

Name: Ushasree J

Reg.No: 21BPS1491

Campus: VIT Chennai

DA Assignment 3

The growth of supermarkets in most populated cities is increasing and market competitions are also high. The dataset is one of the historical sales of supermarket company which has recorded in 3 different branches for 3 months data. Predictive data analytics methods are easy to apply to this dataset.

Attribute information

Invoice id: Computer-generated sales slip invoice identification number

Branch: Branch of supercenter (3 branches are available identified by A, B and C).

City: Location of supercenters

Customer type: Type of customers, recorded by Members for customers using member cards and Normal for those without member cards.

Gender: Gender type of customer

Product line: General item categorization groups - Electronic accessories, Fashion accessories, Food and beverages, Health and beauty, Home and lifestyle, Sports and travel

Unit price: The price of each product in \$

Quantity: Number of products purchased by the customer

Tax: 5% tax fee for customers buying

Total: Total price including tax

Date: Date of purchase (Record available from January 2019 to March 2019)

Time: Purchase time (10 am to 9 pm)

Payment: Payment used by the customer for the purchase (3 methods are available – Cash, Credit card and Ewallet)

COGS: Cost of goods sold

Gross margin percentage: Gross margin percentage

Gross income: Gross income

Rating: Customer stratification rating on their overall shopping experience (On a scale of 1 to 10)

MySQL workbench

SQL Queries

- Created a Database 'Supermarket' and Table named as 'Supermarket_sales'.

The screenshot shows the MySQL Workbench interface. On the left, the 'SCHEMAS' pane shows a tree view with 'supermarket' expanded, containing 'Tables', 'Views', 'Stored Procedures', 'Functions', 'superstore', and 'sys'. The 'Query 1' editor contains the following SQL code:

```
1 use supermarket;
2 select * from supermarket_sales;
3
4
5
```

The 'Result Grid' shows the following data:

Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross margin percentage
750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	1/5/2019	13:08	Ewallet	522.83	4.761904762
226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22	3/8/2019	10:29	Cash	76.4	4.761904762
631-41-3108	A	Yangon	Normal	Male	Home and life	Electronic accessories	16.2155	340.5255	3/3/2019	13:23	Credit card	324.31	4.761904762	
123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.048	1/27/2019	20:33	Ewallet	465.76	4.761904762
373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2/8/2019	10:37	Ewallet	604.17	4.761904762
699-14-3026	C	Naypyitaw	Normal	Male	Electronic accessories	85.39	7	29.8865	627.6165	3/25/2019	18:30	Ewallet	597.73	4.761904762
355-53-5943	A	Yangon	Member	Female	Electronic accessories	68.84	6	20.652	433.692	2/25/2019	14:36	Ewallet	413.04	4.761904762
315-22-5665	C	Naypyitaw	Normal	Female	Home and lifestyle	73.56	10	36.78	772.38	2/24/2019	11:38	Ewallet	735.6	4.761904762
665-32-9167	A	Yangon	Member	Female	Health and beauty	36.26	2	3.626	76.146	1/10/2019	17:15	Credit card	72.52	4.761904762
692-92-5582	B	Mandalay	Member	Female	Food and beverages	54.84	3	8.226	172.746	2/20/2019	13:27	Credit card	164.52	4.761904762
351-62-0822	B	Mandalay	Member	Female	Fashion accessories	14.48	4	2.896	60.816	2/6/2019	18:07	Ewallet	57.92	4.761904762
529-56-3974	B	Mandalay	Member	Male	Electronic accessories	25.51	4	5.102	107.142	3/9/2019	17:03	Cash	102.04	4.761904762

The screenshot shows the MySQL Workbench interface. On the left, the 'SCHEMAS' pane shows a tree view with 'supermarket' expanded, containing 'Tables', 'Columns', 'Indexes', and 'Foreign Keys'. The 'Columns' pane shows a list of columns for 'supermarket_sales': Invoice ID, Branch, City, Customer type, Gender, Product line, Unit price, Quantity, Tax 5%, Total, Date, Time, Payment, cogs, gross margin percentage, gross income, and Rating. The 'Query 1' editor contains the following SQL code:

```
1 use supermarket;
2 select * from supermarket_sales;
3 select city from supermarket_sales;
4 select `Customer type` from supermarket_sales where city="Naypyitaw";
5
```

The 'Result Grid' shows the following data:

city
Yangon
Naypyitaw
Yangon
Yangon
Yangon
Naypyitaw
Yangon
Naypyitaw
Yangon
Mandalay
Mandalay
Mandalay
Yangon
Yangon

SCHEMAS

Filter objects

supermarket

Tables

supermarket_sales

Columns

- Invoice ID
- Branch
- City
- Customer type
- Gender
- Product line
- Unit price
- Quantity
- Tax 5%
- Total
- Date
- Time
- Payment
- cogs
- gross margin pe
- gross income
- Rating

```

1 • use supermarket;
2 • select * from supermarket_sales;
3 • select city from supermarket_sales;
4 • select `Customer type` from supermarket_sales where city="Naypyitaw";
5

```

Result Grid

Customer type
Normal
Normal
Normal
Member
Member
Member
Normal
Member
Member
Member

Query 1

```

2 • select * from supermarket_sales;
3 • se: Execute the selected portion of the script or everything, if there is no selection
4 • select `Customer type` from supermarket_sales where city="Naypyitaw";
5
6 /* counting*/
7 • select count(`Invoice ID`) from supermarket_sales;
8

```

Result Grid

count(`Invoice ID`)
1000

Query 1

```

8
9 • select Branch, count(`Invoice ID`) as Total_order
10 from supermarket_sales
11 group by Branch
12 order by Total_order desc;

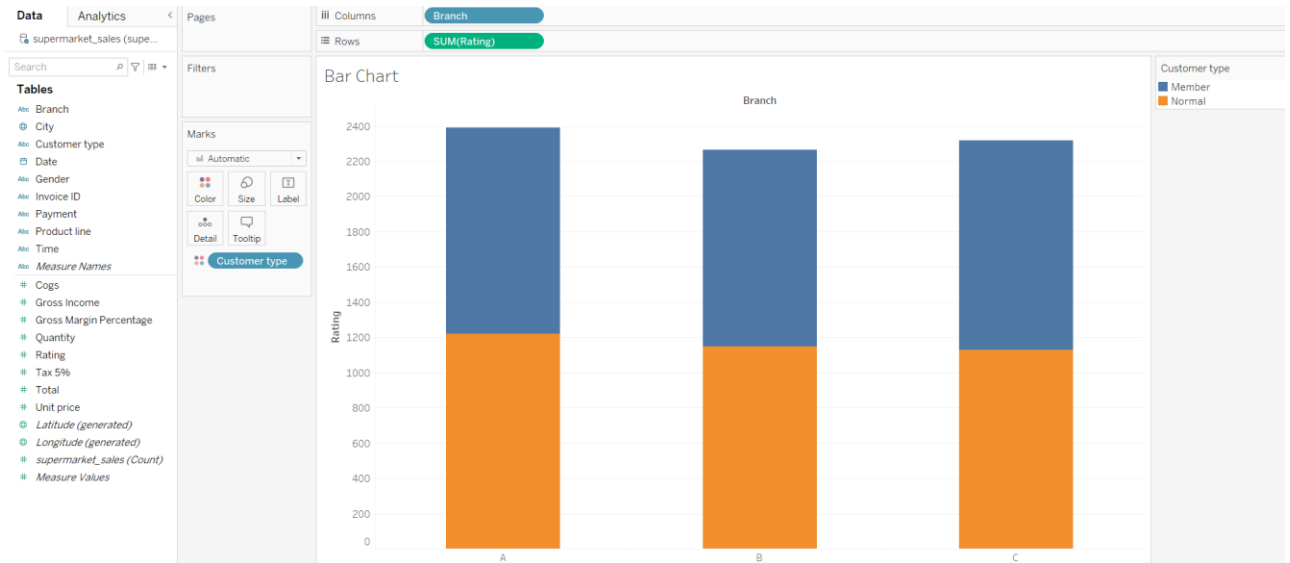
```

Result Grid

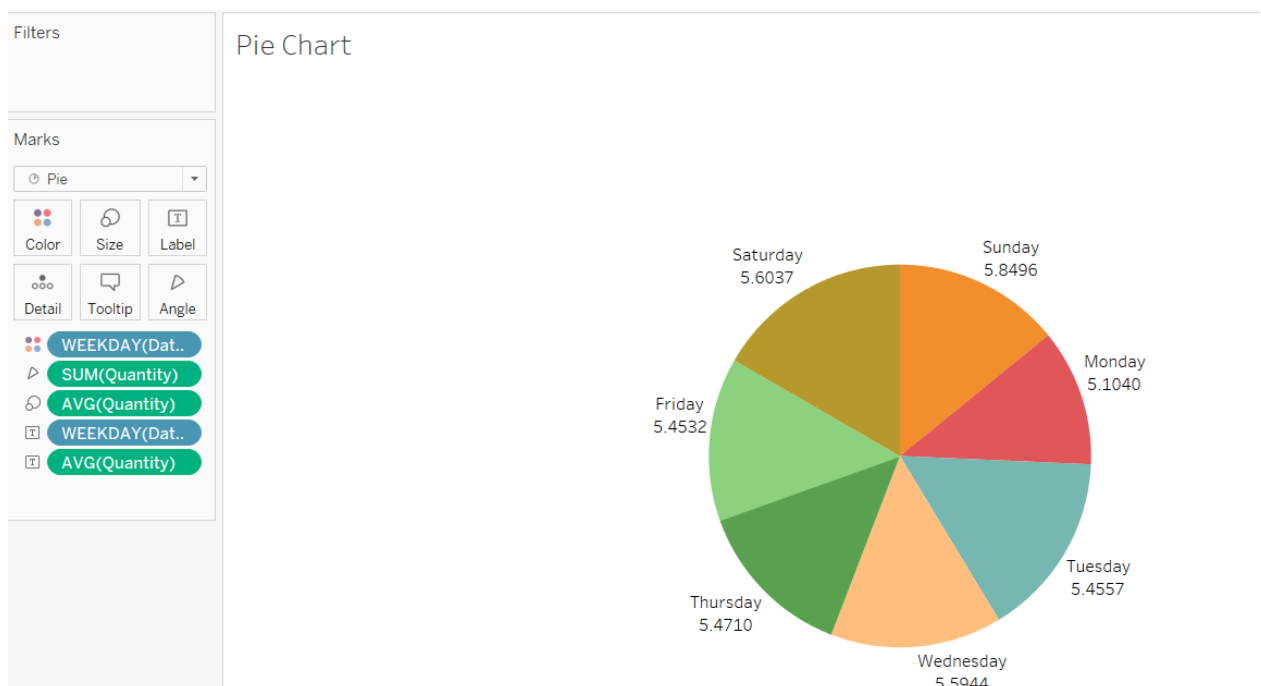
Branch	Total_order
A	340
B	332
C	328

Data Visualisation using Tableau

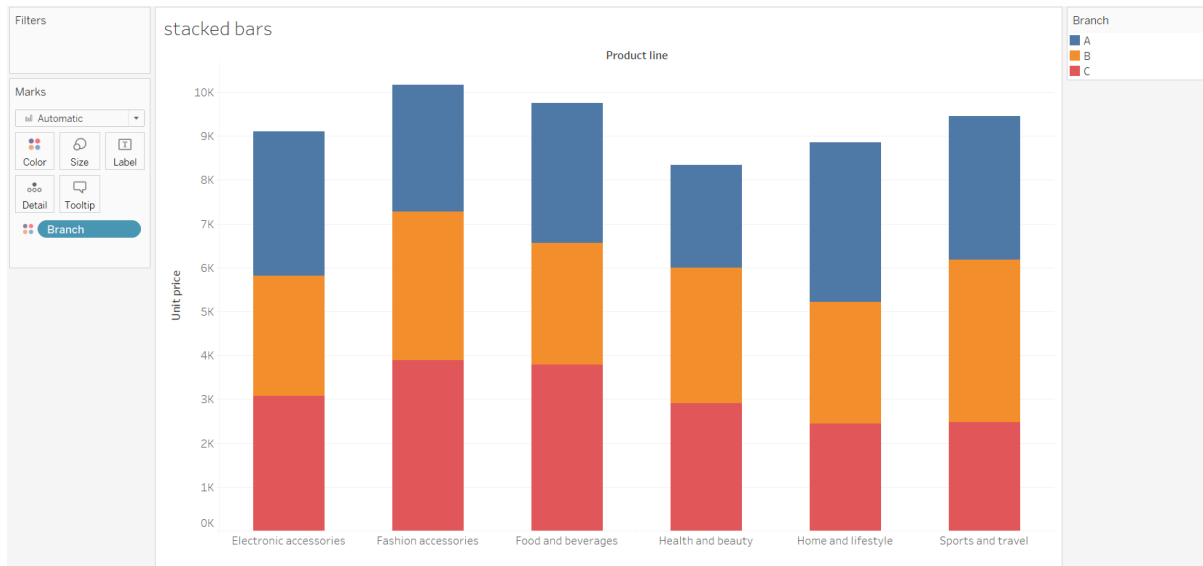
- Bar chart representing Branch Distribution by rating



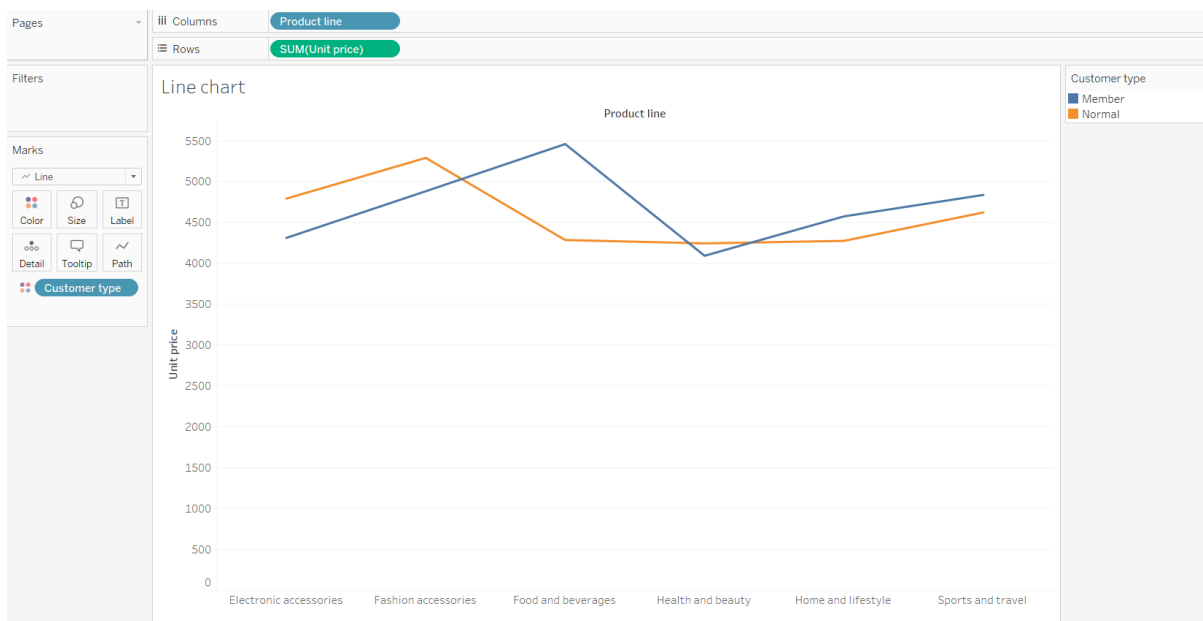
- Pie chart representing Quantity Distribution by week of the days



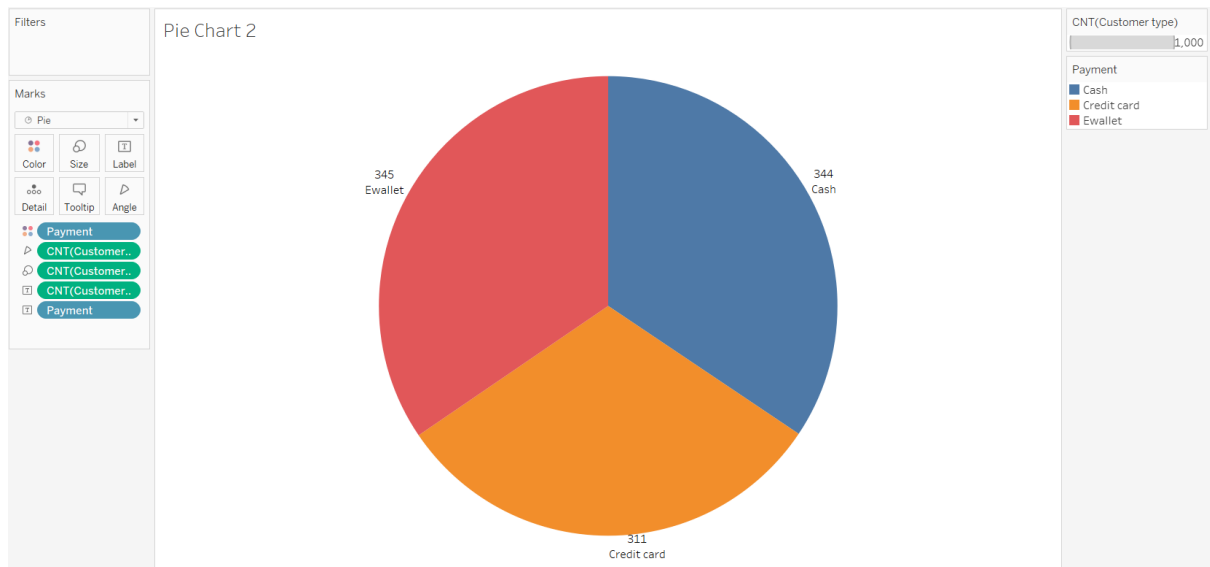
- Stacked Bars representing Product Distribution by Unit price



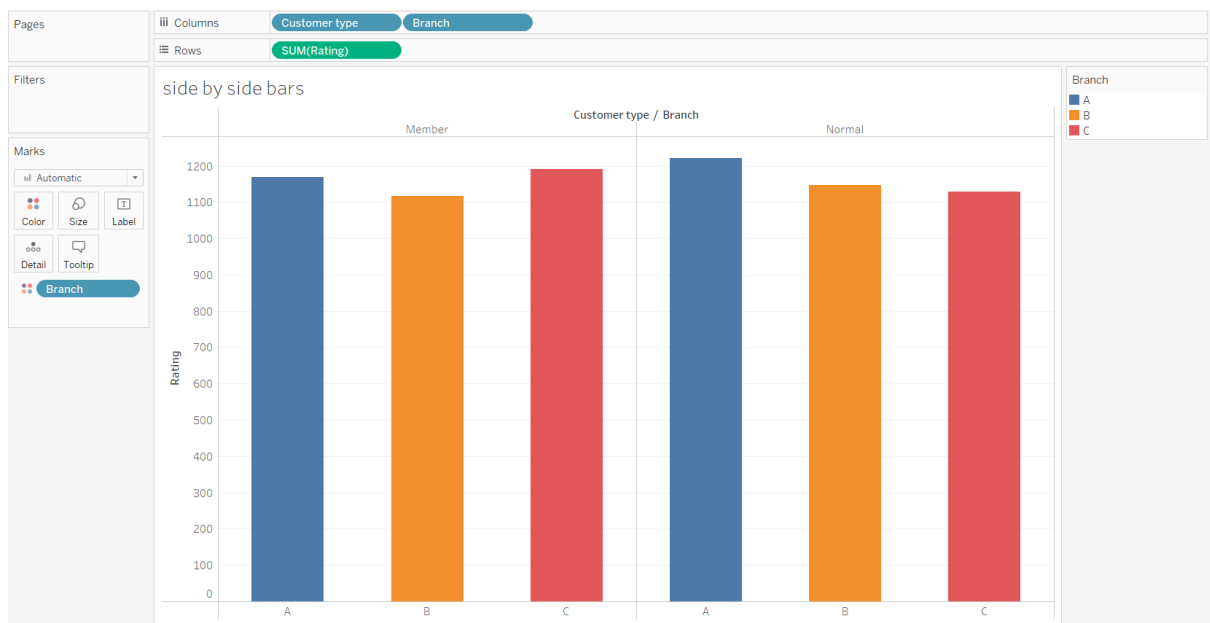
- Line chart representing Profit line Distribution by rating



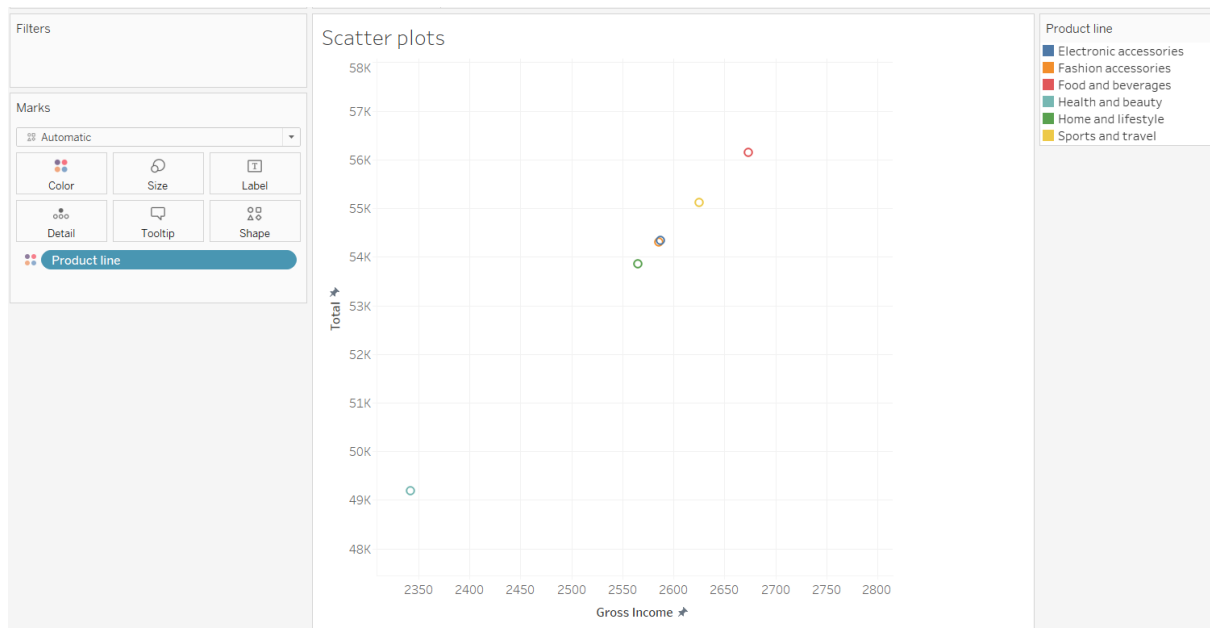
- Pie chart representing Payment method Distribution by Customer type



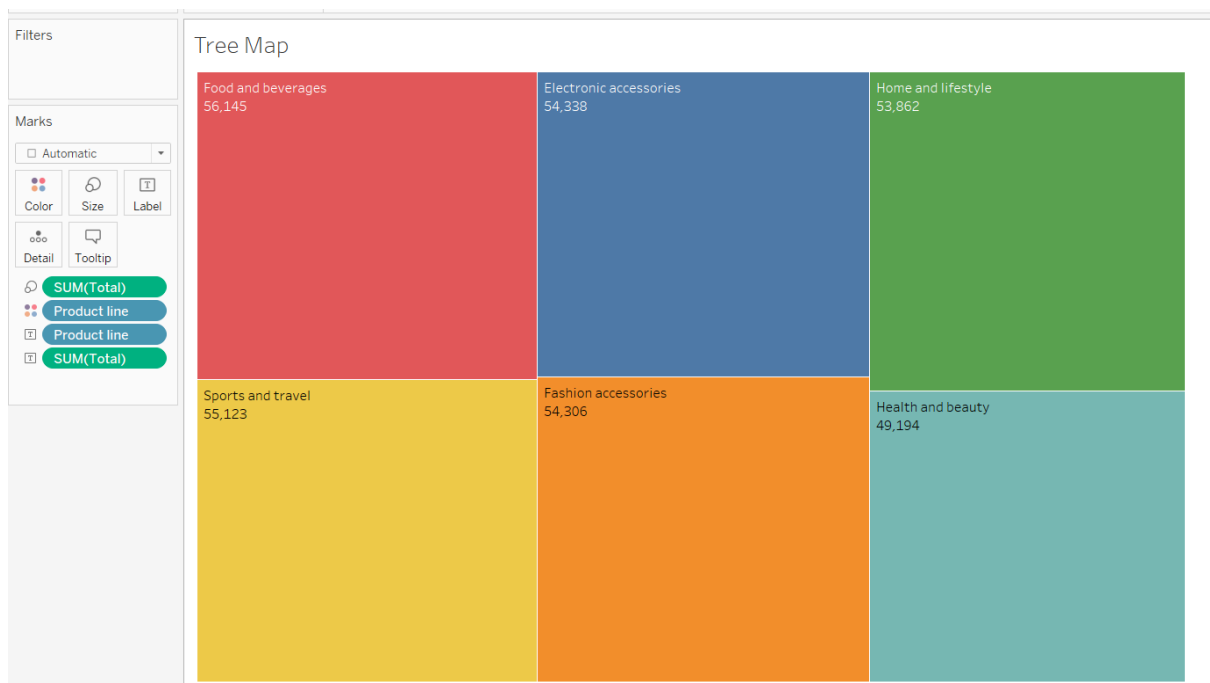
- Side by side Bars representing Branch and Customer Distribution by rating



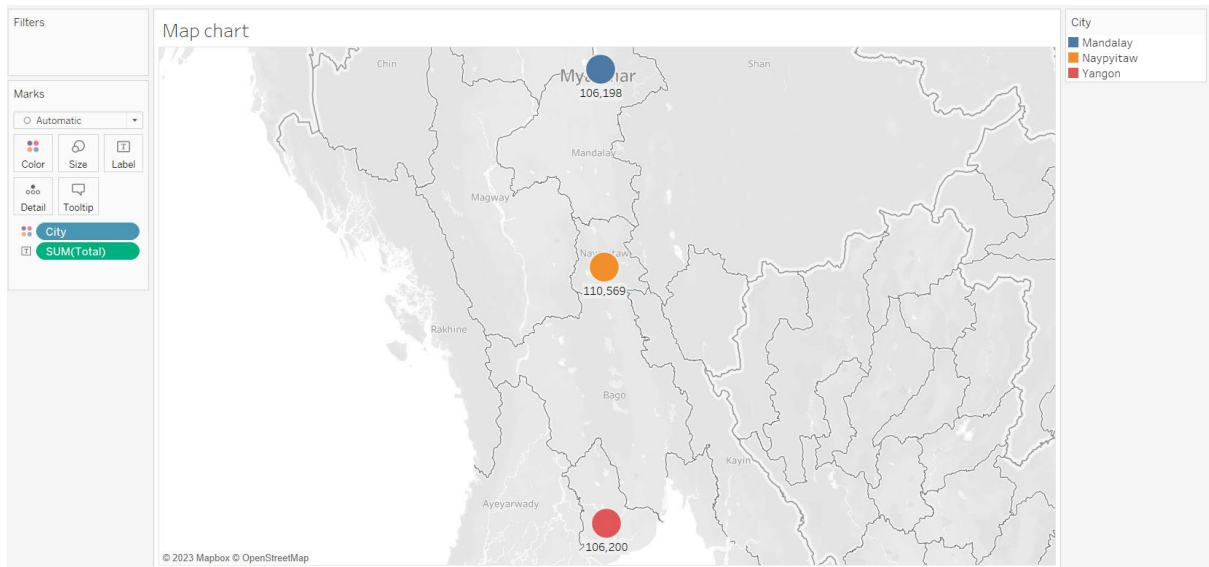
- Scatter plots representing Product Distribution by Grocss Income and Total



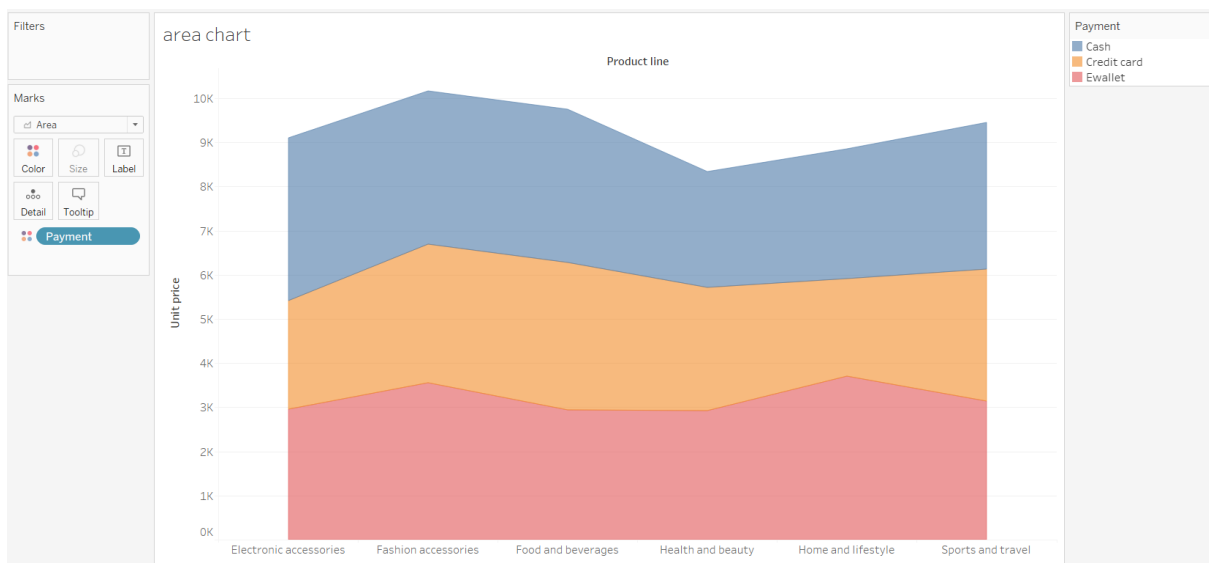
- Tree Map representing Product Distribution by Total sales



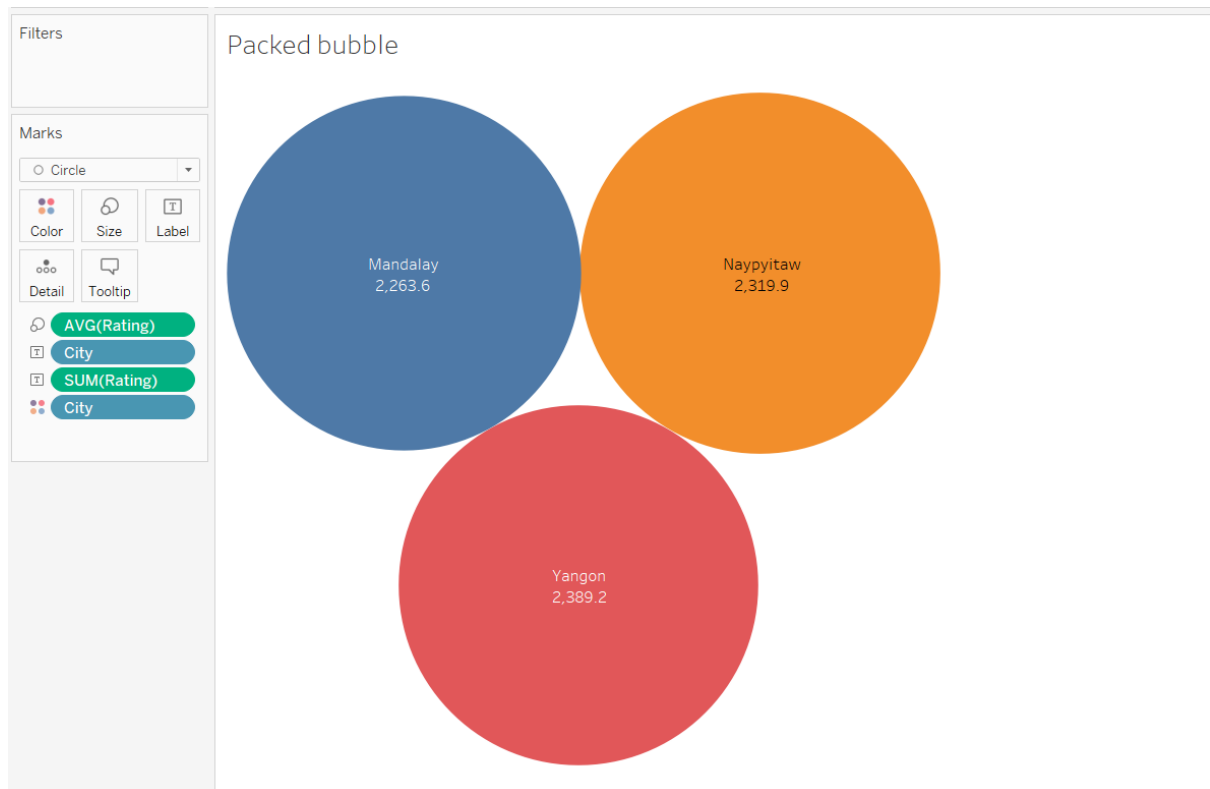
- Map chart representing Sales Distribution by City



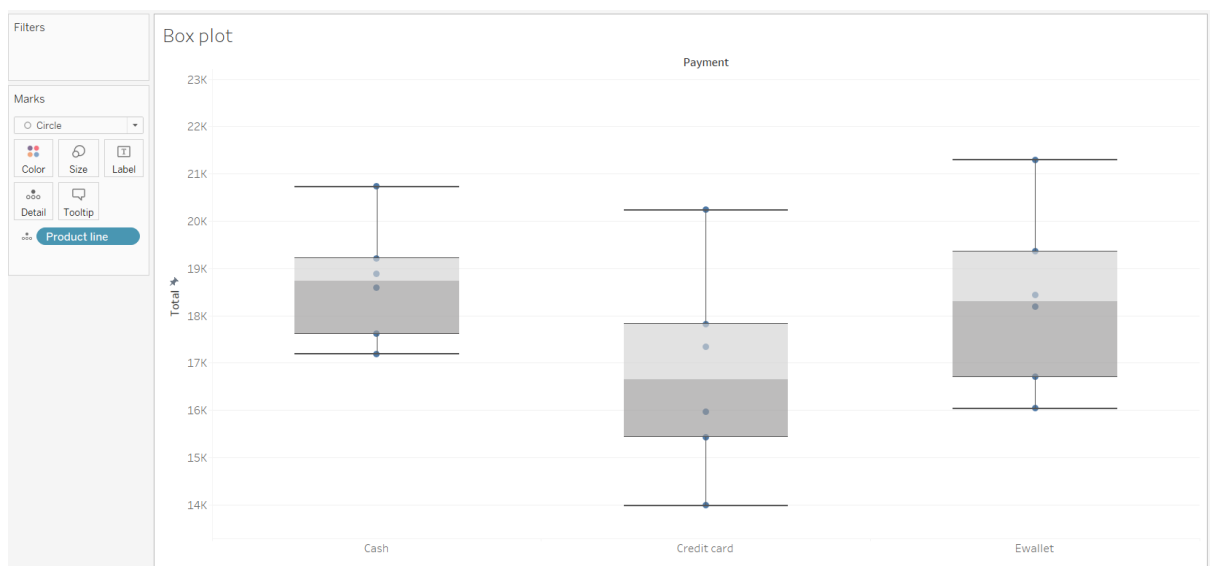
- Area chart representing Payment and Product Distribution by Unit price



- Packed Bubble representing Average Rating Distribution by City



- Box plot representing Payment method Distribution by Total Sales



■ Bubble Chart representing Product line Distribution by Average Quantity

