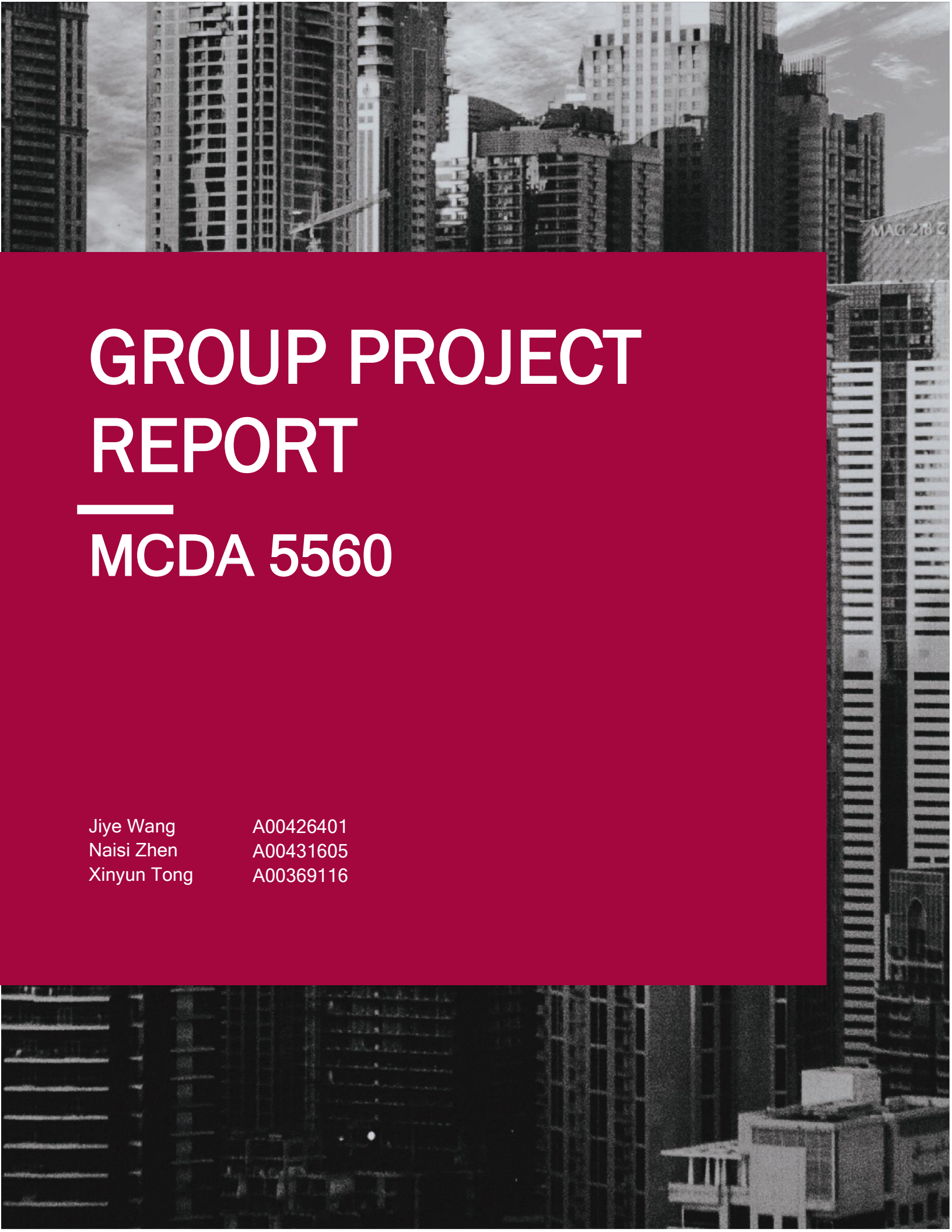


GROUP PROJECT REPORT

MCDA 5560

Jiye Wang	A00426401
Naisi Zhen	A00431605
Xinyun Tong	A00369116

Contents

Executive Summary 3
Introduction 3
Analysis 4
Conclusions 23
Reference 25
Appendix 26
Table of Figures 29

Executive Summary

Key Performance Index (KPI) is an important performance measurement. Without clearly defined KPIs, we wouldn't know if the company was doing good or bad, and to which degree the company was performing. Every business is profit-orientated, we determined that "Sum of Profit" is the most important KPI, assistant by "Sum of Sales", and "Average of Product Base Margin". As there was no "On-Time Delivery Rate" that was provided in the dataset, we used "Order Date" and "Shiping Date" for analysis purposes. We recommend documenting "On-Time Delivery Rate" as a KPI in later business operation. Another critical KPIs that were not provided is Cost of Goods Sold and Operation Cost, which Shipping Cost is a part of. We recommend using them as KPIs in the future to draw the complete business picture.

In data visualization analysis we have concluded the following recommendations: reduce office furniture inventory, focus on technology product and office supplies, clearance out of date products, focus on BC, AB and ON market, target marketing to the high-value corporate customer, look into high return rate product and customer and reduce the processing cost. For further mining on the dataset, we perform association mining and cluster analysis in order to find valuable information on the products and customers. We use python to write the analysis models and change the parameters to find the models with the best performance.

The sales data are up to date and correct, the market interest for old school product is low, the office furniture market continues to shrink and there no major change in the structure for business customer and consumer. Since no Cost of Goods Sold and Operation Cost data were given, from the fact that "Furniture" had high Average of Product Base Margin but with a much lower profit than the others, we assumed "Furniture" had a much higher operation cost (which Average of Shipping Cost is a part of). We also assume there are certain rules in the office products in the sales dataset and the customer segmentation is based on the login information filled by customers.

Introduction

In this project, we have analyzed the sales transaction dataset from one retailer's Canada market (we bet this retailer is Staples). We performed data visualization analysis, association analysis, clustering analysis, SWOT analysis and identified KPI and create a Pivot table to investigate this company's performance covered the year from 2009 to 2012. At the end of this report, we concluded all recommendations associated with our analysis.

From the dataset, we found that there are three main categories of products the company sales, namely "Furniture", "Office Supplies", "Technology"; sales were generated from all Canadian territories over the documented periods.

Row Labels	Sum of Sales	Average of Product Base Margin	Sum of Profit	Average of Shipping Cost
Furniture	5178591	0.598555291	117433	30.8838109
Office Supplies	3752762	0.461270429	518021	7.829828633
Technology	5984248	0.556305085	886314	8.954886199
(blank)				
Grand Total	14915601	0.512513196	1521768	12.83855697

Figure 1 Cateogry SoS/AoPBM/SoP

Sum of Sales (SoS): measures the total revenue generated by selling the products, returned product excluded. Higher sales usually lead to higher profit.
Average of Product Base Margin (AoPBM): is the rate of the margin of the sell price, calculated as **(sell price – cost)/sell price**. The higher AoPBM, the more money can be generated from that product.
Sum of Profit (SoP): has been determined as the most important KPI for this project as businesses are a for-profit entity. Normally it's the SoS deducted by all costs. The higher of SoP, the more successful a business it is.

In term of volume of sales, all three categories are not much different from each other, average around 5 million. In term of Average of Product Base Margin, "Furniture" topped over the others with a nearly 60% Avg. Margin, a huge contradictory to the generated profit.
In term of profitability, all three departments are performing well, but "Furniture" generated a much lower "Sum of Profit" than the other two while the "Sum of Sales" was at a bigger or same level of the other two.

Analysis

SWOT analysis for the business model

Strengths: • Broad Geographic Coverage • Broad Product/Brand Coverage • High Margin on Product Sold	**Weaknesses:** • Lack of Star Products • Lack of Geographic Focus • High Operation Cost that eats a lot of margin
Opportunities: • Product Vertical Integration • Increase Shipping Efficiency/Effectiveness	**Threats:** • Inventory Cost • Product Value Depreciation over time

Figure 2 Business Model SWOT

Data visualization analysis

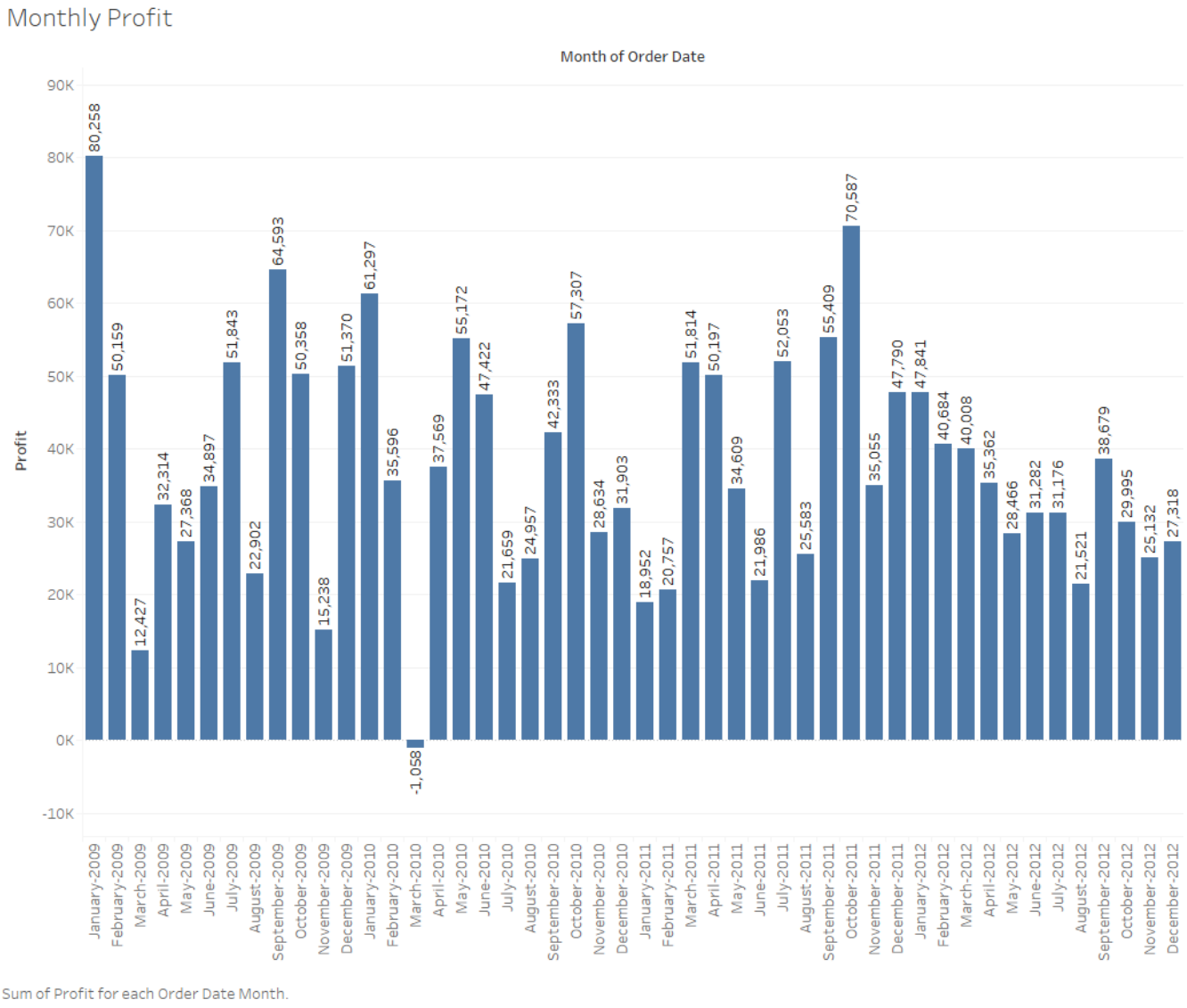


Figure 3 Monthly Profit Break Down

At the beginning of data visualization analysis, we have created a monthly profit break down bar chart to investigate the monthly profit performance. We can observe from the bar chart, from January 2019 to December 2012 all the months this company has positive profit except March 2010 which is losing money.

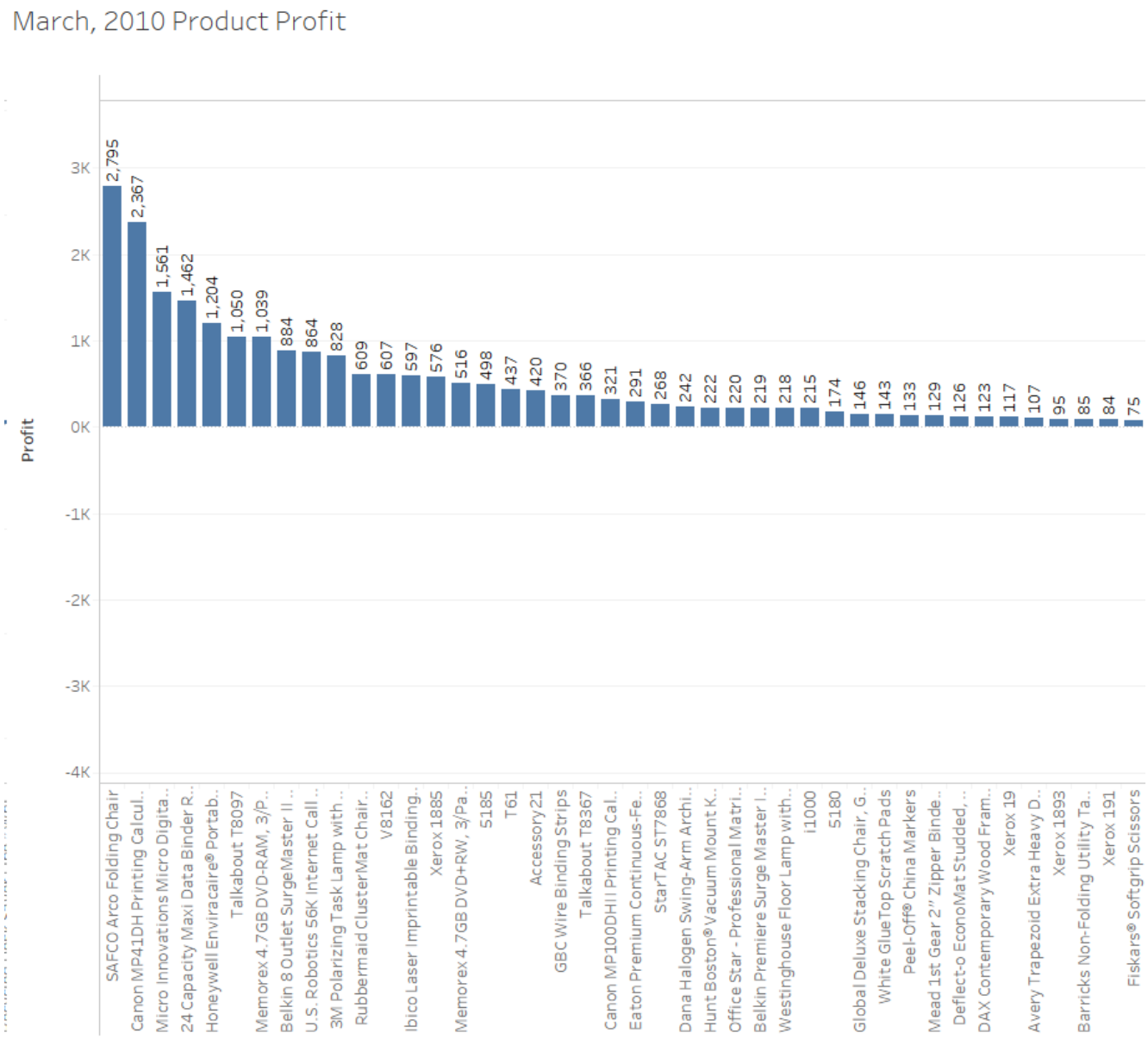


Figure 4 Category Profit Break Down

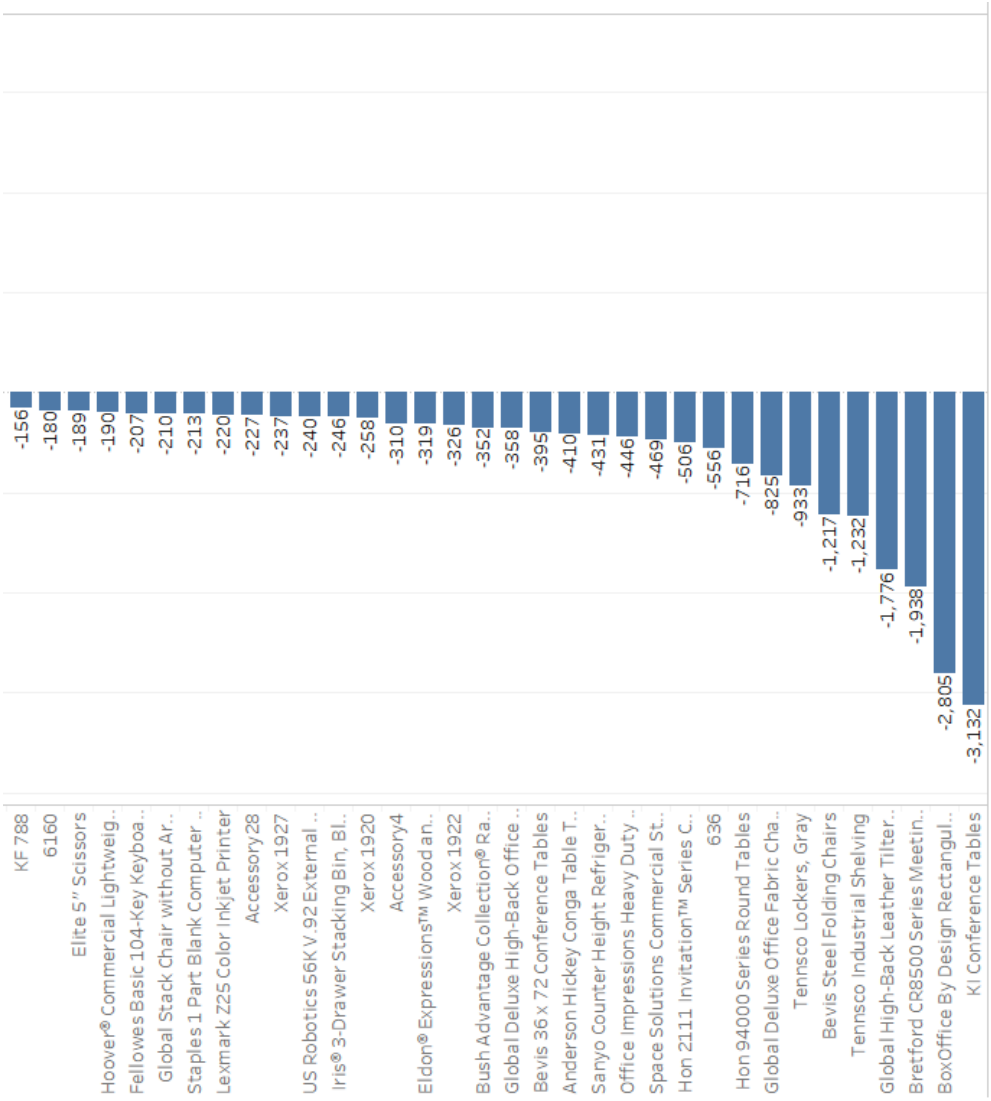


Figure 5 Category Profit Break Down

For further investigate the profit performance in March 2010, we created a bar chart to show all the products sold within March 2010 with net profit. From the chart, we can see “SAFCO Arco Folding Chair”, “Canon MP41DH Printing Calculator” and “Micro Innovations Micro Digital Wireless Keyboard and Mous, Gray” generated the most profit for that month. But “KI Conference Table”, “BoxOffice By Design Rectangular and Half-Moon Meeting Room Tables” and “Bretford CR8500 Series Meeting Room Furniture” are loose most of the profit for that month.

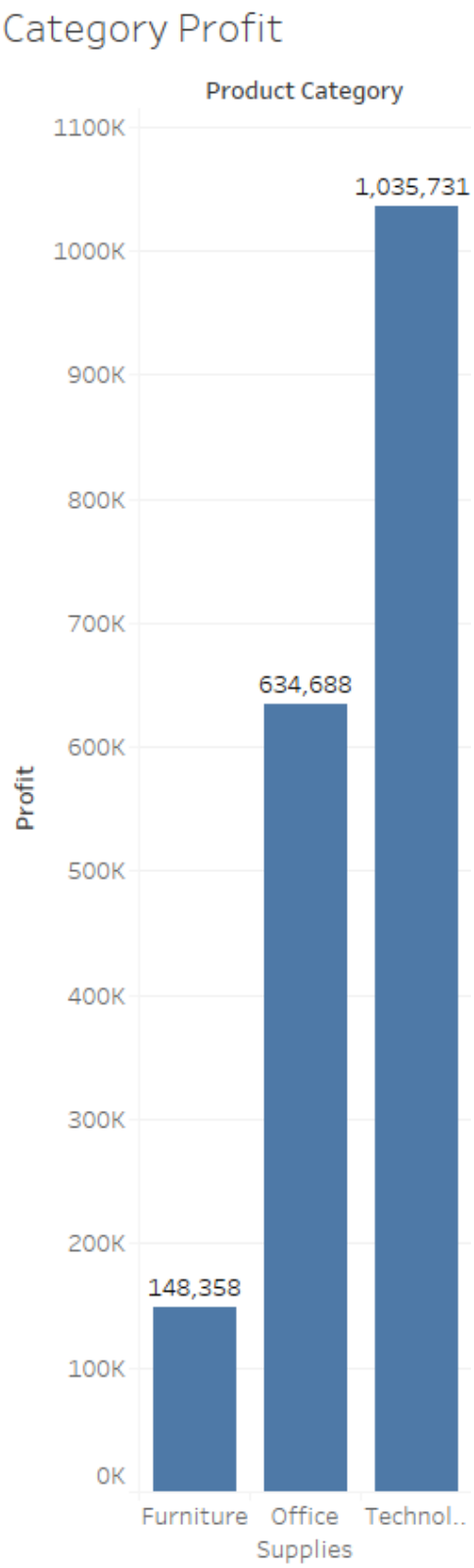


Figure 6 Category Profit Break Down

In this Category Profit Break Down bar chart, we can notice technology product generated the highest profit which is $1,035,731. The second one is office supplies which generated $634,688 profit. Furniture has the lowest performance which only generated $148,358 profit.

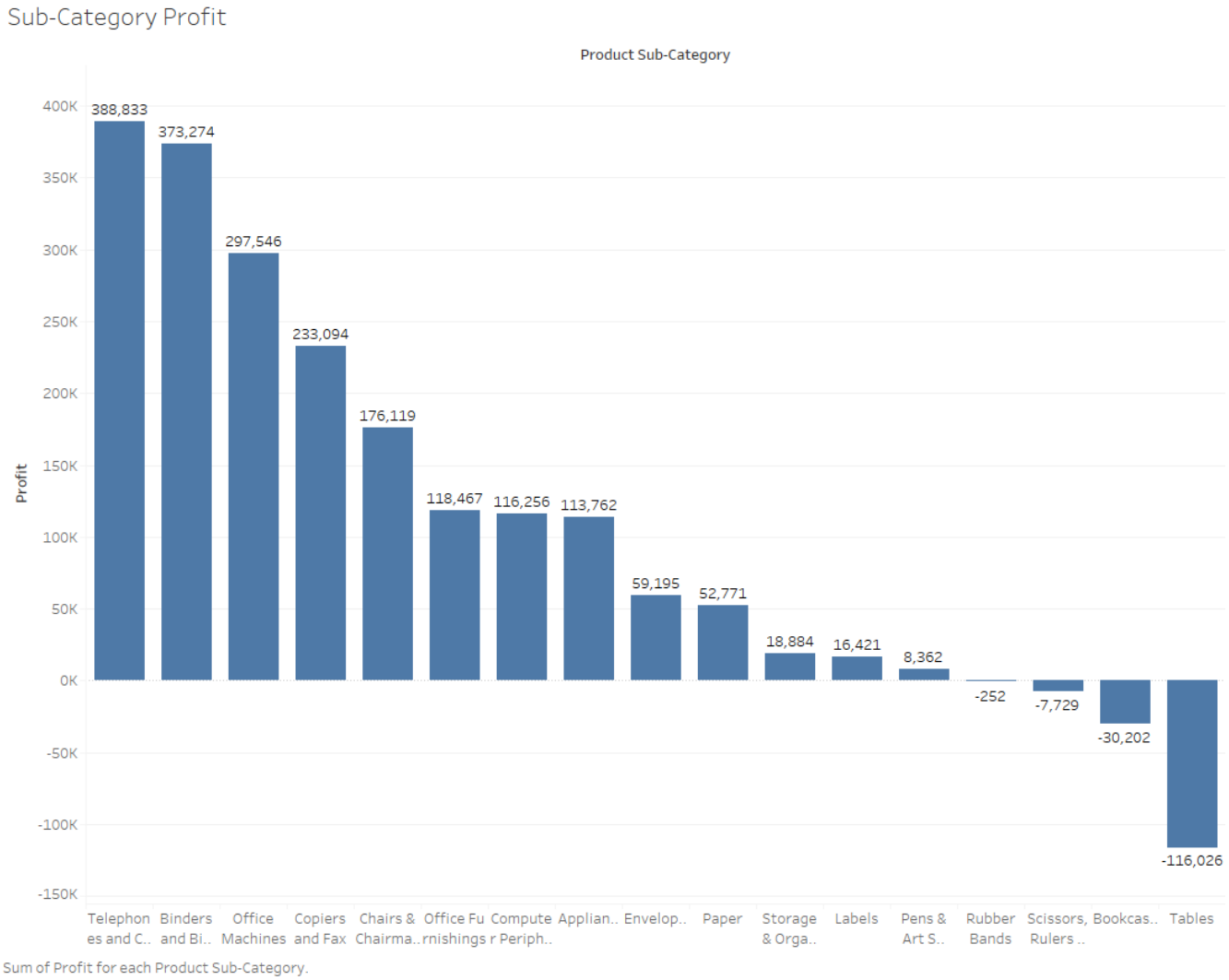


Figure 7 Sub-Category Profit Break Down

From this Sub-Category Profit Break Down chart, we can know “Telephones and Communication” is the highest profitable sub-category which earned $388,833 profit. The second high sub-category is Binders and Binder Accessories” which earned a total $373,274 profit. “Tables” has the worst performance which loose $116,026 profit and “Bookcase” total loose $30,202 in profit.

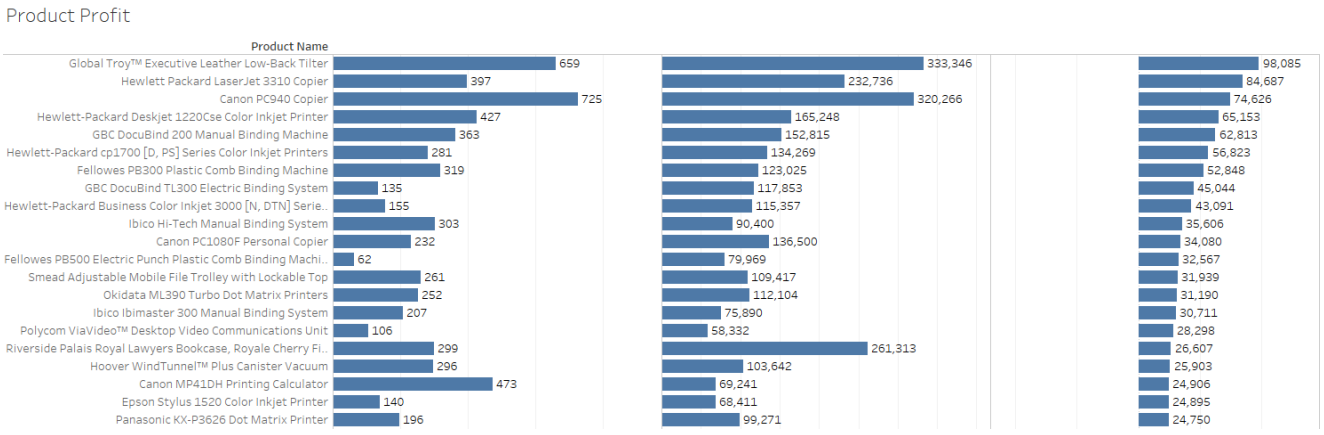


Figure 8 Product Profit Break Down

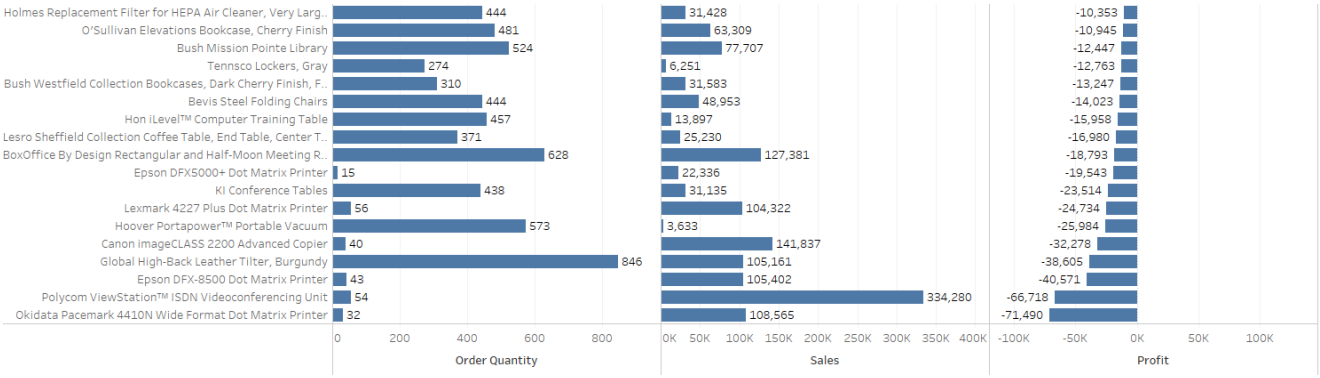
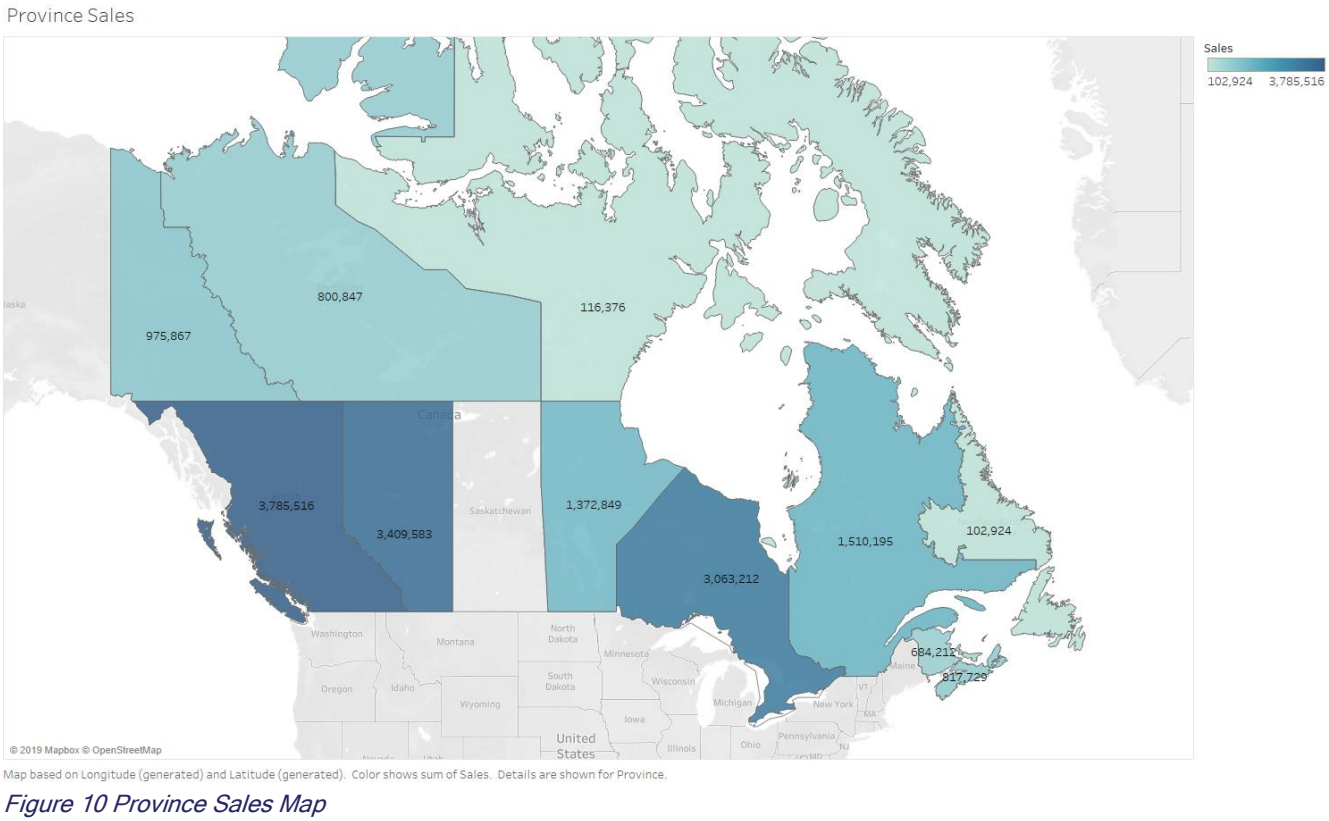


Figure 9 Product Profit Break Down

We look into further in products level, "Global Troy™ Executive Leather Low-Back Tilter" is the most profitable product which sold 659 with sales amount is $333,346 and generated a total $98,085 profit. "Hewlett Packard Laser Jet 3310 Copier" is the second profitable product which sold 397 with sales amount $232,736 and generated $84,687 profit. "Okidata Pacemark 4410N Wide Format Dot Matrix Printer" total sold 32 with sales $108,565 but loose $71,490 profit. "Polycom ViewStation™ ISDN Videoconferencing Unit" total sold 54 with sales $334,280 and lose $66,718 in profit.



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Sales. Details are shown for Province.

Figure 10 Province Sales Map

This province sales map told us that BC is the most important province for this company which total has $3,785,516 in sales, the second one is AB which has a total sales of $3,409,583. ON is the third large market in which total sales are $3,063,212.

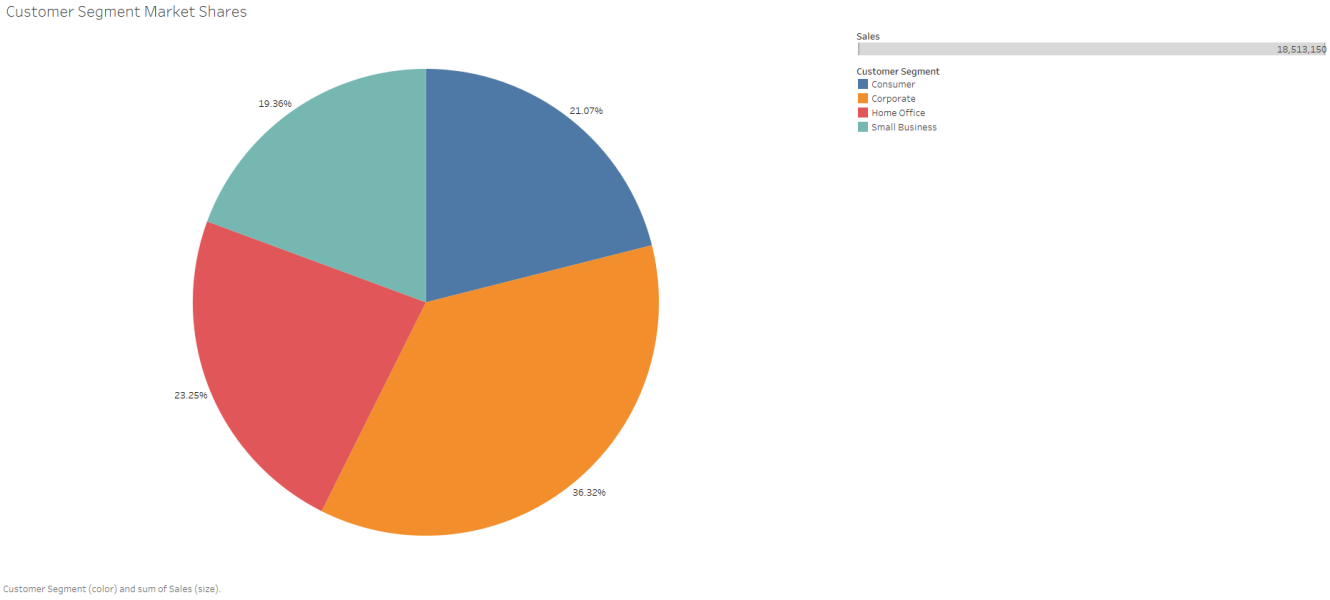


Figure 11 Customer Segment Shares

In the above Customer Segment Shares pie chart, we can see the largest customer segment is Corporate total occupied 36.32%. The second large customer segment is Home Office total occupied 23.25%. The third large customer segment is Consumer total occupied 21.07%. The smallest customer segment is Small Business which occupied 19.36%.

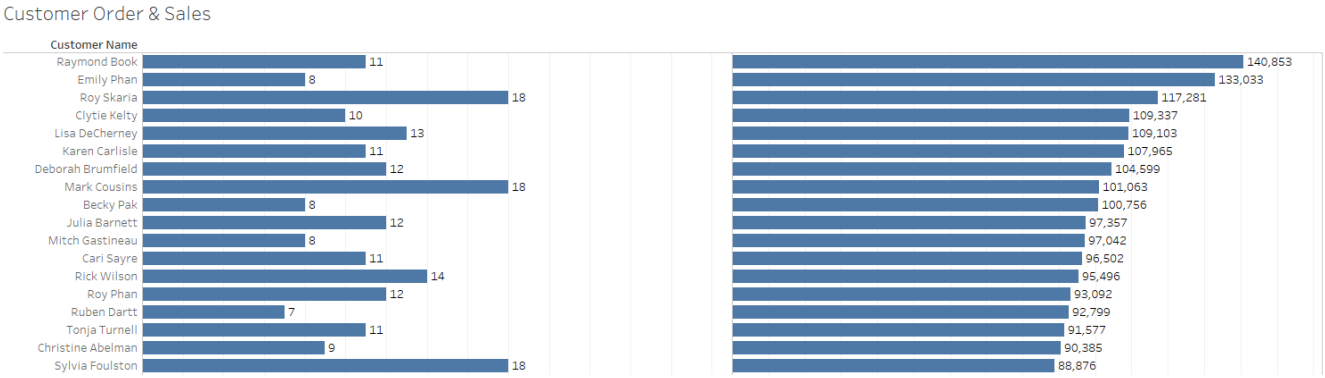
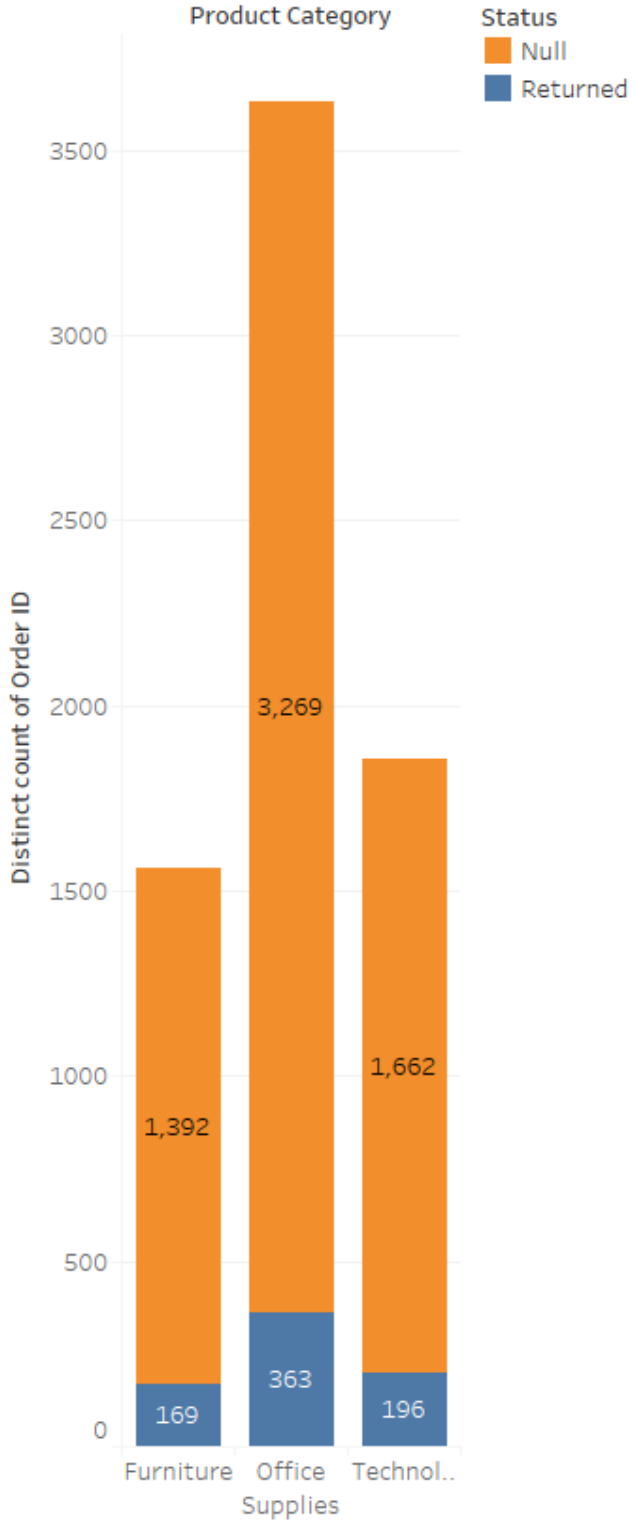


Figure 12 Customer Order & Sales

We created a customer order and sales chart and sorted by total sales amount. This chart let us know who has contributed the most sales for the company. Raymond Book has placed total of 11 orders and contributes $140,853 in sales. Emily Phan has placed total of 8 orders and contributes $133,033 in sales. Roy Skaria has placed 18 orders and contributes total $117,281 in sales.

Category Return



Distinct count of Order ID for each Product Category. Color shows details about Status. The view is filtered on Status, which keeps Null and Returned.

Figure 13 Category Return

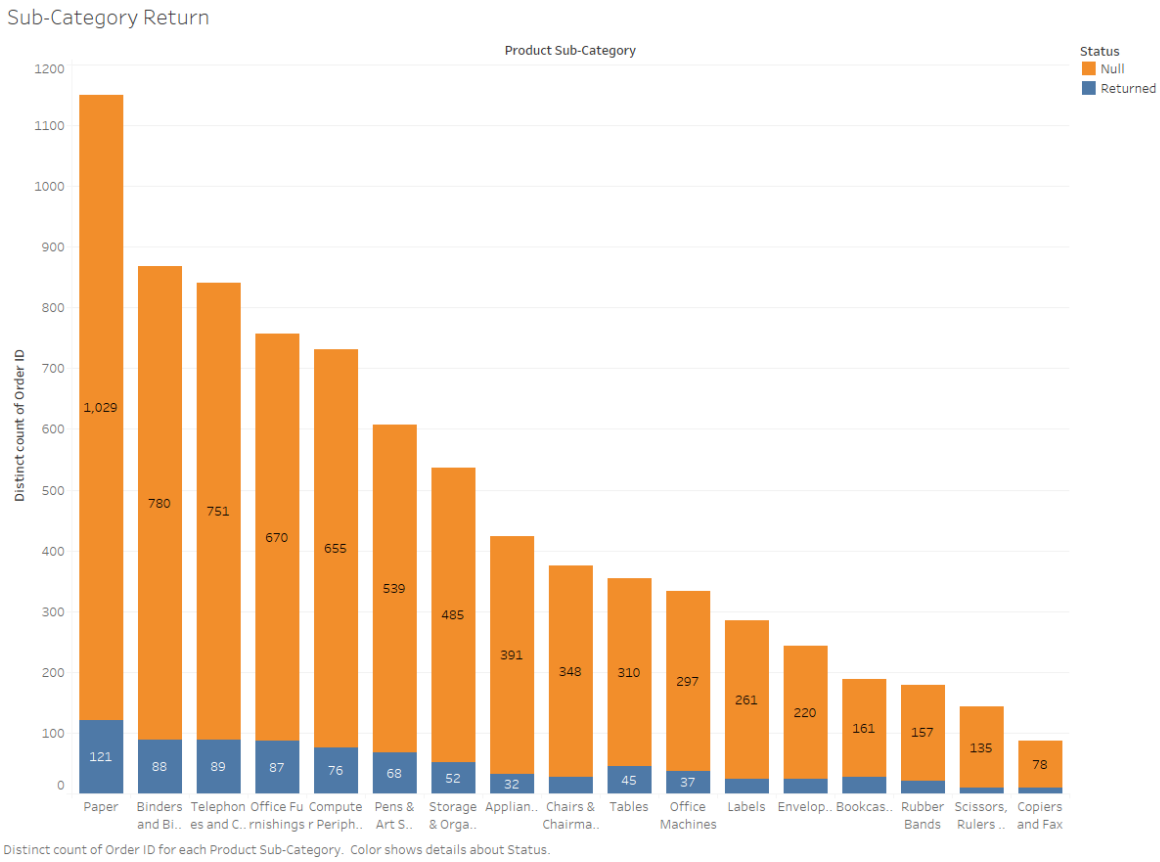
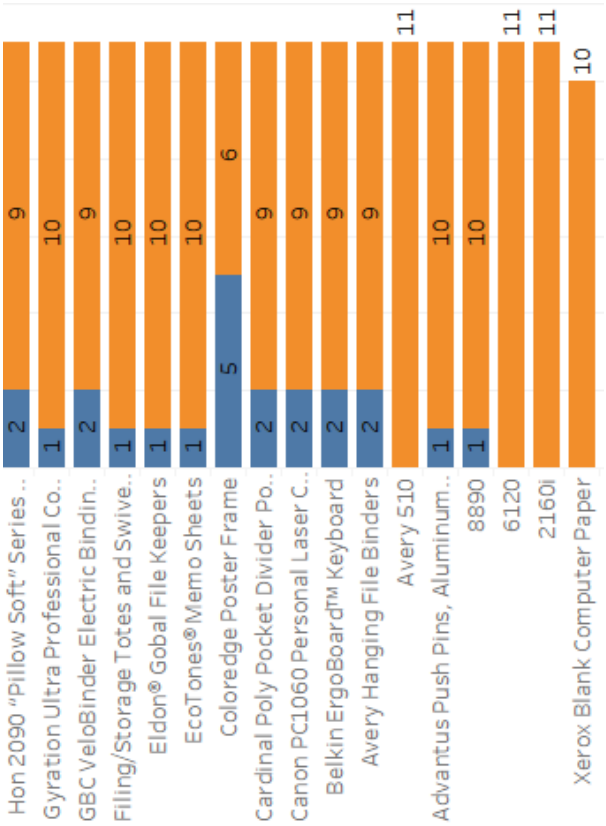
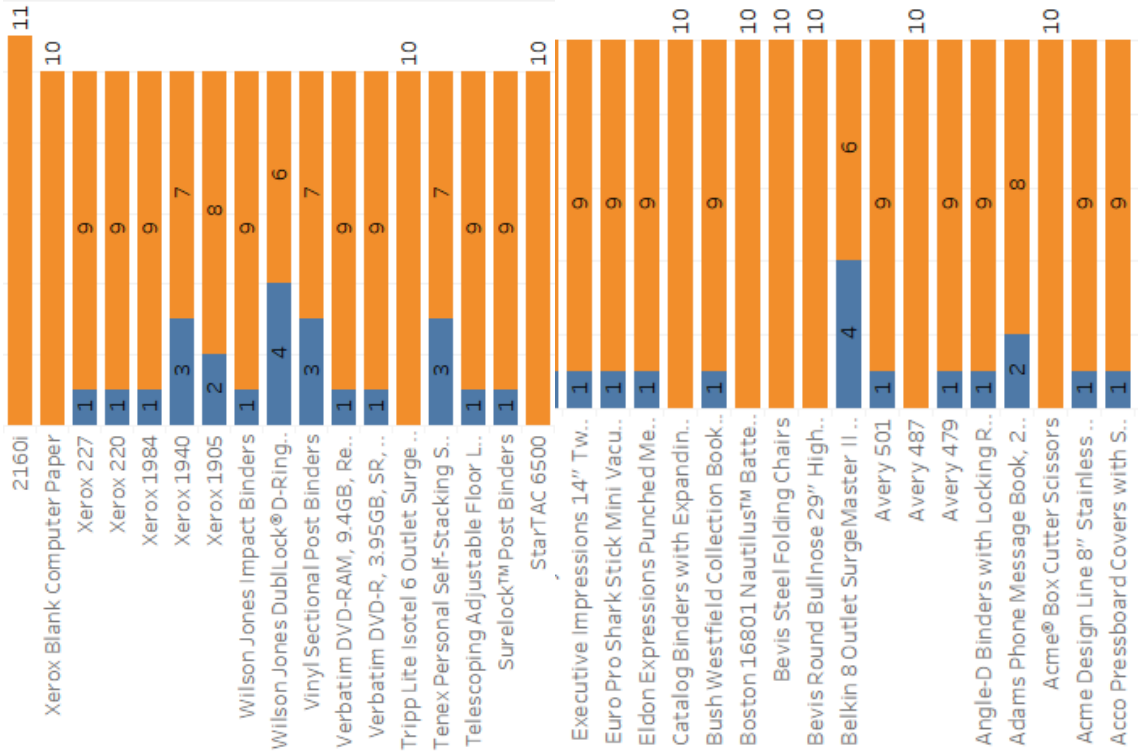


Figure 14 Sub-Category Return

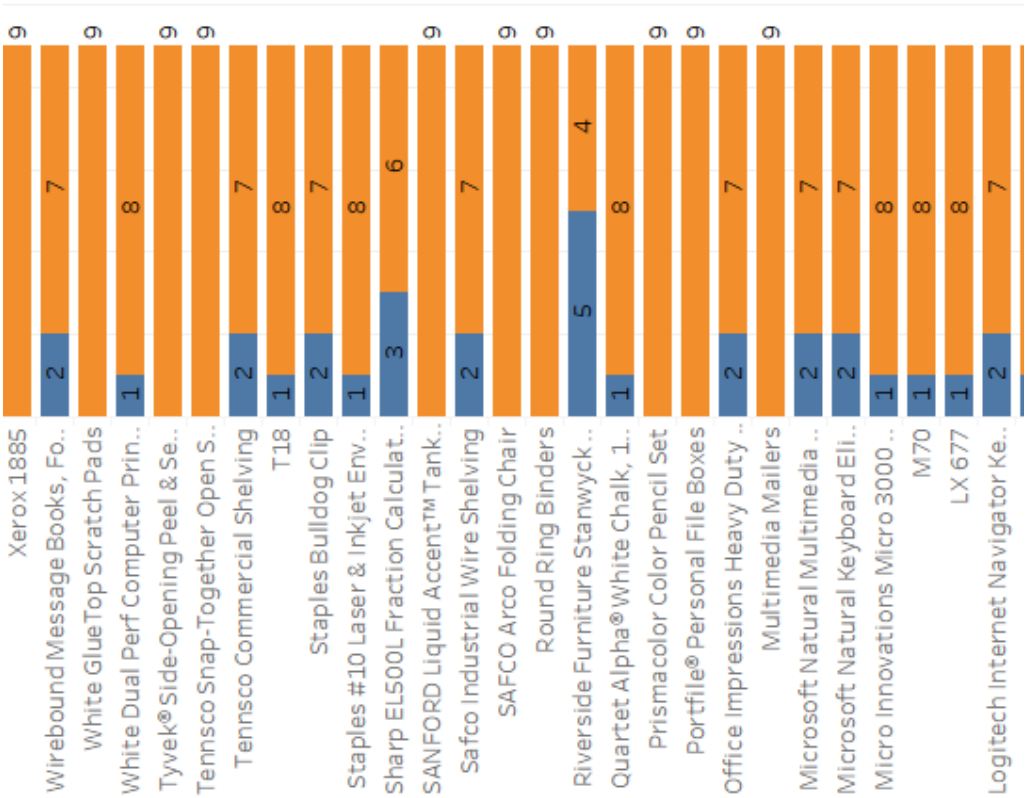
The return ratio in both category return and sub-category return all remains a reasonable ratio. There is nothing to worry about.



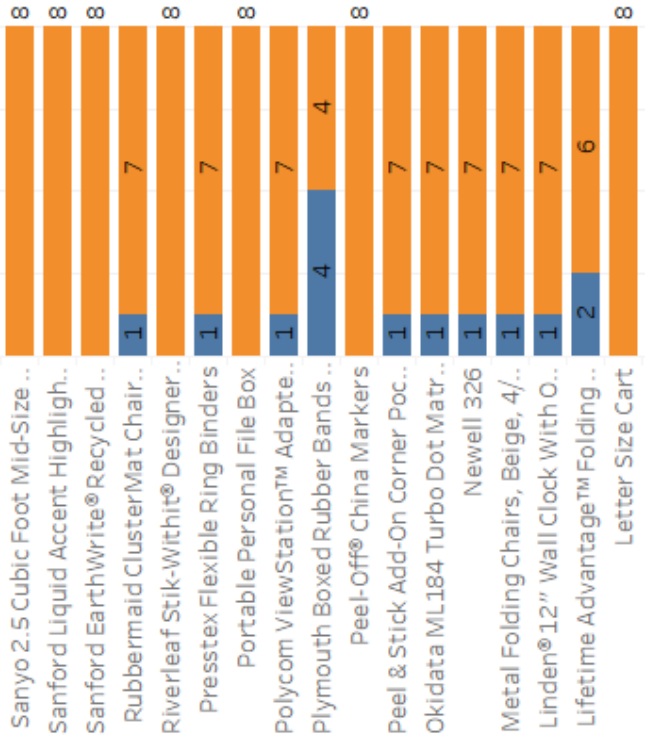
Product return ratio chart can tell us the return ratio of each product. For example, Coloredge Poster Frame has total 11 order but five of them are returned.



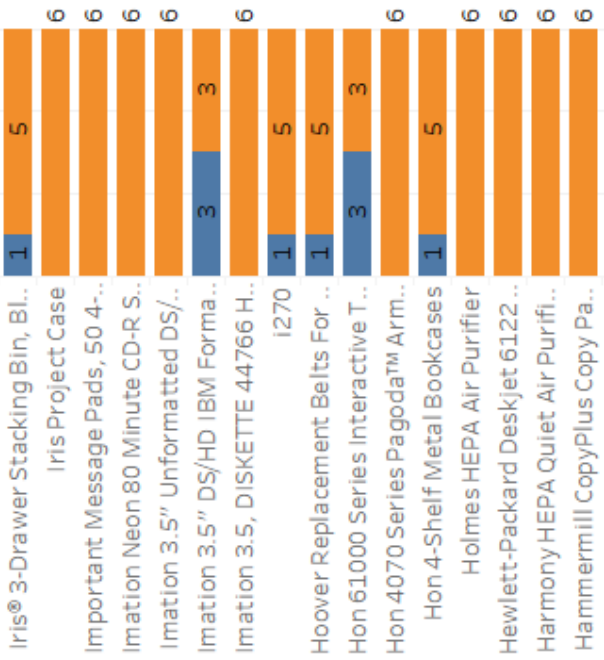
Wilson Jones DubLock D-Ring Binder and Belkin 8 Outlet SurgeMaster II Gold Surge Protector each has total 10 orders but 4 of them are returned.



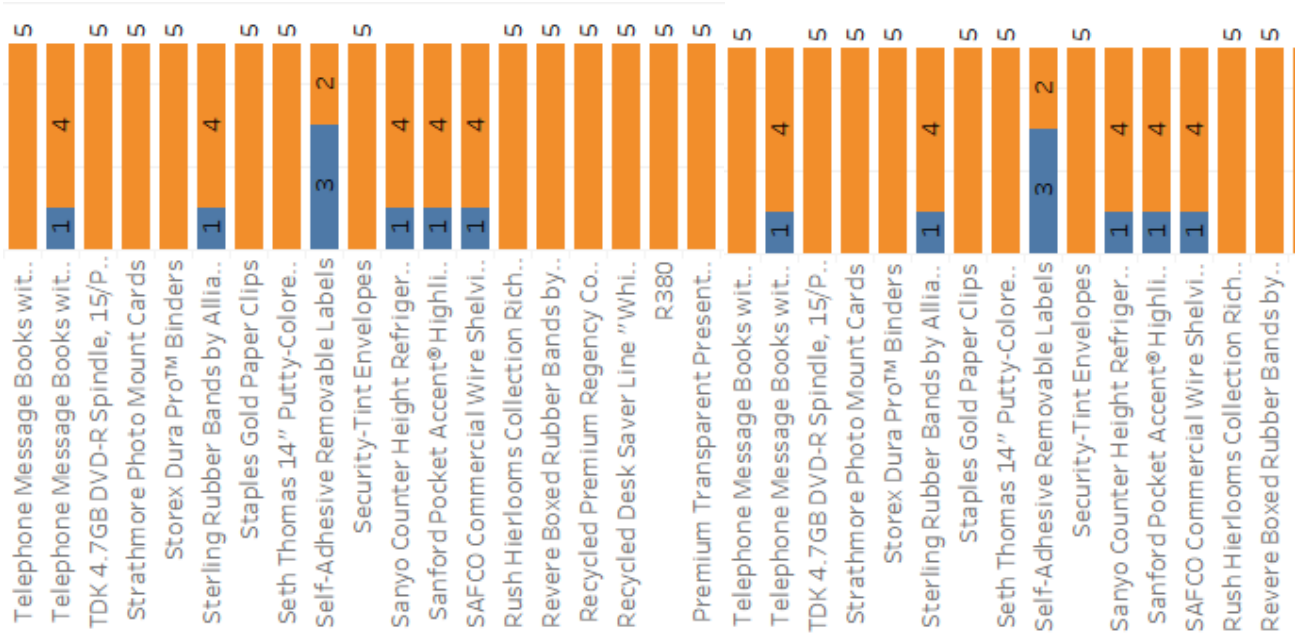
Riverside Furniture Stanwyck Manor Table Series has total of 9 orders but five of them are returned.



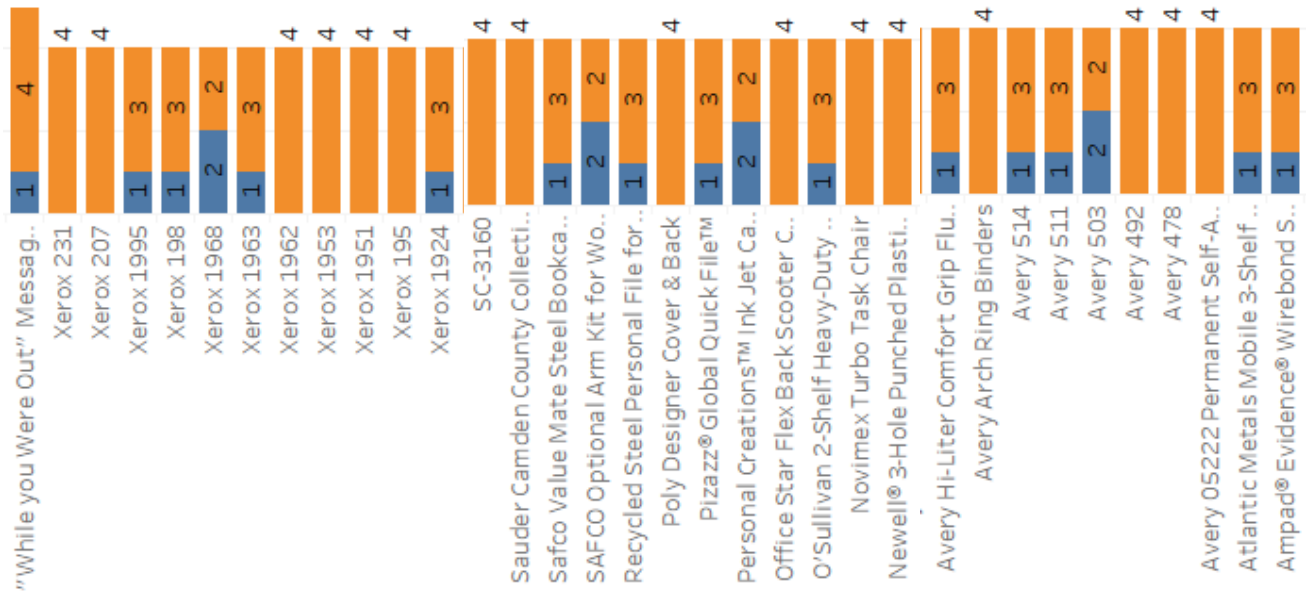
Plymouth Boxed Rubber Bands by Plymouth has total of eight orders but four of them are returned.



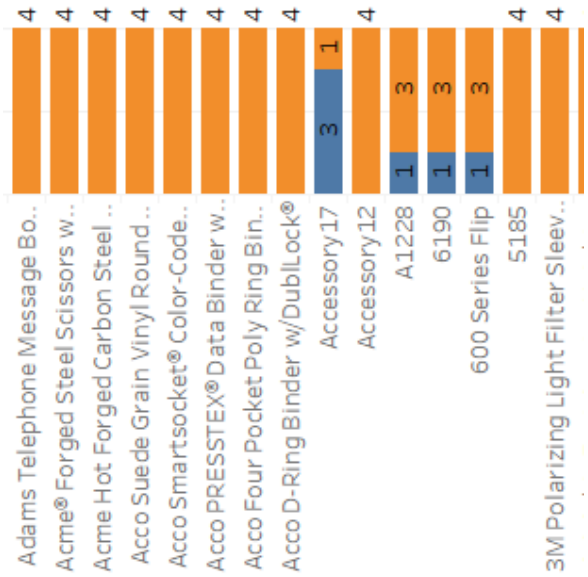
Imation 3.5" DS/HD IBM Formatted Diskettes, 10/Pack and Hon 61000 Series Interactive Training Table total has six orders but half of them are returned.



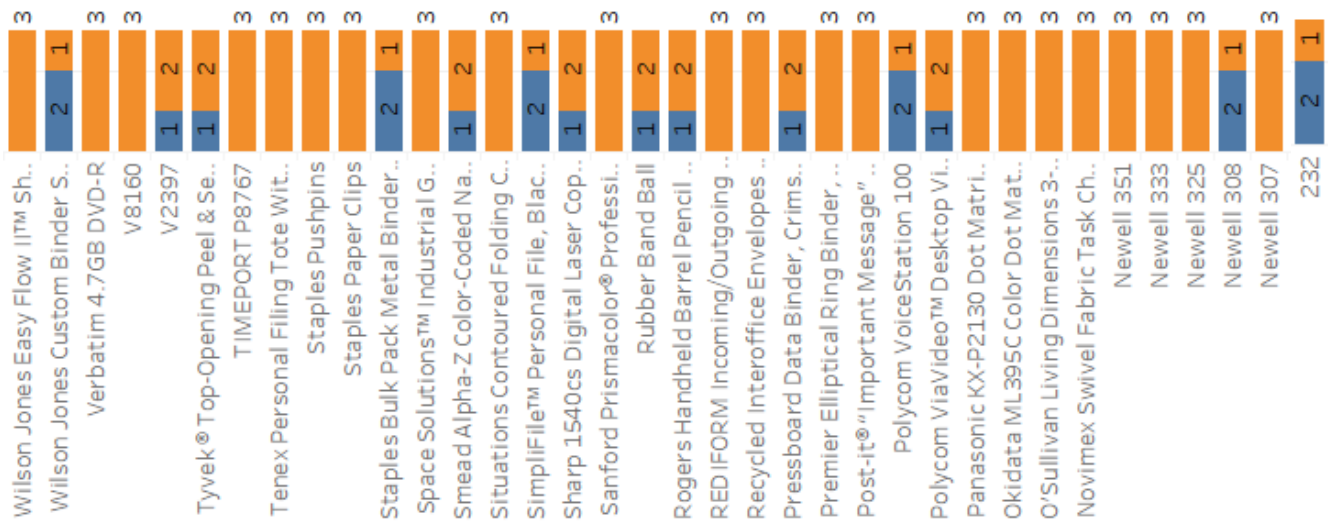
Eldon Cleatmat Plus™ Chair Mats for High Pile Carpet and Sefl-Adhesive Removable Labels each has total five orders but three of them are returned.



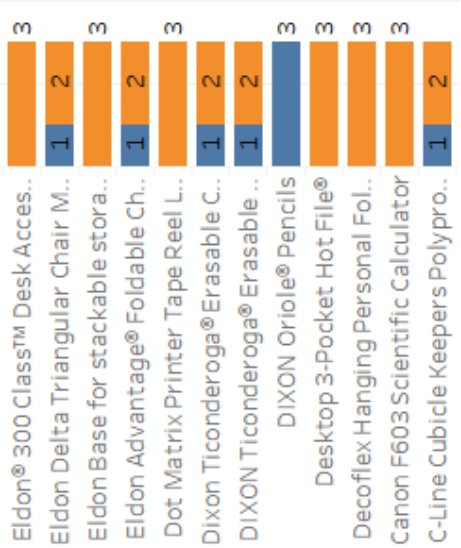
Xerox 1968, SAFCO Optional Arm Kit for Workspace Cribbage Stacking Chair, Personal Creations™ Ink Jet Cards and Labels and Avery 503 each has total four orders but half of them are returned.



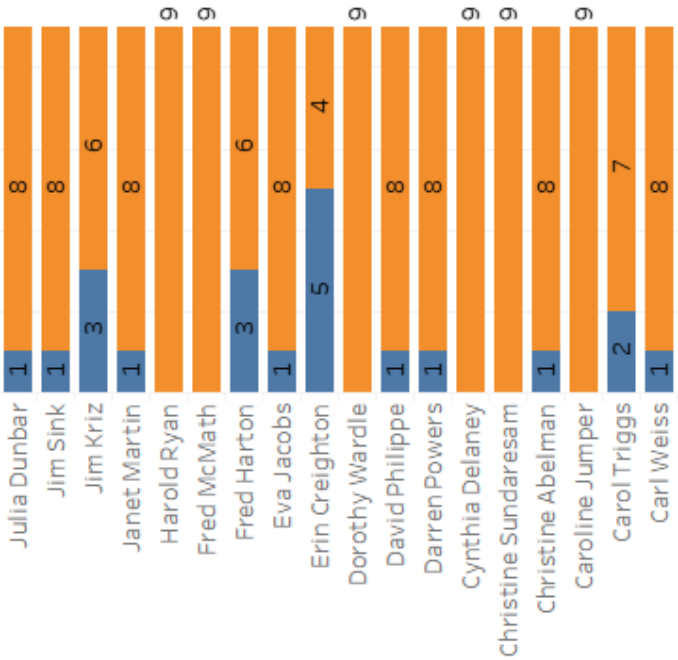
Accessory17 has total of four orders but three of them returned.



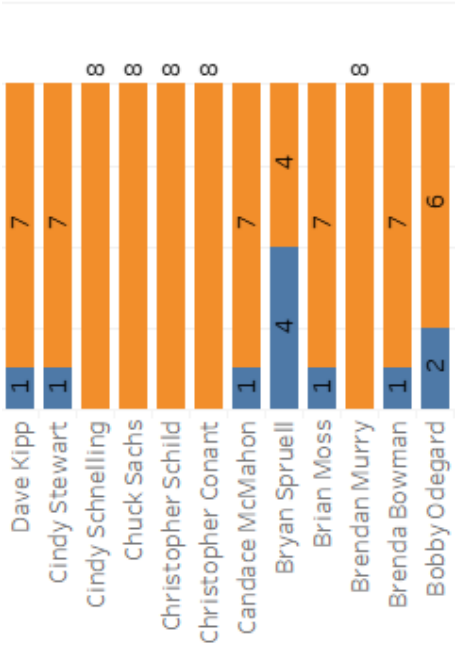
Wilson Jones Custom Binder Spines & Labels, Staples Bulk Pack Metal Binder Clips, SimpliFile™ Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h, Polycom VoiceStation 100, Newell 308 and 232 each has three orders but two of them are returned.



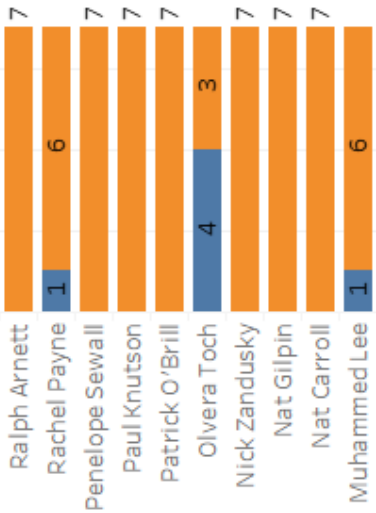
DIXON Oriole Pencils has three orders but all three orders are returned.



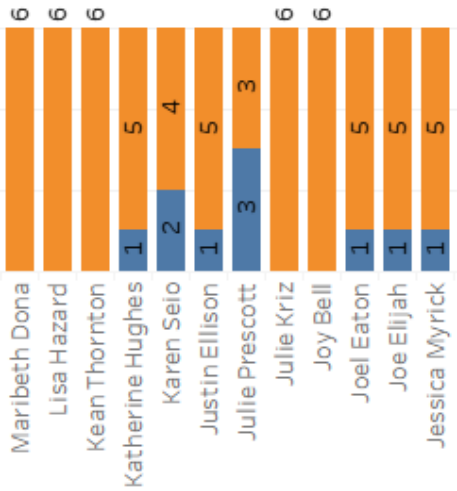
We have also created a customer return bar chart to show the return ratio of a customer's history orders. Customer Erin Creighton has total placed nine orders but returned five.



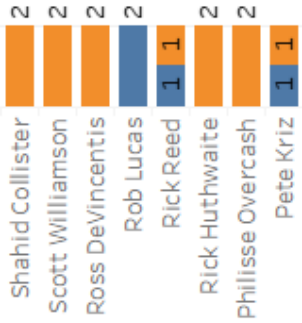
Customer Bryan Spruell has placed total of eight orders but returned four.



Customer Olvera Toch has placed a total seven orders but returned four.



Customer Julie Prescott has placed six orders, but half of them are returned.



Customer Rob Lucas has placed two orders, all of them are returned.

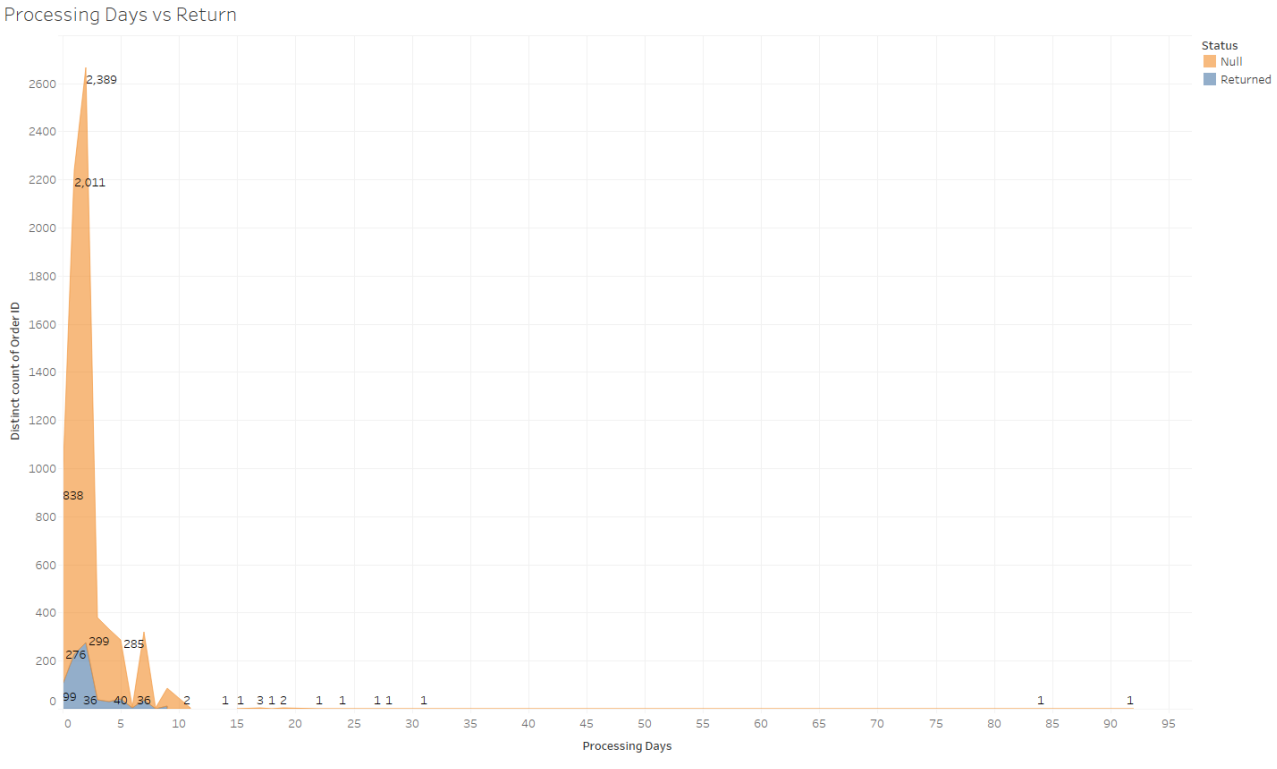


Figure 15 Processing Days and Return

We created a graph to investigate the relationship between return and processing days (The day ship out minus the day received order). Most orders are ship out within two days, the return order amount only related to the total order amount which means their customer is not sensitive to the processing days.

Product association and customer cluster analysis

Association Mining

Firstly, we performed association mining based on orders. However, even if we set the support value to 0.0002 and the confidence value to 0.2, there was still no rules. As a result, we tried to perform association mining based on customer names and saved the results into a csv file.

item1	item2	support	confidence	lift
('Micro Innovations 104 Keyboard',)	('Belkin 105-Key Black Keyboard',)	0.00503778	0.36363636	19.2484848
('Belkin 105-Key Black Keyboard',)	('Micro Innovations 104 Keyboard',)	0.00503778	0.26666667	19.2484848
('Snap-A-Way¬Æ Black Print Carbonless Ruled Speed Letter, Triplicate',)	('Belkin 105-Key Black Keyboard',)	0.00503778	0.23529412	12.454902
('Belkin 105-Key Black Keyboard',)	('Snap-A-Way¬Æ Black Print Carbonless Ruled Speed Letter, Triplicate',)	0.00503778	0.26666667	12.454902
('Advantus Push Pins, Aluminum Head',)	('#10 White Business Envelopes,4 1/8 x 9 1/2',)	0.00377834	0.27272727	16.6573427
('#10 White Business Envelopes,4 1/8 x 9 1/2',)	('Advantus Push Pins, Aluminum Head',)	0.00377834	0.23076923	16.6573427
('Global Leather and Oak Executive Chair, Black',)	('12 Colored Short Pencils',)	0.00377834	0.21428571	13.0879121
('12 Colored Short Pencils',)	('Global Leather and Oak Executive Chair, Black',)	0.00377834	0.23076923	13.0879121
('Hoover WindTunnel,Ñ¢ Plus Canister Vacuum',)	('12-1/2 Diameter Round Wall Clock',)	0.00377834	0.23076923	15.2692308
('12-1/2 Diameter Round Wall Clock',)	('Hoover WindTunnel,Ñ¢ Plus Canister Vacuum',)	0.00377834	0.25	15.2692308
('LX 677',)	('1726 Digital Answering Machine',)	0.00377834	0.33333333	14.7037037
("O'Sullivan Elevations Bookcase, Cherry Finish",)	('1726 Digital Answering Machine',)	0.00377834	0.21428571	9.45238095
('Southworth 25% Cotton Premium Laser Paper and Envelopes',)	('1726 Digital Answering Machine',)	0.00377834	0.23076923	10.1794872
('Tenex Contemporary Contur Chairmats for Low and Medium Pile Carpet, Computer, 39" x 49"',)	('1726 Digital Answering Machine',)	0.00377834	0.25	11.0277778
('GBC Standard Plastic Binding Systems Combs',)	('3M Organizer Strips',)	0.00377834	0.23076923	15.2692308
('3M Organizer Strips',)	('GBC Standard Plastic Binding Systems Combs',)	0.00377834	0.25	15.2692308
('Accessory8',)	('6120',)	0.00377834	0.3	21.6545455
('6120',)	('Accessory8',)	0.00377834	0.27272727	21.6545455
('Durable Pressboard Binders',)	('80 Minute CD-R Spindle, 100/Pack - Staples',)	0.00377834	0.21428571	8.50714286
('Eldon Expressions Mahogany Wood Desk Collection',)	('80 Minute CD-R Spindle, 100/Pack - Staples',)	0.00377834	0.25	9.925
('Xerox 220',)	('80 Minute CD-R Spindle, 100/Pack - Staples',)	0.00377834	0.3	11.91
('8860',)	('Bevis 36 x 72 Conference Tables',)	0.00377834	0.25	9.02272727
('Ibico Covers for Plastic or Wire Binding Elements',)	('8860',)	0.00377834	0.3	19.85
('8860',)	('Ibico Covers for Plastic or Wire Binding Elements',)	0.00377834	0.25	19.85
('GBC Standard Therm-A-Bind Covers',)	('Accessory23',)	0.00377834	0.2	31.76
('Accessory23',)	('GBC Standard Therm-A-Bind Covers',)	0.00377834	0.6	31.76
('Global Adaptabilities,Ñ¢ Conference Tables',)	('Accessory35',)	0.00377834	0.25	14.1785714
('Accessory35',)	('Global Adaptabilities,Ñ¢ Conference Tables',)	0.00377834	0.21428571	14.1785714
('Unpadded Memo Slips',)	('Accessory36',)	0.00377834	0.375	19.85
('Accessory36',)	('Unpadded Memo Slips',)	0.00377834	0.2	19.85
('Xerox 1949',)	('Accessory41',)	0.00377834	0.25	16.5416667

Figure 16 Product Association Rules

There are 5496 orders of 1263 products in total, and only 795 customer names. However, the result still shows that there are only 4 rules with support value more than 0.005 and other 159 rules are all with support value of 0.00377834, which indicate that there are no actually convincing rules in the sales dataset. As we can see that the confidence values are relatively low, around 0.2 to 0.4. Some of the rules are with high lift value, which is because the items in item2 column are actually auxiliary products of items in item1 column. So, we can come to the conclusion that there are no valuable rules between products.

K-Means Cluster Analysis

We notice that all customers are labeled different segmentations as consumer, home office, small business and corporate. By taking 4 factors into consideration, which are the number of orders, total sales, total profit and total product, we performed k-means cluster analysis to examine if the current segmentations are precise and valuable.

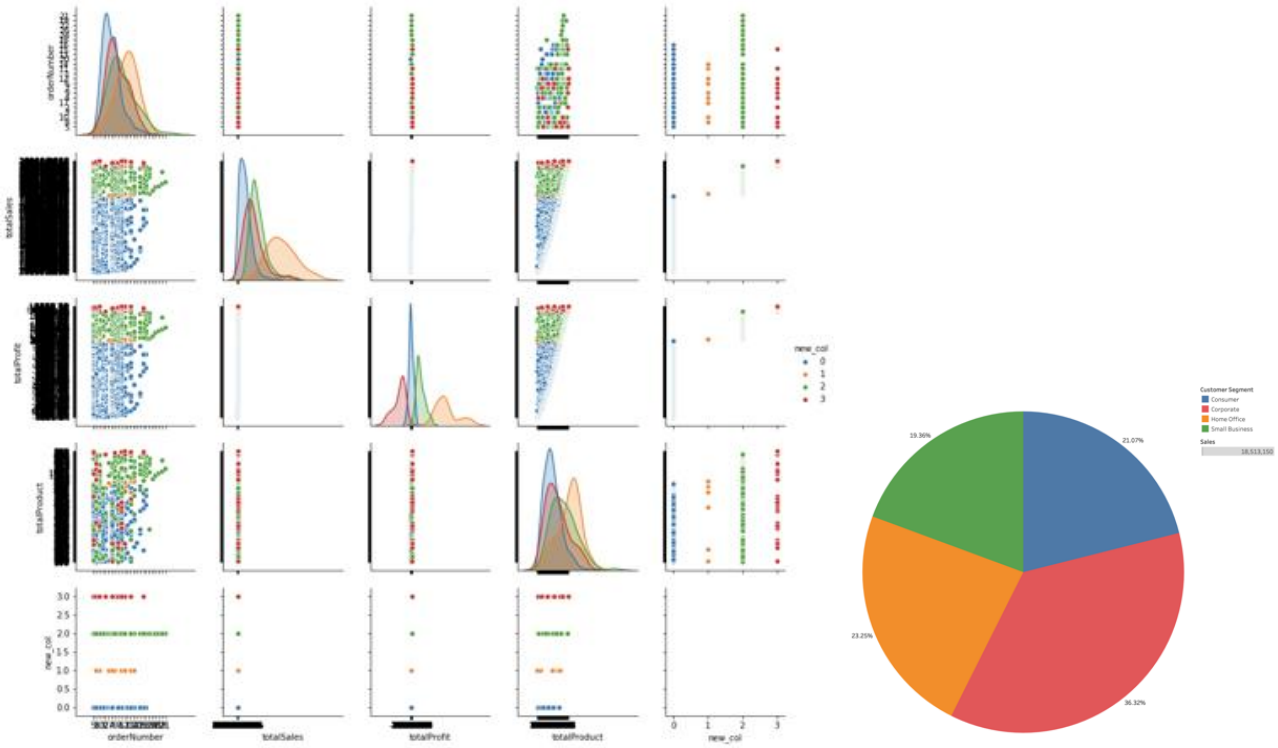


Figure 17 Customer Classification

Firstly, we set the k value of 4 as the same as the current segmentation system. By observation of the result pair plot, the percentage of consumer level customers (blue) are found as the biggest part of customers, more than 50%. And the home office (yellow) as well as the corporate (red) customers are only of a small percentage. The segmentations are considerably different from the current ones. Furthermore, we found that 6 is the best k value to differentiate the customers.

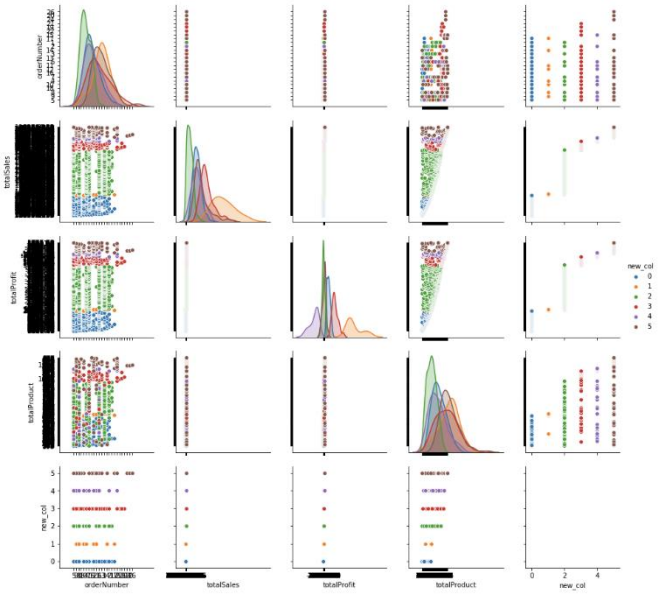


Figure 18 Customer Classification

The consumer level customers can be further differentiated by the number of purchased products. And the business level customers can also be further analysis and segmented into different categories.

Pivot table for shipping cost analysis

Row Labels	Sum of Sales	Average of Product Base Margin	Sum of Profit	Average of Shipping Cost
⊟Furniture	5178591	0.60	117432.99	30.88
⊞Delivery Truck	4154946	0.663349282	-8175.450119	48.88311033
⊞Express Air	159759	0.541239669	20659.51177	13.66872
⊞Regular Air	863886	0.533406897	104948.9246	13.23519411
⊟Office Supplies	3752762	0.46	518021.46	7.83
⊞Delivery Truck	338966	0.593866667	4270.32	53.9376
⊞Express Air	476419	0.463690476	65515.72	7.291043771
⊞Regular Air	2937377	0.458374936	448235.418	7.033572697
⊟Technology	5984248	0.56	886313.52	8.95
⊞Delivery Truck	1730966	0.468447489	273549.9674	28.66762557
⊞Express Air	548241	0.577310606	61255.0559	6.861515152
⊞Regular Air	3705041	0.564962073	551508.492	6.57534134
⊞(blank)				
Grand Total	14915601	0.513	1521767.960	12.839

Figure 19 Shipping Cost Pivot Table

Using data from provided Sales-Canada dataset, we found out the Average of Shipping Cost (AoSC) had eaten a lot of SoP, especially in “Furniture” category.
It’s interesting to notice that all three main categories having lower air shipping cost (both Express Air and Regular Air) than by Delivery Truck which was counter intuitive. By the nature of business, “Furniture” mainly used Delivery Truck which drove AoSC much higher than the other two categories. Since we are not accessing all other operation information, it would be a future focal point to investigate that if it’s a result of inefficiently used of delivery truck. It would be crucial to the company to cut down operational cost like these.

Conclusions

Based on data visualization we can conclude the following recommendations.
The company should reduce the inventory for office furniture since the furniture is normally large which required higher cost for transportation and storage but at the same time furniture are fast to be out of style. As a result, furniture is most likely to lose profit when the company decides to clearance. The company should stay focus on technology product and office supplies market, slowly reduce the inventory share of furniture. Some products the company should be paying attention to. “Okidata Pacemark 4410N Wide Format Dot Matrix Printer” and “Polycom ViewStation™ ISDN Videoconferencing Unit” which is an out of date inventory, rare business continue dot matrix printer also nobody uses ISDN video conferencing system in today. As the depreciation will continue the company should get rid of that inventory as soon as possible by clearance and sale and ensure do not bring more inventory of those products. The province sales map give a recommendation to focus on the most important market in BC, AB and ON in case to maintain the profit performance. Corporate is the most important customer segment the company should pay attention to, but Home Office, Consumer and Small Business they are mostly equal which is also important to the company. We have the Customer Order & Sales chart, we suggest the company to filter the top sales customer and provide target promotion marketing plan to boost the sales. The return ratio varies for each

products, we suggest the company look into flagged product with high return rate to see if there is any quality issue with the product. For the return rate with the customer, currently the order amount still low for most customer, we recommend the company to set up criteria on return rate for the customer, if a customer's return rate exceeds the criteria the company should look into to see why the customer is not satisfied with their order. Since most customers most are business customer and not sensitive to the processing speed, we suggest the company to cut some cost for processing in order to reduce the operational cost.

The company is currently selling products in a large number of types. The cost of inventory and shipping may be considerably high and the promotion strategy of those products will be hard to implement. The company can be more focus on specific categories or products, dropping the product lines which few customers purchase with little profit.

The company also need to further segment the customers and check with individual customers to provide more precise products and services. The current segmentation system isn't generating any valuable information.

The company should try to implement the recommendations and KPI as soon as possible and monitor the new progress. The company should thrive to become a data driven company. That is to use data to:

- Clearance old inventories of losing money products;
- Bundle related products to increase sales;
- Increase brand and star product awareness;
- Implement customer loyalty program and gather useful data for promotion purposes.

Periodical data analysis will make sure the company runs on the right track.

Reference

1. Akash, R. (2018) *How to perform different operations using Tableau (Visualizations, OLAP etc.)*. Retrieved on 8th Aug, 2019
from: https://www.youtube.com/watch?v=8iYpxiipNHQ

Appendix

Association Mining

```
from efficient_apriori import apriori
import pandas as pd
from graphviz import Digraph
import csv

data = []

#with open('FuTeaShop_Sale.csv','r') as csvFile:
#    reader = csv.reader(csvFile)
#    for row in reader:
#        data.append([row[0],row[3]])

#with open('Shop_Order.csv','r') as csvFile:
#    reader = csv.reader(csvFile)
#    for row in reader:
#        data.append([row[1],row[5]])

#with open('Sales_Canada1.csv','r') as csvFile:
#    reader = csv.reader(csvFile)
#    for row in reader:
#        data.append([row[1],row[17]])

with open('Sales_Canada2.csv','r') as csvFile:
    reader = csv.reader(csvFile)
    for row in reader:
        data.append([row[11],row[17]])

#data = data[1:]
transaction = []
tmp = 0
tmpbsk = []

for row in data:
    if row[0] != tmp:
        transaction.append(tmpbsk)
        tmp = row[0]
        tmpbsk = []
        tmpbsk.append(row[1])
    else:
```

```
        tmpbsk.append(row[1])

count = 0


transactions = transaction[1:]

for row in transactions:
    if len(row) > 1:
        count = count +1

print(count)

#itemsets, rules = apriori(transactions, min_support=0.0005,  min_confidence=0.2)
itemsets, rules = apriori(transactions, min_support=0.0035,  min_confidence=0.2)
print(rules)
#print(itemsets)

rule_dt = []
for row in rules:
#    print(row.lhs,row.rhs,row.support,row.confidence,row.lift)
    rule_dt.append([row.lhs,row.rhs,row.support,row.confidence,row.lift])

df = pd.DataFrame(rule_dt, columns = ['item1','item2','support','confidence','lift'])
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
print(rule_dt)
print(df)
df.to_csv('rules.csv')
```

K-Means Cluster

```
import numpy as np
import csv
from sklearn.cluster import KMeans
import pandas as pd
import seaborn as sns

data = []
sum0 = 0
sum1 = 0
sum2 = 0
sum3 = 0
```

```
with open('customers.csv','r') as csvFile:
    reader = csv.reader(csvFile)
    for row in reader:
        data.append([row[0],row[1],row[2],row[3]])
        sum0 = sum0+float(row[0])
        sum1 = sum1+float(row[1])
        sum2 = sum2+float(row[2])
        sum3 = sum3+float(row[3])

avg0 = sum0/795
avg1 = sum1/795
avg2 = sum2/795
avg3 = sum3/795

c_names = ['OrderNumber','totalSales','totalProfit','totalProduct']
cluster = pd.DataFrame(data)
cluster[0] = cluster[0].map(lambda x: float(x)/avg0)
cluster[1] = cluster[1].map(lambda x: float(x)/avg1)
cluster[2] = cluster[2].map(lambda x: float(x)/avg2)
cluster[3] = cluster[3].map(lambda x: float(x)/avg3)

num_clusters = 6
km_cluster = KMeans(n_clusters=num_clusters, max_iter=300, n_init=40, \
                    init='k-means++',n_jobs=-1)

result = km_cluster.fit_predict(cluster)
print(result)

cluster_result = pd.DataFrame(data, columns=
['orderNumber','totalSales','totalProfit','totalProduct'])
cluster_result['new_col'] = result

sns.pairplot(cluster_result, hue = "new_col")
```

Table of Figures

Figure 1 Cateogry SoS/AoPBM/SoP 4
Figure 2 Business Model SWOT 4
Figure 3 Monthly Profit Break Down 5
Figure 4 Category Profit Break Down 6
Figure 5 Category Profit Break Down 7
Figure 6 Category Profit Break Down 8
Figure 7 Sub-Category Profit Break Down 9
Figure 8 Product Profit Break Down 9
Figure 9 Product Profit Break Down 10
Figure 10 Province Sales Map 10
Figure 11 Customer Segment Shares 11
Figure 12 Customer Order & Sales 11
Figure 13 Category Return 12
Figure 14 Sub-Category Return 13
Figure 15 Processing Days and Return 20
Figure 16 Product Association Rules 21
Figure 17 Customer Classification 22
Figure 18 Customer Classification 22
Figure 19 Shipping Cost Pivot Table 23