Dhiman Sarkar
MS Business Analytics ‘21
Loyola Marymount University
Los Angeles, CA

CRM Analytics on Superstore Dataset

Summary:

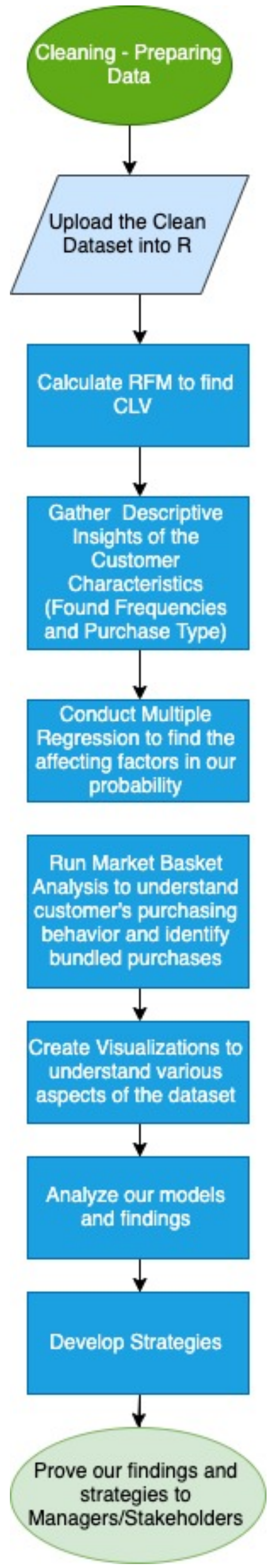
For our project we referred to the superstore dataset to conduct our analysis and create various models to study the relations with respect to the variables. In our methodologies we calculated the customer lifetime value by utilizing the Recency-Frequency-Monetary (RFM) model. This allowed us to understand the highly profitable customers and their trends with respect to their purchasing behaviors. Subsequently, a multiple regression was performed where it allowed us to analyze the relationships amongst the various independent variables to the dependent variables and identify the ones that display a high correlation. In addition, we also carried out the basket analysis which illustrated the combination of products which are purchased during a transaction. This helped us in identifying the association between the different items and the consumption trends. Besides R, we also made use of Tableau to generate various visualizations that depicts the different trends such as the most popular segments by the number of orders, growth in sales revenue by year, most profitable customers by segments, order priority trends by the year, median profits by sub-category etc. By grouping the different variables, it enabled us to understand the correlations better and verify our findings.

Steps of the Flow Chart:

1. First, we cleaned and prepared the data.
2. Upload the clean dataset into R.

3. Then we calculated the RFM for the customer lifetime value calculation.
4. After that, we found descriptive insights from the dataset regarding various customer characteristics. Specifically, we found the frequencies of region and purchase type for the customers.
5. We then ran a multiple regression to find the factors that affect the profitability of customers.
6. After the multiple regression, we ran a Basket Analysis to understand our customers' purchasing behavior and what products are often purchased together.
7. We ran a series of visualizations to understand the various other aspects of our data set.
8. We analyzed our models and developed strategies for the managers.

Flow Chart:



Part 1 : Descriptive Insights from the Data Set:

Market Frequency Table (All 7 Markets)

Var1	Freq
APAC	11002
LATAM	10294
EU	10000
US	9994
EMEA	5029
Africa	4587
Canada	384

Region Frequency Table (All 13 Regions)

Var1	Freq
Central	11117
South	6645
EMEA	5029
North	4785
Africa	4587
Oceania	3487
West	3203
Southeast Asia	3129
East	2848
North Asia	2338
Central Asia	2048
Caribbean	1690
Canada	384

Country Frequency Table (Top 10 most frequent countries only)

Var1	Freq
United States	9994
Australia	2837
France	2827
Mexico	2644
Germany	2065
China	1880
United Kingdom	1633
Brazil	1599
India	1555
Indonesia	1390

State Frequency Table (Top 10 most frequent states only)

Var1	Freq
California	2001
England	1499
New York	1128
Texas	985
Ile-de-France	981
New South Wales	781
North Rhine-Westphalia	719
Queensland	717
San Salvador	615
Pennsylvania	587

City Frequency Table (Top 10 most frequent cities only)

Var1	Freq
New York City	915
Los Angeles	747
Philadelphia	537
San Francisco	510
Santo Domingo	443
Manila	432
Seattle	428
Houston	377
Tegucigalpa	362
Jakarta	337

Part 2: Modelling-Based Insights:

Customer Lifetime Value Calculations and Analysis:

We used the RFM (Recency, Frequency, Monetary) model to calculate the customer lifetime value for each of our individual customers. However, recency, frequency, and monetary were not available in the data set so we needed to create those columns and append them to our dataframe.

Frequency is represented by n in the dataset. Recency is represented by last_date. Monetary is represented by total_profit. The code for calculating and appending each of these is shown below.

Recency(last_date):

```
#define recency
recency <- rfm.df %>%
  group_by(Customer.Name) %>%
  summarise(last_date = max(Order.Date))
rfm.df <- recency %>% left_join(rfm.df)
```

Frequency(n):

```
#define frequency
frequency <- rfm.df %>%count(Customer.Name)
frequency
rfm.df <- frequency %>% left_join(rfm.df)
```

Monetary(total_profit):

```
#define monetary
total_profit <- data.df %>%
  group_by(Customer.Name) %>%
  summarise(total_profit = sum(Profit),
            total_discount = sum(Discount),
            total_sales = sum(Sales))
data.df <- total_profit %>% left_join(data.df)
```

Before calculating CLV, we had to create bins and labels for the distributions of recency, frequency, and monetary. The breaks we used for labeling were the min, 1st quartile, mean, 3rd quartile, and max of each respective distribution. The code for this process is shown below.

```
rfm.df %>% mutate(frequency=n,recency=last_date,monetary=total_profit)->rfm.df
head(rfm.df)

summary(rfm.df$frequency)
rfm.df %>% mutate(Findex=cut(frequency,breaks=c(28,58,67,77,109),labels=c("1","2","3","4")))->rfm.df

summary(rfm.df$recency)
date1 <- as.Date("2015-01-01", tz="UTC")
as.numeric(difftime(date1, rfm.df$recency, units="days"))->rfm.df$Rdays

summary(rfm.df$Rdays)
rfm.df %>% mutate(Rindex=cut(frequency,breaks=c(0,7,17,36,430),labels=c("1","2","3","4")))->rfm.df

summary(rfm.df$total_profit)
rfm.df %>% mutate(Mindex=cut(frequency,breaks=c(-6153,1040,1834,2673,8674),labels=c("1","2","3","4")))->rfm.df
```

After cutting and labeling the data, we could go ahead and calculate CLV. For our RFM formula, we decided to weigh recency, frequency, and monetary equally at 33.33% each. The calculation is shown below:

```
rfm.df %>% mutate(CLV=33.33*as.numeric(Findex)/5+33.33*as.numeric(Rindex)/5+33.33*as.numeric(Mindex)/5)->rfm.df
summary(rfm.df$CLV) #33.33% Frequency, 33.33% Recency, 33.33% Monetary
#Findex = frequency index, Rindex = recency index, Mindex = monetary index
```

After determining how to value our customers, we set out to understand who our most valuable customers are. In other words, we want to understand what describes our ideal target customers. Based on the data available to use, we determined two questions we wanted to answer.

The first question we asked was “where are they?”. This information is available in the dataset as location data such as market, region, country, city and state. By calculating and ranking the frequency of these data points in our top customers dataset, we determined where our most valuable customers live and/or do business. Secondly, we asked “what are they buying?”. Specifically, we wanted to know what product categories and subcategories are most popular amongst our most valuable customers. Again, we accomplished this by calculating and ranking the frequency of product categories and subcategories in our top customers dataset.

The code we used to answer our first question -- “where are they?”-- is shown below along with the output and our interpretation.

```r
#Locations of our most valuable customers?
rfm.df %>% group_by(Customer.Name,clmns=CLV) %>% summarize(N=n(),Region=Region,Market=Market,Country=Country)->CLV2.df
CLV2.df #includes Region, Market,and Country. Customer Names repeats if the customer has ordered from multiple locations
#also shows frequency of each location
CLV2.df[order(CLV.df$clmns,decreasing=TRUE),]->CLV2.df
head(CLV2.df,n=1000)->Top1000_CLV_Location
Top1000_CLV_Location #shows one line for each order a top customer placed

#Most frequent country of top customers(1000 orders)
table(Top1000_CLV_Location$Country)
country_freq.df<-as.data.frame(table(Top1000_CLV_Location$Country))
country_freq.df[order(country_freq.df$Freq,decreasing=TRUE),]->country_freq.df
head(country_freq.df,n=10)->Mostvaluable_countries
Mostvaluable_countries #shows 10 countries that appear most in top 1000 orders of top customers

#Most frequent market of top customers
market_freq.df<-as.data.frame(table(Top1000_CLV_Location$Market))
market_freq.df[order(market_freq.df$Freq,decreasing=TRUE),]->market_freq.df
head(market_freq.df,n=10)->Mostvaluable_markets
Mostvaluable_markets

#Most frequent region of top customers
region_freq.df<-as.data.frame(table(Top1000_CLV_Location$Region))
region_freq.df[order(region_freq.df$Freq,decreasing=TRUE),]->region_freq.df
head(region_freq.df,n=10)->Mostvaluable_regions
Mostvaluable_regions
```

In order to determine the most important locations, we calculated the frequency of different location data in a sample of 1000 orders of our top customers. This sample was not random. It was simply the top 1000 orders of our customers ranked by their CLV. The frequency tables below show the top markets, regions, and countries as well as their frequency. For example, in the sample of 1000 orders, the United States appeared 169 times. This is equivalent to 16.9% of the orders.

Most Frequent Countries of Top Customers:

Var1	Freq
United States	169
France	50
Australia	49
Mexico	46
United Kingdom	45
India	26
Italy	26
Indonesia	25
Turkey	25
China	23

Most Frequent Regions of Top Customers:

```
          Var1 Freq
       Central  180
         North  101
         South   87
          EMEA   68
        Africa   63
          East   55
       Oceania   53
          West   53
Southeast Asia   51
  Central Asia   32
```

<u>Most Frequent Markets of Top Customers:</u>

```
  Var1 Freq
    EU  182
    US  169
  APAC  164
 LATAM  143
  EMEA   68
Africa   63
Canada    6
```

The code we used to answer our second question -- "What are they buying?" -- is shown below along with the output and our interpretation.

```
#Create df with Customer Name, CLV, product category, and sub category
rfm.df %>% group_by(Customer.Name,clmns=CLV) %>% summarize(N=n(),Product_Category=Category,Sub_Category=Sub.Category)->CLV3.df
CLV3.df#includes product category and sub-category
CLV3.df[order(CLV3.df$clmns,decreasing=TRUE),]->CLV3.df
head(CLV3.df,n=1000)->Top1000_CLV_Products

#Most frequent product category
category_freq.df<-as.data.frame(table(Top1000_CLV_Products$Product_Category))
category_freq.df[order(category_freq.df$Freq,decreasing=TRUE),]->category_freq.df
head(category_freq.df,n=3)->Mostvaluable_productcategories #only 3 product categories
Mostvaluable_productcategories

#Most frequent sub-category
sub_freq.df<-as.data.frame(table(Top1000_CLV_Products$Sub_Category))
sub_freq.df[order(sub_freq.df$Freq,decreasing=TRUE),]->sub_freq.df
head(sub_freq.df,n=20)->Mostvaluable_subcategories
Mostvaluable_subcategories
```

Again we filtered our top customers database for 1000 orders. This time we included the category and subcategory data of each. We then calculated the frequency of each category and subcategory and sorted the resulting frequency tables. Just as above, the frequency in the tables below is out of 1000. For example, a frequency of 601 for office supplies means that 60.1% of the orders in our sample were of the product category office supplies.

Most Frequent Categories of our Top Customers:

Var1	Freq
Office Supplies	601
Technology	201
Furniture	198

Most Frequent Subcategories of our Top Customers:

Var1	Freq
Binders	126
Art	108
Storage	95
Accessories	68
Chairs	68
Furnishings	66
Paper	62
Phones	61
Labels	50
Bookcases	45

Interpretation:

According to our location analysis, 16.9% of our sample orders were placed in the United States. This was far more than any other country. The next highest was France at 5%. However, this does not mean we should solely focus our attention on the United States. According to the region frequency table, 18.2% of the orders in our sample were placed in the EU(European Union), 16.4% in the APAC (Asia-Pacific), and 14.3% in the LATAM(Latin America) region. In conclusion, it is clear that the United States should be a point of emphasis for our marketing and sales teams but it's difficult to narrow our most valuable customers down to one region.

As for our product analysis, the office supplies category was the clear winner. 60.1% of the orders in our sample fell into the office supplies category, followed by Tech at 20.1% and Furniture at 19.8%. Diving deeper, the most frequent subcategories were much more evenly distributed, with only two subcategories accounting for more than 10% of the total. Not surprisingly, most of these subcategories belong to the office supplies category.

Multiple Regression Analysis and Insights:

We ran two multiple regressions to understand how the various factors affect sales and profitability. We also used these regressions to test for seasonality by using the re-coded day of the week and month of the year variables to see if there was a significant difference caused by changes in month or day of week. For brevity, we are only attaching the code and output from the final model we ran for each profit and sales. Our process was to initially run a naive model that included the factorized variables for day of the week and month of the year as well as the variables relating to location, product type, discount, sales, and shipping cost. However, we were unable to include Product name or ID and Customer Name or ID because the number of categories would cause the computer to crash. We would then test for outliers and remove the first 10. Afterwards, we would run the AIC to find the most efficient model. Finally, we would make any necessary tweaks to the AIC model to remove unnecessary factors.

Profit Code:

```
#Test without order priority

profit.m4<- lm(Profit~Sales+Discount+Sub.Category+Quantity+Country+Shipping.Cost

           +Region, data = data.df[c(-28613,-43454,-15887,-8899,-43137,-42768,-38849,-28814,-
11890,-22944),])

summary(profit.m4) # no change, but fewer variables, so it is better
```

Profit Output:

```
Call:
lm(formula = Profit ~ Sales + Discount + Sub.Category + Quantity + 
    Country + Shipping.Cost + Region, data = data.df[c(-28613, 
    -43454, -15887, -8899, -43137, -42768, -38849, -28814, -11890, 
    -22944), ])

Residuals:
    Min      1Q  Median      3Q     Max 
-7273.8   -30.4    -4.8    37.3  4875.5 

Coefficients: (7 not defined because of singularities)
                                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)                          5.911e+01  1.924e+01   3.073 0.002123 ** 
Sales                                2.014e-01  2.117e-03  95.096  < 2e-16 ***
Discount                            -2.512e+02  4.938e+00 -50.874  < 2e-16 ***
Sub.CategoryAppliances              -2.099e+01  4.269e+00  -4.916 8.85e-07 ***
Sub.CategoryArt                      1.400e+00  3.289e+00   0.426 0.670394    
Sub.CategoryBinders                  1.664e+01  3.153e+00   5.278 1.31e-07 ***
Sub.CategoryBookcases               -3.625e+01  3.908e+00  -9.276  < 2e-16 ***
Sub.CategoryChairs                  -2.759e+01  3.537e+00  -7.801 6.26e-15 ***
Sub.CategoryCopiers                 -8.982e+00  4.019e+00  -2.235 0.025449 *  
Sub.CategoryEnvelopes                6.257e+00  3.857e+00   1.622 0.104736    
Sub.CategoryFasteners                8.460e+00  3.869e+00   2.186 0.028785 *  
Sub.CategoryFurnishings              3.283e+00  3.590e+00   0.914 0.360518    
Sub.CategoryLabels                   6.402e+00  3.789e+00   1.690 0.091128 .  
Sub.CategoryMachines                -4.386e+01  4.520e+00  -9.704  < 2e-16 ***
Sub.CategoryPaper                    4.026e+00  3.507e+00   1.148 0.251010    
Sub.CategoryPhones                  -2.288e+01  3.559e+00  -6.429 1.29e-10 ***
Sub.CategoryStorage                 -1.212e+01  3.248e+00  -3.731 0.000191 ***
Sub.CategorySupplies                -1.899e+00  3.862e+00  -0.492 0.622889    
Sub.CategoryTables                  -1.981e+02  5.624e+00 -35.224  < 2e-16 ***
Quantity                            -5.309e+00  3.040e-01 -17.466  < 2e-16 ***
CountryAlbania                      -4.028e+01  4.503e+01  -0.895 0.370996    
CountryAlgeria                      -2.982e+01  2.157e+01  -1.383 0.166806    
CountryAngola                       -2.495e+01  2.295e+01  -1.087 0.276998    
```

```
CountryEcuador                      -2.253e+01  2.717e+01  -0.829 0.406891    
CountryEgypt                        -3.432e+01  2.011e+01  -1.707 0.087865 .  
CountryEl Salvador                  -3.502e+01  2.013e+01  -1.739 0.081984 .  
CountryEquatorial Guinea            -5.005e+01  8.378e+01  -0.597 0.550204    
CountryEritrea                      -3.726e+01  1.017e+02  -0.366 0.714078    
CountryEstonia                      -2.364e+01  4.812e+01  -0.491 0.623218    
CountryEthiopia                     -1.883e+01  5.670e+01  -0.332 0.739808    
CountryFinland                      -3.525e+01  2.598e+01  -1.357 0.174810    
CountryFrance                       -4.976e+01  1.962e+01  -2.536 0.011214 *  
CountryGabon                        -3.558e+01  4.230e+01  -0.841 0.400226    
CountryGeorgia                      -4.455e+01  3.721e+01  -1.197 0.231210    
CountryGermany                      -4.100e+01  1.969e+01  -2.082 0.037302 *  
CountryGhana                        -2.596e+01  2.364e+01  -1.098 0.272115    
CountryGuadeloupe                    1.106e+00  5.347e+01   0.021 0.983498    
CountryGuatemala                    -4.216e+01  2.041e+01  -2.066 0.038830 *  
CountryGuinea                       -2.563e+01  3.838e+01  -0.668 0.504204    
CountryGuinea-Bissau                -4.059e+01  5.081e+01  -0.799 0.424383    
CountryHaiti                         2.080e+01  2.366e+01   0.879 0.379384    
CountryHonduras                     -1.210e+01  2.019e+01  -0.599 0.548986    
CountryHong Kong                    -1.480e+02  1.492e+02  -0.992 0.321156    
CountryHungary                      -4.029e+01  3.289e+01  -1.225 0.220635    
CountryIndia                        -1.551e+01  1.939e+01  -0.800 0.423776    
CountryIndonesia                     1.266e+00  1.948e+01   0.065 0.948187    
CountryIran                         -3.772e+01  2.850e+01  -1.324 0.185651    
CountryIraq                         -3.488e+01  2.883e+01  -1.210 0.226324    
CountryIreland                      -6.419e+00  2.361e+01  -0.272 0.785774    
CountryIsrael                       -4.898e+01  3.120e+01  -1.570 0.116427    
CountryItaly                        -3.344e+01  1.999e+01  -1.672 0.094461 .  
CountryJamaica                      -2.345e+01  3.173e+01  -0.739 0.459896    
CountryJapan                        -1.259e+02  1.463e+02  -0.861 0.389482    
CountryJordan                       -3.654e+01  3.542e+01  -1.032 0.302298    
CountryKazakhstan                    4.648e+01  3.138e+01   1.481 0.138557    
CountryKenya                        -3.844e+01  2.381e+01  -1.614 0.106441    
CountryKyrgyzstan                   -3.147e+01  3.576e+01  -0.880 0.378843    
CountryLebanon                      -1.230e+01  4.943e+01  -0.249 0.803459    
```

```
CountryNetherlands                   -5.572e+01  2.066e+01  -2.697 0.006995 **
CountryNew Zealand                   -9.853e+00  1.989e+01  -0.495 0.620367
CountryNicaragua                     -4.112e+01  2.027e+01  -2.029 0.042474 *
CountryNiger                         -3.284e+01  2.877e+01  -1.142 0.253650
CountryNigeria                        3.471e+01  1.993e+01   1.741 0.081682 .
CountryNorway                        -1.795e+01  2.546e+01  -0.705 0.480695
CountryPakistan                      -5.131e+01  2.116e+01  -2.425 0.015332 *
CountryPanama                        -1.845e+01  2.077e+01  -0.889 0.374222
CountryPapua New Guinea               2.566e+01  3.571e+01   0.719 0.472384
CountryParaguay                      -3.682e+01  4.522e+01  -0.814 0.415540
CountryPeru                           1.682e+01  2.254e+01   0.746 0.455540
CountryPhilippines                   -1.763e+01  1.988e+01  -0.887 0.375126
CountryPoland                        -3.381e+01  2.930e+01  -1.154 0.248485
CountryPortugal                      -7.326e+01  2.592e+01  -2.826 0.004713 **
CountryQatar                         -1.991e+01  4.502e+01  -0.442 0.658340
CountryRepublic of the Congo         -5.398e+01  6.075e+01  -0.889 0.374219
CountryRomania                       -4.154e+01  2.991e+01  -1.389 0.164900
CountryRussia                        -2.783e+01  2.883e+01  -0.965 0.334535
CountryRwanda                        -3.138e+01  3.029e+01  -1.036 0.300236
CountrySaudi Arabia                  -3.505e+01  2.889e+01  -1.213 0.225038
CountrySenegal                       -2.217e+01  2.327e+01  -0.953 0.340647
CountrySierra Leone                  -3.823e+01  4.116e+01  -0.929 0.353069
CountrySingapore                     -2.395e+01  2.246e+01  -1.066 0.286318
CountrySlovakia                      -5.585e+01  5.723e+01  -0.976 0.329055
CountrySlovenia                       9.066e+01  8.623e+01   1.051 0.293110
CountrySomalia                       -5.909e+01  2.734e+01  -2.162 0.030644 *
CountrySouth Africa                  -3.194e+01  2.016e+01  -1.584 0.113196
CountrySouth Korea                   -1.523e+02  1.463e+02  -1.041 0.298004
CountrySouth Sudan                   -3.862e+01  1.017e+02  -0.380 0.704150
CountrySpain                         -2.422e+01  2.012e+01  -1.204 0.228743
CountrySri Lanka                     -4.760e+01  5.670e+01  -0.839 0.401201
CountrySudan                         -3.172e+01  2.409e+01  -1.317 0.187841
CountrySwaziland                     -3.758e+01  1.017e+02  -0.370 0.711713
CountrySweden                        -1.970e+01  2.163e+01  -0.911 0.362308
```

```
CountryTrinidad and Tobago            3.247e+01  3.207e+01   1.013 0.311295
CountryTunisia                       -2.912e+01  4.358e+01  -0.668 0.504053
CountryTurkey                         1.817e+01  2.827e+01   0.643 0.520282
CountryTurkmenistan                  -1.319e+01  3.982e+01  -0.331 0.740391
CountryUganda                         5.963e+01  3.025e+01   1.972 0.048672 *
CountryUkraine                       -3.261e+01  2.884e+01  -1.131 0.258218
CountryUnited Arab Emirates           4.925e+01  4.704e+01   1.047 0.295189
CountryUnited Kingdom                -9.364e+00  1.937e+01  -0.483 0.628883
CountryUnited States                 -1.069e+01  1.922e+01  -0.556 0.578237
CountryUruguay                       -2.986e+01  3.483e+01  -0.857 0.391368
CountryUzbekistan                    -4.143e+01  3.457e+01  -1.198 0.230750
CountryVenezuela                     -7.843e+00  2.209e+01  -0.355 0.722561
CountryVietnam                       -2.379e+01  2.097e+01  -1.134 0.256635
CountryYemen                         -5.323e-01  3.811e+01  -0.014 0.988857
CountryZambia                        -1.707e+01  2.364e+01  -0.722 0.470189
CountryZimbabwe                       5.775e+01  2.500e+01   2.310 0.020892 *
Shipping.Cost                        -8.266e-02  1.721e-02  -4.803 1.57e-06 ***
RegionCanada                                 NA         NA      NA       NA
RegionCaribbean                              NA         NA      NA       NA
RegionCentral                         1.668e+01  3.905e+00   4.272 1.94e-05 ***
RegionCentral Asia                           NA         NA      NA       NA
RegionEast                            2.915e+00  3.644e+00   0.800 0.423869
RegionEMEA                            5.649e+00  2.041e+01   0.277 0.781937
RegionNorth                                  NA         NA      NA       NA
RegionNorth Asia                      1.262e+02  1.447e+02   0.872 0.383364
RegionOceania                                NA         NA      NA       NA
RegionSouth                          -1.130e-01  4.313e+00  -0.026 0.979102
RegionSoutheast Asia                         NA         NA      NA       NA
RegionWest                                   NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 141.3 on 51108 degrees of freedom
Multiple R-squared:  0.3457,    Adjusted R-squared:  0.3435
F-statistic: 157.9 on 171 and 51108 DF,  p-value: < 2.2e-16
```

Commentary on Profit Output:

Based on the p-value of 2.2e-16 being lower than 0.05, we can assume that the findings of our model are significant. However, the adjusted R-Squared of 0.3435 means that only 34.35% of the deviation is explained by the model. This means that the model is weak in its predictive power. This weakness is possibly caused by the variation in profitability within each category. The number of different products and the limitations of our computers prevented us from being able to incorporate the specific items in the model. Some notable findings from this regression were that the hypothetical sale (the intercept) estimates a base profit of $59.11. However, this is more important for prediction rather than for explaining relationships. The following findings are all statistically significant, meaning they have a p-value of less than 0.05 and are meaningful. We found that each dollar increase in sales leads to an increase in profitability by $0.20. Also as expected, an increase in discount of 1 percent leads to a decrease in profit of $25.12. Interestingly, an increase in quantity of 1 unit leads to a decrease in profit of $5.31. This may be caused by an increase in shipping cost caused by the new item or by the variation in profit for each item. Each dollar increase in shipping cost leads to a decrease in profit of $0.08. The notable region category is the central region, which has an increased profit of $16.68 compared to the baseline. Although there are too many countries to mention, two significant findings are that buyers in the Netherlands are $55.72 less profitable than other buyers and buyers in Portugal are $73.26 less profitable than the average buyer. Although most of the subcategories were significant and indicated a decreased profitability, the Binders subcategory indicated an increased profitability of $16.64 compared to the average transaction. The notably negative subcategory was Copiers. People who bought items in the Copiers subcategory were predicted to be $89.82 less profitable than the baseline buyer. This could be caused by the wide variation in the profitability of the items in the copier category. To reiterate, although the findings of our profit model were significant, the predictive power is low because it seems that profitability is more closely related to specific customers or items rather than broad categories and it was beyond our computers' capabilities to process the broad number of categories. Seasonality was not a factor for profit as neither day of the week nor month of the year led to a significant difference in profit.

Sales Regression Code:

Our sales regression model process was similar to our regression process for profit. We ran a naive model, removed outliers, and ran the AIC. We then tweaked the AIC to reduce unnecessary factors. We did not include profit, Product Name, or Customer Name in our AIC or in our naive model.

```
sales.m2 <- lm(Sales~Shipping.Cost+Order.Priority+Sub.Category+Quantity
        +Ship.Mode+Discount+Market+Region,data=data.df[c(-28613,-43454,-15887,-8899,-43137,-42768,-38849,-28814,-11890,-22944),] )

summary(sales.m2)
```

Sales Regression Output:

```
Call:
lm(formula = Sales ~ Shipping.Cost + Order.Priority + Sub.Category +
    Quantity + Ship.Mode + Discount + Market + Region, data = data.df[c(-28613,
    -43454, -15887, -8899, -43137, -42768, -38849, -28814, -11890,
    -22944), ])

Residuals:
    Min      1Q  Median      3Q     Max
-3385.9   -82.2   -21.3    41.9 22183.7

Coefficients: (4 not defined because of singularities)
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)           -201.24392    8.52126 -23.617  < 2e-16 ***
Shipping.Cost            5.83309    0.02663 219.068  < 2e-16 ***
Order.PriorityHigh     127.48767    5.34919  23.833  < 2e-16 ***
Order.PriorityLow      124.44363    8.03635  15.485  < 2e-16 ***
Order.PriorityMedium   196.87326    5.46888  35.999  < 2e-16 ***
Sub.CategoryAppliances 138.02158    8.64110  15.973  < 2e-16 ***
Sub.CategoryArt        -58.27008    6.66485  -8.743  < 2e-16 ***
Sub.CategoryBinders    -51.96243    6.37895  -8.146 3.85e-16 ***
Sub.CategoryBookcases  153.11414    7.89687  19.389  < 2e-16 ***
Sub.CategoryChairs      74.67454    7.16408  10.423  < 2e-16 ***
Sub.CategoryCopiers    178.52918    8.11656  21.996  < 2e-16 ***
Sub.CategoryEnvelopes  -55.37128    7.82214  -7.079 1.47e-12 ***
Sub.CategoryFasteners  -70.72071    7.84365  -9.016  < 2e-16 ***
Sub.CategoryFurnishings -39.56694   7.27894  -5.436 5.48e-08 ***
Sub.CategoryLabels     -76.82400    7.68159 -10.001  < 2e-16 ***
Sub.CategoryMachines   129.96231    9.14592  14.210  < 2e-16 ***
Sub.CategoryPaper      -61.31696    7.11169  -8.622  < 2e-16 ***
Sub.CategoryPhones     101.98150    7.20970  14.145  < 2e-16 ***
Sub.CategoryStorage      2.31109    6.58915   0.351   0.7258
```

```
Sub.CategorySupplies     -45.18876    7.83072  -5.771 7.94e-09 ***
Sub.CategoryTables       265.43771   11.28568  23.520  < 2e-16 ***
Quantity                  27.19718    0.60508  44.948  < 2e-16 ***
Ship.ModeSame Day          5.68164    6.44894   0.881   0.3783
Ship.ModeSecond Class     45.53292    4.40128  10.345  < 2e-16 ***
Ship.ModeStandard Class   68.93707    4.05582  16.997  < 2e-16 ***
Discount                 -89.12355    6.26428 -14.227  < 2e-16 ***
MarketAPAC                 1.29501    6.77063   0.191   0.8483
MarketCanada               8.40697   15.28166   0.550   0.5822
MarketEMEA                 0.40915    5.86596   0.070   0.9444
MarketEU                 -10.30567    8.70406  -1.184   0.2364
MarketLATAM              -50.14649    8.83702  -5.675 1.40e-08 ***
MarketUS                 -16.88393    6.73327  -2.508   0.0122 *
RegionCanada                    NA         NA      NA       NA
RegionCaribbean            9.13714   10.37180   0.881   0.3783
RegionCentral             13.98180    7.10412   1.968   0.0491 *
RegionCentral Asia        21.26009    8.26157   2.573   0.0101 *
RegionEast                10.14133    7.39880   1.371   0.1705
RegionEMEA                      NA         NA      NA       NA
RegionNorth               21.22836    8.38148   2.533   0.0113 *
RegionNorth Asia           5.97724    7.97257   0.750   0.4534
RegionOceania              5.23233    7.11137   0.736   0.4619
RegionSouth               16.86747    7.35782   2.292   0.0219 *
RegionSoutheast Asia            NA         NA      NA       NA
RegionWest                      NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 287 on 51240 degrees of freedom
Multiple R-squared:  0.6538,    Adjusted R-squared:  0.6536
F-statistic:  2482 on 39 and 51240 DF,  p-value: < 2.2e-16
```

Commentary on Sales Regression Output:

Although the adjusted r-squared is slightly lower than the naive model, we chose the suggested model because it reduced the correlation between some of the factors that were used. This model has a p-value of 2.2e-16 which is well below 0.05, so the findings of the model are significant. The explanatory power of the model is 65.36%, which means that this model accounts for a good amount of the deviation from expectations. Part of the missing explanatory power may come from specific customers or products, but we could not include these in our model because of computer limitations. All of the following findings mentioned are significant, meaning they have a p-value of below 0.05 and are meaningful indicators of changes in sales dollars. The intercept is negative $201.24, but this number is significant for modelling purposes rather than as an indicator of base sales value. For each increase in Shipping Cost of $1.00, we have found an increase in sales of $5.80. This relationship could be because an increase in shipping cost is caused by increases in the quantity of items sold. We also found that the Order Priority type also had an effect on sales. In comparison to a Critical Order Priority, High, Low, and Medium orders led to increases in sales of $127.48, $124.44, and $196.87, respectively. The three product subcategories that increase the sales values the most are Tables, Copiers, and Bookcases. These increase sales amount in comparison to a baseline purchase by $265.44, $178.53, and $153.11, respectively. The three subcategories that lead to the greatest decrease in sales amount compared to a baseline purchase are Labels, Fasteners, and Paper at a decrease of $76.80, $70.72, and $61.31. Quantity was also a significant factor, with each increase in quantity of an item sold leading to an increase in sales amount by $27.20. Second class and Standard class shipments lead to increases in sales dollars of $45.52 and $68.94 in comparison to a first class shipment. With each percentage increase in discount, we noticed a decrease in sales dollars of

$89.12. Customers in the Central Asia Region increased sales dollars the most, with an increase of $21.26, compared to the baseline customer. Customers ordering to the LATAM and US markets had significantly lower sales values than the baseline customer, with a reduction of $50.15 in LATAM and $16.88 in the US. We noticed no seasonality either by day of the week or month of the year in our sales analysis, as there was no significant difference in sales caused by a specific month. Some of the variation in sales dollars may be caused by the specific products sold or the specific customer purchasing, but our computers were unable to run a regression with that many categories.

Basket Analysis:

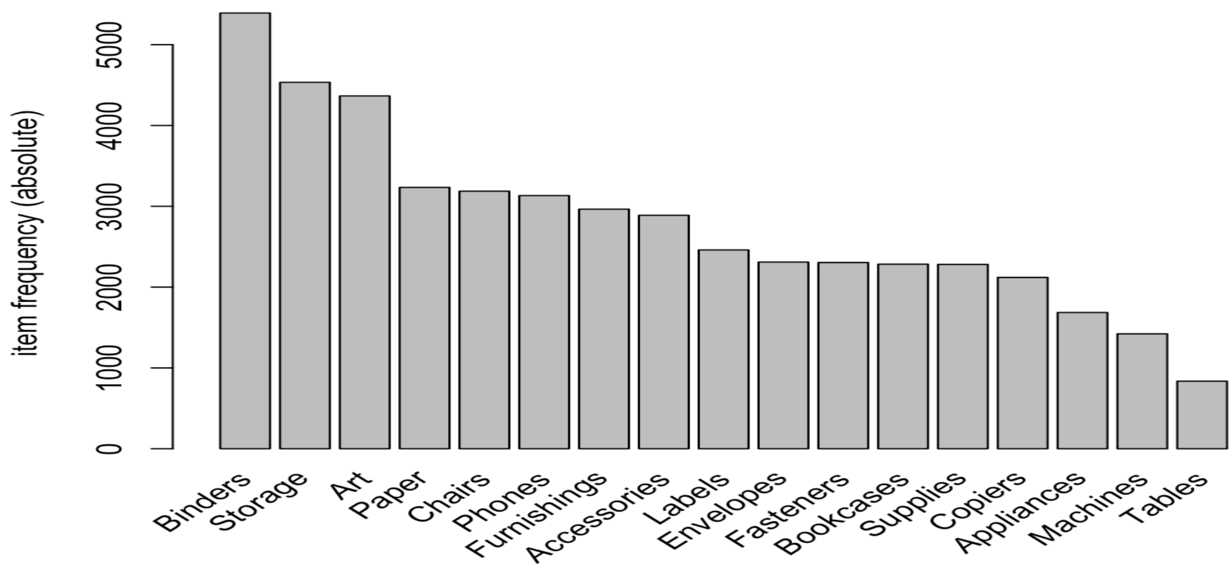
In order to understand which types of products were often purchased together, we also ran a market basket analysis. To run this analysis, we created a subset of the data set to organize the information by Order ID and product SubCategory. We chose Order ID to understand each transaction, since we know that some customers make multiple purchases. We also chose Subcategory rather than a specific product because it would be easier to detect general trends. We then transformed the data into transactional format. After experimenting with support and confidence levels, we settled on one that would give a handful of rules that would allow us to form simple business suggestions. After selecting the subset of rules with the appropriate confidence and support levels, we sorted by lift to find the rules that had the strongest chance of occurring in reality. This would also help in our presentation for providing easy pointers to salespeople.

Basket Analysis Code:

```r
# Market Basket Analysis for Group Presentation

library(arules)
library(arulesViz)


arm <- read.csv("group.csv")
dim(arm)  #51290 rows, 24 Cols, Row.ID is the rowid column from excel & not available in Data Dictionary. Need to remove it.
head(arm)
arm <- arm[,-1]
dim(arm)  #51290 rows, 23 Cols, Order.ID is the key column.

str(arm) #Most of the cols are character/ Numeric data type. Our consideration will be sub-category
summary(arm)

#Lets find unique order id count
length(unique(arm$Order.ID)) #--> 25035 unique Order.ID
length(unique(arm$Sub.Category)) #17 unique sub category --> important for us

#Converting the DF to transactional format.
#Using read.transactions
#create mini DF (only required columns in  factor format)
#to a csv file and then read it using read.transactions function
#removes any irregularities possible in the data

arm_mini <- arm[,c("Order.ID","Sub.Category")] # ---> making subset(don't care about customer)
head(arm_mini)
write.csv(arm_mini,"transdata",row.names = F) #force the DF into csv
transdata<-read.transactions(file="transdata",format="single",sep=",",cols=c("Order.ID","Sub.Category"), rm.duplicates = T, header = T)

arm_transactions<-as(transdata,"transactions")
summary(arm_transactions)
#visually showing most frequent items

itemFrequencyPlot(arm_transactions,topN=20,type="absolute")

# find some initial rules
arm.rules <- apriori(arm_transactions, parameter=list(support=0.0006, conf=0.5,
                                                       target="rules"))
inspect(arm.rules)

#discover strong relationship between Appliances and Binders
top_lift<-head(arm.rules, n=5, by= "lift")
inspect(top_lift)

#Mapping the top 5 rules by lift
group.hi<-head(sort(arm.rules,by="lift"),5)
plot(group.hi,method="graph",control=list(type="items"))
```

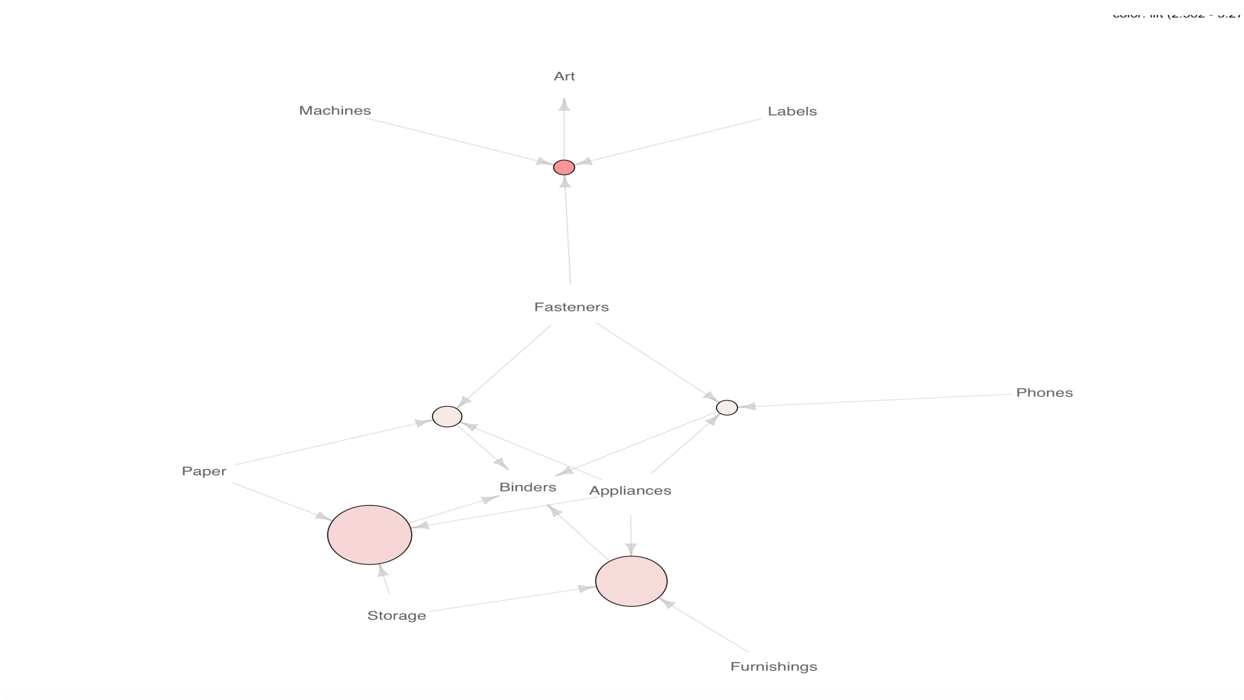
Basket Analysis Output:

Frequency plot:



Initial Rules sorted by lift:

```
    lhs                                  rhs       support      confidence coverage    lift     count
[1] {Fasteners,Labels,Machines}       => {Art}     0.0006391053 0.5714286  0.001118434 3.276618 16
[2] {Appliances,Paper,Storage}        => {Binders} 0.0012382664 0.6078431  0.002037148 2.822209 31
[3] {Appliances,Furnishings,Storage} => {Binders} 0.0011184342 0.5957447  0.001877372 2.766036 28
[4] {Appliances,Fasteners,Paper}      => {Binders} 0.0007189934 0.5625000  0.001278211 2.611682 18
[5] {Appliances,Fasteners,Phones}     => {Binders} 0.0006391053 0.5517241  0.001158378 2.561649 16
```



Basket Analysis Commentary:

After formatting the data, we counted the frequency of purchase of each subcategory, so we would have a good idea of what types of items would appear in the rules together. The five rules we selected and sorted by lift are shown in the table above. These show that the strongest lift is in a rule that does not occur as frequently as others (Fasteners, Labels, Machines) -> Art. However, it will provide a helpful suggestion in our Managerial Implications portion. We also noticed that items in the Binders subcategory frequently appear in the right hand side, while items in the Appliance subcategory appear on the left hand side in four of the five rules we have selected. The appearance of Binders in the rules is unsurprising, since it is the most frequently purchased subcategory of item. However, it is surprising that Appliances would appear so often since they are not as frequently purchased. The support is relatively small for all of the rules because there are over 25,000 transactions, so the number of occurrences of a specific incident are relatively small. However for all of our rules, the confidence level is between 0.5 and 0.6.

This means that when items on the left hand side are purchased, the item on the right hand side will also be purchased between 50% and 60% of the time. Based on the lift, we can say that customers that purchase Fasteners, Labels, and Machines are 3.28 times more likely to purchase Art than the average customer. Possibly, they are buying the other items to label and hang the painting. We can say that customers who purchase Appliances, Paper, and Storage are 2.82 times more likely to purchase Binders than the average customer. Customers who purchase Appliances, Furnishings, and Storage are 2.78 times more likely to purchase Binders than the average customer. Customers who purchase Appliances, Fasteners, and Paper are 2.61 times more likely to purchase Binders than the average customer. Finally, we see that customers who purchase Appliances, Fasteners, and Phones are 2.56 times more likely to purchase Binders than the average customer.

The above graph is a depiction of the rules. The increased size of the circle is a sign of increased support, while the increased lift is shown by the darker color. The arrow pointing from the circle shows the right hand side item, while the arrows pointing to the circle indicate the left hand side items.

Part 3: Managerial Implications

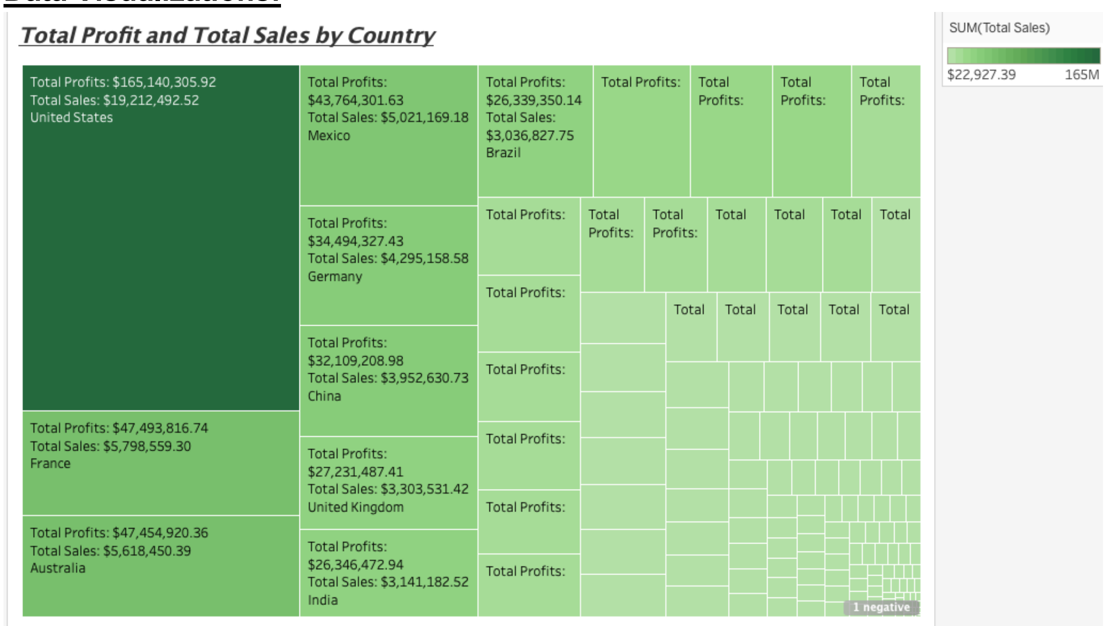
Based on the two regressions and the product analysis we ran, we can provide some suggestions for business managers. One of the primary issues is that the products within the same category are similar, but have hugely different effects on profitability. For example, the most profitable and least profitable products available are both Copiers. Rather than expecting the sales representatives to inherently understand the differences in profitability in slightly different models, a more effective technique would be to phase out unprofitable products.

Phasing out unprofitable products, especially high-priced items would free space for the more profitable products while allowing salespeople to generate profit. This would also make future analysis of sales trends easier, because the profitability of various customers, product subcategories, and markets would not be skewed by items that generate losses.

Another critical managerial takeaway from the regression analysis would be that the drivers of sales dollars are not the same as the drivers of profitability. Specifically profitable customers and products aside, factors like quantity, shipping cost, order priority, discounts offered, and the customers' location can all drive sales dollars or counts upwards while decreasing profits. The quantity sold can increase profits, but if a large number of the items sold cause losses, the large sale is more harmful than a small sale at a loss. Similarly, discounts can increase the quantity of sales, but may push a profitable good into being a loss generator. Rather, a manager should use discounts to either get rid of excess inventory or to drive a cross-sale by promoting common market baskets. In this situation Shipping Cost, Shipping Mode and Order Priority all also contributed negatively to the profitability of a sale. This is because these factors add additional costs to an item. Finally, it is important for managers to understand how seasonality may affect their sales. In this case, the data set was not seasonal. However, specific products or product lines may be seasonal even if the general sales trend is not.

From our basket analysis, we had some key takeaways. Understanding what items or types of items are commonly purchased together will enable a manager to increase cross-selling. By driving cross-selling, a manager can improve their profitability by increasing their share of the particular customer. Three techniques to do this are discounting items, bundling, and placement in a store. Discounting a specific item that is known to form part of a basket is a good technique because it encourages the customer to purchase the item and feel like they have gotten a deal.Discounting is most effective when offering a large discount on a cheaper item. In this dataset, discounting Fastener products, which appear in three of the five rules and are low cost, would be an effective way to drive cross-selling.This discount would allow the manager to stimulate three types of cross-selling at little cost. Another way market baskets can be used to drive sales would be through explicitly bundling products. In this dataset Fasteners, Labels, and Machines are often sold with Art. This could be because customers who want to hang up a piece of art must buy fasteners, labels, and a labelling machine. By creating a story like this, we can see that offering a bundle to people who want to buy art or accomplish another specific task would be an effective way to drive sales. Finally, managers can drive cross-selling by placing items that are frequently purchased together near each other in the store or in the recommendation section on a webpage. In this dataset, binders are frequently sold in baskets that include appliances. By placing binders near appliances in the store or on the website, the manager can drive cross-selling.

Data Visualizations:



Description: The above Treemap displays the most Total Profits along with the Total Sales by country. By the dense colorization we can identify the countries with the highest amount of total sales. In this case the United States, France, Australia, Mexico, Germany, China, United Kingdom and India are amongst the top countries by the most sales. The United States ranks first in both the domains with raking up Total Profits as $165,140,305.92 and Total Sales as $19,212,492.52.

Highest Shipping Cost Countries
Sorted by Total Shipping Cost and Total Quantity

Country	Shipping.C..	Quantity
United States	$238,173.79	37,873
France	$95,387.81	10,804
Australia	$100,359.02	10,673
Mexico	$67,659.85	10,011
Germany	$63,965.18	7,745
China	$78,957.02	7,081
United Kingdom	$53,580.27	6,161
Brazil	$38,170.73	6,148
India	$61,780.72	5,758
Indonesia	$43,948.62	5,237
Italy	$29,970.47	4,126
Spain	$29,848.09	3,240
Turkey	$11,664.16	3,024
Dominican Republic	$14,588.39	2,736
El Salvador	$18,276.06	2,734

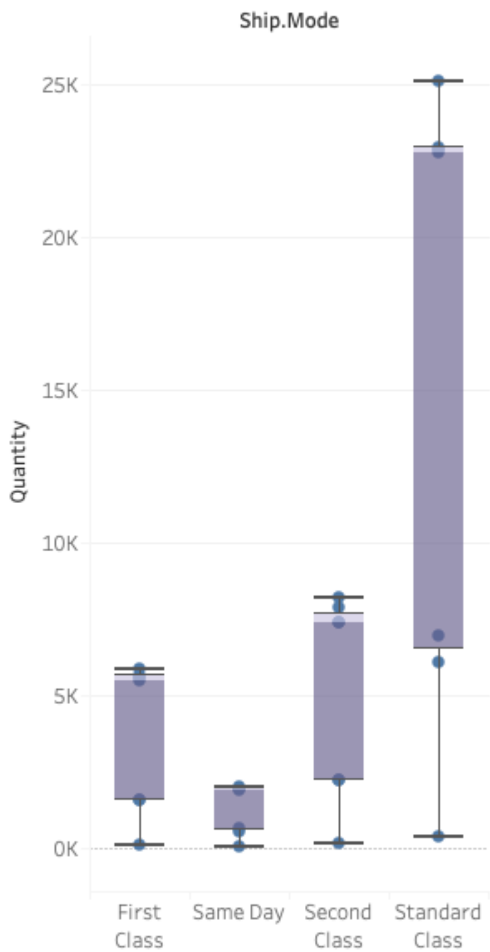
Description: The above table lists the countries with the highest shipping costs and the respective quantities of consignments shipped. Here again the United States paves the way by having the highest number of consignments shipped (37,873). As a result, they also account for the high shipping cost incurred ($238,173).

Highest Shipping Cost by Cities
Sorted by Total Shipping Cost and Total Quantity

City	Shipping.C..	Quantity
New York City	$26,948.17	3,417
Los Angeles	$18,777.07	2,879
Philadelphia	$12,080.88	1,981
Seattle	$11,581.45	1,590
Manila	$11,506.06	1,661
San Francisco	$10,753.24	1,935
Sydney	$11,183.66	975
Jakarta	$10,735.47	1,226
London	$9,890.22	1,033
Santo Domingo	$9,220.42	1,649
Managua	$9,393.94	1,233
Mexico City	$8,860.27	1,143
Gold Coast	$8,824.94	728
Brisbane	$8,405.25	689
Bangkok	$7,654.49	1,083

Description: Here the table lists down the cities across the world by the highest shipping costs and the total quantity of consignments shipped. Through our immediate observation we notice that out of the top 15 cities 5 of them are from the United States which explain our observation from the previous table where the United States ranks first in the classification of the countries. Here New York City is number one city with the largest quantity shipped(3,417) along with the highest Shipping Cost ($26,984.17)

Whisker Plot for Total Shipment Quantity by Class (Across Markets)



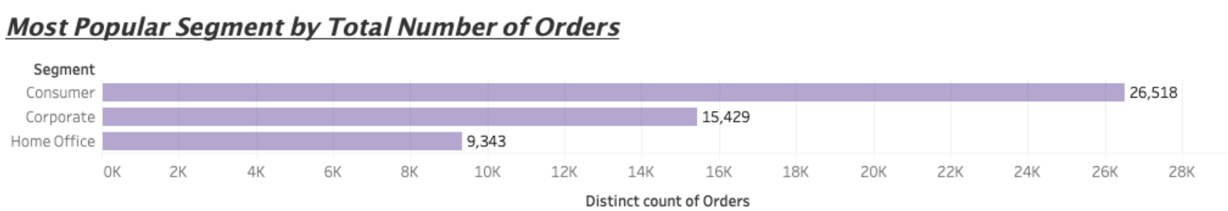
Readings: In the box plot above, The information with respect to the whiskers are listed below.

FIRST CLASS	Market	Quantity
Lower Whisker	Canada	139
Upper Whisker	APAC	5,871
Median	EU	5,518

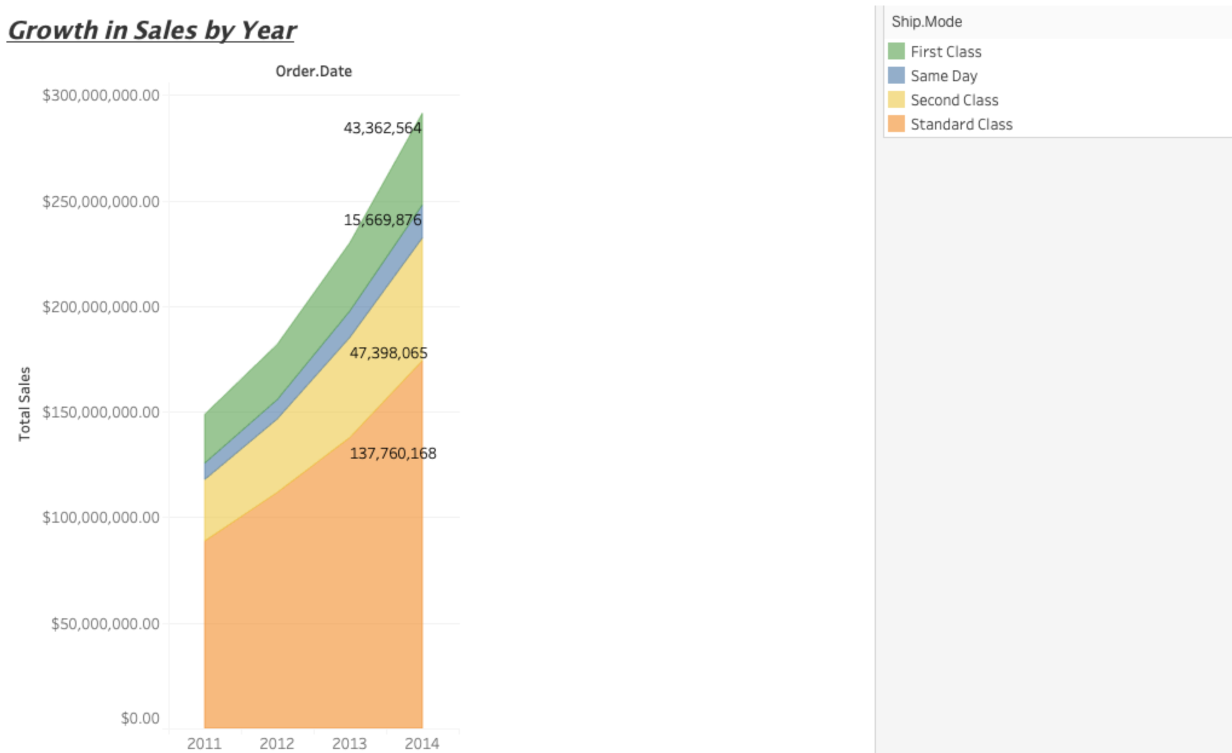
SAME DAY	Market	Quantity
Lower Whisker	Canada	70
Upper Whisker	LATAM	2,030
Median	EU	1,904

SECOND CLASS	Market	Quantity
Lower Whisker	Canada	200
Upper Whisker	APAC	8,238
Median	EU	7,421

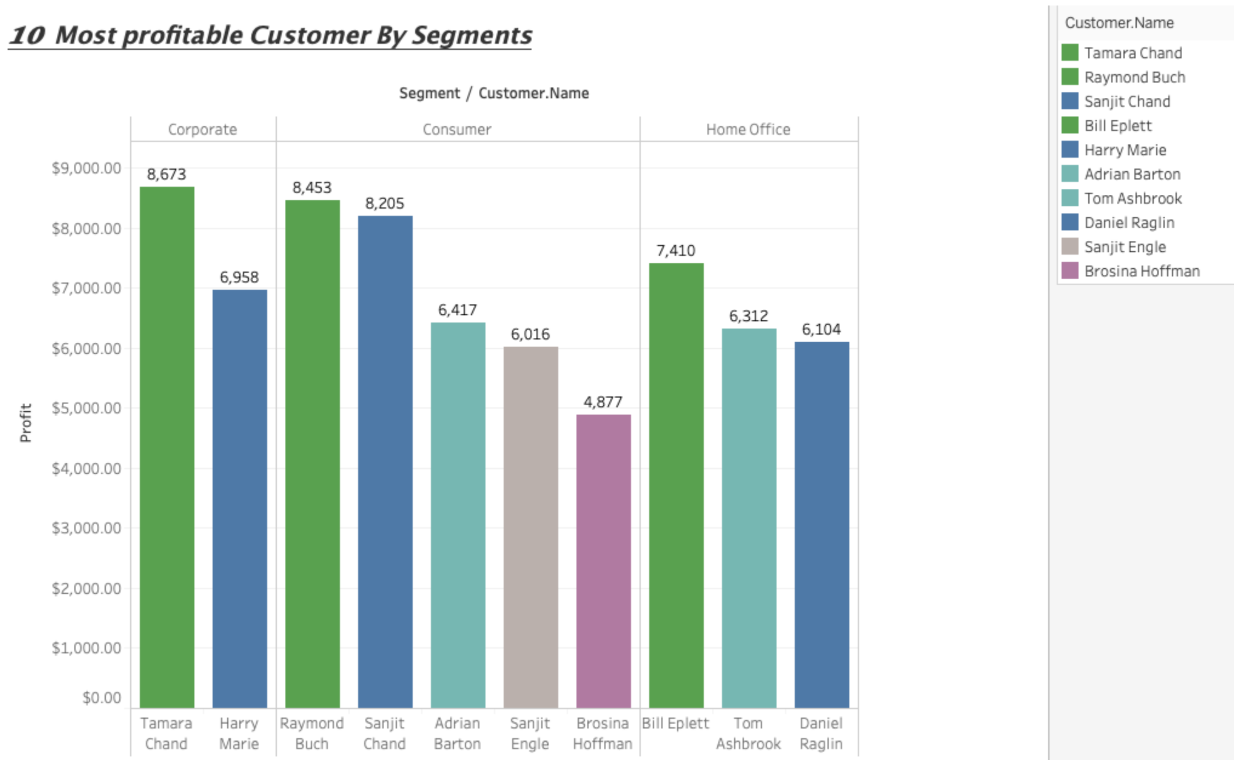
STANDARD CLASS	Market	Quantity
Lower Whisker	Canada	424
Upper Whisker	APAC	25,140
Median	US	22,797



Description: In this next horizontal bar graph, we can study the most popular segment in relation to the total number of orders. We can clearly see that Consumer is the most popular segment with 26,518 orders; followed by the Corporate Segment with 15,429 orders and the Home Office being the least Popular with 9,343.



Description: In the area graph for growth in sales by year, we can see a steady increase in the total sales across all the shipping modes and for the years 2011 to 2014. The standard class shipping constitutes the highest amount of sales across the entire spectrum. It has a staggering sales of $137,760,168 in the year 2014. The next big segment is constituted by second class with $47,398,065 in the year 2014 followed by first class and same day with $43,362,564 and $15,669,876 in the year 2014 respectively.



Description: The displayed side-by-side bar graphs depicts the 10 most profitable customers grouped by the purchases made in the three different segments. We find that Tamara Chand is the most profitable customer representing the Corporate segment with a profit of $8,673, followed by Raymond Buch from the Consumer segment generating $8,453 in profits and Bill Eplett from the Home Office segment generating 7,410 in profits. These the top three customers for Superstore can be identified as their highly valued customers as they are extremely profitable for the company.

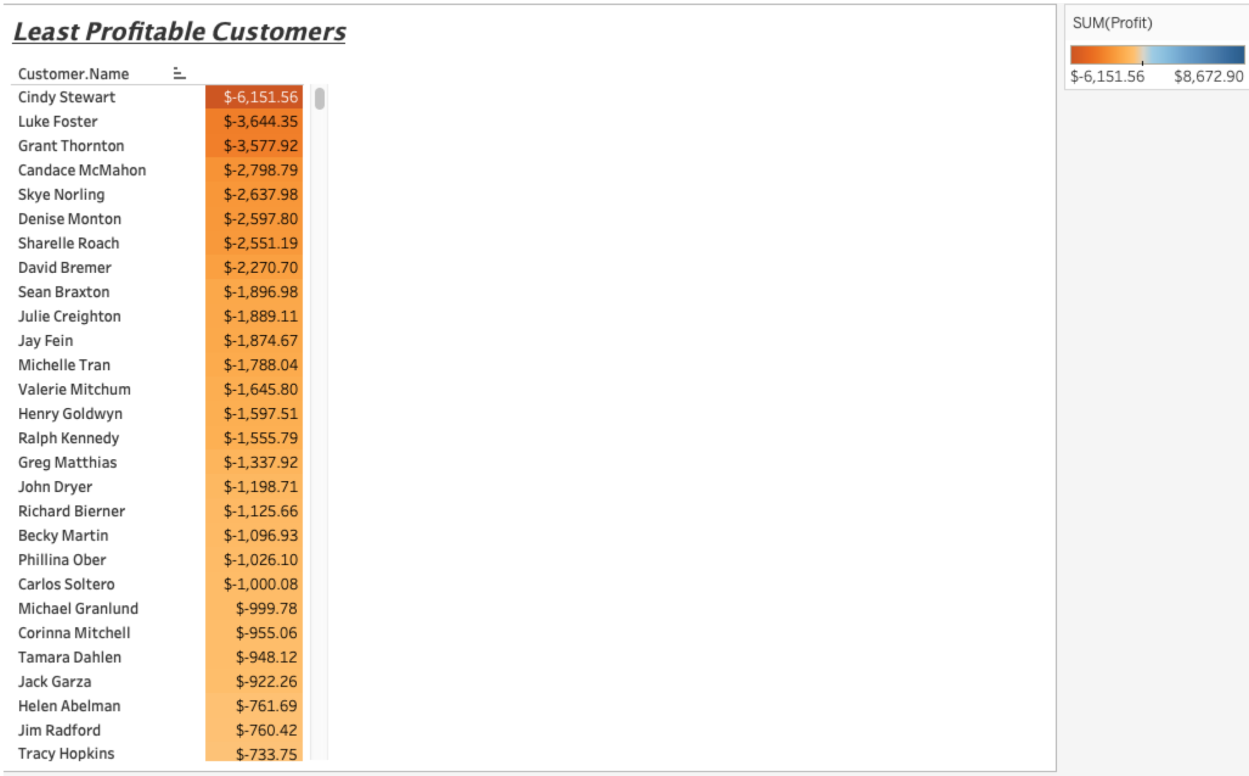
Most Profitable Customers

Customer.Name	
Tamara Chand	$8,672.90
Raymond Buch	$8,453.05
Sanjit Chand	$8,205.38
Hunter Lopez	$7,816.57
Bill Eplett	$7,410.01
Harry Marie	$6,958.29
Susan Pistek	$6,484.41
Mike Gockenbach	$6,458.68
Adrian Barton	$6,417.28
Tom Ashbrook	$6,311.98
Jane Waco	$6,265.85
Daniel Raglin	$6,103.97
Sanjit Engle	$6,015.78
Bill Shonely	$5,968.91
Ellis Ballard	$5,848.75
Nathan Mautz	$5,789.12
Christopher Conant	$5,603.33
Keith Dawkins	$5,486.17
John Huston	$5,395.82
Greg Tran	$5,214.13
Rick Wilson	$5,177.43
Todd Sumrall	$5,093.97
Dianna Wilson	$5,029.09
Carlos Daly	$5,028.91
Adam Bellavance	$4,979.98
Jill Fjeld	$4,877.69
Brosina Hoffman	$4,876.94
Janet Lee	$4,730.69

SUM(Profit)

$-6,151.56 $8,672.90

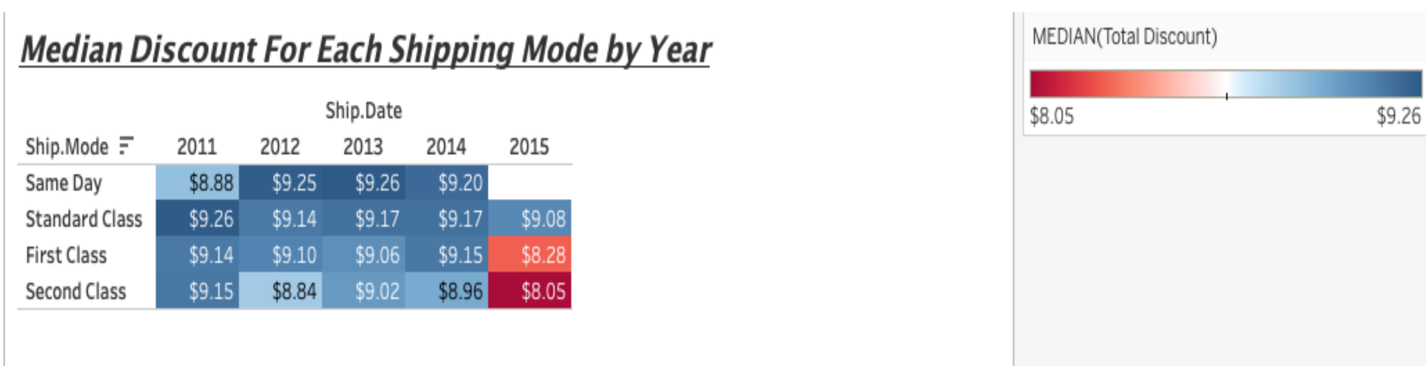
Description: The above is the list of the most profitable customers supporting our previous finding of Tamara Chand and Raymond Buch being the top two most profitable customers for the company. After calculating the profitable customer's customer lifetime value if they show a high value for the long term association with the company, superstore must continue to maintain a strong customer satisfaction score with such customer in order for their retention.



Least Profitable Customers

Customer.Name	SUM(Profit)
Cindy Stewart	$-6,151.56
Luke Foster	$-3,644.35
Grant Thornton	$-3,577.92
Candace McMahon	$-2,798.79
Skye Norling	$-2,637.98
Denise Monton	$-2,597.80
Sharelle Roach	$-2,551.19
David Bremer	$-2,270.70
Sean Braxton	$-1,896.98
Julie Creighton	$-1,889.11
Jay Fein	$-1,874.67
Michelle Tran	$-1,788.04
Valerie Mitchum	$-1,645.80
Henry Goldwyn	$-1,597.51
Ralph Kennedy	$-1,555.79
Greg Matthias	$-1,337.92
John Dryer	$-1,198.71
Richard Bierner	$-1,125.66
Becky Martin	$-1,096.93
Phillina Ober	$-1,026.10
Carlos Soltero	$-1,000.08
Michael Granlund	$-999.78
Corinna Mitchell	$-955.06
Tamara Dahlen	$-948.12
Jack Garza	$-922.26
Helen Abelman	$-761.69
Jim Radford	$-760.42
Tracy Hopkins	$-733.75

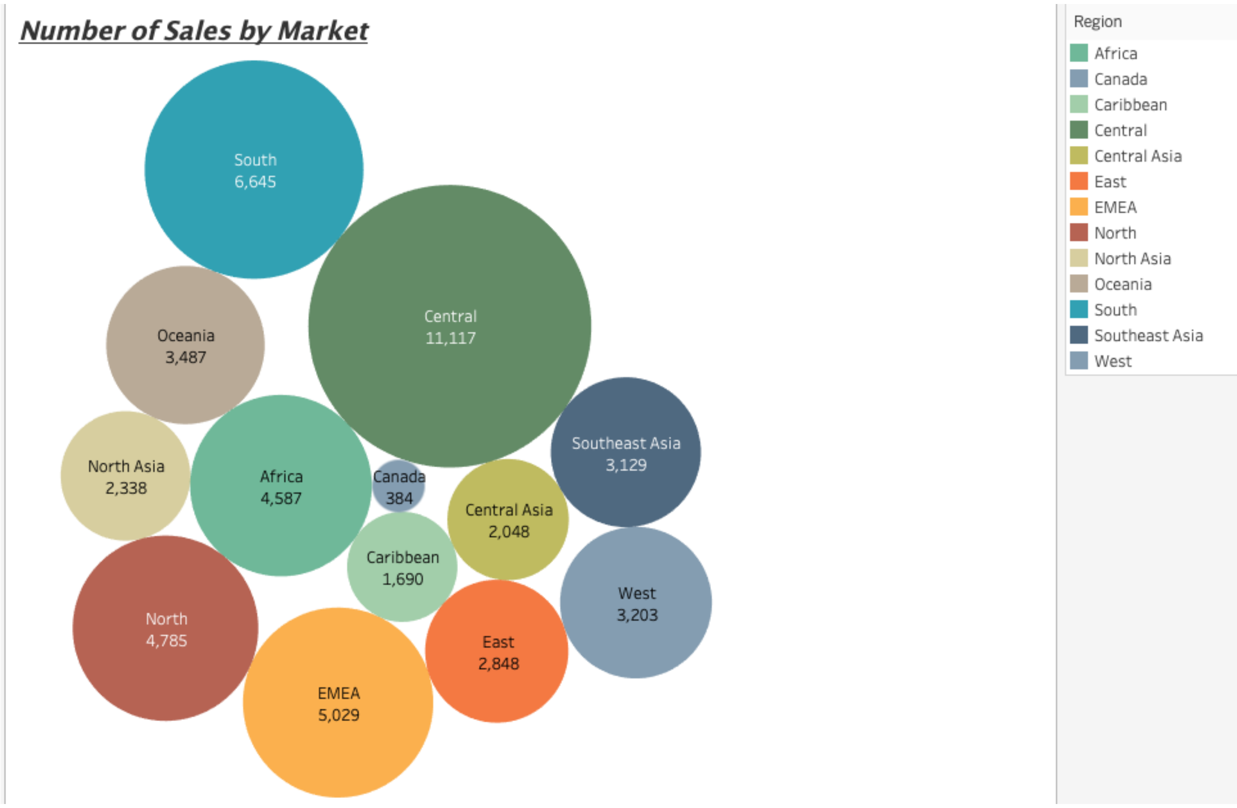
Description: The above list depicts the non-profitable clients for the company. As per the list Cindy Stewart caused the superstore a $6,151.56 loss. Clients cursing high negative defits to the company can reduce its overall profitability. For such customers, if they end up yielding negative customer lifetime value then it is best for the company to skillfully divest from such accounts. This would reduce the future bleeding of the company and help divert the valuable financial resources to the customers who have a high customer lifetime value.



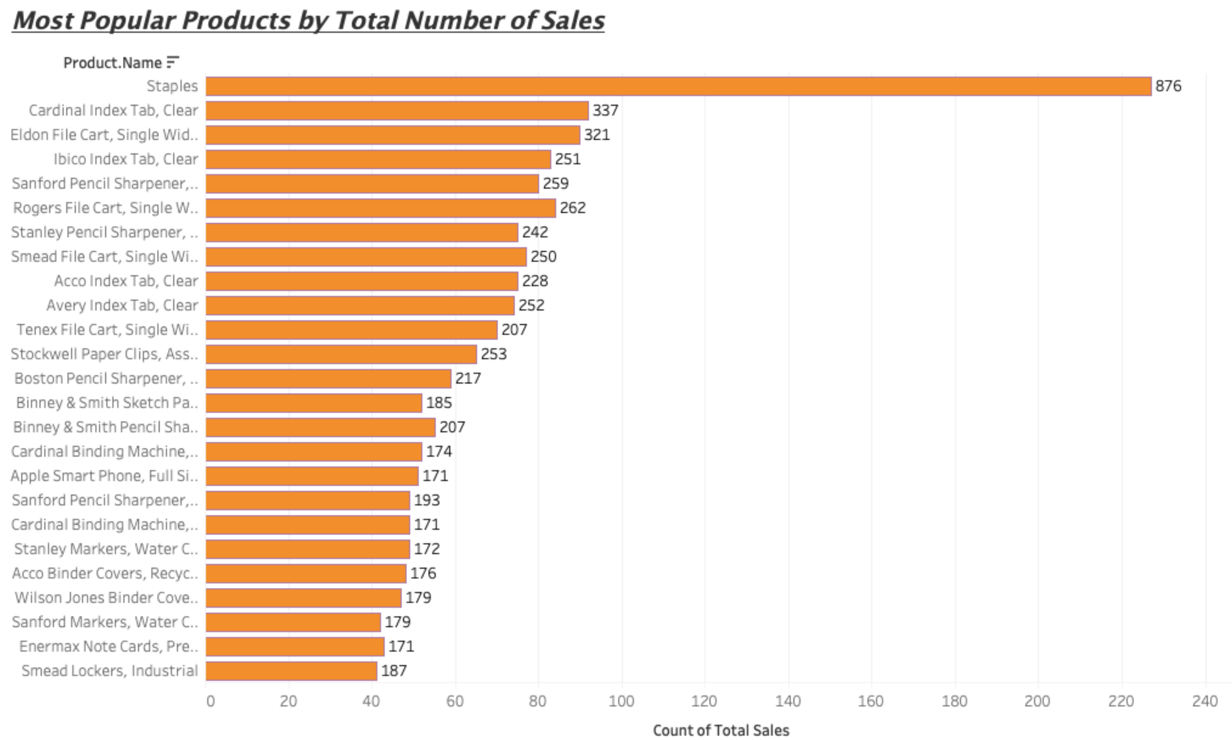
Median Discount For Each Shipping Mode by Year

Ship.Mode	2011	2012	2013	2014	2015
Same Day	$8.88	$9.25	$9.26	$9.20	
Standard Class	$9.26	$9.14	$9.17	$9.17	$9.08
First Class	$9.14	$9.10	$9.06	$9.15	$8.28
Second Class	$9.15	$8.84	$9.02	$8.96	$8.05

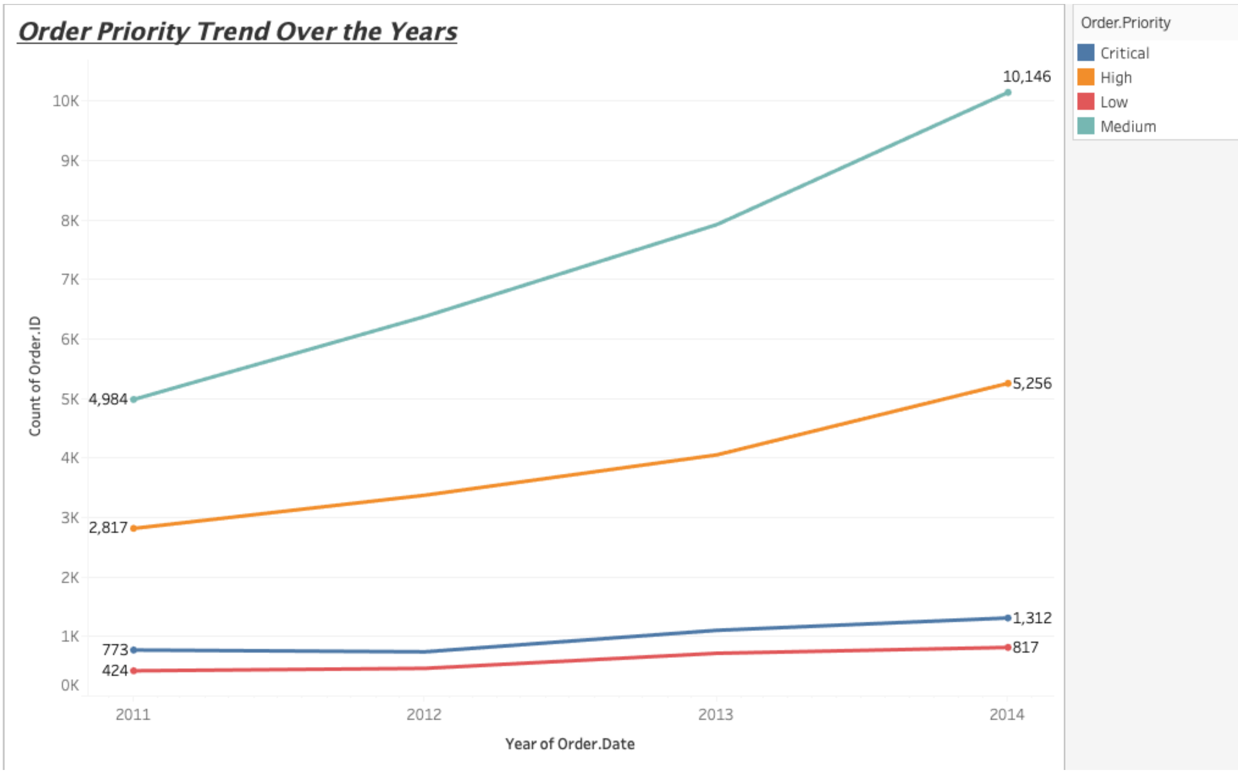
Description: The heat map showcases the median discount for the different shipping modes as per the respective year. Here we notice that the highest amount of discounts were given in the years 2011 and 2013 for $9.26 on a shipment for Standard Class and Same day shipping respectively. It seems that the discounts given out in the most recent year are going down and have given out as little as $8.05 in 2015. This can be looked at a positive light, as the less amount is being utilized by the company to give large discounts. Hence, the company is saving more than before.



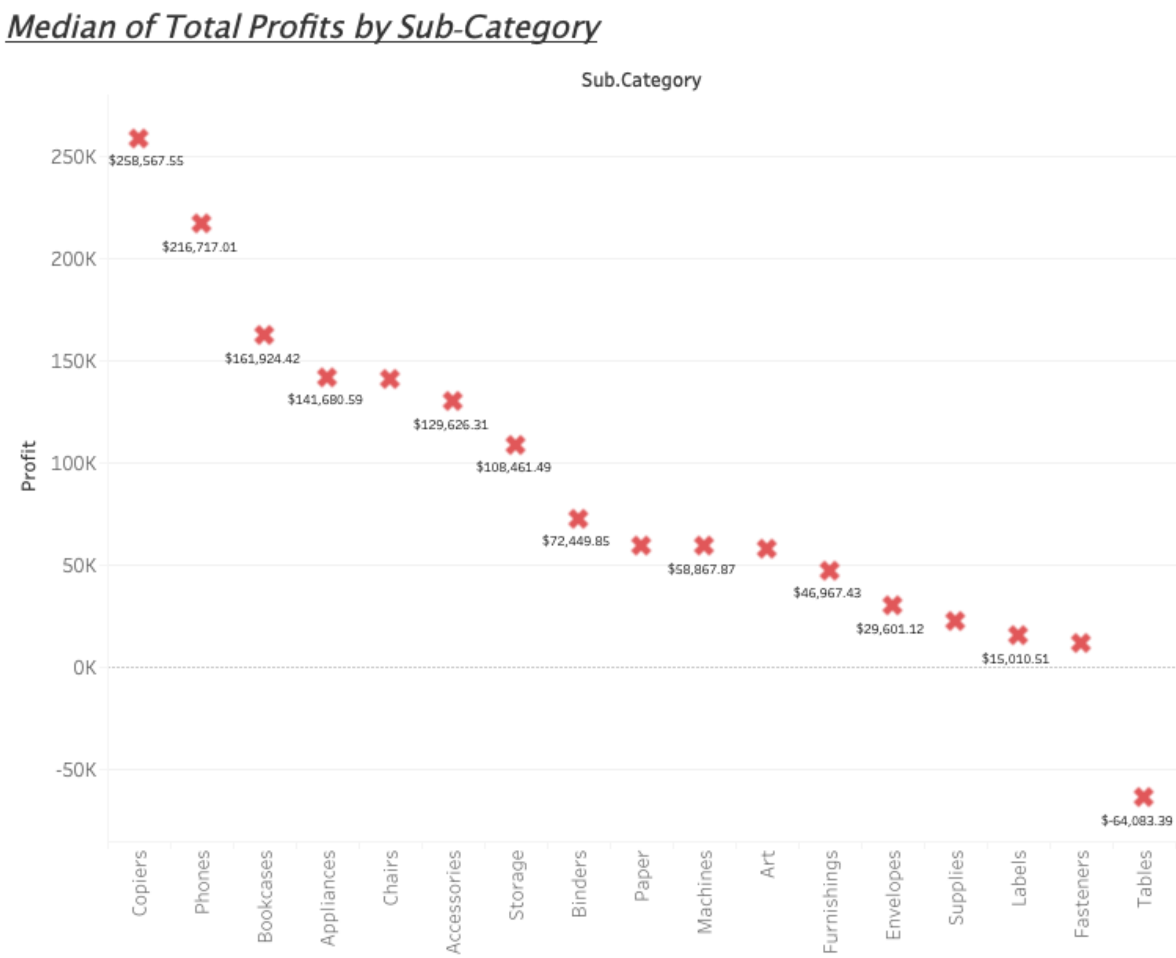
Description: In the above visualization, we can infer the size of the different regional markets purely based on the number of sales. The largest market size by sheer number sales is Central at 11,117 total sales followed by South with 6,645 total sales. And the smallest market is Canada with only 384 total number of sales.



Description: This visualization of the horizontal bar graph showcases the most popular products by the total number of sales. By the statistics we can see that Staples is the most popular item at the store selling a total of 876 units. There is a huge difference in the numbers of item one and item two by which we can estimate the high demand and the popularity of the staples. The number two and three most popular products are of the Cardinal Index Tab,Clear with 337 units solds and Eldon File Cart, Single Width with 321 units sold.



Description: The order priority trendline depicts the rise in the order priority trend from the year 2011 - 2015. Each and every type of order priority has seen a rise in the number of orders over the years. Out of which the highest jump was witnessed for the medium priority orders where it showed a rise from 4,984 to 10,146 followed by high priority order showing progression from 2,817 to 5,256 over the years. The critical and low priority orders remain low as compared to the other two.



Description: The visualization above illustrates the median of the total profits laid out by the different sub-categories. The highest median of the profits out of all the categories is for the copiers at $258,567.55 whereas the sub-category tables show a negative value indicating a loss of $64,083.39. If tables continue to yield a steady loss over the years then the company should do a thorough analysis on whether to address any fault lines or consider eliminating them from their selection.