**SCHOOL OF ENGINEERING AND TECHNOLOGY**

**ASSIGNMENT FOR THE**

**BSC (HONS) IS (BUSINESS ANALYTICS); YEAR 3**

**ACADEMIC SESSION MARCH 2021**

**IST3144: PROBLEM SOLVING USING ANALYTICS TOOLS**

**DEADLINE: Week 12 (2 July, Friday 2pm) via eLearn by Group Leader\* only**

**STUDENT NAME\*: \_\_\_Chua Wen Soong\_\_\_\_\_ STUDENT ID: \_\_\_18032573\_\_\_**

**STUDENT NAME: \_\_\_\_\_\_\_Ng Jia Hui\_\_\_\_\_\_\_\_ STUDENT ID: \_\_\_17106386\_\_\_**

**STUDENT NAME: \_\_\_\_Ryan Liew Kok Lam\_\_\_ STUDENT ID: \_\_\_17114950\_\_\_**

**INSTRUCTIONS TO CANDIDATES**

# 

* This assignment will contribute 30% to your final grade.

**Learning Outcome:**

LO1: Apply a useful analysis framework to solve a business problem.

LO3: Construct a data analytics solution from real world problems using analytics tools.

**IMPORTANT**

The University requires students to adhere to submission deadlines for any form of assessment. Penalties are applied in relation to unauthorized late submission of work.

* Coursework submitted after the deadline but within 1 week will be accepted for a maximum mark of 40%.
* Work handed in following the extension of 1 week after the original deadline will be regarded as a non-submission and marked zero.

**Lecturer’s Remark** (Use additional sheet if required)

We …...Chua Wen Soong (18032573), Ng Jia Hui (17106386), Ryan Liew Kok Lam (17114950)....... (Name/Student ID) received the assignment and read the comments.

**Wen Soong, Jia Hui, Ryan Liew** (17/06/21)

**Academic Honesty Acknowledgement**

“We …Chua Wen Soong (18032573), Ng Jia Hui (17106386), Ryan Liew Kok Lam (17114950)..... (student name/Student ID) verify that this paper contains entirely our own work. I have not consulted with any outside person or materials other than what was specified (an interviewee, for example) in the assignment or the syllabus requirements. Further, I have not copied or inadvertently copied ideas, sentences, or paragraphs from another student. I realize the penalties (refer to page 16, 5.5, Appendix 2, page 44 of the student handbook diploma and undergraduate programme) for any kind of copying or collaboration on any assignment.”

**Wen Soong, Jia Hui, Ryan Liew** (17/06/21)

**IST3144 Problem Solving using Analytics Tools**

**Group Assignment**

**An Analysis on Company XYZ’s Business Portfolio for Future Growth**

**STUDENT NAME\*: \_\_\_Chua Wen Soong\_\_\_\_\_ STUDENT ID: \_\_\_18032573\_\_\_**

**STUDENT NAME: \_\_\_\_\_\_\_Ng Jia Hui\_\_\_\_\_\_\_\_ STUDENT ID: \_\_\_17106386\_\_\_**

**STUDENT NAME: \_\_\_\_Ryan Liew Kok Lam\_\_\_ STUDENT ID: \_\_\_17114950\_\_\_**

**Evaluation Rubric:**

| Assessment criteria | Weight (%) | Excellent (15-11) | Good (10-8) | Satisfactory (7-5) | Unsatisfactory (4-0) |
| --- | --- | --- | --- | --- | --- |
| Plan the analysis, address the problem and determine the best way to evaluate the data sets | 20 | All elements of methodology or framework are skillfully described and applied. | All elements of methodology or framework are appropriately described and applied. | Critical elements of methodology or framework are appropriately described and applied. | Description of inquiry design demonstrates a misunderstanding of methodology or framework. |
| Construct a data analytics solution using an analytics tools | 10 | Provides thorough and accurate description of data analytics solution. Skillfully use tools, variables, and report an insight into the steps of the logic. | Provides accurate description of data analytics solution. Appropriate use tools, variables and clear structure report. | Provides simple and mostly accurate descriptions of data analytics solution. There are occasional, minor errors and produce simple structure report. | Draws fundamentally incorrect conclusions about what the data mean for analytics. Lack of claim to support the logic and inconsistent structure report. |
| Remarks: | | | | | |

Table of Contents:

[**1. INTRODUCTION**](#_pjhbuqvcpzdq) **[Peter] 3**

[**2. DATA EXPLORATION, CLEANING AND MANIPULATION**](#_l48wvx24lrh6) **[Peter] 3**

[**3. CURRENT BUSINESS PROGRESS**](#_aplwoaaj85qm) **4**

[[3.1] Where are their incomes mainly coming from?](#_4nx9sev3x3me) [Ryan] 4

[[3.1.1] Analysis on Product Categories](#_q3q27smxdico) 4

[[3.1.2] Analysis on Product Subcategories](#_57uzaotf67hm) 4

[[3.1.3] Analysis on Product Categories Based on Time](#_tsvjnhp3kqvz) 5

[[3.1.4] Analysis on different Store Types:](#_59txgbp5wyps) [Peter] 6

[[3.1.5] Analysis on Product Categories within each Store Type](#_j5q0alvzwk7c) 6

[[3.2] Where/What is allowing them to make a big loss?](#_ua48wxcbyho8) 7

[[3.2.1] Income vs Loss of each Store Type](#_f43b2twni525) 7

[[3.2.2] Analysis of Product Category in each Store Type](#_bmu7b1832is4) 7

[[3.2.3] Analysis on Product Subcategories in each Store Type](#_m5hv0qnathqg) 8

[**4. TARGET MARKET**](#_xkixis135afe) **[Jia] 9**

[[4.1] Which generation is the most popular one?](#_ituzm0ukxvss) 10

[[4.2] In that generation, which gender contributes the most?](#_4x7zwj4blji6) 10

[[4.3] In that gender category, where are they mostly located?](#_ncxdzdle8xo6) 11

[[4.4] In that city code, what stores are available?](#_6a2yhjctn19y) 11

[[4.5] In that store, what is the most preferred product category bought?](#_cnv9gblx8yl1) 12

[[4.6] In that product category, what is the most preferred product sub-category bought?](#_6h6gcl1o5med) 12

[**5. SUGGESTIONS**](#_cv6e0wdozi16) **[Jia] 13**

[**6. CONCLUSION**](#_m8ic96diay9d) **[Ryan] 14**

[**7. REFERENCES**](#_67kfqr6lmfd4) **15**

[**8. APPENDIX**](#_rrhisfs51g8c) **[Ryan] 16**

[Appendix I](#_ffbi2id8mp14) 16

[Appendix II](#_u7evhbwkxcbp) 17

[Appendix III](#_nvqi3dga3wd2) 18

[Appendix IV](#_wnnl96lvzb08) 19

**AN ANALYSIS ON COMPANY XYZ’S BUSINESS PORTFOLIO**

**FOR FUTURE GROWTH**

Chua Wen Soong Ng Jia Hui Ryan Liew Kok Lam  
18032573@imail.sunway.edu.my 17013686@imail.sunway.edu.my 17114950@imail.sunway.edu.my

# 1. INTRODUCTION

Company XYZ runs a business with multiple products sold at diverse stores and locations. For a couple of years, this company has been operating without a clear identification of their target market, source of income and loss as well as their best selling product. As many other retail businesses are entering the market, company XYZ has gained many competitors and the CEO is now worried whether the business plan is strategically placed to stand out in the market. With that, assuming the current year is 2014, this paper aims to provide a clear insight on their current business progress in terms of their main sources of income, and the aspects that are allowing them to make losses. Additionally, the analysis will also help to determine their target market which consists of the customers who provide them with the most income. With such information gathered, it would help the employees in company XYZ to determine way forwards in order to boost their sales and retain their customers to set themselves up for future success.

# **2.** **D**ATA EXPLORATION, CLEANING AND MANIPULATION

Before any analysis was done, data exploration was first carried out to have a clearer understanding of each dataset provided as it allows users to have a clear overview of individual attributes which would in return help with the analysis of data [[1]](#kix.mcmslainte44).

For data exploration, it was found that the variable total\_amt has positive and negative values and with that, the positive values were taken as **income** while the negative values were taken as **returns**, whereas the sum of the two was taken as **profit**. Income and returns are assumed to be normally distributed with a slight skew due to some outliers in the data (refer to [Figure 21](#x95ke65b51pk) and [Figure 22](#an9lsoq8e9sq)).

Cleaning was performed in the “Customers” and “Transactions” datasets. In customers, 4 rows were deleted as they had missing values in either City or Gender. In the Transactions dataset, all duplicate rows (2175 rows) were deleted. A total of 20878 rows of data was left in the Transactions dataset.

Manipulation was performed by adding a “Quarter” column which contains the quarters the transactions took place for time based analyses. Additionally, “prod\_subcat\_code” in the “Transactions” table was renamed to “prod\_sub\_cat\_code” to merge with the “Product” dataset. Next, “customer\_id” in the “Customer” table was also renamed to “cust\_id” to merge with the “Transactions” dataset.

# 3. CURRENT BUSINESS PROGRESS

This section would cover the main income sources and the main losses of this organisation.

## [3.1] Where are their incomes mainly coming from?

### [3.1.1] Analysis on Product Categories

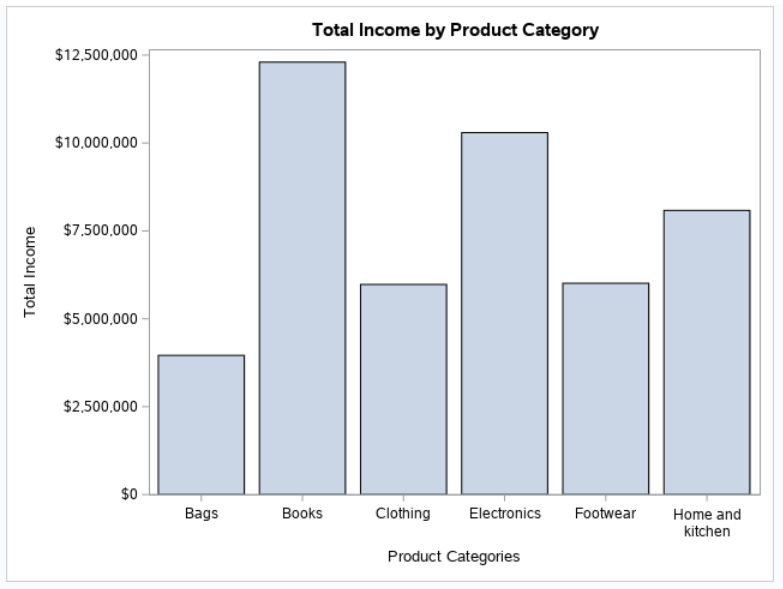
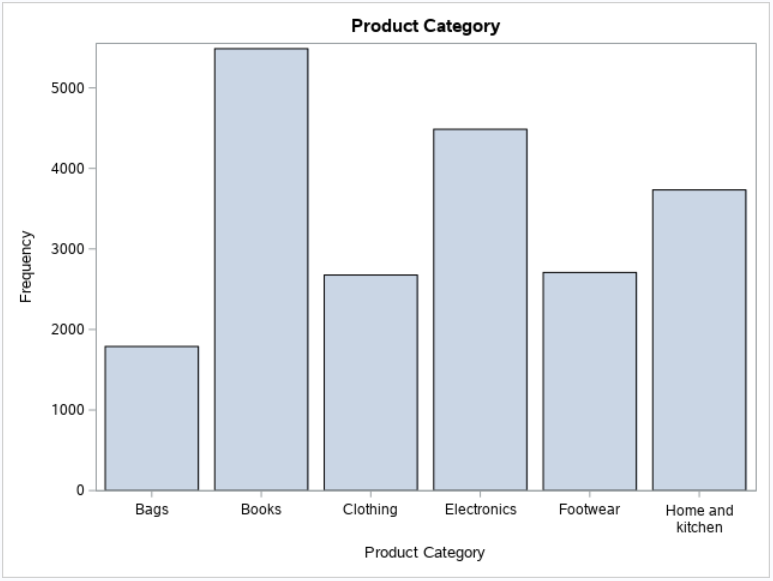
 Figure 1: Products sold by Product Category Figure 2: Total Income by Product Category

Figure 1 shows the frequency of each product category bought, with the highest being Books followed by Electronics, Home and kitchen, Footwear, Clothing, and lastly bags. This same pattern is shown in Figure 2, which shows the total income gained from each product category, with the highest being Books and Electronics that return amounts surpassing $10 million.

### [3.1.2] Analysis on Product Subcategories

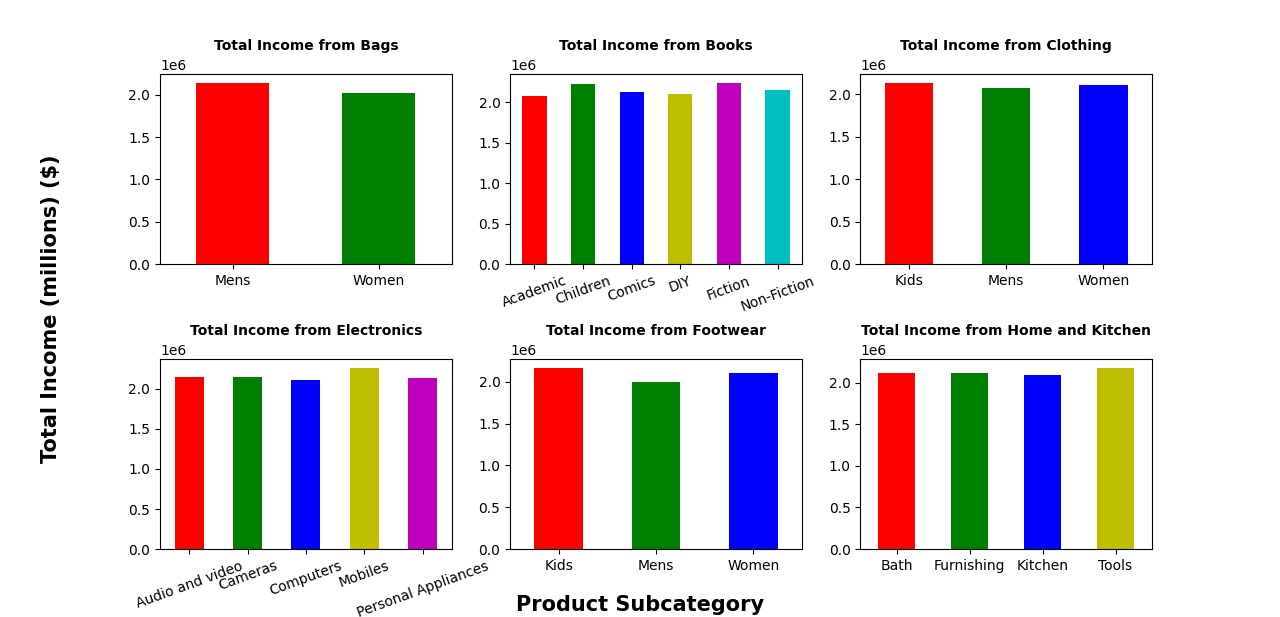


Figure 3: Total Income by Product Subcategories

Looking at Figure 3, as all product subcategories return similar amounts within their own product category, it can mean that the customers are satisfied with the products and there is not much of a product deficiency.

### [3.1.3] Analysis on Product Categories Based on Time

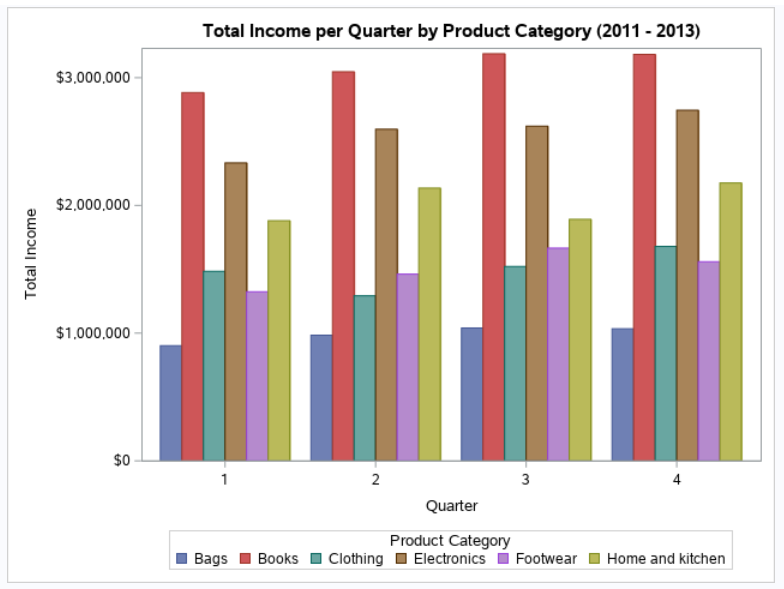


Figure 4: Total Income per Quarter by Product Category

Figure 4 shows a similar pattern to the product category bar graphs above in all four quarters, with Books being the highest followed by Electronics, with both generating over $2 million per quarter. 2014 was excluded due to transactions only taking place until February. This means that books are the most preferred product by the customers and is the best selling product in Company XYZ.

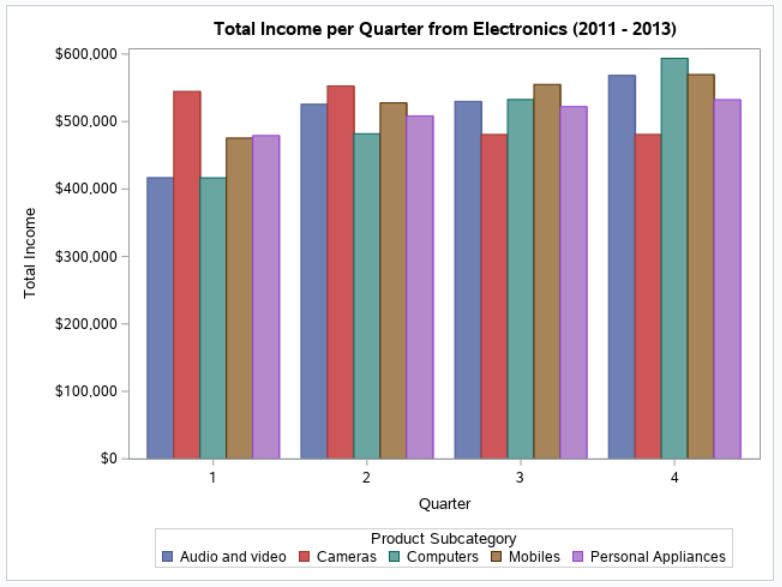
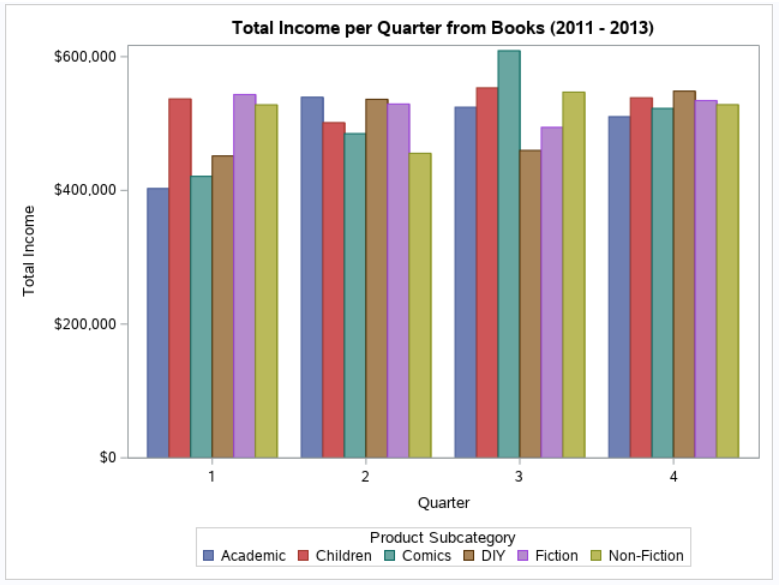


Figure 5: Total Income per Quarter from Books Figure 6: Total Income per Quarter from Electronics

Diving down into the Books category in Figure 5, Fiction generated the most income during the first quarter, while Academic books took over in the second quarter. Comics are seen to have a spike in generating income during the third quarter. In the last quarter, the different subcategories of books return similar amounts of income.

As for the Electronics category in Figure 6, Cameras appear to be the more popular choice during the first two quarters, while Computers and Mobiles return more income during the last two quarters. Personal Appliances are consistently returning around $400,000 to $500,000 throughout the year, while Audio and Video electronics seem to have an uptrend throughout the year together with Computers. All in all, both the graphs show that there is a change in best selling product category in each quarter and this data can be useful when deciding which subcategory to promote more during that certain quarter.

### [3.1.4] Analysis on different Store Types:

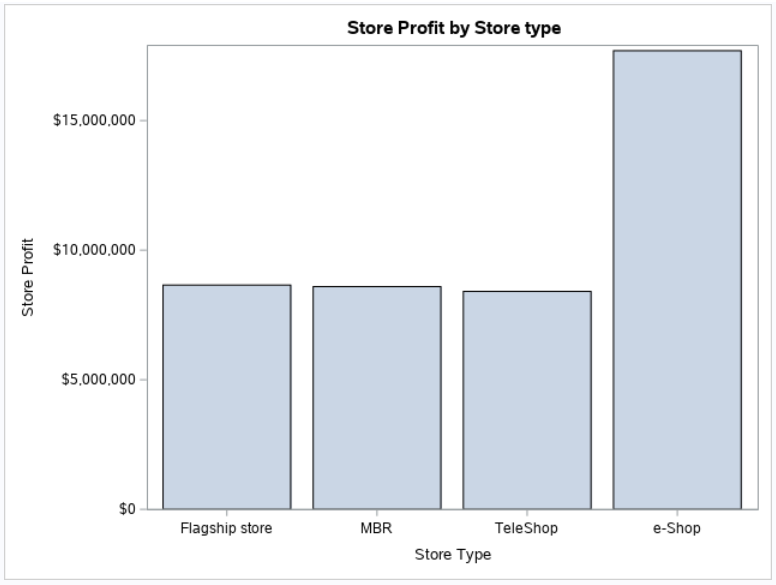
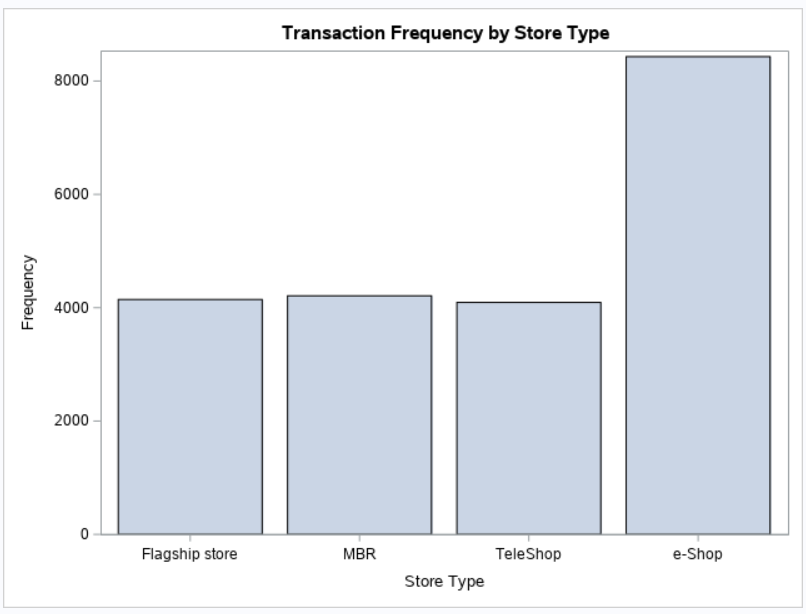


Figure 7: Transaction Frequency by Store Type Figure 8: Store Profit by Store Type

In Figure 7 and Figure 8, the income and frequency of transactions in different store types appear to have a similar trend. Flagship store, MBR, and TeleShop have about $10 million in income; e-Shop has nearly $20 million in income. An interesting note is that although MBR has a higher frequency of transactions than Flagship store, MBR has a lower profit. It is hypothesised that this difference is caused by return transactions.

### [3.1.5] Analysis on Product Categories within each Store Type

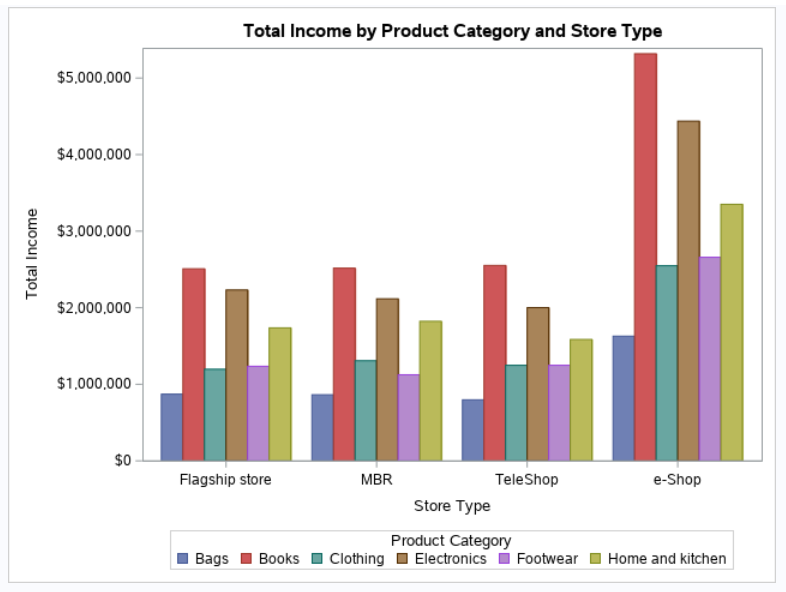


Figure 9: Total Income by Product Category and Store Type

From Figure 9, it can be noted that e-Shops return the most income while the Flagship store, MBR, and Teleshop return similar amounts of income based on the product categories sold. Similar to the income graphs above, it can be seen that Books are the favoured product category followed by Electronics once more. Each product category is also seen to be doing similarly across all stores.

## [3.2] Where/What is allowing them to make a big loss?

Here, we provide two variables to identify losses. First, the negative values in total\_amt in the main dataset are considered as Returns. Second, a ratio of Returns to Income was created as return to income Ratio.

### [3.2.1] Income vs Loss of each Store Type

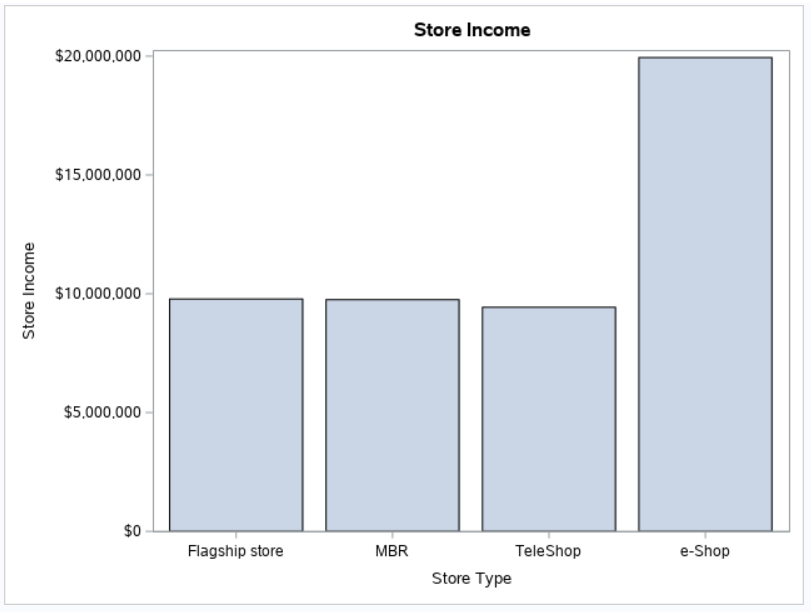


Figure 10: Store Income by Store Type Figure 11: Store Returns by Store Type

The two bar charts above compare the trends between store income and store returns across the different store types. It is obvious that although MBR has a higher return amount than Flagship store, it has a lower amount of income than Flagship store. Furthermore, Flagship store, MBR and Teleshop have a return rate of about 10% of its income, but e-Shop seems to have a slightly higher return rate. Thus, it can be seen that the proportion of store income and store returns are not equal across the stores.

#### 

### [3.2.2] Analysis of Product Category in each Store Type

This section drills down to the return to income ratios of each product category by store type.

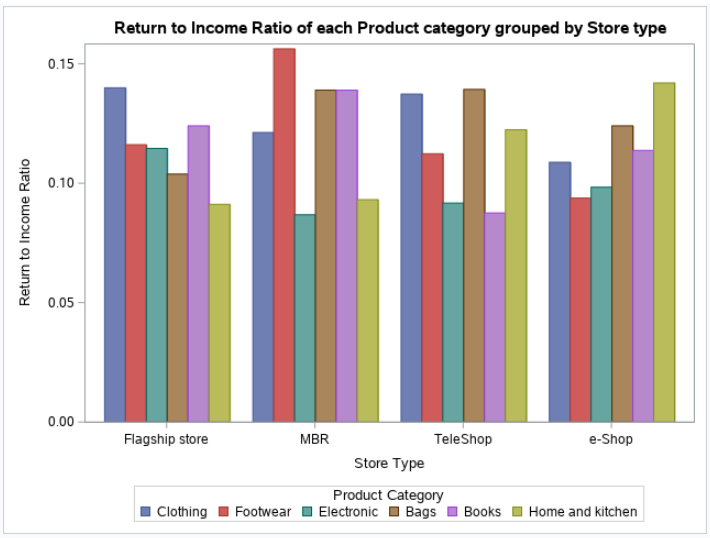


Figure 12: Return to Income Ratio of each Product Category Grouped by Store Type

From Figure 12, we can identify problematic product categories in each store type. The product categories with a high return to income ratio would be causing most loss in the organisation. For example, in MBR stores, Footwear is returned quite often relative to its income generated. At the same time, e-shops tend to do very well on Footwear.

### [3.2.3] Analysis on Product Subcategories in each Store Type

Further analysis was done on product subcategories within each store type.

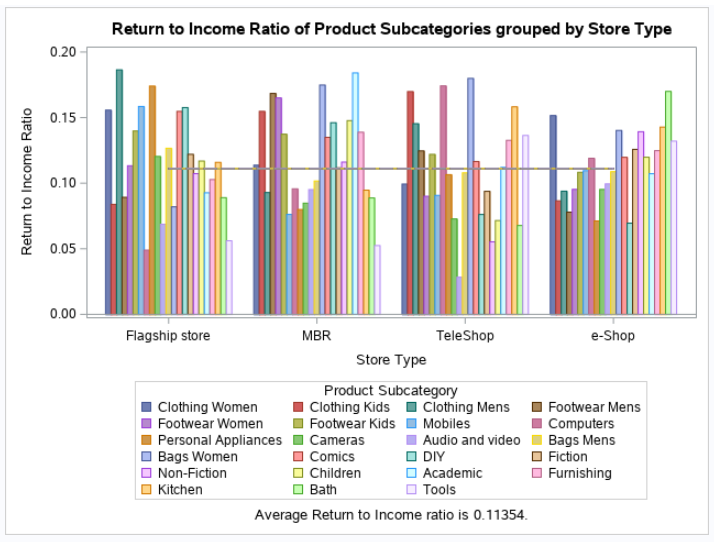


Figure 13: Return to income ratio of Subcategories by Store Type

Figure 13 shows the return to income ratio of subcategories in each store type. The line in the middle is the average return to income ratio in the period. From the figure, we can see that each product subcategory has varying return to income ratio across each store. For example, personal appliances are returned very often in flagship store whereas they are not returned often in e-Shop.

Taking the [average of the income to return ratio](#de4s0o859avh) as a benchmark, it alerts the product subcategories that are problematic with return to income ratio above the average ratio line.

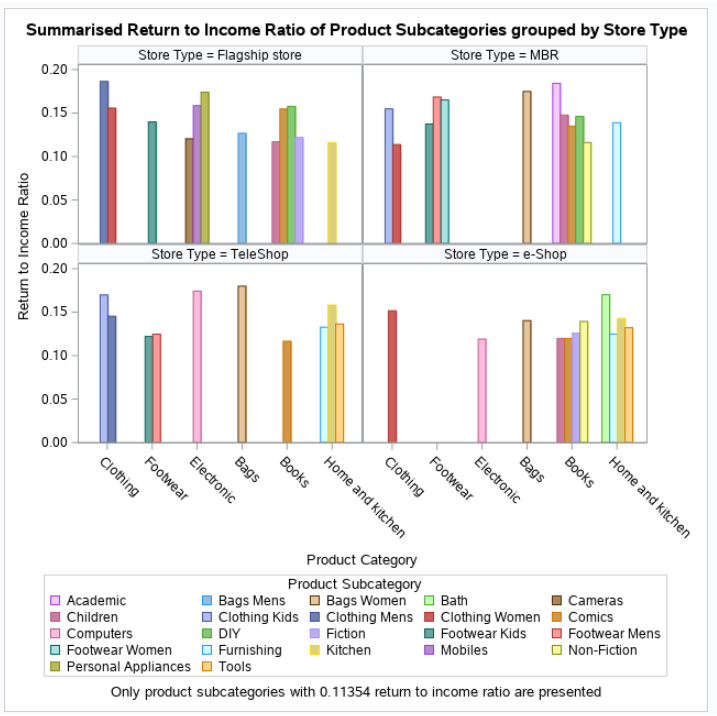


Figure 14: Return to income ratio of Subcategories grouped by Store Type

Figure 14 is a summarised version of figure x showing product subcategories with more than 0.11354 return to income ratio. 45 subcategories are shown to have more than 0.11354 return to income ratio. Out of these 45 subcategories, TeleShop has 10, e-Shop has 11, MBR and Flagship store have 12. Home and Kitchen category has a lower return to income ratio in MBR and Flagship store compared to TeleShop and e-Shop. The subcategory comics has a return to income ratio higher than 0.11354 in all store types. Audio and Visual subcategory has a return to income ratio lower than 0.11354 in all store types.

# 4. TARGET MARKET

This section will mainly cover the market that contributes most to company XYZ over the years.

## [4.1] Which generation is the most popular one?

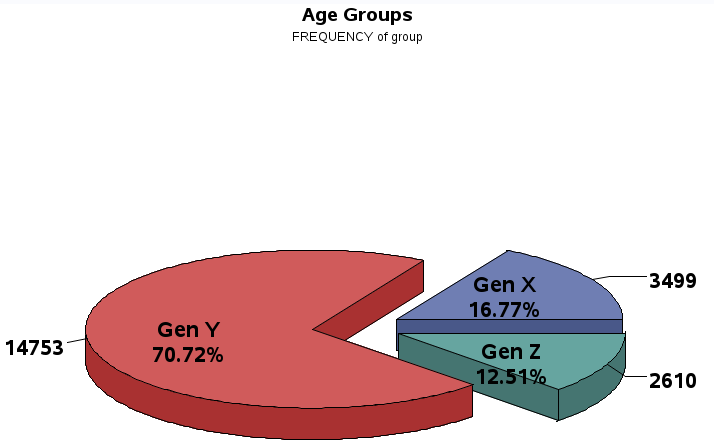


Figure 15: Proportion of Company XYZ’s Customers by Generation

Based on Figure 15, the current market of company XYZ comprises mostly Gen Y, where it is categorized by customers born between the years of 1981 and 1996 as stated in [[2]](#kix.86zjutl45vvw). It can also be found that this generation, Gen Y, has a relatively high purchasing power as found in [[3]](#kix.90eet5ei63r3). Hence, with that, it also explains why Gen Y contributes the most to sales.

## [4.2] In that generation, which gender contributes the most?

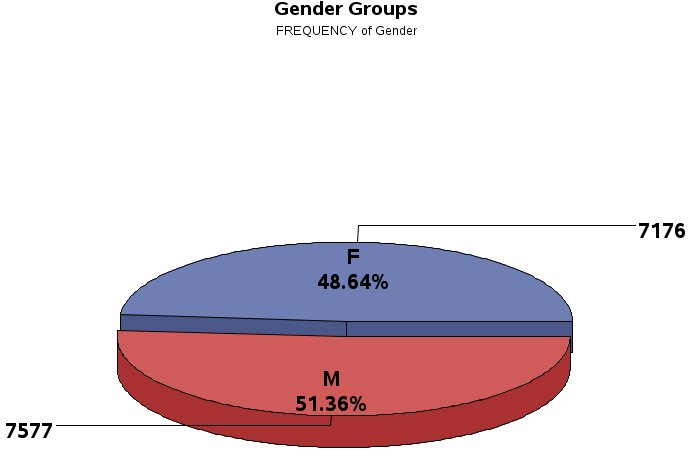


Figure 16: Proportion of Company XYZ’s Customers by Gender

Based on Figure 16, it shows that more males in Gen Y are contributing to company XYZ’s sales over the years as compared to females. After identifying the proportion of gender in Gen Y contributing to company XYZ’s sales, the next identification would be on their location, most preferred store type, product category and product sub-category.

## [4.3] In that gender category, where are they mostly located?

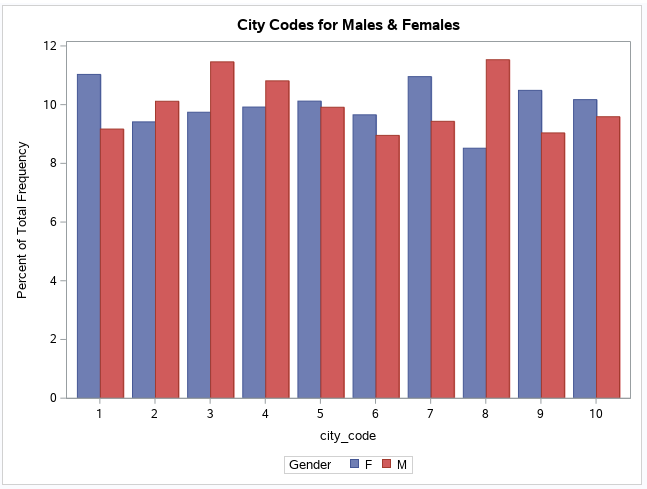


Figure 17: Proportion of Generation Y’s Genders by City Code

Based on Figure 17 it is shown that males from Gen Y are mostly located in city code 8 while females are in city code 1.

## [4.4] In that city code, what stores are available?

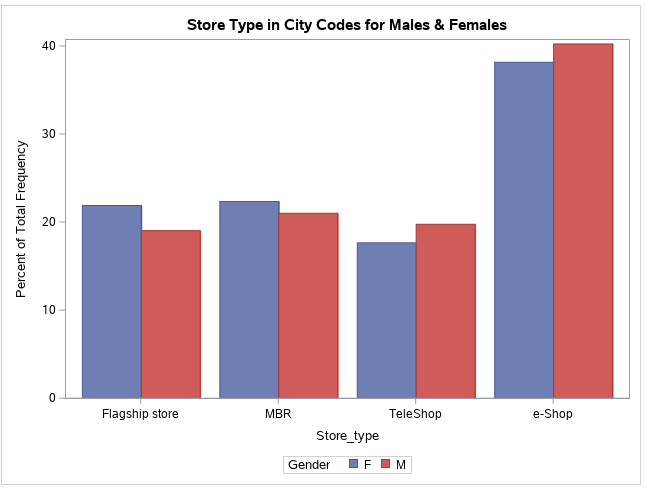


Figure 18: Proportion of Generation Y’s Genders by Store Type

Based on Figure 18, it is shown that males from city code 8 and females from city code 1 both prefer to shop via e-shops. With statistics shown above, it can also mean that store types like flagship stores, MBR and teleshops are less accessible to the target market in their area. Hence, e-shops are mostly chosen by the target market for convenience and easy access purposes.

## [4.5] In that store, what is the most preferred product category bought?

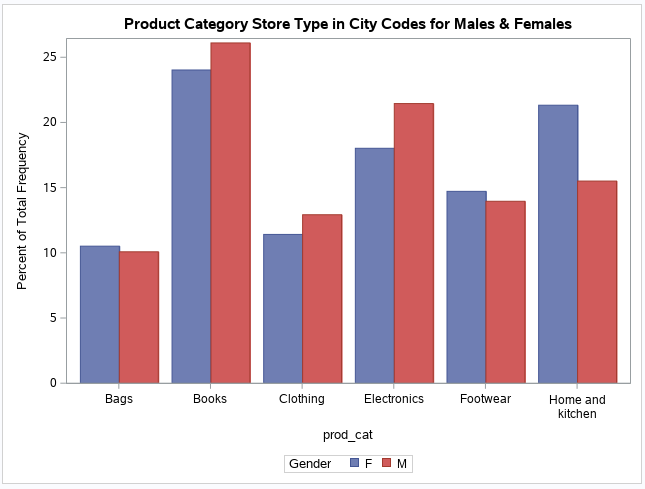
****

Figure 19: Proportion of Gen Y’s Genders by Product Categories in e-Shops

Based on Figure 19, it can now be identified that males and females frequently bought product categories from e-Shops are books. This can mean that both genders are more into books than other product categories.

## [4.6] In that product category, what is the most preferred product sub-category bought?

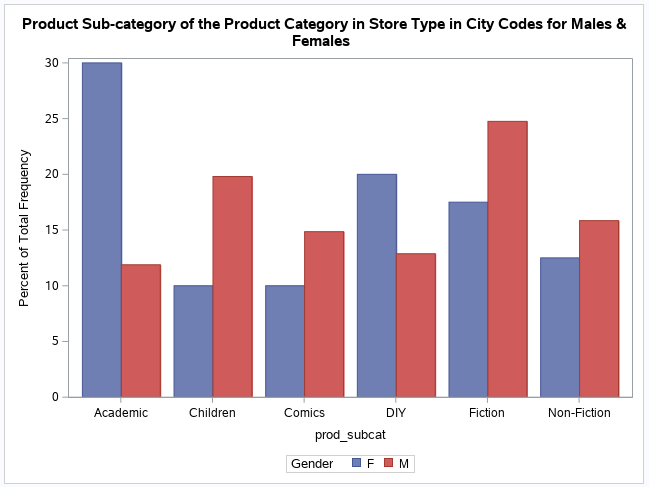


Figure 20: Proportion of Gen Y’s Genders by Product Subcategories for Books in e-Shops

Based on Figure 20, it is proven that even though company XYZ’s target market gender both prefer books, males and females both prefer fiction and academic books respectively over other books subcategories due to the significant difference when doing the comparison with one another.

# 5. SUGGESTIONS

After conducting a series of analyses on company XYZ’s business progress in terms of their income, loss, most valued customer group and most preferred product, it is found that **books** are the most successful product category that has consistently generated the **highest amount of income** compared to any other product category throughout the years, and in every store type. Looking at loyal customers from the Target Market section, **Males** living in **City Code 8** prefer **Fiction books**, which are the highest selling book subcategory in the **first quarter** according to [Figure 5](#l5nm36ysftjp). On the other hand, **Females** living in **City Code 1** prefer **Academic books** which are the highest selling book subcategory in the **second quarter**. Additionally, the loyal customers’ preferred store is **e-Shops**, which is also the store type that sold the highest number of books.

With the understanding of company XYZ’s current progress, we can now identify that the marketing department is progressing well on the promotion of fiction books in e-shops to males in city code 8 and academic books in e-shops in city code 1. In order for future growth, company XYZ can start with focusing on **online seasonal promotions** in their e-Shops.

For example, a **spring discount event** can be held in the first quarter of the year that comes with discounted prices on Fiction books, and a **summer discount event** in the second quarter with discounted Academic books, with both personalized events catering towards the target market which would lead to an increase of sales. Discounts can also be added to other products that are doing well in each quarter. An example would be to also have cameras in the spring and summer discount events as it was the highest selling Electronics product in the first and second quarters according to [Figure 6](#l5nm36ysftjp). An additional suggestion would be to include **e-vouchers** to people who purchase discounted books that can be used to incentivise customers to purchase more products during the discount events [[4]](#kix.458loqsawce2).

Next, company XYZ can invest in **getting ambassadors** to help increase the outreach of their products to many more customers by having the ambassadors **showcase** their most preferred product on their social media. A unique **promo code** can be given to them that allows people to obtain discounts on products when used while shopping in company XYZ. Additionally, the more the code is used, the ambassador will get more commission which incentivises them to continuously promote company XYZ’s products. With this, they will be able to increase sales through word-of-mouth and personal incentives provided to motivate their customers.

Other than that, by identifying the number of returns, company XYZ can now look into their supply chain management department to ensure that there will be **less product deficiency** and have their marketing department to **not over-promise their customers** so that there will be less number of products returned and better reviews provided on a certain product. Additionally, company XYZ should **review the subcategories** that have **alarming return rates** (refer to [Figure 14](#rjf7levvi268)). The high differences between interstore categories imply that it may be better for company XYZ to focus on different products depending on the store type. For example, the home and kitchen category has a high return rate overall in e-Shop and teleShop compared to MBR and Flagship store. Thus, company XYZ should **focus on improving product quality of the home and kitchen category in MBR and Flagship stores**. An alternative solution would be focusing on non-traditional retail by **adopting a customer centered strategy** [[5]](#5m6cjm914ky8) to reduce return rates. Company XYZ can work on **narrowing down the problematic factors** in customer engagement and interactive experiences in the different store types so that customers would provide more feedback and insights on returned items.

Based on [Appendix IV](#79grtitwp145), we can observe that overall, company XYZ has earned $5,317,564.98 of income from solely just selling 5,999 books in e-Shops alone throughout the years of 2011 to 2014 February with 9.55% of Gen Y contributing to the sales of books in e-shops. If company XYZ successfully adopts the suggestions provided above on ways to further enhance the sales on the target customer group with their most preferred product category, it is estimated that the proportion of purchased books from e-Shops can **increase by 10%** and result in earning a total of **$5,317,564.98** with a total of **6,598** number of books sold in **e-Shops** in the following year, **2015**  to first breakeven. Then in the following year, 2016, it is estimated to **increase another extra 5%** so that they will be able to gain more profit. With all these suggestions provided, company XYZ will be able to enhance the business’ portfolio and ultimately improve their businesses. With a strategic business portfolio, a company can then only perform well and can increase financial returns for the company [[6]](#6s280s1g37om).

# 6. CONCLUSION

In conclusion, with the analyses performed on company XYZ, we are able to identify its sources of income and losses as well as their most valued customer target group.

Company XYZ’s target/largest market was found to be people from Generation Y who live in city code 8. Most of their transactions are in the Books category and took place in e-Shops, with the highest proportion of those transactions being Academic books for females, and Fiction books for males.

A few suggestions were proposed to improve the efficiency of company XYZ. First, a marketing strategy such as seasonal promotions for books that caters to a certain demographic should be employed to improve sales in e-Shop. Second, another marketing strategy that targets getting ambassadors to showcase the products and allow other customers to use a unique promo code from the ambassadors to purchase at a lower price. Third, company XYZ should investigate the factors that cause high return rates on problematic products in [Figure 14](#rjf7levvi268).

With the proposed way forward for company XYZ to have better marketing strategies and supply chain management, it can help increase their income and reduce losses for the year 2015 by adopting effective marketing strategies on the right target market for their best selling product categories and subcategories.

# 7. REFERENCES

[1] V. Kotu and B. Deshpande, “Data Science: Concepts and Practice,” 2nd ed, p. 40, 2019. [Online]. Available: <http://asolanki.co.in/wp-content/uploads/2019/04/Data-Science-Concepts-and-Practice-2nd-Edition-3.pdf> [Accessed June 20, 2021].

[2] Kasasa, “Boomers, Gen X, Gen Y, and Gen Z Explained,” Jan.13, 2021. [Online]. Available: <https://www.kasasa.com/articles/generations/gen-x-gen-y-gen-z>. [Accessed June 20, 2021].

[3] G. Ordun, “Millennial (Gen Y) Consumer Behavior, Their Shopping Preferences and Perceptual Maps Associated With Brand Loyalty,” *Canadian Social Science*, vol. 11, no. 4, pp. 40-55, April 2015. [Online]. Available: <https://www.researchgate.net/publication/298506301_Millennial_Gen_Y_Consumer_Behavior_Their_Shopping_Preferences_and_Perceptual_Maps_Associated_With_Brand_Loyalty> [Accessed June 28, 2021].

[4] X.T. Li, “Could Deal Promotion Improve Merchants’ Online Reputations?

The Moderating Role of Prior Reviews,” *Journal of Management Information Systems*, vol. 33, no. 1, pp. 171-201,June 2016. [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/07421222.2016.1172450> [Accessed June 27, 2021].

[5] M. Childs, T. Blanchflower, S. Hur, D. Matthews, “Non-traditional marketplaces in the retail apocalypse: investigating consumers' buying behaviours,”  
International Journal of Retail & Distribution Management, vol. 48, no. 3, 2020. [Online]. Available: <https://www-emerald-com.ezproxy.sunway.edu.my/insight/content/doi/10.1108/IJRDM-03-2019-0079/full/html#sec006> [Accessed June 27, 2021].

[6] S.J. Lin and J.R. Lee, “Configuring a Corporate Venturing Portfolio to Create Growth Value: Within-portfolio Diversity and Strategic Linkage,” *Journal of Business Venturing*, vol. 26, pp. 262-286, Dec 2009.[Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0883902609001074> [Accessed June 27, 2021].

# 8. APPENDIX

## Appendix I

**Further information on Datasets of company XYZ**

Company XYZ has three datasets, Customer, Product, and Transactions which contain data based on what they are named.

The Customer dataset contains 4 columns, namely customer\_Id which contains unique IDs for each customer, DOB which contains the customers’ date of birth, Gender which notes the customers’ gender, and city\_code which contains the code of the city where each customer lives in. Gender had only two values which are either ‘M’ for male or ‘F’ for female, while city\_code only had 10 unique city codes. It was noted that some values of Gender and city\_code were missing, and the rows containing the missing values were deleted accordingly.

The Product dataset contains 4 columns, which are prod\_cat\_code which contains the product category code, prod\_cat which contains the name of the product category, prod\_sub\_cat\_code which contains the product subcategory code, and prod\_subcat which contains the name of the product subcategory. There are only 6 product categories, each with a different number of product subcategories, with some product subcategories sharing the same code despite representing a different product category.

The Transactions dataset contains 10 columns, which are transaction\_id which contains unique IDs for each transaction, cust\_id which references customer\_Id from the Customer table, tran\_date which contains the date each transaction took place, prod\_subcat\_code which references prod\_sub\_cat\_code from the Product table, prod\_cat\_code which references the same column in the Product table, Qty which contains the quantity of products bought or returned, Rate which contains the price of each product, Tax which is a fixed percentage of the next column (about 10.5%) which is total\_amt which contains the total amount paid or returned in each transaction, and lastly Store\_type which contains the type of stores where the transactions took place, which could be in a Flagship store, MBR, TeleShop or an e-Shop.

## Appendix II

**Additional Graphs for Data Exploration**

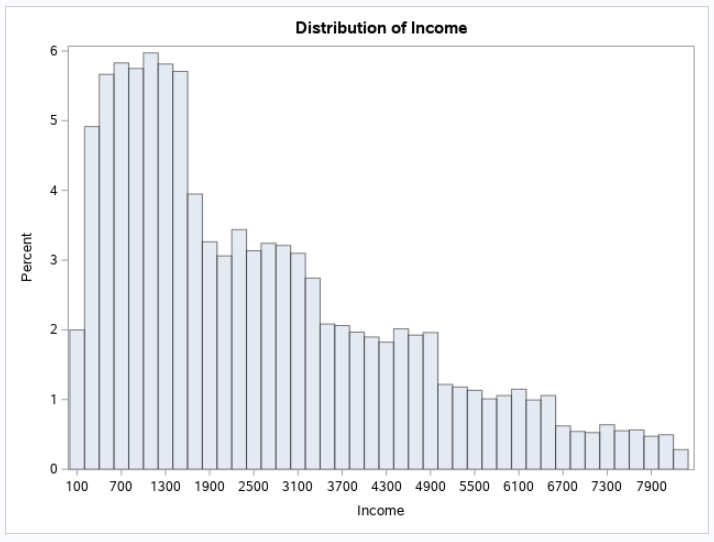
****

Figure 21: Distribution of Income

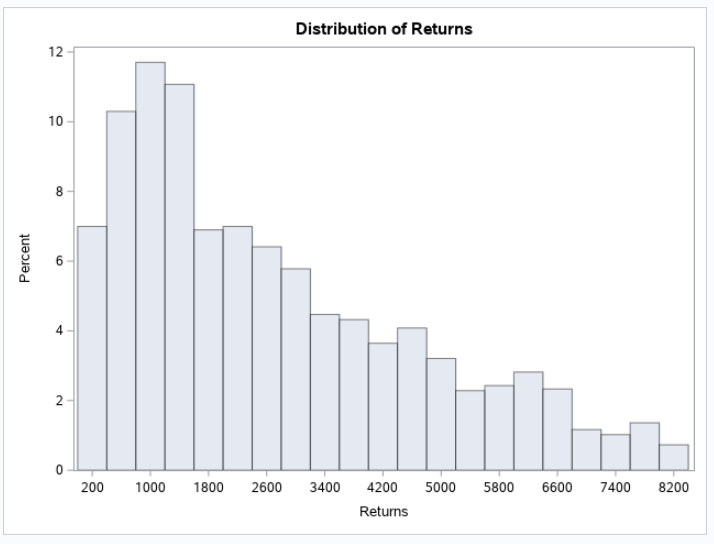
****

Figure 22: Distribution of Returns

## Appendix III

Additional Graph for Returns

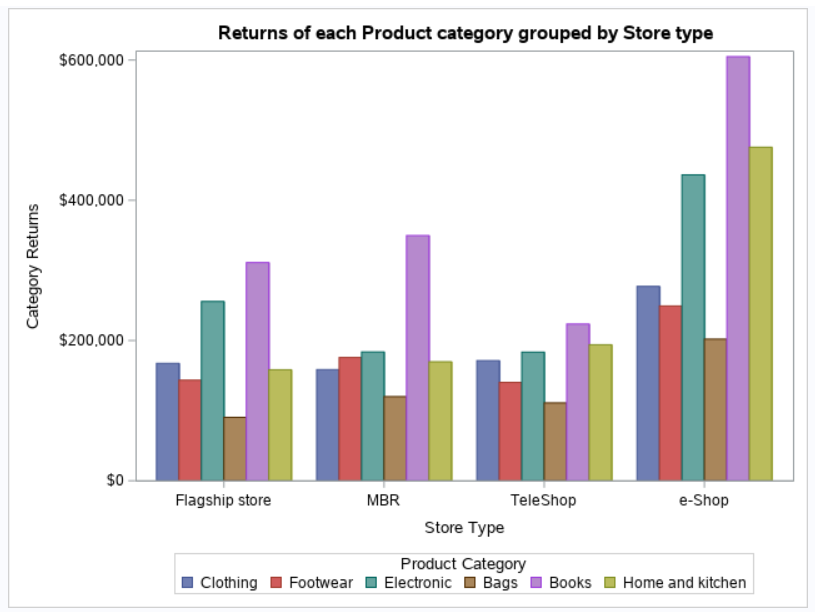


Figure 23: Returns of each product category grouped by store type

## Appendix IV

**Additional Tables for Suggestions**

Table 1: Average Price per Product in each Product Category within each Store Type

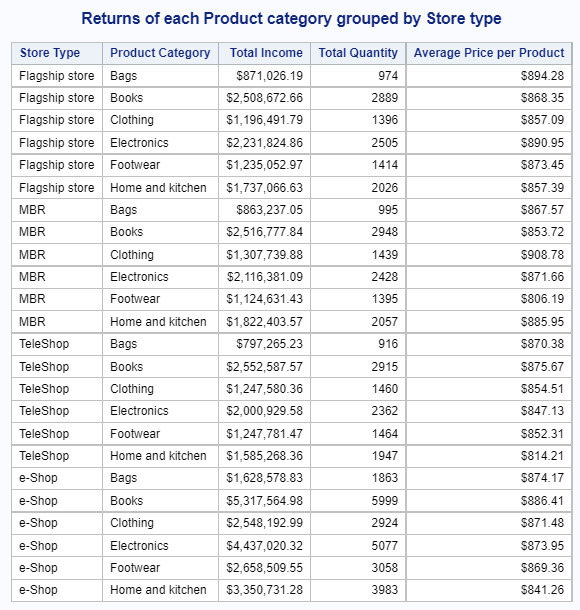


Table 2: Assumptions and predictions for growth in Company XYZ in the following year.

| Assumption:  Every other factor contributing to the income gained other than the proportion of Gen Y purchasing books from e-Shops is constant | | | |
| --- | --- | --- | --- |
|  |  |  |  |
|  | Year | % Income Increase | Income ($) |
| Income | 2014 (Feb) | 1.00 | 5,317,564.98 |
| 2015 | 1.10\* | 5,849,321.48 |
| 2016 | 1.15 | 6,726,719.70 |
|  |  |  |  |
|  | Year | % Amount Increase | Amount (#) |
| # of books [Overall] | 2014 (Feb) | 1.00 | 5 999 |
| 2015 | 1.10\* | 6 598 |
| 2016 | 1.15 | 7 587 |

\*Estimated increase in book sales from Gen Y through e-Shops in 2015.

**Context for Table 2:**

The total income accumulated from 2011 to 2014 for sales generated from books bought by Gen Y through e-Shops is $5,317,564.98 and we estimate to increase the book sales from Gen Y in 2015 by 10%, which will amount to $5,849,321.48. The estimation for the percentage of income increase can be proven in [Table 3](#5safxkgwnwfc), where it shows the total income for each year. From 2011 to 2012, there is an increase of 11.07% while from 2012 to 2013, there is a decrease of 9.02%. With the suggestions stated above, we hope to at least have company XYZ to breakeven in 2015 and then grow another 5% in 2016.

The same estimation of 10% increment is used to estimate the percentage of total number of books sold by 2015 from Gen Y through e-Shops.

Table 3: Income from Books bought by Gen Y according to Year

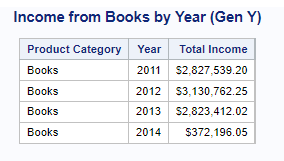


Table 4: Total Income and Returns

