **Exploratory Data Analysis**
- Customer Clustering Insights
- Conclusion



# Demographic Analysis

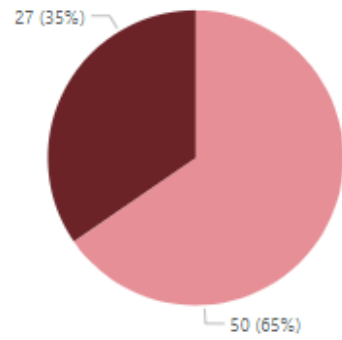
Age Distribution by Gender

Location 1: ● Women ● Men



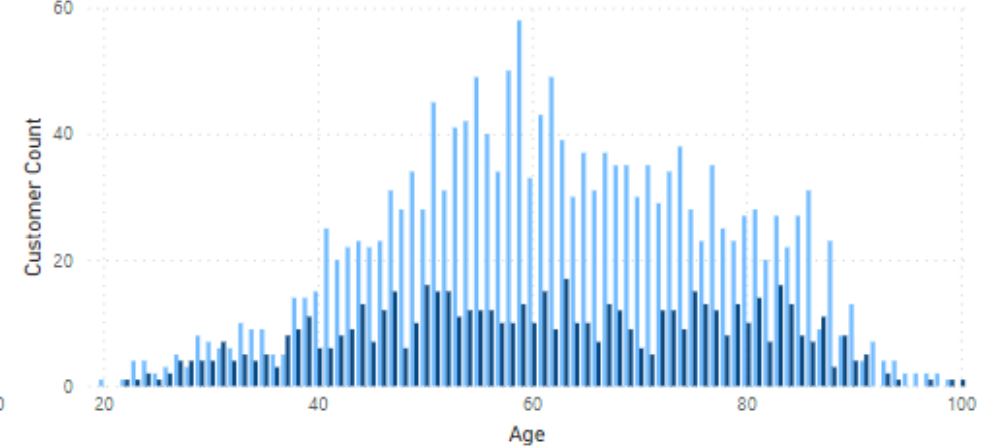
Gender Distribution

Location 1: ● Women ● Men



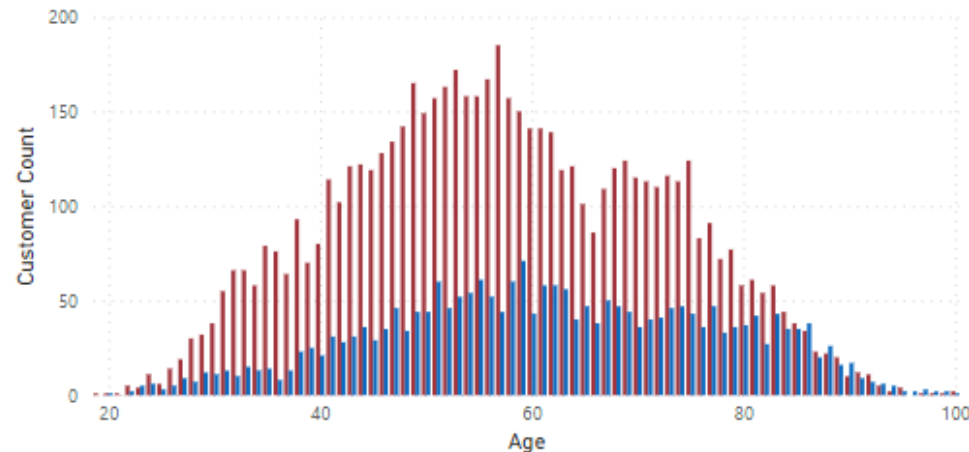
Age Distribution by Gender

Location 2: ● Women ● Men



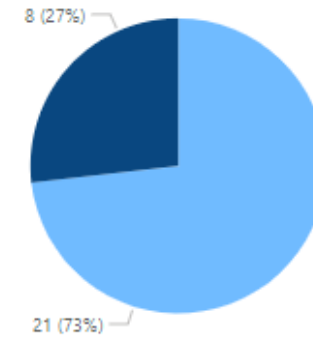
Overall Age Distribution

● Location 1 ● Location 2



Gender Distribution

Location 2: ● Women ● Men

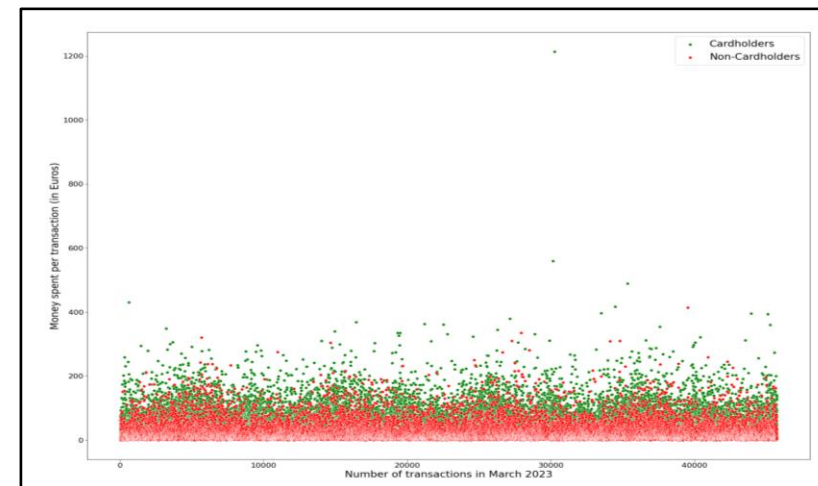
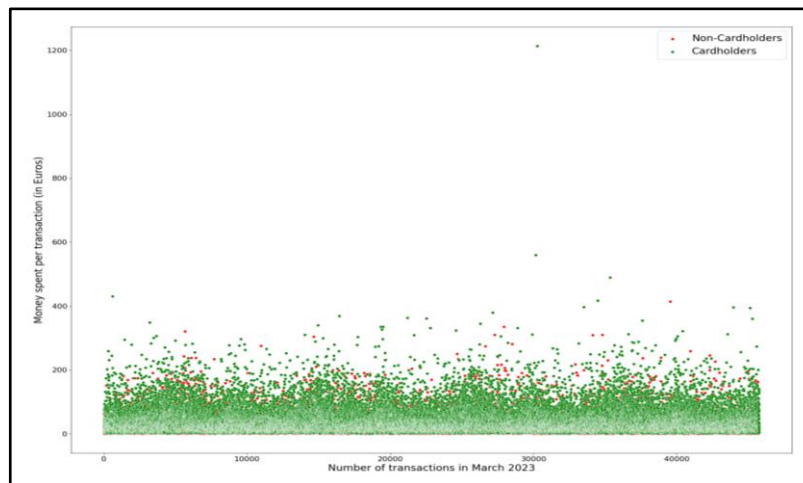




# Transactional Analysis

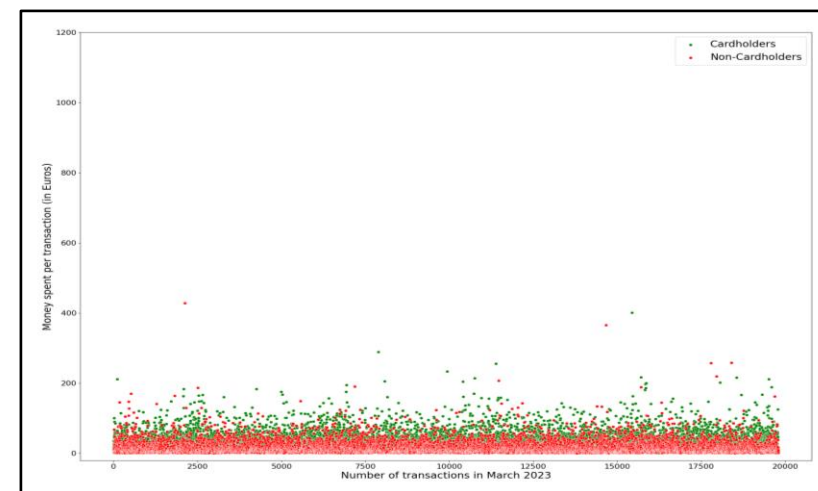
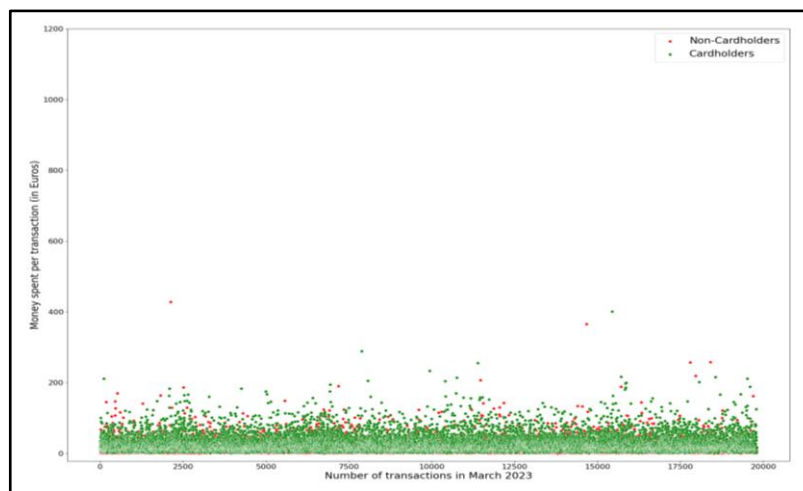
## LOCATION 1

	Cardholders	Non-Cardholders
Items on Offer	4.3	2.1
Total Items	19	11
Amount Spent	45	26
% Items on Offer	23%	19%
Transactions	29,000	16,700



## LOCATION 2

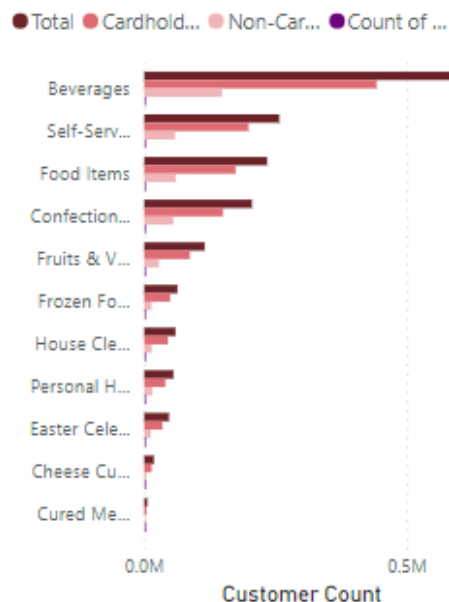
	Cardholders	Non-Cardholders
Items on Offer	3.8	1.6
Total Items	13	8
Amount Spent	28	18
% Items on Offer	30%	19%
Transactions	19,200	7,600



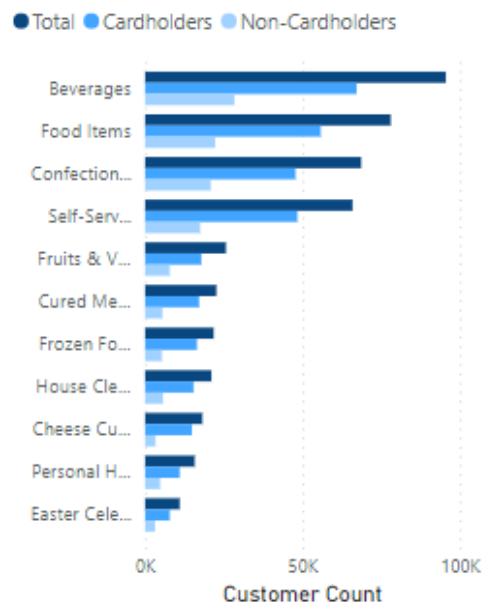


# Item Sector Analysis

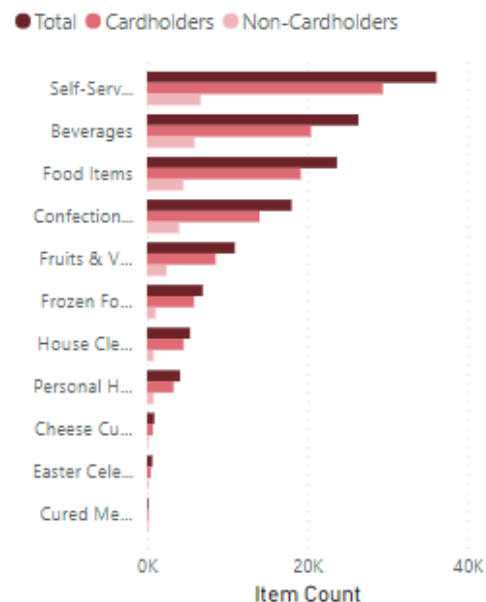
Location 1: Total Sector Expenditure



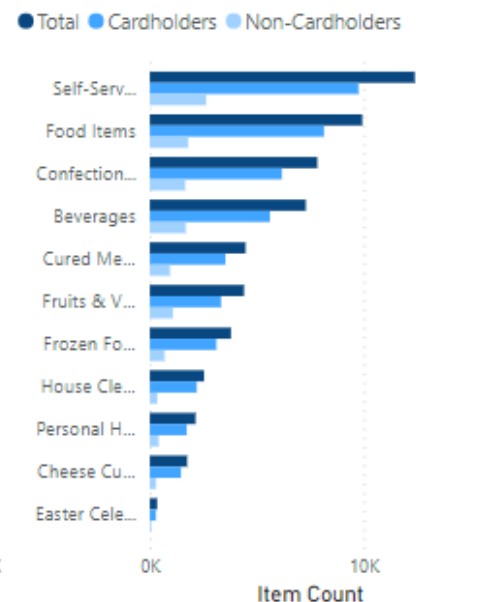
Location 2: Total Sector Expenditure



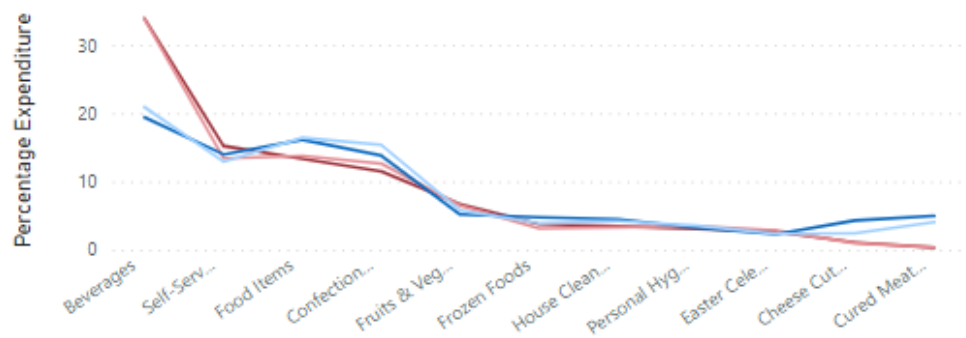
Location 1: Items on Offer by Sector



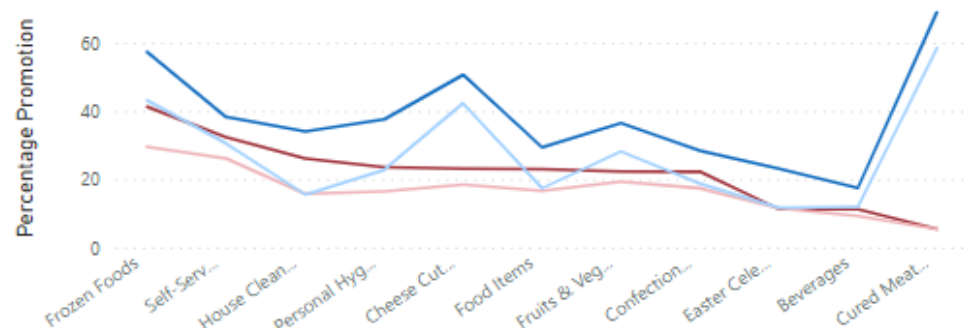
Location 2: Items on Offer by Sector



Location 1 & Location 2: ● Cardholders ● Non-Cardholders ● Cardholders ● Non-Cardholders



Location 1 & Location 2: ● Cardholders ● Non-Cardholders ● Cardholders ● Non-Cardholders

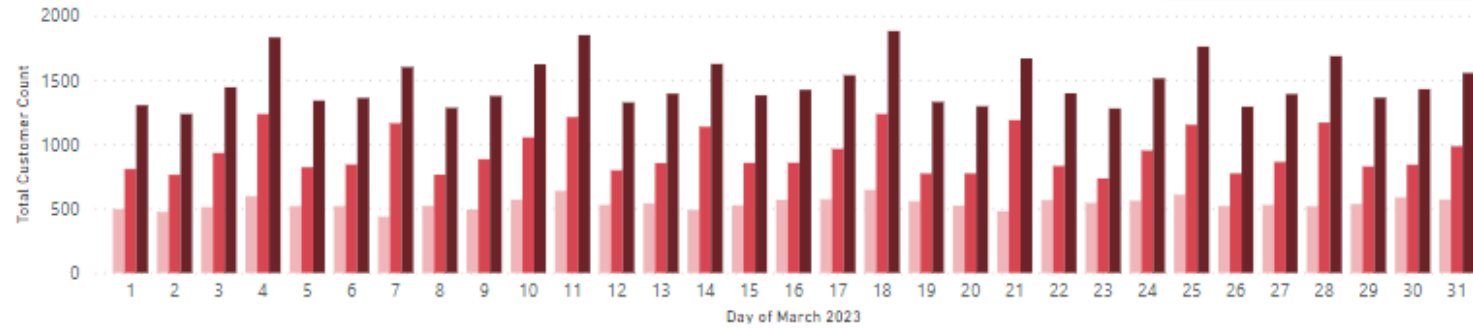




# Daily Analysis

Customer Count by Day

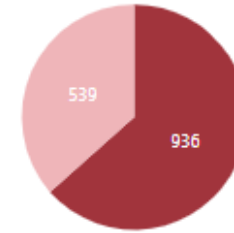
Location 1: Non-Cardholders Cardholders All Customers



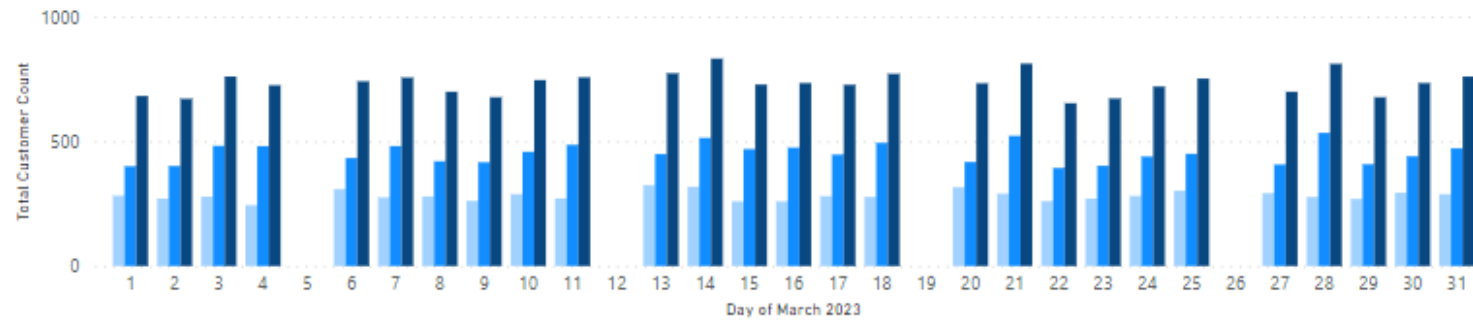
Friday

Location 1:

Cardholders Non-Cardholders

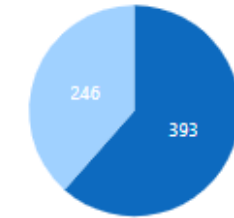


Location 2: Non-Cardholders Cardholders All Customers



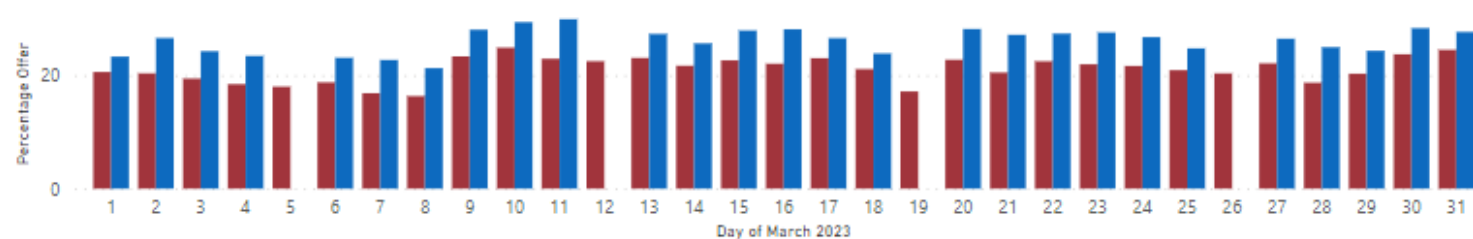
Location 2:

Cardholders Non-Cardholders



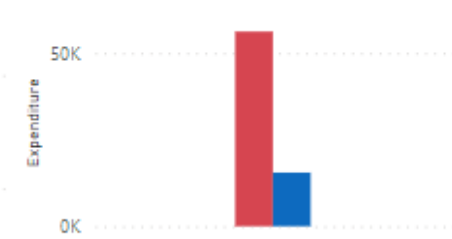
Percentage of Items on Offer by Day

Location 1 Location 2



Average Expenditure (in Euros)

Location 1 Location 2



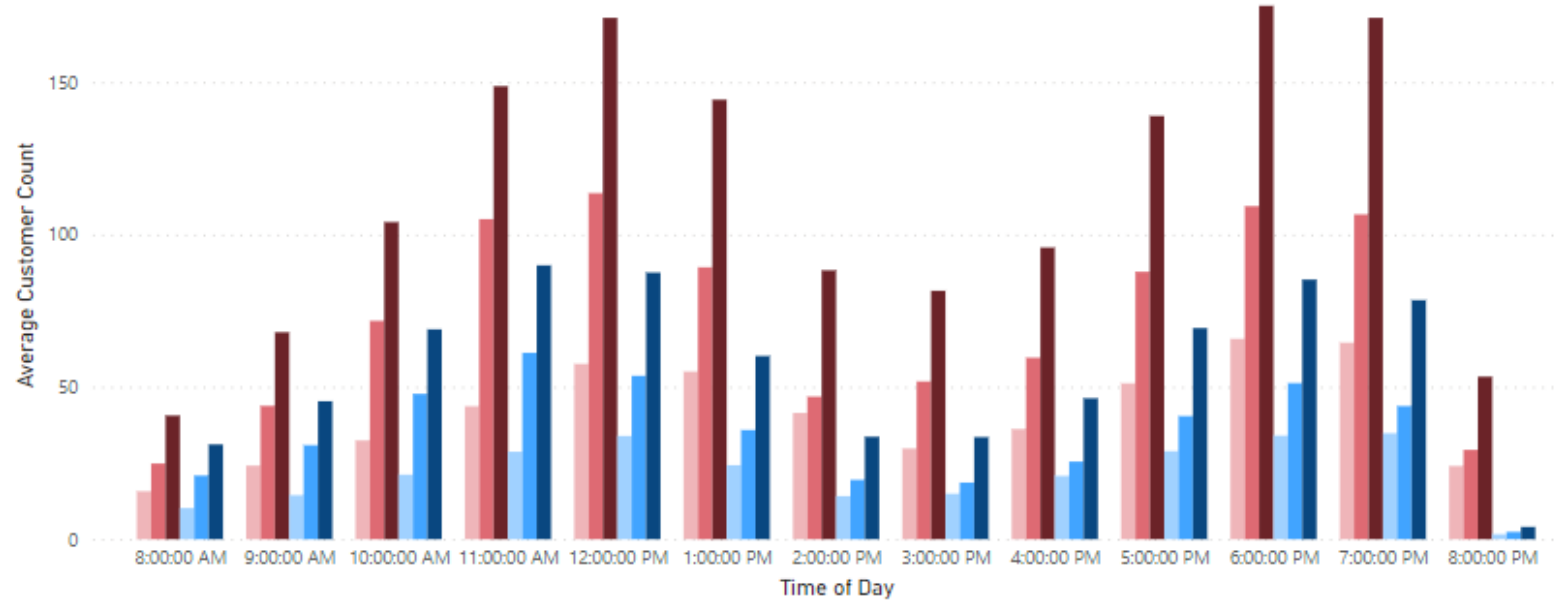




# Hourly Analysis

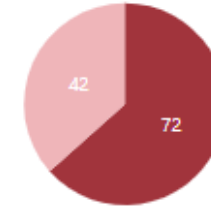
Customer Count by Time of Day

Location 1 & Location 2: Non-Cardholders Cardholders All Customers Non-Cardholders Cardholders All Customers



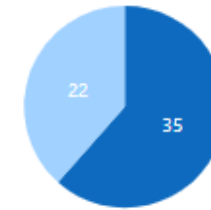
Location 1:

Cardholders Non-Cardholders



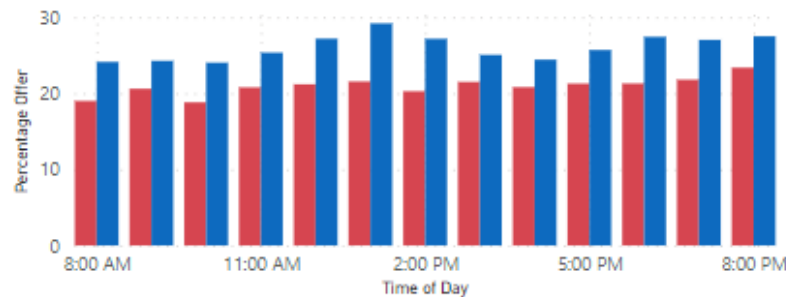
Location 2:

Cardholders Non-Cardholders



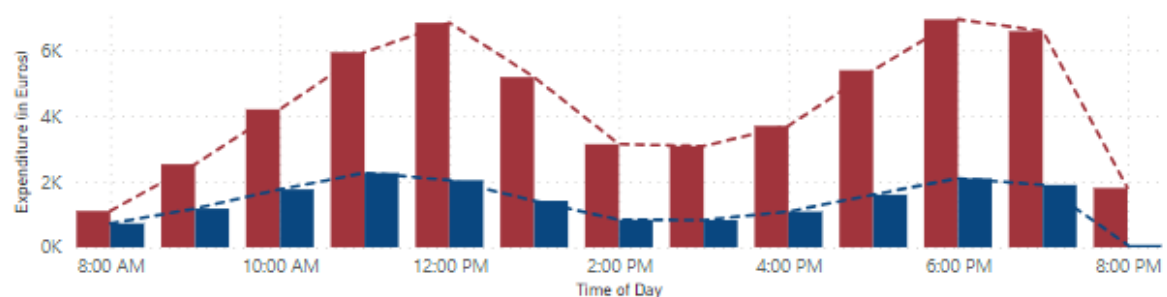
Percentage of Items on Offer by Time of Day

Location 1 Location 2



Expenditure by Time of Day

Location 1 Location 2 Location 1 Location 2





# Table of Contents

- Introduction
- Exploratory Data Analysis
- **Customer Clustering Insights**
- Conclusion

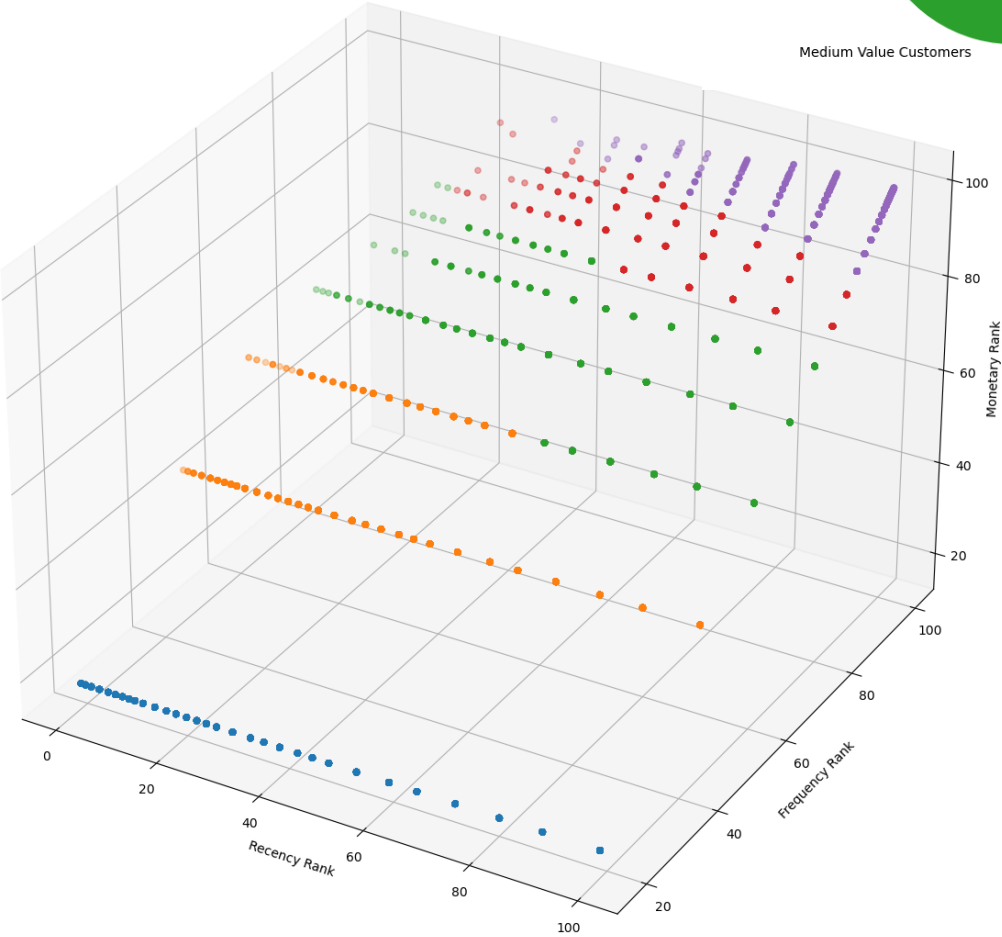
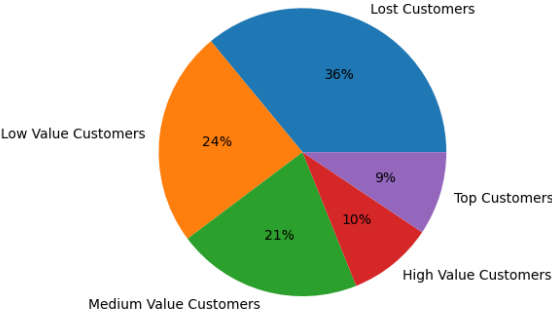


# Recency-Frequency-Monetary (RFM)

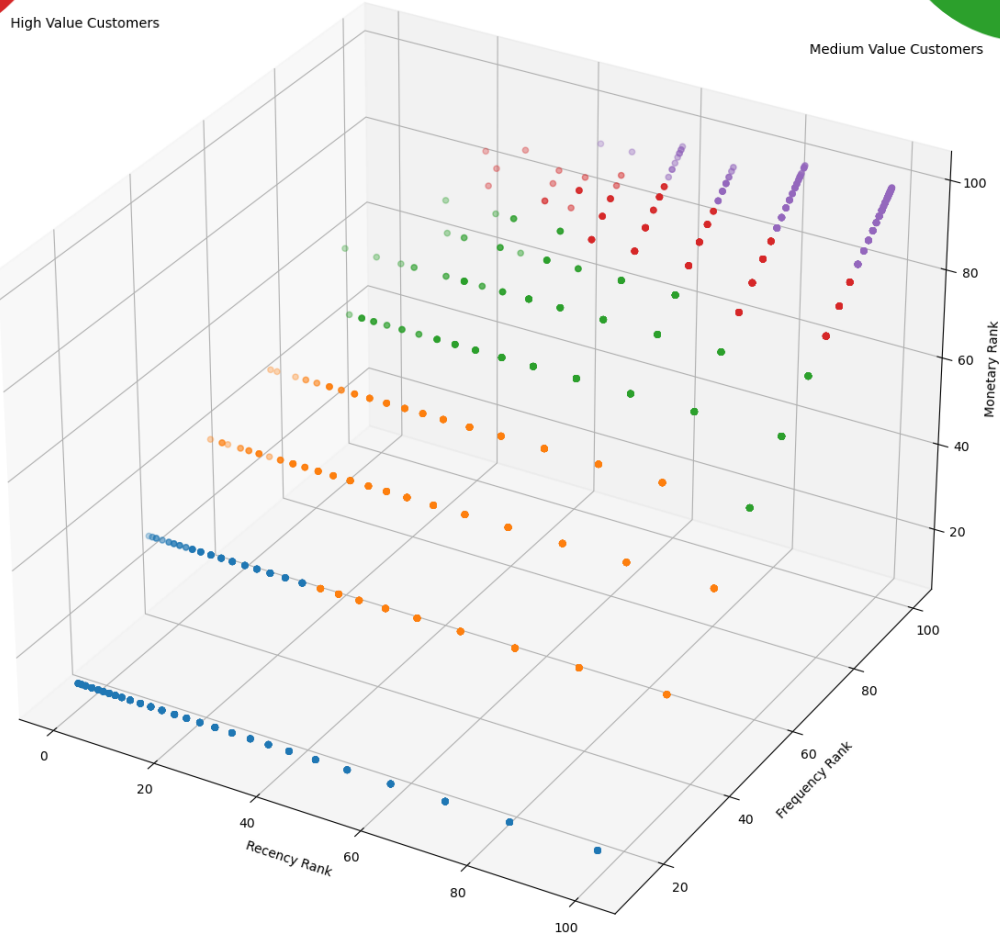
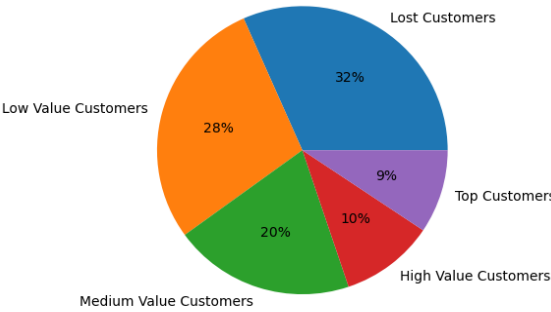
- **Customer Base under Investigation: Cardholders**
- **Two Main Parts:**
  - **Predefined Segmentation**
  - **K-Means Clustering**
- **Predefined Segmentation**
  - **10% Recency, 30% Frequency, 60% Monetary**
  - **RFM Score on 5 Point Scale**
  - **Rank Separation at 1.6, 3.0, 4.0, 4.5**
- **K-Means Clustering into 2 & 4 Clusters**



# Location 1



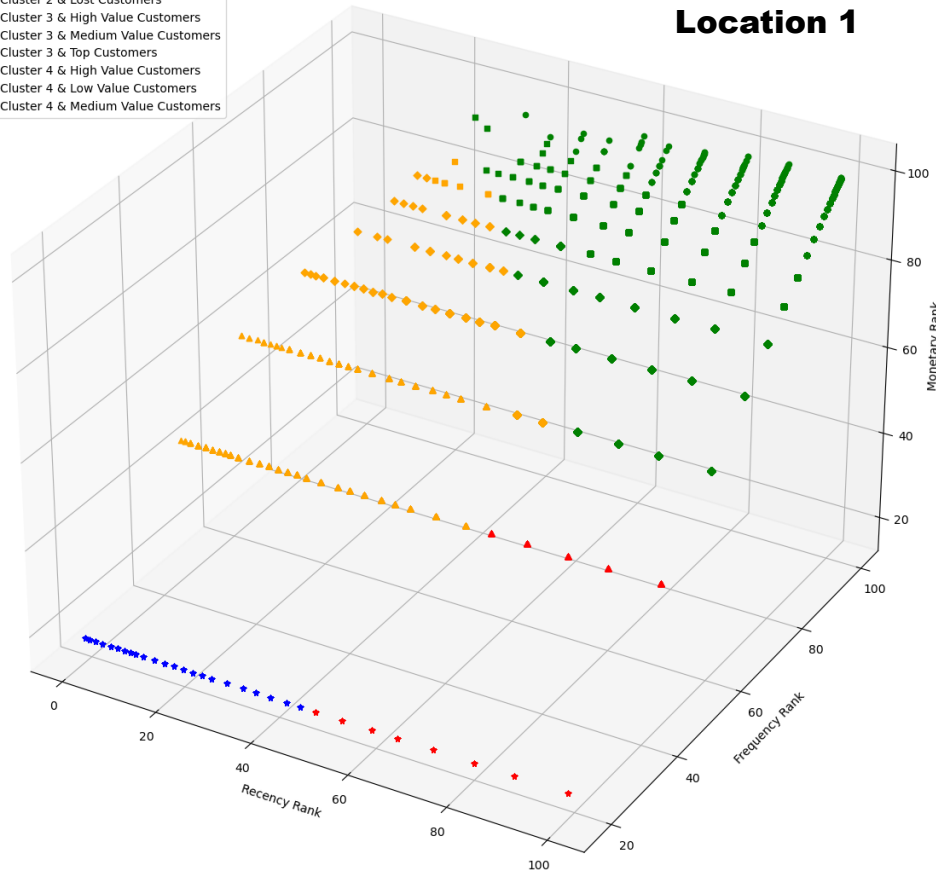
# Location 2





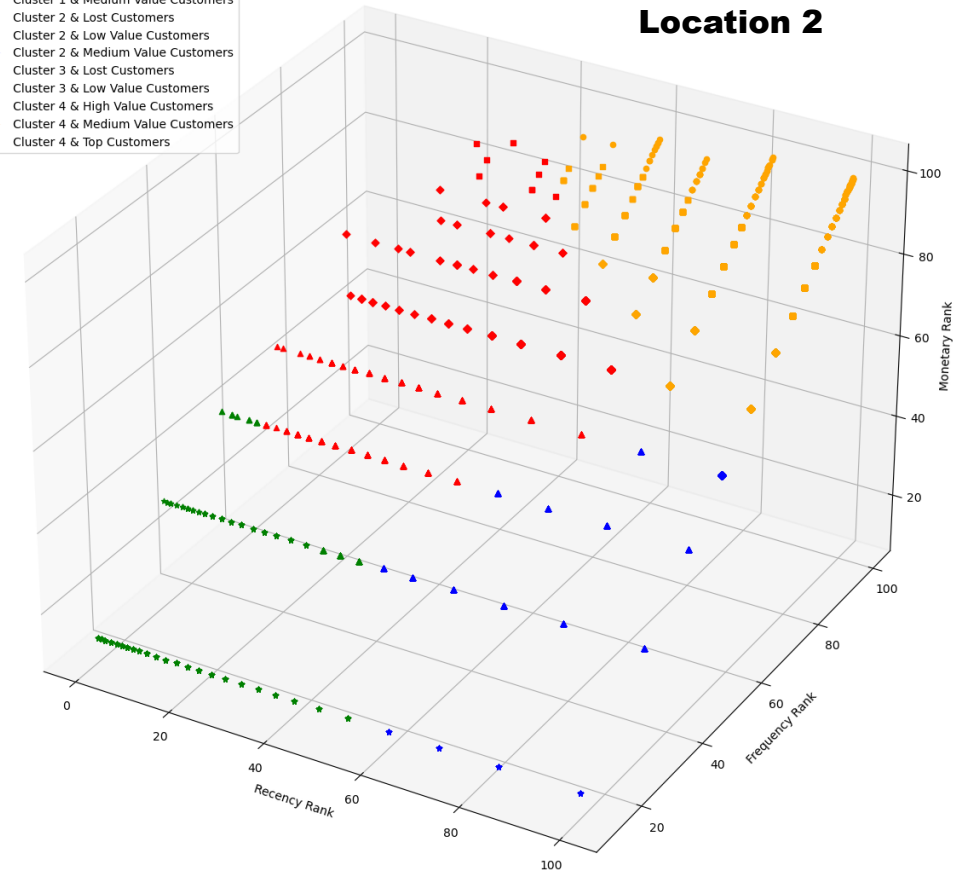
RFM-Based K-Means Clustering Results & Predefined Segments (3D)

- K-Means Clusters & Predefined Segments:
- Cluster 1 & Lost Customers
  - Cluster 1 & Low Value Customers
  - Cluster 2 & Lost Customers
  - Cluster 3 & High Value Customers
  - Cluster 3 & Medium Value Customers
  - Cluster 3 & Top Customers
  - Cluster 4 & High Value Customers
  - Cluster 4 & Low Value Customers
  - Cluster 4 & Medium Value Customers



RFM-Based K-Means Clustering Results & Predefined Segments (3D)

- K-Means Clusters & Predefined Segments:
- Cluster 1 & High Value Customers
  - Cluster 1 & Low Value Customers
  - Cluster 1 & Medium Value Customers
  - Cluster 2 & Lost Customers
  - Cluster 2 & Low Value Customers
  - Cluster 2 & Medium Value Customers
  - Cluster 3 & Lost Customers
  - Cluster 3 & Low Value Customers
  - Cluster 4 & High Value Customers
  - Cluster 4 & Medium Value Customers
  - Cluster 4 & Top Customers

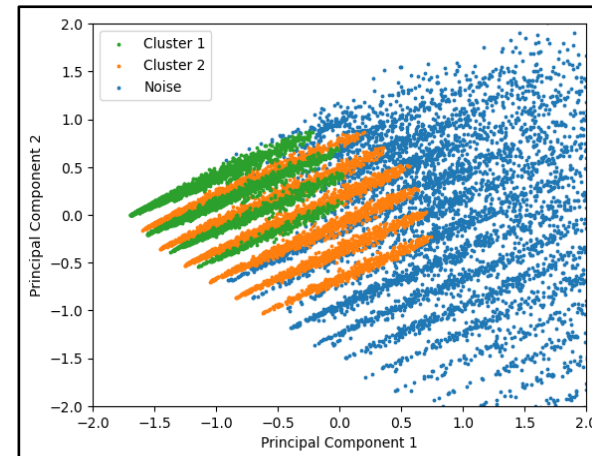
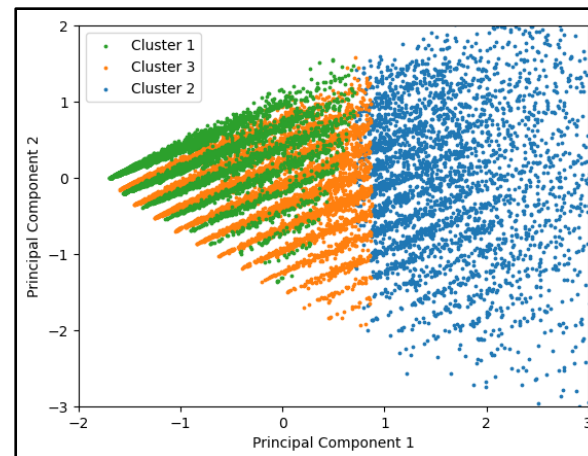
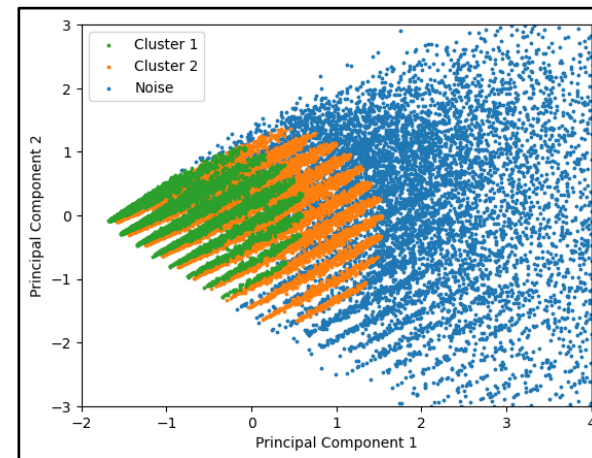
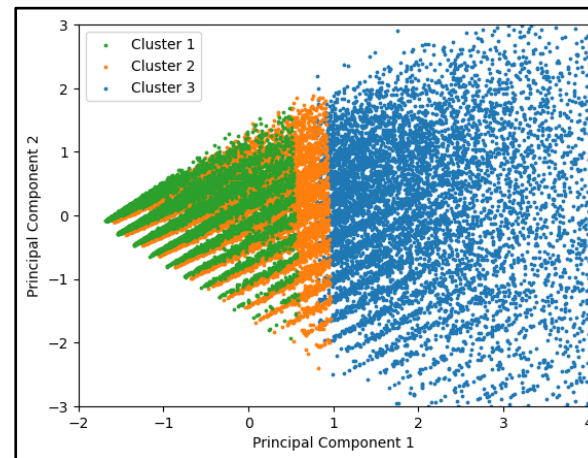


**Cluster 1 & Cluster 2: Low Frequency, Low Expenditure, Recently Visited**  
**Cluster 2 & Cluster 3: Lowest Frequency, Lowest Expenditure, Not Recently Visited**  
**Cluster 3 & Cluster 4: Highest Frequency, Highest Expenditure, Recently Visited**  
**Cluster 4 & Cluster 1: Average Frequency, High Expenditure, Not Recently Visited**



# Transactional Clustering

- **K-Means Clustering with k=2: High & Low Expenditure Clusters**
- **K-Means Clustering with k=3: High Expenditure Cluster, Low Expenditure Cardholder & Non-Cardholder Clusters**
- **DBSCAN Clusters:**
  - **High Expenditure: Noise**
  - **Low Expenditure Cardholder: Orange**
  - **Low Expenditure Non-Cardholder: Green**
- **Blue: Smallest, High Expenditure, 81% & 80% Cardholders, 72% & 71% Cardholders**
- **Orange: Largest, Low Expenditure Cardholders, 99% & 92% Cardholders**
- **Green: Low Expenditure Non-Cardholders, 100% & 99% Non-Cardholders**



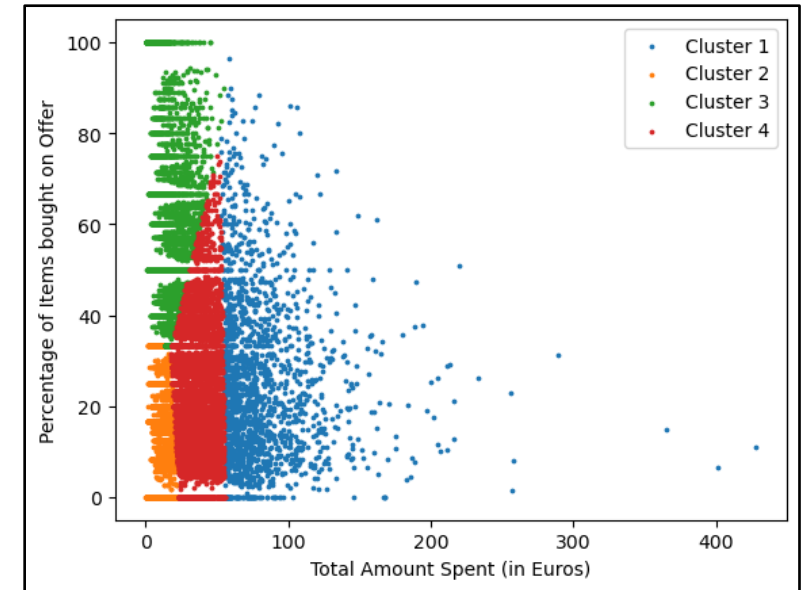
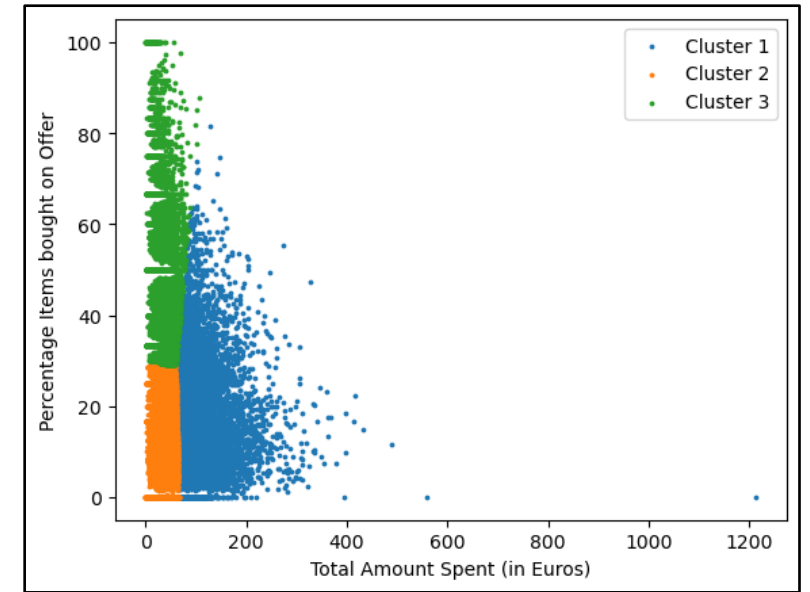
**K-Means**

**DBSCAN**



# Promotional Clustering

- **Location 1 (64% Cardholders): 3 Clusters**
- **Location 2 (62% Cardholders): 4 Clusters**
- **Blue Clusters: Smallest, 80% Cardholders**
  - **% Items on Offer: 18% & 26%**
- **Orange Clusters: Biggest, 50% Cardholders**
- **Green Clusters: 67% & 71% Cardholders**
  - **% Items on Offer: 49% & 61%**
- **Red Cluster: Average Shopper (71% Cardholders)**





# Table of Contents

- Introduction
- Exploratory Data Analysis
- Customer Clustering Insights
- Conclusion





# Interpretations & Expectations

- **Cardholders spend more than Non-Cardholders**
- **Women more Brand-Loyal than Men**
- **Foot Traffic highest on the Weekends**
  - **Tuesday Anomaly**
- **50-65 Year Olds, the most Brand-Loyal Age Group**
- **Men prefer Weekends and Evenings to shop**



# Implications & Applications

- **Associating Customer Characteristics, Items, External Factors**
- **Understanding & Categorizing Customer Behaviors**
- **Specifically understanding Cardholding Customer Base**
- **Increasing Cardholding Customer Base**
- **Targeted Promotions**

**THANK YOU FOR LISTENING!**

**PLEASE FEEL FREE TO ASK  
ANY QUESTIONS!**



UNIVERSITÀ  
DEGLI STUDI  
DELL'AQUILA