Soutenance Finale

Wuyuan XU. Yifei DAI,
Zhaoyang CHEN, Zhifan LIN

UE : IF36

Responsable UE : Joris FALIP, Margaux COEURET

Semestre : P24

Dataset

two dataset:

- sales data of a large supermarket chain in the United States for four years
- the sales data from three Walmart chains in Myanmar.

Reasons for selecting datasets:

- Multi-dimensional sales data
- International perspective

five aspects:

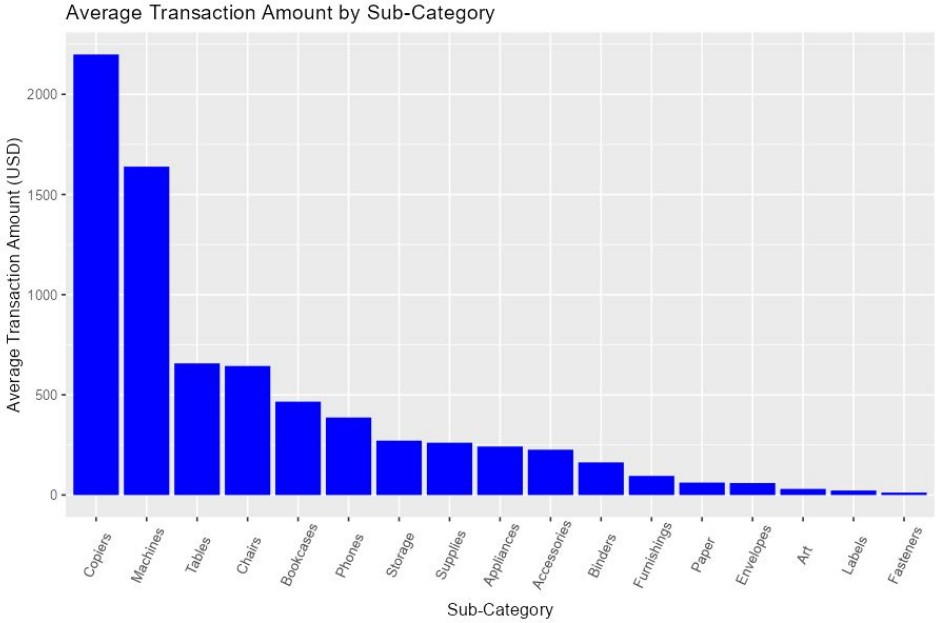
order dimension, customer dimension, product dimension, time dimension, and geographical dimension.

Expected results and applications:

identify the key factors affecting sales and customer behavior, optimize market strategies, and improve operational efficiency

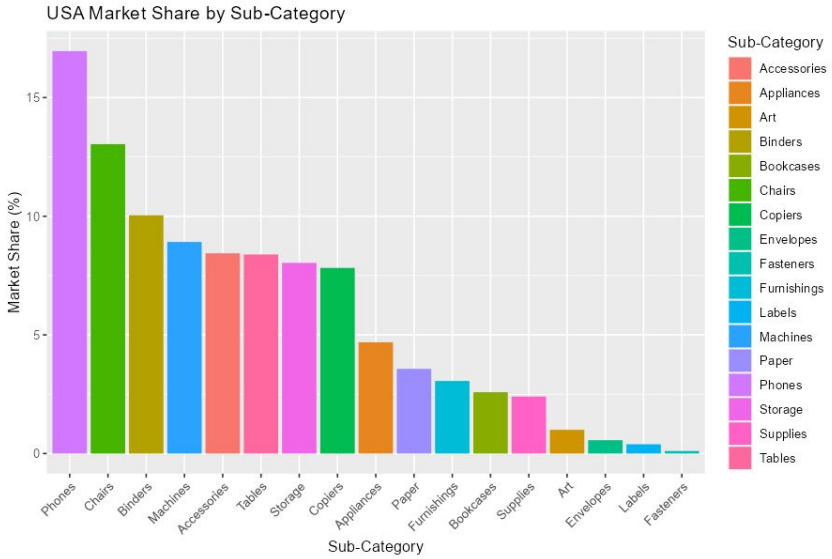
help companies find better competitive strategies in the international market

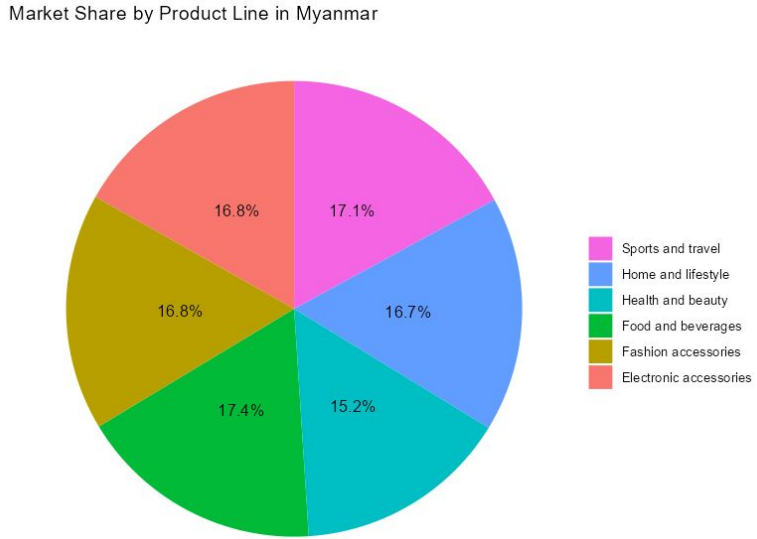
What is the relationship between transaction amount and product category in the U.S.?



- "Copiers" and "Machines" categories have the highest transactions
- Reflects ongoing demand for modern office equipment
- Office supplies and labels show stable demand

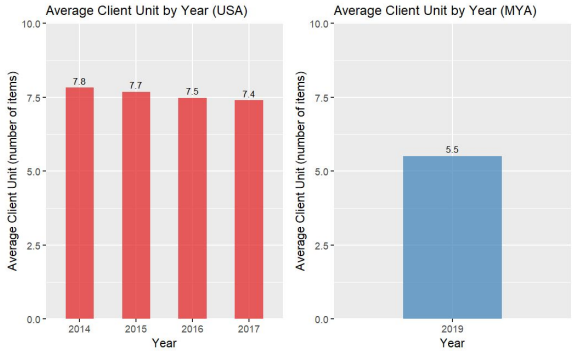
Which product categories are most popular in the U.S. and Vietnam markets?





U.S. Market: The phones category dominates, highlighting the high demand for tech products.
Vietnam Market: Sports and travel categories are significant, reflecting a preference for healthy and active lifestyles.

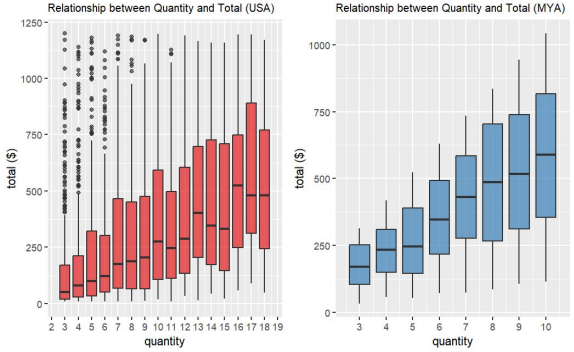
What are the annual average client units of supermarkets in the United States and Myanmar?



USA: The average client units vary slightly from year to year, with a downward trend from 2014 to 2017.

MYA: For the year 2019, the average client unit is 5.5 items.

What is the relationship between the number of items purchased by a customer in a single transaction and the total price?

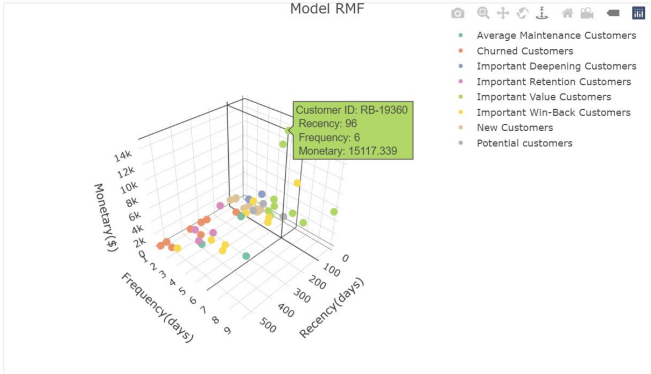


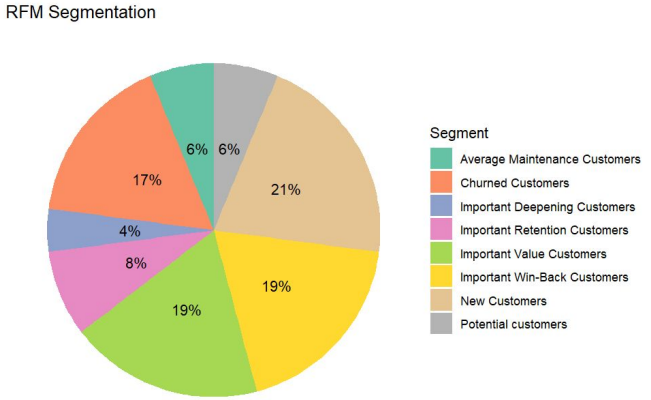
Positive Relationship: Both countries show a **positive correlation**.

Variability: The USA exhibits **more price variability**, while MYA shows a more **predictable, linear increase** in price variability.

Regional Differences: Price dispersion is greater in the USA for the same quantity of items, possibly due to differences in **item prices**, **promotions**, or **purchasing behaviors**.

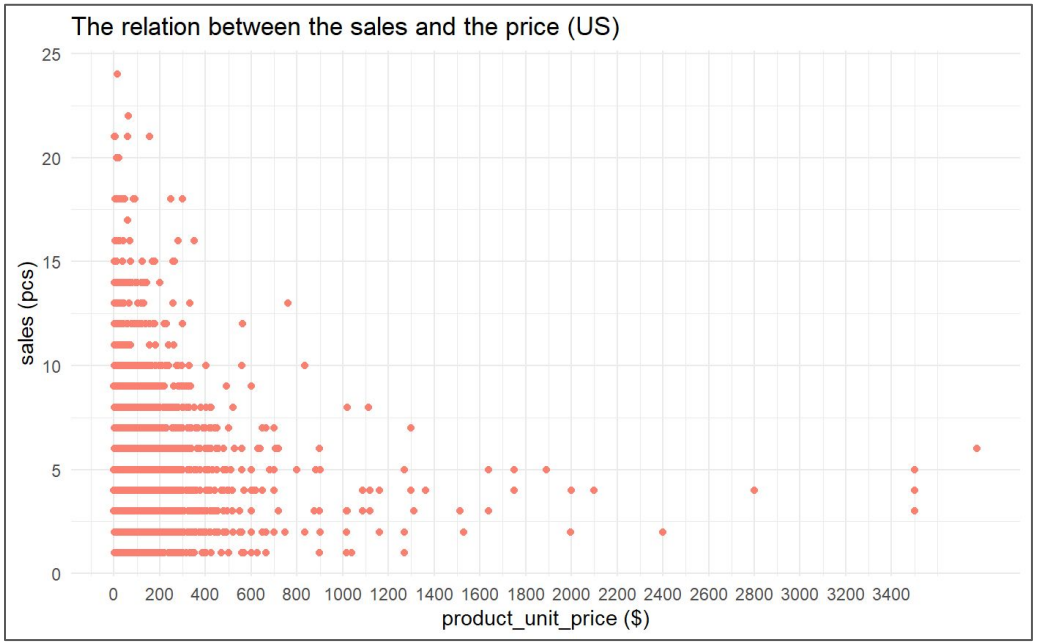
How to classify customers based on their consumption behavior and to formulate corresponding strategies?





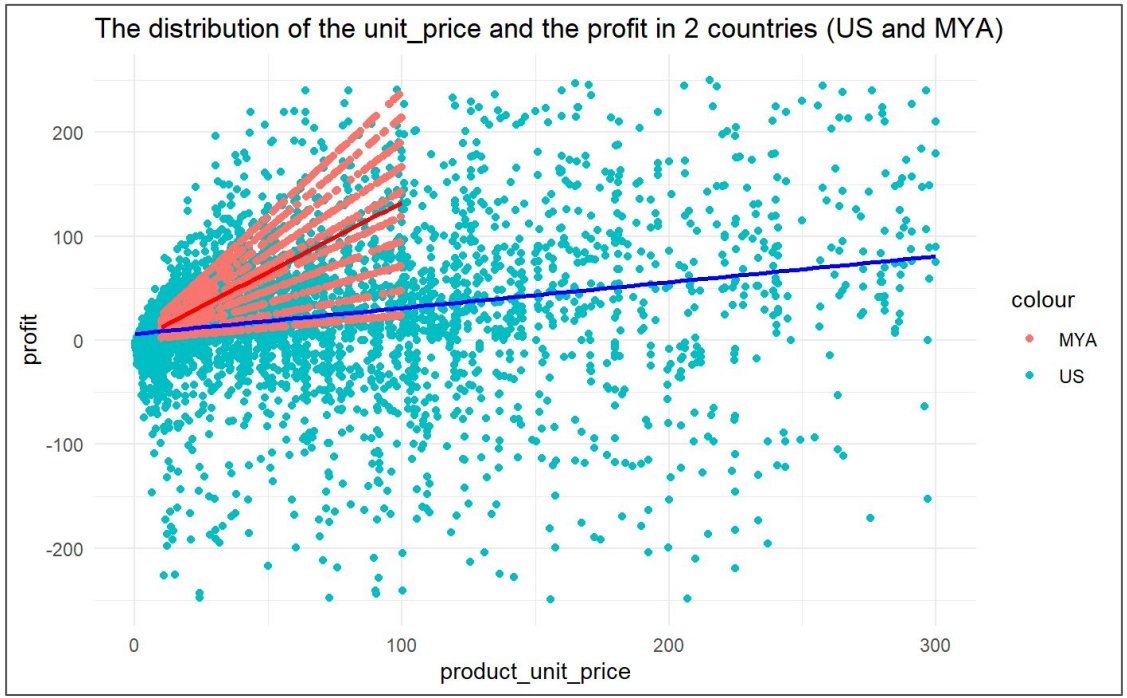
The three types of customers (**New Customers**, **Important Win-Back Customers**, **Important Value Customers**) represent the largest proportion.

What is the relationship between unit price and the sales in the US market ?



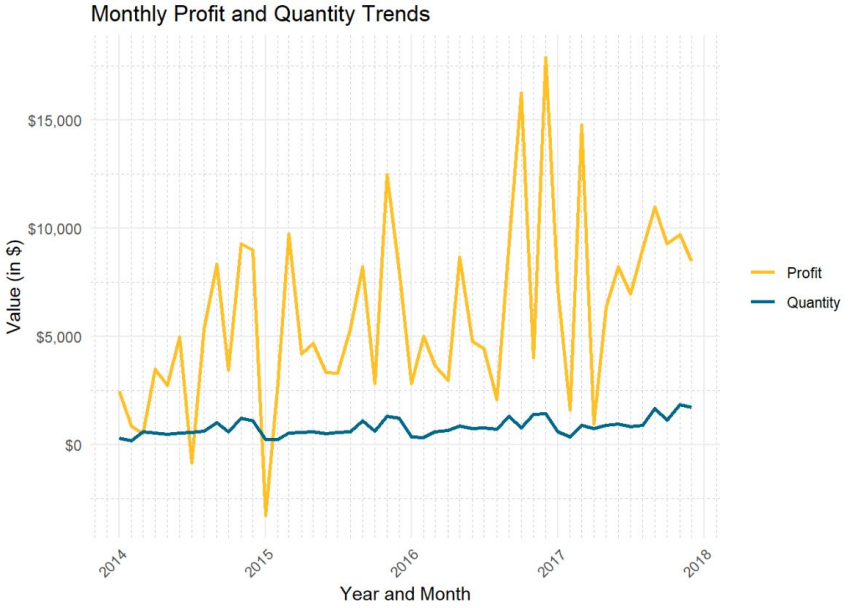
- Products costing less than $500 sell better.
- Lower prices do not always generate more sales.
- The sales can sometimes be increased thanks to lower price.
- The expensive products always have bad performance in sales.

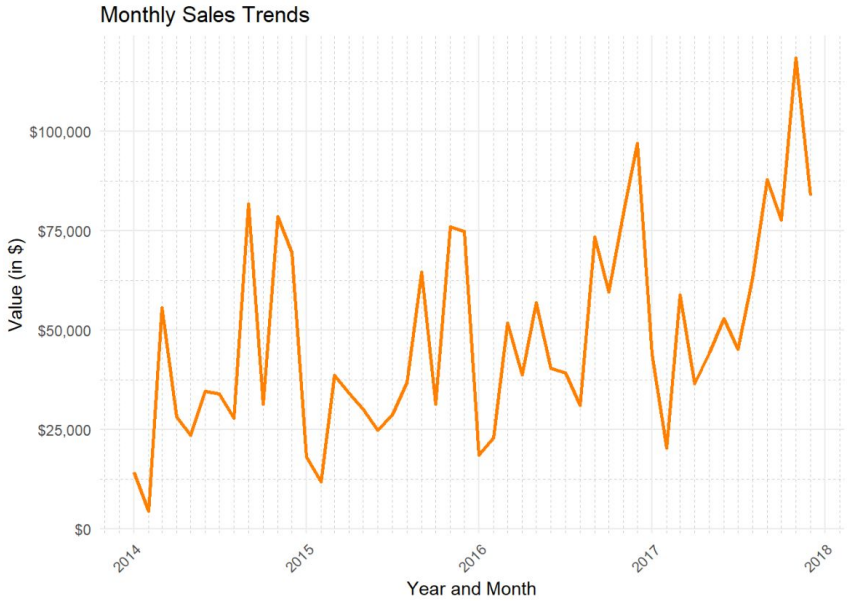
What is the relationship between unit price and the profit ? Can we find the difference between the US and the MYA market ?



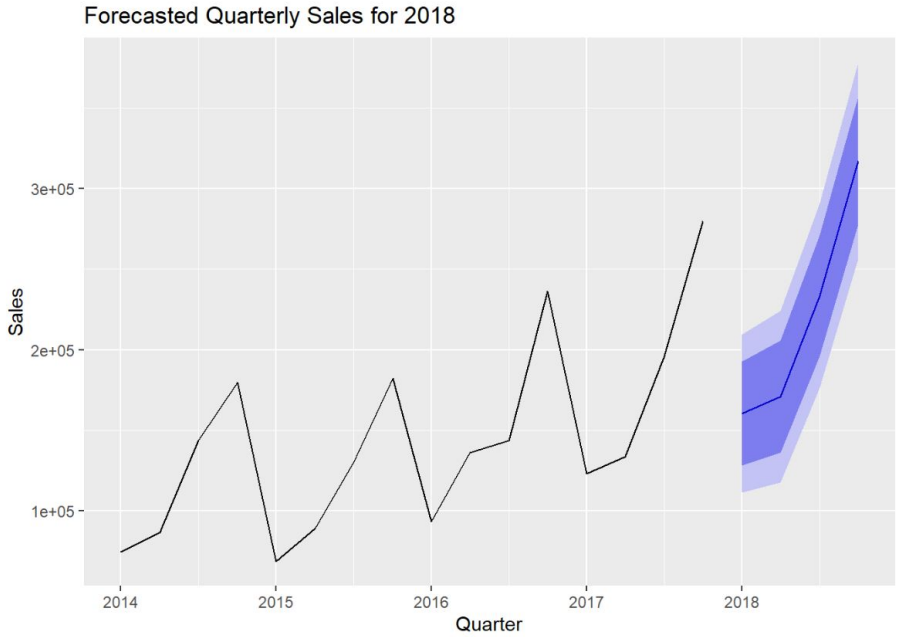
- Higher price sensibility, wider distribution of product prices and lower profit margins
- Higher competitiveness, higher concentration of the profit margin.
- Lower competitiveness, higher distribution of the profit margin.

What is the sales trend each month and what is the growth rate (or decline rate) analysis?



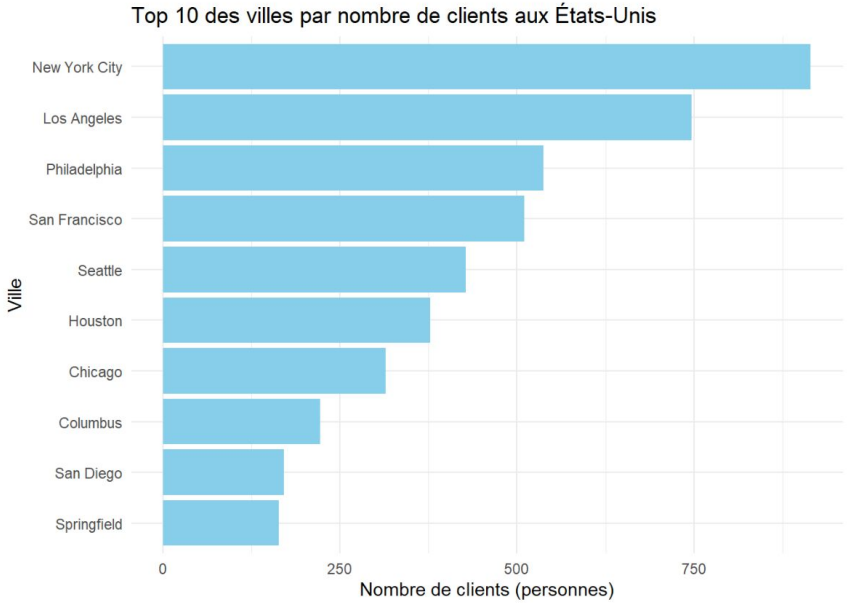


Time series model



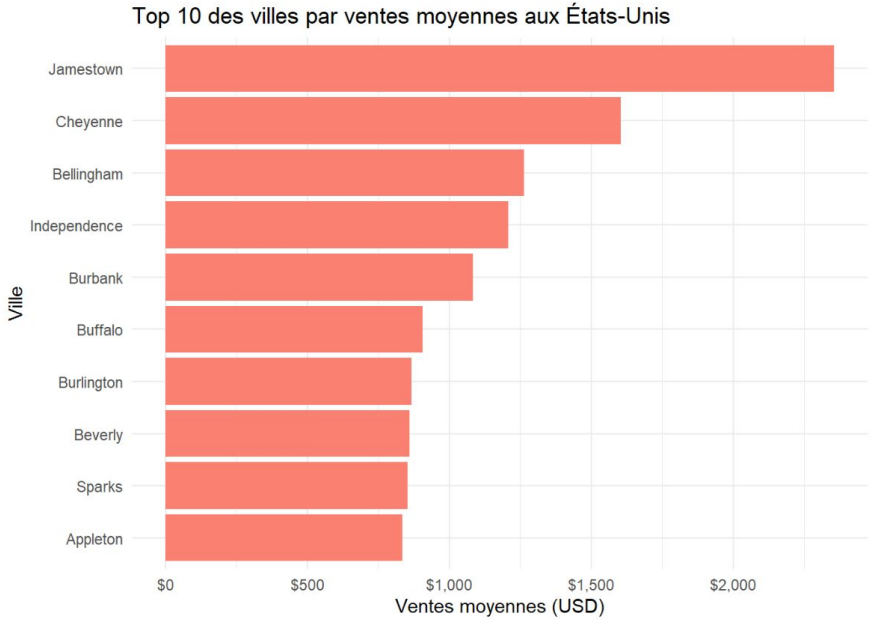
- Use the ARIMA time series analysis model to make a quarterly sales forecast for 2018.
- discontinuous between Q4 2017 and Q1 2018 because the forecast values are range values and no specific points connected in the line graph can be found.
- Light colors indicate that the forecast value has a 95% probability of falling within the range, and dark colors are 95 percent.
- Actual sales data from 2014 to 2017 showed obvious seasonal fluctuations.

Which cities in the United States do customers mainly come from?



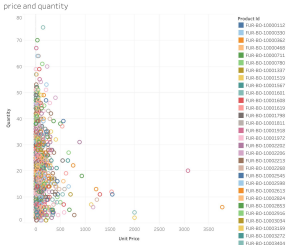
- We can see that New York and Los Angeles are the two cities with the highest sales and customer traffic.
- At the same time, Philadelphia, San Francisco, Seattle, Houston, and Chicago have much higher sales and customer traffic than other cities.
- focus on these cities and make sales plans.

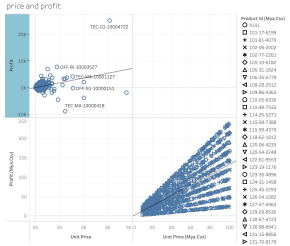
Which city has the highest average customer purchasing power?



- determine which city has the highest average purchasing power.
- selects the top 10 cities for display
- Each bar represents the average sales in a given city. The city name is shown on the vertical axis, while the average sales in dollars is shown on the horizontal axis.

tableau





[if36soutenance | Tableau Public]

1. I tried to realize two of the diagrams with the tool tableau public. I imported the two datasets, and joined them using ‘full outer join’.
2. Then I created a calculation worksheet. I dragged the unit price to the column, and the quantity to the row. Finally I added the attribute ‘product ID’ to the color module, to show different colors for different products.
3. For the last diagram, I dragged the unit price to the column, and the profit to the row. Finally I added the attribut ‘product ID’ to the shape module and to the size module. I wanted to use the color module, but it doesn’t support the trend line. It occurs many parallel trend lines. (parallel to x-axis)

Advantage

1. Tableau is much easier to use for the beginners, because you don't have to remember all the APIs in ggplot2 and in shiny. All the manipulations are intuitional. And an experience in programmation is not required. All the operating steps are similar to the Power BI.
2. Like the Power BI service, Tableau offers us a way to easily publish our diagrams on the Internet and enable others to read without difficulty, while the shiny and the ggplot can't be published directly, and other users have to install R, Rstudio and import the package shiny to execute the ui.R and the server.R. And after the efforts, then they can see the visualisations. So for those who don't have a R environment, it's a problem for shiny.

Weakness

1. In the Tableau, the ui part and the server part adhere together. So sometimes it's difficult to modify the data. In Shiny, we can create 2 files: ui.R and server.R. We can consider the ui.R as the Html DOM, and the server.R as the Javascript, who can change the content of Html elements.
2. Limited functions. Because tableau is easy to use, and it use the the operating method of Power BI. So it can't handle a big amount of data, and can't realize many functions in the shiny and ggplot2. Just as what I met during the jointure of the two tables, I couldn't use the 'full outer join'. So finally the two datasets can't be put in the same diagram. Instead, they are separated in two diagrams with their respective scale in x-axis and y-axis.

Merci pour votre écoute !

Questions ?