**Assignments: ALY6080 90325 Integrated Experiential Learn SEC 03 Summer 2023 CPS [BOS-1-HY]**

**Module 3 Assignment — XN Project: Digging Deeper**

Logo

Description automatically generated

**NORTHEASTERN UNIVERSITY**

**College of Professional Studies, Boston, MA, 02215.**

**Submitted by**

Zihan Ma

ma.zihan1@northeastern.edu

**Instructor**

 Prof. Valerie A. Atherley

**Date**

07/23/2023

**Exploratory Data Analysis of the XN Project**

**Abstract:**

This document outlines an in-depth Exploratory Data Analysis (EDA) conducted on two datasets to inform future project planning. Preliminary findings, methodologies, and data preparation tasks are discussed, underscoring the critical role of quality data in generating insightful results.

1. **Introduction**

We are exploring two datasets: 'Locally\_Inspired\_monthly\_inventory\_2023-05-01.xlsx,' which shows inventory changes in May, and 'Inspired Order\_1'1'23-5'31'23.xlsx,' documenting orders placed from January to May. These datasets provide a wealth of information for financial analysis, forecasting, sales analysis, and customer behavior studies.

1. **Methodology**

The EDA process consists of four steps: applying summary() and str() functions, identifying unique column values, ascertaining number distributions for numeric columns, and investigating missing data across both datasets.

1. **Results:**

***monthly\_inventory dataset:***

See the screenshot result in Appendix A

1. The summary() and str() functions revealed minimal missing data, indicating the dataset's generally high quality. (See A 1.1)
2. The unique values demonstrate a wide distribution across almost all columns, hinting at diverse inventory types and transactions. (See A 1.2)
3. While most number distributions center on lower prices, some outliers were temporarily omitted for clarity. Some negative or zero values require further clarification. (See A 3.1, 3.2)
4. Missing values, highlighted below, will be addressed in upcoming meetings. (See A 4.1)

***Order dataset:***

See the screenshot result in Appendix B

1. The Order dataset exhibits a substantial amount of missing data, necessitating thorough cleaning. (See B 1.1~1.3)
2. Unique value exploration unearthed intriguing results: (See B 2.1~2.3)
   1. Financial Status: Nearly all orders are fully paid (1681 out of 1687), indicating effective payment collection mechanisms and potentially high customer satisfaction as refund requests are minimal. This is crucial for maintaining positive cash flow and profitability.
   2. Fulfillment Status: Almost all orders are fulfilled (1685 out of 1687), which suggests effective order processing and inventory management. High fulfillment rate is a key factor in customer satisfaction and can greatly impact customer retention and reputation.
   3. Accepts Marketing: More than half of the customers accept marketing (1043 out of 1687), providing a good opportunity for promoting new products, offers, or upselling. A receptive audience for marketing efforts can support sales growth and customer engagement.
   4. Source: Most orders are made through point-of-sale (POS) systems (1598 out of 1687), indicating a strong brick-and-mortar presence. This could inform marketing and sales strategies, such as focusing on in-store promotions or improving the in-store customer experience.
3. Visualization of Subtotal, Shipping, Taxes, and Total columns show a trend toward smaller transaction amounts, with a few outliers indicating larger transactions. Numerous zero-values, particularly in Shipping, might reflect free services or promotions. Missing data must be scrutinized further. (See B 3.1~1.3)
4. While the high volume of missing data initially appears concerning, some could be meaningful, such as unregistered VIP customers. However, most missing data requires clarification from the client. (See B 4.1~1.3)
5. **Discussion**

Our EDA has enabled us to unearth several valuable insights and potential areas for further investigation. It has highlighted issues that need addressing and possible hypotheses to test during subsequent analysis stages.

1. **Conclusion**

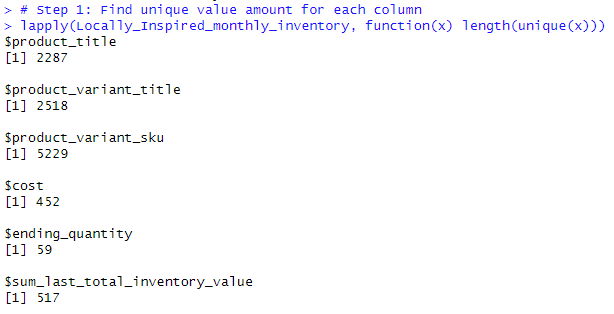
The EDA has provided an illuminating overview of the data and highlighted potential areas of interest for further exploration. Our team is prepared to dive deeper into the analysis, addressing the data gaps and continuing our exploratory journey.

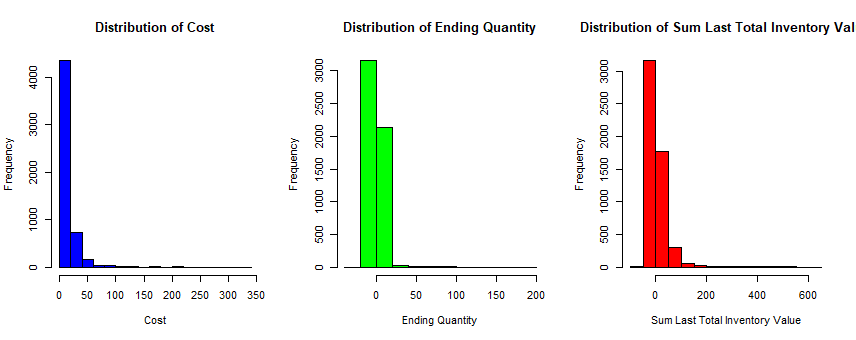
**Appendix A**

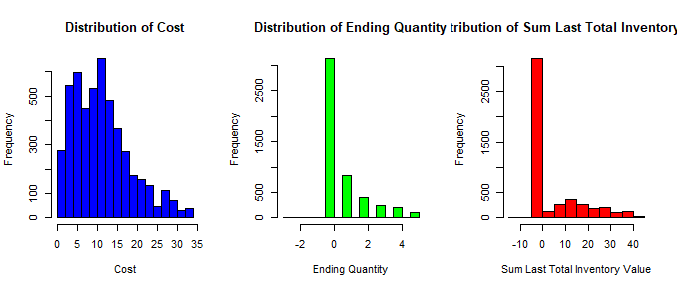
**Screenshort result of monthly\_inventory dataset**

A 1.1A screenshot of a computer

Description automatically generated

A 2.1  


A 3.1  


A 3.2  


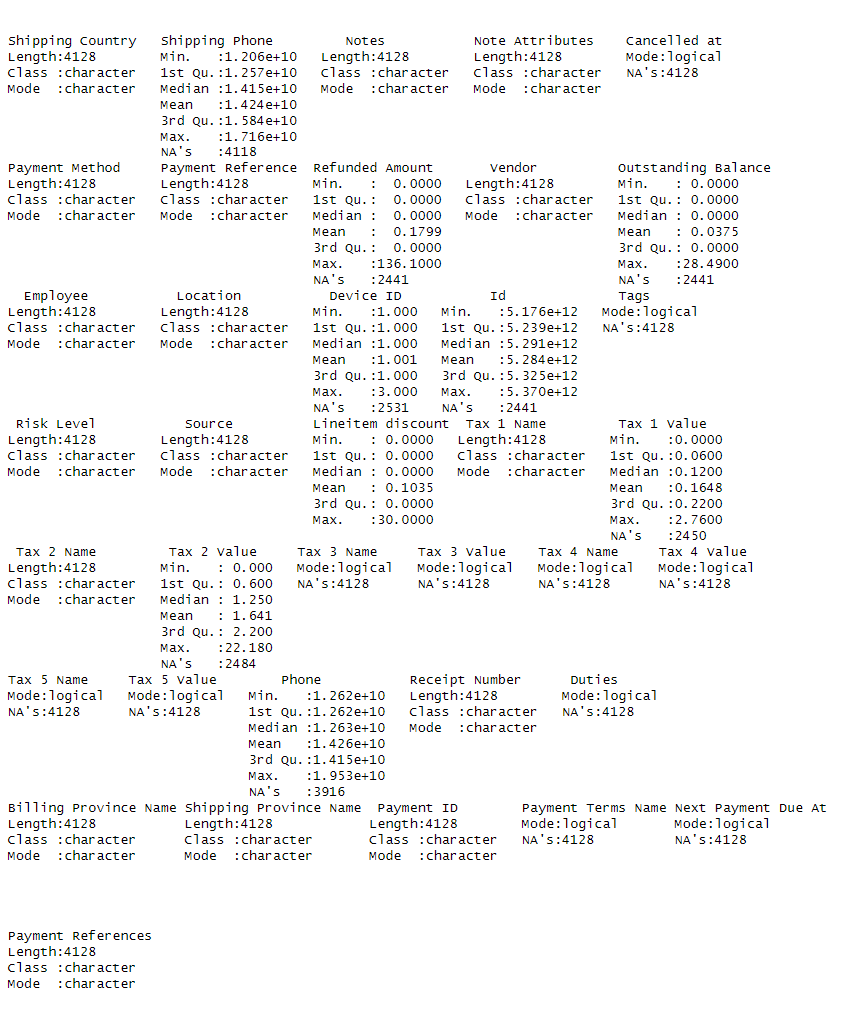
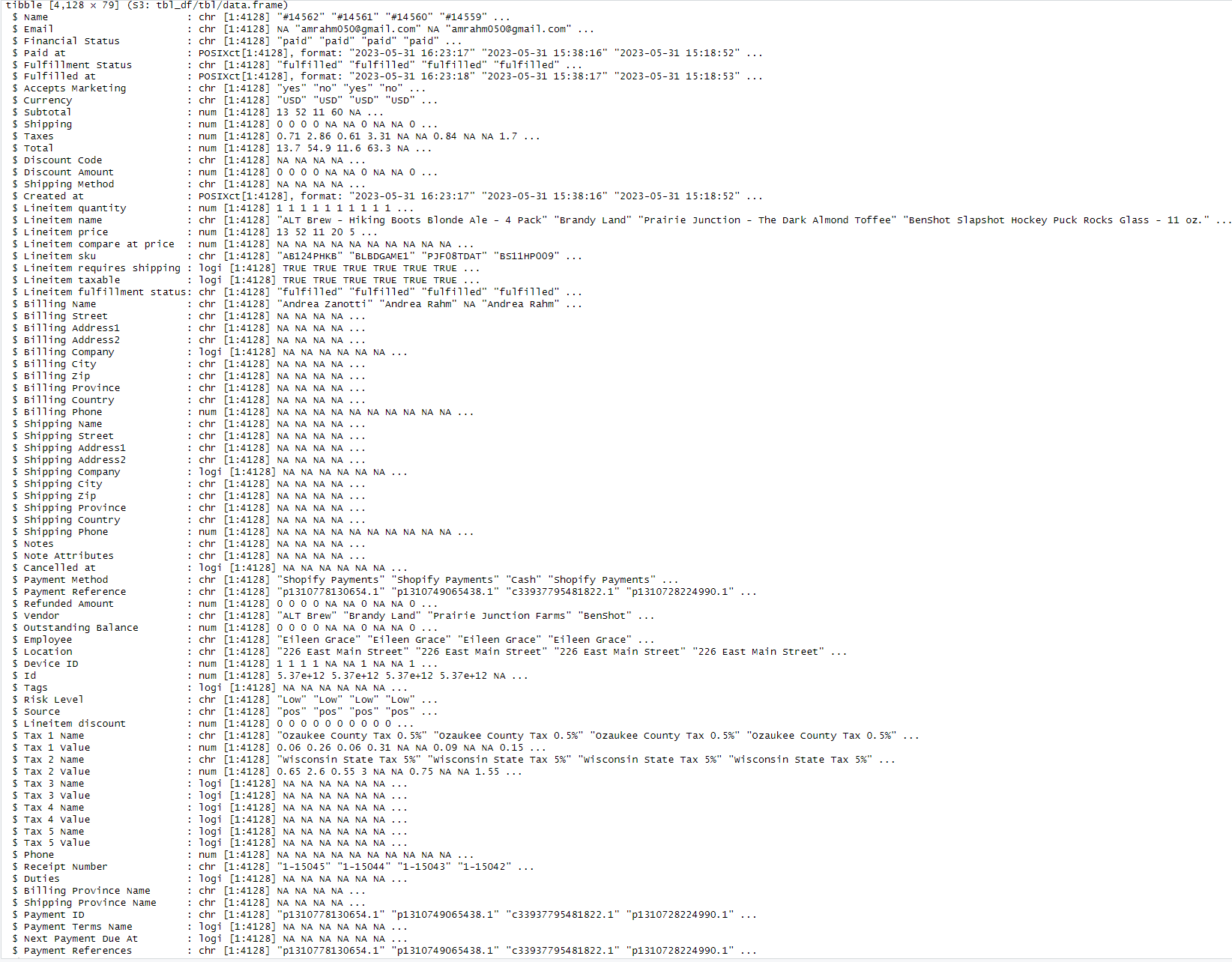
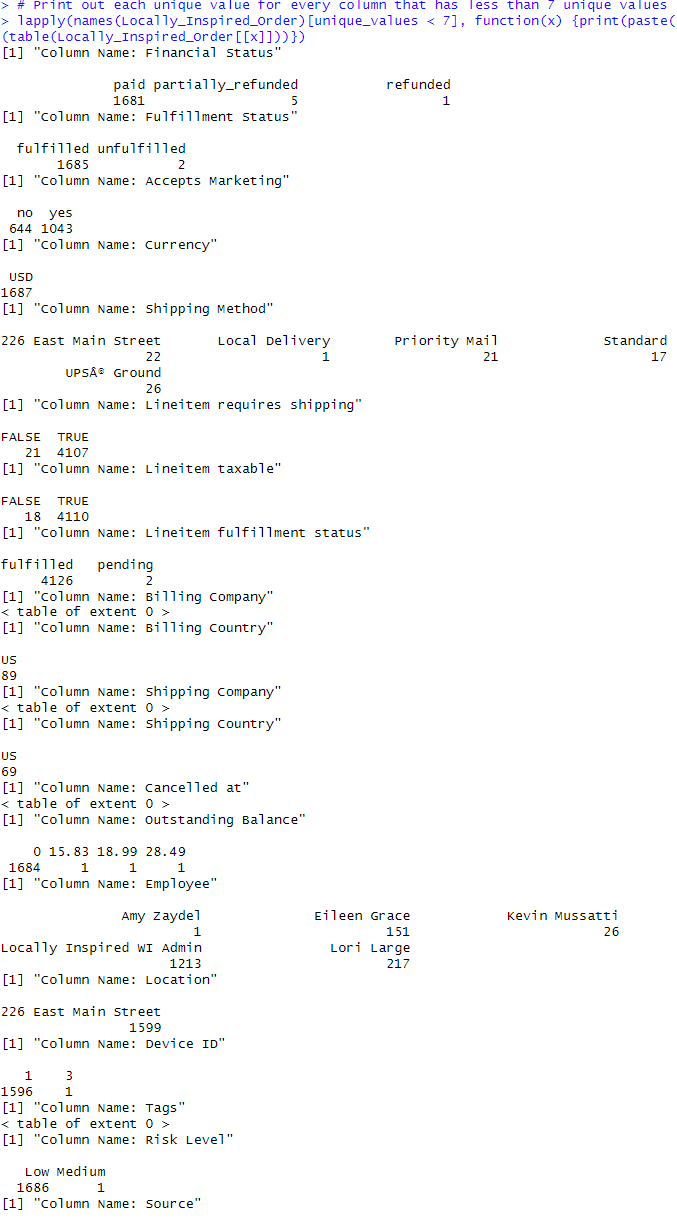
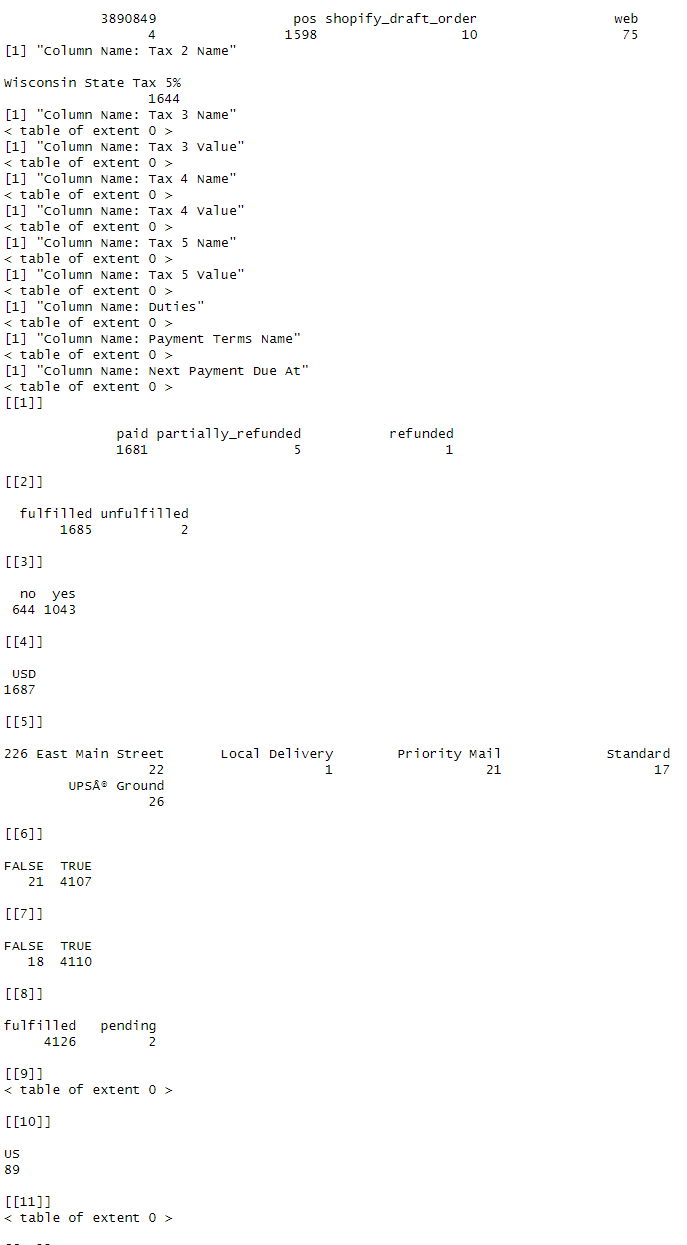
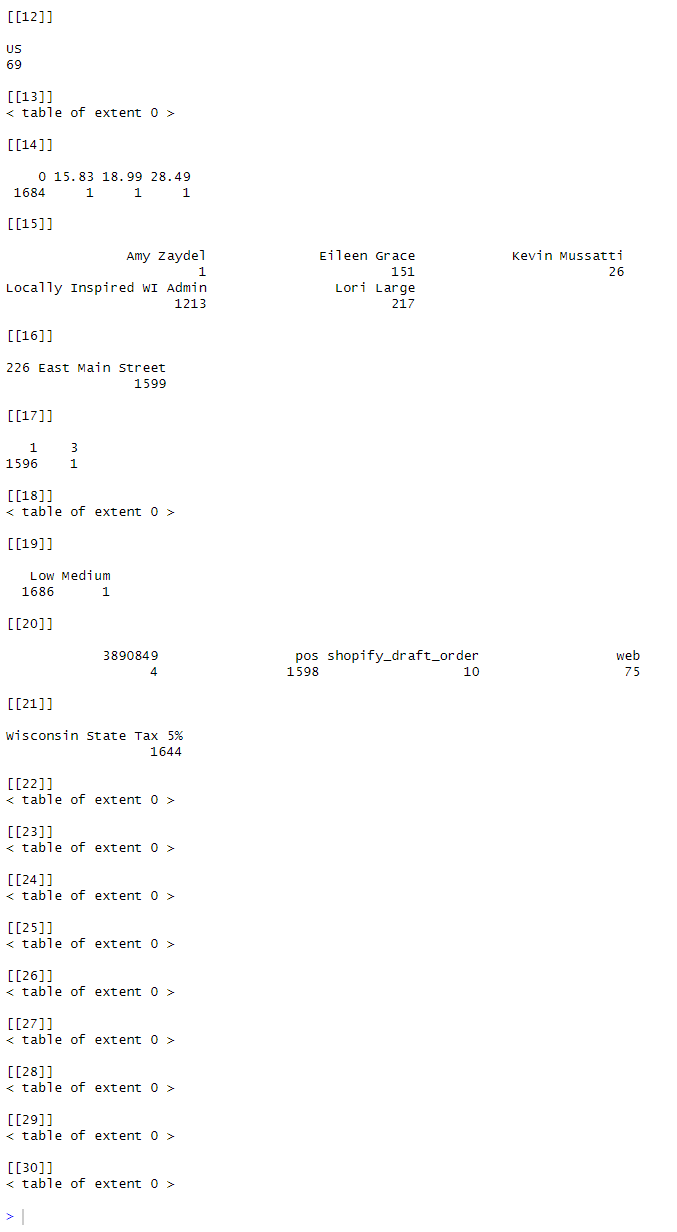
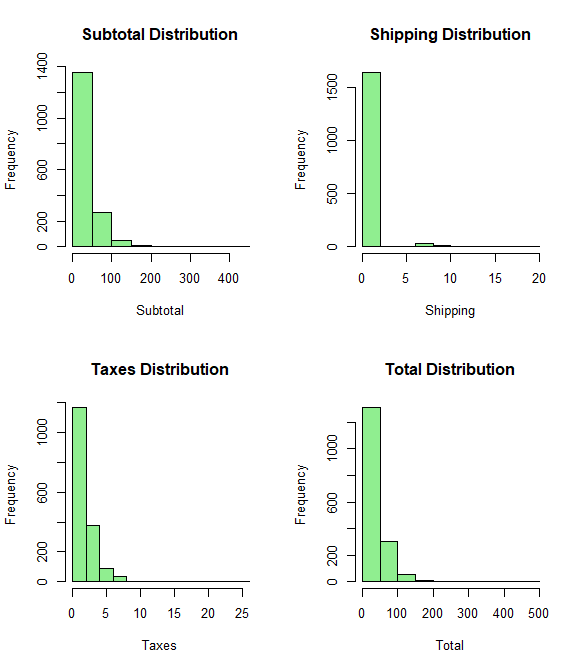
A4.1  
A screen shot of a computer

Description automatically generated

**Appendix B**

**Screenshot result of Order dataset**

B 1.1  
A screenshot of a computer screen

Description automatically generated  
B 1.2  
  
B 1.3  
  
B 2.1  
  
B 2.2  
  
B 2.3  
  
B 3.1  
  
B 4.1  
A screen shot of a computer code

Description automatically generated