

The background features two large, overlapping, curved lines. One line is a light green color and the other is a light blue color. They are positioned in the top-left and bottom-right corners of the slide, framing the central text.

STERLING E-COMMERCE COMPANY

Exploratory Data Analysis Project

Introduction

- Sterling E-Commerce is a one-stop online shopping destination for a wide range of high-quality products.
- The company's mission is to provide their customers with the best shopping experience possible, with a focus on quality, affordability, and convenience.
- Products across various categories such as
 - health & sports,
 - men's and women's fashion,
 - computing,
 - entertainment,
 - mobiles & tablets,
 - appliances,
 - beauty & grooming,
 - home & living,
 - soghaat,
 - school & education,
 - books, and more.

Problem Statement

- The Director of Sterling contacted our team (of Data Scientists) because they are interested in leveraging the power of their data to gain insights into the business and improve their efficiency.
- Sterling wants to better understand their customers' needs, preferences, and behavior. Specifically, want to identify any patterns or trends.
- Sterling believes that leveraging the power of data can help them make more informed decisions and boost effectiveness.
- They want to understand their data, so they can optimize their product offerings, streamline operations, and enhance Sterling's customer experience

Data Dictionary

- **Category** – The category of goods
- **City** – The city where customers are ordering from
- **County** – The county where customers are ordering from
- **Cust Id** – Customer ID
- **Customer Since** – The date when the customer first order.
- **Date of Order** – The date when the customer placed an order.
- **Full Name** – Customer full name
- **Gender** – (Male, Female)
- **Item Id** – Item ID
- **Order Id** – Order ID
- **Payment Method** – The payment platform the customer used
- **Place Name** – The exact location where the customer are ordering from
- **Ref Num** – Reference Number
- **Region** – The region where the customer are ordering from
- **State** - The state where the customer are ordering from
- **User Name** – The customer's username
- **Zip** – Customer's zip code
- **Qty Ordered** – The quantity of goods ordered
- **Total** – Total amount paid by customer

Data Set Given

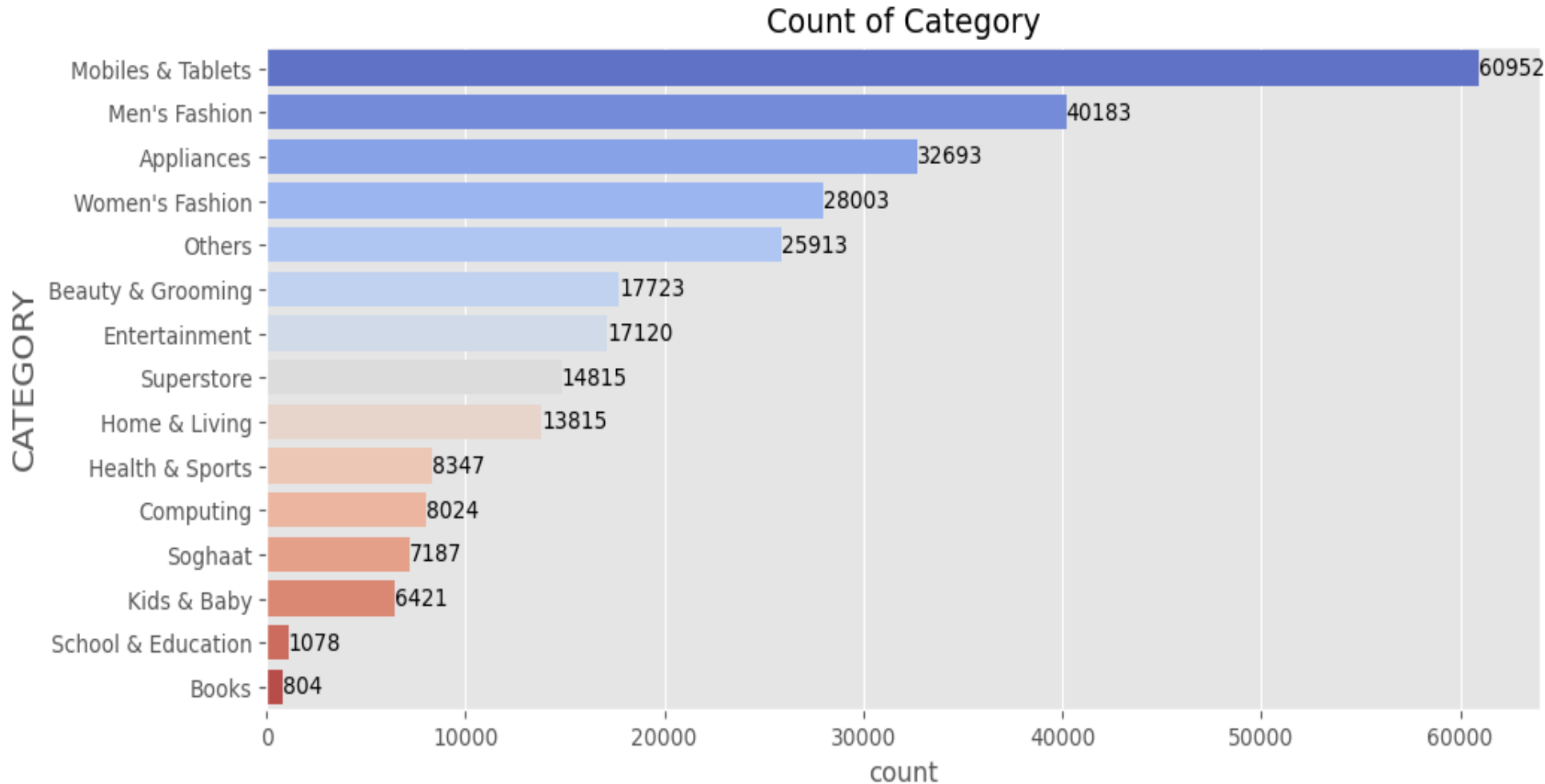
- Rows : 283083
- Columns :19
- Datatypes : Floats (2), Integers(5), Objects(10), Date(2)
- Duplicates : 0
- Missing Values: 5 (only on the order_id) – drop the entire rows
- Zeros on the Total : 18924 rows × 27 columns – replaced the 0.0 with Nan and transformed with the median grouping by the category.

FEATURE ENGINEERING (Creating 8 More columns)

- Split the Date into Quater Month, Month_name and Days
- Create a New Column for seasons using the Months and 'if' functions
- Create a Customer loyalty column using the 'customer since' column

Exploratory Data Analysis

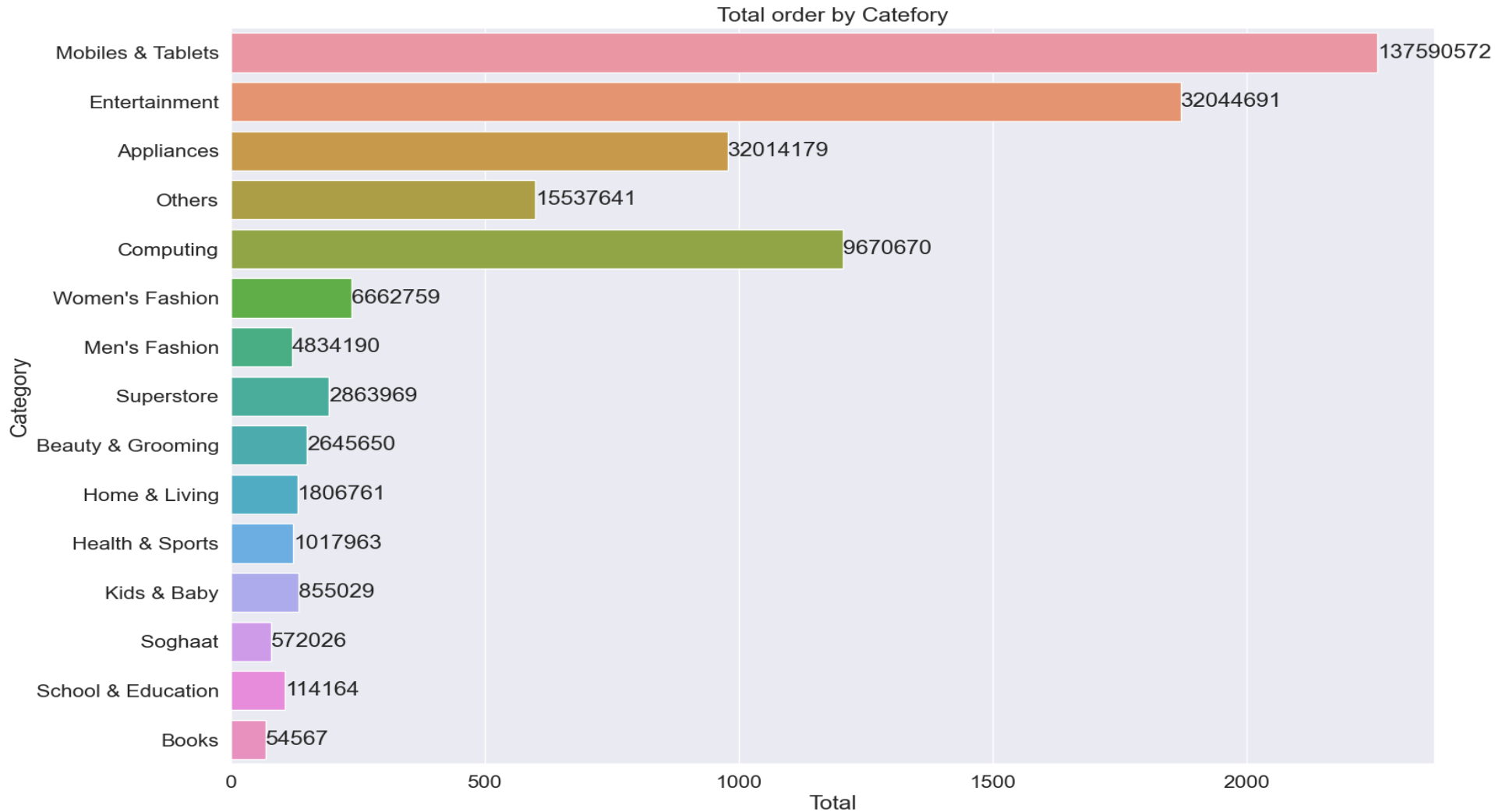
Univariate, Bivariate and Multivariate Analysis was used to explore the Data after cleaning



Insight: The Mobile & Tablets, men's fashion and Appliances are the top 3, while Books, School/education items are the least items ordered

Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Data after cleaning

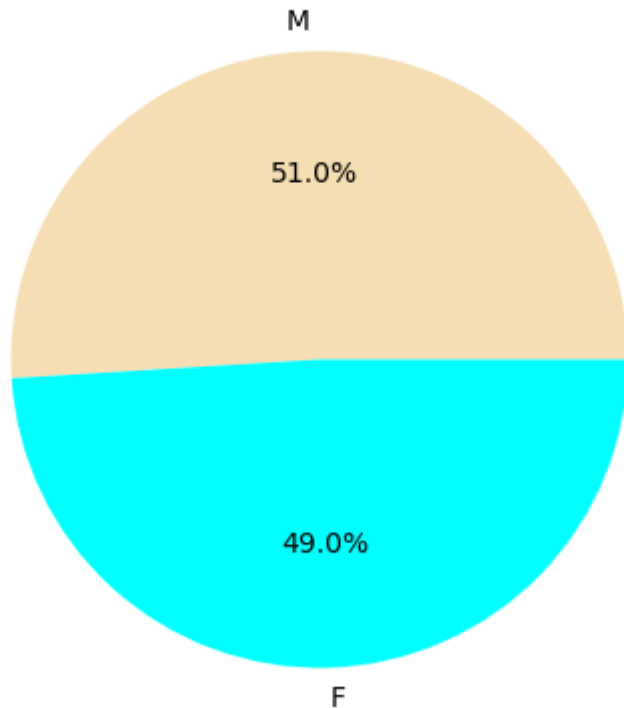


Insight:
Mobile/Tablets,
Entertainment,
Appliances and
computing are mostly
the products ordered

Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Data after cleaning

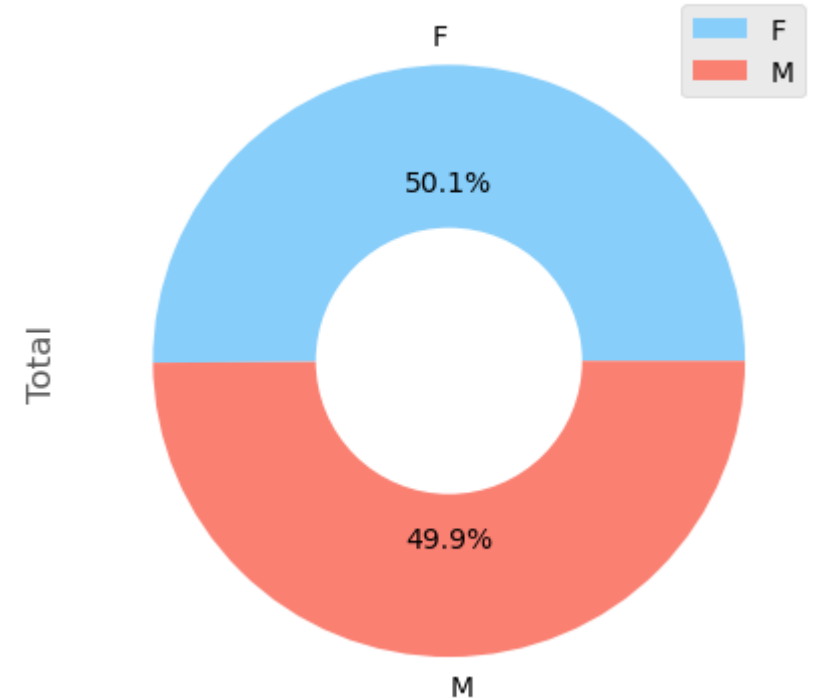
Gender Distribution



Insight:

We have an almost equal distribution in the Gender, with male slightly higher than female.

Pie Chart of Total order by Gender

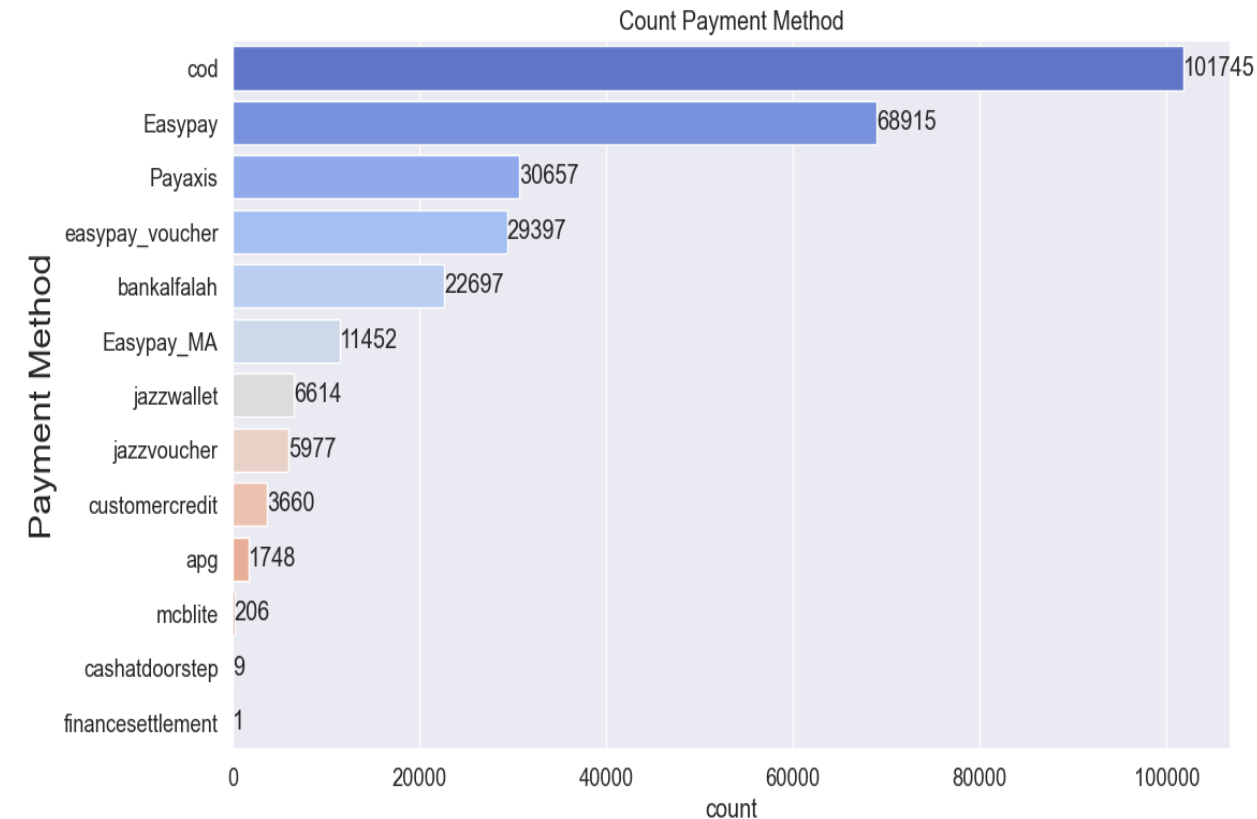


Insight:

The Female Gender have the most order, although the difference is infinitesimal.

Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Data after cleaning

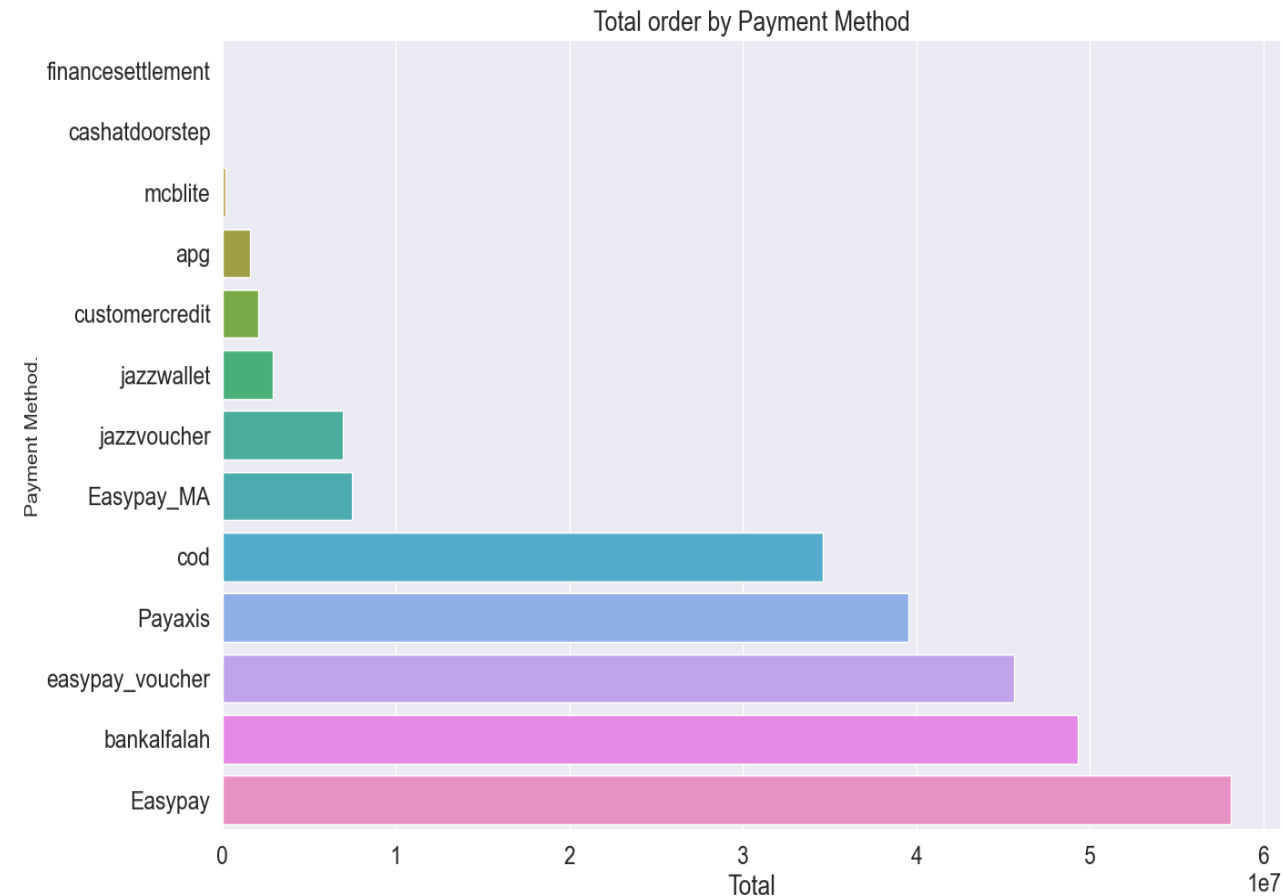


Insight:

Cod, easy pay are outstanding in our dataset and mostly used for customer payment, and finance settlement, cash at doorstep, and mcbllite are the least used

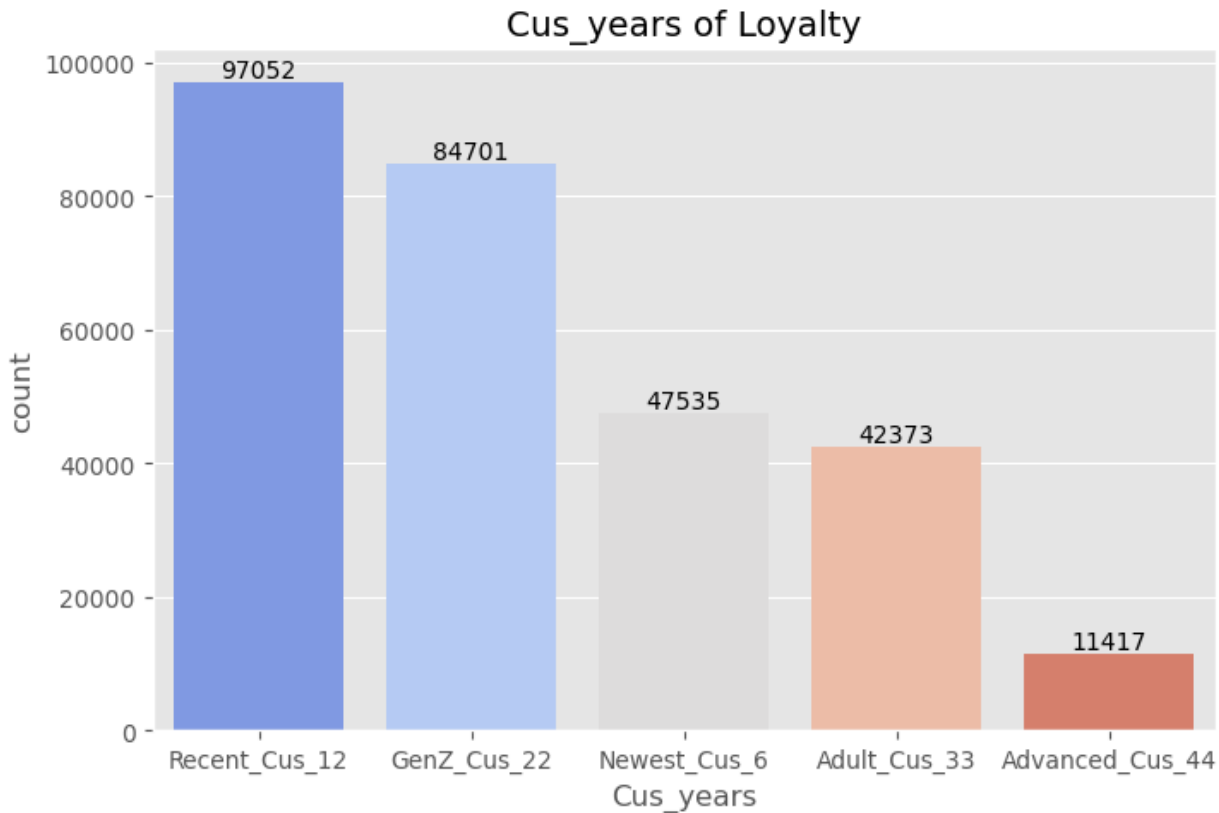
Insight:

Easypay, Bankafalah and easypay_voucher are the most payment methods used against the Total orders.



Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Data after cleaning



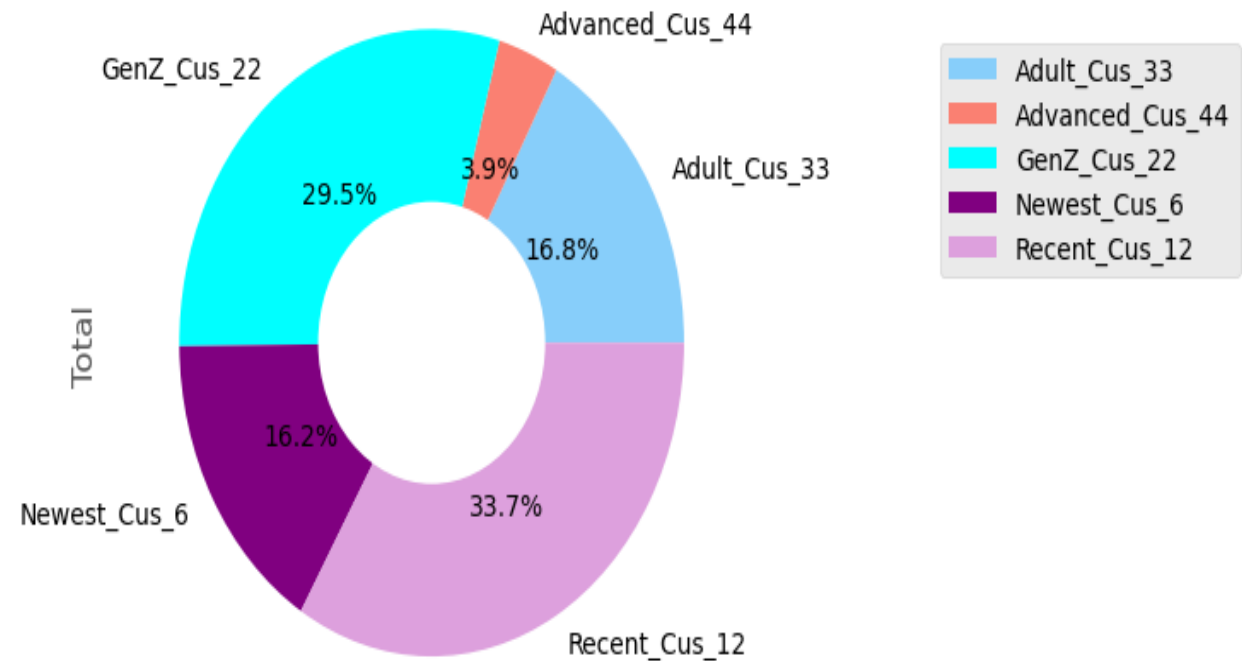
Insight:

Recent (12yrs) and GenZ (22) customers are more in numbers, with advanced customers(44yrs) as the least

Insight:

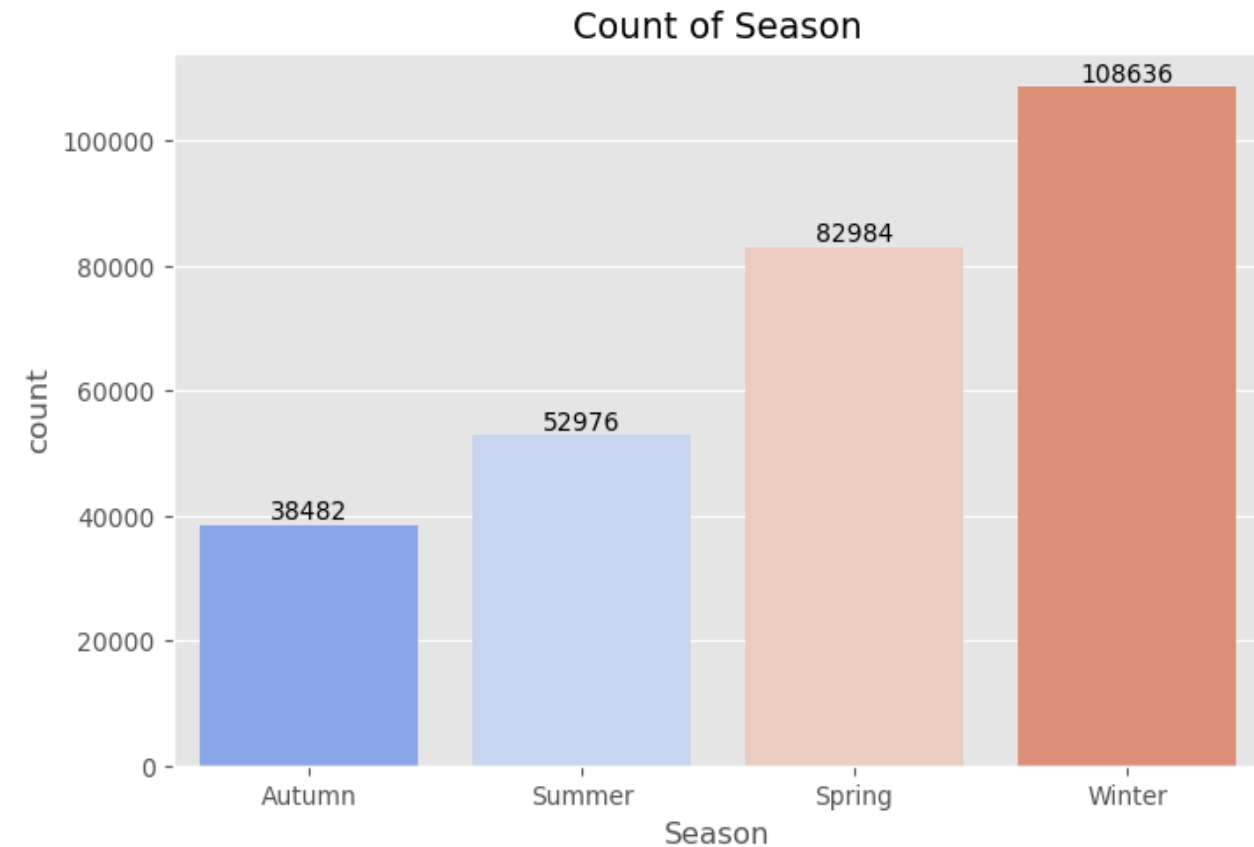
The recent_cus(12yrs) are most followed by Genz_customers (22yrs)

Pie Chart of Total order by Cus_Years



Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Date after cleaning



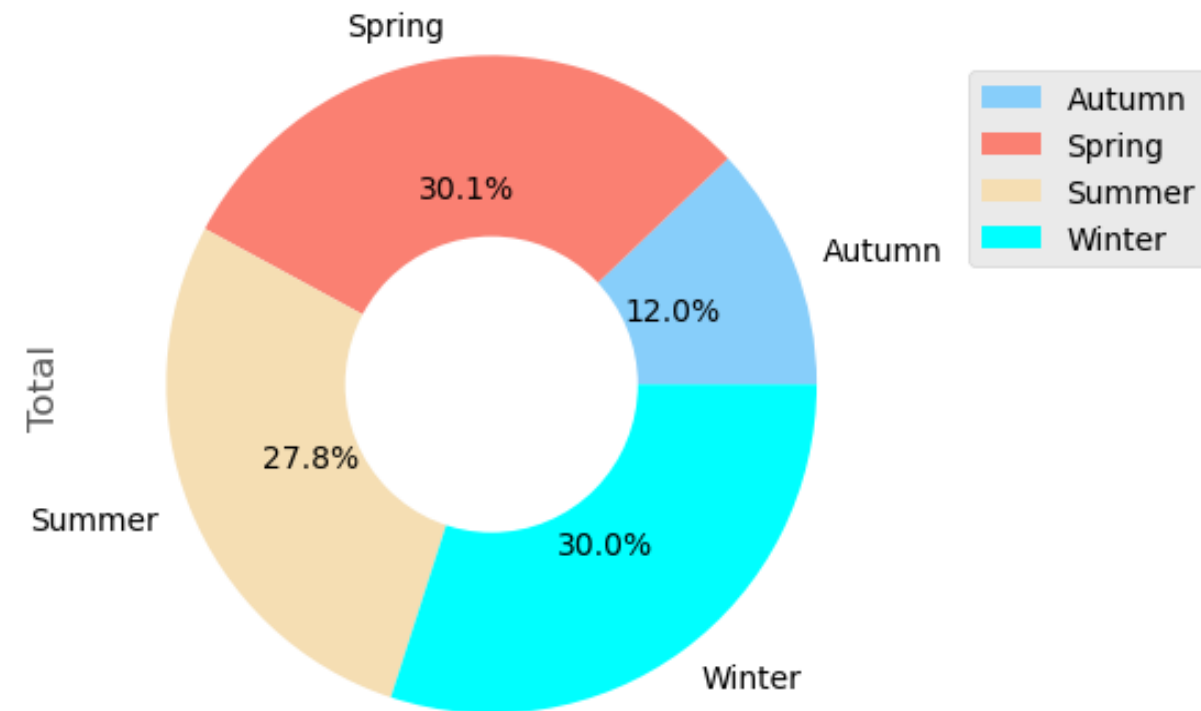
Insight:

Winter and Spring are most prominent and possibly the busiest for sterling E-commerce while Summer and Autumn are light

Insight:

Most of the orders are made in Spring and Winter

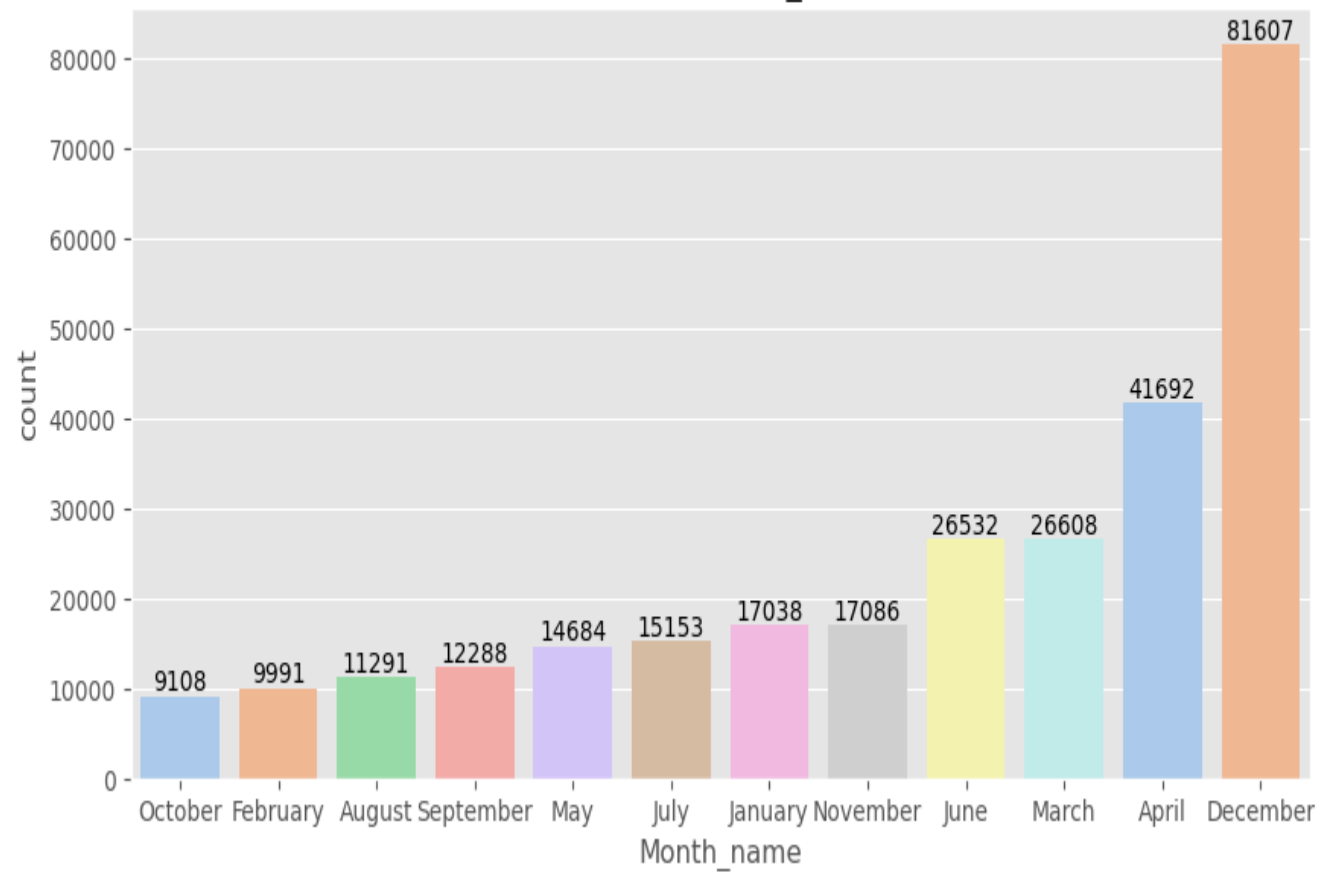
Pie Chart of Total order by season



Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Date after cleaning

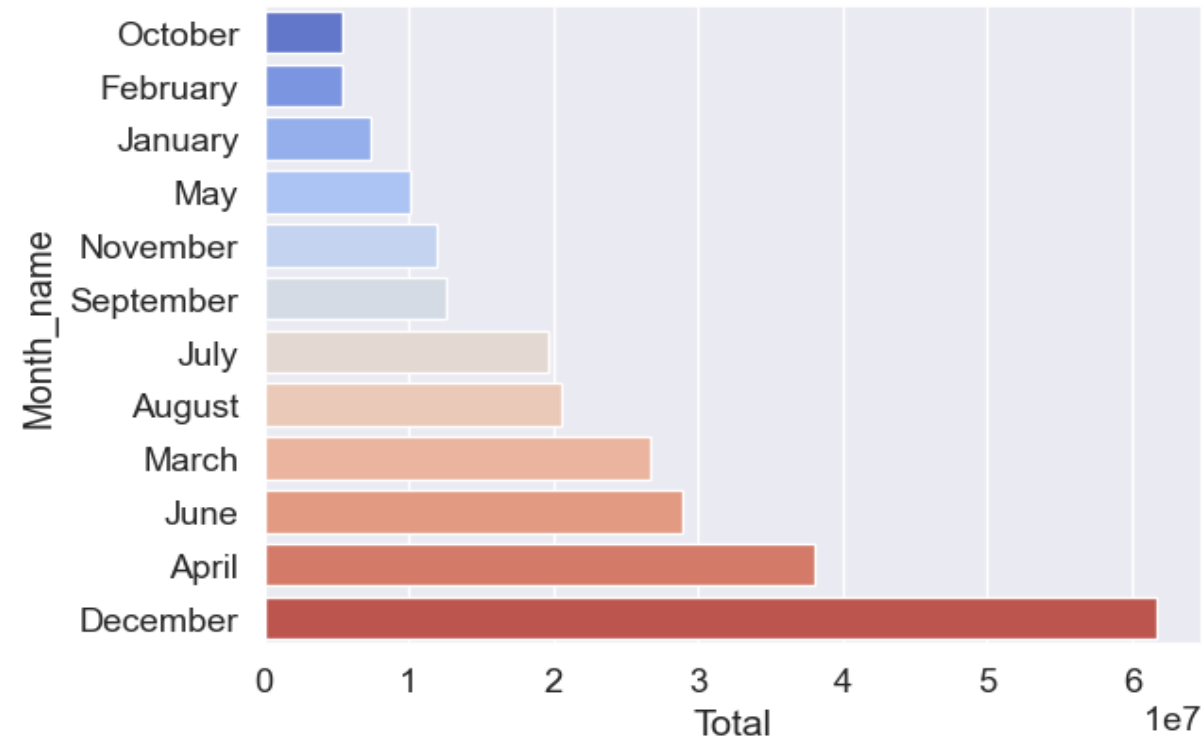
Count of Month_name



Insight:

December, April and June have the highest Total orders, while October, January and February are the least

Total order per month



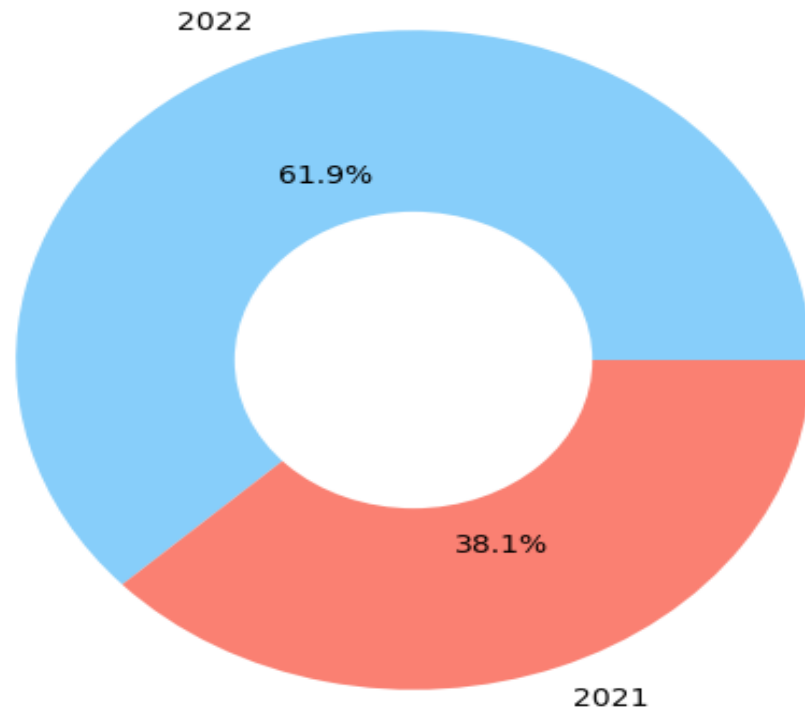
Insight:

December is Winter, while march and April are spring in most seasons and this is the period Sterling E-commerce experience most of their orders. October, February are the most quiet periods.

Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Data after cleaning

Count of Year Distribution

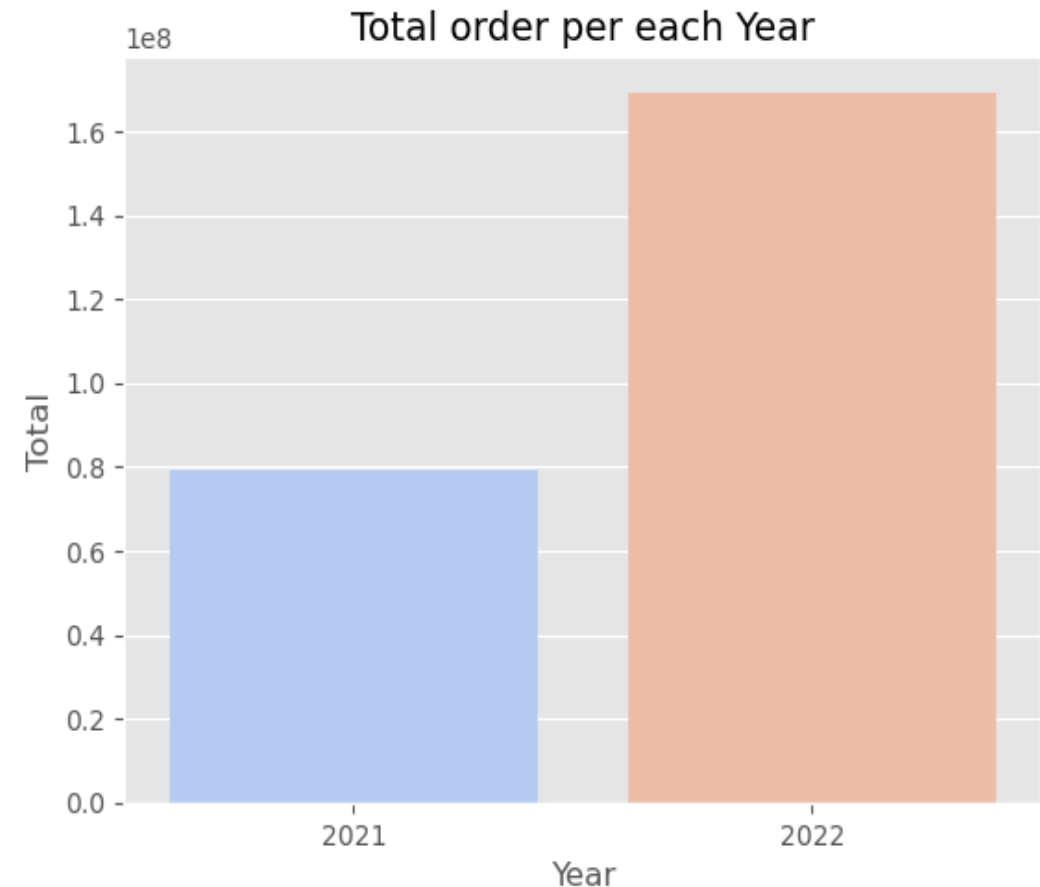


Insight:

Year 2022 has more activities than 2021 in our dataset

Insight:

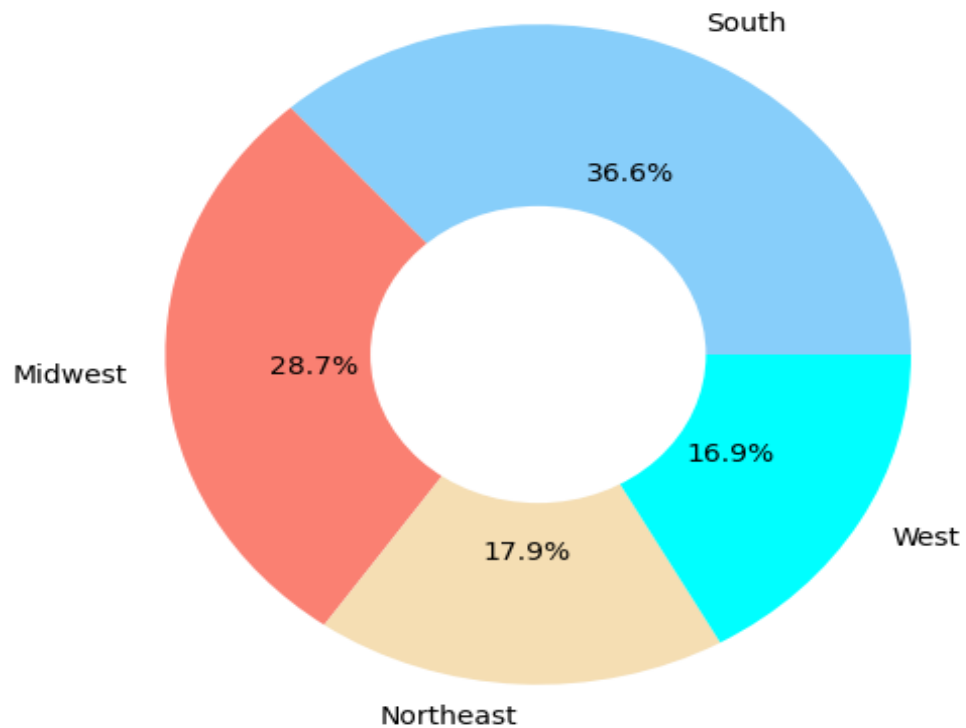
We have more order in 2022 than the previous year of 2021



Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Data after cleaning

Count of Region/Distribution



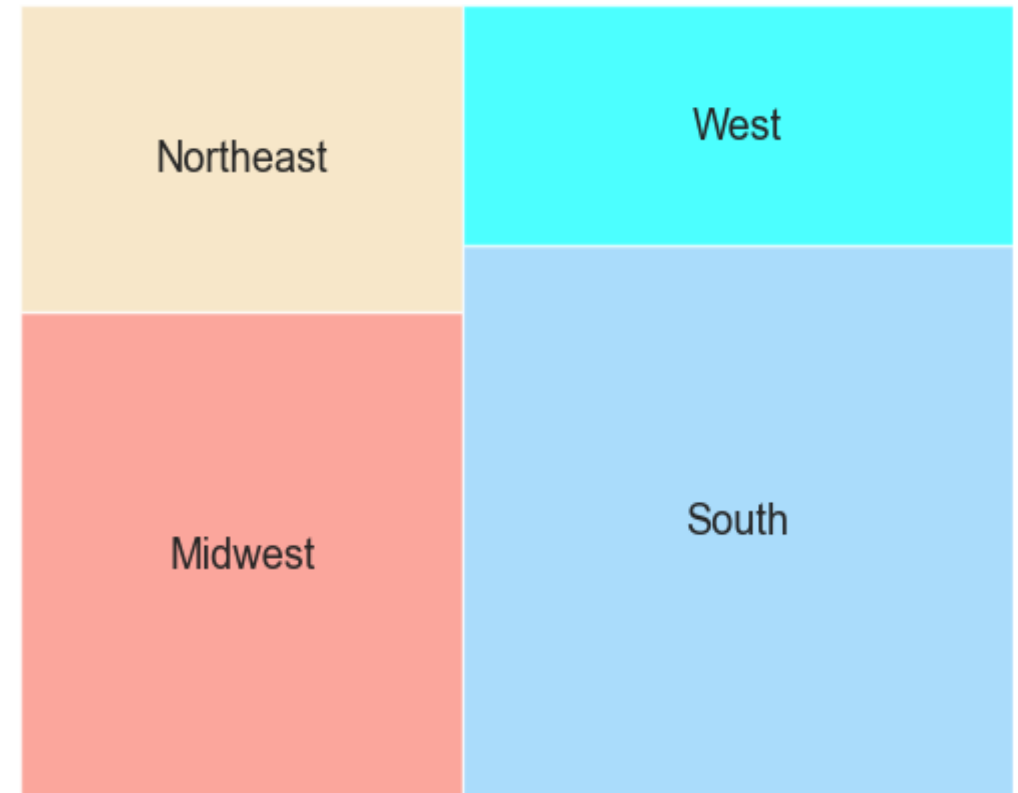
Insight:

South region followed by Midwest region have the most number of customers, with the least being the west.

Insight:

The total order mostly are coming also from the south and then midwest, the Northeast and west remains the least

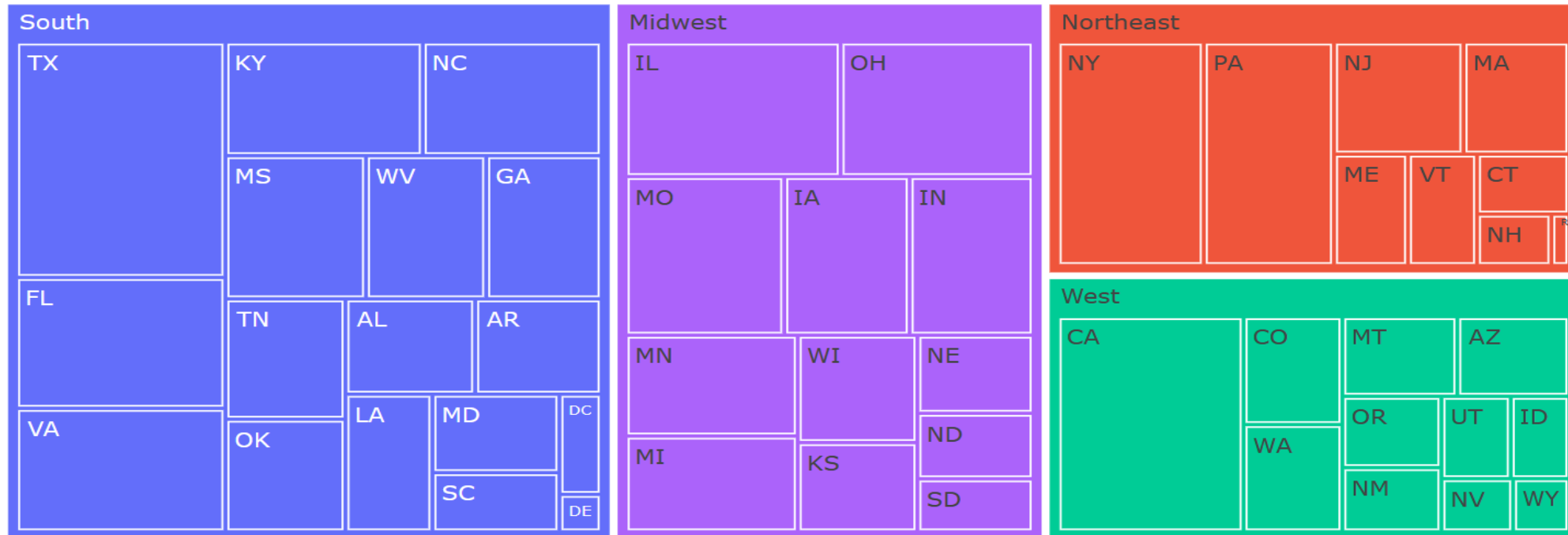
Total order by Region



Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Data after cleaning

Overview of Total amount paid by Customer by Region & State



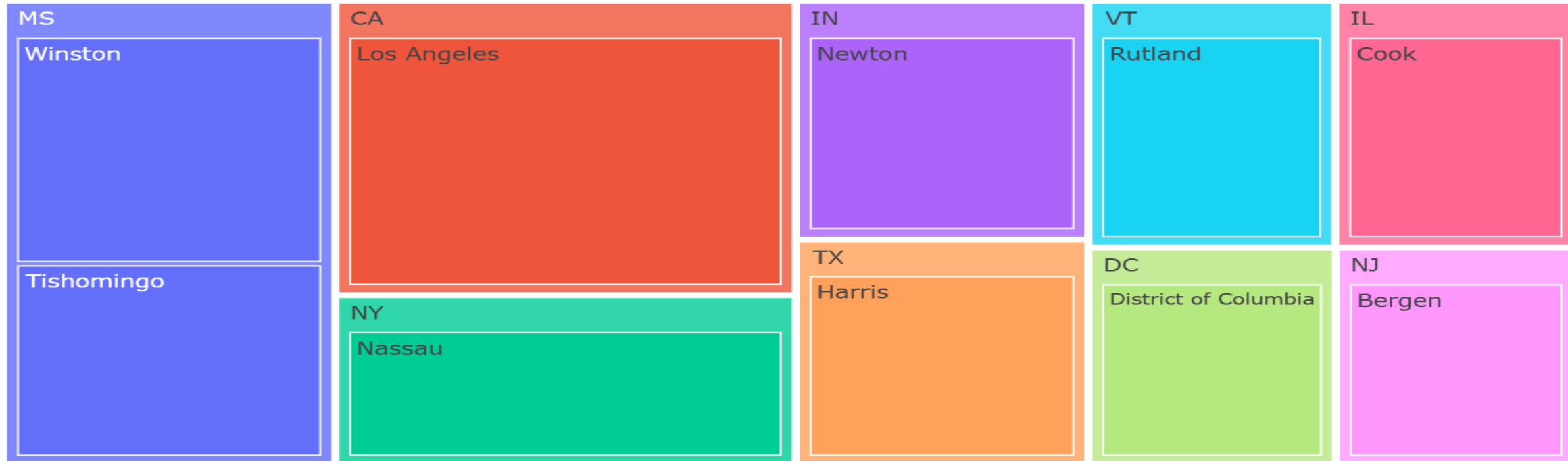
Insight:

Further seeing the details in the total order, we can see the cities with highest/lowest order by state in the various regions. South have Texas and Florida, west have California as the highest sales

Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Data after cleaning

Overview of Total order by Customer by Top_10 County

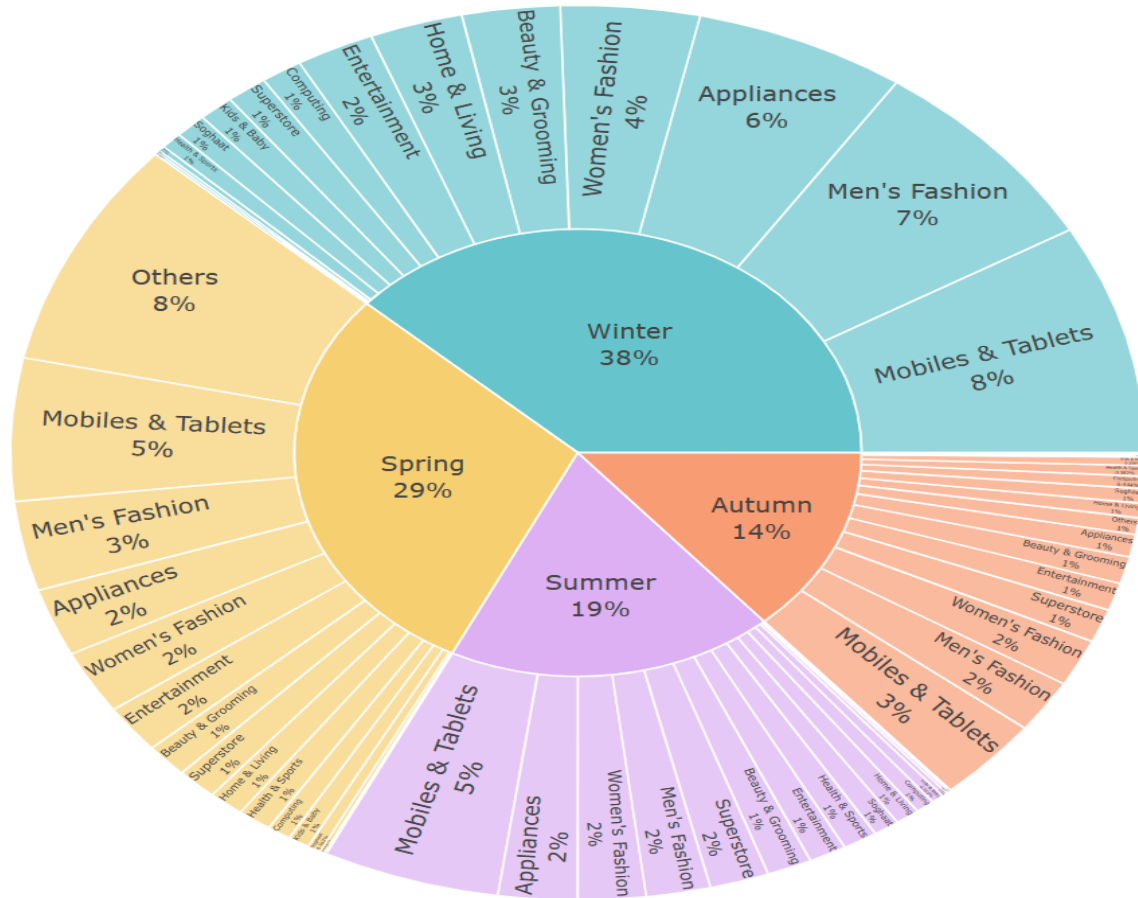


Insight:

L.A. County returns the highest Total of order and by research is the most populous county in the United States. 9,861,224 residents estimated in 2022

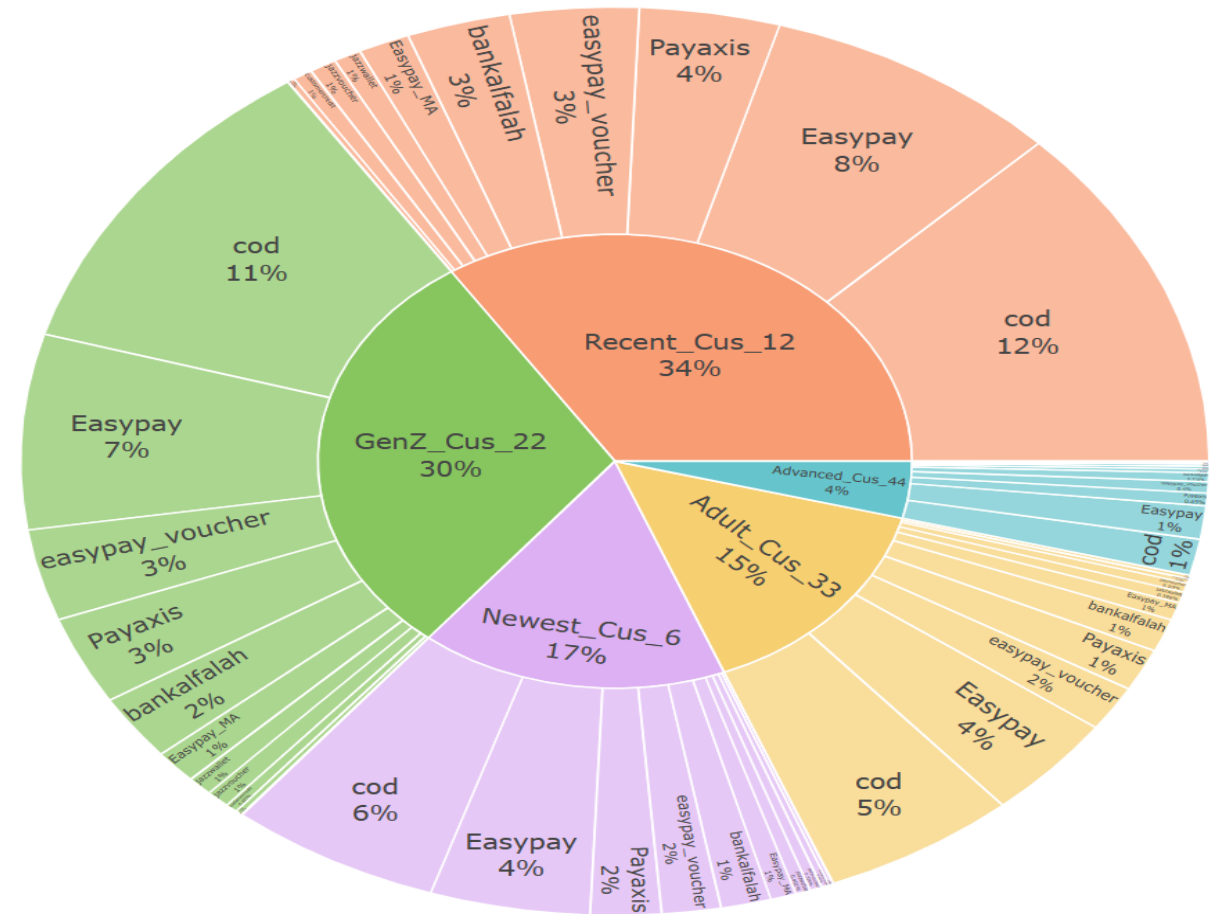
Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Date after cleaning



Insight:

Most of the orders are in the Winter, and Spring and still remains Mobiles & tablets, men and women fashions are also part of the top orders across seasons

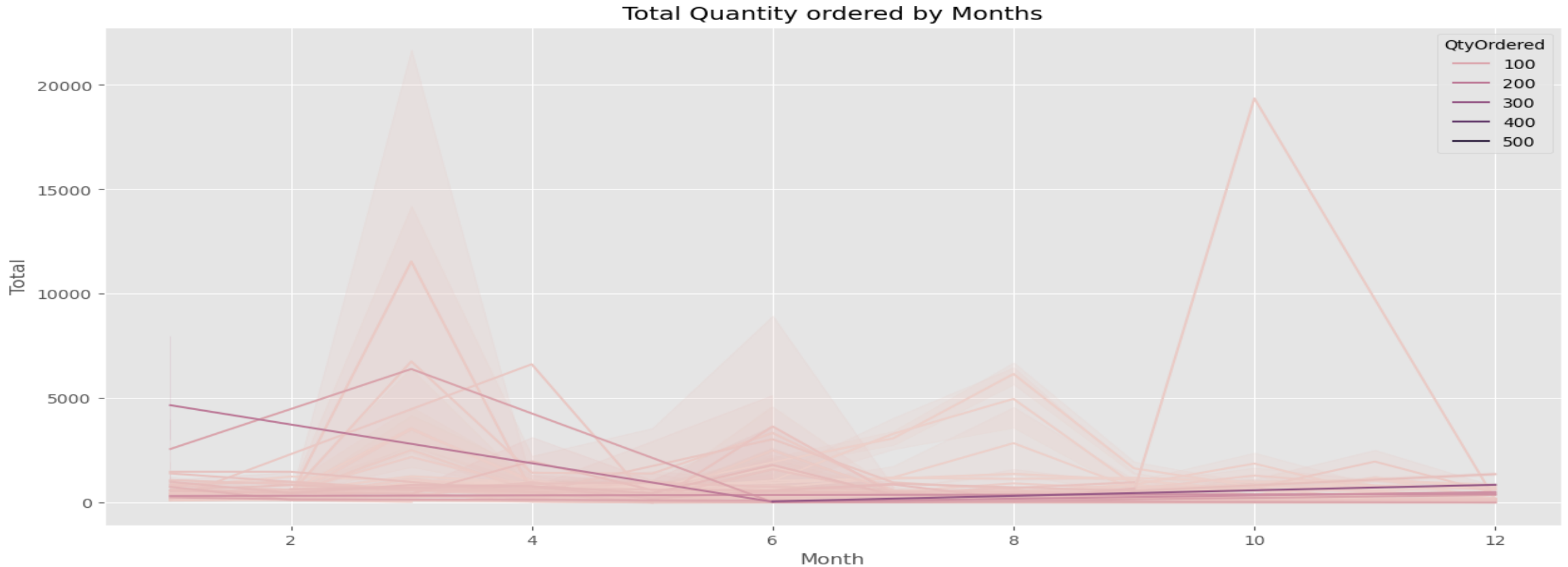


Insight:

Comparing the customer loyalty against payment methods, it would seem almost everyone prefers the COD and easypay across all ages.

Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Date after cleaning

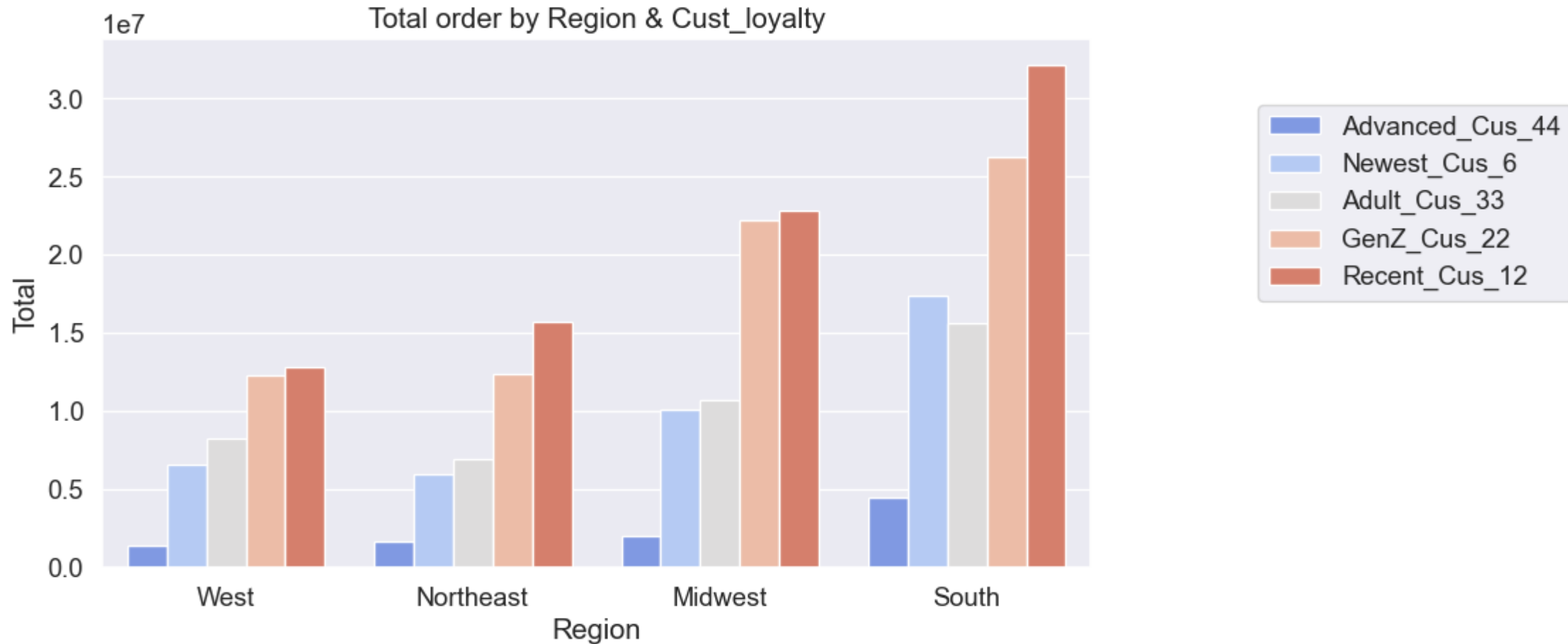


Insight:

Most of the total quantity ordered are highest in December, march and April, which is also the Winter and Spring months.

Exploratory Data Analysis

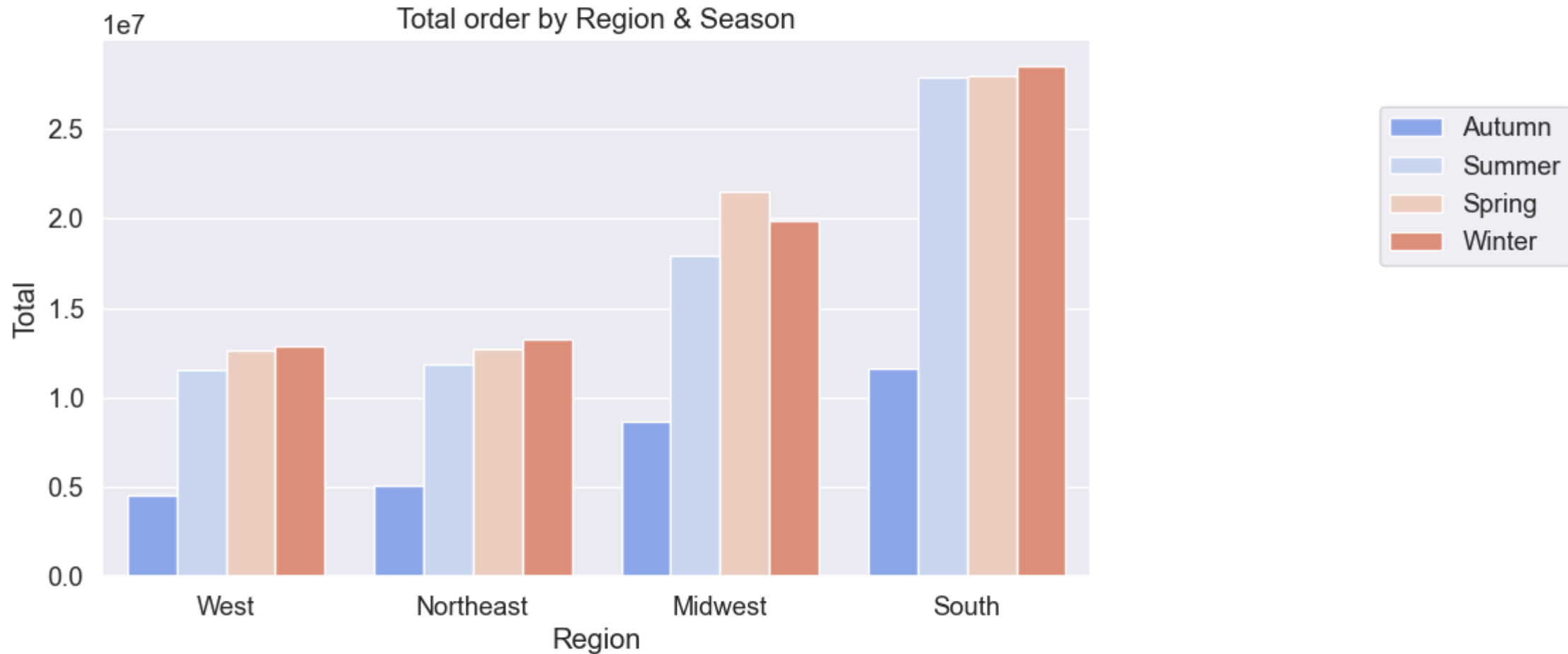
Univariate, Bivariate and Multivariate Analysis was used to explore the Data after cleaning



Insight:
The Customer loyalty is equally distributed by region as previously seen with most sales coming from the South.

Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Date after cleaning



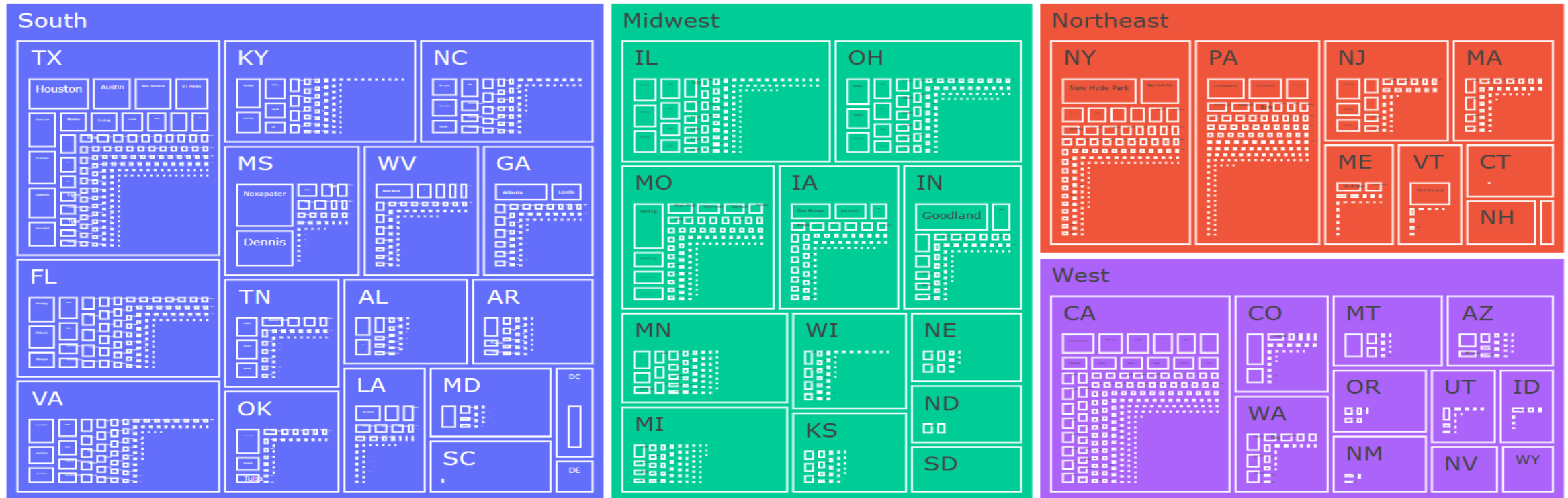
Insight:

Autum months appears to b the lowest for all Total orders as every other season seems to be equally distributed

Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Data after cleaning

Overview of Total amount paid by Customer by Region, State & City



Insight:

We have broken down the Total sales by State and we can see the regions, distribution of the states and city using the interactive map, clearly showing we have Texas and Florida (South), Illinois and Ohio (Midwest), New York and California in the Northeast and Western Region

Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Data after cleaning

INSIGHT AND RECOMMENDATIONS

- The Sales are mostly in the Southern Region followed by Midwestern region, having the Northeast and West as the least areas for customer order
- The winter and the spring are the seasons where most orders do come
- There is no clear distinctive difference between customer Gender and they are almost equal with subtle differences both in count and Total orders
- Payment method mostly used are cod and easy pay, although most Total orders were returned from easypay, bankafalah and easypayvoucher
- On the category of items sold, the Mobile& Tablets, men's fashion and appliances are highest in numbers, however, on the total orders, entertainment and computing together with mobile and tables are mostly ordered.
- Advanced and Adult customer orders are declining, perhaps this should be a recommendation to stock more items used by aging people.

Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis was used to explore the Data after cleaning

RECOMMENDATIONS

- Sterling E-commerce should consider collecting data on the cost of items to be able to determine the profit margins on each products
- Because of the reason above, we could not estimate if the items on sales are returning profit or loss
- We also recommend stocking old people items as we noticed a declining order for Advanced and Elderly Customers
- Perhaps also having a customer loyalty point system in their business as we noticed some usernames which are reoccurring in the dataset, but could not explore more on them (e.g. Jugonzalez ordered 2,524 times between 2021 and 2022)
- Expanding to High-Performing Regions: The region with the largest orders come from the South, followed by the Midwest and the least come from the West. Marketing Campaigns
- Enhancing Customer Engagement: The highly loyal customers and the ones with a long-term relationship with the company could be offered personalized rewards, incentives, or exclusive benefit