# 

1. **Project Objectives**

This project aims to show off the skills and analyses that may be required of a data analyst for an online retailer management.

The primary objective of an online electronic retailer management is to ensure the smooth process of running and managing sales of goods and services through the internet. They aim to ensure that customers are happy with the goods and services provided while the retailers remain profitable.

Based on this primary objective of an electronic retailer management and the data that has been provided, these few business questions have been identified for exploration:

1. How is the overall health of the business?
   1. What is the overall profit?
   2. What is the overall sales?
   3. What is the overall quantity of items ordered?
   4. What is the distribution of these key variables (profit/sales/quantity) across the world?
2. What is the breakdown of category/sub-category of items in terms of the key variables?
   1. What items should the retailers be focused on to make the most profit or sales?
   2. What items should the retailers stock up on?
   3. What items should the retailers be increasing or reducing discounts on?
3. How is our quality assurance?
   1. How much time does it take for the customer to place an order to the day the items are shipped out?
   2. Are these duration in accordance with our standard based on the type of shipping mode?
   3. Are there any countries in particular that the management wants to focus on?
4. **Data Preparation**

The data consist of 20,281 rows of entries. At the first glance, the current data state seems complete and clean. However, after going through a few checks, there seem to be some discrepant information. The following data verification and cleaning process are done:

|  |  |
| --- | --- |
| **Data Verification** | **Result & Cleaning Process** |
| Blank Data Points | There are no blank data points in the dataset. |
| Duplicated Rows | There are no duplicated rows of data in the dataset. Each cell in the row was combined into one cell and a conditional formatting was used to highlight duplicated values to determine if there were exact duplicated rows. |
| Spelling Errors | ‘Accessories’ was misspelt as ‘Accessoires’ in the original file. Final data file was edited to display the correct spelling. |
| Data Types | Data types were checked and verified when data was input into tableau. |
| Logic Checks (Order Date and Ship Date) | A logic check was done for order date and ship date. Order date should be before ship date as the customer needs to order before the shipping occurs. Hence if the order date comes after shipping date, these entries (N = 268) are marked out and are not used for calculation involving duration of order to ship date. Visualizations not involving this duration will not have these entries removed as the other information might still be correct and crucial for other analyses. |

The following data preparation work was done on tableau:

1. Grouping of data
   1. Category and sub-category of items were combined in the same hierarchy for better visualization and flexibility.
2. Calculated fields
   1. Duration from order date to ship date was created using the datediff function.
   2. A binary variable was created and used in filters to ensure that the order date comes before shipping date for calculations involving duration from order to ship date.
3. **Exploratory Data Analysis and Visualization**

Overall statistics were generated for the crucial variables. Pie charts were also generated based on the consumer and corporate segment. The crucial variables are namely:

1. Overall Profit
2. Overall Sales
3. Overall Quantity

Majority of the profit, sales and quantity ordered are by consumers (Blue) and not corporate (Orange).

Chart

Description automatically generated

*Figure 1. Overall Statistics*

A map based on the profit and sales of the retailers is also generated. From the map, it can be seen that:

1. Not all states are operating at a profit.
2. United States is the country with the highest profit margin.
3. Niger is the country with the most loss.

Map

Description automatically generated

*Figure 2. Profit Map*

A combined visualization of profit margin and discount separated by categories is generated.

1. As profit is one of the most important aspect for the retailers, a profit by categories of item was explored.
   1. Camera category has the highest sum of profit.
   2. Audio & HiFi category has the lowest sum of profit.
2. The average discount by categories was also generated.
   1. Audio & HiFi category has the largest discount given on average.
   2. TV category has the smallest discount given on average.
3. From this visualization, we can conclude that when separated by category, the largest amount of profit is not necessarily dependent on the discount given.

Chart, bar chart

Description automatically generated

*Figure 3. Profit and Discount by Categories*

A combined visualization of sales and quantity ordered separated by categories is generated.

1. From the sales, we can see that
   1. Camera category has the highest sales.
   2. Audio and HiFi category has the lowest sales.
2. The quantity ordered by categories was also generated. A similar pattern emerges when compared to sales.
   1. Camera category has the highest quantity ordered.
   2. Audio and HiFi category has the lowest quantity ordered.
3. We can conclude that
   1. At a category level, the sales and quantity ordered are of the same pattern.
   2. The more the quantity ordered of a category, the more net sales there are.

Chart, bar chart

Description automatically generated

*Figure 4. Sales and Quantity by Categories*

To analyze the items at a greater detail, the same combined visualization of profit margin and discount is generated. This time, it is separated by sub-categories.

1. Profit by sub-categories was explored.
   1. TV has the highest sum of profit, followed by digital cameras and speakers.
   2. TV accessories is making a loss.
2. The average discount by sub-categories was also generated.
   1. HiFi accessories has the largest discount, followed by radios, and speakers.
   2. Projectors has the least discount.
3. We can conclude that
   1. Once again, the discount and profit are not necessarily dependent on each other. This is very prominent from looking at HiFi Accessories, where an average discount percentage of 31% only brought about the lowest profit for sub-categories at $4,184.
   2. The discounts given are not exactly a good tell-tale sign of profit generation.
   3. TVs are good items to sell for retailers to earn a profit.
   4. Retailers should ensure that they have sufficient TV stock in their inventory. However, it is crucial to not purchase too much as the exact quantity of TVs sold is not high. (Refer to Figure 6)

*Chart, bar chart

Description automatically generated*

*Figure 5. Profit and Discount by Sub-categories*

To analyze the items at a greater detail, the same combined visualization of sales and quantity ordered is generated. This time, it is separated by sub-categories.

1. Sales by sub-categories chart was generated.
   1. TV has the highest sum of sales, followed by speakers and digital cameras.
   2. HiFi accessories has the lowest sum of sales
2. Quantity by sub-categories chart was generated.
   1. Camera accessories has the highest quantity ordered, followed by speakers and TV accessories.
   2. HiFi accessories has the lowest quantity ordered.
3. We can conclude that
   1. Unlike the pattern at the category level, at the sub-category level, the highest number of quantity ordered does not necessarily mean the largest amount of sales.
   2. Retailers should stock up on camera accessories if they are just focused on the quantity of items sold.
   3. Retailers should also stock up on speakers looking at its high quantities demanded and high sales.

Chart, bar chart

Description automatically generated

*Figure 6. Sales and Quantity by Sub-categories*

We then look at the time taken for each of the ship mode. Note that in this visualization, those that have the orders’ ship date earlier than the order date are removed.

From this visualization we can conclude that

1. The express ship mode takes on an average of 0.03 Days
2. The prime ship mode takes on an average of 2.27 Days
3. The basic ship mode takes on an average of 3.23 Days
4. The standard ship mode takes on an average of 5.05 Days

Chart, bar chart

Description automatically generated

*Figure 7. Time Taken from Order Date to Ship Date by Ship Mode*

The same analysis is done but with the breakdown of category to check if this pattern appears regardless of the categories of items ordered.

Chart, bar chart

Description automatically generated

*Figure 8. Time Taken from Order Date to Ship Date by Ship Mode in Different Categories*

A tree map of the proportion of ship mode is generated to understand the proportions. In general

1. The standard ship mode takes around 63% of the total orders.
2. The basic ship mode takes around 23% of the total orders.
3. The prime ship mode takes around 18% of the total orders.
4. The express ship mode takes around 7% of the total orders.

Chart, treemap chart

Description automatically generated

*Figure 9. Tree map of Proportion of Ship Mode*

1. **Dashboard**

This dashboard shows the crucial topline overall information for the online electronic retailer management. Our first business question – how is the overall health of the business - is answered here.

At a glance, the management would be able to know their overall profit, sales and quantity all over the world. These key variables the management might be interested in can also be identified from the profit map by hovering the mouse over.

Chart

Description automatically generated with low confidence

*Figure 10. Topline Dashboard*

There is also interactivity in this topline dashboard. Any of the variables can be clicked and filtered to see the details. If the management is curious to know on how China is doing, they can click on the country on the map to filter for information on the country.

Chart

Description automatically generated

*Figure 11. Topline Dashboard – China Filter*

If the management is curious to see the profit and loss of the corporate segment, they can also click or filter from the sidebar.

Map

Description automatically generated with medium confidence

*Figure 12. Topline Dashboard – Corporate Filter*

The next dashboard shows a visualization of 4 variables – Profit, average discount percentage, sales, and quantity ordered – for the different sub-categories of items. We can take a deep-dive to answer our second business question – What is the breakdown of category/sub-category of items in terms of these 4 variables.

From analyzing these 4 variables, the dashboard can help management to prioritize which sub-categories of items to focus on accordingly.

Chart, bar chart

Description automatically generated

*Figure 13. Categories of Items Dashboard - Profit, Discount, Sales, Quantity*

Our last dashboard helps to tackle the question on our quality assurance. To ensure quality assurance, the time taken from order date to ship date should be within the set standards.

Chart, treemap chart

Description automatically generated

*Figure 14. Quality Assurance Dashboard*

If the management is interested in a particular country, they can click on the map to check the details of it. For example, the management might be interested in increasing its outreach to China. The dashboard can aid in this analysis. After filtering, the basic ship mode for China is making an overall loss of $9,219. The management can then investigate to ensure that it is not due to poor quality, service or product that has caused such issues.

Chart

Description automatically generated

*Figure 15. Quality Assurance Dashboard – China/Basic Ship Mode Filter*