**The Hang Seng University of Hong Kong**

2022-2023 Semester 1

**MSIM 4101:**

Decision Analytics

**Report Title:**

Customer Segmentation Using Transaction Data

Class: L02

Group: Group 5

Group Members:

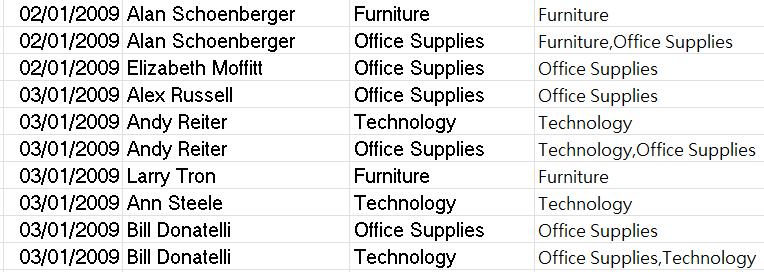
| s198189 | KWONG Cheuk Hang |
| --- | --- |
| s198210 | NG Suen Qiu Atalie |
| s197671 | LAU Hei Yee |
| s198193 | LAM Yuet Ying |
| s198202 | LI Siu Fong |

1. **Introduction**

AGA is a furniture and electrical appliances online store that trades with merchants and consumers. It was established in 2009 and headquartered in Canada. We are planning a promotion for AGA to understand customers' needs and consumption patterns to help AGA increase its sales and profits. We will analyze three different aspects: transaction, customer, and profit.

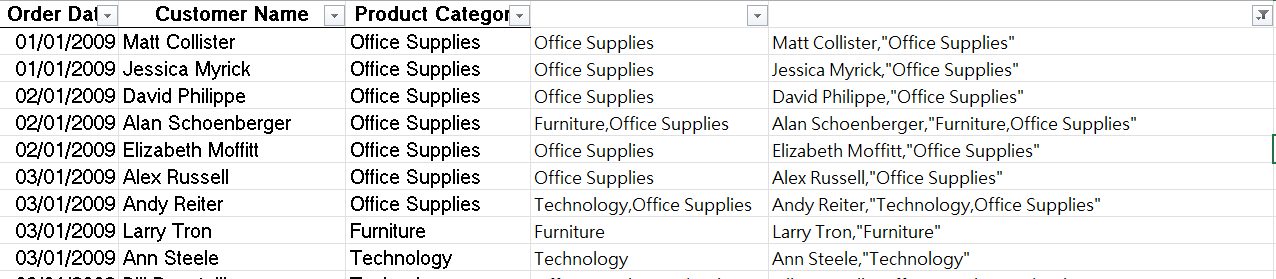
Analyzing the company's product categories to understand the frequent items and sell bundles with infrequent to bring up the sales. Also, analyzing customers to segment them and using a focus strategy to retain customers. Carry out an assumption analysis based on AGA's past profits, and recommend the company use the strategy of small profit but quick turnover to increase sales.

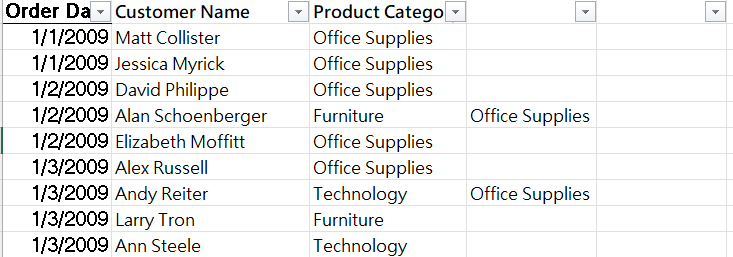
1. **Data Process**
2. Select Customer Name, Order Date and Product Category column into a new sheet.
3. Advanced filter function to create unique record in three columns(but there might **same customer** buy different product category at the **same day**)
4. Enter formula “=IF(Y2<>Y1,Z2,AA1 & "," &Z2)”in the right column , which can show like to below as a example



1. Enter formule “=IF(Y2<>Y1,CONCATENATE(Y2,",""",AA2,""""),"")” in the right column and present like the below example



1. Add filter in the column and select all except (“blank”) in the fifth column
2. Using the Text to Column function that separates the text by delimiter the comma. And the data will become the below.



1. **Transaction Analysis**

According to graphic 1.c, there are similar profits every year, especially in 2009 which had the highest profit which is $223562.57. The trend is that there is a sharp decrease between 2009 and 2010 and then slowly increased in 2012 as you can refer to graphic 1.d.

To make effective promotion, we will focus on the popular item.

Look at the graphics 1.e and 1.f, Technology is the total highest profit by product category every year. For the sub-category, the office machine made the highest sales which is $122455.35. So, we may promote these more popular products. Facilitating the company's most popular products as hot products to attract the audience to recognize our company. It may increase AGA's popularity and attract new customers.

Next, graphic 1.g shows that the West had the highest order quantity which is 60018 orders. If AGA wants to increase the sales of fewer order regions, we recommend the company can have more promotions or do surveys to focus on their needs. For example, graphic 1.h displays the popular products of sub-categories in Nunavut. AGA can promote the most popular products such as computer peripherals and labels, or give more discounts on them. It can balance the order quantity within different regions and increase the company's profits.

After plotting the scat chart from graphic 1.i and viewing data from graphic 1.j in detail, the total number of order quantities is proportional to profit. That means increasing the order quantity can increase profits. We recommend the company organize more discount events. Such as buy 1 get 1 free or buy more for more discount to increase order quantity.

We also plot the bar chart shown in graphic 1.k to find out Corporate has the highest order quantity in the customer segment. Take the West region as an example. Corporate takes the largest share with 34.2% shown in the graphic 1.l and Technology made the highest profit in 4 years in the West which is shown in graphic 1.m. To bring more profit, AGA can enter more attractive products to the region that customers are interested in.

We have been using the Apriori Algorithm for finding the association rules. After we processed the data, we found out there were 5505 unique transactions over 4 years. We would like to apply the association rule in the transaction data to help to show the probability of the relationship and discover sales correlations in transactional data.

First, we will calculate the support of different itemsets as support is an indicator of how frequently the itemset appears in the itemset. We have set the minimum support as 10%. According to Table 1.a, the calculation is shown in the following:

Let O, F and T become Office Supply, Furniture, and Technology.

1. Support(F) = P(F) = 1559/5505 = 0.2832 >10%
2. Support(T) = P(T) = 1859/5505 = 0.3377 >10%
3. Support(O) = P(O) = 3625/5505 = 0.6585 <10%

We can see that the support of Furniture, Technology, and Office Supply is higher than the minimum support. This indicates that three of them are the frequent 1-itemset.

Next, we will calculate the 2-itemsets which are shown below:

Let O, F and T become Office Supply, Furniture, and Technology.

1. Support(OT) = P(OT) = 737/5505 = 0.1339 >10%
2. Support(OF) = P(OF) = 605/5505 = 0.1099 >10%
3. Support(FT) = P(FT) = 316/5505 = 0.0574 <10%

We found that the support of Office Supply with Technology and support of Furniture with Office Supply are also higher than the minimum support, which shows that those are the frequent 2 itemsets in the data.

Furthermore, we have calculated the confidence to confirm how often the rule has been found to be true. We have set the minimum confidence level to 30%. Based on the previous calculation, we know their support, and the confidence level calculation has shown below:

1. Confidence (O→T) = P(T|O) = 0.1339/0.3377 = 0.2033
2. Confidence (T→O) = P(O|T) = 0.1339/0.6585 = 0.3964 >30%
3. Confidence (O→F) = P(F|O) = 0.1099/0.2832 = 0.1669
4. Confidence (F→O) = P(O|F) = 0.1099/0.6585 = 0.3881 >30%

The confidence of people who buy Technology will also buy Office Supply is higher than the minimum confidence level. It can be sure that AGA customers who bought Technology will also buy Office supplies altogether.

On the other hand, the confidence of people who bought Furniture will also buy Office Supplies also higher than the minimum confidence level. We can say that customers who bought Furniture will also buy Office Supplies altogether.

Therefore, we have generated two rules that satisfy the minimum support and the minimum confidence at the same time, which are:

1. Customers who bought Technology will also buy Office Supplies. (T→O)
2. Customers who bought Furniture will also buy Office Supplies. (F→O)

1. **Customer Analysis**

The customer behavior of AGA will be examined in this section. We would like to learn about customer purchasing habits. By determining the target group, a target marketing strategy may be developed.

Firstly, to determine loyal customers and vulnerable customers. Looking at Graph 2.a, the scatter plot shows the customer segment measure with a count of product categories and the total profit. There are four customer segments including Consumer, Corporate, Home Office, and Small Business. In this chart, Corporate owns the largest variety of products and the highest total profit. It performs extremely well compared to other customer segments. So, Corporate is the loyal customer segment. While Small Businesses are the vulnerable customer segment. Although the plotting of small business and consumer is very closed horizontally, consumers made a higher profit than the Small Business segment. Besides, referring to Graph 2.b, the bar chart illustrates the total order quantity by customer segment, it also shows the same result. Corporate achieve the highest order quantity, $78.52k. Home Office is in second place with a $52.81k profit. Consumers are the second last, which made $41.76k profit to AGA. Small businesses own the lowest, $41.69k. Therefore, **Corporate is the loyal customer segment, and Small Businesses are the vulnerable customer segment**.

Second, to determine the ranking of regional customers' purchasing power. According to Graph 2.c, A variety of product categories and order quantities brought by different regions is shown. Customers in the West have the largest purchase power compared to other regions.

The ranking is as follows: Ontario, Prairie, Atlantic, Quebec, Yukon, and Northwest Territories.AND Nunavut placed last. Again, refer to the bar chart in Graph 2.d shows the profit of each region's customer segment. Corporations generate the largest profit in each location. West customers typically have the highest purchasing power. Customers from the West consistently earn the largest profit among the four customer segments. Customers from Yukon, the Northwest Territories, and Nunavut generate less profit for AGA when compared to other regions. So, the **Customer in the West has the highest purchasing power.**

Besides, to find out the customer who buys the many order quantities, and if the customer places the most orders, the company will make the greatest money from that customer. Refer to Graphic 2.e, a bar chart showing the top-5 Best Customers of AGA, ranked by order quantity. When generating the bar chart, we found that Durren Budd is placed first, with 41 orders. But we found there is a repeat name customer, Durren Budd, who is from Yukon and placed 6 orders. After removing the 6 orders from Darren Budd, he comes to the second. Ed Braxton who placed 38 orders came in first place. And Brad Thomas, Carlos Soltero, and Patrick Jones are the remaining Top 3, 4, 5 Best customers. The average order quantity of the Top-5 Best Customers is 34.2 orders. Afterward, we use excel to Qlik to generate the customer's total profit ranking bar chart in Graphic 2.f. It found the first place, and Emily Phan has contributed $7.57k profit. Ed Braxton, who placed the highest order, placed at the 9th, has contributed $4.15k profit. Darren Budd placed the second high order and placed the 13th. He has contributed $3.99k profit. Therefore, it is found that **customers who buy many order quantities do not mean to contribute the highest profit.**

As our project will focus on the customers' order quantity. We would like to find out the purchase habits of the Top-3 Best Customers: Ed Braxton, Darren Budd, and Brad Thomas. The product categories bought from Darren Budd are displayed in Graph 2.g. Darren Budd had mainly purchased furniture and technology. The percentage of product categories is presented in Graphic 2.h. The portion between Technology and Furniture is 8:2. The three 1-itemsets are found in his transaction. He has bought furniture, technology, and office supplies independently. For Ed Braxton, his purchase record and the percentage of product categories are shown in Graphic 2.i and Graphic 2.j. Ed Braxton has bought three product categories. Office supply purchases account for about 50% of all purchases, followed by technology at around 35% and furniture at about 20%. In his purchase record, we found three frequent 1-itemsets and two frequent 2-itemsets. He has made individual purchases of furniture, technology, and office supplies independently. Also, he made two frequent 2-itemset purchases: office supplies with technology and office supplies with furniture. For Brad Thomas, his purchase record and percentage of product categories have been displayed in Graphic 2.k and Graphic 2.l. He has purchased Office Supplies, Technology, and Furniture. The percentage of Office Supplies, Technology, and Furniture is 47.4%, 31.6%, and 21.1%. Same as Ed, Brad has purchased three frequent 1-itemsets and two frequent 2-itemsets too. Therefore, the **frequent itemset that we have found is the transaction analysis part adopted in the customer’s purchase record.**

In conclusion, we found that Corporate is the loyal customer segment. Small Businesses are the vulnerable customer segment. But actually other than Corporate, the three segments, i.e. Consumer, Home Office, and Small Business do not perform well.More marketing schema should be placed in the three segments, in order to win more customers and profit in these three fields. Customers in the West usually have the highest purchase power. More promotions should be done in regions that have low sales. We found that customers who buy many order quantities do not mean to contribute the highest profit. Also, the purchase habits of customers, have been determined in the last part, transaction analysis. The frequent itemset is found in the customer's purchase record. It has been verified that the frequent itemsets are constantly found in the customer transaction records.

1. **Profit Analysis**

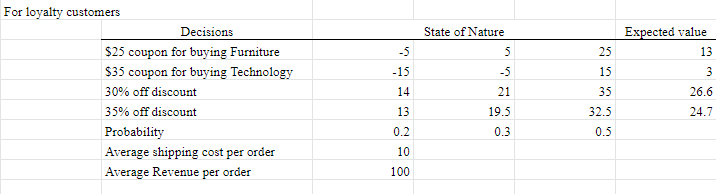
In this section we will analyze the profit of the AGA. We would like to learn about The sale of our goods. Then find the trend of each product and provide promotion strategies to some of our customers.

From our analysis, we find that the quantity of Office Supplies has a significantly greater profit among the three categories of Furniture, Office Supplies, and Technology from 2009 to 2012. A sum of 44.32k of Furniture, 118.28k of Office Supplies, and 52.17k of Technology is sold and a profit of $268.15k, $202.35k, and $343.39k are made respectively during the four years.

In order quantity of sub-category, besides Others, Paper has 14.6% of orders, Binders and Binder Accessories has 10.9% order and Telephone and communication has 10.5%. For the Sub-category in Office Supplies, Paper has the best sales volumes while the Scissors, rules, and trimmers are the worst with 30.87k and 3.93k respectively. For Technology, The best seller is Telephones and communication devices with 22.97k, and copies and fax are the worst with only 2.13k. Last but not least is Furniture, the number one seller is Office Furnishing with 20.65k and the least popular is bookcases with the amount of 4.89k.

We found that the relationship between quantity and profit made is not always positive. This may be due to the fact that the price of each unit of product varies greatly. For example, the average price of technology is much higher than the average price of office supplies. However, the shipping cost is mostly based on weight and dimension. We tend to give promotion discounts based on the quantity of the order.

In promotion strategies, we set up a few alternatives for both the group of loyal customers and vulnerable customers.



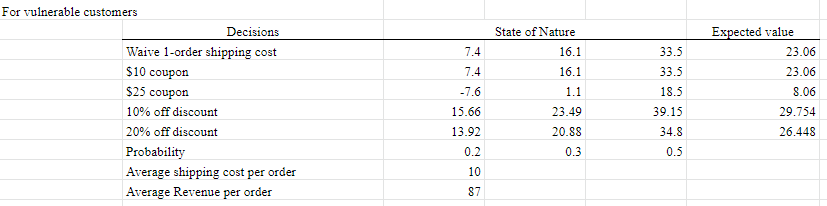
For loyal customers. The alternative we set up is as follows:

Alternative 1: $25 coupon for buying Furniture

Alternative 2: $35 coupon for buying Technology

Alternative 3: 30% off discount

Alternative 4: 35% off discount



For vulnerable customers. The alternative we set up is as follows:

Alternative 1: Waive 1-order shipping cost

Alternative 2: $10 coupon

Alternative 3: $25 coupon

Alternative 4: 10% off discount

Alternative 5: 20% off discount

For the loyalty customers we would like to provide an exclusive discount offer to ensure their contribution to our company. We would like to waive the shipping cost of the order when the bulk buying quantity is over or equal to 30. As we find that the average order quantity of loyalty customers is 26, so they will only need to buy 4 more products per order to enjoy the discount. Therefore, we expect the probability of them accepting our offer will be close to 100%. And as the average profit and shipping cost of loyalty customers is $100 and $10 respectively. We expected the change of profit per order will be increased to [(30-26)\*$100-$10]\*100% = $390. Even if the customers have a lower chance (85%) of accepting the offer, the expected change of profit per order will still increase to [(30-26)\*$100-$10]\*85% = $331.5.

1. **Recommendation**

After analyzing the data of the company, we suggest the company set up bundles of products from different product categories. It lets customers buy the bundling Sales package at a lower price for a set of goods and services from our company. For example, bundle up Office Machine from Technologies together with Paper from Office Supplies. In a normal scenario, customers who buy Office machines may not buy paper as they may have stock in their inventory. However, when the office machine and paper are bundled together, the customer will tend to buy paper as well. As Paper is necessary for office machines, they would be willing to stock more in their inventory.

Besides, we recommend that AGA companies can combine the popular and unpopular products together. According to Graphic 3.f, we found out that Paper is the most popular product, Binders and Binder Accessories is the second and Pen and Art Supplies are the third. We think AGA can combine with the products that complement. For example, Paper complements Scissors Rulers and Trimmer which people will always use together. It can help AGA to reduce the inventory of unpopular products. In addition, they can use a relatively lower price than each of the unit prices to attract customers purchasing more products than customer expected, so that increases AGA total profit.

Besides, AGA could give discounts to different kinds of customers. Based on the alternatives mentioned in the profit analysis, 30% off discounts could be offered to its loyalty customers. And, for the vulnerable customers, AGA could offer 10% off discounts for each order of the vulnerable customers. From this, AGA would still create $26.6 and $29.754 in the expected value from the loyalty and vulnerable customers respectively, and increase the customers’ willingness to continue purchasing the products of AGA.

Except for the two main customer segments: loyalty and vulnerable customers, we found that the best-sales region is West, and the ship mode of customers from the West, mostly chose the Regular Air. Therefore, we also recommend that the company could partner with cargo airlines or join some plans of them, in order to waive on the shipping cost and can still gain profits.

We recommend opening showrooms in various areas. We discovered from customer analysis that the West always has the highest sales. Since AGA is an internet retailer, less people are aware of it and its offerings. Creating a showroom could improve the reputation of the business. AGA can raise awareness by placing their showrooms in public places like malls and exhibitions. The public can view the office chairs, storage, and other items on display. It is anticipated that displaying the products to the public will boost sales across Canada.

Last but not least, we suggest AGA offer a special promotion package for primary school or high school. Referring to Graphic 3.c, we found that a part of the orders came from Paper, Binders and Binder Accessories, Pens and Art supplies, which are stationary. AGA can offer a promotion package for schools. When schools purchase a large amount of paper, binders, pens and art supplies, AGA can give some discounts for schools, for example, to waive shipping costs or to give out some more products to schools. This package allows schools to freely match the quantity and stationery products. Also, it is a good way to promote the brand to schools. It can also raise the order quantity at one time.

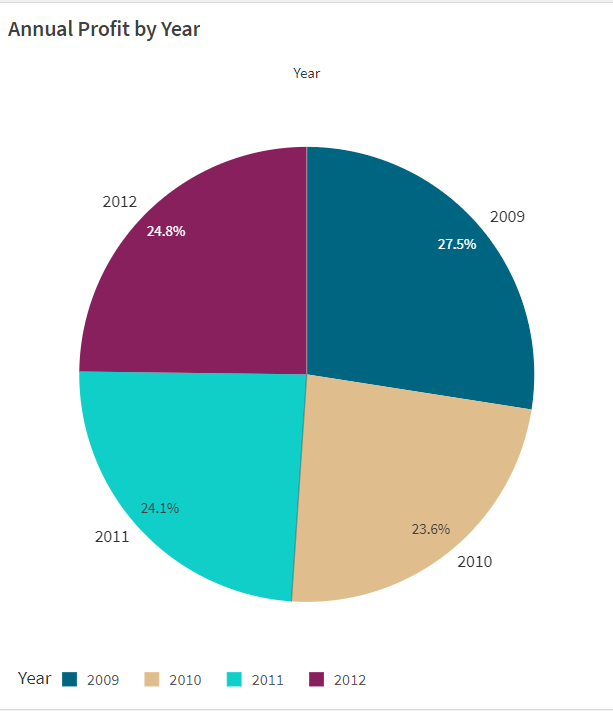
1. **Appendix**

| Item | Support counts |
| --- | --- |
| Furniture | 1559 |
| Technology | 1859 |
| Office Supplies | 3625 |

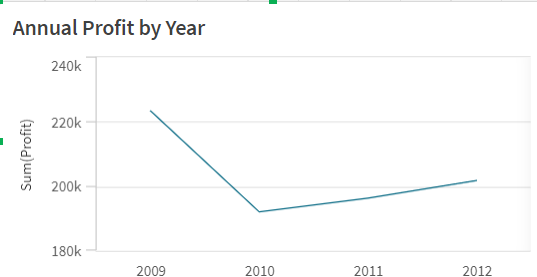
Table 1.a

| Item | Support counts |
| --- | --- |
| OT | 737 |
| OF | 605 |
| FT | 316 |

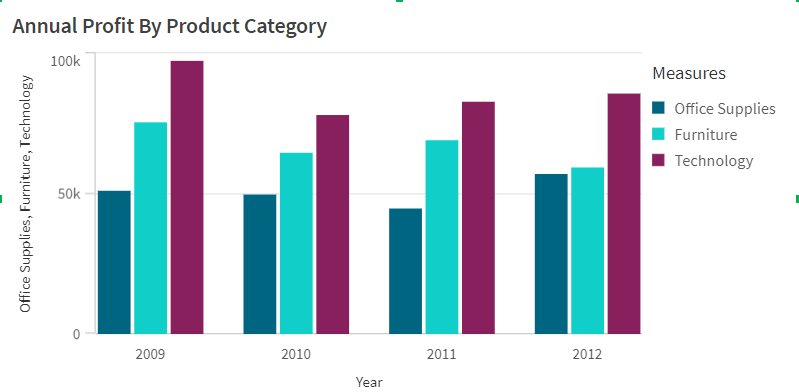
Table 1.b



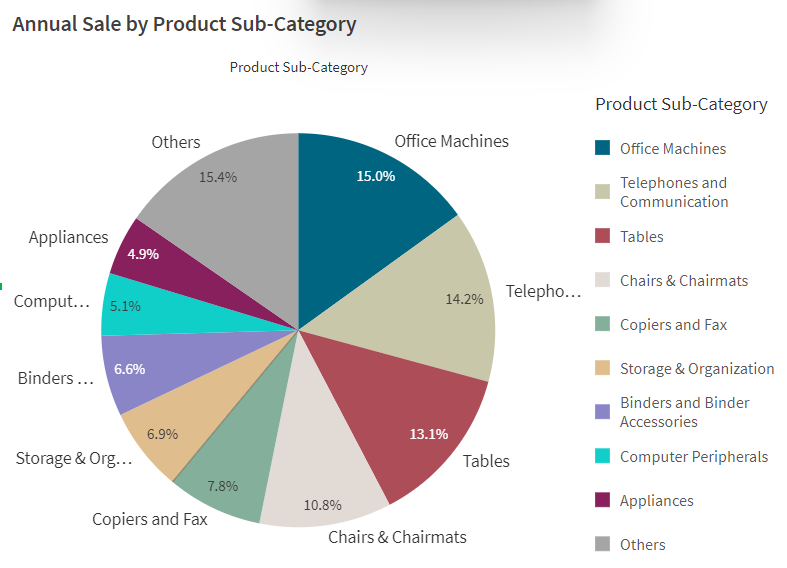
Graphic 1.c



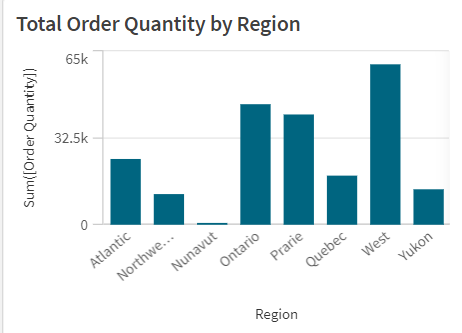
Graphic 1.d



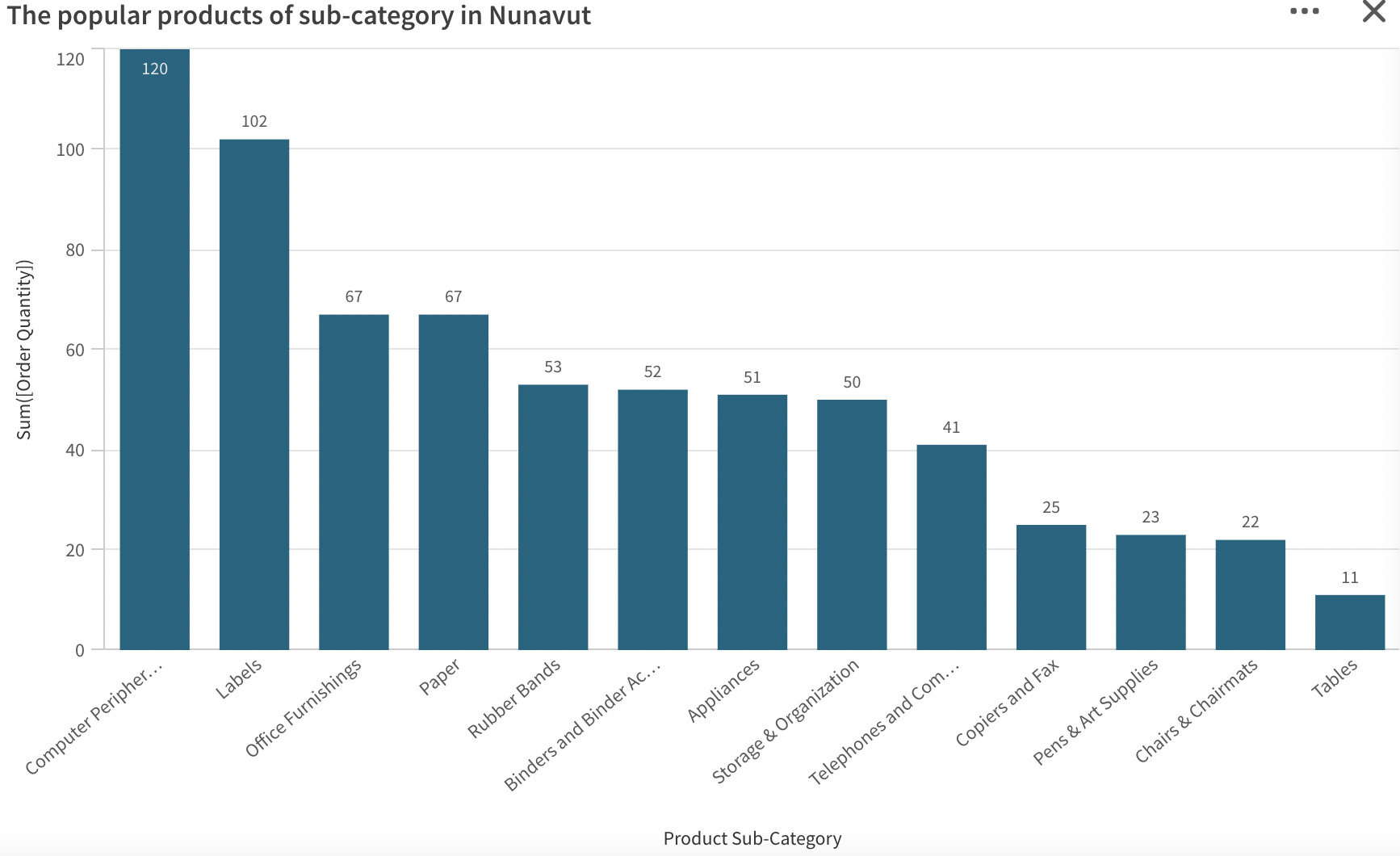
Graphic 1.e



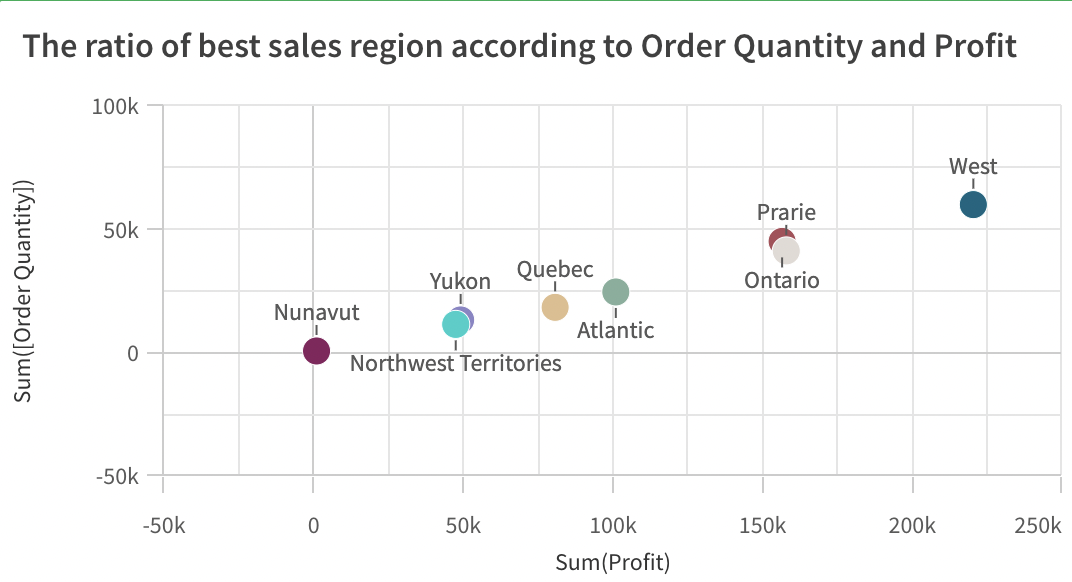
Graphic 1.f



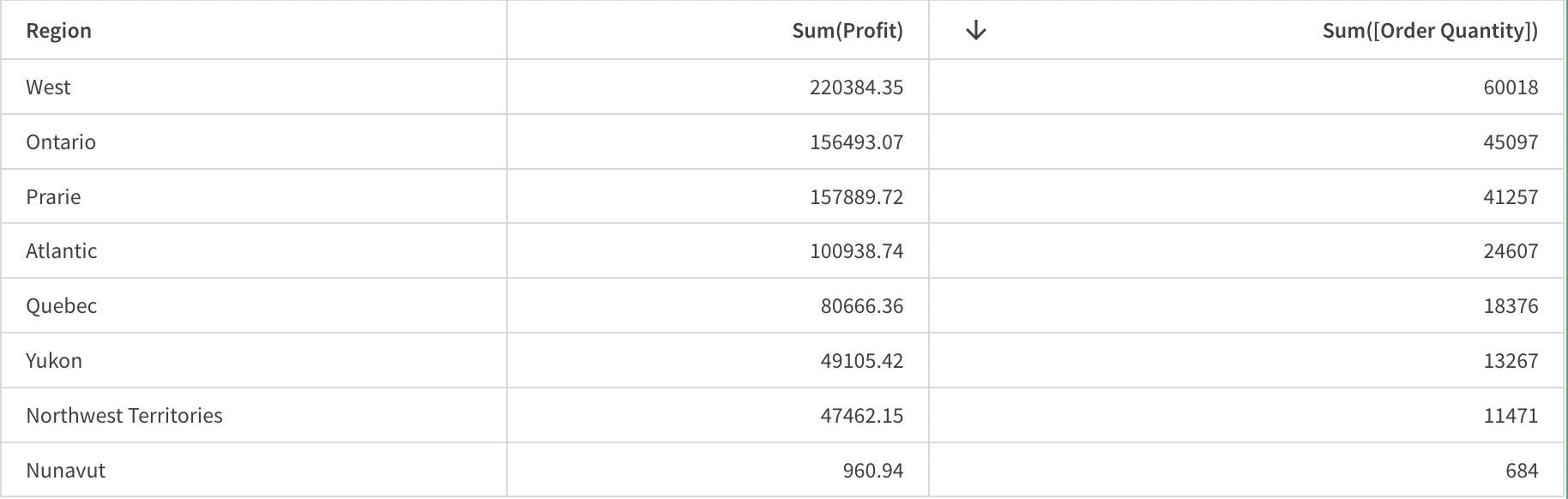
Graphic 1.g



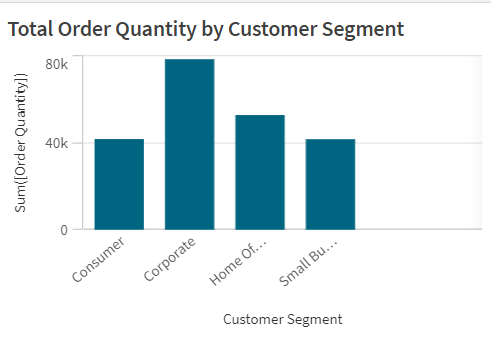
Graphic 1.h



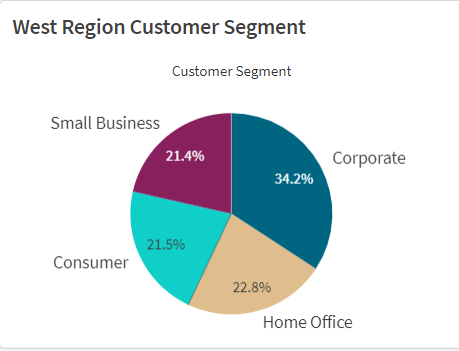
Graphic 1.i



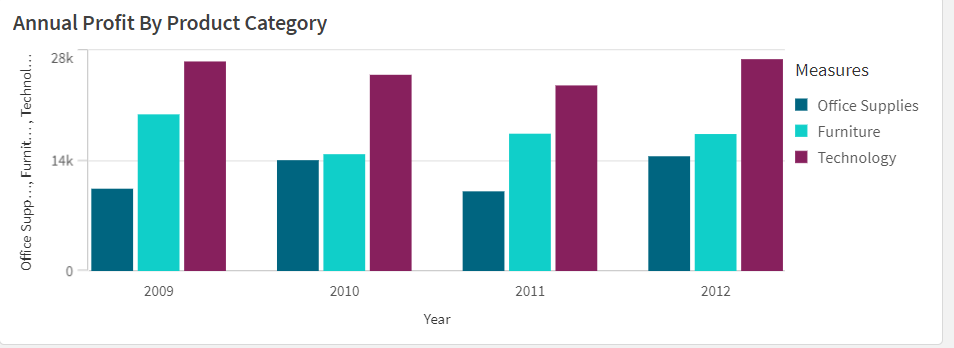
Graphic 1.j Data of graphic 1.e



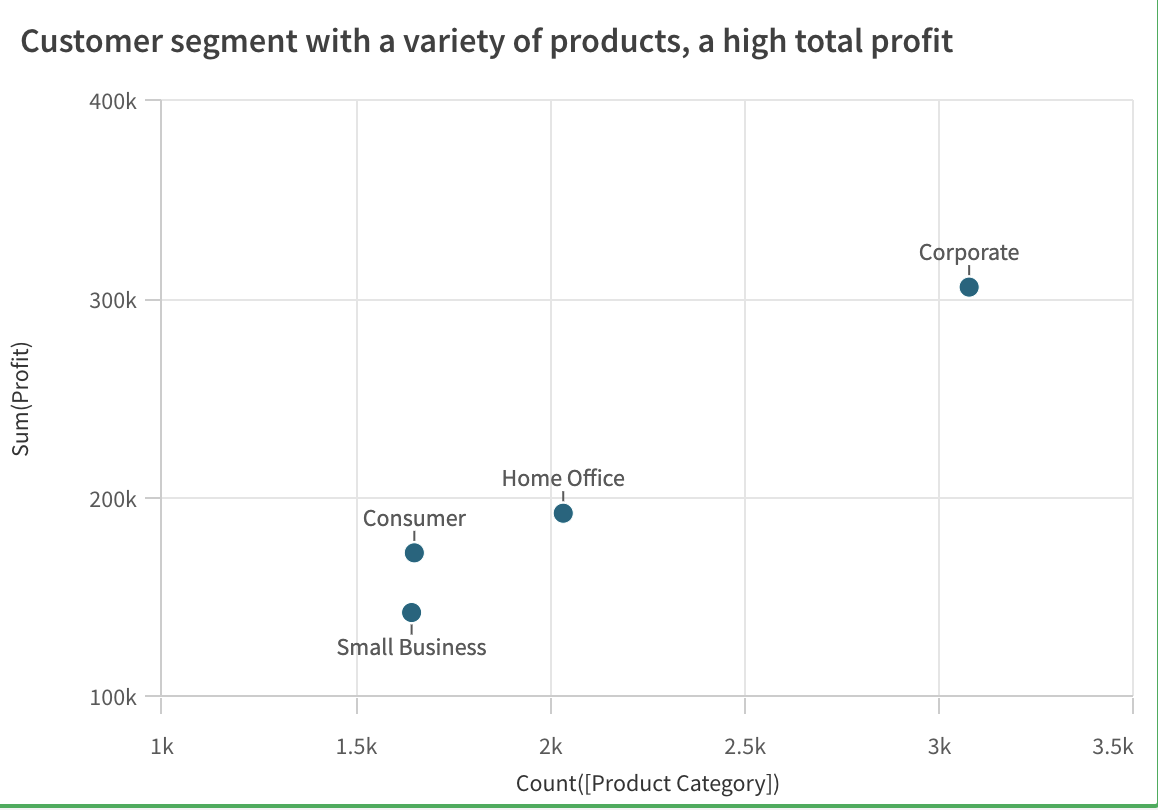
Graphic 1.k



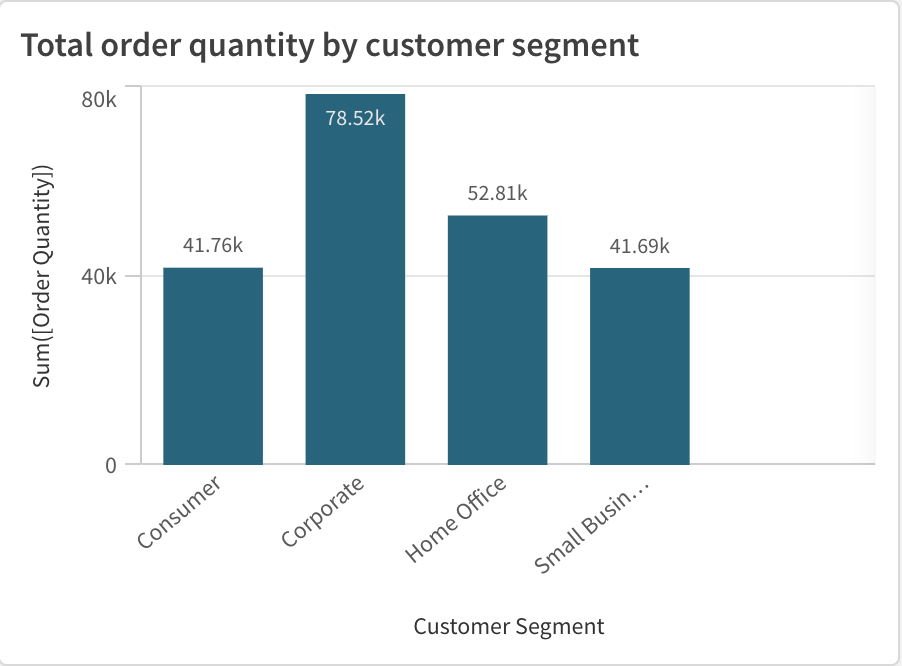
Graphic 1.l



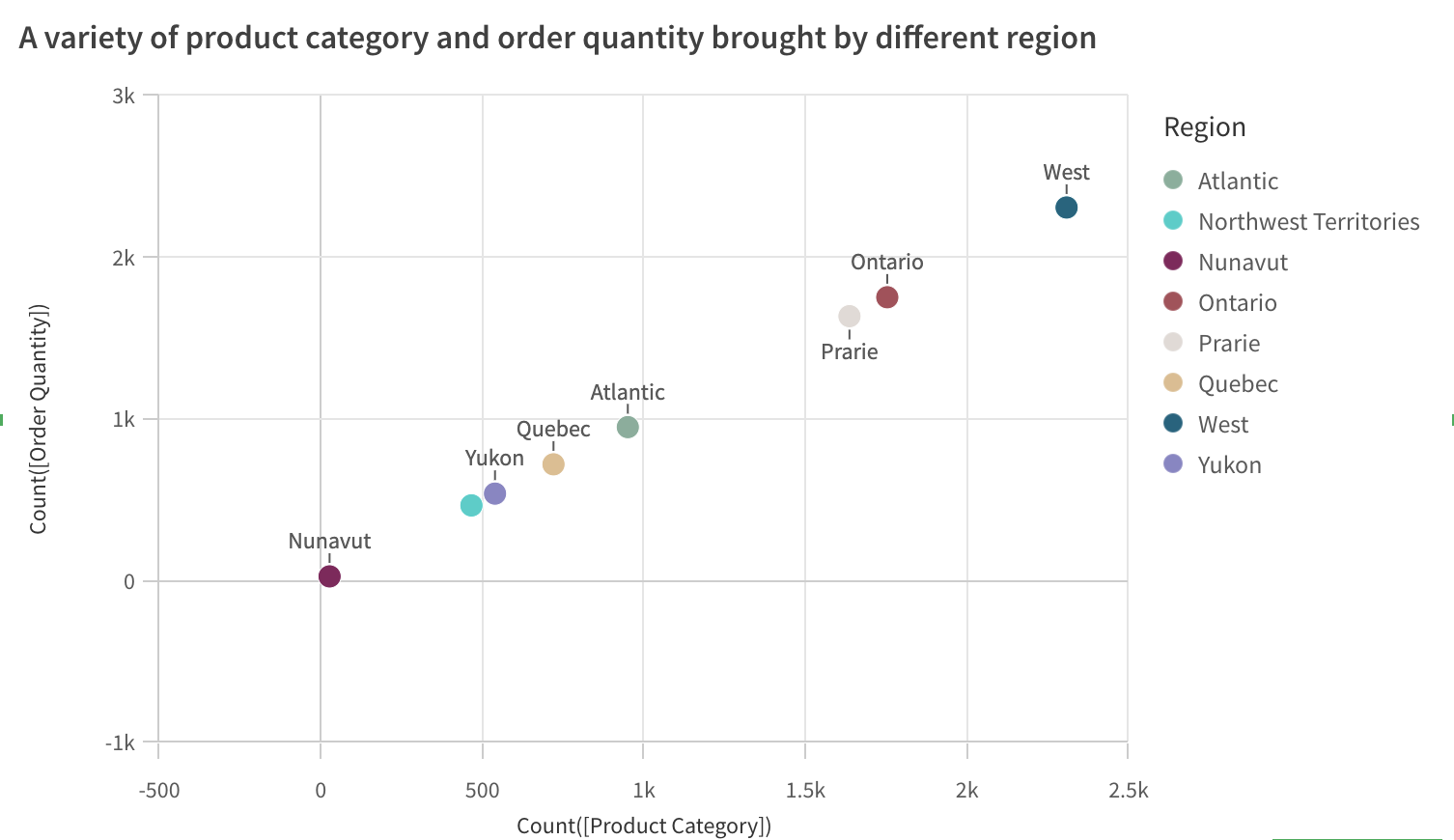
Graphic 1.m



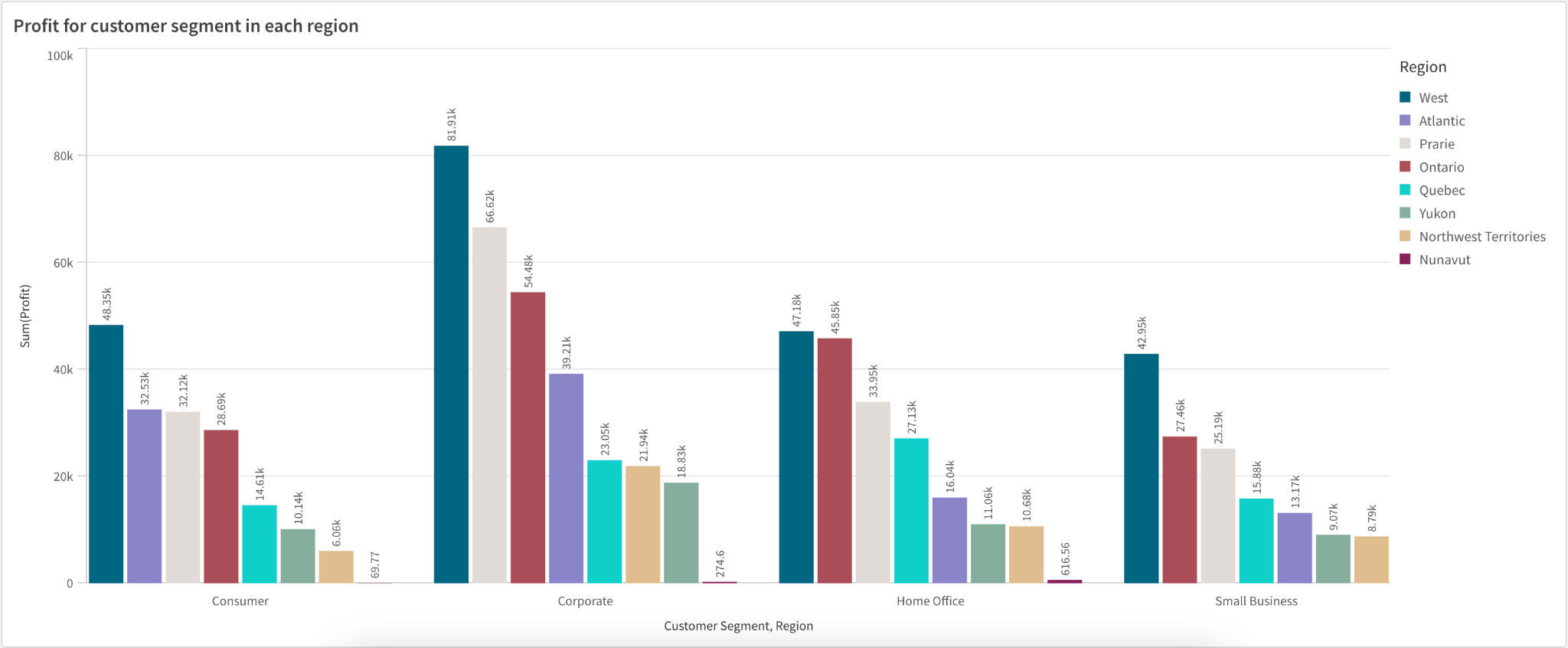
Graphic 2.a



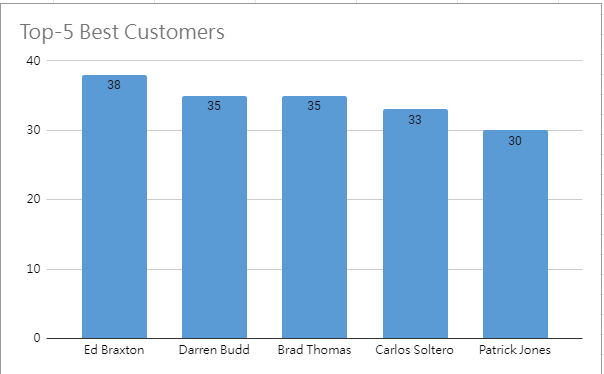
Graphic 2.b



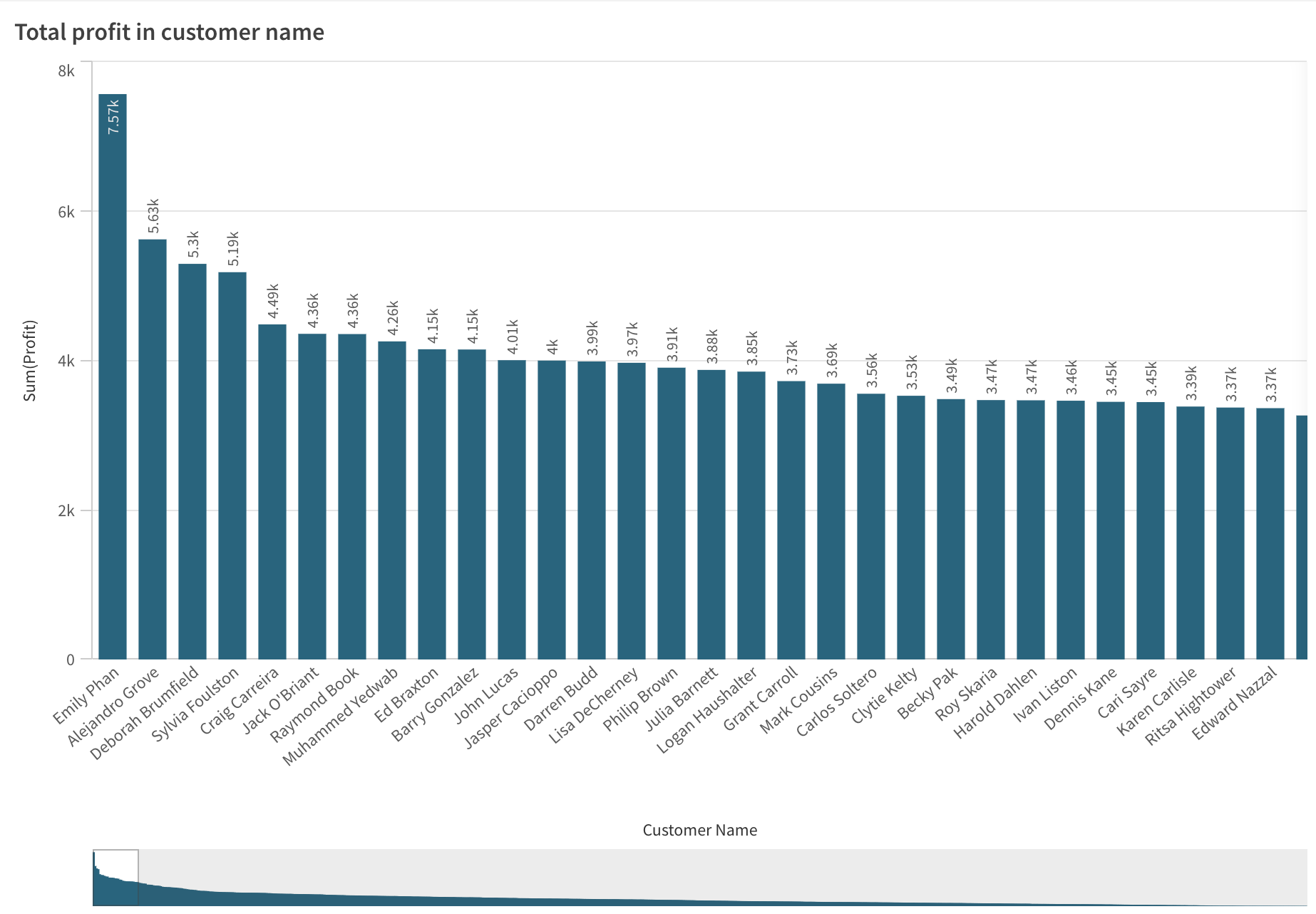
Graphic 2.c



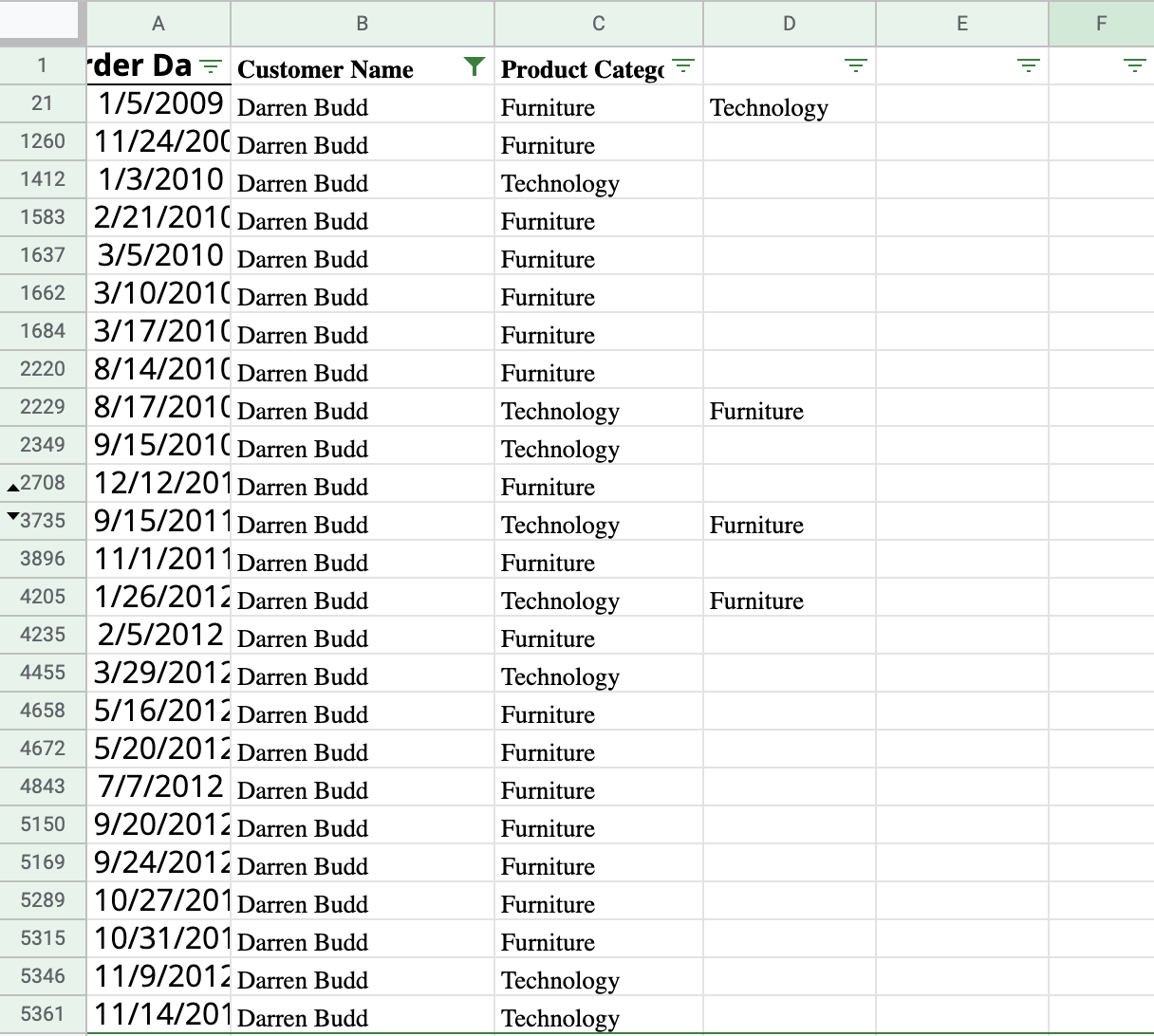
Graphic 2.d



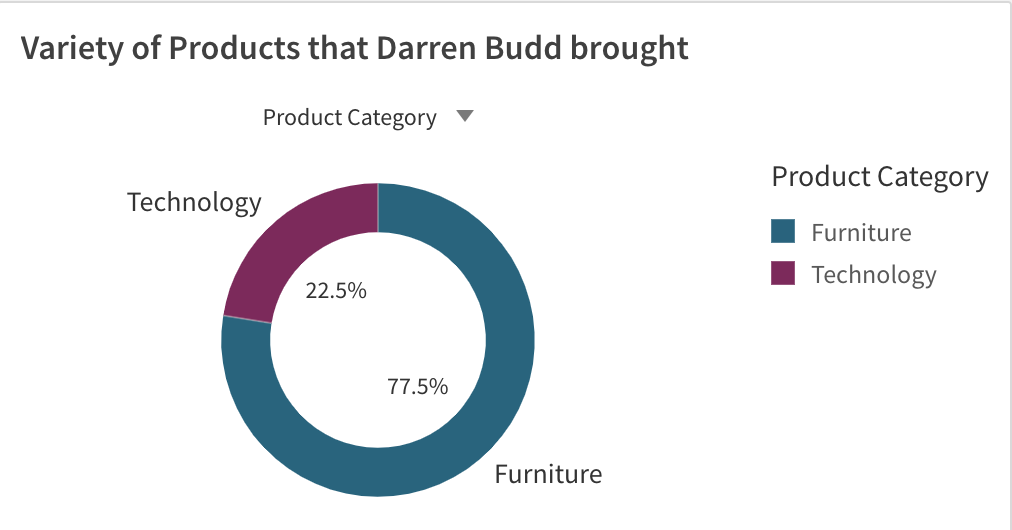
Graphic 2.e



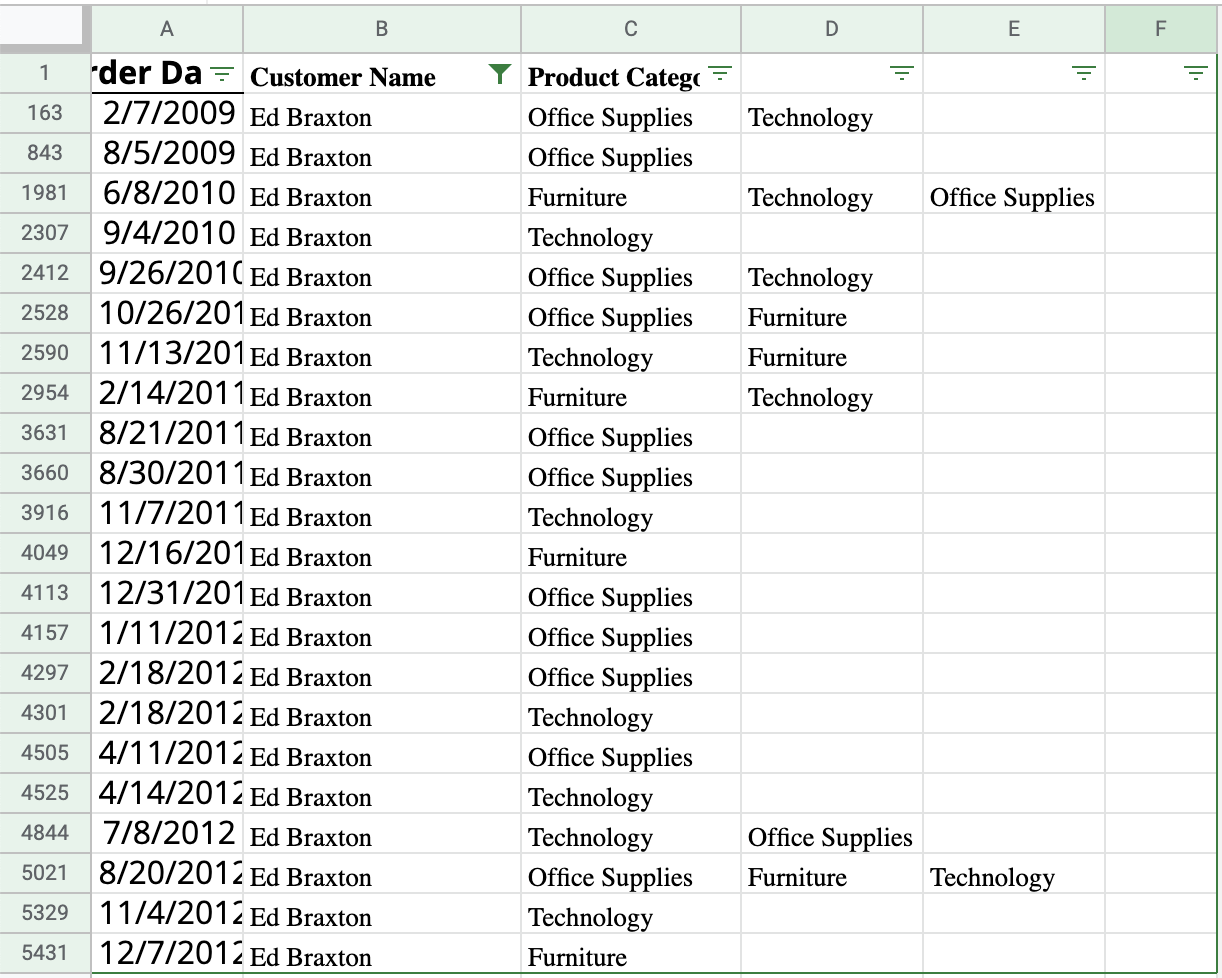
Graphic 2.f



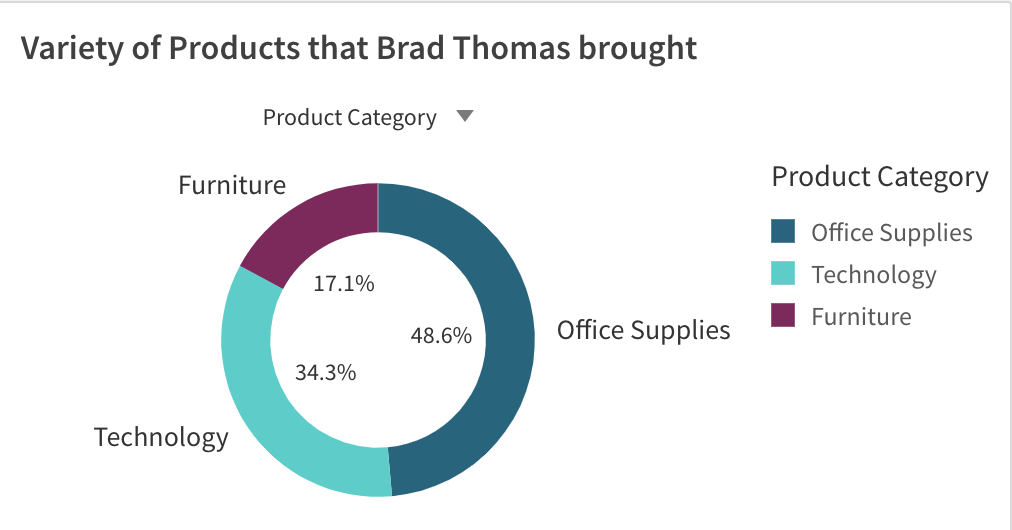
Graphic 2.g Product category bought from Darren Budd



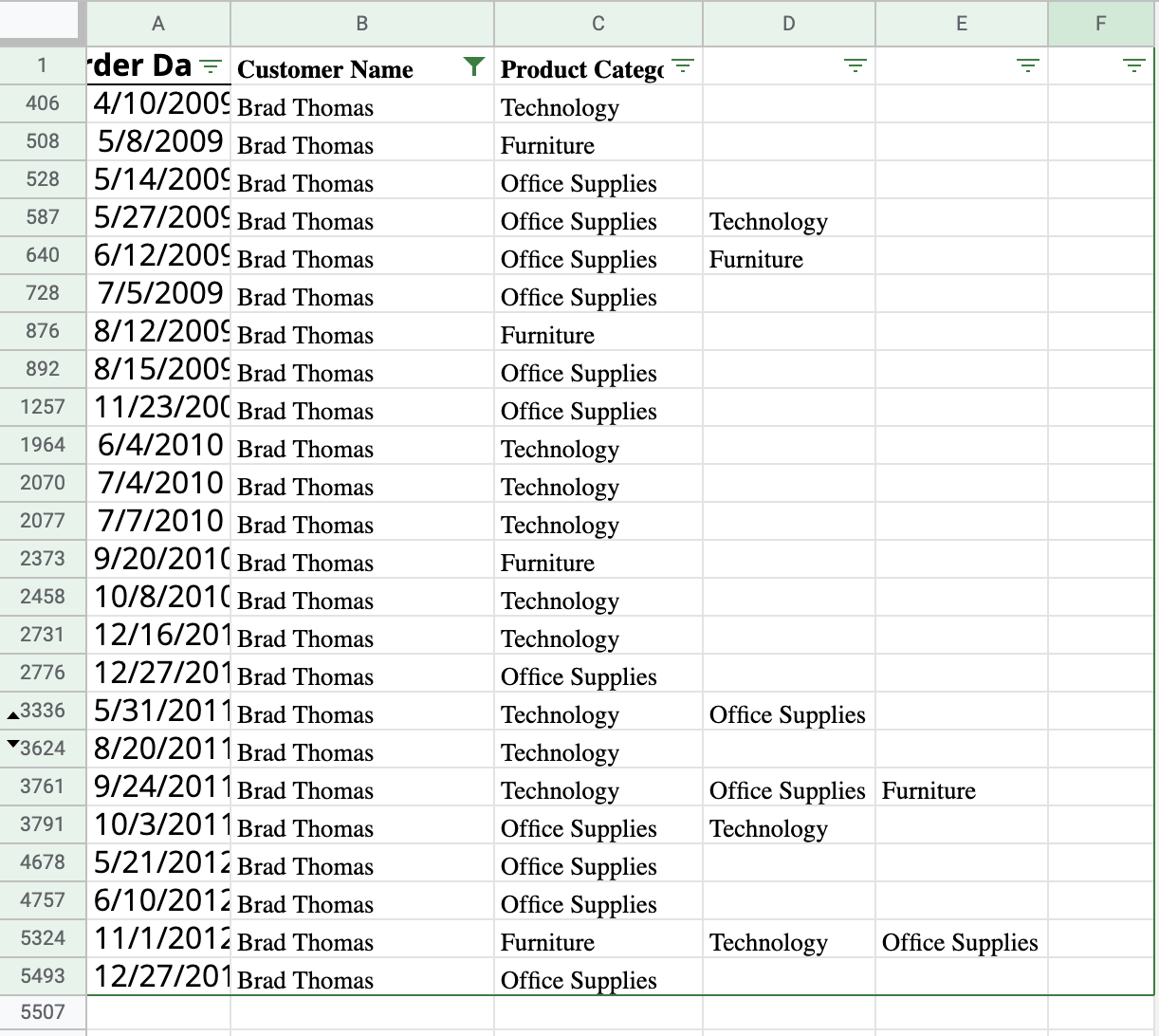
Graphic 2.h



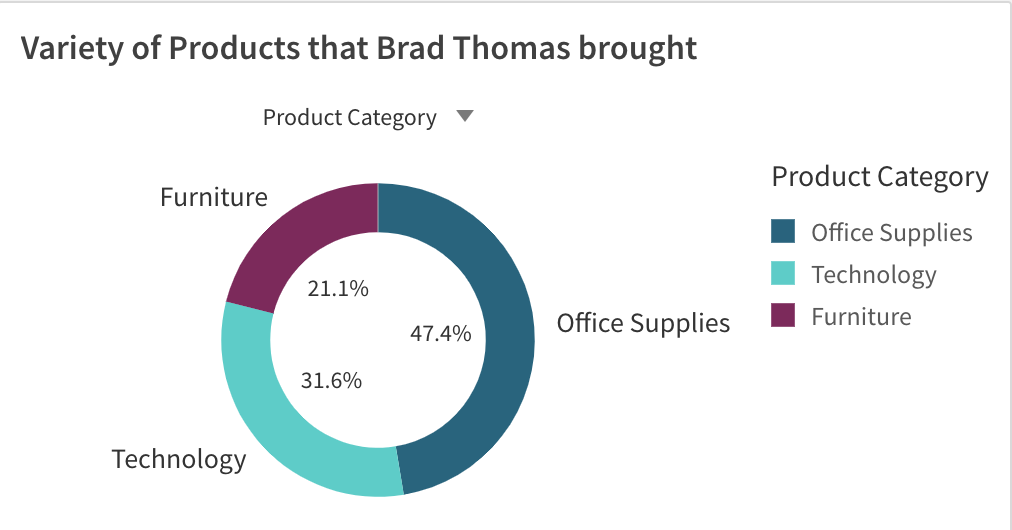
Graphic 2.i Product category bought from Ed Braxton



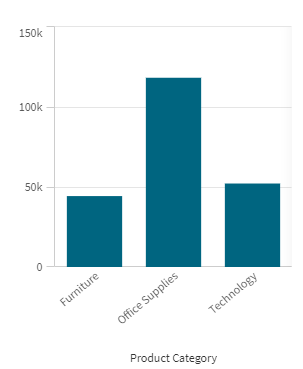
Graphic 2.j



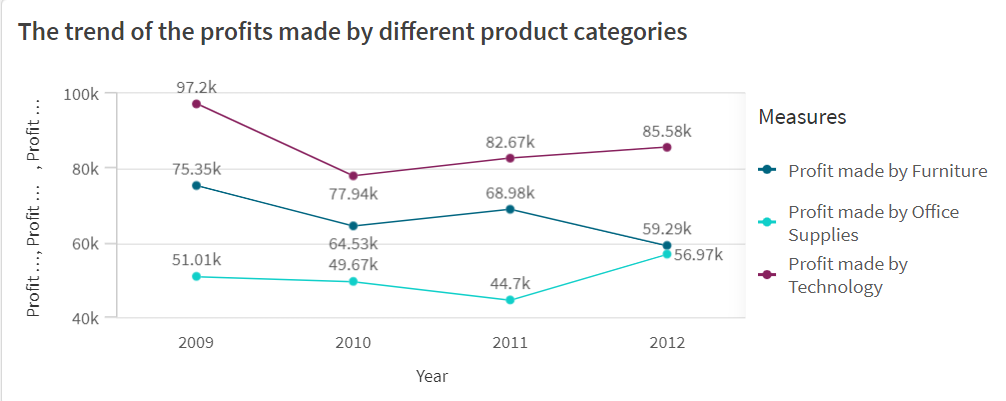
Graphic 2.k Product category bought from Brad Thomas



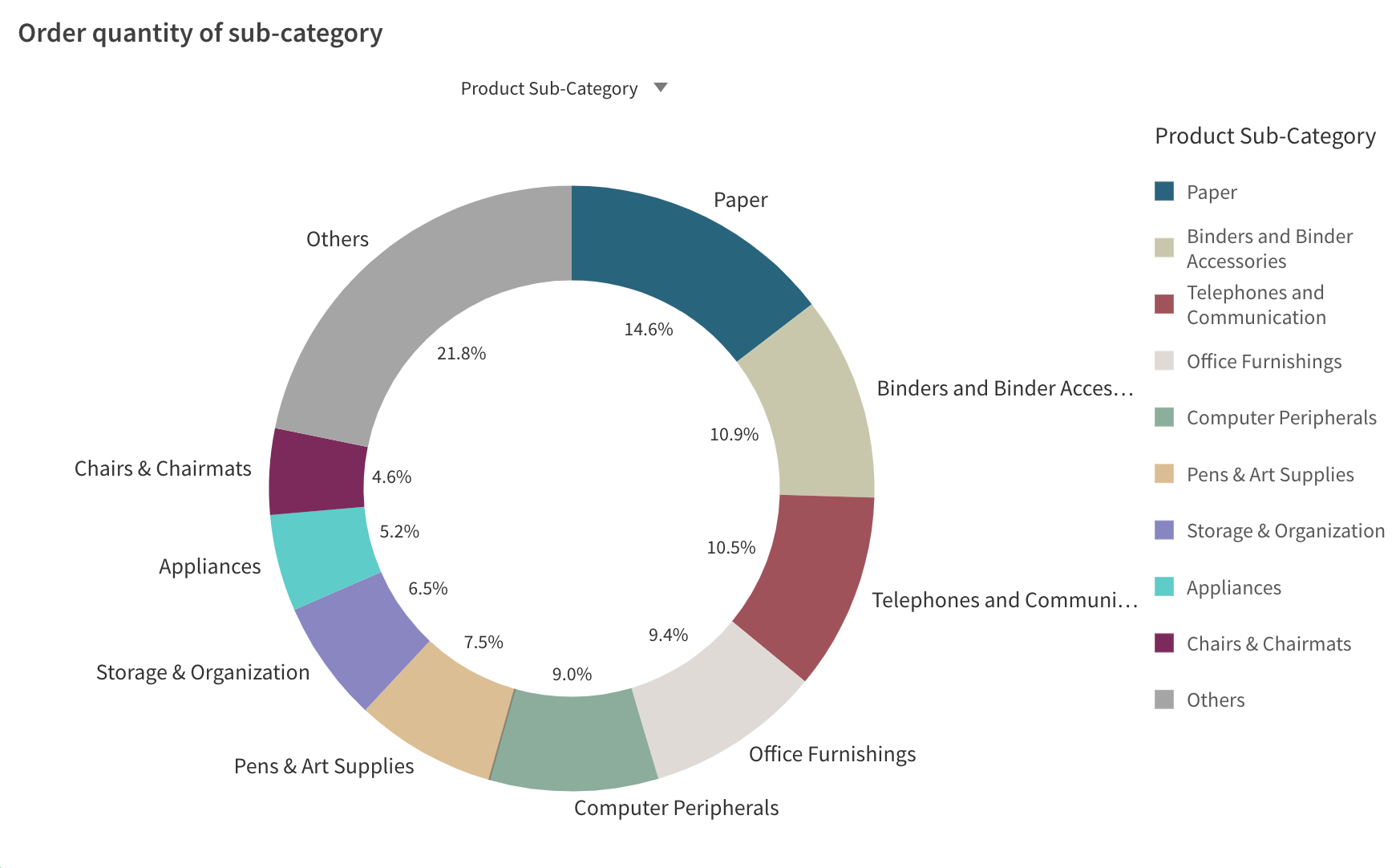
Graphic 2.l



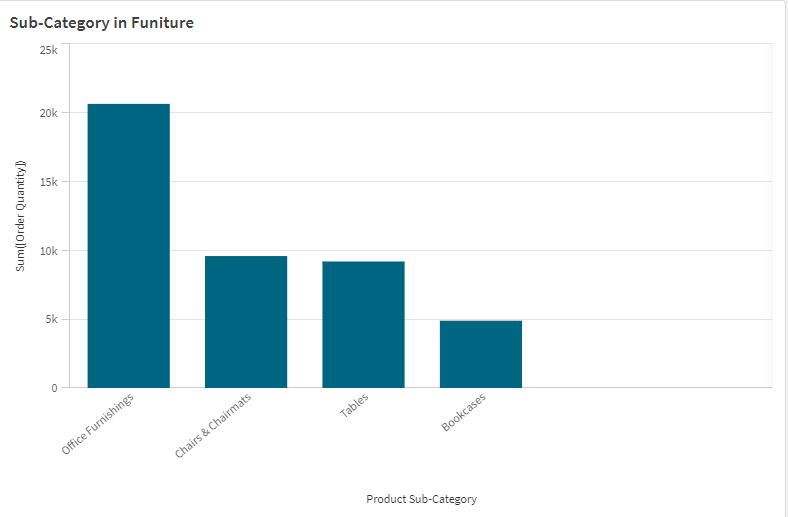
Graphic 3.a



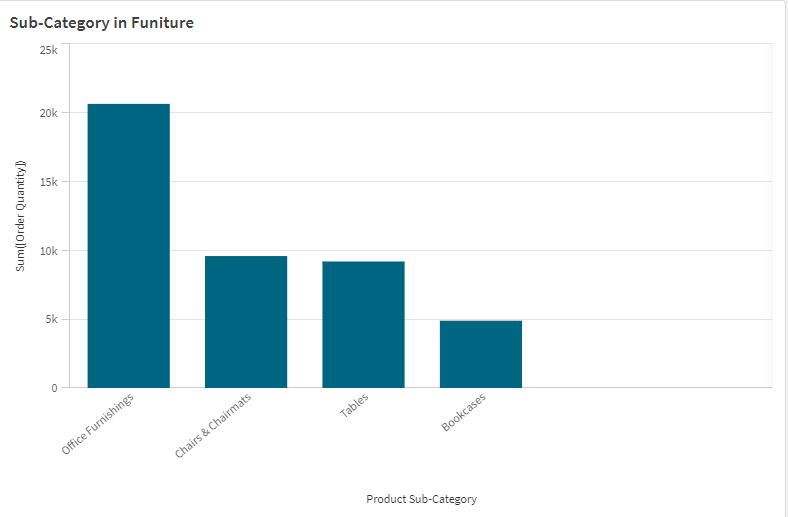
Graphic 3.b



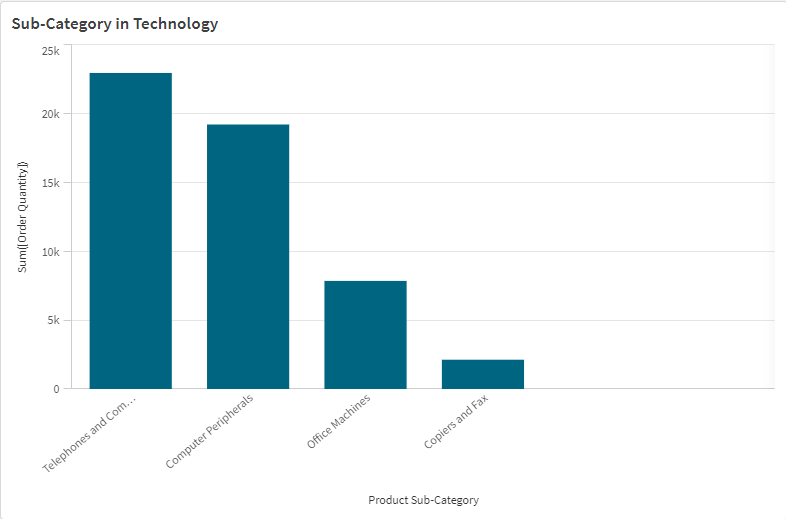
Graphic 3.c



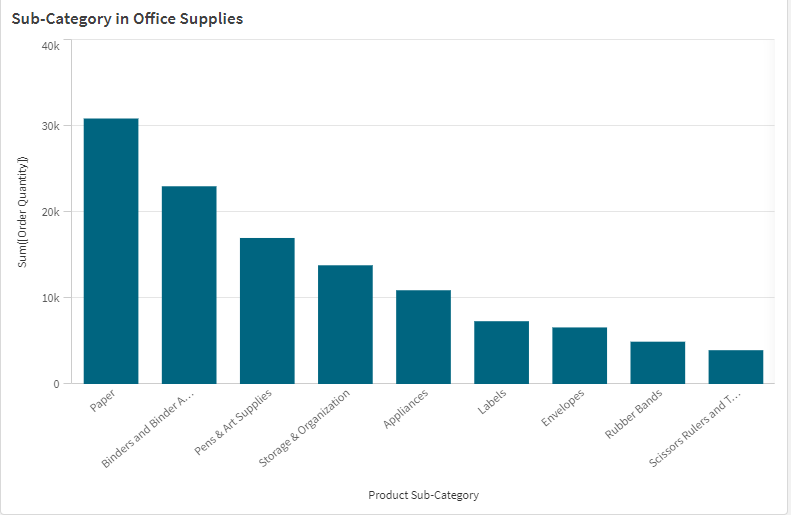
Graphic 3.d



Graphic 3.e



Graphic 3.f



Graphic 3.g