**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 the sake of simplicity, I have used list wise deletion for missing values in the excel spreadsheet.

I have also not removed any outliers.

**Univariate Analysis**

**Purchase Frequency (Figure 1 & 2):**

I have concluded that there is a large cluster of customers that have only made one purchase. The average customer will make around 30 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.

**Number of Days from Recent Purchase (Figure 3 & 4):**

Large cluster of customers that have made their most recent purchase in the last 30 days. Average customer has made their most recent purchase ~175 days from the last date recorded in the excel file.

**Number of Days from First Purchase (Figure 5 & 6):**

There is a large cluster of customers who have made their first purchase within 100 days of the first recorded date. The average customer has made their first purchase ~205 days from the last day recorded in the excel file.

**Total Revenue per Customer (Figure 7):**

Large cluster 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 insight.

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

Mean is 2178

Median is 677

Since the median is much lower than the mean, this tells me the distribution is skewed right.

**Univariate Graphs:**

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

A screenshot of a cell phone

Description automatically generated

**Figure 2: Most common purchase frequencies**

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

**Figure 3: Boxplot of all days elapsed since each customers last purchase**

A screenshot of a cell phone

Description automatically generated

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

A screenshot of a cell phone

Description automatically generated

**Figure 5: Boxplot of all days elapsed since each customers first purchase**

A screenshot of a cell phone

Description automatically generated

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

A screenshot of a cell phone

Description automatically generated

**A screenshot of a cell phone

Description automatically generatedFigure 7: Revenue of** **customers. Applied square root to y axis as feature engineering**

**Bivariate Analysis**

**Purchase Frequency vs. Number of Days from Last Purchase (Figure 8):**

Observations lead me to believe that there is some kind of a correlation between a higher purchase frequency and a lower number of days from the last purchase.

**Purchase Frequency vs. Number of Days from First Purchase (Figure 9):**

Observations lead me to believe that there is some kind of a correlation between a higher purchase frequency and a higher number of days from the first purchase.

**Purchase Frequency vs. Total Revenue Spent per Customer (Figure 10):**

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

**Number of Days from Last Purchase vs. Number of Days from First Purchase (Figure 11):**

Lots of customers who have only purchased once and never returned. Ideally, all of the dots would be in the top left corner.

**Revenue per Customer vs. Number of Days from Last Purchase (Figure 12):**

Correlation between higher revenue and lower number of days from last purchase. Some outliers that have a high revenue but have 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.

**Revenue per Customer vs. Number of Days from First Purchase (Figure 13):**

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.

**Bivariate Graphs:**

A screenshot of a cell phone

Description automatically generated

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

A screenshot of a cell phone

Description automatically generated

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

A close up of a map

Description automatically generated

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

A screenshot of a cell phone

Description automatically generated

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

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**

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**