

Name: Ushasree J

Reg.No: 21BPS1491

Campus: VIT Chennai

DA Assignment 1

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)

■ Creating a Supermarket_sales Data Module

| Invoice ID | Branch | City | Customer type | Gender | Product line | Unit price |
|-------------|--------|-----------|---------------|--------|------------------------|------------|
| 750-67-8428 | A | Yangon | Member | Female | Health and beauty | 74.69 |
| 226-31-3081 | C | Naypyitaw | Normal | Female | Electronic accessories | 15.28 |
| 631-41-3108 | A | Yangon | Normal | Male | Home and lifestyle | 46.33 |
| 123-19-1176 | A | Yangon | Member | Male | Health and beauty | 58.22 |
| 373-73-7910 | A | Yangon | Normal | Male | Sports and travel | 86.31 |
| 699-14-3026 | C | Naypyitaw | Normal | Male | Electronic accessories | 85.39 |
| 355-53-5943 | A | Yangon | Member | Female | Electronic accessories | 68.84 |
| 315-22-5665 | C | Naypyitaw | Normal | Female | Home and lifestyle | 73.56 |
| 665-32-9167 | A | Yangon | Member | Female | Health and beauty | 36.26 |
| 692-92-5582 | B | Mandalay | Member | Female | Food and beverages | 54.84 |
| 351-62-0822 | B | Mandalay | Member | Female | Fashion accessories | 14.48 |
| 529-56-3974 | B | Mandalay | Member | Male | Electronic accessories | 25.51 |
| 365-64-0515 | A | Yangon | Normal | Female | Electronic accessories | 46.95 |

■ Removing Unwanted Columns (Invoice ID)

| Invoice ID | Branch | City |
|-------------|--------|-----------|
| 750-67-8428 | A | Yangon |
| 226-31-3081 | C | Naypyitaw |
| 631-41-3108 | A | Yangon |
| 123-19-1176 | A | Yangon |
| 373-73-7910 | A | Yangon |
| 699-14-3026 | C | Naypyitaw |
| 355-53-5943 | A | Yangon |
| 315-22-5665 | C | Naypyitaw |
| 665-32-9167 | A | Yangon |

- Splitting Date column into 4 Different columns (i.e Year , Month, day, day of the week)

IBM Cognos Analytics * Supermarket_s ... data module

Grid Relationships Custom tables

Search

Supermarket_s ...

- # Row Id
- abc Branch
- City
- abc Customer type
- abc Gender
- abc Product line
- Unit price
- Quantity
- Tax 5%
- Total
- Date

| Total | Date | Time | Pay |
|----------|------------|----------|-----|
| 548.9715 | 2019-01-05 | | Ewa |
| 80.22 | 2019-03-08 | | Cas |
| 340.5255 | 2019-03-03 | | Cre |
| 489.048 | 2019-01-27 | | Ewa |
| 634.3785 | 2019-02-08 | | Ewa |
| 627.6165 | 2019-03-25 | | Ewa |
| 433.692 | 2019-02-25 | | Ewa |
| 772.38 | 2019-02-24 | | Ewa |
| 76.146 | 2019-01-10 | 17:15:00 | Cre |

Filter...
Create navigation path...
Split...
Hide from users
Remove
Format data...
Clean...
Sort descending
Sort ascending
Properties

IBM Cognos Analytics * Supermarket_s ... data module

Grid Relationships Custom tables

Search

Supermarket_s ...

- # Row Id
- abc Branch
- City
- abc Customer type
- abc Gender
- abc Product line
- Unit price
- Quantity
- Tax 5%
- Total
- Date
- Year
- Month
- Day
- Day of the week
- Time

| Date | Year | Month | Day | Day of the week |
|------------|------|-------|-----|-----------------|
| 2019-01-05 | 2019 | 1 | 5 | Saturday |
| 2019-03-08 | 2019 | 3 | 8 | Friday |
| 2019-03-03 | 2019 | 3 | 3 | Sunday |
| 2019-01-27 | 2019 | 1 | 27 | Sunday |
| 2019-02-08 | 2019 | 2 | 8 | Friday |
| 2019-03-25 | 2019 | 3 | 25 | Monday |
| 2019-02-25 | 2019 | 2 | 25 | Monday |
| 2019-02-24 | 2019 | 2 | 24 | Sunday |
| 2019-01-10 | 2019 | 1 | 10 | Thursday |
| 2019-02-20 | 2019 | 2 | 20 | Wednesday |
| 2019-02-06 | 2019 | 2 | 6 | Wednesday |
| 2019-03-09 | 2019 | 3 | 9 | Saturday |
| 2019-02-12 | 2019 | 2 | 12 | Tuesday |

- Creating New Column (i.e Sales Column)
- Calculation : $\text{Sales} = \text{Total} / (1 + \text{tax})$

IBM Cognos Analytics * Supermarket_s ... data module

Grid Relationships Custom tables

Search

customer type

- abc Gender
- abc Product line
- Unit price
- Quantity
- Tax 5%
- Total**
- 🕒 Date
- 🕒 Year
- 🕒 Month
- 🕒 order date
- abc Day of the week
- 🕒 Time

| ↑↓ | Tax 5% | Total |
|----|---------|----------|
| | 26.1415 | 548.9715 |
| | 3.82 | 80.22 |
| | 16.2155 | 340.5255 |
| | 23.288 | 489.048 |
| | 30.2085 | 634.3785 |
| | 29.8865 | 627.6165 |
| | 20.652 | 433.692 |
| | 36.78 | 772.38 |
| | 3.626 | 76.146 |
| | 8.226 | 172.746 |

Filter...
Create calculation...
Create data group...
Hide from users
Remove
Format data...
Clean...
Sort descending
Sort ascending
Properties

IBM Cognos Analytics * Supermarket_s ... data module

Grid Relationship

Search

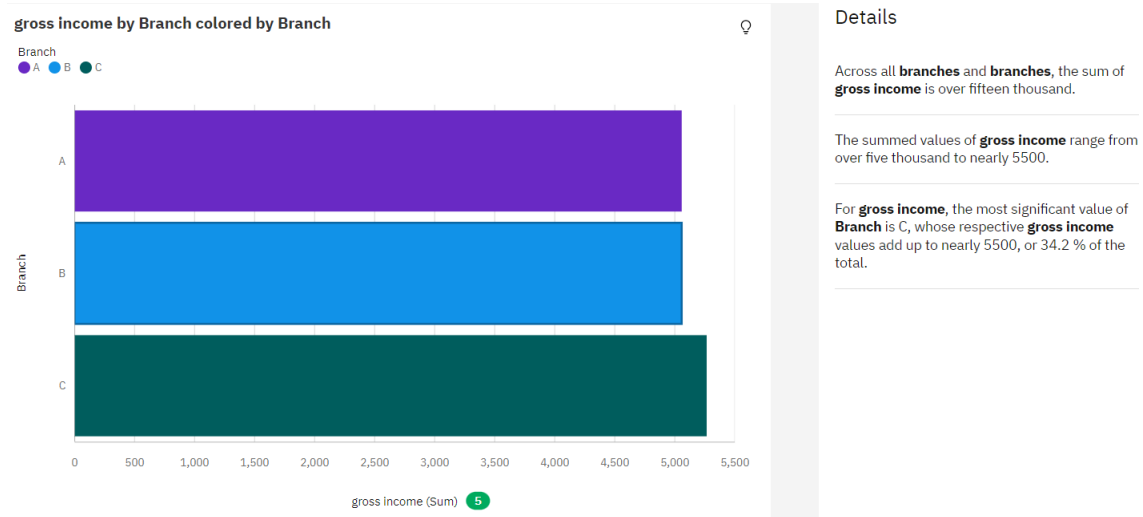
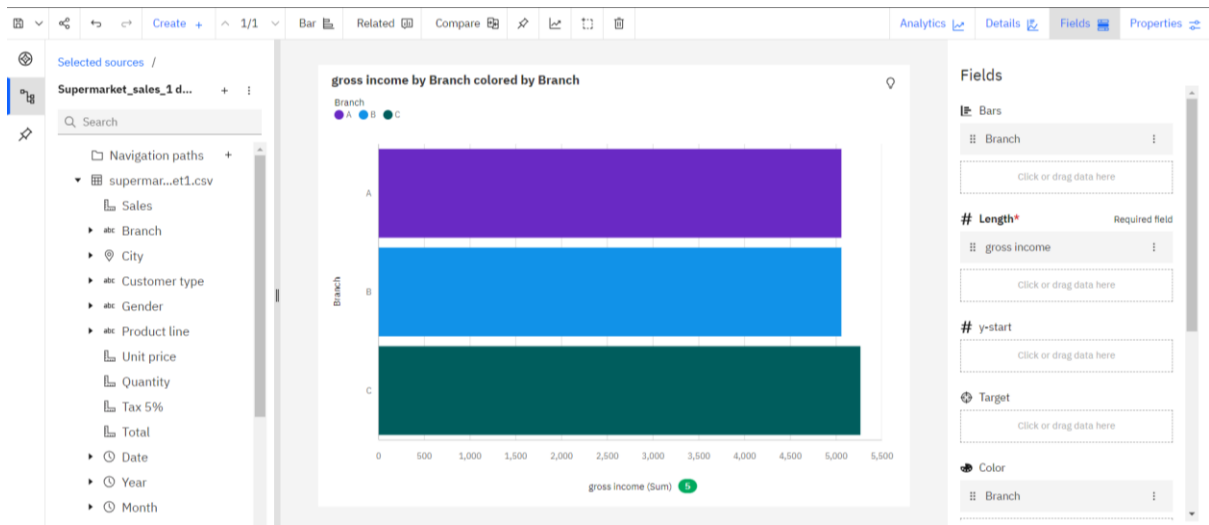
Sales

- # Row Id
- abc Branch
- 📍 City
- abc Customer type
- abc Gender
- abc Product line
- Unit price
- Quantity
- Tax 5%
- Total
- 🕒 Date
- 🕒 Year
- 🕒 Month
- 🕒 order date
- abc Day of the week

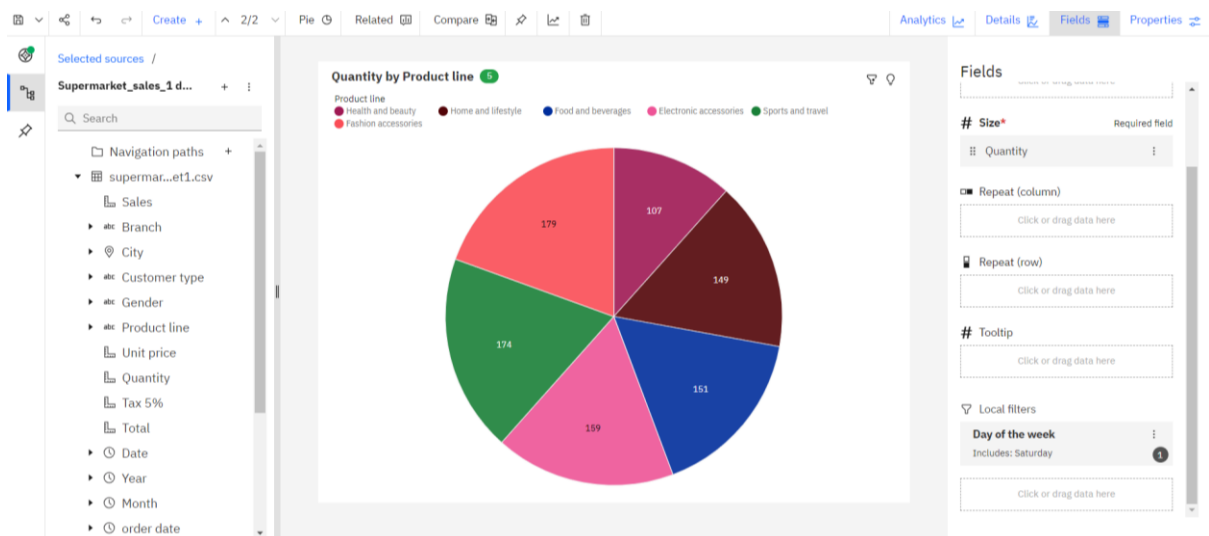
| ↑↓ | Sales |
|----|--------|
| | 576.11 |
| | 85.04 |
| | 357.74 |
| | 513.34 |
| | 665.59 |
| | 658.50 |
| | 455.34 |
| | 810.16 |
| | 80.77 |
| | 181.97 |
| | 64.71 |
| | 113.24 |
| | 259.23 |

Visualizations (and Analytics):

1. Branch Distribution: Bar chart

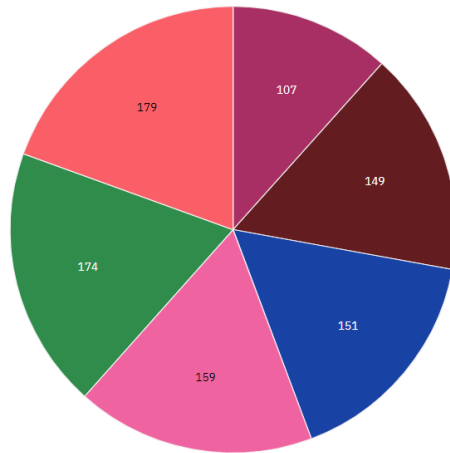


2. Product line Distribution(by Quantity): Pie Chart



Quantity by Product line 5

Product line
● Health and beauty ● Home and lifestyle ● Food and beverages ● Electronic accessories ● Sports and travel
● Fashion accessories



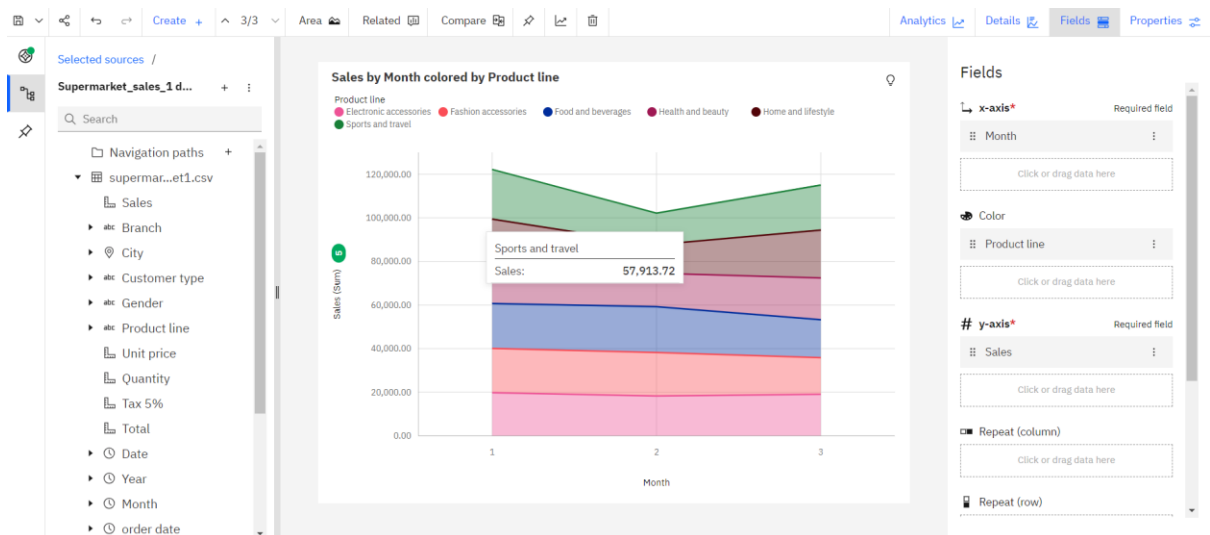
Details

Over all **product lines**, the sum of **Quantity** is 919.

Quantity ranges from 107, when **Product line** is Health and beauty, to 179, when **Product line** is Fashion accessories.

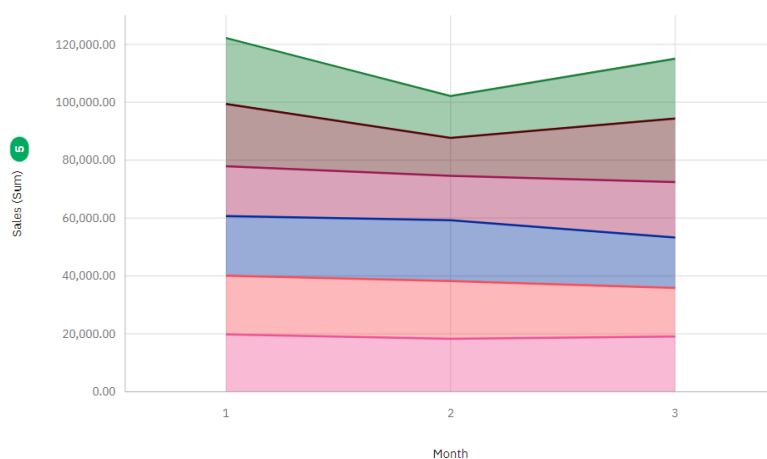
For **Quantity**, the most significant values of **Product line** are Fashion accessories, Sports and travel, Electronic accessories, Food and beverages, and Home and lifestyle, whose respective **Quantity** values add up to 812, or 88.4 % of the total.

3. Total Sales Over Time(months): Area Plot



Sales by Month colored by Product line

Product line
● Electronic accessories ● Fashion accessories ● Food and beverages ● Health and beauty ● Home and lifestyle ● Sports and travel



Details

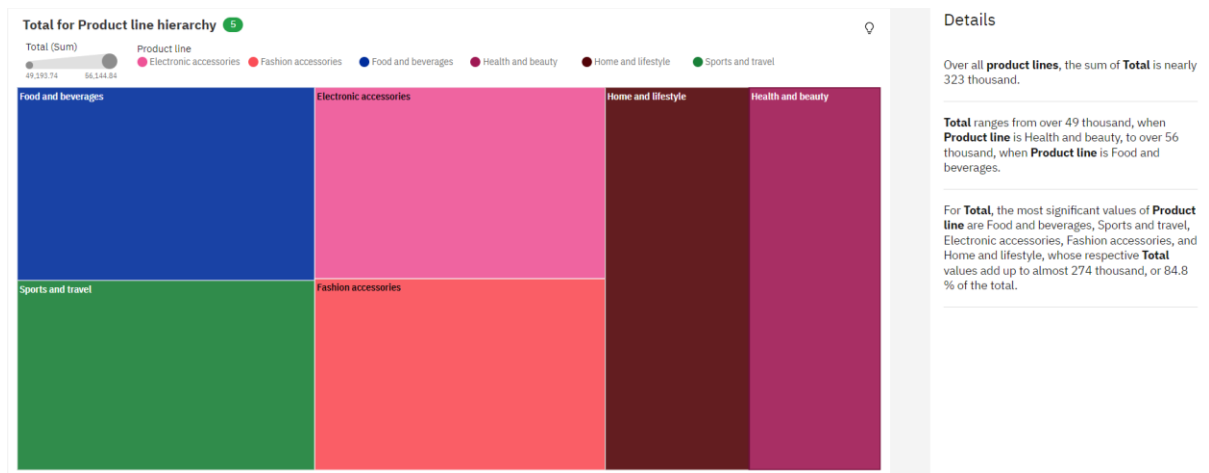
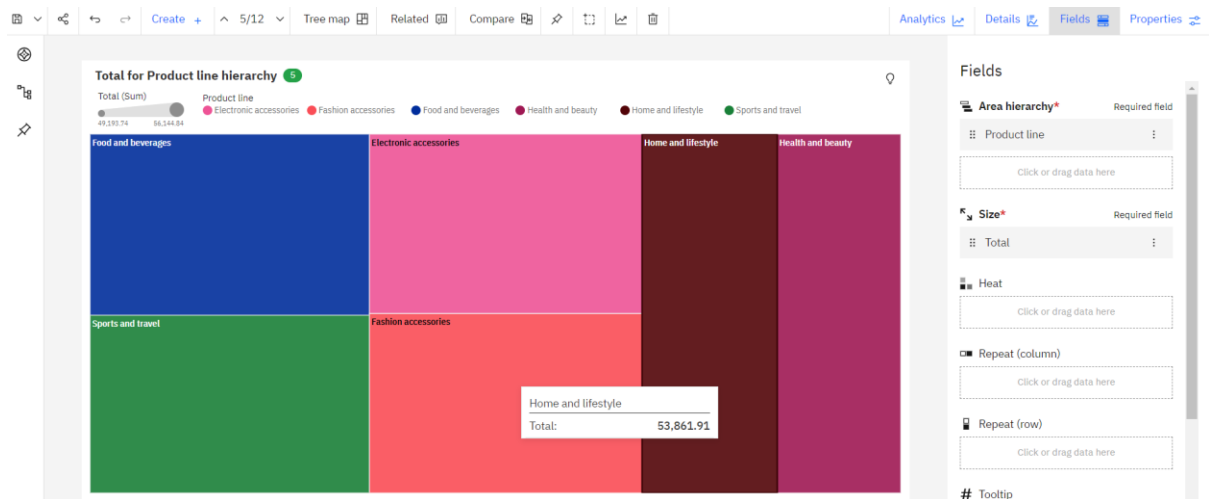
Across all **months** and **product lines**, the sum of **Sales** is over 339 thousand.

The summed values of **Sales** range from over 13 thousand to almost 23 thousand.

For **Sales**, the most significant values of **Month** are 1 and 3, whose respective **Sales** values add up to over 237 thousand, or 69.9 % of the total.

For **Sales**, the most significant values of **Product line** are Food and beverages, Sports and travel, Electronic accessories, Fashion accessories, and Home and lifestyle, whose respective **Sales** values add up to nearly 288 thousand, or 84.8 % of the total.

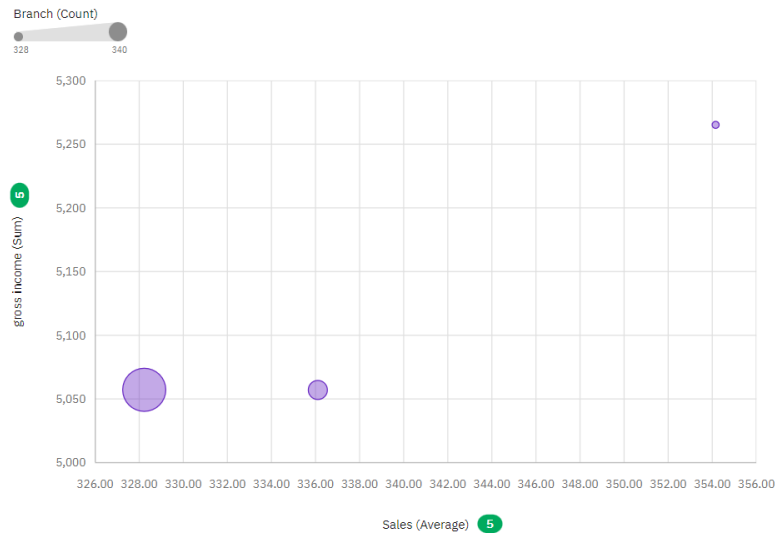
4. Product Line Distribution (by Total): Tree map



5. Total Sales by Branch: Bubble Chart



Sales by gross income with points for Branch



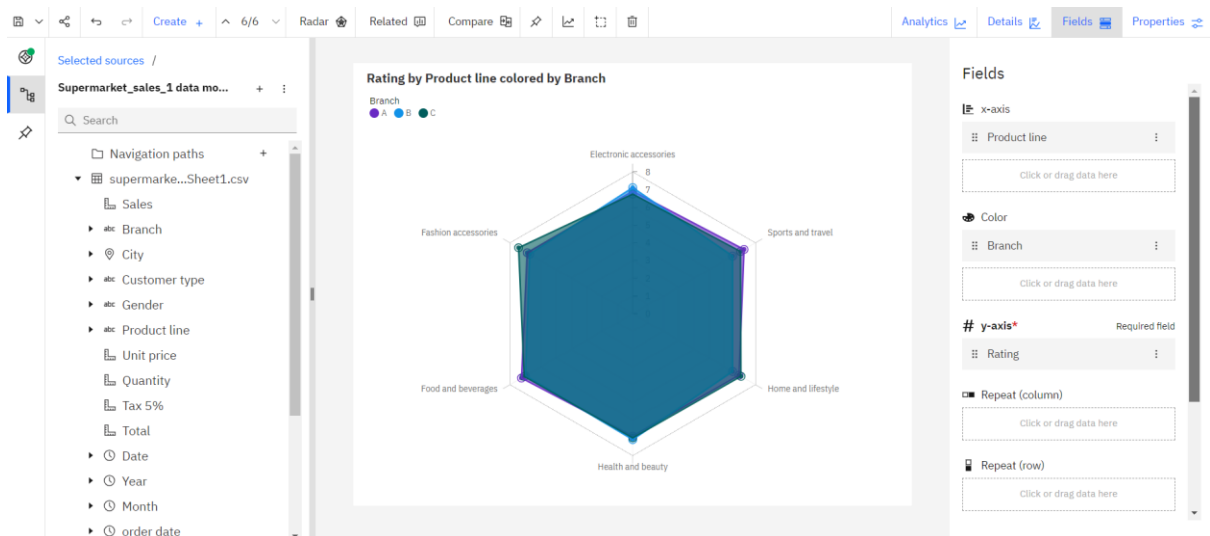
Details

The total number of results for **Sales**, across all **sales**, is a thousand.

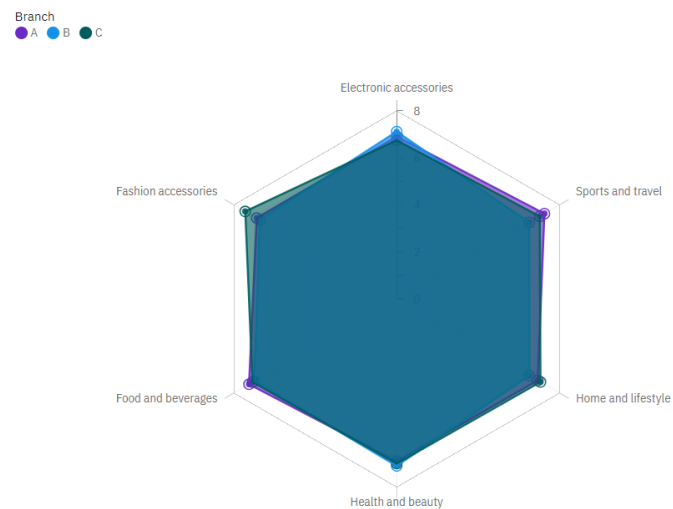
The total of **gross income** is over fifteen thousand.

gross income and **Sales** have a strong positive linear association, being gross income = 2215 + 8.577 * Sales

6. Customer Satisfaction Attributes: Radar Chart



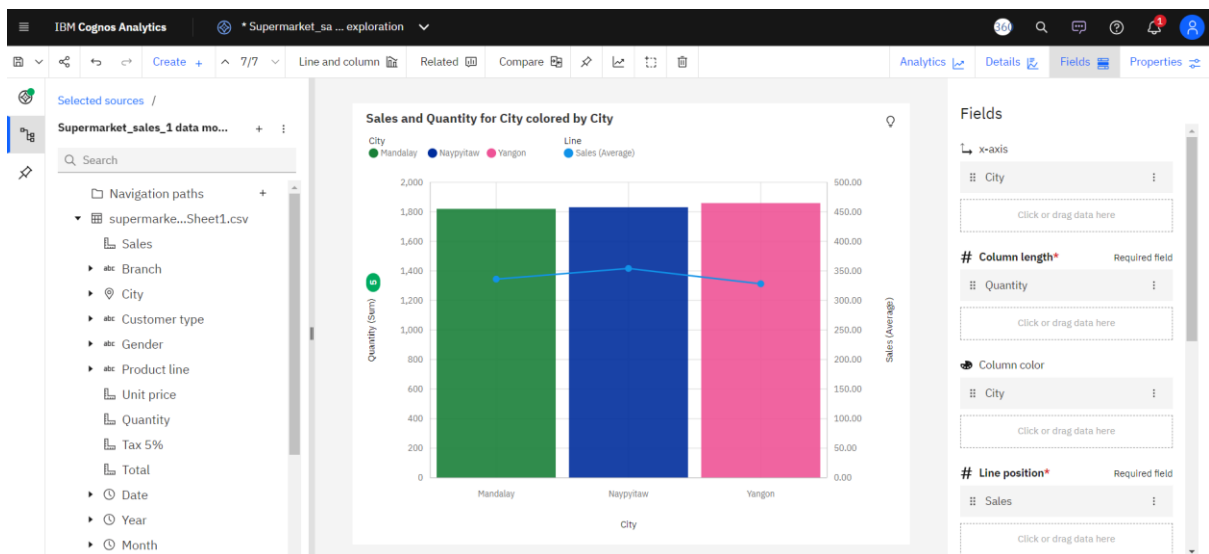
Rating by Product line colored by Branch



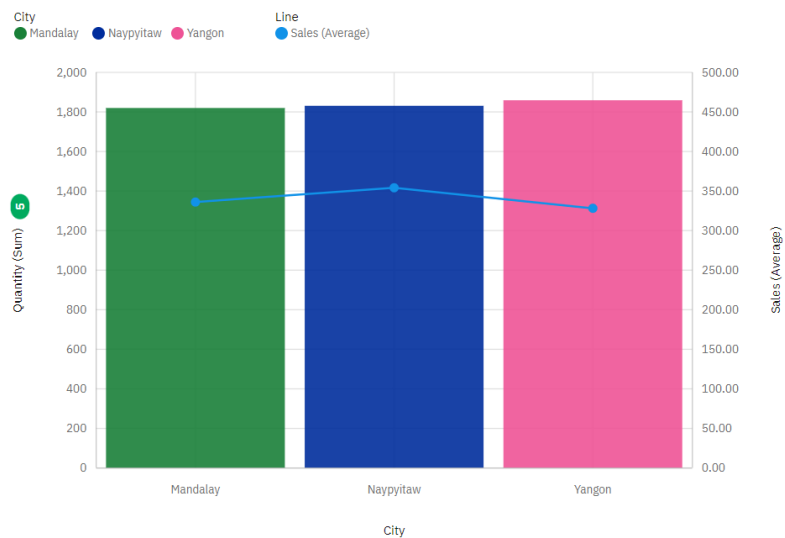
Details

The overall number of results for **Rating** is a thousand.

7. Sales and Quantity Comparison: Line and Column Chart



Sales and Quantity for City colored by City



Details

Across all **cities** and **cities**, the sum of **Quantity** is over 5500.

The summed values of **Quantity** range from 1820 to 1859.

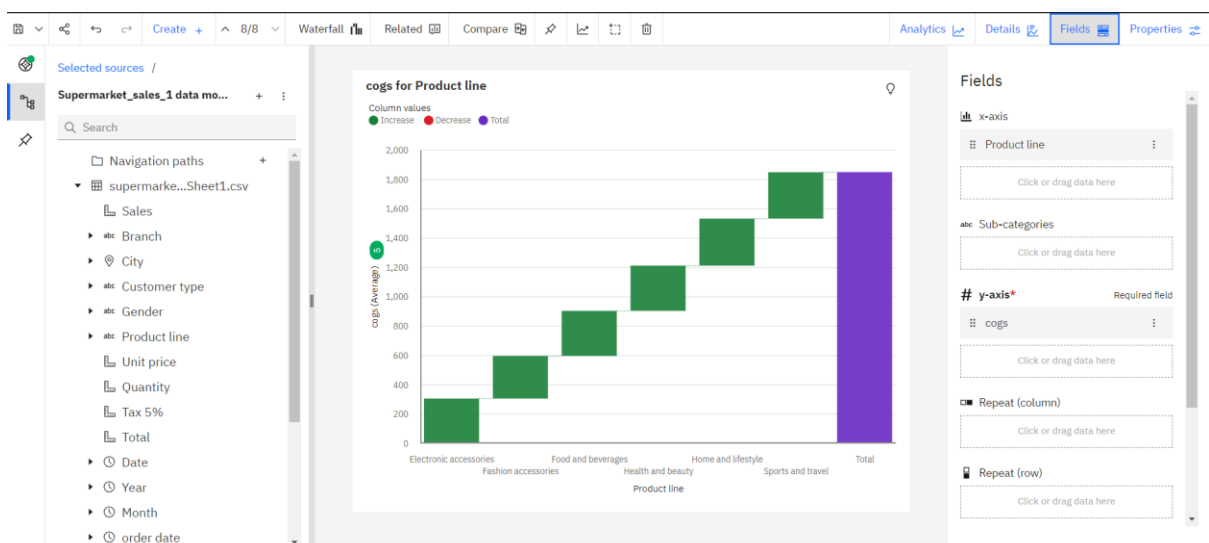
For **Quantity**, the most significant value of **City** is Yangon, whose respective **Quantity** values add up to nearly two thousand, or 33.7 % of the total.

Over all **cities**, the average of **Sales** is 339.3.

The average values of **Sales** range from 328.2, occurring in Yangon, to 354.2, in Naypyitaw.

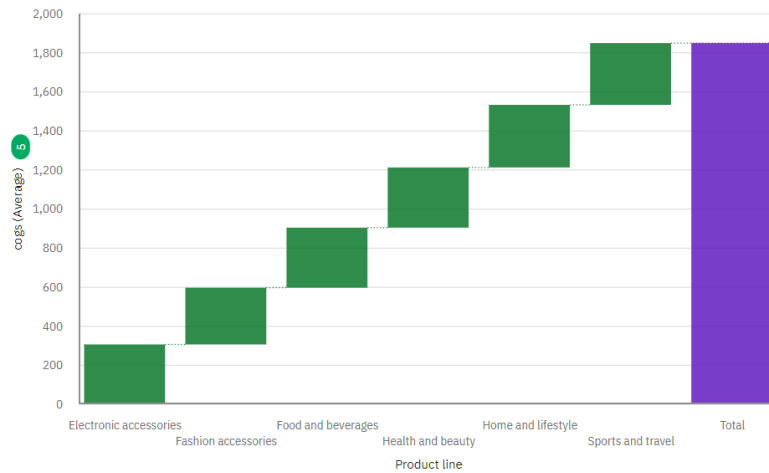
Yangon is the most frequently occurring category of **City** with a count of 340 items with **Sales** values (34 % of the total).

8. Cost of Goods Sold(COGS) Breakdown: Waterfall Chart



cogs for Product line

Column values
 ● Increase ● Decrease ● Total



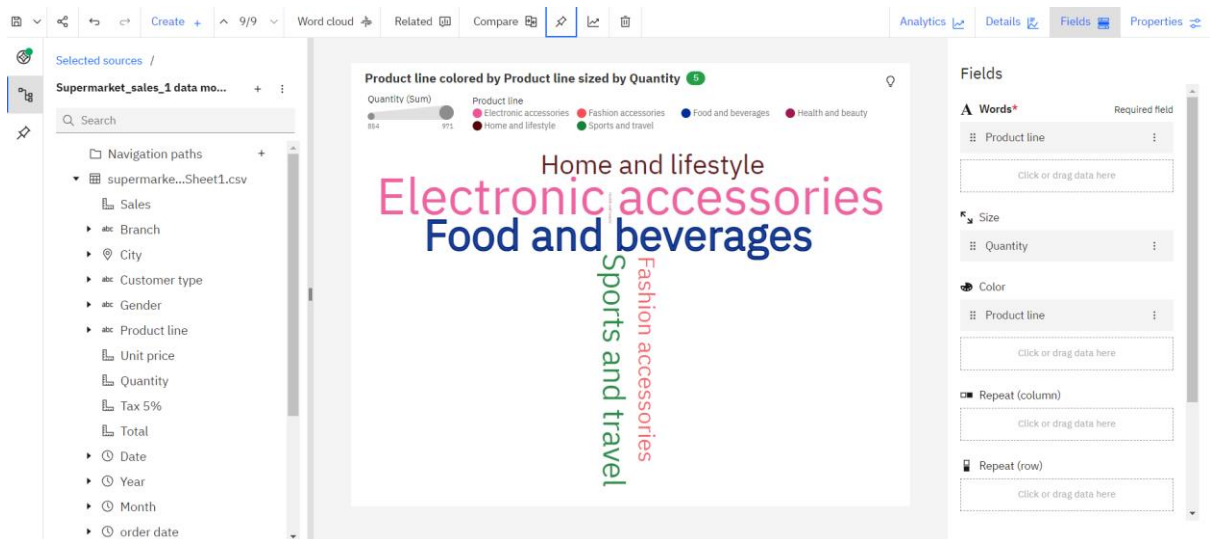
Details

Over all **product lines**, the average of **cogs** is 307.6.

The average values of **cogs** range from 290.6, occurring when **Product line** is Fashion accessories, to 320.6, when **Product line** is Home and lifestyle.

Fashion accessories (17.8 %), Food and beverages (17.4 %), Electronic accessories (17 %), Sports and travel (16.6 %), and Home and lifestyle (16 %) are the most frequently occurring categories of **Product line** with a combined count of 848 items with **cogs** values (84.8 % of the total) .

9. Product Line Keywords: Word Cloud



Product line colored by Product line sized by Quantity

Quantity (Sum)
 ● Electronic accessories ● Fashion accessories ● Food and beverages ● Health and beauty
 ● Home and lifestyle ● Sports and travel



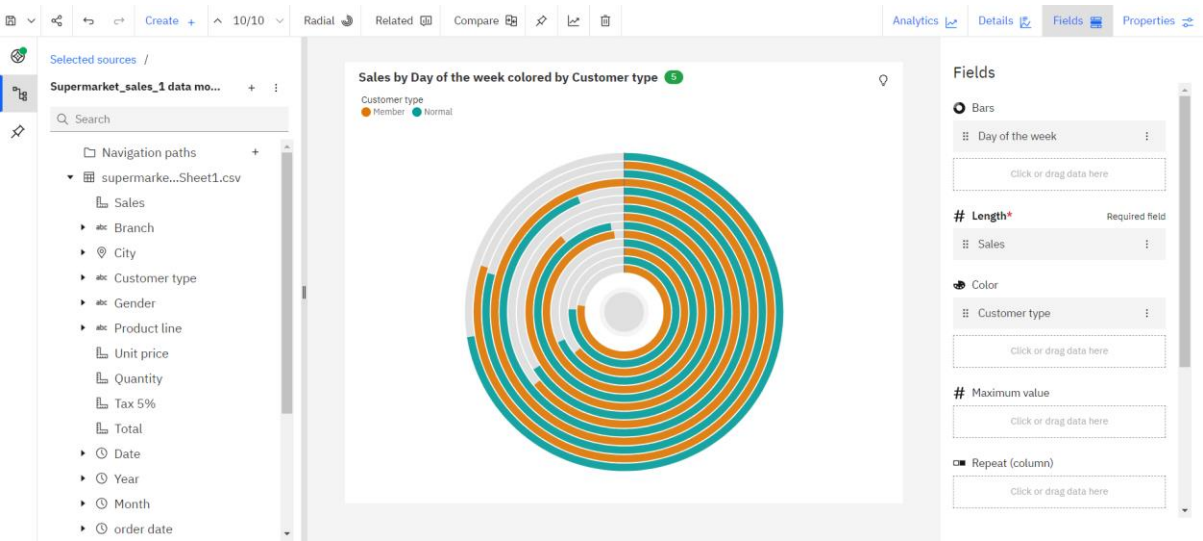
Details

Across all **product lines** and **product lines**, the sum of **Quantity** is over 5500.

The summed values of **Quantity** range from 854 to 971.

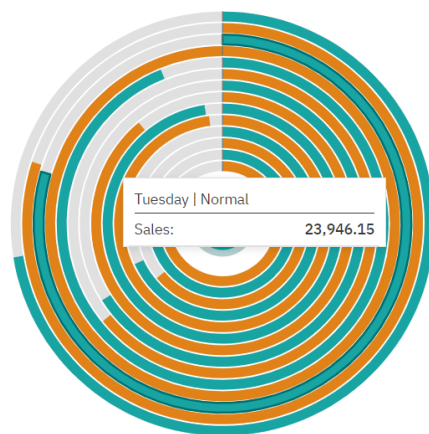
For **Quantity**, the most significant values of **Product line** are Electronic accessories and Food and beverages, whose respective **Quantity** values add up to nearly two thousand, or 34.9 % of the total.

10. Sales Distribution (by day of the week): Radial Chart



Sales by Day of the week colored by Customer type 5

Customer type
● Member ● Normal



Details

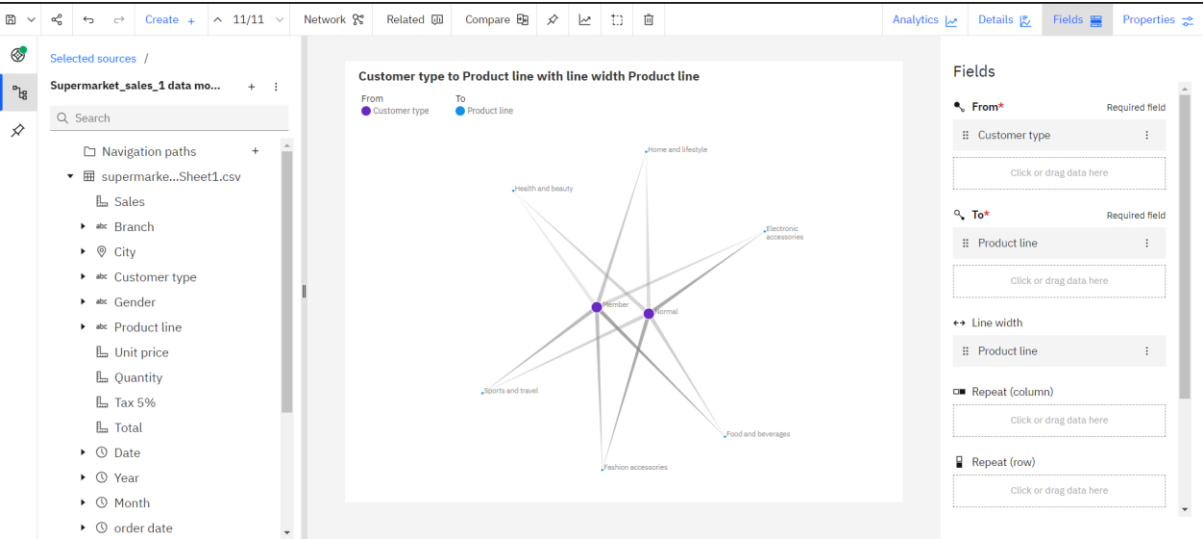
Across all **day of the weeks** and **customer types**, the sum of **Sales** is over 339 thousand.

The summed values of **Sales** range from over 19 thousand to over 30 thousand.

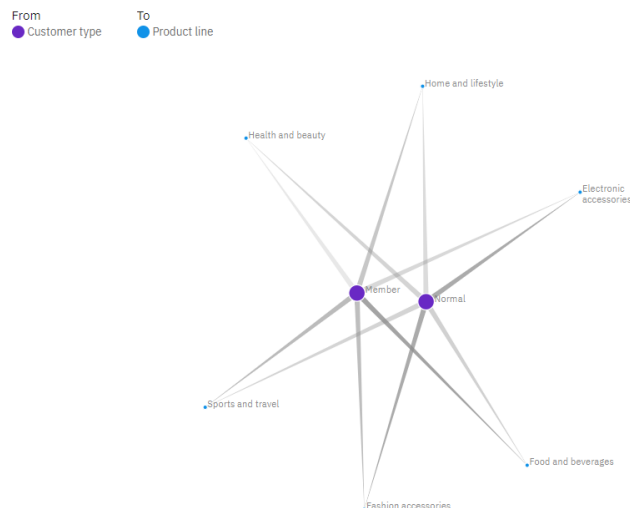
For **Sales**, the most significant value of **Customer type** is Member, whose respective **Sales** values add up to almost 173 thousand, or 50.8 % of the total.

For **Sales**, the most significant values of **Day of the week** are Saturday and Tuesday, whose respective **Sales** values add up to over 113 thousand, or 33.3 % of the total.

11. Customer-Product Interactions: Network Chart



Customer type to Product line with line width Product line



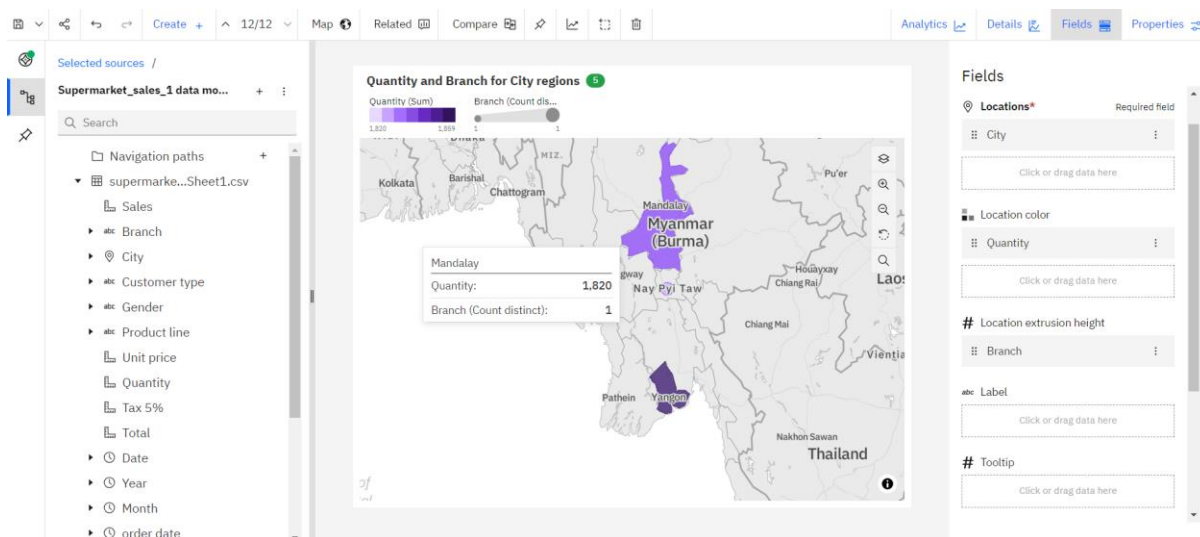
Details

The total number of results for **Product line**, across all **customer types**, is a thousand.

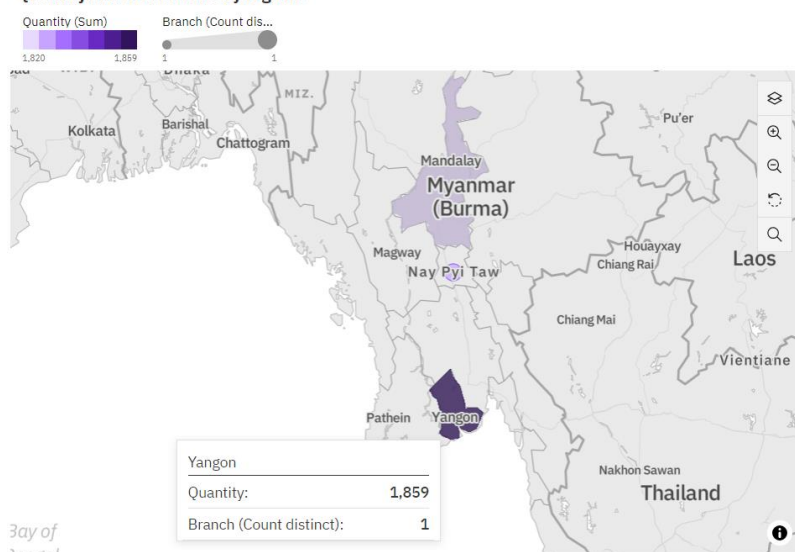
Fashion accessories (17.8 %), Food and beverages (17.4 %), Electronic accessories (17 %), Sports and travel (16.6 %), and Home and lifestyle (16 %) are the most frequently occurring categories of **Product line** with a combined count of 848 items with **Product line** values (84.8 % of the total) .

Member is the most frequently occurring category of **Customer type** with a count of 501 items with **Product line** values (50.1 % of the total).

12. Geographic Sales Distribution: Map Chart



Quantity and Branch for City regions



Details

Across all **cities**, the sum of **Quantity** is over 5500.

Quantity ranges from 1820, in Mandalay, to 1859, in Yangon.