

# OLAP Cube

Business Intelligence

# Typical Application of DWH

## Data Warehousing for Business Intelligence

- Stores **structured historical data** for analysis
- Enables businesses to identify **patterns & trends**

## From Raw Data to Meaningful Insights

- Data Mining** helps extract patterns from large datasets
- Multi-dimensional models** organize data for deeper analysis

ID	Name	Age	Income	Children	Car	Spent
12	Peter	45	€ 65,000	2	Mini Van	€ 210.00
15	Gabriel	28	€ 53,000	0	Coupe	€ 30.00
...	...	...	...	...	...	...
122	Claire	40	€ 52,000	1	Mini Van	€ 250.00



### Example Insight:

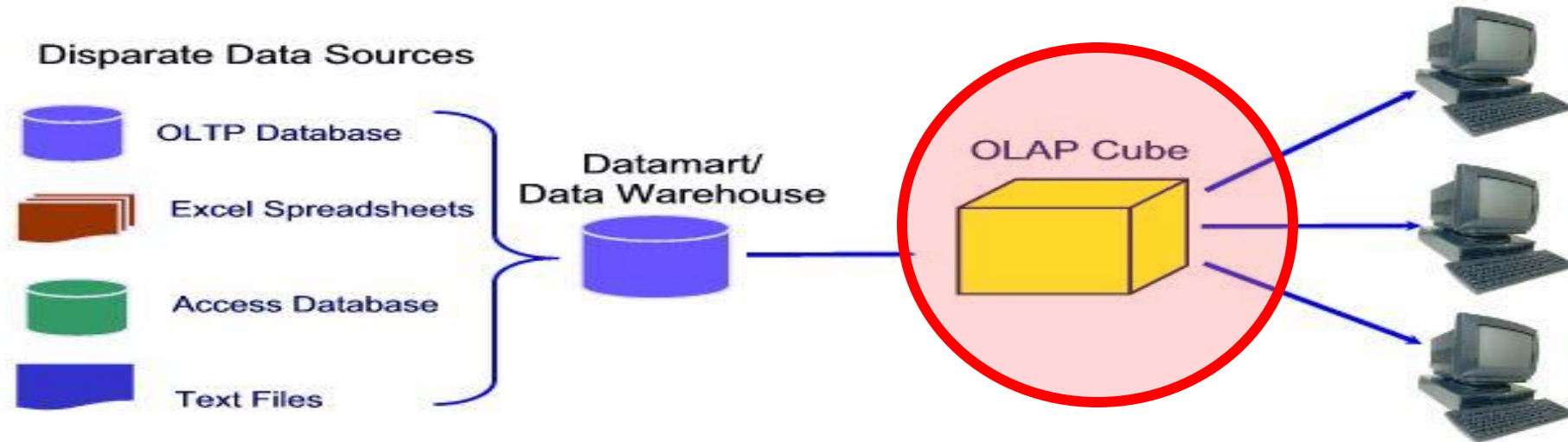
*"Customers **older than 35**, with **at least 1 child**, and driving a **minivan** tend to spend **more than €100** on groceries."*

# Why Multi-Dimensional Analysis?

- Businesses need to **analyze trends across different factors** (e.g., age, income, car type)
- A simple table is **not enough** to handle complex queries efficiently

# OLAP Cube

- OLAP (**O**n**L**ine **A**nalytical **P**rocessing)
- Represents **front-end analytics** based on a DW repository
- Enables **multi-dimensional** analysis
- Used for **reporting** and **strategic decision-making**

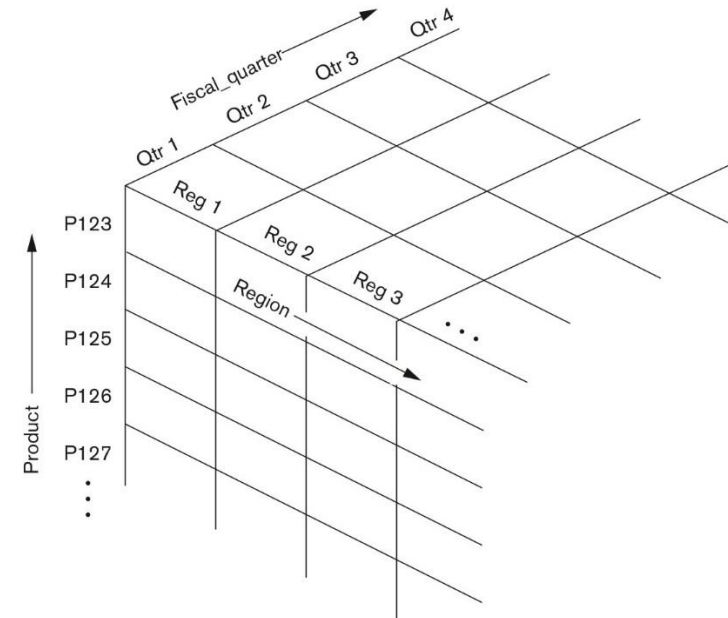


# Dimensional Analysis with OLAP Cubes

- Example of Two- Dimensional vs. Multi- Dimensional (3D typically called "Data Cube")

	Region			
	Reg 1	Reg 2	Reg 3	...
Product				
P123				
P124				
P125				
P126				
⋮				

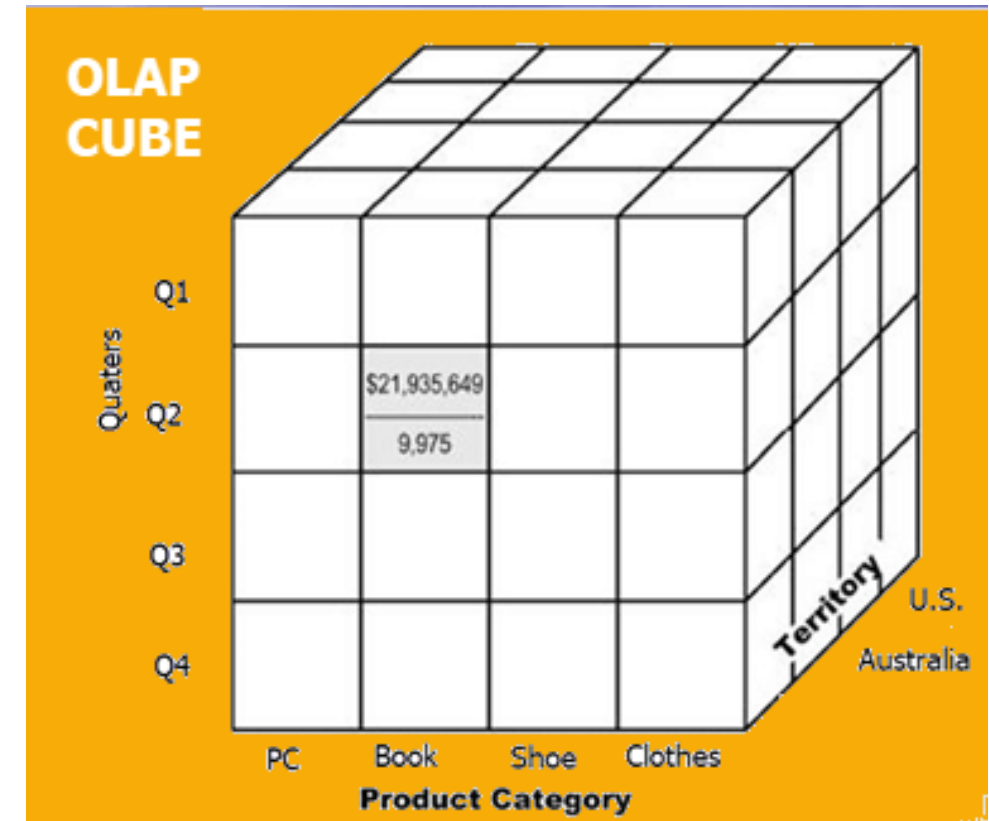
**Figure 29.2** A two-dimensional matrix model.



**Figure 29.3** A three-dimensional data cube model.

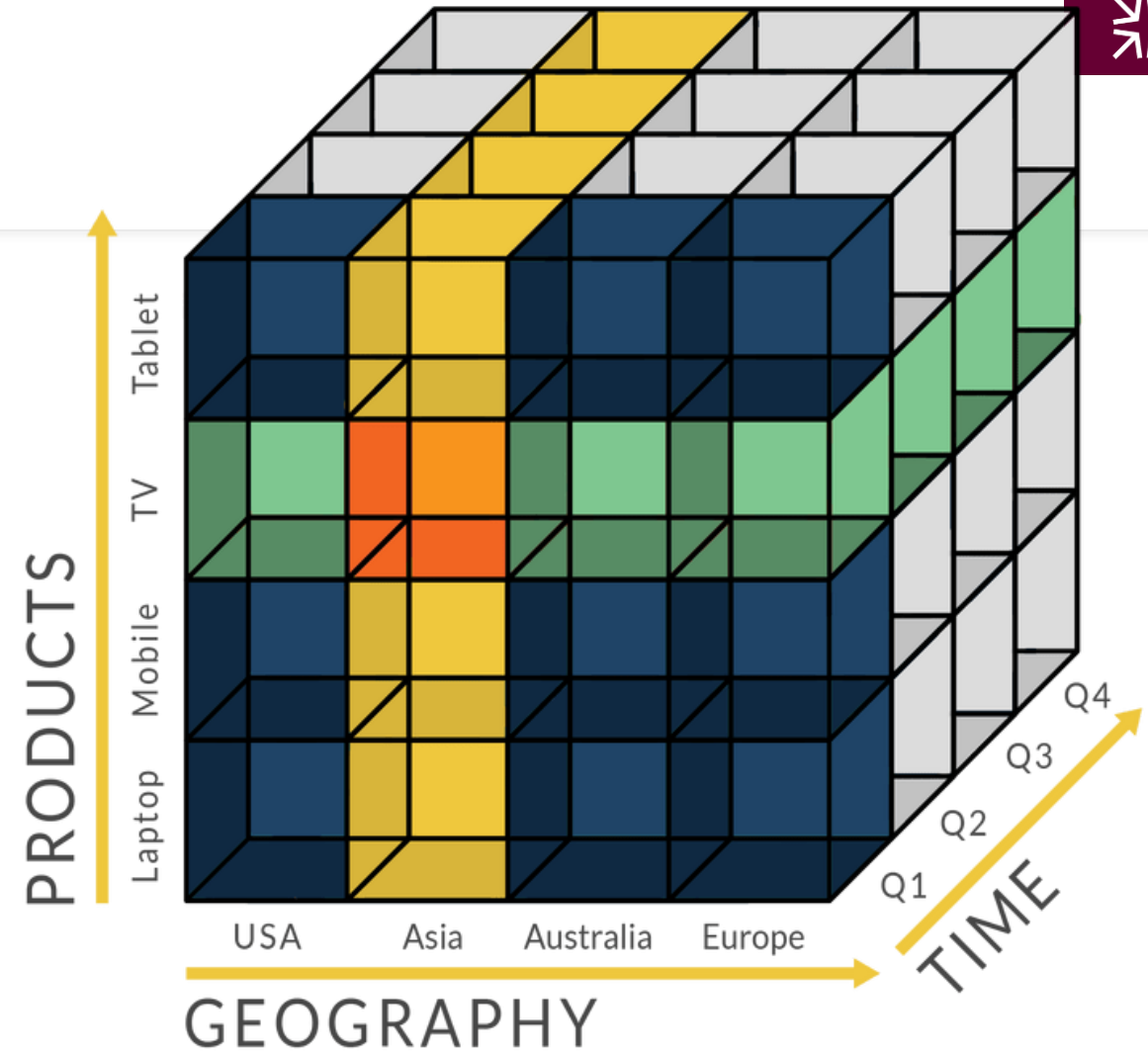
# OLAP Cube

- The OLAP Cube consists of numeric facts called measures which are categorized by dimensions.
- Theoretically no limit for the number of dimensions
- Typical cubes have 4-12 dimensions
- Also called Hypercube when more than 3-dimensions.
- The cube can store and analyze multidimensional data in a logical and orderly manner.



# OLAP Cube

- Structure:
  - N- dimensional Array
  - Nested Lists
- Essentially are generalized **group by** queries





A cube consists of **cells**



A given combination of dimension values



A cell can be empty (no data for this combination)



A **sparse** cube has few non-empty cells

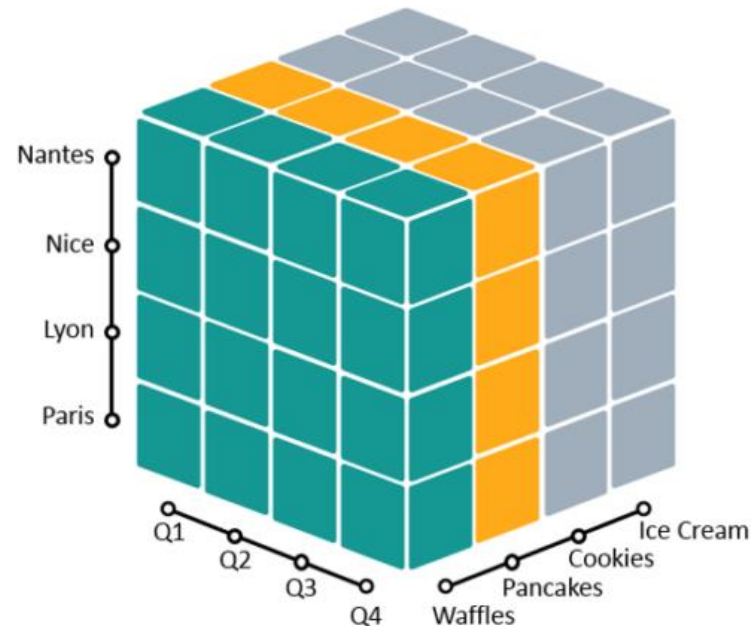


A **dense** cube has many non-empty cells



Cubes become sparser for many/large dimensions

### 3-dimensional Cube in Data Warehousing Example with dimensions Location, Quarter, Product categories

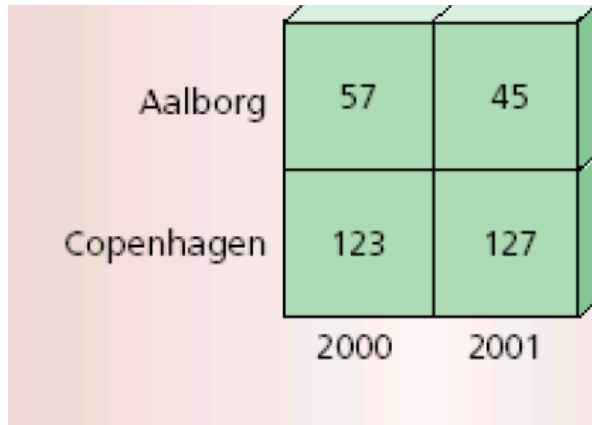


Example: You want to summarize sales by specific product, by the time, and by store location. These are data cube dimensions.

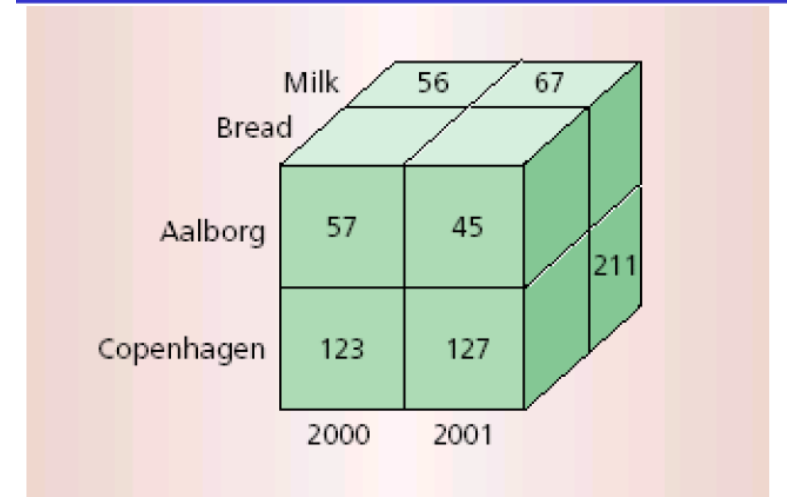


# OLAP cube

Store	Product	Time	Sales
Aalborg	Bread	2000	57
Aalborg	Milk	2000	56
Copenhagen	Bread	2000	123
...	...	...	...



Summarized by  
Store and Time  
*2-dimensional*

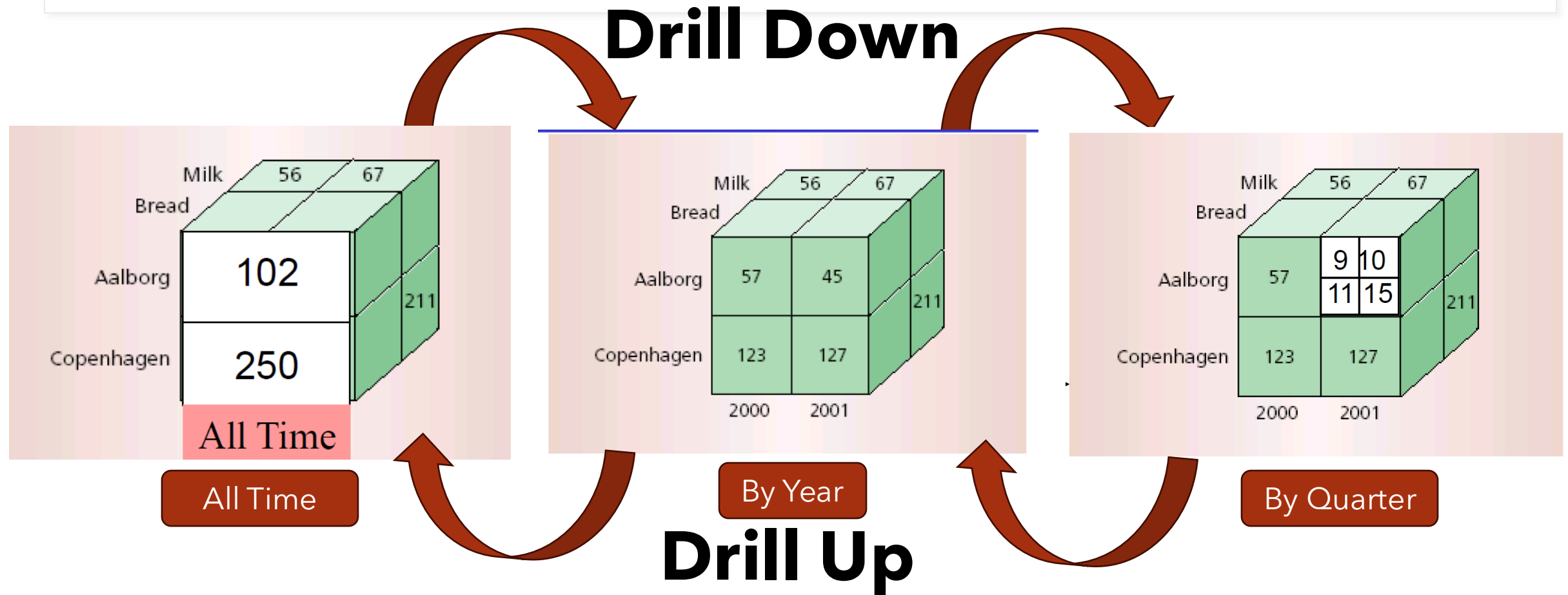


Summarized by  
Store, Time and Product  
*3-dimensional*

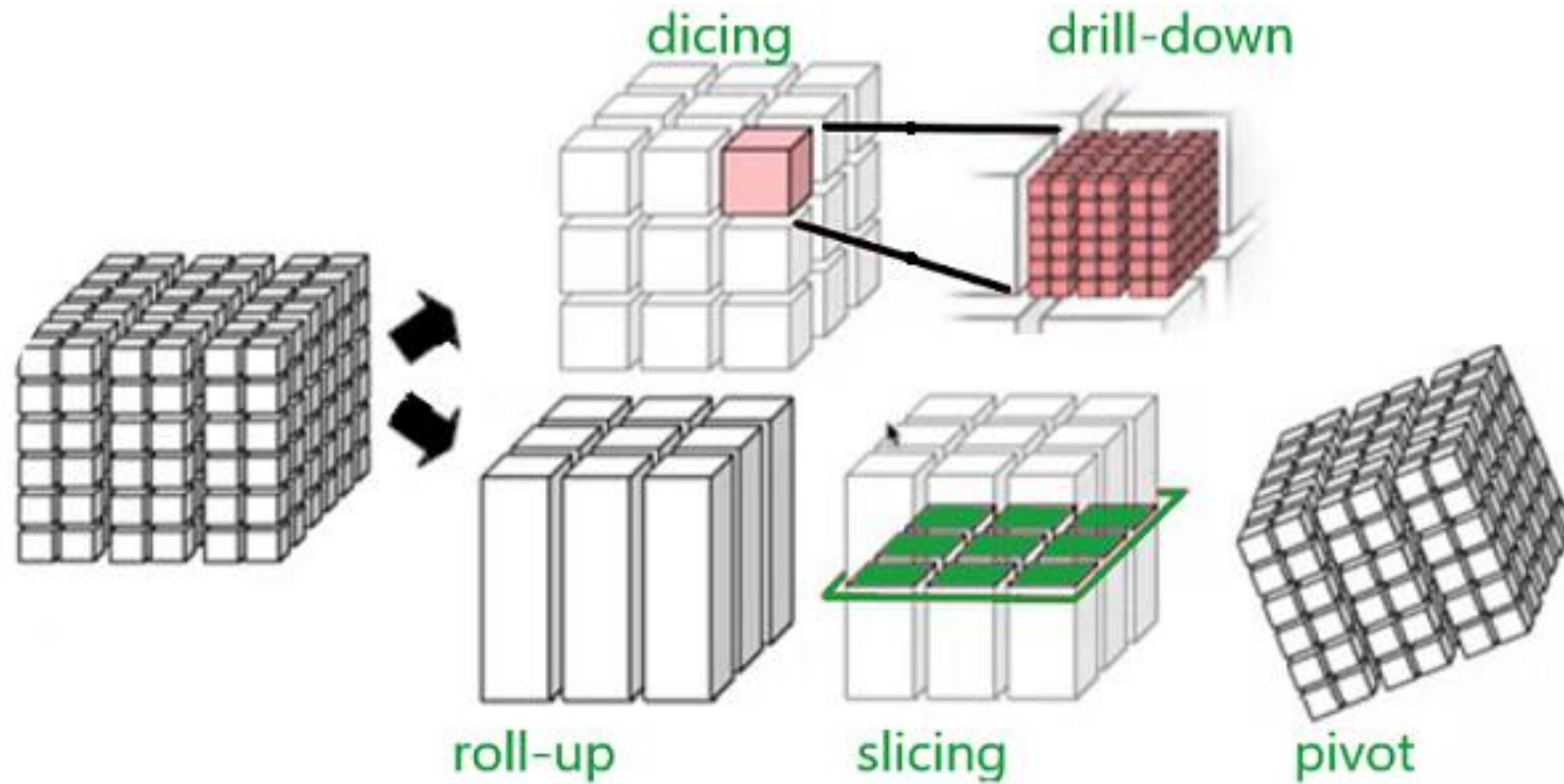
# Drill Up and Drill Down

- You can **drill down** and **drill up/ roll up** to explore different aspects of your business and move between levels of information.
- For example, you can examine revenue for an entire product line and then drill down to see revenue for each individual product in the line.
- When you finish viewing individual product revenue, you can drill back up.

- You can **drill down** and **drill up/ roll up** to explore different aspects of your business and move between levels of information.



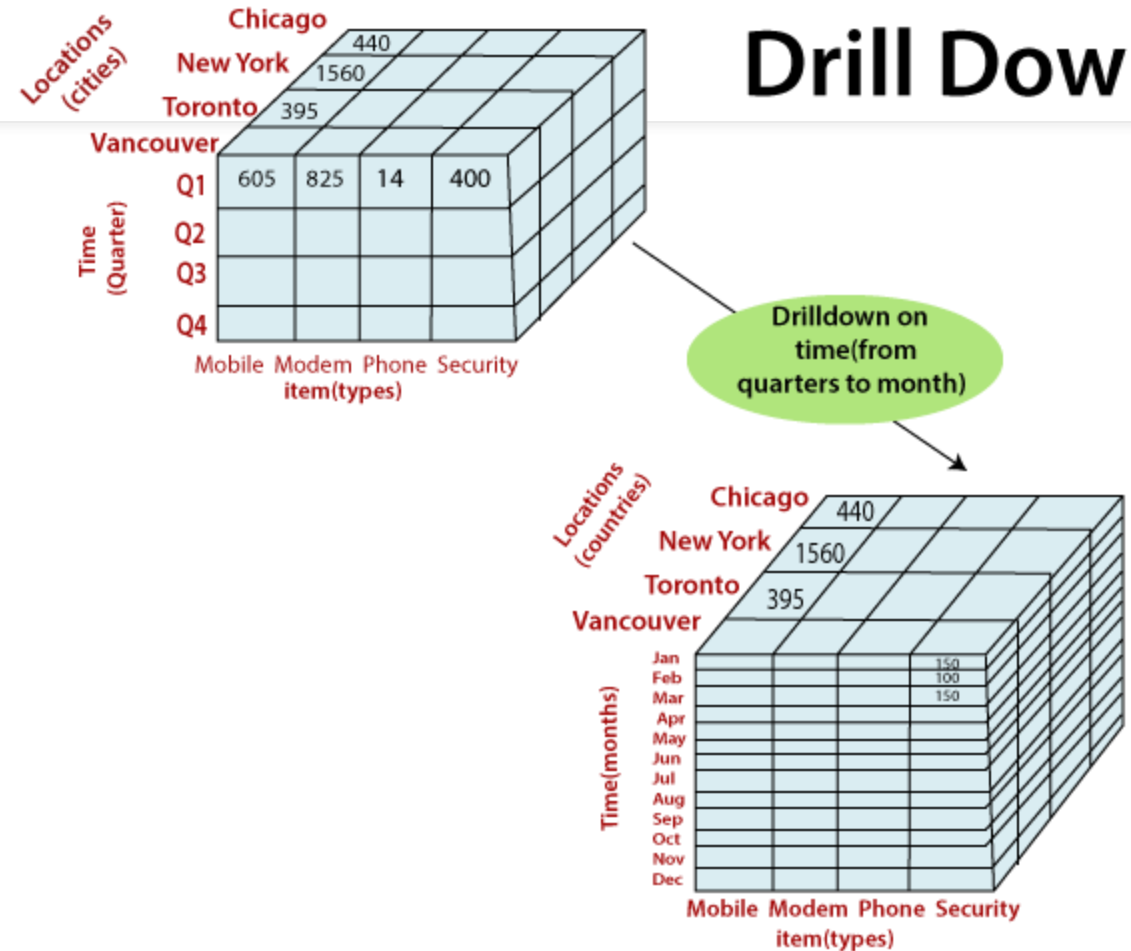
# OLAP Cube Operations



# Drill Down

- Drill-down is like zooming-in on the data cube. It navigates from less detailed record to more detailed data.

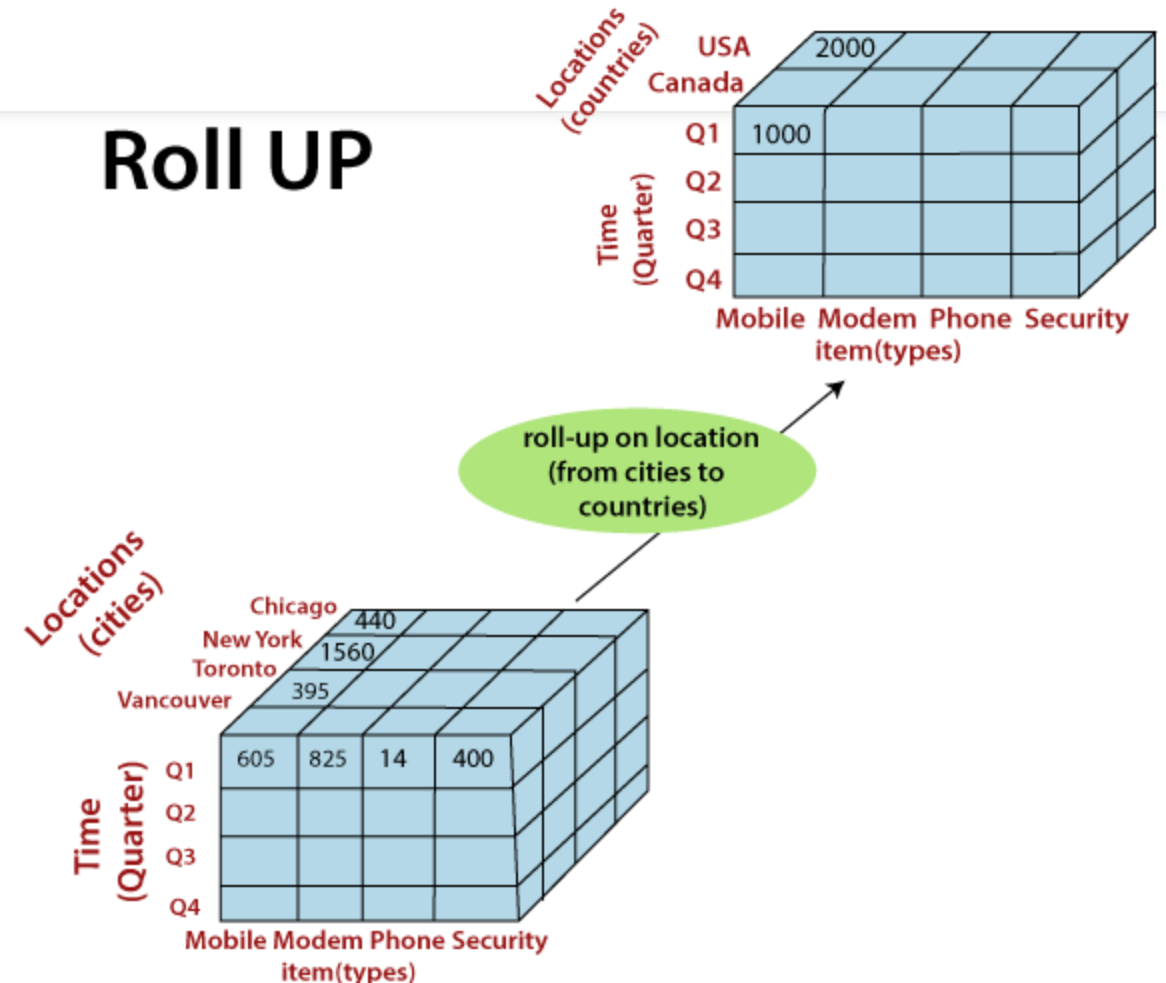
## Drill Down



# Roll Up

- The roll-up operation (**also known as drill-up or aggregation operation**) performs aggregation on a data cube, by climbing up the concept **hierarchies**, i.e., dimension reduction.
- Roll-up is like **zooming-out** on the data cubes.

## Roll UP



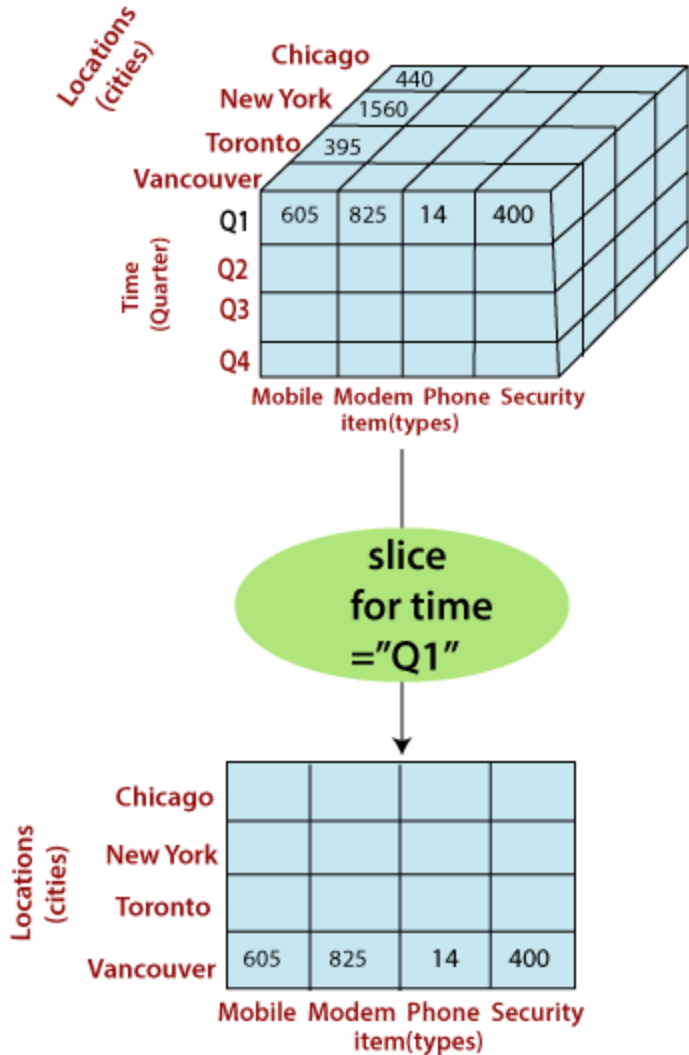
# Understanding Hierarchies

- A Hierarchy is a systematic way of organizing the Members of a Dimension into a logical tree structure that defines parent-child aggregation relationships.

## ***Examples***

- **Date Hierarchy:** Year → Qtr → Month → Day
- **Location Hierarchy:** Country → City → Area...
- **Product Hierarchy:** Category → Sub-category → Product

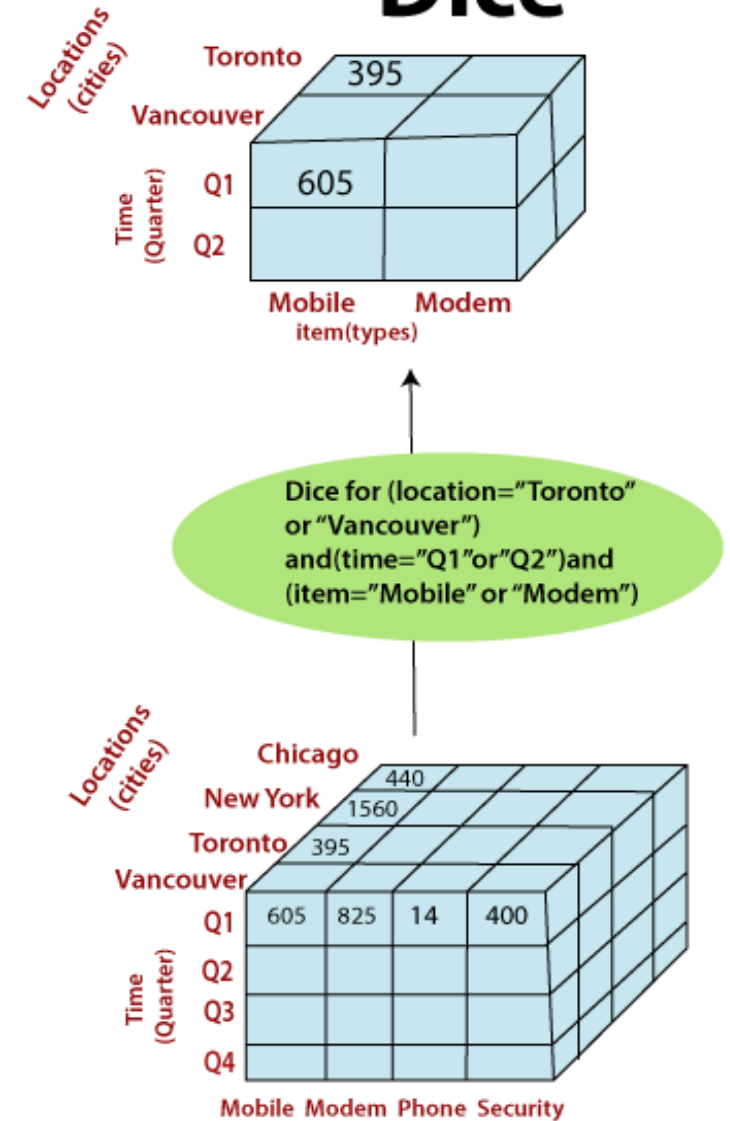
# Slice



## Slice and Dice:

- Applying filters on a single-dimension resulting in a 2-dimensional site (**slice**)
- Applying filters on 2 or more dimensions (**dice**)

# Dice



# Pivot

- The pivot operation is also called a rotation.
- Pivot is a visualization operations which *rotates* the data axes in view to provide an alternative presentation of the data.
- It may contain swapping the rows and columns or moving one of the row-dimensions into the column dimensions.

## Pivot

Locations  
(cities)

Chicago				
New York				
Toronto				
Vancouver	605	825	14	400

Mobile Modem Phone Security  
item (types)



Item  
(types)

Mobile				605
Modem				825
Phone				14
Security				400

Chicago New Toronto Vancouver  
York  
Location (cities)



# OLAP Cube on MS Excel

Class Activity

# Sales\_dataset.csv

Download the dataset and manual from LMS.  
Open a new MS Excel workbook and load the dataset.

Identify Facts and Dimensions.

# List of Columns in Sales.csv

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

# Facts and Dimensions

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

# Facts and Dimensions

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

This column has a single value which is **useless** for analysis hence, I can choose to delete this!

Acting as a  
Fact Table  
where  
Invoice ID is  
the PK

# Brainstorm Dimensional Queries

Think in perspective of a business problem

# Dimensional Queries (Examples)

1. Find the total sales by branch and customer type.
2. Find the lowest rated product lines. (Bottom 2)
3. Find the highest selling product lines by month
4. Total sales trend by month for each branch.
5. Daily total sales time-series trend for highest selling branch. Identify the day of highest sales.
6. Find the trends in gender-wise sales by identifying the customer expenditure for different product lines. Which product line is popular among each gender?

# Understanding the Queries

	Query	Dimension	Fact	Aggregate
1	Find the total sales by branch and customer type.	Branch, Customer type	Total	SUM
2	Find the lowest rated product lines.	Product lines	Rating	AVG
3	Find the highest selling product lines by month	Product line Month	Quantity	SUM
4	Total sales trend by month for each branch	Branch Month	Total	SUM
5	Daily total sales time-series trend for highest selling branch. Identify the day of highest sales.	Date	Total	SUM
6	Find the trends in gender-wise sales by identifying the customer expenditure for different product lines. Which product line is popular among each gender?	Product line Gender	Total	SUM

# Understanding the Queries

	Query	Dimension	Fact	Aggregate
1	Find the total sales by branch and customer type.	Branch, Customer type	Total	SUM
2	Find the lowest rated product lines.	Product lines	Rating	AVG
3	Find the highest selling product lines by month	Product lines Month	Total	SUM
4	Total sales trend by month for each branch	Branch Month	Total	SUM
5	Daily total sales time-series trend for highest selling branch. Identify the day of highest sales.	Date	Total	SUM
6	Find the trends in gender-wise sales by identifying the customer expenditure for different product lines. Which product line is popular among each gender?	Product line Gender	Total	SUM

# Understanding the Queries

	Query	Dimension	Fact	Aggregate
1	Find the total sales by branch and customer type.	Branch, Customer type	Total	SUM
2	Find the lowest rated product lines.	Product lines	Rating	AVG
3	Find the highest selling product lines by month	Product lines Month	Quantity	SUM
4	Total sales trend by month for each branch	Branch Month	Total	SUM
5	Daily total sales time-series trend for highest selling branch. Identify the day of highest sales.	Date	Total	SUM
6	Find the trends in gender-wise sales by identifying the customer expenditure for different product lines. Which product line is popular among each gender?	Product line Gender	Total	SUM

# Understanding the Queries

	Query	Dimension	Fact	Aggregate
1	Find the total sales by branch and customer type.	Branch, Customer type	Total	SUM
2	Find the lowest rated product lines.	Product lines	Rating	AVG
3	Find the highest selling product lines by month	Product lines Month	Quantity	SUM
4	Total sales trend by month for each branch	Branch Month	Total	SUM
5	Daily total sales time-series trend for highest selling branch. Identify the day of highest sales.			
6	Find the trends in gender-wise sales by identifying the customer expenditure for different product lines. Which product line is popular among each gender?	Product line Gender	Total	SUM

# Understanding the Queries

	Query	Dimension	Fact	Aggregate
1	Find the total sales by branch and customer type.	Branch, Customer type	Total	SUM
2	Find the lowest rated product lines.	Product lines	Rating	AVG
3	Find the highest selling product lines by month	Product lines Month	Quantity	SUM
4	Total sales trend by month for each branch	Branch Month	Total	SUM
5	Daily total sales time-series trend for highest selling branch. Identify the day of highest sales.	Date	Total	SUM
6	Find the trends in gender-wise sales by identifying the customer expenditure for different product lines. Which product line is popular among each gender?	Product line Gender	Total	SUM

# Understanding the Queries

	Query	Dimension	Fact	Aggregate
1	Find the total sales by branch and customer type.	Branch, Customer type	Total	SUM
2	Find the lowest rated product lines.	Product lines	Rating	AVG
3	Find the highest selling product lines by month	Product lines Month	Quantity	SUM
4	Total sales trend by month for each branch	Branch Month	Total	SUM
5	Daily total sales time-series trend for highest selling branch. Identify the day of highest sales.	Date	Total	SUM
6	Find the trends in gender-wise sales by identifying the customer expenditure for different product lines. Which product line is popular among each gender?	Product line Gender	Total	SUM

# Business Problem: Sales Analysis

- Note that most of the dimensional queries are directed towards an **analysis of sales**.
- Remember, the dimensional queries come *after* you have identified a **business problem to solve**.
- For this exercise, we have a variety of queries not necessarily targeted for sales analysis. This is a demo exercise to strengthen our understanding.

# How do we find the trends and present the insights?

- Start by creating a pivot table.
- Turn your table into a suitable chart. (We can begin with a chart but it is good to visualize the numbers in a tabular form as well)
- Know that there is more than 1 way to present the same information.
- Let's try to solve these dimensional queries by creating pivot tables/charts using Power Pivot tool on MS Excel.

# Pivot Table and Pivot Chart Fields

PivotTable Fields

Active

All

Choose fields to add to report:

sales\_dataset

☐ Invoice ID
☐ Branch
☐ City
☐ Customer type

Search

Drag fields between areas below:

Filters

Columns

Rows

Values

PivotChart Fields

Active

All

Choose fields to add to report:

sales\_dataset

Search

Drag fields between areas below:

Filters

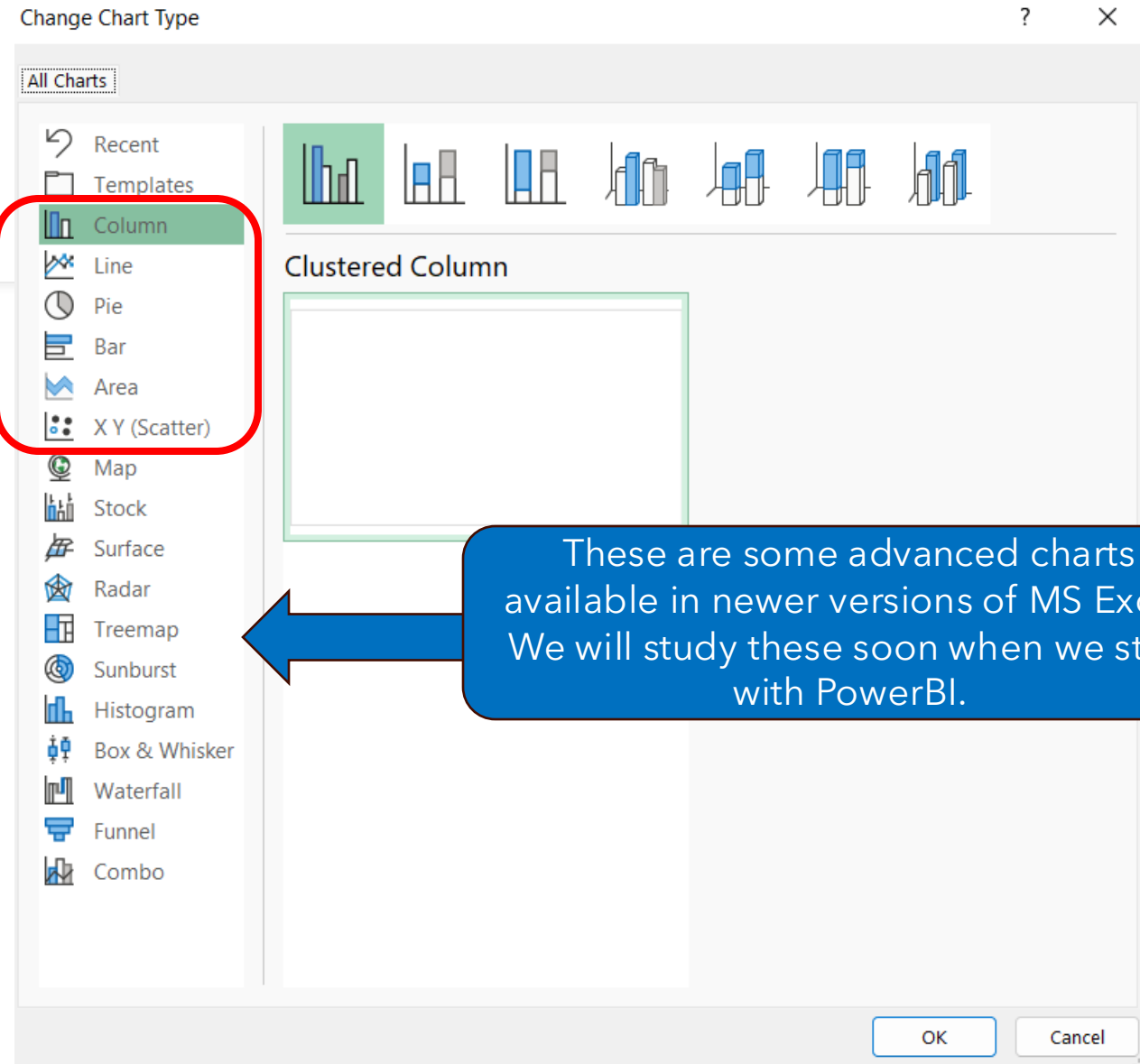
Legend (Series)

Axis (Categories)

Values

For now, we will stick to simple charts from these options. Remember, the easier it is for the viewer to understand your visualizations, the better is.

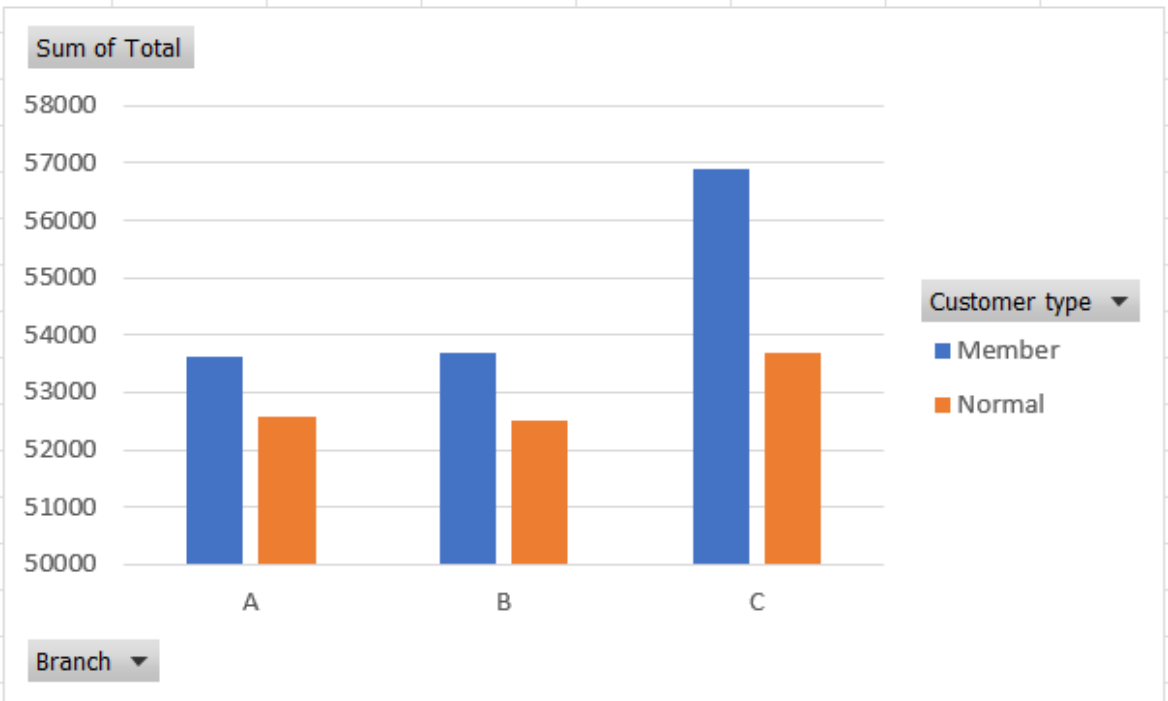
# Chart Types



These are some advanced charts available in newer versions of MS Excel. We will study these soon when we start with PowerBI.

# 1. Find the total sales by branch and customer type.

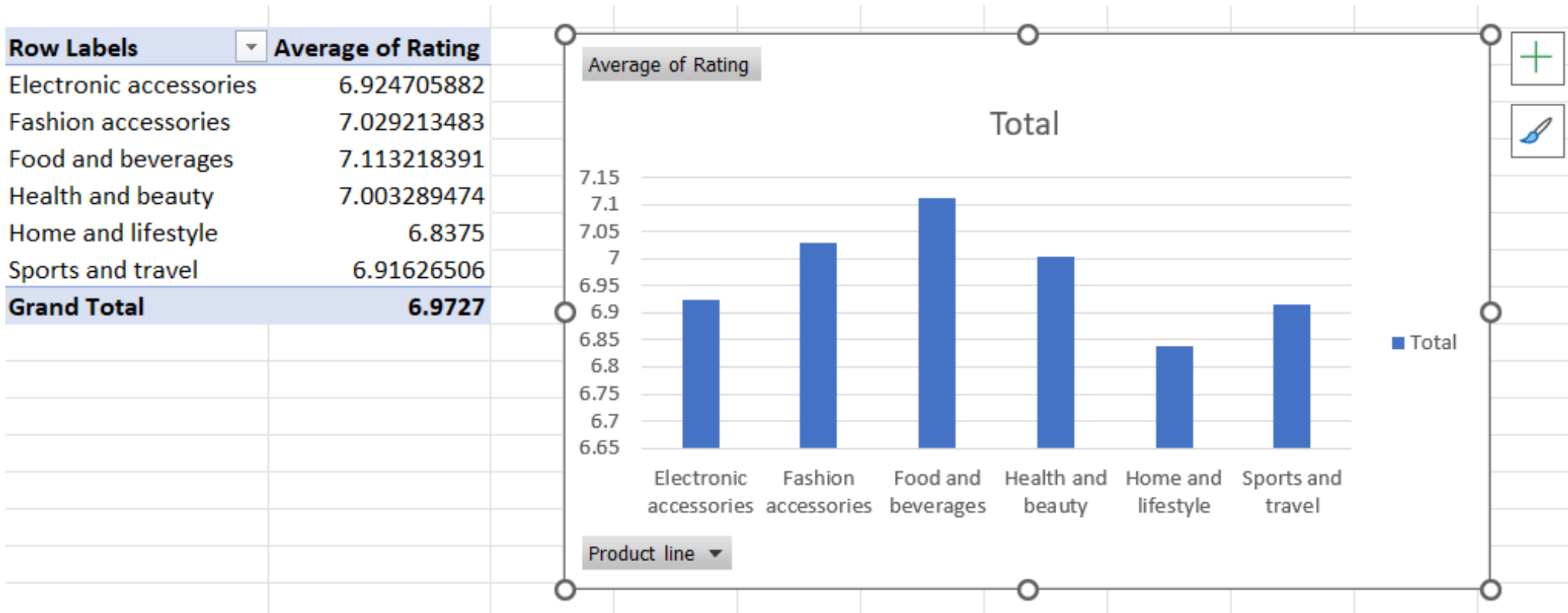
Sum of Total	Column Labels ▾		
Row Labels ▾	Member	Normal	Grand Total
A	53637.4755	52562.895	106200.3705
B	53704.686	52492.986	106197.672
C	56881.2825	53687.424	110568.7065
Grand Total	164223.444	158743.305	322966.749



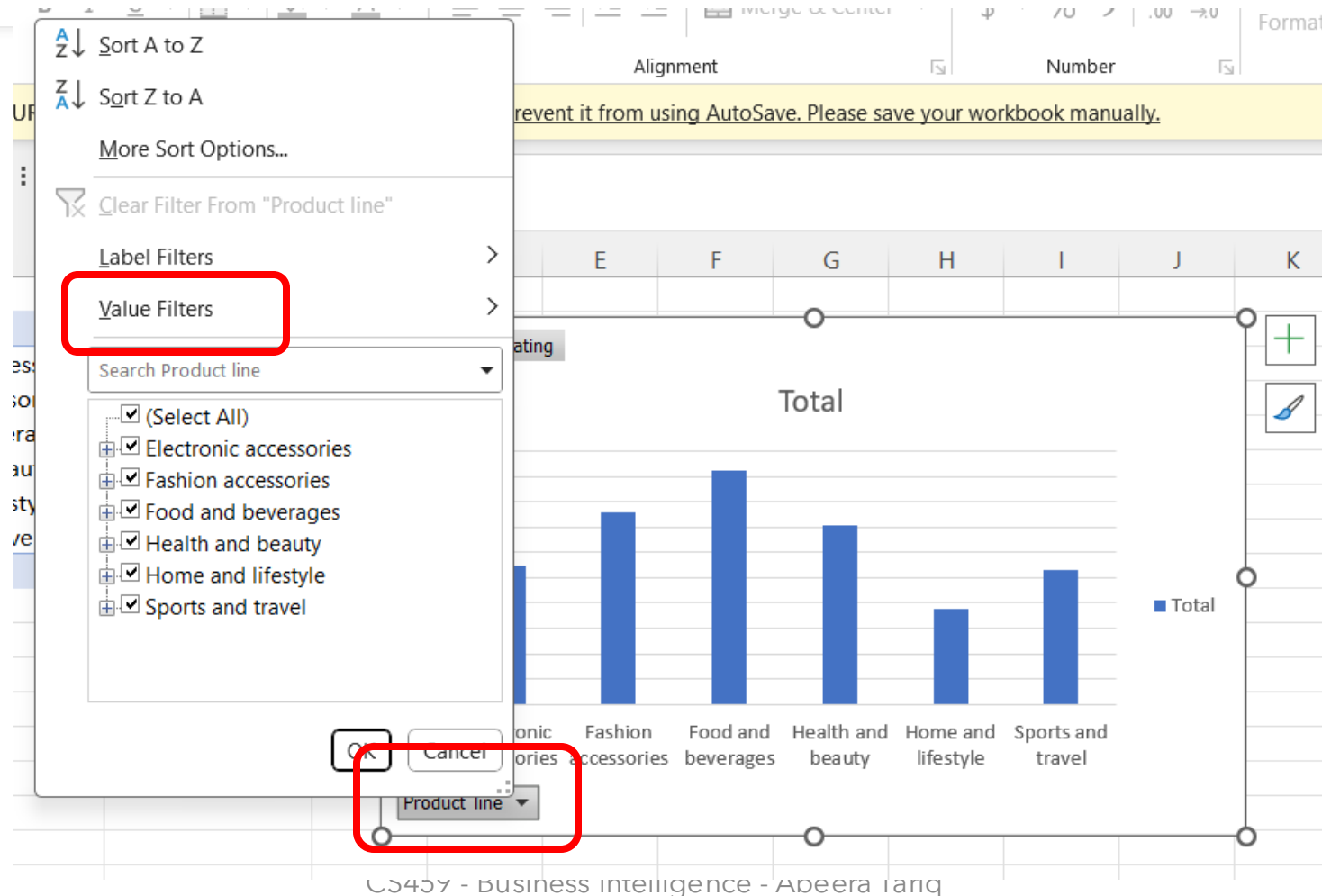
It is assumed that all students have an idea how to modify the design and look of visualizations (axes, labels, colors, etc.).

The better the presentation, the easier it is to read the chart.

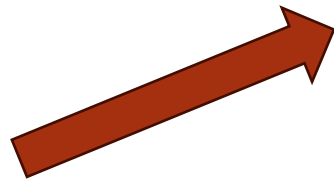
## 2. Find the lowest rated product lines. (Bottom 2)



# We can add Value Filters to any field added to the chart.



# Bottom 2



Top 10 Filter (Product line) ? X

Show

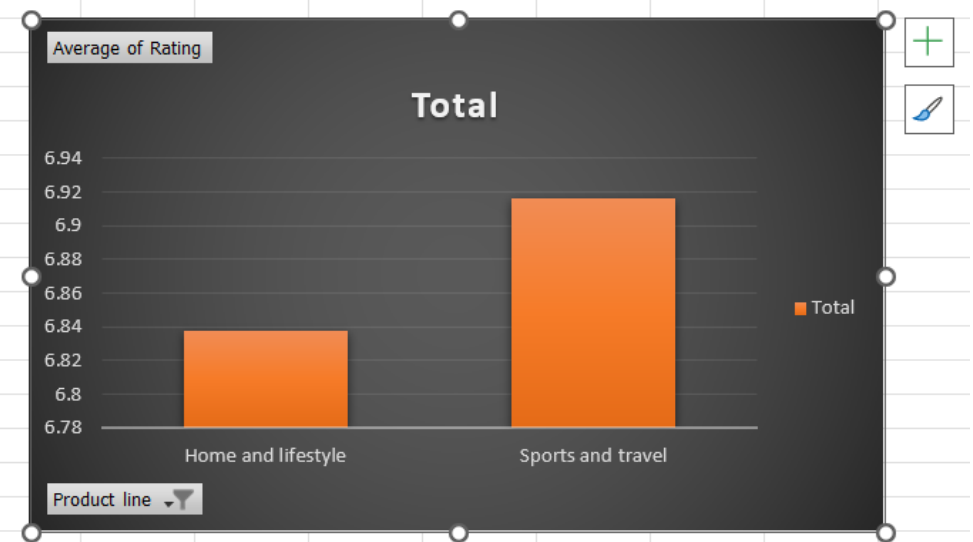
Bottom 2 Items by Average of Rating

OK Cancel



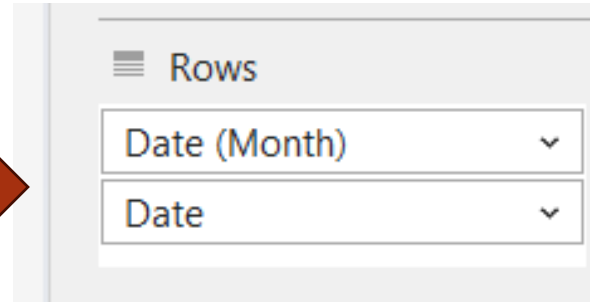
Sort A to Z  
Sort Z to A  
More Sort Options...  
Clear Filter From "Product line"  
Label Filters  
Value Filters  
Search Product line  
[X] (Select All)  
[X] Electronic accessories  
[X] Fashion accessories  
[X] Food and beverages  
[X] Health and beauty  
[X] Home and lifestyle  
[X] Sports and travel  
OK Cancel  
Clear Filter  
Equals...  
Does Not Equal...  
Greater Than...  
Greater Than Or Equal To...  
Less Than...  
Less Than Or Equal To...  
Between...  
Not Between...  
Top 10...

Row Labels	Average of Rating
Home and lifestyle	6.8375
Sports and travel	6.91626506
Grand Total	6.877607362



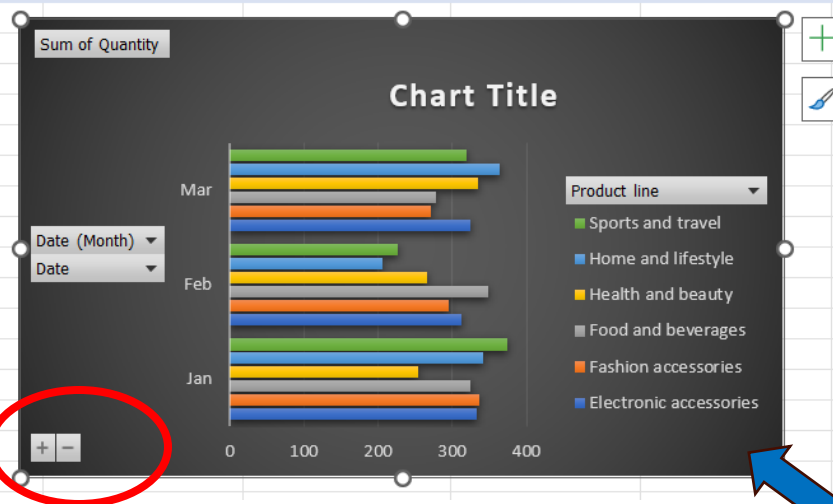
### 3. Find the highest selling product lines by month

Notice when you drag DATE dimension to any of the PivotTable/Pivot Chart Fields, it automatically breaks it into available option.



### 3. Find the highest selling product lines by month

Sum of Quantity	Electronic accessories	Fashion accessories	Food and beverages	Health and beauty	Home and lifestyle	Sports and travel	Grand Total
Jan	333	336	325	254	342	375	1965
Feb	313	295	349	266	205	226	1654
Mar	325	271	278	334	364	319	1891
Grand Total	971	902	952	854	911	920	5510



Active All

Choose fields to add to report:

Search

**sales\_dataset**

- ☐ Invoice ID
- ☐ Branch
- ☐ City
- ☐ Customer type

Drag fields between areas below:

**Filters**

**Legend (Series)**

Product line

**Axis (Categories)**

Date (Month)

Date

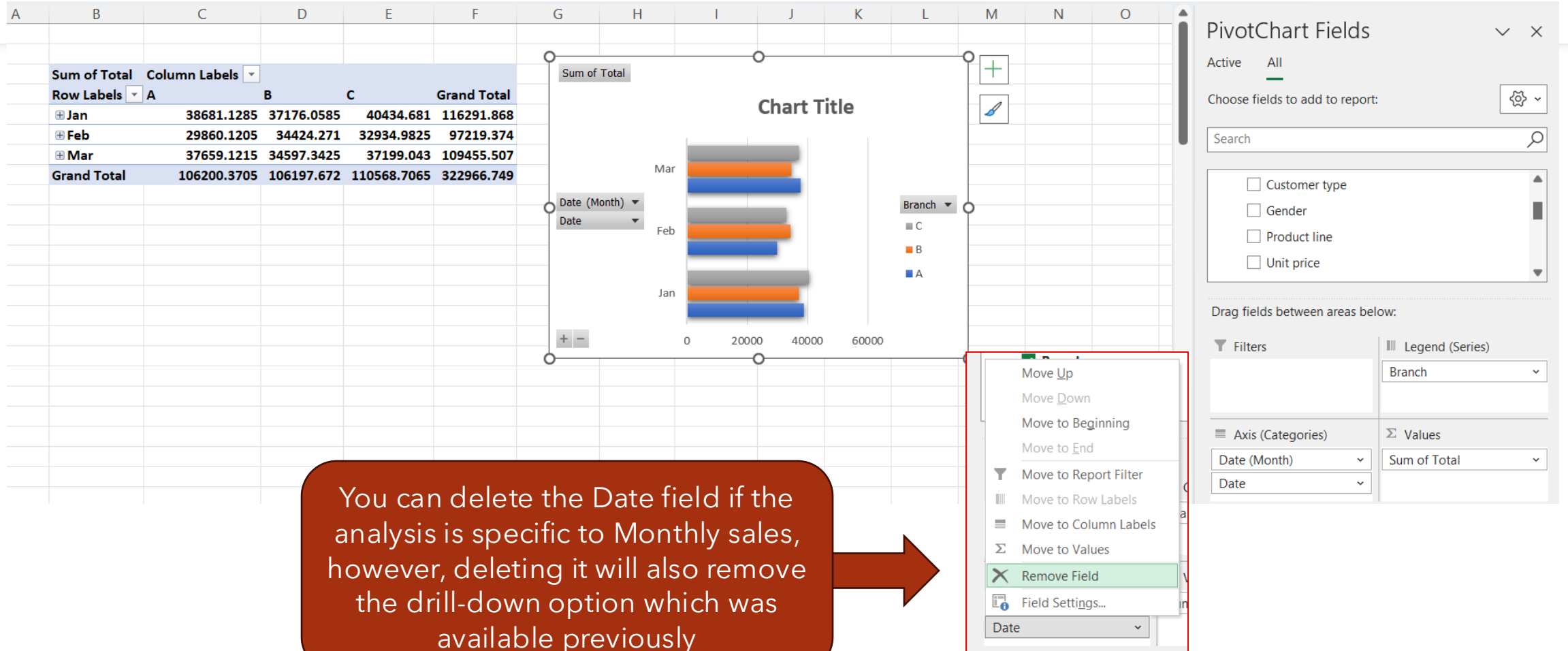
**Values**

Sum of Quantity

Try the + and - options.  
This will allow you to drill down and up across the date dimension with the available breakdown i.e. Date and month

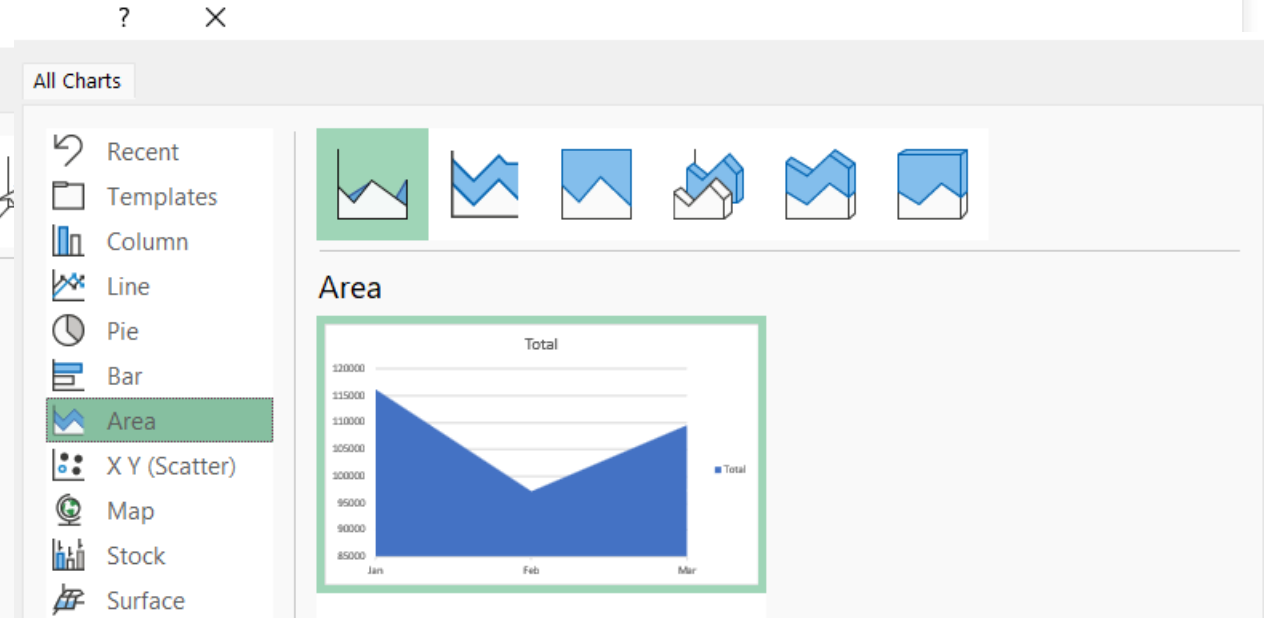
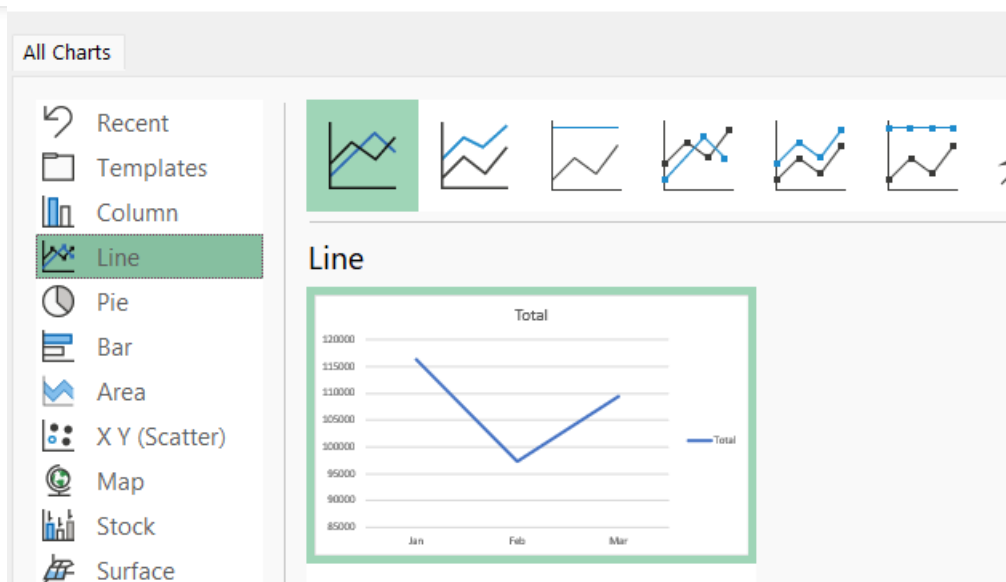
This is a BAR chart.  
Reading the bars instead of columns makes it easier to do a comparison.

## 4. Total sales trend by month for each branch.



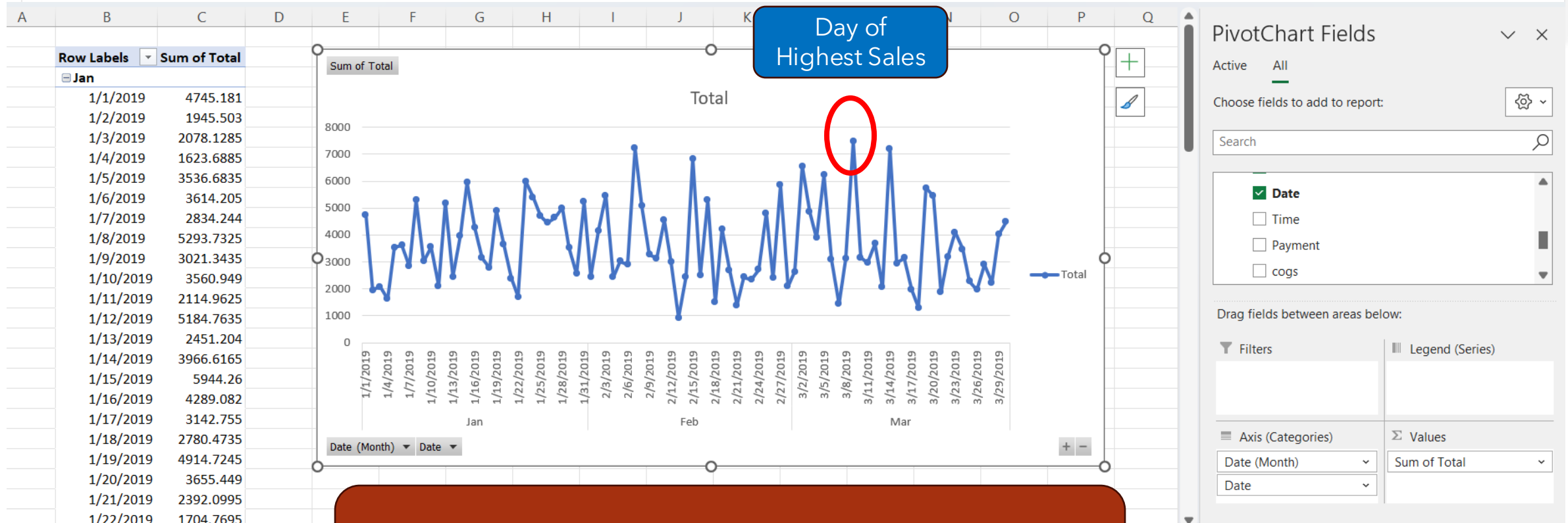
## 5. Daily total sales time-series trend for highest selling branch. Identify the day of highest sales.

Insert Chart



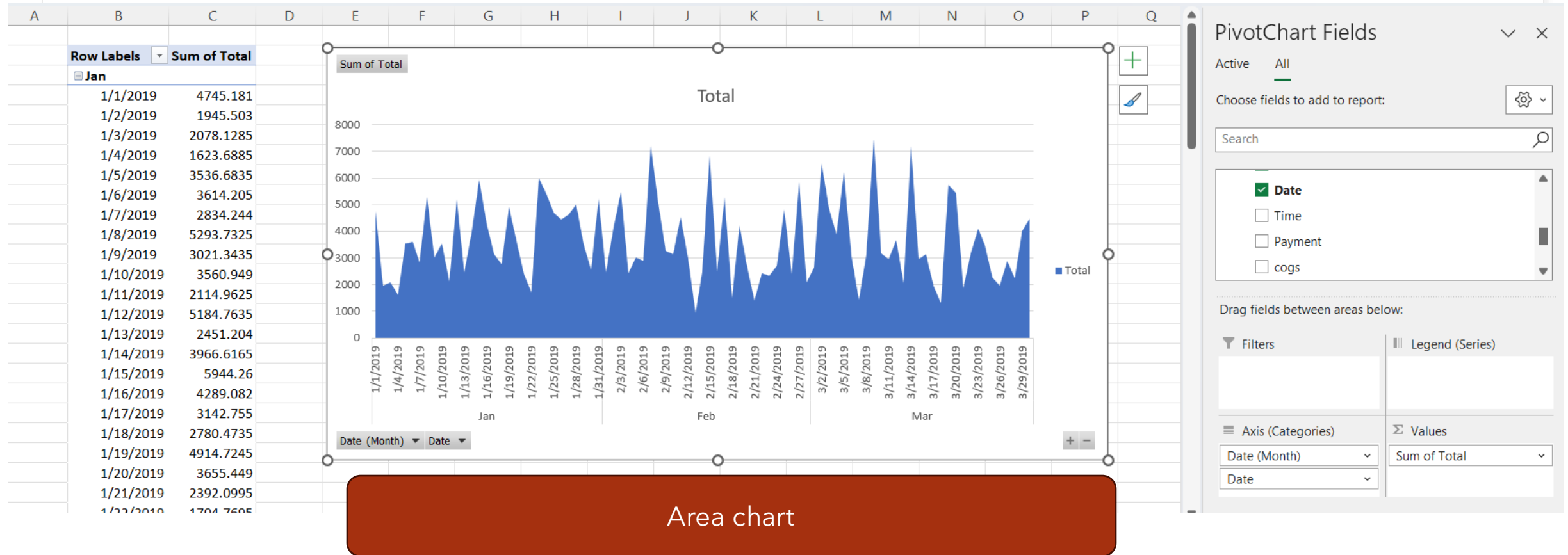
Line charts and Area charts are good choices for a time-series analysis.

## 5. Daily total sales time-series trend for highest selling branch. Identify the **day** of highest sales.



Through this line chart, the peaks are easier to read and present analysis of daily sales trends.

## 5. Daily total sales time-series trend for highest selling branch. Identify the **day** of highest sales.



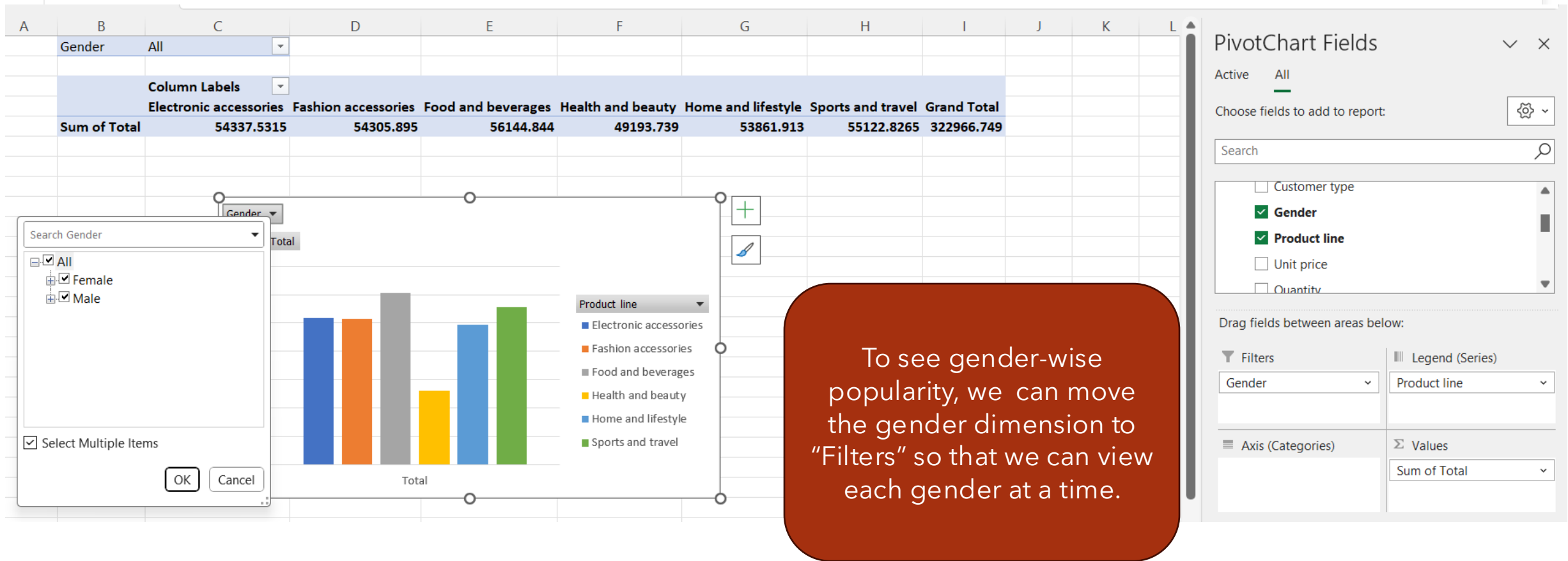
Customer expenditure = purchases = Total

Let's try to see the gender-wise total sales



Gender-wise, product-wise total expenditure

## 6. Find the trends in gender-wise sales by identifying the customer expenditure for different product lines. Which product line is popular among each gender?



## Another Business Problem

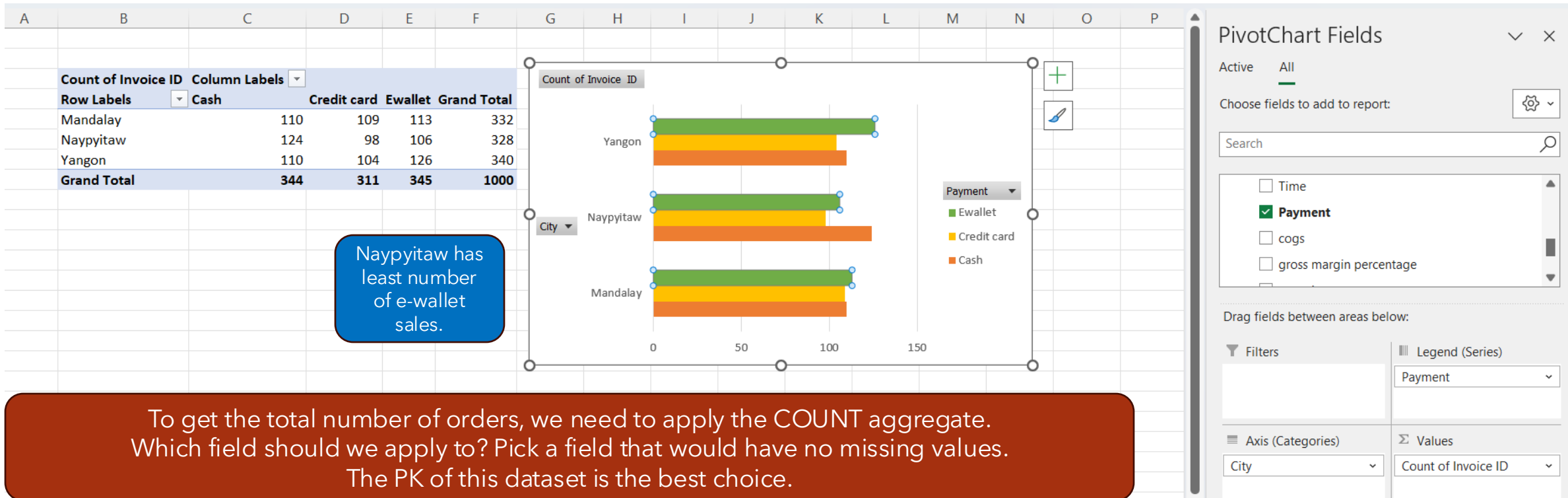
# Real-Life Scenario: E-Wallet Expansion

Assume that the company expanding relations with an *e-wallet platform* and want to identify the *city/cities* that need attention.

Although we do not have platform-specific details about the e-wallet option, the aim is to have more customers opt for the e-wallet payment option as compared to others in identified *city/cities*.

As part of the solution, you are required to present a suggestion to the decision makers based on the insights gained.

# Payment - City - No. of orders

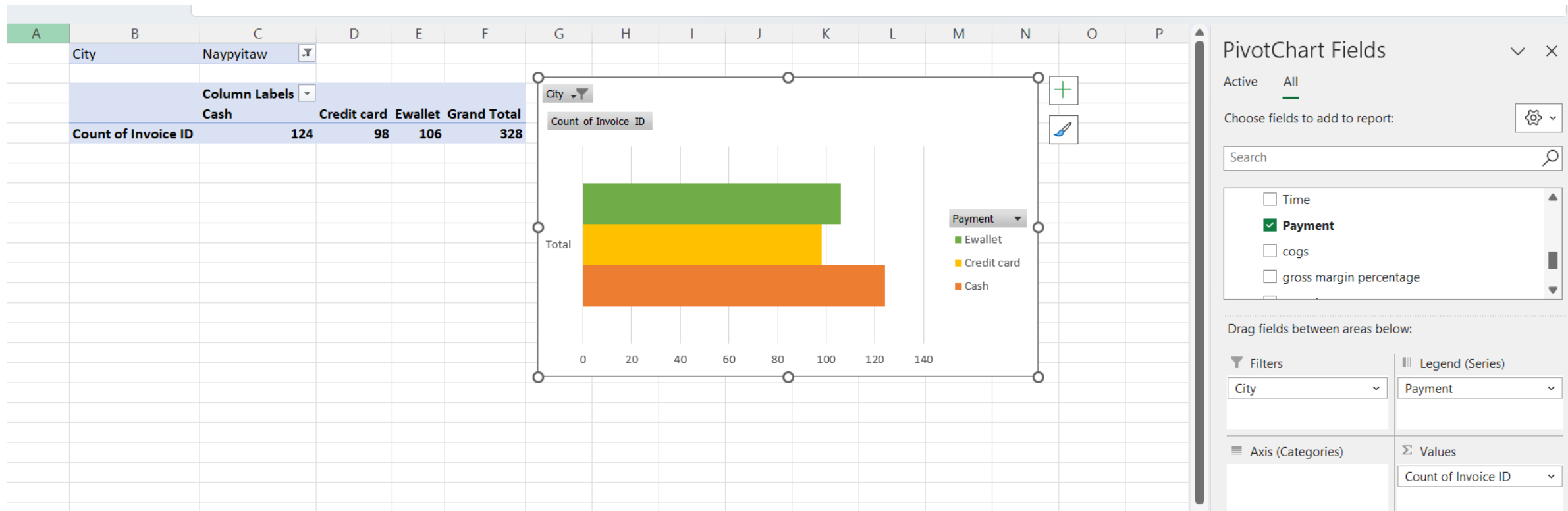


# For the city identified, how do we convince the customers to opt for this option?

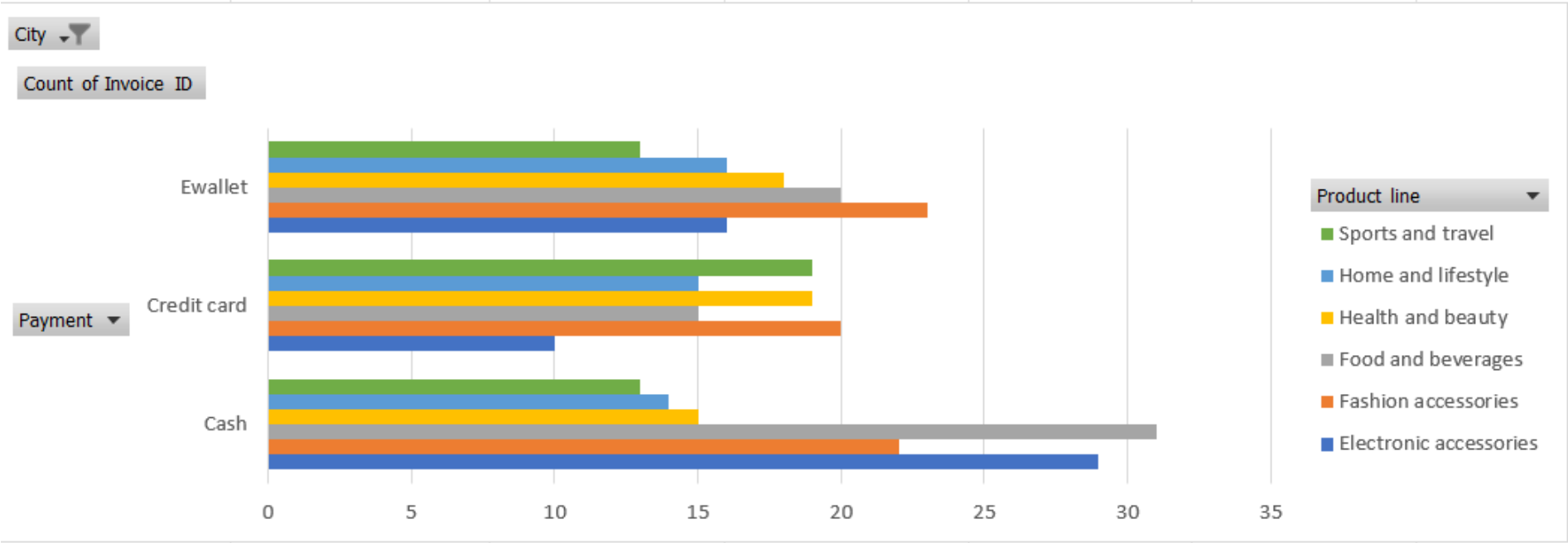
## Incentivizing E-Wallet Purchases

- **Discount Strategy:** Offering discounts on purchases made via e-wallet can drive adoption and boost overall sales.
- **Branch-Level Impact:** A branch-wide discount encourages sales across all products, increasing revenue.
- **Pilot Approach:** Since the e-wallet partnership is new, a small-scale test is advisable.
- **Targeted Strategy:** Focus on popular products in specific cities to maximize impact, as existing demand makes it easier to shift customers to e-wallet payments.

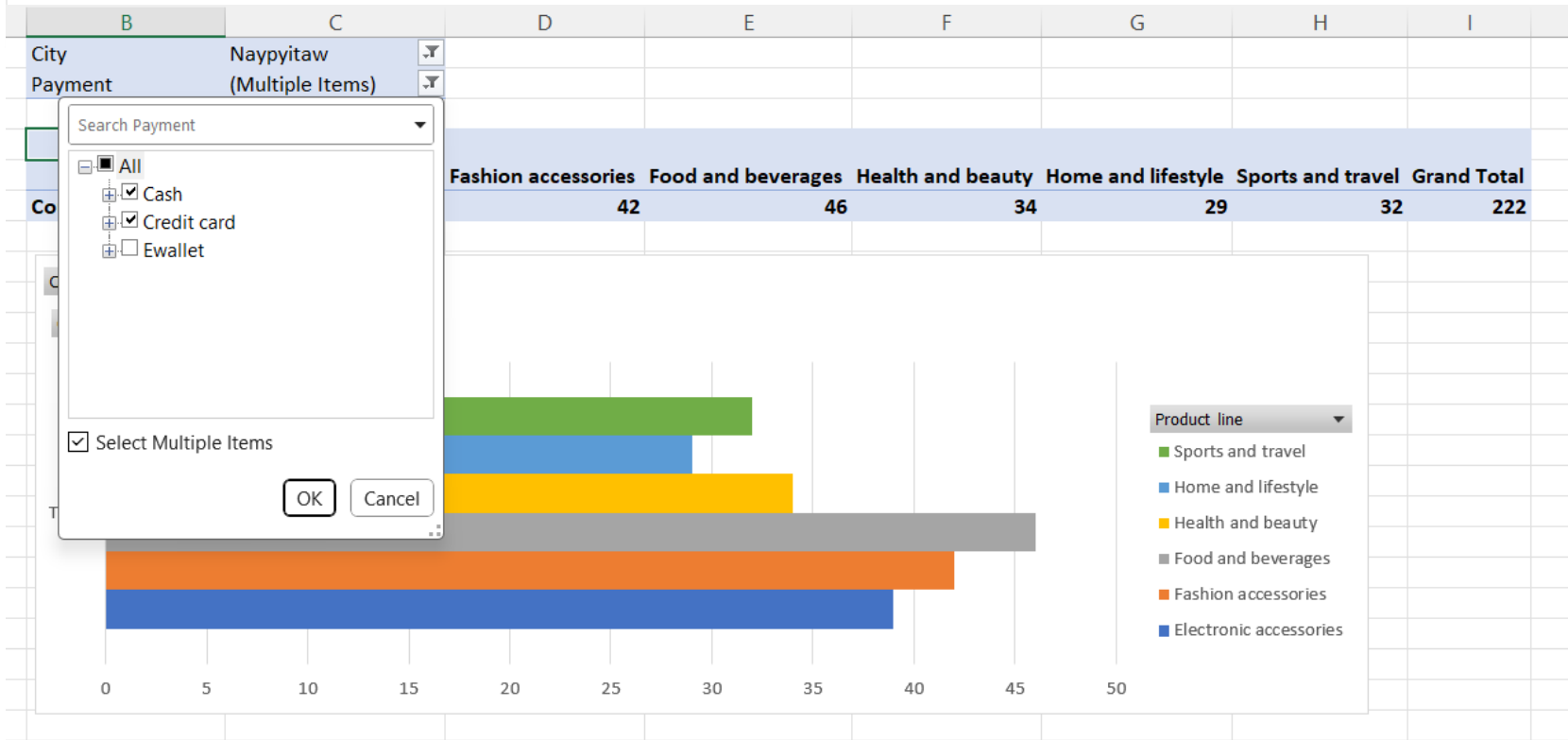
# Filter the city



# Naypyitaw sales by payment option and product lines



# Naypyitaw Sales by Product Lines for other payment options excluding e-wallet



Food and beverages have high sales in Naypyitaw but payment is made through other methods. This is a good place to introduce the discount strategy. We know our product sales are good, but now we need to convince customers to opt for e-wallet payment option.



# Solving Business Problems

- For any business problem, there may be **more than one** recommended solution! You need to find the best one!
- The job of a BI Analyst is not simply a matter of creating a few charts or visuals. ***It is about solving a business problem!***

# Solving Business Problems

- It requires a deep understanding of the business, the specific business domain and the company's data.
- The goal is to unravel meaningful ***insights*** by conducting a thorough **multi-dimensional analysis** using the identified ***facts*** and ***dimensions***, which will facilitate the ***business decision making process***.

# Profit/Loss Analysis

MyRetailStore

# Explore the dataset and the problem

- Understand the trends of sales and profit.
- Plot on a single chart to see how each is affected by the other.
- Numerical columns → Scatter plots
- Investigate the distribution and detect prominent outliers
- Conduct a multi-dimensional analysis after identifying pain areas i.e. loss.

# Drill-Down Strategy for Loss Analysis

💡 *A BI Analyst iterates through multiple analytical approaches before finalizing insights.*

- 🔍 **Step 1:** Identify the **Greatest Loss**
  - Use **time series analysis** to determine when it occurred.
- 📈 **Step 2:** Analyze **Profit Trends**
  - Examine category and sub-category patterns.
- 📍 **Step 3:** Locate **Problem Areas**
  - Identify regions and cities contributing to losses.
- 🛑 **Step 4:** Pinpoint **Root Causes**
  - Identify the specific **products, reasons, and affected customers.**
- ⚡ **Step 5:** Develop **Solutions** – Recommend strategies to mitigate future losses.

# OLAP Cube Pros

- ✓ **Faster Queries** – Pre-aggregated data speeds up retrieval and reporting.
- ✓ **Multi-Dimensional Analysis** – Enables slicing, dicing, drill-down, and roll-up for deep insights.
- ✓ **Pre-Calculated Aggregations** – Reduces on-the-fly computations, improving performance.
- ✓ **Better Decision-Making** – Provides interactive, real-time analysis for strategic insights.
- ✓ **Optimized for Large Data** – Handles complex queries efficiently without slowing down.

# OLAP Cube Cons

- ✗ **High Cost & Maintenance** - Expensive to build, maintain, and requires skilled personnel.
- ✗ **Performance Overhead** - Slow to update and process large datasets.
- ✗ **Storage Challenges** - Consumes significant space due to pre-aggregated data.
- ✗ **Limited Granularity** - Cannot drill down to transaction-level details.
- ✗ **Complex Navigation** - Non-technical users may find it hard to browse.
- ✗ **Lack of Flexibility** - Not easily adaptable to dynamic business needs.