Technology has led to the expansion of the online e-commerce market which is expected to grow at a CAGR of almost 29% to $10.87 trillion during 2021-2025. How can e-commerce analytics further aid in this explosive growth and help firms derive more value from their marketing strategies?

Overall thoughts

• How can e-commerce firms identify suitable cross-selling opportunities based on historical transactions?
• How can e-commerce firms decide on the best sales target for a particular city, region, state or product category?
• How can analytics help companies to decide what changes they can make to their marketing mix?
• To answer the questions above, the team aims to build a model using Rstudio that incorporates drivers and meaningful targets to help firms determine sales targets. Visualization and market basket analysis will also be utilized to identify actionable insights that can be translated to marketing strategies.

Hypotheses

• The team hypothesized that marketing efforts will have varying impacts on different regions which will ultimately affect the firm's profitability. Certain drivers that may drive these differences include the discount rates on products, individual product profitability and the number of e-commerce orders.
• Research in the field has suggested that Sales may be affected by Profits, Margins, Quantity, Marketing and variable expenses in the consumer value chain.
• A firm's performance and sales can be heavily affected by its products. The ability to promote and sell multiple products together will greatly benefit the firm in driving sales.

Proposed approach (R-Studio Analysis)

• After reviewing the dataset, the team discovered that there were no null values present and that data preparation was performed to obtain the margin of the products using sales less profits. This allows the team to analyze the synergistic attributes of the variables onto sales and whether products with negative margin should be discontinued or not.
• Support Vector Model (SVM) using Linear Analysis was performed to obtain the best fit between train and test splits of 70-30 to allow for model testing onto 4 Regions in the United States (Central, East, South, West) based on the data provided and 3 different product Categories (Technology, Office Supplies, Furniture) using a 4-dimensional hyperplane visualization to generate the Mean Value of the predicted optimal Sales targets. The position and orientation of the hyperplane is influenced by Support Vectors, which are data variables established as Profit, Product Margins, Quantity Ordered and Customer Discounts.
• In this model, a k-fold cross validation was performed on trainset data in SVM to continuously cross-validate the dataset variables the model 10 times and resample/repeat the procedure 3 times with control parameters to obtain the optimal trainset model to obtain predicted sales using cross-validation method for testing of the 30% split test set.
• Market-Basket Analysis or Association Rule is used to determine the correlation and co-occurrences between each of the products. For this approach to work, transactional information such as transaction codes and product codes will be required. This approach segregates the data into various parameters: support, confidence, lift, and minlen. Support refers to the frequency of the item set's appearance as a comparison to the whole dataset, confidence refers to the frequency of the item set's appearance as a comparison to the primary item, lift refers to association between the two items and determines the confidence that the secondary item will be purchased together with the primary item, and minlen refers to the minimum number of items to be considered in the rule.
• Data visualization will also be utilized to examine profitability based on states, cities and individual products to identify factors that may have contributed to losses in certain ecommerce regions.

Deriving insights/solutions

• The team observed that certain products such as SAFCO Boltless Steel Shelving had a negative profit due to the high margin required to manufacture and deliver products to the customers. (Margin of $562, Sales of $454, Profit of -$108). The SVM model (Appendix A & B) showed that sales across regions and categories showed different Sales target using the Mean values derived using the optimal Quantity, Margins, Discounts and Profits, which are detailed in the table as shown:

Region	**East**	**West**	**South**	**Central**	**Category**	**Office Supplies**	**Technology**	**Furniture**
Sales target (USD$)	283.69	274.68	285.92	229.32	**Sales target (USD$)**	172.05	461.1	328.12

• The SVM model showed that technology needed a higher sales target to maximize sales, possibly due to the higher margins required in producing the items. From the Team's analysis, Office Supplies had an average margin of $103.04 while Technology had a product margin of $354.21 and Furniture with a margin of $309.58. (Appendix C) The increased product margin resulting from the technology sector can be attributed to the high Research and Developments to keep

abreast of the growing consumption and need for new technologies in the tech sector. In 2008, Apple spent USD$1.109 billion on R&D and generated $11.84 in 2009 gross profit per dollar spent on R&D, as shown in their Return of Return Capital (RORC). However, this hurdle can be overcome with increased Economies of Scale and the Metaverse adoption.

• Compared to the other regions, the central regions yielded a lower sales target, while the South and East Regions showed a higher sales target. In-depth analysis of the dataset showed that the Central regions were given on average an additional 9 to 13 basis points of discount compared to the East, West and South Regions. This pivoted profits to be lower on average for the Central Regions and to be 9 to 30 dollars lower per transaction when compared to the other 3 regions. (Appendix D) According to a Research Study performed by The Central American Group, the ecommerce channel accounted for 8% of all retail sales in Panama (Central America). Despite the low Technology penetration in Central America, it is expected that with increased technology consumption in Central America, as more than 60% of the population in the region will be using the internet, Central America has the potential to generate a large number of online consumers and hence should focus on expanding the commerce outreach to the region. Hence, greater discounts given to the region can be seen to attract future consumers of the e-commerce industry.

• The cross validation of the model onto the train and test set was successful with employability of the data points attributed to R-Squared of 0.99, showing high data and variable fit onto the hyperplane in question when analysing sales targets. Low error values of less than 30 were observed across all models despite the dataset having high data values, showing accuracy of SVM in sales analysis. However, the team recognizes the limitations of SVM includes the utilizations of models that would be subjected to hindsight bias, as various macroeconomic scenarios or errors resulting in fluctuations in Sales scores prediction was not factored in this model, resulting in possible lack of accuracy in data comparison.

• For the *Market Basket Analysis*, the team first analysed based on product level. Upon conducting the analysis, the team discovered that repeated transactions within this dataset were not common. For example, there were only up to 2 similar and repeated transactions for each itemset analysed. This shows that the product offerings within this dataset were too diverse, and it would be unreliable to make recommendations based on individual products. Thus, the team decided to focus on Sub-Category and Category where more insights can be derived. (Refer to Appendix E)

• From the Sub-Category analysis, it was found that the Sub-categories within this dataset was diverse as well. However, from the results generated, it was discovered that itemsets with a support level of 0.41-0.8% and confidence level of more than 50% contained a secondary sub-category item “Binders”. This shows that Binders are often a secondary item and together with a primary item. (Refer to Appendix F)

• From the Category analysis, it was found that Office Supplies were often purchased together Furniture and Technology products. The support level derived was 18%, with the total number of transactions appearing more than 300 times out of 3312 transactions. Looking at confidence level, it was found that Technology and Furniture as primary products have a higher confidence level of more than 50%. Thus, this shows that the customers often purchase Office Supplies as a secondary product to complement their primary purchase of Furnitures and Technology products. (Refer to Appendix G)

• From the above analysis, the team recommends the e-commerce firm to add product recommendations within the Office Supplies category. One Sub-Category that the firm can recommend is Binders, as it is a popular secondary product that customers purchase. Based on these categories, the firm can then recommend various products that might potentially catch the eye of the customer. For example, by recommending Binders, the firm will be able to suggest various types such as Hard Cover Binders, Soft Cover Binders, and many more. Furthermore, bundle pricings can be added for categories with lower support and confidence level to encourage customers to purchase them as a set and boost sales.

• An analysis of ecommerce profitability done using data visualization revealed New York, Seattle and Los Angeles to be the most profitable ecommerce hubs in the United States (Appendix H) while Chicago, Burlington and Philadelphia are the most unprofitable cities (Appendix I).

• Electronic products such as phones, copiers and machines were revealed to have contributed to much of the profits in the top 3 cities (Appendix J, K & L) whereas binders seem to be the driving force behind most of the losses in the least profitable cities (Appendix M, N & O).

• Using Chicago city as a case study, an investigation revealed that binders have the second highest sales (Appendix P) among all the product offerings but suffer high losses. One explanation for this is the high discount rate offered by the firm (Appendix Q). Secondly, certain binder products such as Ibico and Fellowes appear to contribute significantly to the losses (Appendix R).

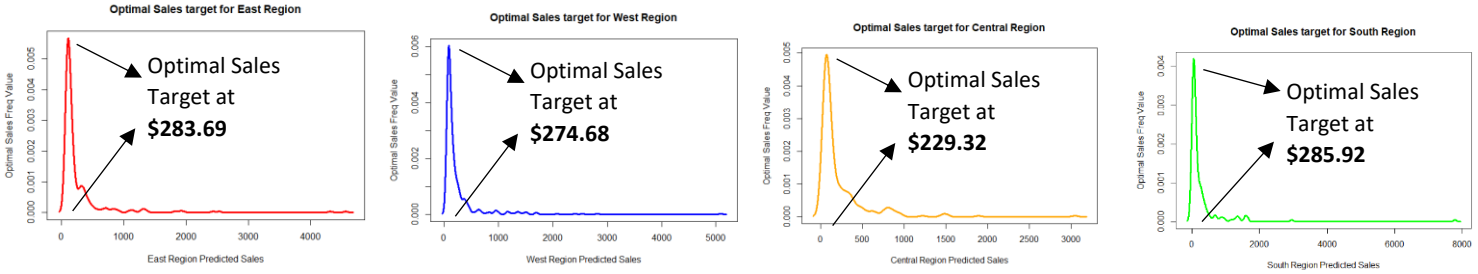
• With these insights, the team recommends that the ecommerce firm intensify their marketing efforts in highly profitable cities and states with a strong focus on the most profitable/popular product offerings. As for the least popular cities, firms can consider changing their product mix in those cities or perform an adjustment to their existing pricing and discount policies to ensure that sales can at least break even.

References

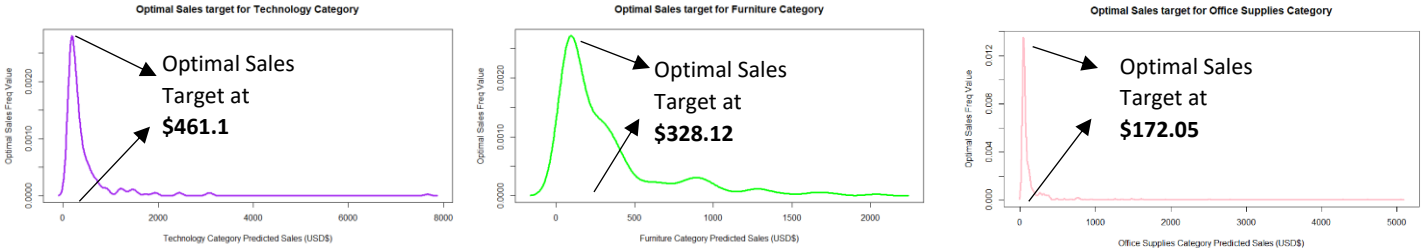
1. SPEC INDIA. (2021, January 15). Market basket analysis in R: A detailed and extensive guide. SPEC INDIA. Retrieved January 13, 2022, from https://www.spec-india.com/blog/market-basket-analysis-in-r
2. Kadimisetty, A. (2018, May 13). Association rule mining in R. Medium. Retrieved January 13, 2022, from https://towardsdatascience.com/association-rule-mining-in-r-ddf2d044ae50
3. Rpubs (n.d.) TuneGrid and TuneLength in Caret. Accessed 12 January 2021. Retrieved January 13, 2022, from https://rpubs.com/Mentors_Ubiqum/tunegrid_tunelength
4. McClure, B. (2021, May 19). R&D spending and profitability: What's the link? Investopedia. Retrieved January 13, 2022, from https://www.investopedia.com/articles/fundamental-analysis/10/research-development-rorc.asp
5. Colantuoni, S. (2021, June 16). The Electronic Commerce Market in Central America reached $ USD 3.7 billion during 2020. The Central American Group. Retrieved January 13, 2022, from https://www.thecentralamericangroup.com/electronic-commerce-in-central-america/

Appendices

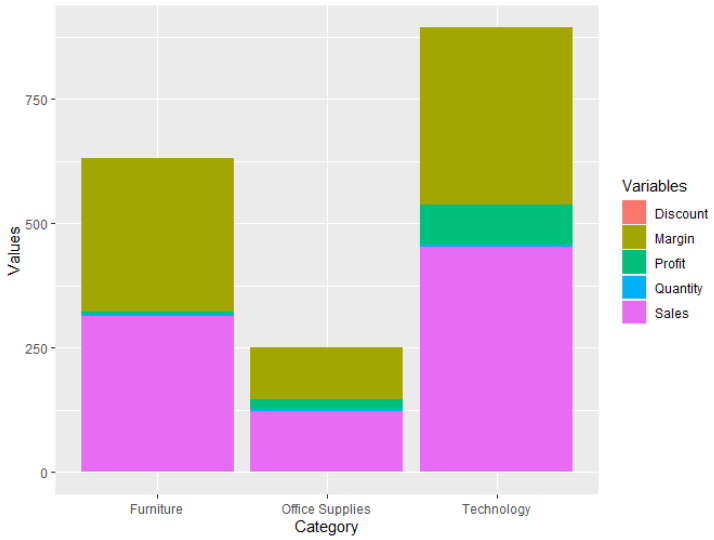
Appendix A: SVM Predicted Sales target across US Regions (East, South, West, Central)



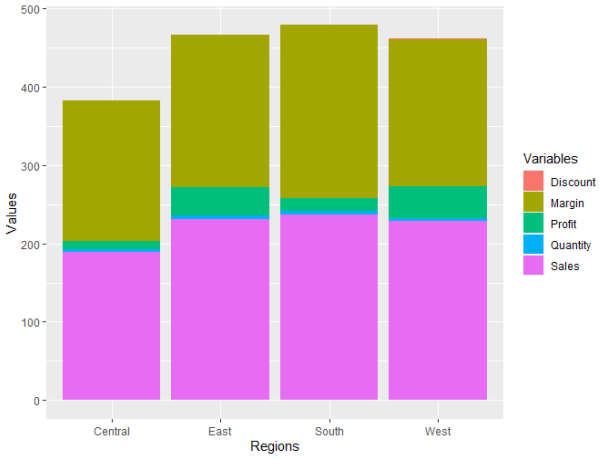
Appendix B: SVM Predicted Sales target across Categories (Technology, Office Supplies, Furniture)



Appendix C: Differences in Sales across categories attributed to Discounts, Quantity, Margin and Profit (Technology, Office Supplies, Furniture)



Appendix D: Differences in Sales across US regions attributed to Discounts, Quantity, Margin and Profit (East, South, West, Central)



Appendix E: Market-Basket Analysis Results based on Product ID

	LHS	RHS	support	confidence	coverage	lift	count
1	{FUR-FU-10003832}	{OFF-BI-10001359}	0.001185536	0.6666667	0.001778305	224.9333	2
2	{OFF-BI-10001359}	{FUR-FU-10003832}	0.001185536	0.4000000	0.002963841	224.9333	2
3	{OFF-AP-10000938}	{OFF-PA-10002195}	0.001185536	0.6666667	0.001778305	281.1667	2
4	{OFF-PA-10002195}	{OFF-AP-10000938}	0.001185536	0.5000000	0.002371073	281.1667	2
5	{OFF-AR-10001919}	{OFF-LA-10004544}	0.001185536	0.6666667	0.001778305	281.1667	2
6	{OFF-LA-10004544}	{OFF-AR-10001919}	0.001185536	0.5000000	0.002371073	281.1667	2
7	{FUR-CH-10002017}	{OFF-AP-10001563}	0.001185536	0.6666667	0.001778305	281.1667	2
8	{OFF-AP-10001563}	{FUR-CH-10002017}	0.001185536	0.5000000	0.002371073	281.1667	2
9	{TEC-PH-10000702}	{TEC-AC-10002049}	0.001185536	0.5000000	0.002371073	168.7000	2
10	{TEC-AC-10002049}	{TEC-PH-10000702}	0.001185536	0.4000000	0.002963841	168.7000	2
11	{TEC-PH-10004447}	{FUR-TA-10001039}	0.001185536	0.5000000	0.002371073	210.8750	2
12	{FUR-TA-10001039}	{TEC-PH-10004447}	0.001185536	0.5000000	0.002371073	210.8750	2

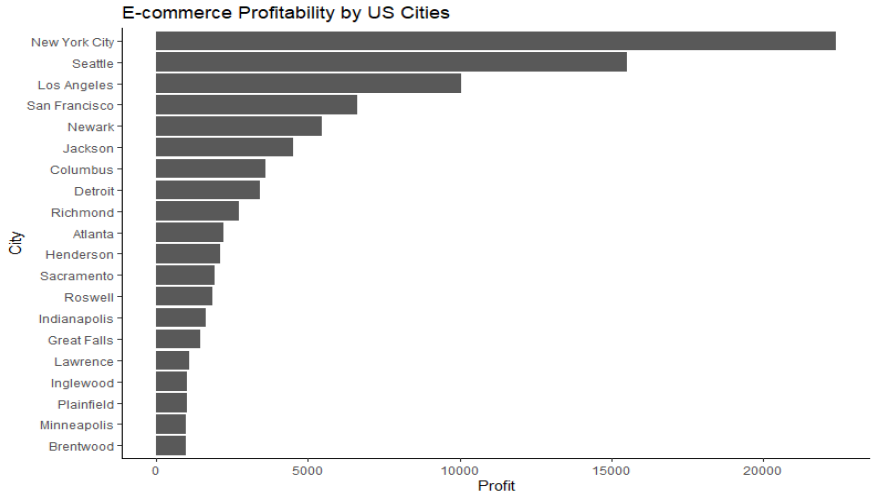
Appendix F: Market-Basket Analysis Results based on Sub-Category

	LHS	RHS	support	confidence	coverage	lift	count
1	{Fasteners,Storage}	{Binders}	0.004742146	0.5000000	0.009484292	1.952546	8
2	{Fasteners,Furnishings}	{Binders}	0.004742146	0.6153846	0.007705987	2.403134	8
3	{Appliances,Chairs}	{Binders}	0.006520451	0.5500000	0.011855365	2.147801	11
4	{Accessories,Appliances}	{Binders}	0.007113219	0.6315789	0.011262596	2.466374	12
5	{Appliances,Storage}	{Binders}	0.006520451	0.5500000	0.011855365	2.147801	11
6	{Appliances,Furnishings}	{Binders}	0.008298755	0.5000000	0.016597510	1.952546	14
7	{Appliances,Chairs,Furnishings}	{Binders}	0.004149378	0.7777778	0.005334914	3.037294	7
8	{Appliances,Binders,Chairs}	{Furnishings}	0.004149378	0.6363636	0.006520451	3.676526	7
9	{Appliances,Binders,Furnishings}	{Chairs}	0.004149378	0.5000000	0.008298755	4.847701	7

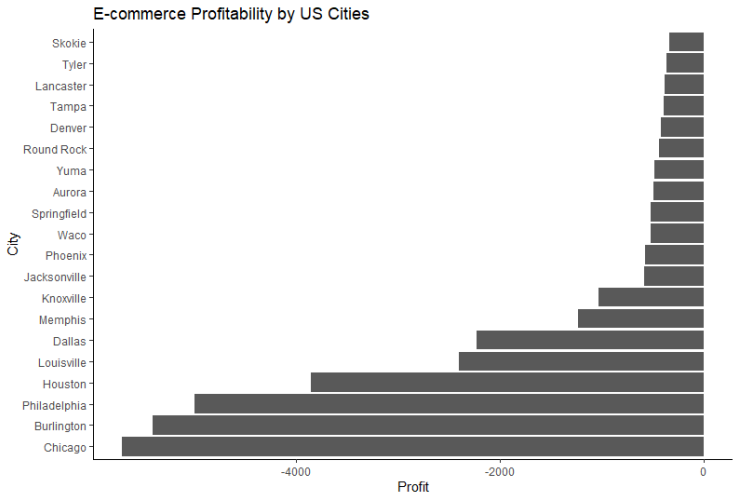
Appendix G: Market-Basket Analysis Results based on Category

	LHS	RHS	support	confidence	coverage	lift	count
1	{Technology}	{Furniture}	0.10077060	0.3231939	0.3117961	0.9667166	170
2	{Furniture}	{Technology}	0.10077060	0.3014184	0.3343213	0.9667166	170
3	{Technology}	{Office Supplies}	0.18731476	0.6007605	0.3117961	0.7967633	316
4	{Office Supplies}	{Technology}	0.18731476	0.2484277	0.7540012	0.7967633	316
5	{Furniture}	{Office Supplies}	0.18020154	0.5390071	0.3343213	0.7148624	304
6	{Office Supplies}	{Furniture}	0.18020154	0.2389937	0.7540012	0.7148624	304
7	{Furniture,Technology}	{Office Supplies}	0.06816835	0.6764706	0.1007706	0.8971744	115
8	{Office Supplies,Technology}	{Furniture}	0.06816835	0.3639241	0.1873148	1.0885459	115
9	{Furniture,Office Supplies}	{Technology}	0.06816835	0.3782895	0.1802015	1.2132592	115

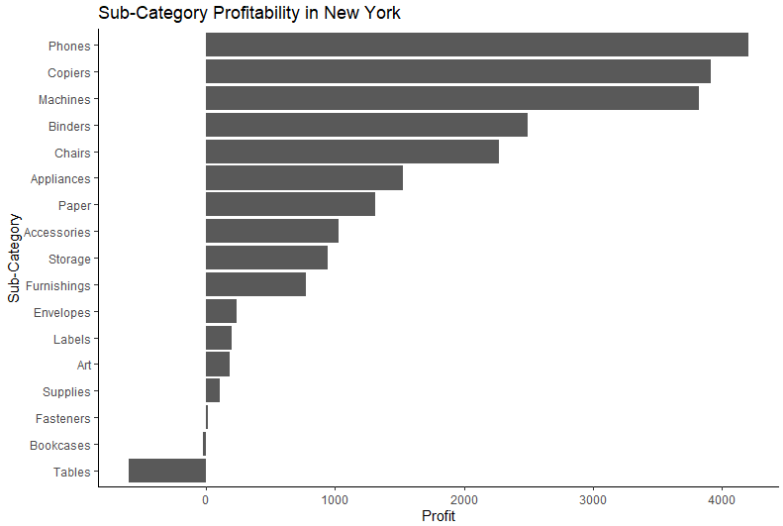
Appendix H: Most profitable US cities (E-commerce)



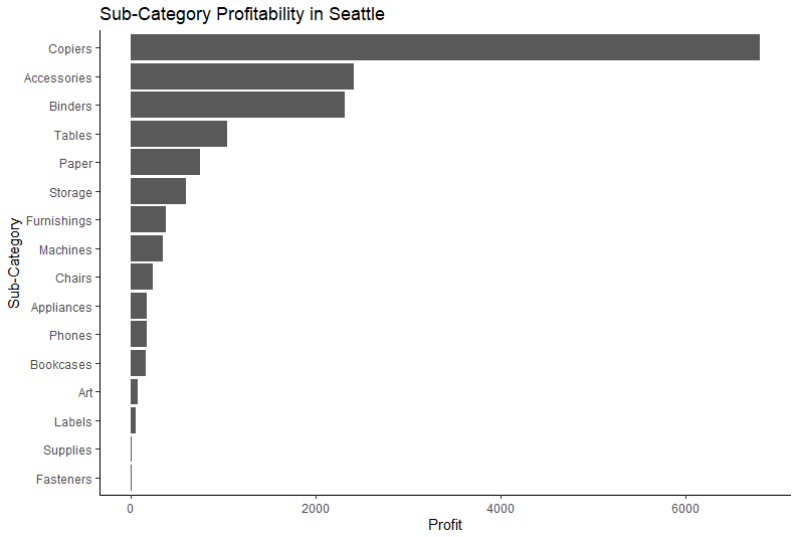
Appendix I: Least profitable US cities (E-commerce)



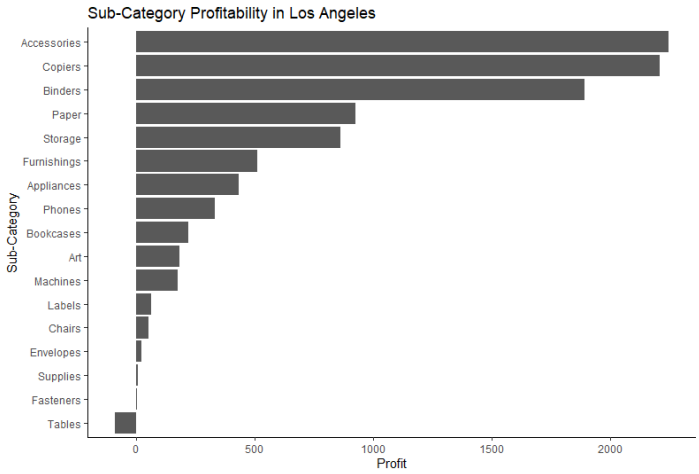
Appendix J: Sub-Category Profits (New York City)



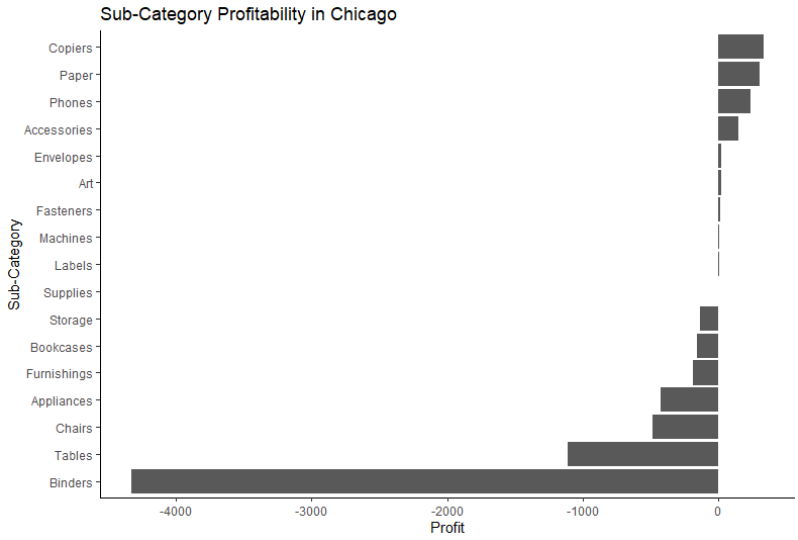
Appendix K: Sub-Category Profits (Seattle)



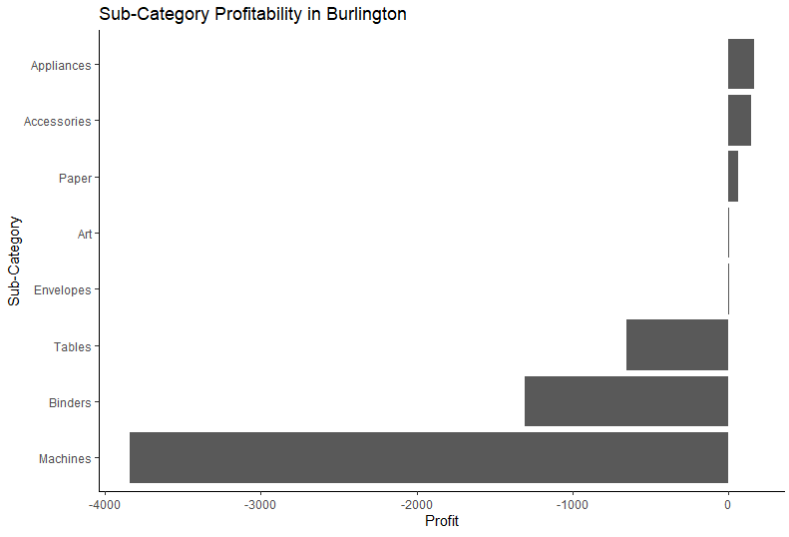
Appendix L: Sub-Category Profits (Los Angeles)



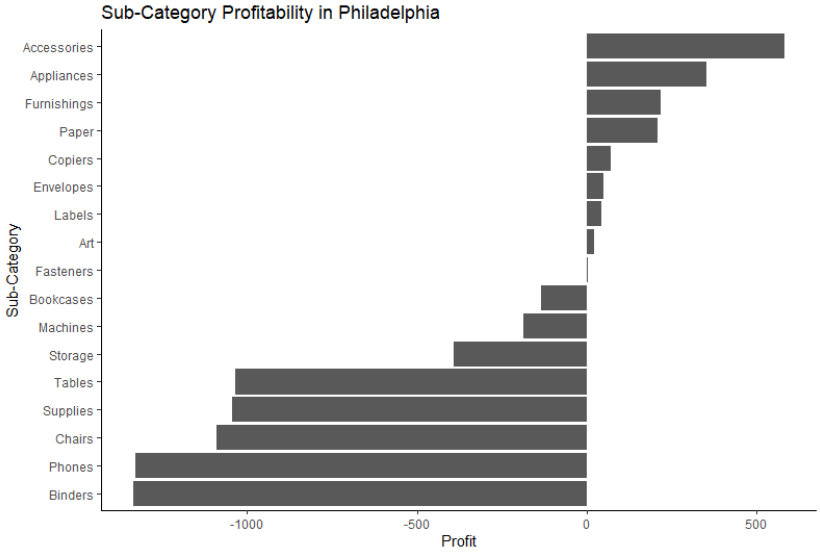
Appendix M: Sub-Category Profits (Chicago)



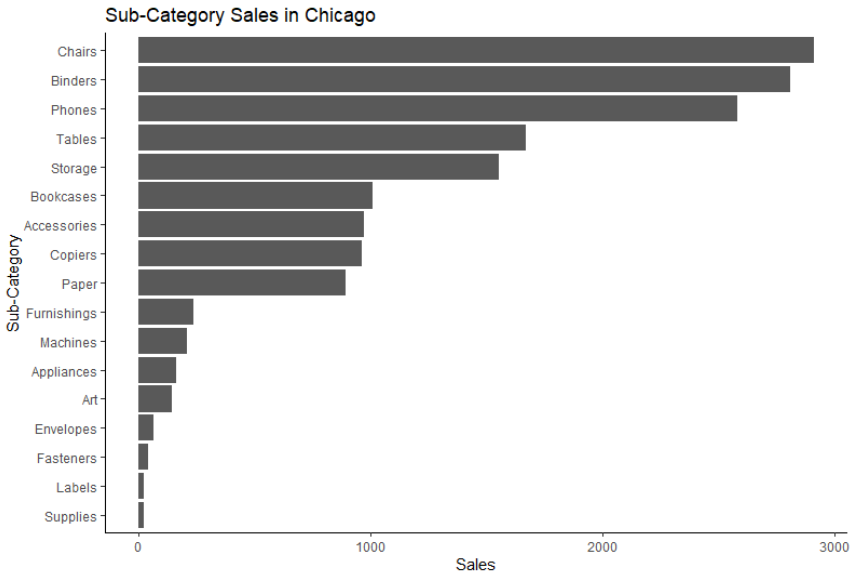
Appendix N: Sub-Category Profits (Burlington)



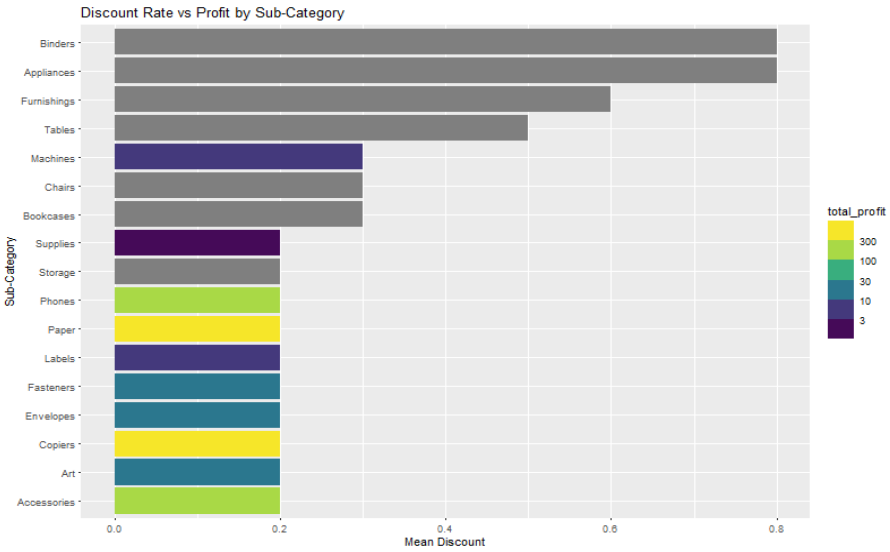
Appendix O: Sub-Category Profits (Philadelphia)



Appendix P: Sub-Category Sales in Chicago



Appendix Q: Discount Rate vs Profit by Sub-Category (Chicago)



Appendix R: Profitability of Individual Binders (Chicago)