

**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\*: GOH YIK HONG      STUDENT ID: 18004150**

**STUDENT NAME: CHAN WEI WEI      STUDENT ID: 16052748**

**STUDENT NAME: CHAN WEI CHEE      STUDENT ID: 16052755**

**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 Chan Wei Wei 16052748, Chan Wei Chee 16052755, Goh Yik Hong 18004150 received the assignment and read the comments ..... *Wei Chee Yik*...28/6/21..... (Signature/date)

**Academic Honesty Acknowledgement**

"We Chan Wei Wei, Chan Wei Chee, Goh Yik Hong 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." ..... *Wei Chee Yik*.....28/6/21.....(Student's signature / Date)

**IST3144 Problem Solving using Analytics Tools  
Group Assignment**

**Report Title: Analysing Retail Data to Improve Sales**

**STUDENT NAME\*: GOH YIK HONG**  
**STUDENT NAME: CHAN WEI WEI**  
**STUDENT NAME: CHAN WEI CHEE**

**STUDENT ID: 18004150**  
**STUDENT ID: 16052748**  
**STUDENT ID: 16052755**

**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:					

# ANALYSING RETAIL DATA TO IMPROVE SALES

Chan Wei Wei  
16052748

Chan Wei Chee  
16052755

Goh Yik Hong  
18004150

## INTRODUCTION

For retailer organizations, data can be a game-changer that enables them to stay ahead of their competitors. Retailers who use data analytics are able to drive competitive advantage by gathering, analyzing their customer experience and motivations and analyzing market trends to effectively market their products. In this scenario, retail customer data is valuable as it can provide insights into the behavior tendencies of customers. Thus, the retail dataset that consists of the day-to-day transaction history of different store types from 2011-2014 is used in this study. The purpose of our analysis is to help the retail organization in visualizing the sales performances over the years and better understand their customer needs. The insights provided in the analysis allows the retail organization to strategize and focus on important business assets as well as maximize their sales profits.

The main research question of this research is stated below.

**Main Research Question:** How to increase sales performance for each store type?

The main research question is supplemented with Sub-research questions that aid in answering the main research question.

**Sub Research Question:**

1. How is each store type performing over the four years?
2. Which customer segment should be prioritized in each store type?
3. Which product category should be prioritized in each store type?

## A. DATA EXPLORATION

The structure of the dataset can be explained through data exploration procedures using 'PROC CONTENTS' in which the results show that there are 5,647 observations and 4 variables (DOB,city\_code,customer\_id,gender) in the 'Customer' dataset ; 23 observations with 4 variables (prod\_cat, prod\_cat\_code, prod\_subcat and prod\_subcat\_code) in the 'Product' dataset ; and 23,053 observations with 10 variables (customer\_id, prod\_cat\_code, prod\_subcat\_code, qty ,rate, store\_type, tax, total\_amt ,tran\_date , transaction\_id) in the 'Transaction' dataset.

## B. DATA VALIDATION

Data validation steps such as checking for missing values was carried out across all 3 datasets using 'PROC FREQ' to ensure accuracy and quality of analysis. It is found that 'gender' and 'city code' variables contain 2 missing values.

## C. DATA CLEANING AND MANIPULATION

The 3 dataset has been merged to carry out further analysis. Firstly, the 'customer' dataset is merged with the 'transactions' dataset using a similar variable which is 'customer\_id' to produce the 'customer\_trans' dataset. Then, the 'customer\_trans' dataset is merged with the 'products' dataset. The final merged dataset named 'customer\_trans\_product\_clean' with removed missing values is used for further analysis.

After exploring the basic structures of all the datasets, data cleaning was carried out to enhance data quality. It is noticed that the "rate", "qty" and "total\_amt" variables in the 'transaction' dataset contains negative values which indicates it is a refunded transaction. Hence, a total of 4234 observations ( including both original transaction and refund transaction) were being filtered out and output into a temporary dataset called 'refund\_trans'. The remaining 18804 observations in the 'clean\_trans' dataset contain only unique transactions without refund.

## D. GRAPH EXPLORATION

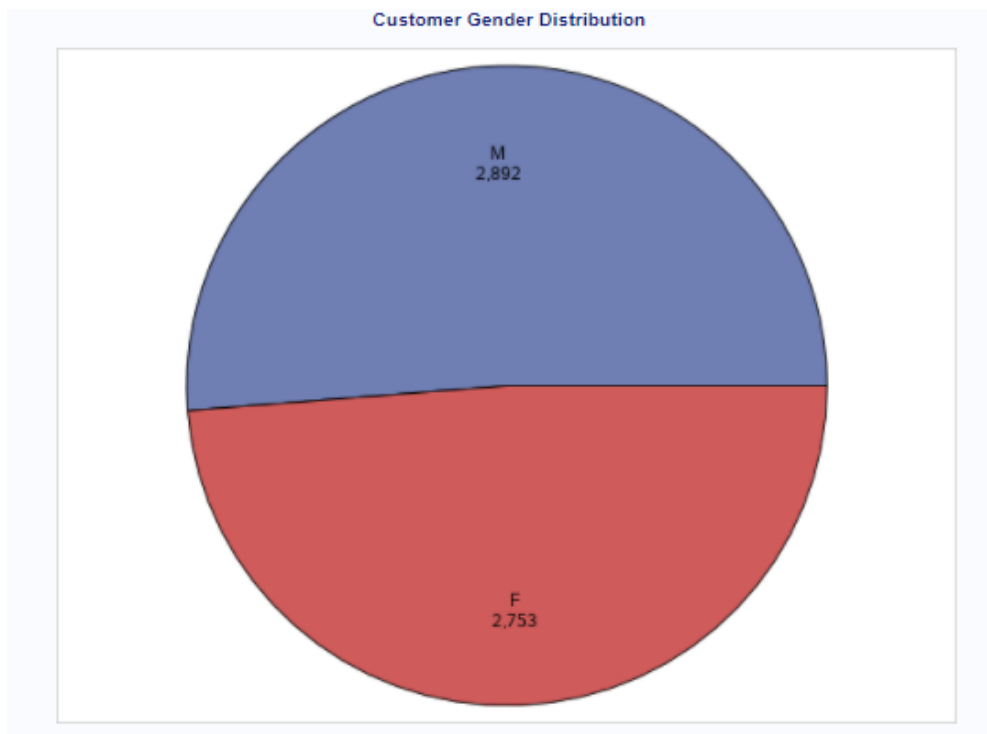


Figure 1 Pie Chart of Customer Gender Distribution in the transactions

The pie chart shown in Figure 1 depicts the customer gender distribution. The customer distribution shows that females with frequency of 2,753 and Male with frequency of 2,892 results in almost equal distribution.



Figure 2 Bar Chart of City Code Distribution where Customers Resides

Figure 2 shows that City Code 3 has the highest frequency of where our customers are based, followed by City Code 4, City Code 5 and City Code 7. City code 6 has the lowest frequency of where customer reside.

Table 1 Descriptive Statistics of Customer Spending

Descriptive Statistics of Customer Spending Based On Transactions History from 2011-2014						
The MEANS Procedure						
Analysis Variable : total_amt						
N	Mean	Median	Mode	Std Dev	Minimum	Maximum
18804	2598.80	2079.61	3049.80	1958.39	77.35	8287.50

Table 1 shows the statistics of individual customer spending. It can be seen that the average amount a customer spends is \$2598.80. The maximum amount that a customer has spent is \$8287.50 and the lowest is \$77.35 in a transaction.

## E. DATA ANALYSIS

Sub Research Question 1: How to increase sales performance for each store type?

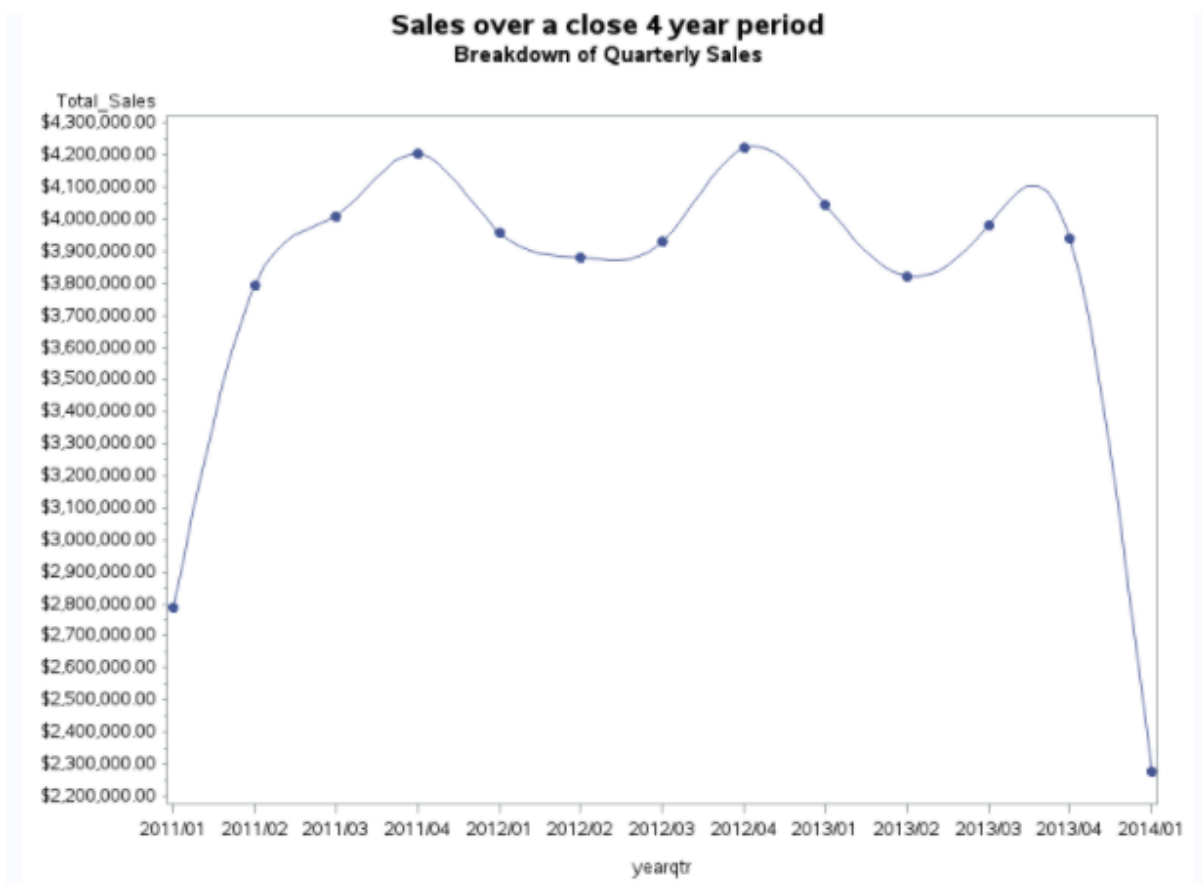


Figure 3 Sales Trends over a close 4 year period

Figure 3 shows the accumulating total sales of all store types (e-shop, Flagship, MBR, Teleshops) by each yearly quarter from 2011 till 2014, a trend can be observed. During the first and second quarter of 2012 and 2013, sales tend to be the lowest. But as the year enters its third quarter, sales begin to rise until eventually it reaches a peak in quarter four.

i. How is each store type performing over the four years?

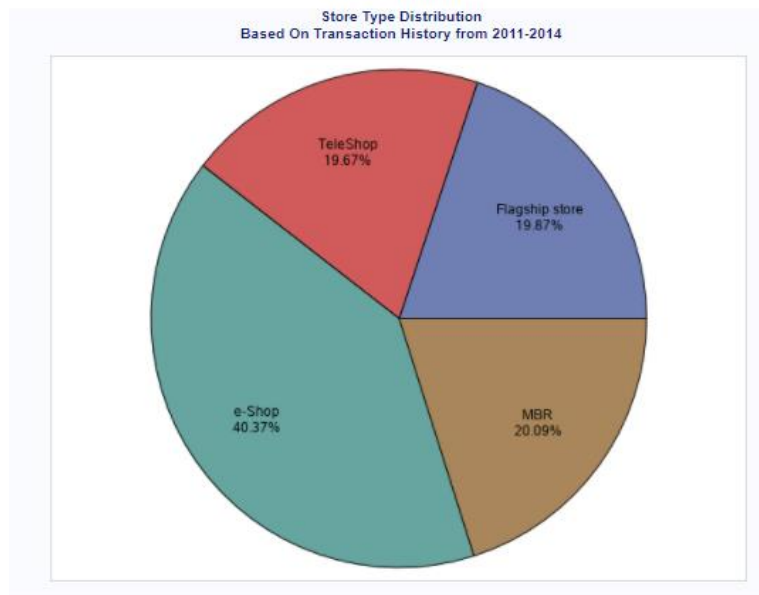


Figure 4 Pie Chart of Store Type Distribution

Figure 4 shows a pie chart illustrating the proportions of transactions among the store types from year 2011 to 2014. It is observed that a large proportion of transactions (40.37%) comes from E-shop. Flagship stores, MBR and Teleshops have very similar transaction proportions (~19%).

Table 2 Accumulating Total Sales by Each Store Type Over the 4 years

Accumulating Total Sales By Each Store Type over the 4 years

store_type	total_sales	transaction_count
Flagship store	\$9,779,248.89	3736
MBR	\$9,743,985.03	3778
TeleShop	\$9,407,669.44	3698
e-Shop	\$19,936,919.41	7592

Table 2 shows the historical sales transactions occurring on each store type. It is observed that E-Shop hold the largest amount of transactions as well as total sales amount. The rest of the shops like Flagship, MBR and Teleshops hold similar transaction amounts (~3700) and similar total sales amounts (~\$9.5 million).

## Sub Research Question 2: Which customer segment should be prioritized in each store type?

### i. Which stores type does each customer age group prefer?

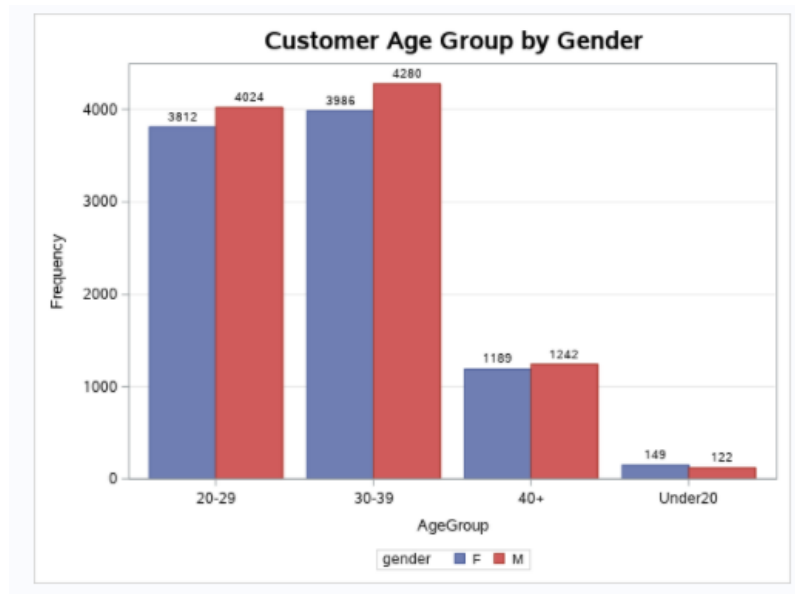


Figure 5 Interleave Bar Chart of Customer Age Group by Gender

The Age variable is computed based on the Date of Birth (DOB) of the customer and year of transaction (tran\_date). Then, the customer is binned into 'AgeGroup' variable accordingly as shown in Figure 5. It shows that gender is equally distributed among each age group. Customers from the age group of 30-39 make up most of the distribution (~8200) followed by the age group of 20-29 (~7800). Customers below the age of 20 have the least portion (~270).

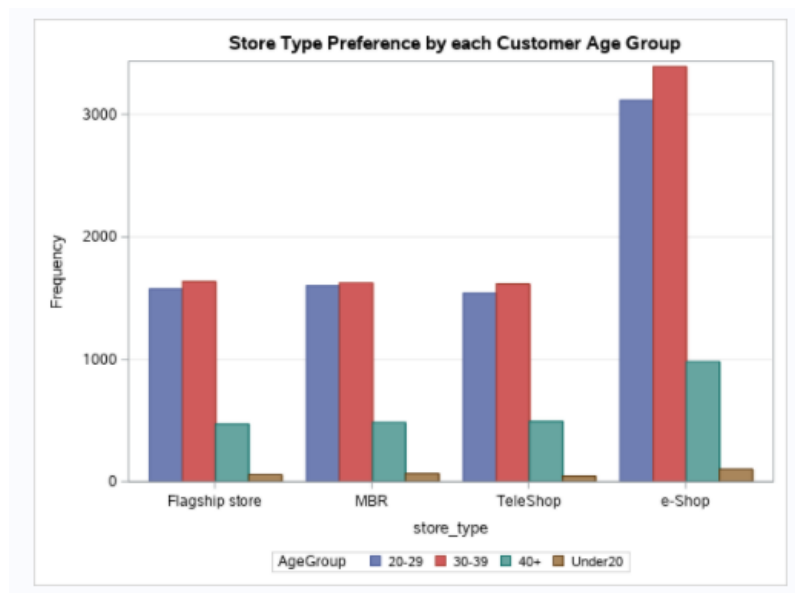


Figure 6 Bar chart of Store Type Preference by each Customer Age Group

The bar chart showing store type preference of each customer age group is plotted as shown in figure 6. It is observed that every age group prefers to shop in E-Shop. MBR, TeleShop and Flagship stores will be their second choice.

## ii. Product Category Preferences by Customer Age Group

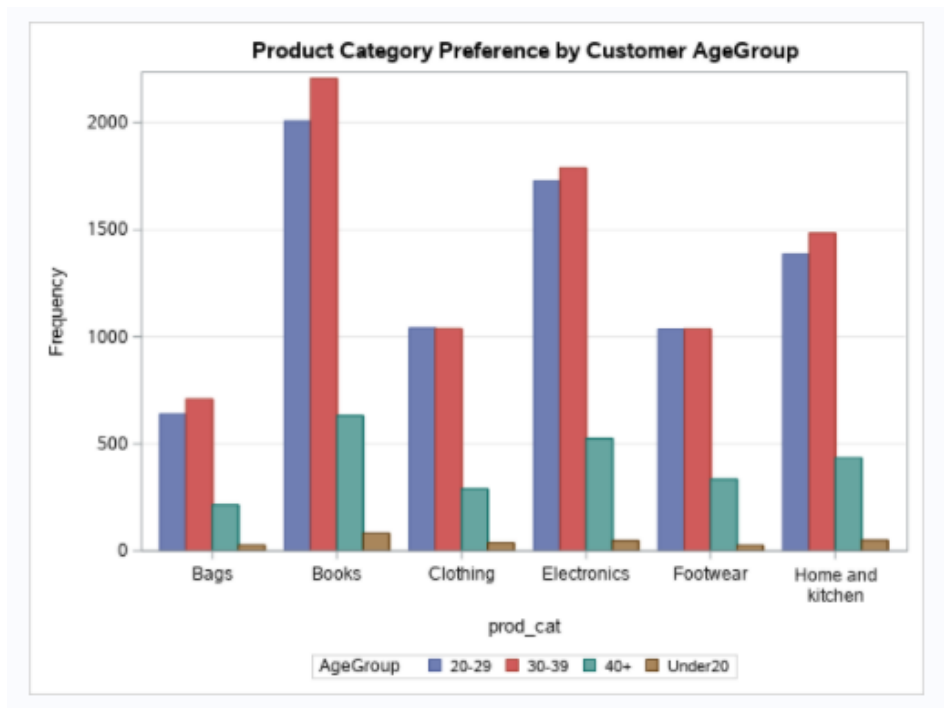


Figure 7 Product Category Preference by Customer AgeGroup

Figure 7 shows a bar chart depicting product category preferences by customer Age group. It can be observed that all customer age groups prefer to buy Books the most, followed by Electronics then Home and Kitchen as compared to clothing, footwear and bags.

## iii. Product category preference by Gender

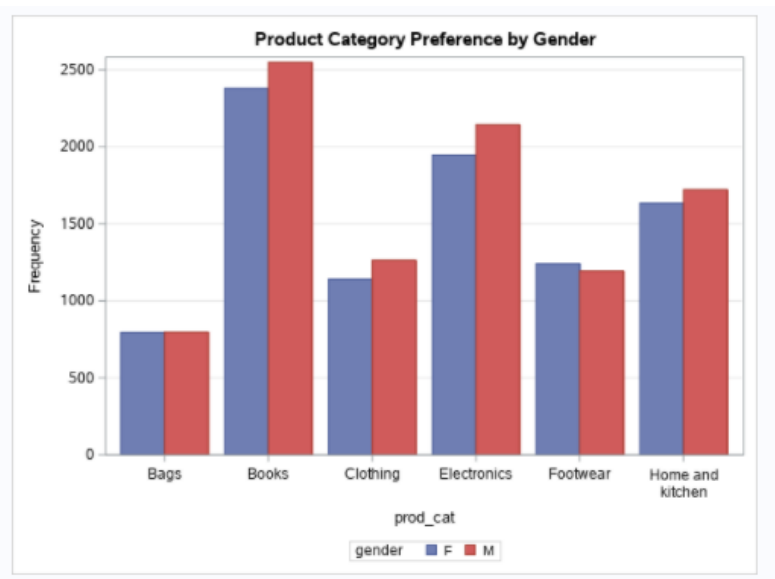


Figure 8 Bar chart of Product Category Preference by Gender

Figure 8 shows that males tend to spend more than females in every product category except footwear. There is a fair amount of difference between the preference of males and females in Books, Electronics, Home and Kitchen and Clothing.



#### iv. Profitable Age Group by each Store Type

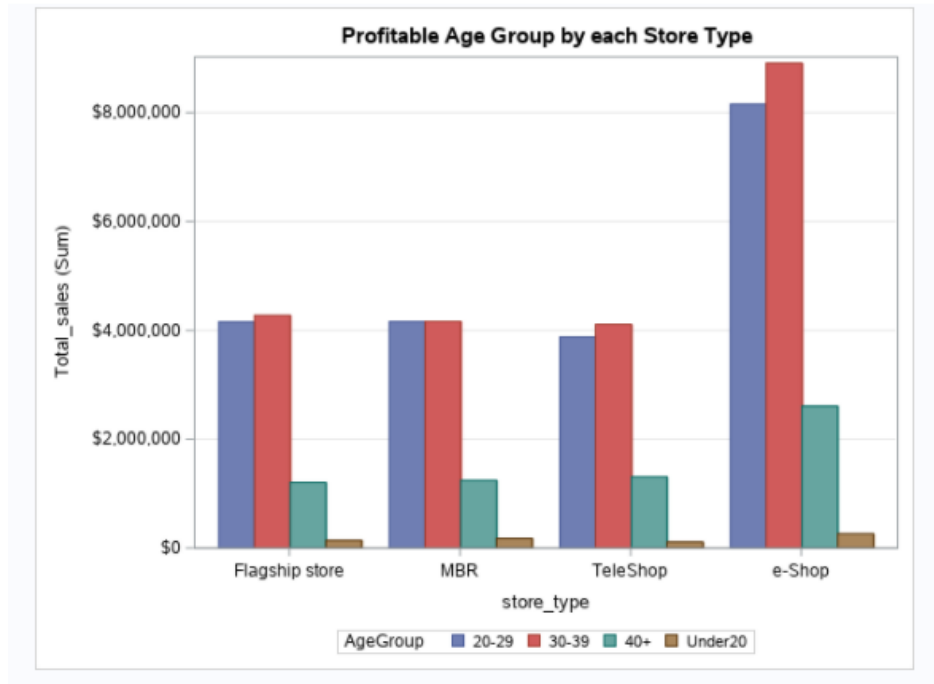


Figure 9 Profitable Age Group by each Store Type

It is observed from Figure 9 that customers from the age group of 30-39 contributed the most in term of total\_sales and frequency count of transactions in each store type with approximately \$8.9 million in E-shop, \$4.2 million in Flagship store, ~\$4.1 million in MBR and Teleshop.

#### v. Analysis of Flagship Store Customer

Table 3 Analysis of Flagship Store Customer

city_code	total_sales	city_count
1	\$916,869.33	365
2	\$964,155.59	350
3	\$1,168,063.46	424
4	\$962,496.99	396
5	\$933,214.49	369
6	\$857,625.86	345
7	\$1,043,405.09	375
8	\$927,297.22	363
9	\$977,730.52	368
10	\$1,028,590.36	381
		3736

Table 3 shows the city code along with the total sales and number of transactions that are generated in flagship store(s) located in each city. It can be observed that flagship store(s) located in City Code 3 produce the highest amount of total sales, \$1.1 million, as well as number of transactions. This is followed by City Code 7 where total sales amassed to \$1.04 million with 375 number of transactions and City Code 10 with \$1.02 million and 381 transactions. The rest of the stores in other cities produced a similar amount of sales (~\$950k) and the city that produced the least amount of sales among the 10 cities is City Code 6 with \$850k.

### RQ3: Which product category should be prioritized in each store type?

#### i. Which product category is the most profitable in each store type?

Table 4 Top 3 profitable product category in e-shop

Top 3 Most Profitable Product Categories in E-Shop		
Obs	prod_cat	TotGross_profit
1	Books	\$4,811,182.00
2	Electronics	\$4,015,403.00
3	Home and kitchen	\$3,030,101.00

As can be seen from the tables 4-7 (refer to table 5-7 in appendix), the top 3 profitable product categories seem to be the same for all shop types. With Books being the most profitable followed by Electronics and Home and Kitchen. Gross profits from E-Shops are the highest among the four different store types due to the high amount of transactions that occur in e-shops (~7592 transactions).

#### ii. Which sub-category of the top 3 profitable product category is profitable?

Table 8 Most Profitable Product Sub-Categories in E-shop

Top 3 Most Profitable Product Sub-Categories in E-shop			
store_type	prod_cat	prod_subcat	tot_gross_profit
e-Shop	Books	Fiction	\$843,985
e-Shop	Books	Children	\$839,635
e-Shop	Books	Academic	\$817,577
e-Shop	Electronics	Personal Appliances	\$869,335
e-Shop	Electronics	Mobiles	\$854,880
e-Shop	Electronics	Audio and video	\$800,119
e-Shop	Home and kitchen	Tools	\$831,590
e-Shop	Home and kitchen	Furnishing	\$744,121
e-Shop	Home and kitchen	Bath	\$727,988

From Table 8 (refer to Table9-11 in appendix) upon a more in depth view into the most profitable product sub categories, it is found that different stores have different popular subcategories of the top 3 profitable product categories. For Books, Fiction and Children's book genres seems to be a very popular option for customers as it was in the top 3 profitable sub-categories for books in three out of the four store types. This is followed by Comics and Non-fiction books which were in the top 3 for two of the four store types.

Moving on to Electronics, Audio and video devices as well as cameras are the most popular and profitable as it appeared in the top 3 profitable sub-categories for electronics in three out of the four store types. Mobiles, Computers and Personal appliances appeared in the top 3 in two out of the four store types.

Lastly, Home and Kitchen. Bath items are the most sought after and profitable, appearing in the top 3 profitable sub-categories for Home and Kitchen in all four store types. This is followed by Furnishing and Tools which was in the top 3 in 3 out of the 4 store types.

### iii. Which product sold the most/least in each store type?

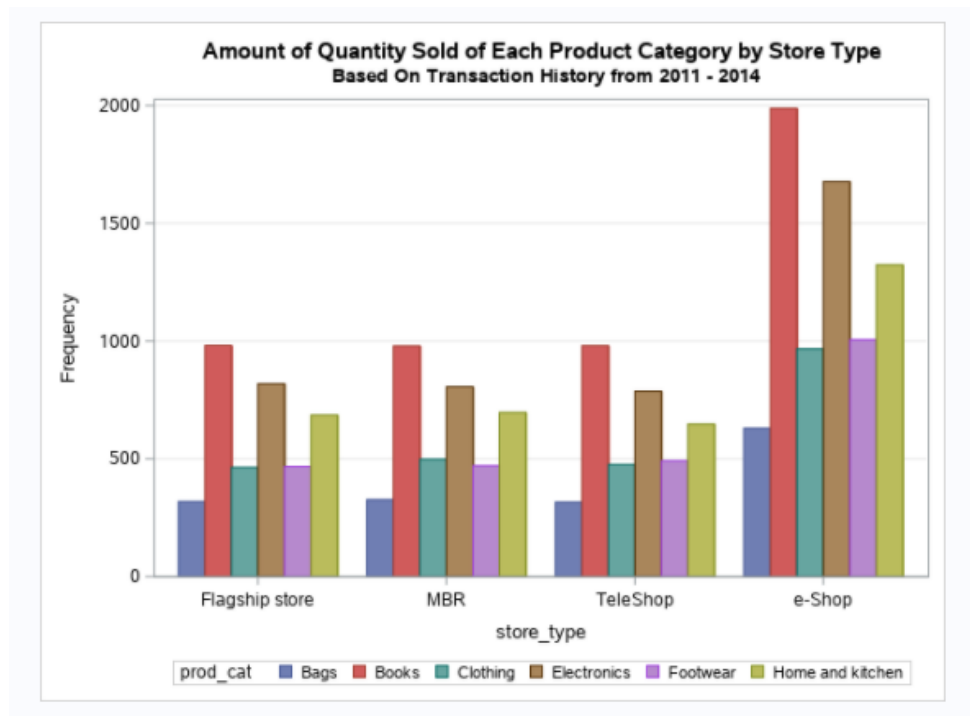


Figure 10 Bar chart of Amount of Quantity Sold of Each Product Category by Store Type

Based on Figure 10, it is observed that Books, Electronics and Home and kitchen product categories are the top 3 most sold categories in all four store types. Whilst Footwear is the fourth most sold category followed by Clothing and Bags.

### iv. Which product is refunded the most in which store type?

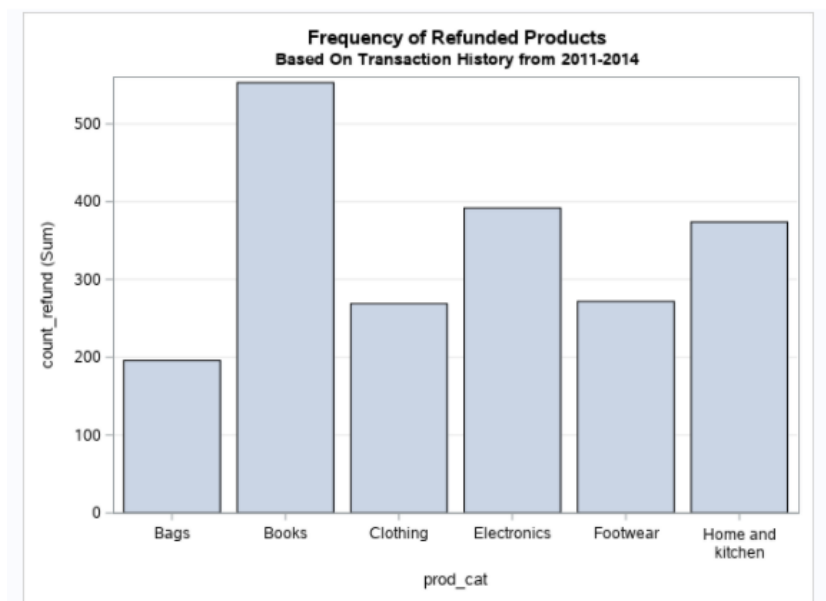


Figure 11 Bar chart of the Frequency of Refunded products

Figure 11 shows the overall refund occurrences of each product category. It could be seen that overall, “Books” have the highest number of refund occurrences with 553 refunds followed by “Electronics” with 392 refunds and “Home and Kitchen” with 374 refunds. Ironically, these three product categories are also the 3 most profitable categories in every store type.

Table 12 Number of Refunds for Each Product Category in E-Shop

store_type	prod_cat	Refund_Freq	total_loss
e-Shop	Books	223	\$604,826.17
e-Shop	Home and kitchen	179	\$475,399.73
e-Shop	Electronics	159	\$436,323.62
e-Shop	Footwear	101	\$249,291.32
e-Shop	Clothing	99	\$275,563.79
e-Shop	Bags	72	\$201,989.58
		833	\$2,243,394.21

Table 12 shows the number of refund occurrences of each product category in E-shop. It could be seen that “Books”, “Electronics” and “Home and Kitchen” are the top 3 most refunded product categories. The same can be said for the other store types (refer to Table13-15 in the appendix). Since refund occurrences of “Clothing” and “Home and Kitchen” in Flagship stores are not much different, we would like to consider both of them taking the third place of the top 3 most refunded products in flagship stores.

## DISCUSSION

### **Performance of the different store types**

It is observed that the retail sales normally hit its peak sales during the fourth quarter of the year, which is between month October to December. One possible logical explanation to this could be due to culture-driven shopping spree and festive seasons, such as Christmas and year-end sales where people would buy lots of things for their friends and family [1]. According to [2], people start shopping online for festive related products five weeks before CNY day. To cater to the high demand, it is highly recommended to order more stocks to ensure that there is sufficient supply to meet seasonal demands.

As discovered earlier, E-shop contains the highest amount of transactions among all the four store types. This could be due to the emerging growth of e-commerce in the 2010's and the benefits provided in terms of convenience and travelling [3]. Hence, retailers should invest more resources in system infrastructure such as website maintenance and make use of cookies to store customer purchasing behaviors and the valuable data can be turned into insights in strategizing advertisement and marketing purposes. The sales of low performing products can be boosted by offering discounts and placing them on the webpage where it is easily viewable by customers. The performance in other store types such as MBR, Flagship store and teleshop is still favorable although the transaction frequency is not as high as compared to the E-shop.

In order to survive the onslaught of rising competitive e-commerce, brick-and-mortar retailers must provide personalized and positive experiences that cannot be experienced in e-shops [4]. With this information, we recommend the retail organization try to create a more personalized service in their physical stores such as Flagship store and MBR and lean further on in-store-only promotions and sales to ensure that customers enjoy a positive experience. This may encourage customers to visit physical stores rather than buying online.

### **Which customer segment should be prioritized in each store type**

Customer segmentation allows the retailer to better understand their customer's needs, marketing efficiency, improve distribution strategies and customer retention. The analysis showed that the main target audience(s) for the retail organization is customers aged between 20-29 to 30-39. Both of these customer segments have generated up to \$41 million of sales (Age group 30-39 \$21mill, 20-29 \$20 mil), which is much higher compared to age groups Under 20 and over 40+. Hence, the retailer should market their products that cater needs of customers that fall in the 20-29 and 30-39 age group.

### ***Which product category should be prioritized in each store type***

As table 4-7 have shown, the most profitable product categories are Books, Electronics and Home and Kitchen. This is true in all of the four different store types. As for the profitable sub-categories, Fiction and Children's books seems to be very popular and profitable within the Books product category, Audio & Video devices, and Camera in the Electronics category and finally Bath, Furnishings and Tools in the Home & kitchen category. With this in mind, the retailer should focus more of their resources on these products. They could also order more stock to ensure there is constant supply to meet demands as these products are bought much more frequently than other products, see figure 10.

Moving on, analysis also has shown that refunded transactions occurred mostly in Books, Electronics and Home & Kitchen. The total combined loss this has caused the retailer is approximately \$13 million. Hence, the retailer must pay more attention to these products and try to minimize the amount of refunds to minimize profit loss.

### **CONCLUSION**

In a nutshell, it can be concluded that E-shop drives the highest sales and generates the most profit. The main target audience(s) will be the customers of the age group 20-29 30-39 that contributed to most of the sales in every store type of the retail shop. Products related to Books, Home and Kitchen and Electronics will be the product categories that need to be prioritized. Overall the performance is quite stable and balanced across each store type as well as product categories while product sub categories do vary across different store types. By getting these insights, the retailers are able to plan a marketing strategy for resource allocation and inventory replenishment more effectively to gain a competitive advantage than its competitors.

## REFERENCES

- [1] Thomas.L, *Retail is having its best holiday season in 6 years* , CNBC, Dec. 2018. Accessed on Jun 27, 2021. [Online]. Available: <https://www.cnn.com/2018/12/26/retail-is-having-its-best-holiday-season-in-6-years.html>
- [2] Tan.J, *How to capture Malaysians' attention during Chinese New Year*, Think with Google, Jan, 2018. Accessed on Jun 27, 2021.[Online]. Available: <https://www.thinkwithgoogle.com/intl/en-apac/marketing-strategies/search/how-to-capture-malaysians-attention-during-chinese-new-year/>
- [3] Kasper, *Importance of E-Commerce and online shopping and why to sell Online*, Medium, Apr 13, 2017. Accessed on Jun 27, 2021. [Online]. Available: <https://medium.com/@nyxonedigital/importance-of-e-commerce-and-online-shopping-and-why-to-sell-online-5a3fd8e6f416>
- [4] Mckamley.C, *Five ways bricks-and-mortar shops can survive the onslaught of e-commerce*, The Conversation, Jun 15, 2018. Accessed on Jun 27, 2021. [Online]. Available: <https://theconversation.com/five-ways-bricks-and-mortar-shops-can-survive-the-onslaught-of-e-commerce-98328>

## APPENDIX

Table 5 Top 3 profitable product category in MBR

Top 3 Most Profitable Product Categories in MBR Stores		
Obs	prod_cat	TotGross_profit
1	Books	\$2,273,431.00
2	Electronics	\$1,915,277.00
3	Home and kitchen	\$1,849,234.00

Table 6 Top 3 profitable product category in Teleshop

Top 3 Most profitable Product Categories in Teleshop		
Obs	prod_cat	TotGross_profit
1	Books	\$2,303,588.00
2	Electronics	\$1,801,820.00
3	Home and kitchen	\$1,428,845.00

Table 7 Top 3 profitable product category in Flagship store

Top 3 Most Profitable Product Categories in Flagship Stores		
Obs	prod_cat	TotGross_profit
1	Books	\$2,270,292.00
2	Electronics	\$2,019,187.00
3	Home and kitchen	\$1,572,006.00

Table 9 Most Profitable Product Sub-Categories in Flagship Stores

Top 3 Most Profitable Product Sub-Categories in Flagship Stores			
store_type	prod_cat	prod_subcat	tot_gross_profit
Flagship store	Books	Children	\$430,070
Flagship store	Books	Non-Fiction	\$413,121
Flagship store	Books	Fiction	\$405,144
Flagship store	Electronics	Mobiles	\$474,820
Flagship store	Electronics	Computers	\$445,241
Flagship store	Electronics	Cameras	\$405,116
Flagship store	Home and kitchen	Bath	\$410,597
Flagship store	Home and kitchen	Kitchen	\$395,511
Flagship store	Home and kitchen	Furnishing	\$384,917

Table 10 Most Profitable Product Sub-Categories in Teleshop

Top 3 Most Profitable Product Sub-Categories in Teleshop			
store_type	prod_cat	prod_subcat	tot_gross_profit
TeleShop	Books	DIY	\$414,071
TeleShop	Books	Children	\$404,256
TeleShop	Books	Comics	\$397,107
TeleShop	Electronics	Cameras	\$402,508
TeleShop	Electronics	Audio and video	\$376,245
TeleShop	Electronics	Personal Appliances	\$360,084
TeleShop	Home and kitchen	Furnishing	\$390,616
TeleShop	Home and kitchen	Bath	\$370,118
TeleShop	Home and kitchen	Tools	\$340,662

Table 11 Most Profitable Product Sub-Categories in MBR Stores

Top 3 Most Profitable Product Sub-Categories in MBR stores			
store_type	prod_cat	prod_subcat	tot_gross_profit
MBR	Books	Comics	\$415,858
MBR	Books	Non-Fiction	\$415,790
MBR	Books	Fiction	\$380,259
MBR	Electronics	Audio and video	\$429,676
MBR	Electronics	Computers	\$388,633
MBR	Electronics	Cameras	\$385,836
MBR	Home and kitchen	Kitchen	\$442,906
MBR	Home and kitchen	Tools	\$413,415
MBR	Home and kitchen	Bath	\$401,829

Table 13 Number of Refunds for Each Product Category in Flagship Store

Number of refunds for each product category in Flagship Store			
store_type	prod_cat	Refund_Freq	total_loss
Flagship store	Books	117	\$311,141.48
Flagship store	Electronics	86	\$255,684.85
Flagship store	Clothing	59	\$167,452.81
Flagship store	Home and kitchen	58	\$158,354.24
Flagship store	Footwear	53	\$143,425.69
Flagship store	Bags	34	\$90,425.47
		407	\$1,126,484.52



Table 14 Number of Refunds for Each Product Category in TeleShop

Number of refunds for each product category in Teleshop			
store_type	prod_cat	Refund_Freq	total_loss
TeleShop	Books	94	\$223,501.72
TeleShop	Electronics	75	\$181,629.96
TeleShop	Home and kitchen	70	\$194,038.00
TeleShop	Footwear	52	\$140,182.51
TeleShop	Clothing	49	\$171,319.20
TeleShop	Bags	46	\$111,034.82
		386	\$1,021,706.21

Table 15 Number of Refunds for Each Product Category in MBR

Number of refunds for each product category in MBR			
store_type	prod_cat	Refund_Freq	total_loss
MBR	Books	119	\$349,729.19
MBR	Electronics	72	\$183,681.94
MBR	Home and kitchen	67	\$169,698.17
MBR	Footwear	66	\$175,744.73
MBR	Clothing	62	\$158,538.77
MBR	Bags	44	\$119,973.16
		430	\$1,157,365.95