MIS304

Information Systems for Business (Advanced)

Assessment 2B – Industry-based report

Students:

Truong Son Vu Nguyen - A00045335

Gilles De Leonardis – A00061017

Dana Mendoza Zuleta – A00056842

Learning Facilitator: Sarah Immanuel

Date of submission: 30/04/2023

Total word count: 2316

Assessment 2 word count: 1152

Table of Contents

Executive Summary

This document discussed the creation of an analysis workbook. The dataset was extracted from the records relative to an e-commerce operating internationally, in the appliances and technology fields, with both retail and trade (offices) customers.

After a brief introduction and justification for selection, the data was cleaned and filtered with the Tableau suite (both Prep and Desktop), to remove errors, discrepancies, and attributes that were not beneficial to the analysis. This way, its quality was improved to offer a solid base for the following steps.

The team then discussed some goals for the analysis, that were identified as profitability growth, customer segmentation, and logistics optimisation. Thus, thanks to the creation of dedicated queries and tables, it was possible to create hypothesis and offer relevant recommendations, to improve the overall profitability and efficiency of the business.

Finally, the document presented the procedure and its results in a storytelling fashion, offering satisfactory, yet concise insights about the business.

1 Introduction

As per the instructions from the Global Superstore's board of directors, the purpose of this report is to offer a preliminary analysis and suggest ways to interpret a dataset about general sales worldwide. The document will address three main aims: improving profitability, optimising logistics, and enhancing customer segmentation. Through analysis, the data will be cleaned and explored within the defined timeframe, identifying patterns, and extracting relevant information. Furthermore, suggestions will be provided to present the findings in a clear, concise, and graphical manner, to support decision-making within the organization. This will lead to specific observations and recommendations presented with a storytelling approach to follow the process and facilitate the understanding by the management team.

1 Dataset overview and the reason for selection

1.1 Data overview

The Global Super Store dataset, collected by Mahalingappa (2020), provides order details of a company that sold various products worldwide on its e-commerce platform from 01/01/2011 to 31/12/2014. The dataset consists of 51,290 rows and 24 columns with the following information:

Columns	**Data types**		**Columns**	**Data types**
Row ID	Numerical		Market	Categorial
Order ID	Categorial		Region	Categorial
Order Date	Categorial		Product ID	Categorial
Ship Date	Categorial		Category	Categorial
Ship Mode	Categorial		Sub-Category	Categorial
Customer ID	Categorial		Product Name	Categorial
Customer Name	Categorial		Sales	Numerical
Segment	Categorial		Quantity	Numerical
City	Categorial		Discount	Numerical
State	Categorial		Profit	Numerical
Country	Categorial		Shipping Cost	Numerical
Postal Code	Categorial		Order Priority	Categorial

This dataset can be accessed via this link:

https://www.kaggle.com/datasets/apoorvaappz/global-super-store-dataset.

1.2 The reason for selection

This dataset is chosen because it provides valuable insights into sales and profitability, customer behaviours, and the trends of e-commerce business over time. E-commerce is a growing, challenging industry, with intense competition, changing consumer behaviours, and fluctuating market trends (Narmadhaa, 2022). Understanding the sales and profitability patterns helps the data analyst to suggest solutions that allow businesses to optimise their operations, identify growth

opportunities, and make data-driven decisions. Additionally, the dataset contains variables such as order date, product category, shipping cost, and profit margin, which allow for a comprehensive analysis of the factors impacting sales and profitability to achieve the aims.

2 Data review and cleaning

2.1 Data review

The dataset appears to be accurate in reflecting the valid values of global sales of a business. Although many Postal Code values are missing, the dataset is considered complete, providing sufficient information for the analysis. However, the consistency of the dataset is not guaranteed due to the typo errors in City, State, and Country fields. Besides, the Dates in this dataset are not in the correct format. All the issues will be addressed in the data cleaning section.

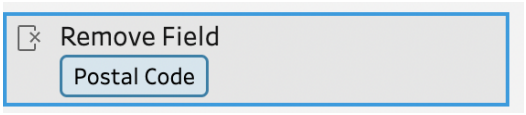
2.2 Data cleaning

Data cleaning is an integral part of data management process; it improves accuracy, reliability, and efficiency. Therefore, Tableau prep builder (TPB) was used to explore and manipulate the data appropriately.

2.2.1 Eliminating outside elements

The postal code column was removed (see Figure 1) using the “remove File function” in TPB, since most of the rows had null values and were irrelevant to this analysis.

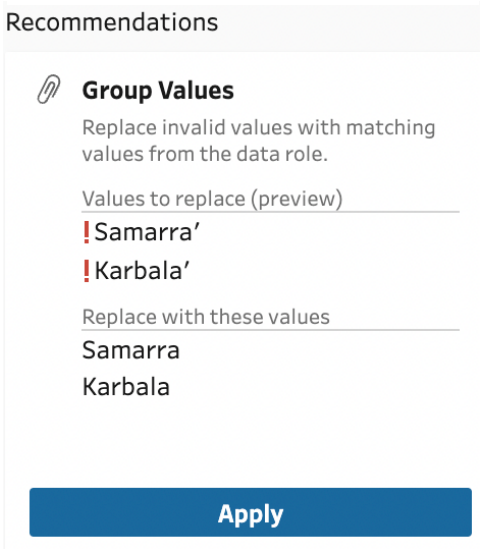
Figure 1
Remove field



2.2.2 Fixing errors

Structural errors such as typographical errors and data categorization were found, e.g., Samarra' was corrected for Samarra in the city column (see Figure 2) as were another 72 values; a typographical error was manually corrected in the Country column and the outliers that did not match the data role and were irrelevant to this analysis in the state and city columns were removed (see Figure 3). Moreover, 24 values were corrected for spelling and pronunciation in the status column.

Figure 2
Spelling errors



Finally, the data categorization of the “city”, “country”, “state”, “order date” and “ship date” columns were also changed (see Appendix A). TPB interprets the data immediately in the document connected to the flow panel by automatically assigning it a data type. However, because different databases are handled differently, the interpretation of TPB may not always be correct (Tableau,n.d). Therefore, the data

type tool was used to change the data type of “order date” and “ship date” from “string” to “date”. Because TPB allows a selection of the data role, the string classification in "country", "city", and "state" was changed. The "city" field was changed to “City”, “State” to “State and Province” since some were actually provinces or towns, and “Country” was changed to “Country and Region” (see Figure 4).

Figure 3
Exclude values

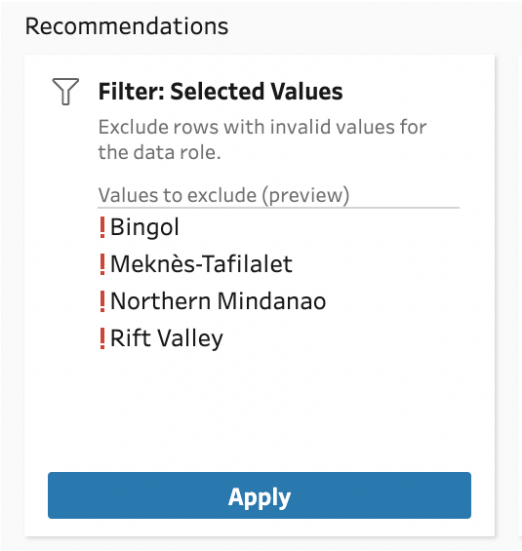
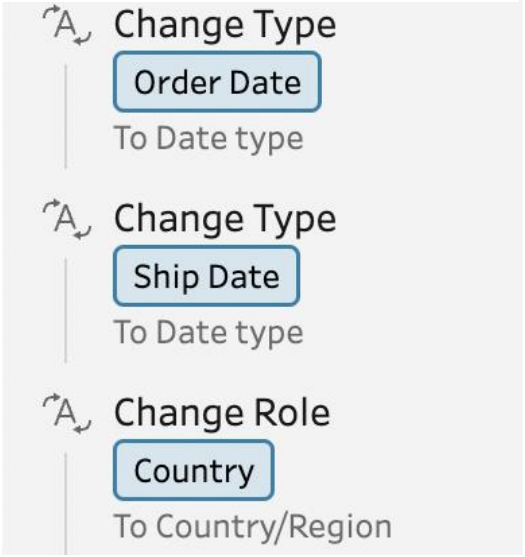


Figure 4
Data type



2.2.3 Filtering outliers

Many outliers in Shipping Costs and Sales can be seen via the below box plots. This could result from the variety of products in categories ranging from cheap to premium items. However, the analysis aims to understand the overall distribution of numbers across categories. Therefore, outliers will be retained for later analysis.

Figure 5
Outliers in Shipping Cost and Sales

3 Data analysis and visualisation plan

The availability of large amount of data represents an unprecedented opportunity for organisations, as it allows to improve and refine the business process management (Liao & Tasi, 2019, p. 1784). Therefore, the data will be analysed to expand over three main aims for a company: Improving its profitability, understanding its audience, and optimising the use of its resources.

3.1 Improving profitability

To improve profitability, it is crucial to understand where the main operating profit comes from, as it generally accounts for most of the total profit of the business (Ruslan et al., 2019, p. 5011).

To increase sales and maximize profit, seasonal sales trends through the years must be identified so that the company can plan inventory and implement marketing strategies accordingly. Therefore, assessing the patterns in sales through the years is fundamental. Using line graphs to show quarterly categories' sales trends allows the data analyst to gain insights about sales patterns from 2011 to 2014 (See Appendix B).

In addition, the contribution of each segment to the annual company profit should be questioned. Understanding this helps the organisation to allocate resources appropriately, ideate targeted marketing strategies, and streamline product development in the segment to gain competitive advantages. To assist the inspection, a stacked bar chart is used to show the profit composition of each component every year (See Appendix C).

3.2 Understanding customer segmentation

The key to a customer-centric approach is knowing the organisation's audience. This way, it is possible to target the most profitable areas with appropriate incentives and increase sales, without affecting the quality perception (Lee & Chen, 2018, n.d.). Within this dataset, this aim can be achieved by identifying which countries have the highest total sales, and how their customers respond to stimuli, such as discounts.

Therefore, to make recommendations while highlighting the relevant patterns, it is possible to create a map with a diverging palette to signify the difference between the countries with the highest sales, and the others (see Appendix D). Also, to visualise the second point, it is possible to create a worksheet with the main markets. Then, each of them can be visualised as a colourblind-friendly bar chart, with the discount split in 10% bins to facilitate the understanding of how discount incentives are relevant within the sales (see Appendix E).

3.3 Optimising logistics

First, to optimise logistics, shipping costs must be reduced, as this translates in better profitability for the company (Ahmed, 2021). Therefore, comparing "ship mode" concurrently with "shipping

cost" by the market can help find solutions for optimising the shipping costs. Two bar charts can be used to visualise this information (see Appendix F).

Another objective to optimise logistics is to improve shipping performance, as it helps businesses improve customer experience and lower operating costs (Stockdale, 2018). Based on the dataset, extra fields, including the "Day To Ship Scheduled", "Days To Ship Actual" and "Ship Status", are created in Tableau to help displaying the trends of "ship status monthly" using a line chart consisting of three trend lines corresponding with the three ship statuses (see Appendix G). This line chart helps the data analyst to have an overview of the businesses' shipping performance.

4 Detailed analysis

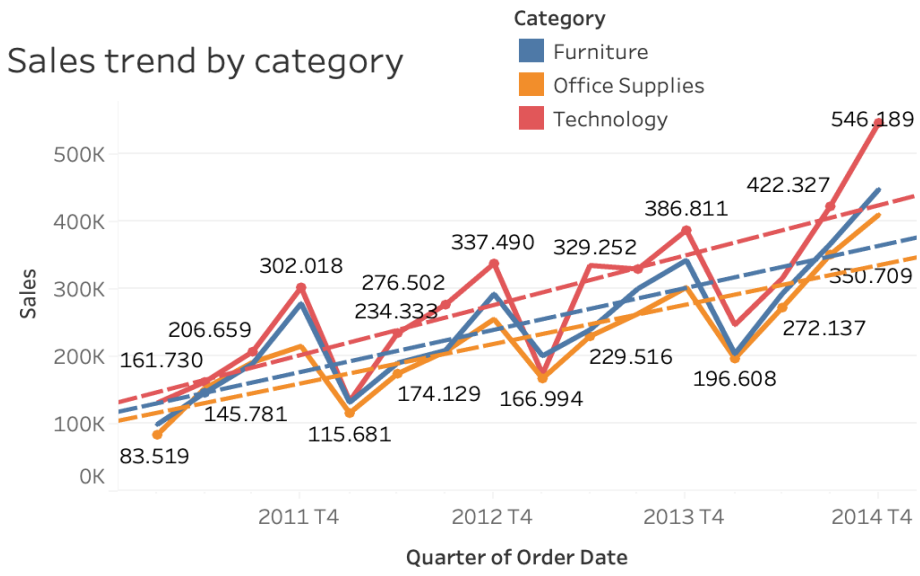
4.1 Improving profitability

As suggested by Bragg (2021) and Bhowmik (2022), analysing sales data, it is possible to predict which products will be successful and which customer segments will generate most profit.

4.1.1 Sales trend by category

Figure 6 shows that technology items had the highest sales, especially in the final quarter of each year. The 2014 holiday shopping season showed a 40% increase compared to previous years. This increase had a domino effect, resulting in a 35% growth in furniture sales and a 30.7% growth in office supply sales. Nonetheless, sales declined after the holiday season, possibly due to overspending. In the first quarter of 2014, post-holiday sales dropped due to "guilty" credit card bills (Dopson, 2022). technology sales fell by 36.3%, while furniture and office supply sales also experienced a decrease.

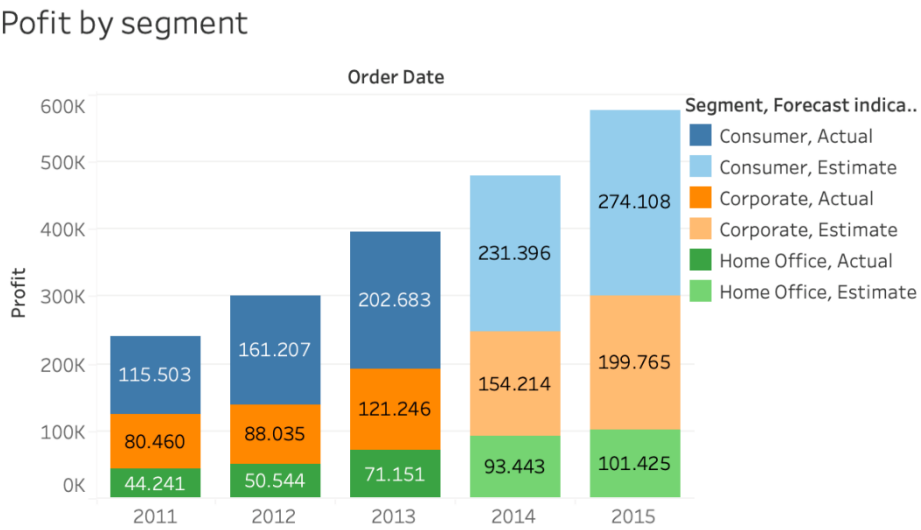
Figure 6
Sales trend by category



According to the seasonal trend analysis, the company needs to enhance its sales and profitability during the off-season. As Galkin (2023) suggests, companies must adopt efficient marketing tactics that cater to customers' needs and preferences during the off seasons to drive sales.

Moreover, appendix C shows that the consumer segment is the most profitable, followed by the corporate and home office segments. Figure 7 further solidifies the notion that this trend will continue. Therefore, it is imperative to prioritize customers retention from the consumer segments when formulating the marketing strategy (Xiemin, 2021).

Figure 7
Profit by segment



Finally, to create successful marketing campaigns and effectively reach the target audience, prioritising ethical practices is crucial. Dopson (2022) states that a multi-channel approach can allow for direct communication with the target customer, in this case, the consumers. However, it is imperative to safeguard customer privacy when collecting personal information, such as email addresses. Companies must balance effective marketing and respect people's inboxes by avoiding excessive emails. RSI Security (2020) strongly recommends implementing necessary measures to prevent spamming practices.

4.2 Understanding customer segmentation

The traditional segmentation approach used divides customers demographically, by country, and behaviourally, by their discount sensitivity, among bestselling countries (Baig et al., 2021). Also, no considerations are made over ethically controversial grounds (e.g., sex and ethnicity), in compliance with most democratic jurisdictions (Kelleher & Tierney, 2018, p. 205).

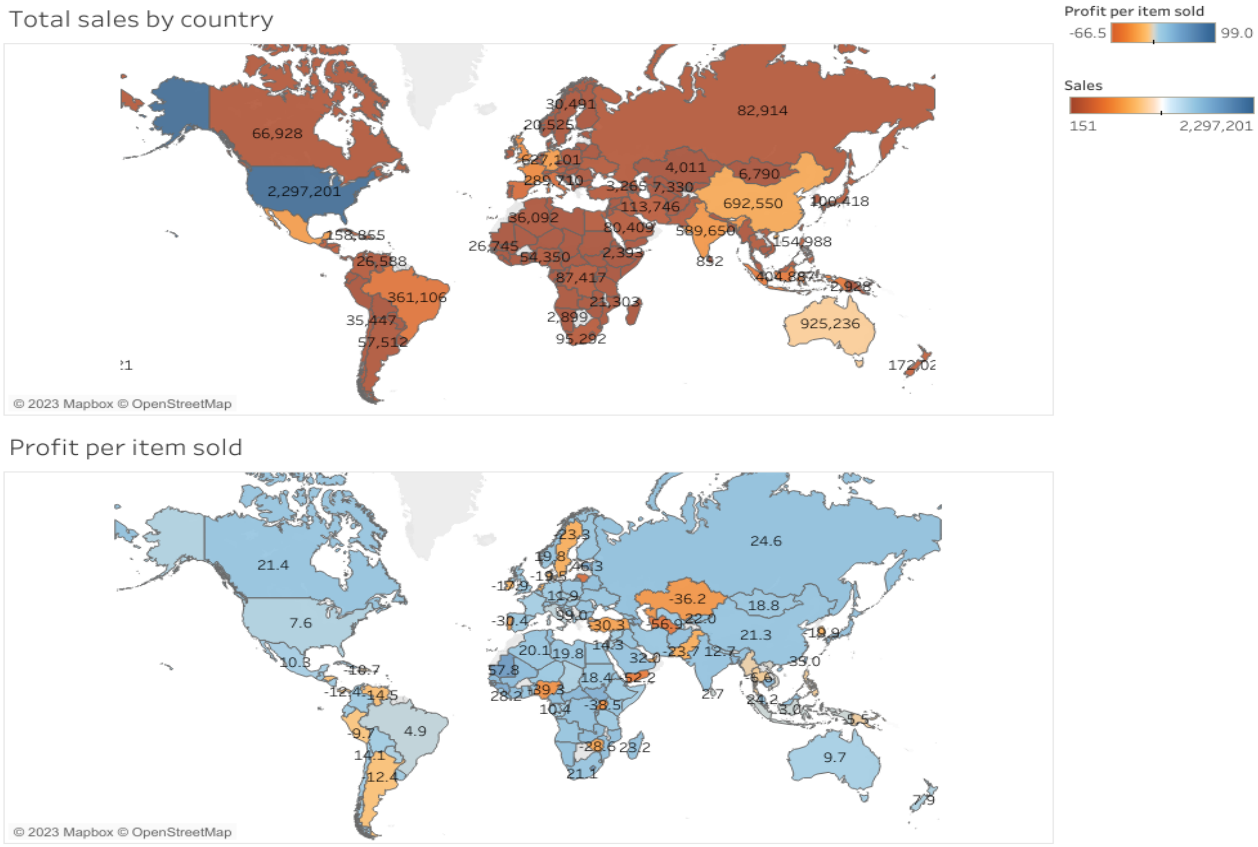
4.2.1 Sales by country

Appendix D shows that the US are the main market, and most of the remaining sales happen in some European countries, Australia, China, and India.

Now, to cut the least grossing segments, the African market could be eliminated. However, further research shows how the “profit per item sold is positive” in these countries (see figure 8), suggesting that this could be a premature abandonment (Olson et al., 2008). Therefore, negative profit areas, such as some central Asia states could be abandoned instead.

Figure 8

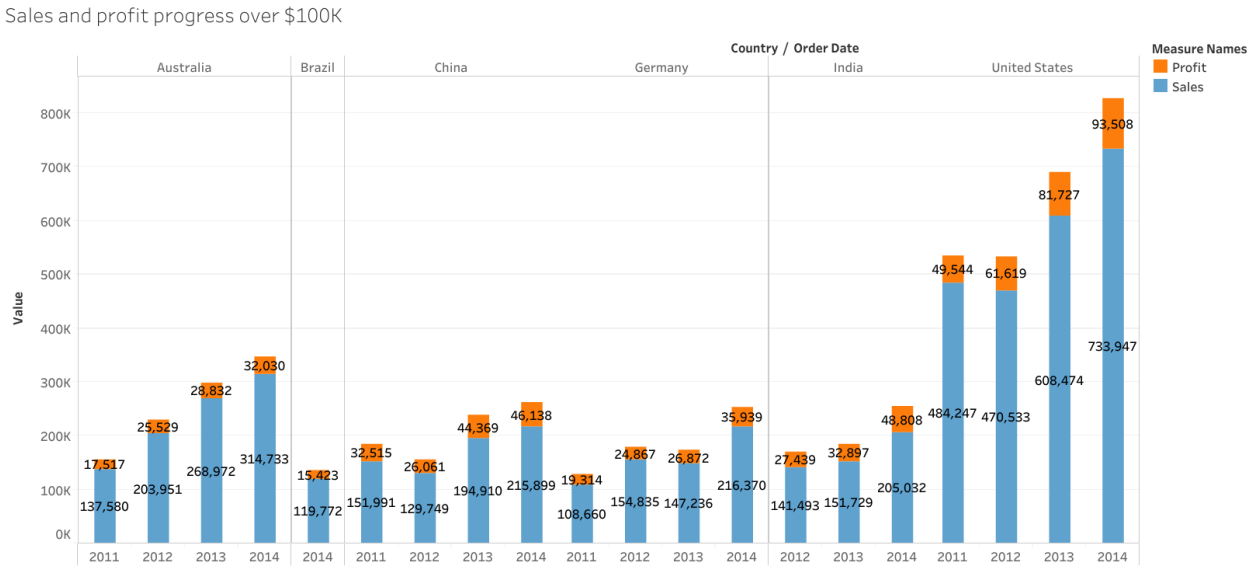
Bestselling countries vs. Profit per item sold



Also, Chinese and Indian customers increased significantly. Hence, it might be worth investing to maintain this growth in the APAC area (see figure 9).

Figure 9

Extract of appendix H, sales and profit progress over $100,000/year



Finally, in South America, Brazilian customers buy significantly more than in neighbouring countries. Hence, the differences in sales policies and offerings should be investigated, to

understand what is causing such a difference and if countermeasures are possible (Roberts, 2021).

4.2.2 Sensitivity to incentives

Appendix E suggests the following:

First, in China, most sales had no discounts, whereas in Australia, most of them had a 10% discount. This justifies differentiated strategies, e.g., 10% discount for Australian buyers, whereas a utilitarian incentive could be used with Chinese customers (Vafainia et al, 2019, p. 77).

Second, Mexican and US customers often buy at retail price, or with discounts around the 20% area. However, the US has a more granular discount policy. Therefore, the organisation should skip discounts under 20% in these markets. Also, in the US, it could develop a variable discount policy, based on their relationship history with the business (Vafainia et al, 2019, p. 78).

Finally, since the US have low profitability per item (see figure 8) the profitability level can be maintained by increasing the retail prices, although this could be ethically controversial (Kelleher & Tierney, 2018, p. 190).

4.3 Optimising logistics

4.3.1 Shipping costs of markets

Figure 10
Shipping mode distribution

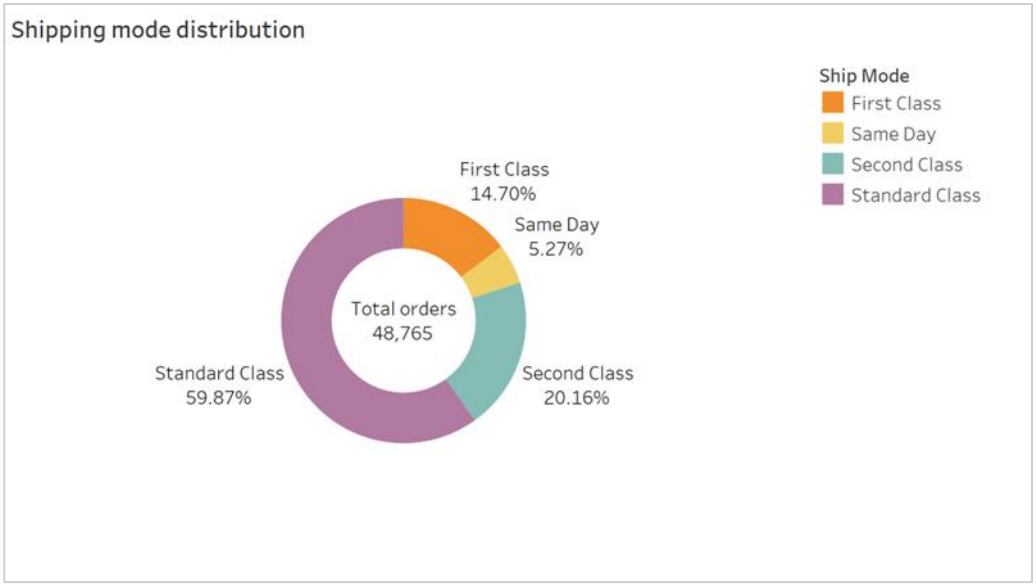
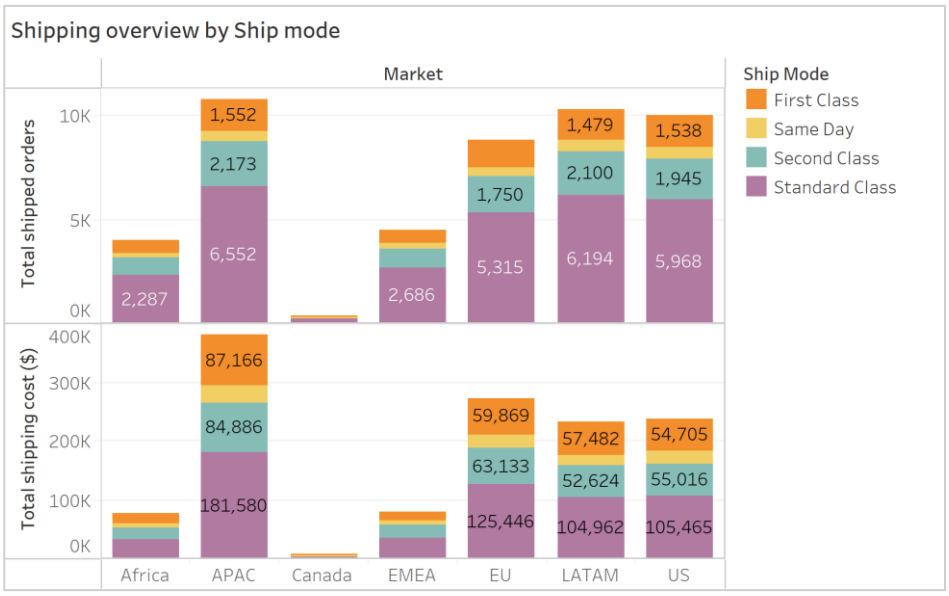


Figure 10 shows that Standard class was the most popular ship mode chosen by customers, as it accounted for nearly 60% of total shipped orders globally. Besides, as seen in Figure 11, Standard class was also the dominant Ship mode in all markets. However, compared to other markets, there was a negative correlation between the number of shipped orders and the total shipping cost in LATAM and US markets. For example, the number of shipped orders using standard class in

LATAM was approximately 1.17 times higher than the figure of the EU, while the total shipping cost in the EU was about 1.2 times higher than that of LATAM.

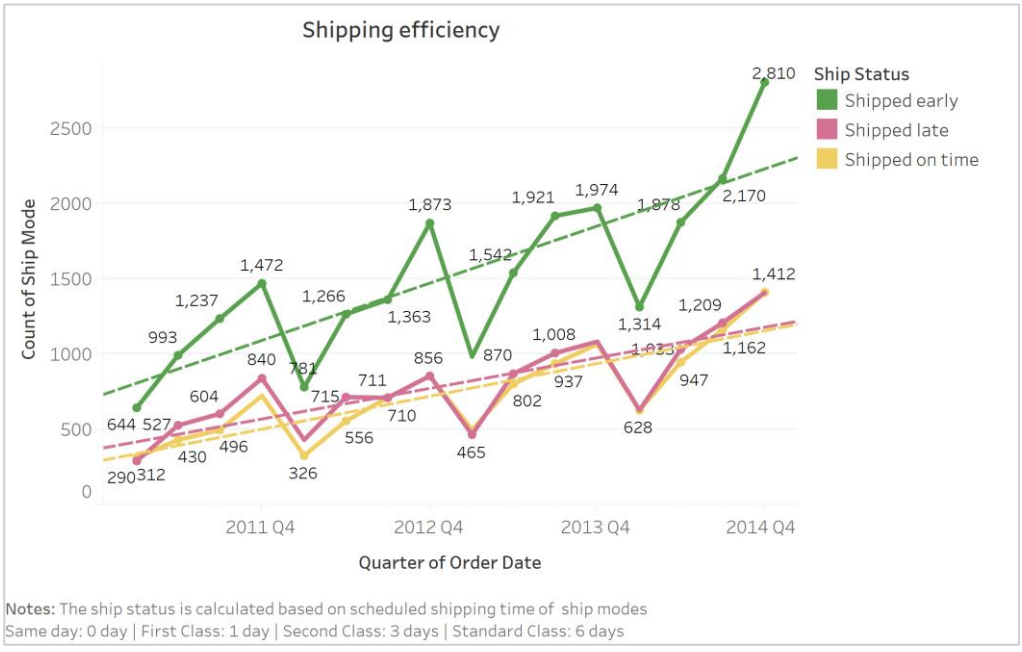
Figure 11
Shipping overview by ship mode



Therefore, higher shipping costs in other markets except LATAM and US is a critical problem that the business should consider. Establishing more regional warehouses, especially in APAC and EU, is essential for the company, as overseas warehouses provide low-cost, quick, and shipping-problem-free delivery of goods to far-off customers (Jayathilaka, 2020).

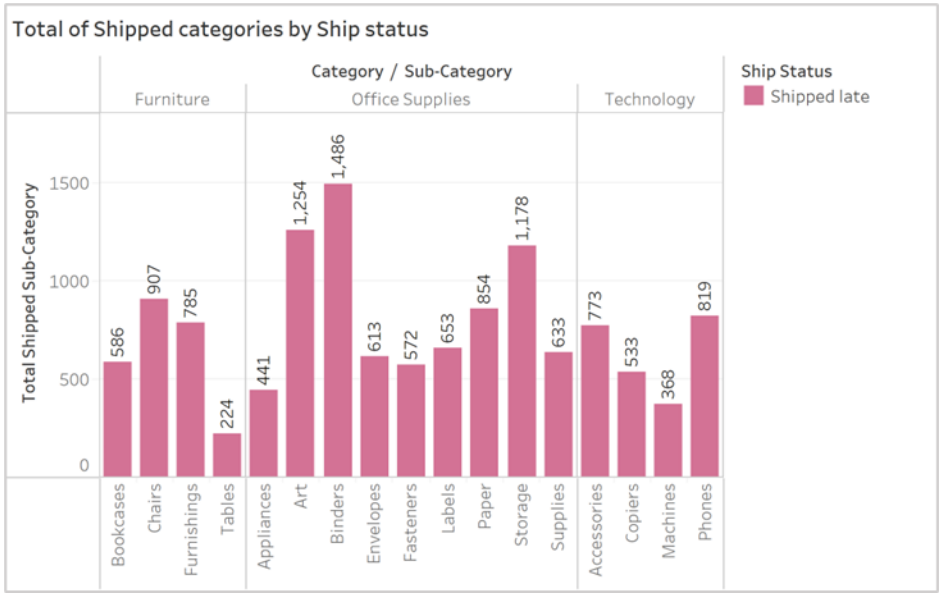
4.3.2 Shipping performance

Figure 12
Shipping Efficiency



It can be observed from Figure 12 that all the trends of Ship status are upwards trends over the years. However, the increased number of orders that were shipped late indicates inefficiency in shipping. The figure rose steadily over four years despite the fluctuations between quarters, and it was always equal to or higher than that of orders shipped on time.

Figure 13
Total of shipped categories by Ship Status



Orders in the Office Supplies category were usually expected to be shipped out late. Specifically, two sub-categories had the longest waiting time, including binders and art, with a total of 1.486 and 1.254 orders, respectively.

Based on the findings, the business should prioritise the improvement of the company's shipping performance, as poor shipping performance will lead to a loss of customers due to dissatisfaction and increased costs for expediting shipping (Saliba, 2022). Therefore, the company needs to optimise inventory management and regularly communicate with suppliers to shorten the lead time for products in the Office Supplies category.

5 Conclusion and recommendations

To sum up, this document found a series of weaknesses and threats that need to be turned into strengths and opportunities.

First, the cleaning process showed that the data collection can be improved. In fact, the loss of data due to format inconsistencies could have undermined the reliability of the information extracted.

Second, within the discussed objectives, several observations should spark a reflection to improve the organisation’s way of doing business. In particular, the sales trends need to be used to timely target the customers with the right product categories. Also, some countries showed great

potential, but this did not always equal to high profit margins. Therefore, the pricing and discounting policies should be revised, with the introduction of new incentive typologies. Lastly, the logistic apparatus came up as the weak link in the organisation, with high costs and low performance. Hence, re-designing the shipping routes and adding local storage facilities is paramount if the goal is to conquer the global market.

Finally, although further, deeper analysis is strongly recommended to improve the accuracy of the recommendations, legal and ethical values must be considered at every stage, to comply with the relevant regulations and protect the organisation image and reputation.

References

Ahmed, N. (2021, July 6). How to reduce shipping costs – effective ways and advice from entrepreneurs [Blog post]. https://www.cloudways.com/blog/reduce-shipping-costs/

Baig, M. R., Govindan, G., & Shrimali, V. R. (2021). *Data science for marketing analytics: a practical guide to forming a killer marketing strategy through data analysis with python (Second).* Packt Publishing. https://lesa.on.worldcat.org/oclc/1268111350

Bragg. (2022, September 13). Sales trend analysis. Accounting Tools. https://www.accountingtools.com/articles/sales-trend-analysis.html

Bhowmik, A. (2022, September 7). Segment Profitability Management - Mirror to Realistic Potentials and Sustained Business growth. Polestar Solutions. https://www.polestarllp.com/blog/segment-profitability-management-for-businesses#:~:text=to%20increase%20profit.-

Dopson, E. (2022, March 11). Beat the Post-Holiday Slump: How to Survive Slow Months in Retail. Shopify. https://www.shopify.com/au/retail/retail-slow-months

Galkin, A. (2019, August 7). 8 Foolproof Ways to Increase Sales and Revenue. Competera.net. https://competera.net/resources/articles/increase-sales-revenue

Jayathilaka, U. R. (2020). Selecting Optimal Overseas Warehouse Location in Global Supply Chain: An Application of Binary Integer Programming. *ResearchBerg Review of Science and Technology, 3*(1), 66–78. Retrieved from https://researchberg.com/index.php/rrst/article/view/56

Kelleher, J. D., & Tierney, B. (2018, April 13). *Data science*. MIT Press. https://ebookcentral-proquest-com.ezproxy.laureate.net.au/lib/think/detail.action?docID=5345177

Lee, J.E., Chen-Yu, J.H. (2018). Effects of price discount on consumers' perceptions of savings, quality, and value for apparel products: mediating effect of price discount affect. *Fash Text 5*(13). https://doi.org/10.1186/s40691-018-0128-2

Liao, S. & Tasi, Y, S. (2019). Big data analysis on the business process and management for the store layout and bundling sales. *Business Process Management Journal, 25*(7), 1783-1801. https://doi-org.torrens.idm.oclc.org/10.1108/BPMJ-01-2018-0027

Mahalingappa, A. (2020). *Global Super Store dataset*. https://www.kaggle.com/datasets/apoorvaappz/global-super-store-dataset

Narmadhaa. (2022). *9 challenges for ecommerce businesses and ways to overcome them*. https://www.zoho.com/en-au/tech-talk/ecommerce-challenges.html

Olson, M. S., van Bever, D., Verry, S. (2008, March). When growth stalls. *Harvard Business Review, 86*(3), 22-23. https://hbr.org/2008/03/when-growth-stalls

Roberts, S. (2021, January 21). *International Marketing Strategies: Examples and Tips.* Cyberclick. https://www.cyberclick.net/numericalblogen/international-marketing-strategies-examples-and-tips#:~:text=Let's%20dive%20in!-

Ruslan, S., Lyubomyr, S., Fedir, H., Sviatoslav, K., Nataliia, P., & Volodymyr, Y. (2019). Improvement of the model for calculating the operating profit of industrial enterprises taking into account the problem of optimal distribution of productive resources. *Traektoriâ Nauki* 5(3), 5011–5017. https://doi.org/10.22178/pos.44-2

RSI Security. (2020, September 29). Ethical Issues in E-Commerce: Handling Customer Data. RSI Security. https://blog.rsisecurity.com/ethical-issues-in-e-commerce-handling-customer-data/

Tableau. (n.d.). Examine Your Data. Help.tableau.com. Retrieved April 15, 2023, from https://help.tableau.com/current/prep/en-us/prep_explore.htm

Tableau. (n.d.). Guide To Data Cleaning: Definition, Benefits, Components, And How To Clean Your Data. Tableau. Retrieved April 15, 2023, from https://www.tableau.com/learn/articles/what-is-data-cleaning#:~:text=Monitoring%20errors%20and%20better%20reporting

Saliba, M. (2022, September 1). The cost of poor on-time delivery performance in supply chain management and how to improve it [Blog post]. https://www.orkestrascs.com/blogs/poor-on-time-delivery-performance

Stockdale, M. (2018). *10 Ways to Improve Your Delivery Performance.* https://www.linkedin.com/pulse/10-ways-improve-your-delivery-performance-michael/

Vafainia, S., Breugelmans, E., & Bijmolt, T. (2019, January 10). Calling Customers to Take Action: The Impact of Incentive and Customer Characteristics on Direct Mailing Effectiveness. *Journal of Interactive Marketing, 45*(1), 62–80.

Jayathilaka, U. R. (2020). Selecting Optimal Overseas Warehouse Location in Global Supply Chain: An Application of Binary Integer Programming. *ResearchBerg Review of Science and Technology, 3*(1), 66–78. Retrieved from https://researchberg.com/index.php/rrst/article/view/56

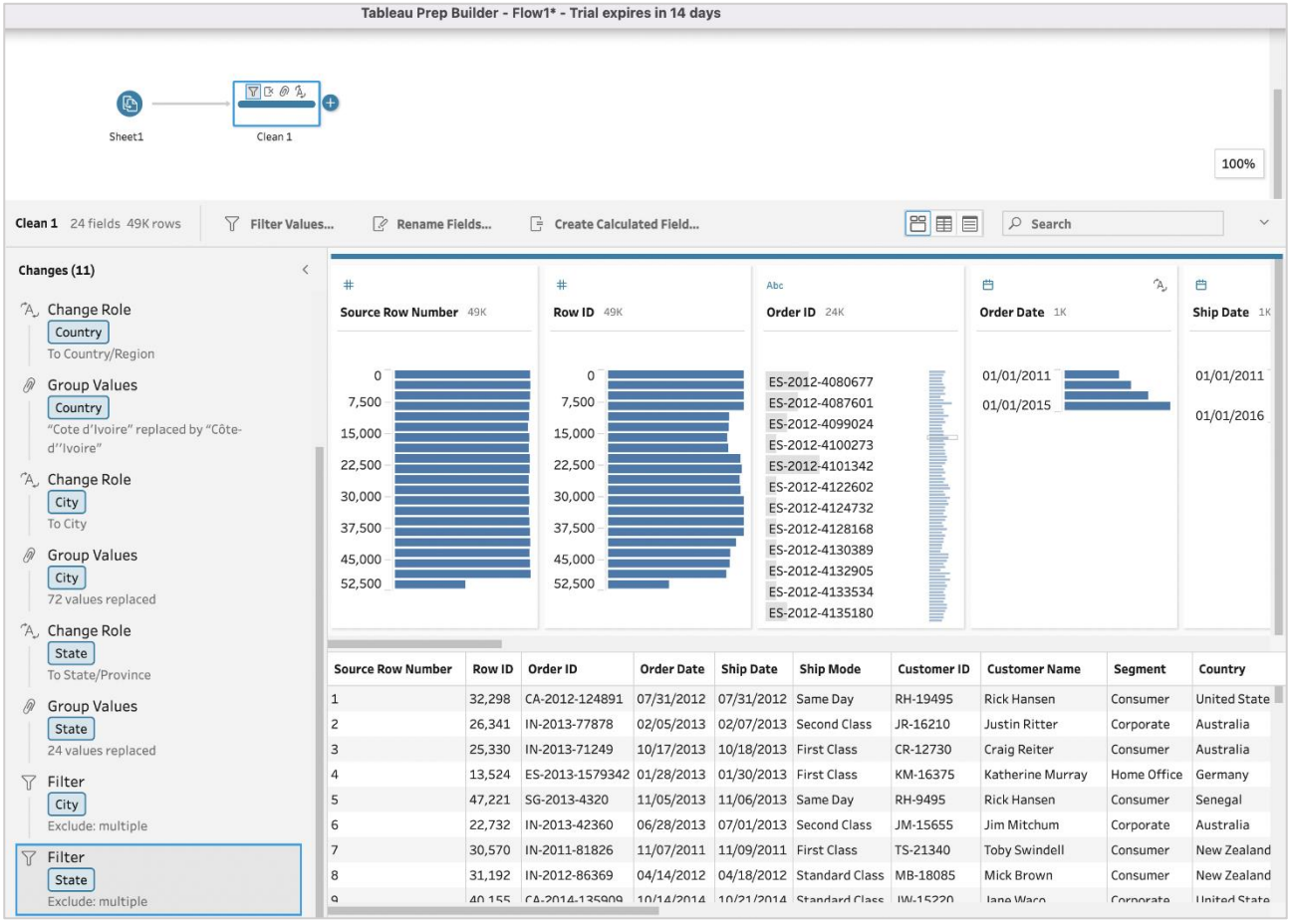
Saliba, M. (2022, September 1). The cost of poor on-time delivery performance in supply chain management and how to improve it [Blog post]. https://www.orkestrascs.com/blogs/poor-on-time-delivery-performance

Xiemin, K. H. (2021, July 20). Exploratory Data Analysis: Super Store. Analytics Vidhya. https://medium.com/analytics-vidhya/exploratory-data-analysis-super-store-cb91c37bcb06

https://www.sciencedirect.com/science/article/pii/S1094996818300707

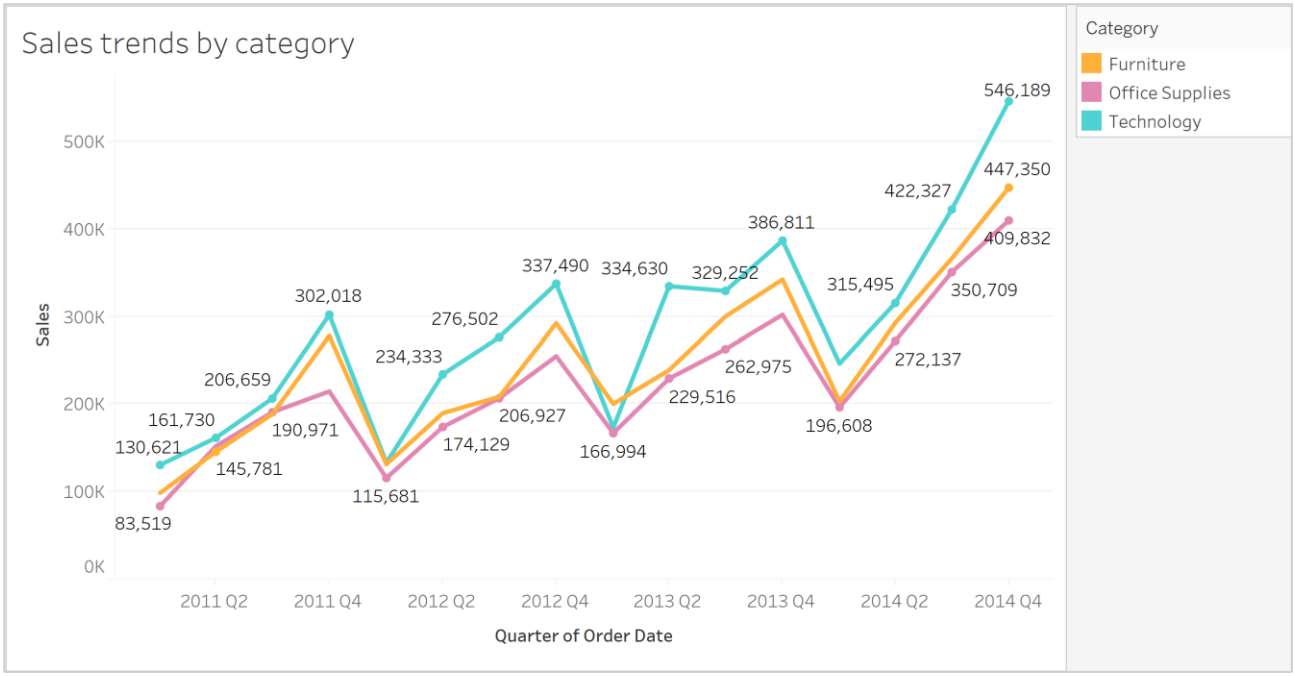
Appendices

Appendix A

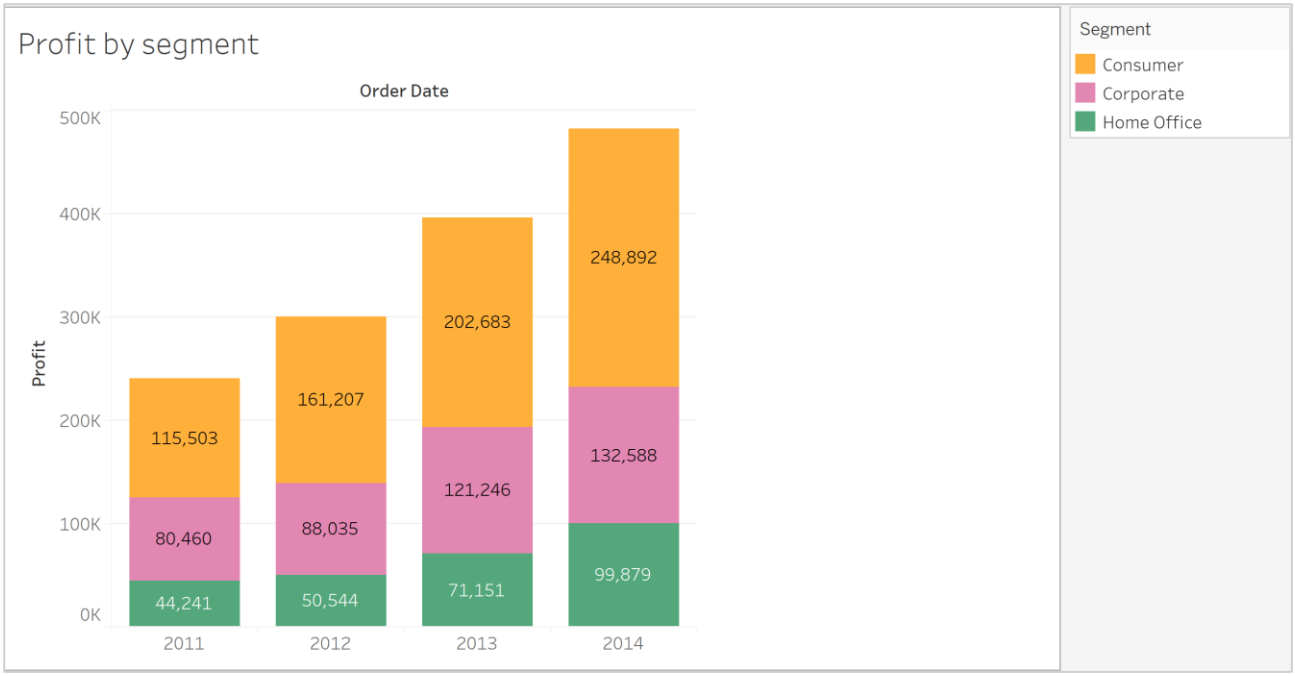


Source Row Number	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country
1	32,298	CA-2012-124891	07/31/2012	07/31/2012	Same Day	RH-19495	Rick Hansen	Consumer	United State
2	26,341	IN-2013-77878	02/05/2013	02/07/2013	Second Class	JR-16210	Justin Ritter	Corporate	Australia
3	25,330	IN-2013-71249	10/17/2013	10/18/2013	First Class	CR-12730	Craig Reiter	Consumer	Australia
4	13,524	ES-2013-1579342	01/28/2013	01/30/2013	First Class	KM-16375	Katherine Murray	Home Office	Germany
5	47,221	SG-2013-4320	11/05/2013	11/06/2013	Same Day	RH-9495	Rick Hansen	Consumer	Senegal
6	22,732	IN-2013-42360	06/28/2013	07/01/2013	Second Class	JM-15655	Jim Mitchum	Corporate	Australia
7	30,570	IN-2011-81826	11/07/2011	11/09/2011	First Class	TS-21340	Toby Swindell	Consumer	New Zealand
8	31,192	IN-2012-86369	04/14/2012	04/18/2012	Standard Class	MB-18085	Mick Brown	Consumer	New Zealand
9	40,155	CA-2014-135909	10/14/2014	10/21/2014	Standard Class	JW-15220	Jane Waco	Corporate	United State

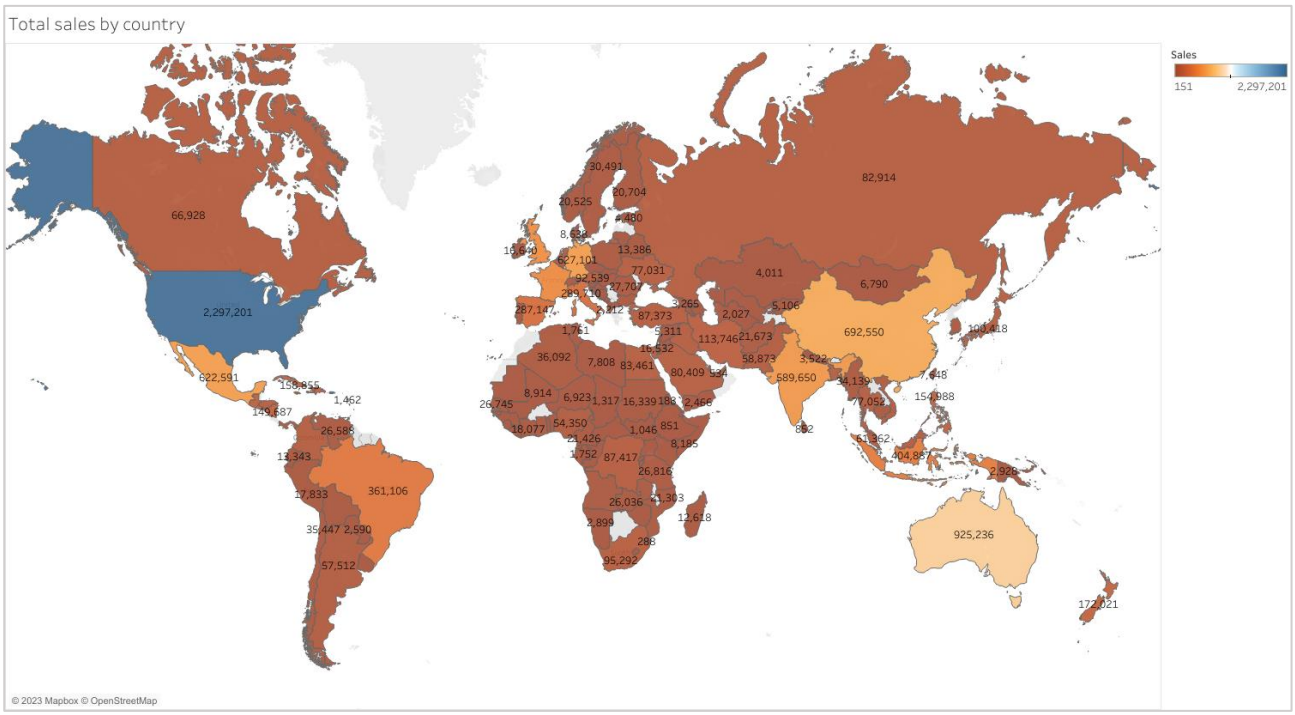
Appendix B


Sales trends by category
Category
Furniture
Office Supplies
Technology
Sales
500K
400K
300K
200K
100K
0K
2011 Q2
2011 Q4
2012 Q2
2012 Q4
2013 Q2
2013 Q4
2014 Q2
2014 Q4
Quarter of Order Date
546,189
447,350
422,327
409,832
386,811
350,709
337,490
334,630
329,252
315,495
302,018
276,502
272,137
262,975
234,333
229,516
206,659
206,927
196,608
190,971
174,129
166,994
161,730
145,781
130,621
115,681
83,519

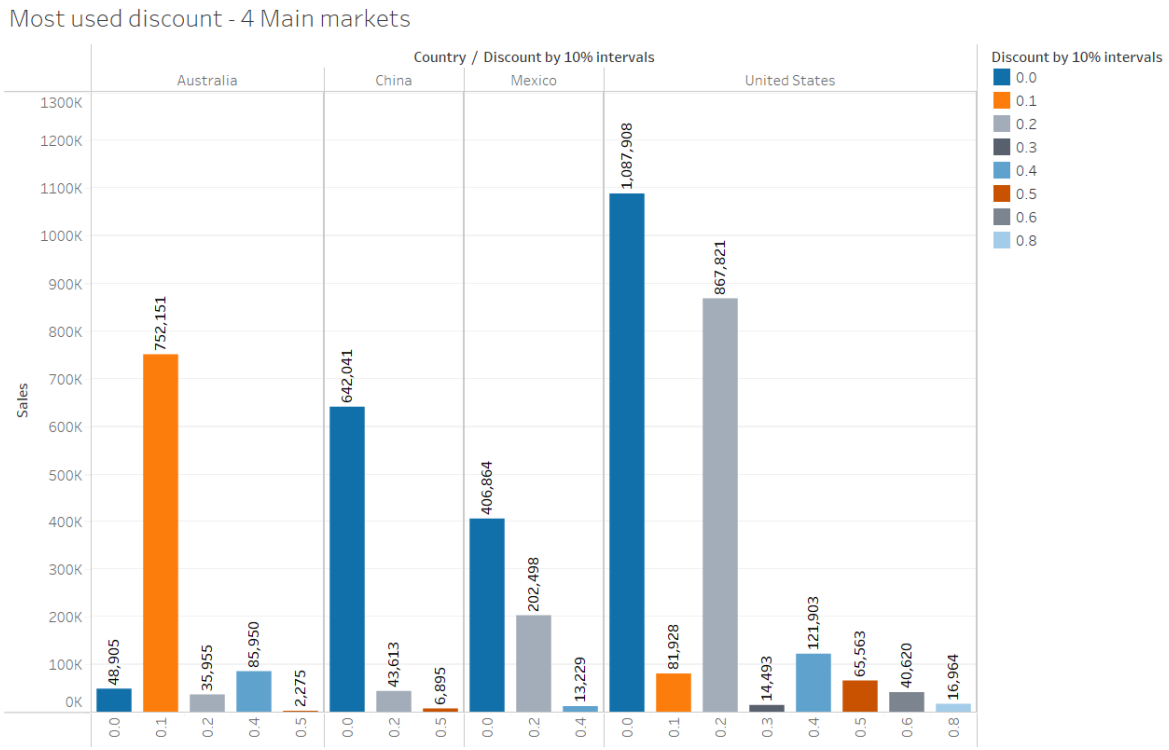
Appendix C


Profit by segment
Order Date
Profit
500K
400K
300K
200K
100K
0K
115,503
80,460
44,241
161,207
88,035
50,544
202,683
121,246
71,151
248,892
132,588
99,879
2011
2012
2013
2014
Segment
Consumer
Corporate
Home Office

Appendix D

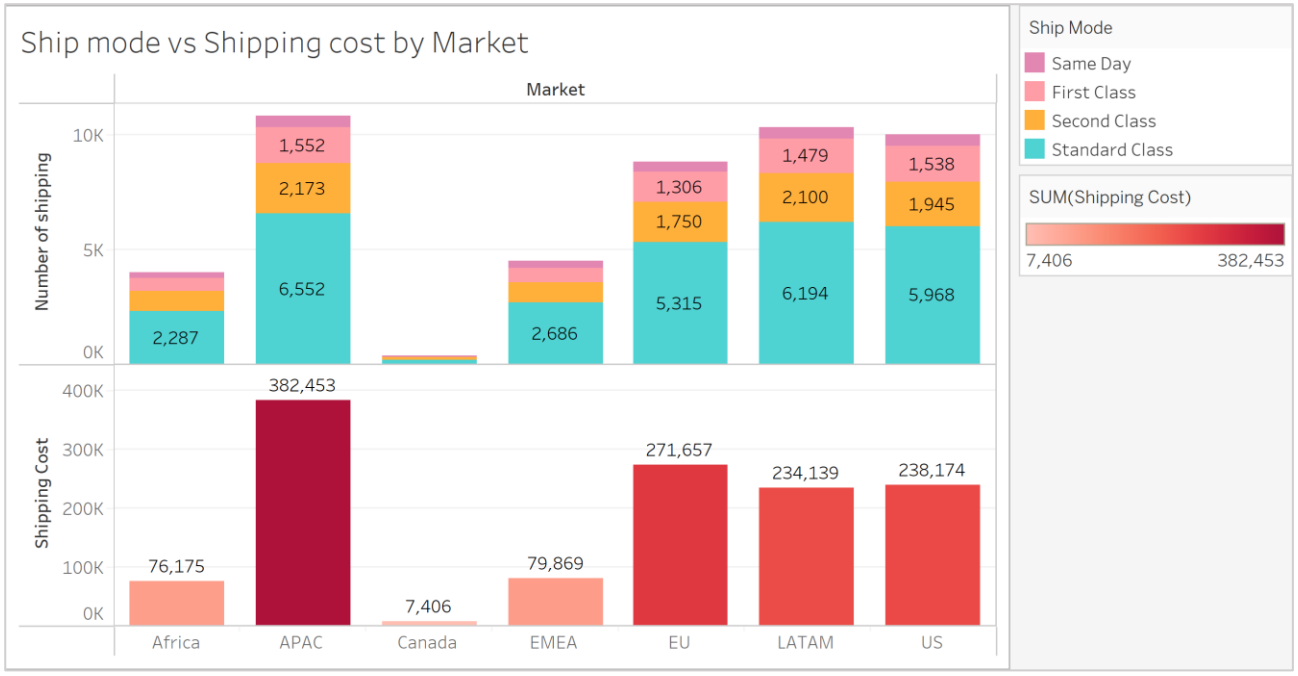

Total sales by country
Sales
151
2,297,201
66,928
2,297,201
82,914
30,491
20,704
20,525
13,386
77,031
4,011
6,790
692,550
5,106
2,027
87,373
113,746
21,673
58,873
80,409
689,650
36,092
7,808
83,461
16,532
8,914
6,923
1,317
16,339
26,745
54,350
18,077
21,426
1,752
87,417
26,816
26,036
12,618
925,236
361,106
57,512
35,447
2,590
26,588
13,343
17,833
1,462
© 2023 Mapbox © OpenStreetMap

Appendix E

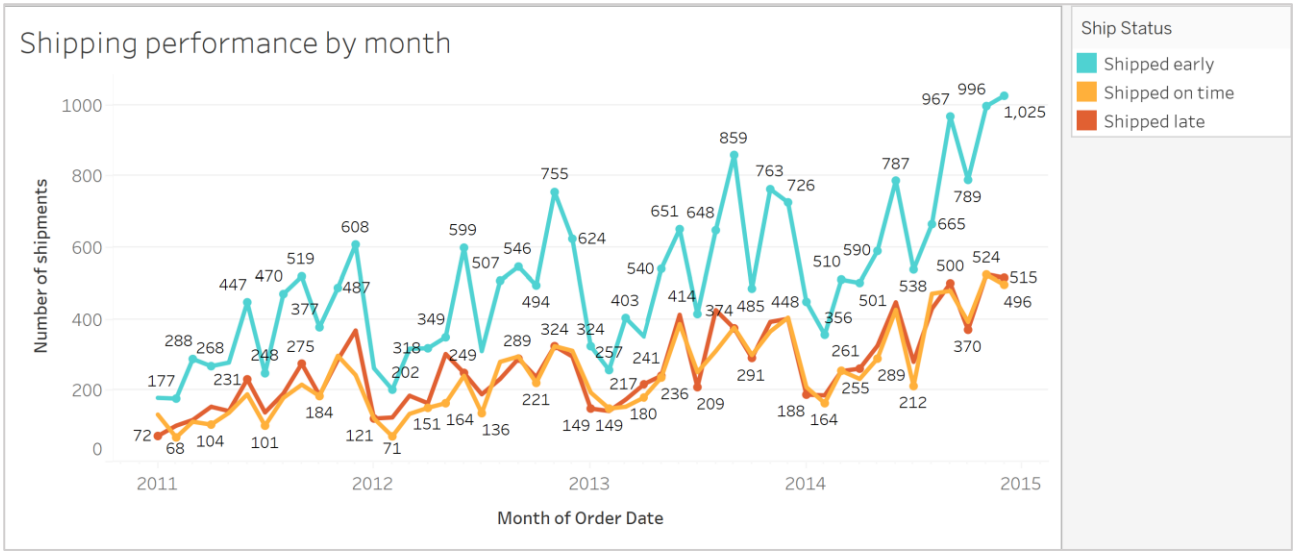


Sum of Sales for each Discount by 10% intervals broken down by Country. Color shows details about Discount by 10% intervals. The view is filtered on Country and Discount by 10% intervals. The Country filter keeps Australia, China, Mexico and United States. The Discount by 10% intervals filter keeps 8 of 8 members.

Appendix F


Ship mode vs Shipping cost by Market
Market
Number of shipping
10K
5K
0K
1,552
2,173
6,552
2,287
2,686
1,306
1,750
5,315
1,479
2,100
6,194
1,538
1,945
5,968
Shipping Cost
400K
300K
200K
100K
0K
382,453
271,657
234,139
238,174
76,175
79,869
7,406
Africa
APAC
Canada
EMEA
EU
LATAM
US
Ship Mode
Same Day
First Class
Second Class
Standard Class
SUM(Shipping Cost)
7,406
382,453

Appendix G


Shipping performance by month
Number of shipments
1000
800
600
400
200
0
2011
2012
2013
2014
2015
Month of Order Date
Ship Status
Shipped early
Shipped on time
Shipped late

Appendix H

Sales and profit progress over $100K