**Basic EDA of Retail Data**

**Introduction:**

The purpose of this exercise is to analyze online retail data collected between December 1st, 2010 and December 9th, 2011. The data are extracted from an excel file that contains purchases from an online retail store.

**The following features have been extracted:**

* FREQUENCY (F) (number of purchases per customer)
* RECENCY (R) (days from the last purchase)
* LIFETIME (T) (days from the first purchase)
* MONETARY\_VALUE (M) (total revenue per customer)

Initial exploratory data analysis for these features included basic univariate statistics and graphs as well as correlation analysis and graphing between the variables extracted (bivariate analysis).

For simplicity, the rows with missing values (3,710 of 541,909, or 0.685%) were deleted in the excel spreadsheet. I have also not removed any outliers.

**Univariate Analysis**

**FREQUENCY (F) - Purchase Frequency:**

The average customer will make 45.2 card swipes across the year of recorded purchases. There are a few outliers that have made well over 100 purchases. Investigating the items that they purchase may give us insight on why they purchase so frequently **(Figure 1)**. Figure 2 shows that there is a large number of customers that have only made one purchase **(Figure 2)**.

A screenshot of a cell phone

Description automatically generated

**Figure 1: Boxplot of number of purchases for each customer**

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

**Figure 2: Most common purchase frequencies**

**RECENCY (R) -** **Number of Days from Recent Purchase:**

Average customer has made their most recent purchase 178 days from the last date recorded in the excel file **(Figure 3)**. Large number of customers that have made their most recent purchase in the last 30 days **(Figure 4)**.

A screenshot of a cell phone

Description automatically generated

**Figure 3: Boxplot of all days elapsed since each customer’s last purchase**

A screenshot of a cell phone

Description automatically generated

**Figure 4: Most common recent purchase days**

**LIFETIME (T) -** **Number of Days from First Purchase:**

The average customer has made their first purchase 182 days from the last day recorded in the excel file **(Figure 5)**. There is a large number of customers who have made their first purchase within 100 days of the first recorded date **(Figure 6)**.

A screenshot of a cell phone

Description automatically generated

**Figure 5: Boxplot of all days elapsed since each customer’s first purchase**

A screenshot of a cell phone

Description automatically generated

**Figure 6: Most common first purchase days**

**MONETARY\_VALUE (M) -** **Total Revenue per Customer:**

Large number of customers who have spent little money. Few customers have spent hundreds of thousands in revenue which are considered outliers. Comparing items purchased between high and low spenders can give us additional insight on customers habits **(Figure 7)**.

Around 5% of the total revenue is spent by the top 2 customers

Around 10% of the total revenue is spent by the top 4 customers

Around 25% of the total revenue is spent by the top 18 customers

Around 50% of the total revenue is spent by the top 183 customers

Around 75% of the total revenue is spent by the top 841 customers**A screenshot of a cell phone

Description automatically generated**

**Figure 7: Money spent per customer sorted in decreasing order. Applied square root to y axis as feature engineering and x axis are the indices of customers ranked by their revenue**

**Bivariate Analysis**

**Purchase Frequency vs. Number of Days from Last Purchase:**

Figure 8 indicates that there is some kind of a correlation between a higher purchase frequency and a lower number of days from the last purchase. This could be interpreted as customers who quit purchasing earlier made fewer purchases **(Figure 8)**.

A screenshot of a cell phone

Description automatically generated

**Figure 8: Relationship between purchase frequency and days from the most recent purchase**

**Purchase Frequency vs. Number of Days from First Purchase:**

Figure 9 indicates that there is some kind of a correlation between a higher purchase frequency and a higher number of days from the first purchase. Customers who started purchasing early tend to spend more money **(Figure 9)**.

A screenshot of a cell phone

Description automatically generated

**Figure 9: Relationship between purchase frequency and days from the first purchase**

**Purchase Frequency vs. Total Revenue Accumulated per Customer:**

There appears to be a strong correlation between the two variables. There are few outliers where few purchases were made but lots of money was spent. These purchases may have been expensive items **(Figure 10)**.

A close up of a map

Description automatically generated

**Figure 10: The relationship between frequency of purchases and money spent**

**Number of Days from Last Purchase vs. Number of Days from First Purchase:**

The points near the diagonal line indicate that there are many customers who have only purchased once or a couple of times. Ideally, all of the dots would be on the top left showing the customers who started early and made their last purchase recently **(Figure 11)**.

A screenshot of a cell phone

Description automatically generated

**Figure 11: Days from most recent purchase to first purchase by customer**

**Revenue per Customer vs. Number of Days from Last Purchase:**

Correlation appears to exist between higher revenue and lower number of days from last purchase. Some outliers have a high revenue but a high number of days from their last purchase. Many who have spent hundreds of thousands of dollars have a low number of days from the last purchase **(Figure 12)**.

A screenshot of a cell phone

Description automatically generated

**Figure 12: Relationship between number of days from last purchase and the total revenue by customer**

**Revenue per Customer vs. Number of Days from First Purchase:**

Higher revenue per customer is associated with a higher number of days from the first purchase. Many who have spent hundreds of thousands of dollars have a high number of days from the first purchase **(Figure 13)**.

A screenshot of a social media post

Description automatically generated

**Figure 13: Relationship between number of days from first purchase and the total revenue by customer**