



# **OLAP Analysis with PowerBI**

## **Data Analysis and Analytics**

***Enrico Gallinucci***

Rimini, 23/05/2022

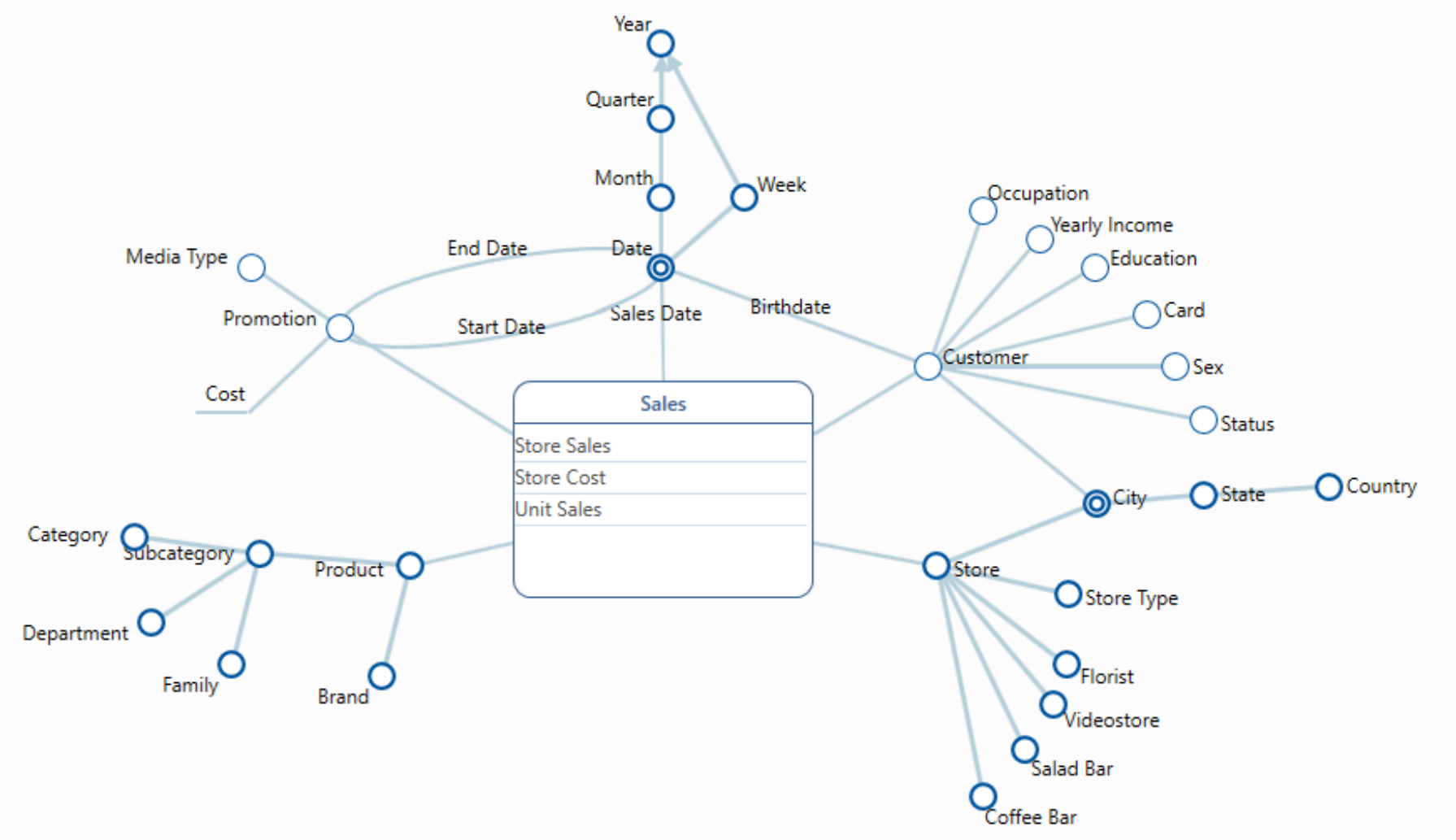
2022/2023

**BOLOGNA BUSINESS SCHOOL**  
Alma Mater Studiorum Università di Bologna

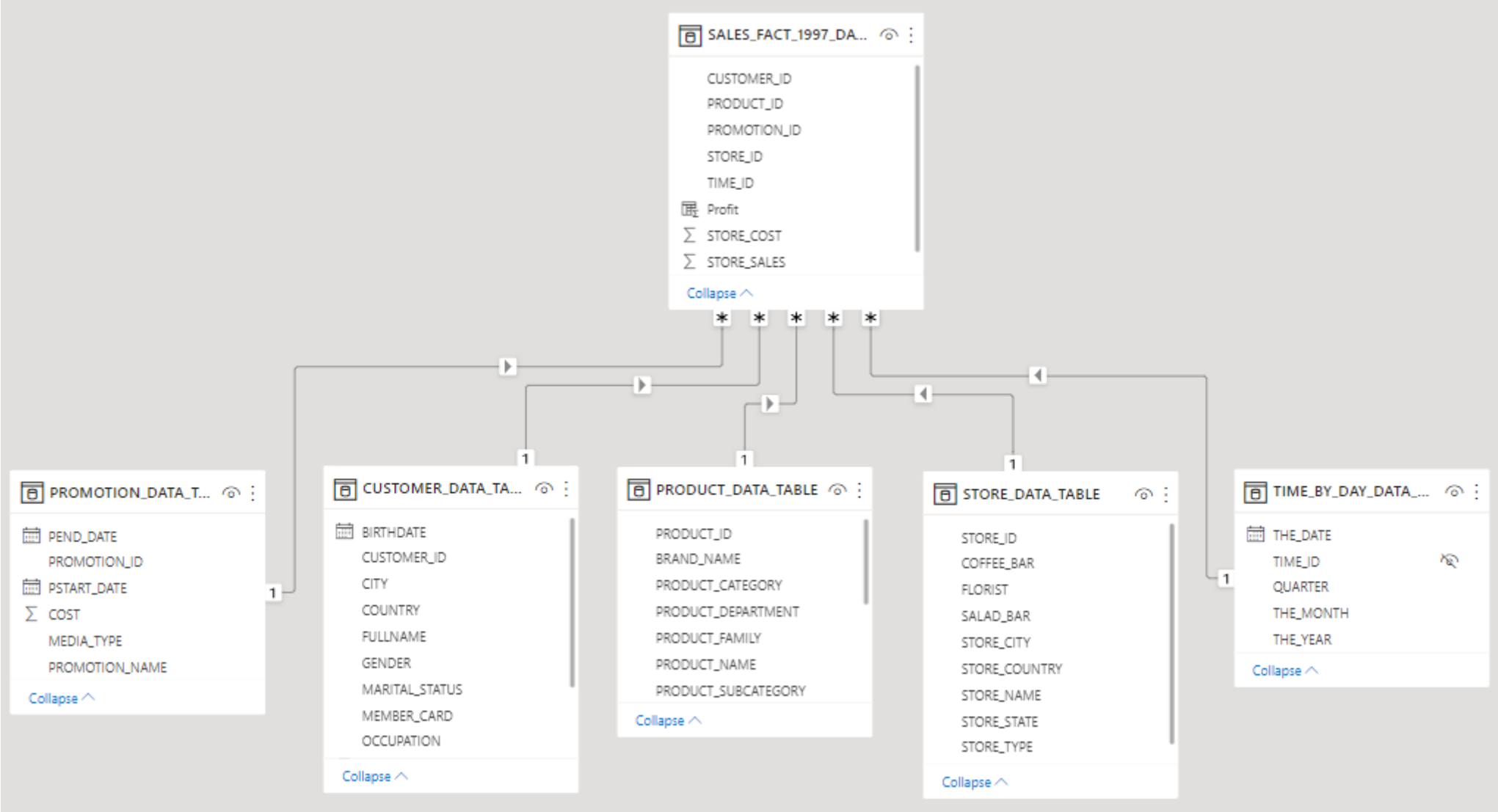
# Use case

Foodmart

# DFM – Foodmart



# Database Structure – Foodmart (Sales)



# Database Structure – Foodmart (Sales)

PROMOTION_ID	PROMOTION_NAME	MEDIA_TYPE	COST	PSTART_DATE	PEND_DATE
1	532 Weekend Markdown	Bulk Mail	14997	1998-03-10 00:00:00	1998-03-13 00:00:00
2	533 Shelf Clearing Days	Sunday Paper, Radio, TV	10196	1998-03-24 00:00:00	1998-03-27 00:00:00
3	534 Mystery Sale	Daily Paper, Radio	6884	1998-04-07 00:00:00	1998-04-10 00:00:00
4	535 Mystery Sale	Street Handout	8527	1998-04-22 00:00:00	1998-04-23 00:00:00
5	536 Bye Bye Baby	Daily Paper, Radio, TV	5816	1998-05-06 00:00:00	1998-05-09 00:00:00
6	537 Big Time Discounts	Daily Paper, Radio, TV	14641	1998-05-21 00:00:00	1998-05-23 00:00:00
7	538 Saving Days	Sunday Paper, Radio, TV	6711	1998-06-04 00:00:00	1998-06-05 00:00:00
8	539 High Roller Savings	Sunday Paper, Radio, TV	9555	1998-06-17 00:00:00	1998-06-19 00:00:00
9	540 Mystery Sale	Cash Register Handout	8553	1998-07-02 00:00:00	1998-07-04 00:00:00
10	541 Price Winners	Sunday Paper, Radio	12251	1998-07-17 00:00:00	1998-07-18 00:00:00
11	542 Green Light Special	Street Handout	12358	1998-07-31 00:00:00	1998-08-01 00:00:00
12	543 Fantastic Discounts	Radio	13746	1998-08-12 00:00:00	1998-08-14 00:00:00
13	544 Saving Days	Product Attachment	8326	1998-08-28 00:00:00	1998-08-29 00:00:00
14	545 Two for One	Daily Paper	6887	1998-09-11 00:00:00	1998-09-14 00:00:00

PROMOTION\_DATA\_T...

- PEND\_DATE
- PROMOTION\_ID
- PSTART\_DATE
- COST
- MEDIA\_TYPE
- PROMOTION\_NAME

[Collapse ^](#)

CUSTOMER\_DATA\_TA...

- BIRTHDATE
- CUSTOMER\_ID
- CITY
- COUNTRY
- FULLNAME
- GENDER
- MARITAL\_STATUS
- MEMBER\_CARD
- OCCUPATION

[Collapse ^](#)

PRODUCT\_DATA\_TABLE

- PRODUCT\_ID
- BRAND\_NAME
- PRODUCT\_CATEGORY
- PRODUCT\_DEPARTMENT
- PRODUCT\_FAMILY
- PRODUCT\_NAME
- PRODUCT\_SUBCATEGORY

[Collapse ^](#)

STORE\_DATA\_TABLE

- STORE\_ID
- COFFEE\_BAR
- FLORIST
- SALAD\_BAR
- STORE\_CITY
- STORE\_COUNTRY
- STORE\_NAME
- STORE\_STATE
- STORE\_TYPE

[Collapse ^](#)

TIME\_BY\_DAY\_DATA...

- THE\_DATE
- TIME\_ID
- QUARTER
- THE\_MONTH
- THE\_YEAR

[Collapse ^](#)

# Database Structure – Foodmart (Sales)

CUSTOMER_ID	CITY	STATE...	COUNTRY	BIRTHDATE	MARITAL...	YEARLY_INCOME	GENDER	MEMBER_CARD	OCCUPATION	FULLNAME	POPULATION	
1	107	Cliffside	BC	Canada	09-MAG-21	M	\$30K - \$50K	M	Bronze	Manual	Barney Velasquez	38000000
2	108	Shawnee	BC	Canada	13-MAG-31	M	\$90K - \$110K	M	Bronze	Management	Kenneth Dubois	38000000
3	109	Newton	BC	Canada	01-AGO-29	S	\$50K - \$70K	M	Bronze	Professional	Kate Maestas	38000000
4	110	West Covina	CA	USA	03-FEB-42	S	\$10K - \$30K	M	Normal	Manual	Josie Underwood	328000000
5	111	Palo Alto	CA	USA	21-OTT-65	M	\$70K - \$90K	F	Bronze	Management	Ramon Strain	328000000
6	112	Santa Monica	CA	USA	24-NOV-48	M	\$30K - \$50K	F	Bronze	Skilled Manual	Pat Azari	328000000
7	113	Sooke	BC	Canada	20-MAG-64	M	\$110K - \$130K	F	Bronze	Professional	Bob Dabit	38000000
8	114	Sooke	BC	Canada	16-APR-43	S	\$150K +	F	Silver	Professional	Marty Carmona	38000000
9	115	Altadena	CA	USA	09-GIU-77	S	\$90K - \$110K	F	Silver	Professional	Gina Saxton	328000000
10	116	Burnaby	BC	Canada	11-OTT-34	S	\$90K - \$110K	M	Bronze	Management	Juan McLaughlin	38000000
11	117	Albany	OR	USA	04-FEB-54	M	\$90K - \$110K	F	Bronze	Professional	Carol Eyster	328000000
12	118	Bremerton	WA	USA	04-OTT-13	S	\$50K - \$70K	M	Bronze	Professional	Nancy Henry	328000000
13	119	N. Vancouver	BC	Canada	26-LUG-57	S	\$10K - \$30K	M	Normal	Manual	Chris Barros	38000000
14	120	Milwaukie	OR	USA	28-GEN-57	M	\$30K - \$50K	F	Bronze	Skilled Manual	Christopher Groome	328000000
15	121	Santa Monica	CA	USA	17-FEB-12	S	\$30K - \$50K	F	Silver	Manual	Anna Hill	328000000

PROMOTION\_DATA\_T...

- PEND\_DATE
- PROMOTION\_ID
- PSTART\_DATE
- COST
- MEDIA\_TYPE
- PROMOTION\_NAME

Collapse ^

CUSTOMER\_DATA\_TA...

- BIRTHDATE
- CUSTOMER\_ID
- CITY
- COUNTRY
- FULLNAME
- GENDER
- MARITAL\_STATUS
- MEMBER\_CARD
- OCCUPATION

Collapse ^

PRODUCT\_DATA\_TABLE

- PRODUCT\_ID
- BRAND\_NAME
- PRODUCT\_CATEGORY
- PRODUCT\_DEPARTMENT
- PRODUCT\_FAMILY
- PRODUCT\_NAME
- PRODUCT\_SUBCATEGORY

Collapse ^

STORE\_DATA\_TABLE

- STORE\_ID
- COFFEE\_BAR
- FLORIST
- SALAD\_BAR
- STORE\_CITY
- STORE\_COUNTRY
- STORE\_NAME
- STORE\_STATE
- STORE\_TYPE

Collapse ^

TIME\_BY\_DAY\_DATA...

- THE\_DATE
- TIME\_ID
- QUARTER
- THE\_MONTH
- THE\_YEAR

Collapse ^

# Database Structure – Foodmart (Sales)

	PRODUCT_ID	BRAND_NAME	PRODUCT_NAME	PRODUCT_SUBCATEGORY	PRODUCT_CATEGORY	PRODUCT_DEPARTMENT	PRODUCT_FAMILY
1	1	Washington	Washington Berry Juice	Juice	Pure Juice Beverages	Beverages	Drink
2	2	Washington	Washington Mango Drink	Flavored Drinks	Drinks	Beverages	Drink
3	3	Washington	Washington Strawberry Drink	Flavored Drinks	Drinks	Beverages	Drink
4	4	Washington	Washington Cream Soda	Soda	Carbonated Beverages	Beverages	Drink
5	5	Washington	Washington Diet Soda	Soda	Carbonated Beverages	Beverages	Drink
6	6	Washington	Washington Cola	Soda	Carbonated Beverages	Beverages	Drink
7	7	Washington	Washington Diet Cola	Soda	Carbonated Beverages	Beverages	Drink
8	8	Washington	Washington Orange Juice	Juice	Pure Juice Beverages	Beverages	Drink
9	9	Washington	Washington Cranberry Juice	Juice	Pure Juice Beverages	Beverages	Drink
10	10	Washington	Washington Apple Juice	Juice	Pure Juice Beverages	Beverages	Drink
11	11	Washington	Washington Apple Drink	Flavored Drinks	Drinks	Beverages	Drink
12	12	Jeffers	Jeffers Oatmeal	Cereal	Breakfast Foods	Breakfast Foods	Food
13	13	Jeffers	Jeffers Corn Puffs	Cereal	Breakfast Foods	Breakfast Foods	Food
14	14	Jeffers	Jeffers Wheat Puffs	Cereal	Breakfast Foods	Breakfast Foods	Food
15	15	Jeffers	Jeffers Grits	Cereal	Breakfast Foods	Breakfast Foods	Food

PROMOTION\_DATA\_T...

- PEND\_DATE
- PROMOTION\_ID
- PSTART\_DATE
- COST
- MEDIA\_TYPE
- PROMOTION\_NAME

Collapse ^

CUSTOMER\_DATA\_TA...

- BIRTHDATE
- CUSTOMER\_ID
- CITY
- COUNTRY
- FULLNAME
- GENDER
- MARITAL\_STATUS
- MEMBER\_CARD
- OCCUPATION

Collapse ^

PRODUCT\_DATA\_TABLE

- PRODUCT\_ID
- BRAND\_NAME
- PRODUCT\_CATEGORY
- PRODUCT\_DEPARTMENT
- PRODUCT\_FAMILY
- PRODUCT\_NAME
- PRODUCT\_SUBCATEGORY

Collapse ^

STORE\_DATA\_TABLE

- STORE\_ID
- COFFEE\_BAR
- FLORIST
- SALAD\_BAR
- STORE\_CITY
- STORE\_COUNTRY
- STORE\_NAME
- STORE\_STATE
- STORE\_TYPE

Collapse ^

TIME\_BY\_DAY\_DATA...

- THE\_DATE
- TIME\_ID
- QUARTER
- THE\_MONTH
- THE\_YEAR

Collapse ^

# Database Structure – Foodmart (Sales)

STORE_ID	STORE_TYPE	STORE_NAME	STORE_CITY	STORE_STATE	STORE_COUNTRY	COFFEE_BAR	VIDEO_STORE	SALAD_BAR	FLORIST
1	1 Supermarket	Store 1	Acapulco	Guerrero	Mexico	0	0	0	0
2	2 Small Grocery	Store 2	Bellingham	WA	USA	1	0	0	0
3	3 Supermarket	Store 3	Bremerton	WA	USA	0	0	1	0
4	4 Gourmet Supermarket	Store 4	Camacho	Zacatecas	Mexico	1	0	1	1
5	5 Small Grocery	Store 5	Guadalajara	Jalisco	Mexico	1	0	0	0
6	6 Gourmet Supermarket	Store 6	Beverly Hills	CA	USA	1	1	1	1
7	7 Supermarket	Store 7	Los Angeles	CA	USA	0	0	0	1
8	8 Deluxe Supermarket	Store 8	Merida	Yucatan	Mexico	1	1	1	1
9	9 Mid-Size Grocery	Store 9	Mexico City	DF	Mexico	0	0	0	0
10	10 Supermarket	Store 10	Orizaba	Veracruz	Mexico	0	0	1	0
11	11 Supermarket	Store 11	Portland	OR	USA	0	0	0	0
12	12 Deluxe Supermarket	Store 12	Hidalgo	Zacatecas	Mexico	1	1	1	1
13	13 Deluxe Supermarket	Store 13	Salem	OR	USA	1	1	1	1
14	14 Small Grocery	Store 14	San Francisco	CA	USA	1	0	0	0
15	15 Supermarket	Store 15	Seattle	WA	USA	1	0	0	0
16	16 Supermarket	Store 16	Spokane	WA	USA	0	0	0	0

PROMOTION\_DATA\_T...

- PEND\_DATE
- PROMOTION\_ID
- PSTART\_DATE
- COST
- MEDIA\_TYPE
- PROMOTION\_NAME

Collapse ^

CUSTOMER\_DATA\_TA...

- BIRTHDATE
- CUSTOMER\_ID
- CITY
- COUNTRY
- FULLNAME
- GENDER
- MARITAL\_STATUS
- MEMBER\_CARD
- OCCUPATION

Collapse ^

PRODUCT\_DATA\_TABLE

- PRODUCT\_ID
- BRAND\_NAME
- PRODUCT\_CATEGORY
- PRODUCT\_DEPARTMENT
- PRODUCT\_FAMILY
- PRODUCT\_NAME
- PRODUCT\_SUBCATEGORY

Collapse ^

STORE\_DATA\_TABLE

- STORE\_ID
- COFFEE\_BAR
- FLORIST
- SALAD\_BAR
- STORE\_CITY
- STORE\_COUNTRY
- STORE\_NAME
- STORE\_STATE
- STORE\_TYPE

Collapse ^

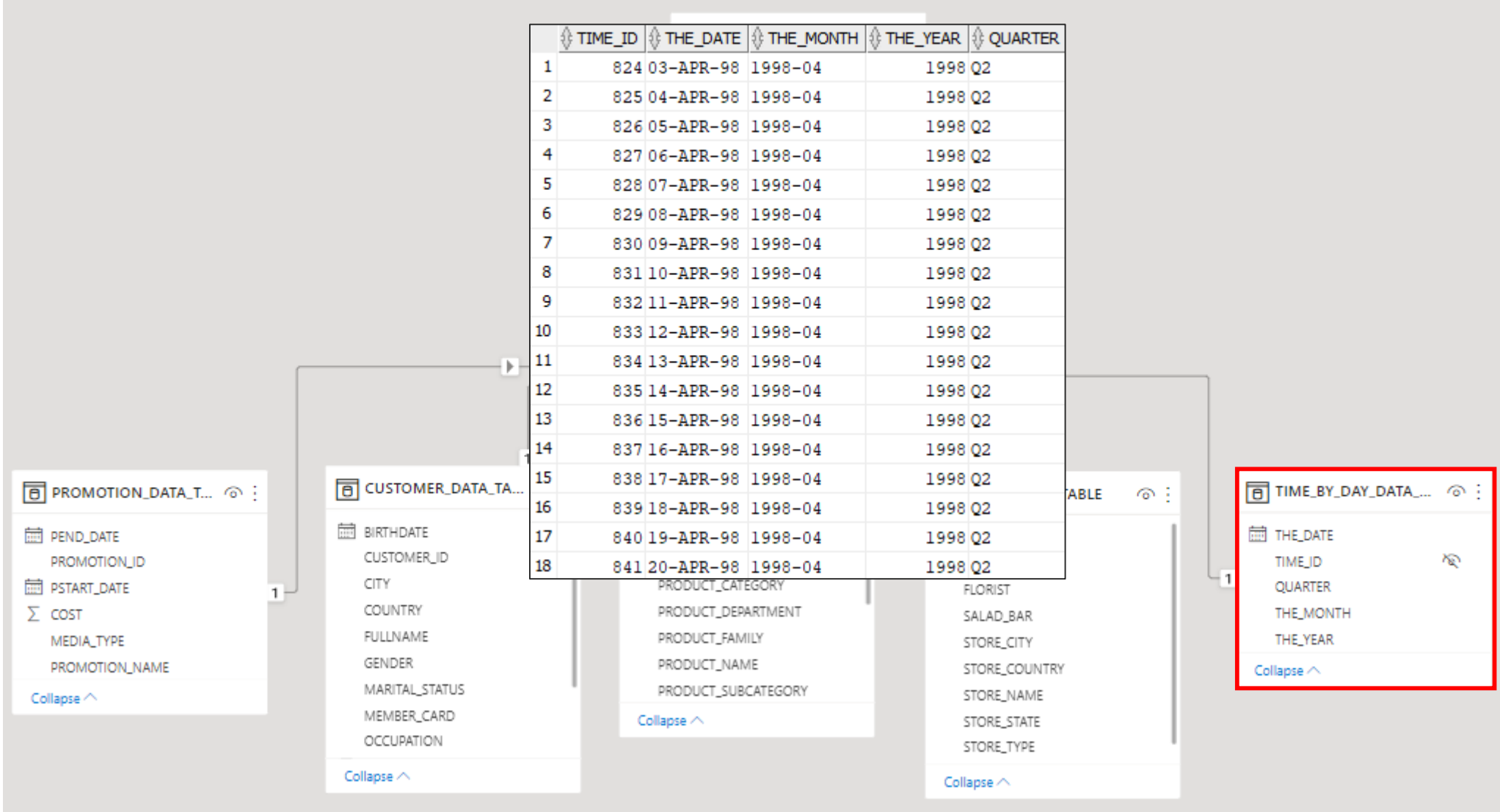
TIME\_BY\_DAY\_DATA...

- THE\_DATE
- TIME\_ID
- QUARTER
- THE\_MONTH
- THE\_YEAR

Collapse ^



# Database Structure – Foodmart (Sales)



# Database Structure – Foodmart (Sales)

The screenshot displays a central fact table, **SALES\_FACT\_1997\_DATA**, which is connected to three dimension tables: **PROMOTION\_DATA**, **TIME\_BY\_DAY\_DATA**, and **SALES\_FACT\_1997\_DATA** (the fact table itself). The fact table columns are: PRODUCT\_ID, TIME\_ID, CUSTOMER\_ID, PROMOTION\_ID, STORE\_ID, STORE\_SALES, STORE\_COST, and UNIT\_SALES. The dimension tables provide hierarchical context for these facts.

**SALES\_FACT\_1997\_DATA** (Fact Table):

PRODUCT_ID	TIME_ID	CUSTOMER_ID	PROMOTION_ID	STORE_ID	STORE_SALES	STORE_COST	UNIT_SALES
1	1	369	4728	501	7	11,4	3,99
2	1	377	9788	1547	13	8,55	4,0185
3	1	414	6666	34	17	8,55	4,1895
4	1	440	5313	413	24	8,55	3,762
5	1	463	916	302	7	11,4	4,902
6	1	474	4461	1839	11	8,55	2,9925
7	1	489	1312	162	3	8,55	3,6765
8	1	500	9169	1435	23	11,4	5,358
9	1	529	5607	501	6	11,4	4,902
10	1	534	456	828	15	11,4	4,332
11	1	570	923	30	15	8,55	2,736
12	1	574	9358	1097	15	8,55	4,275
13	1	576	7704	486	3	5,7	2,508
14	1	590	3441	131	3	8,55	3,42
15	1	594	6248	1860	24	11,4	3,876
16	1	596	5929	496	15	14,25	5,5575
17	1	616	1565	116	24	8,55	4,1895
18	1	617	638	1038	11	8,55	2,9925

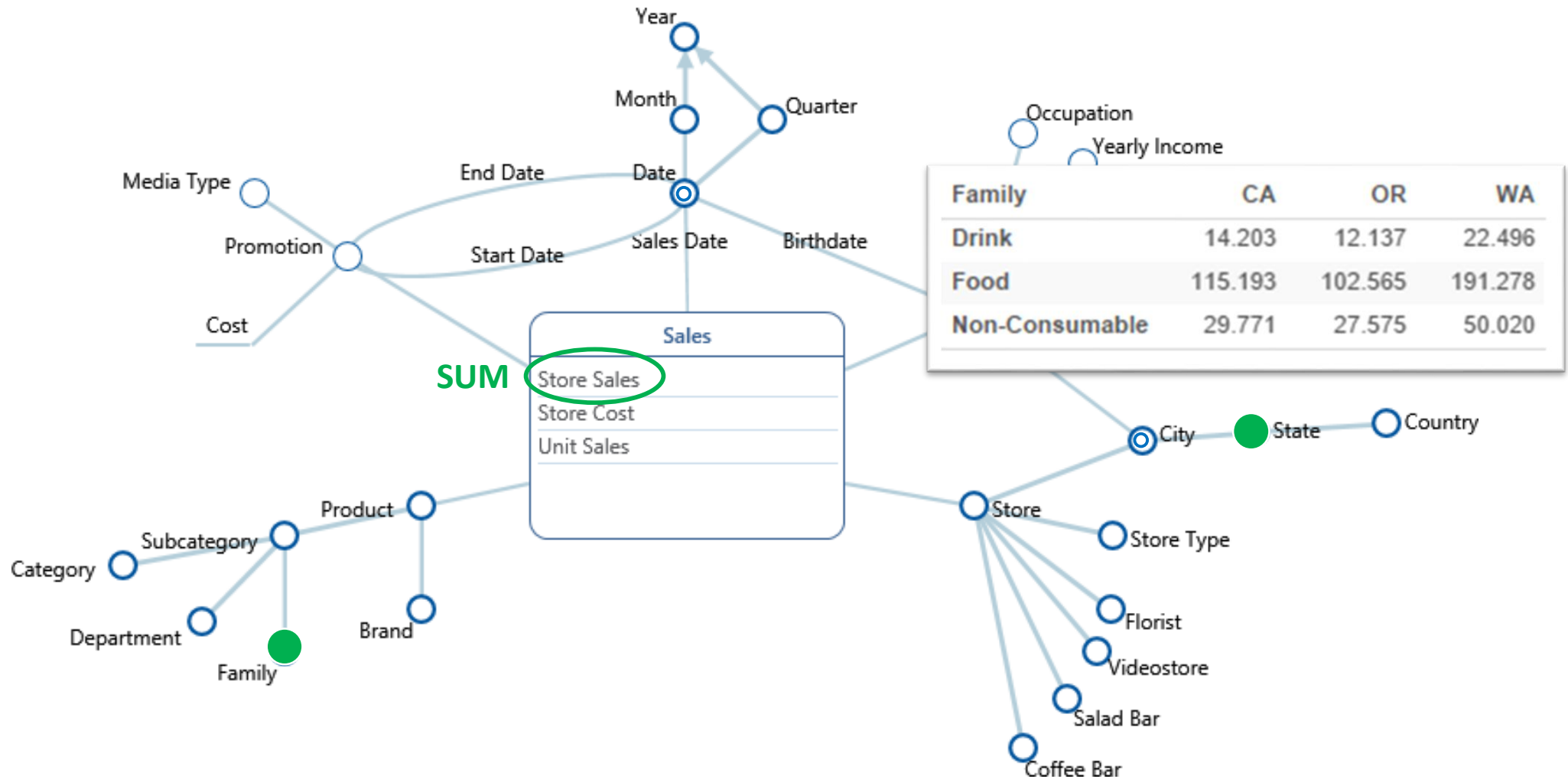
**PROMOTION\_DATA** (Dimension Table):

- PROMOTION\_ID
- PSTART\_DATE
- COST
- MEDIA\_TYPE
- PROMOTION\_NAME

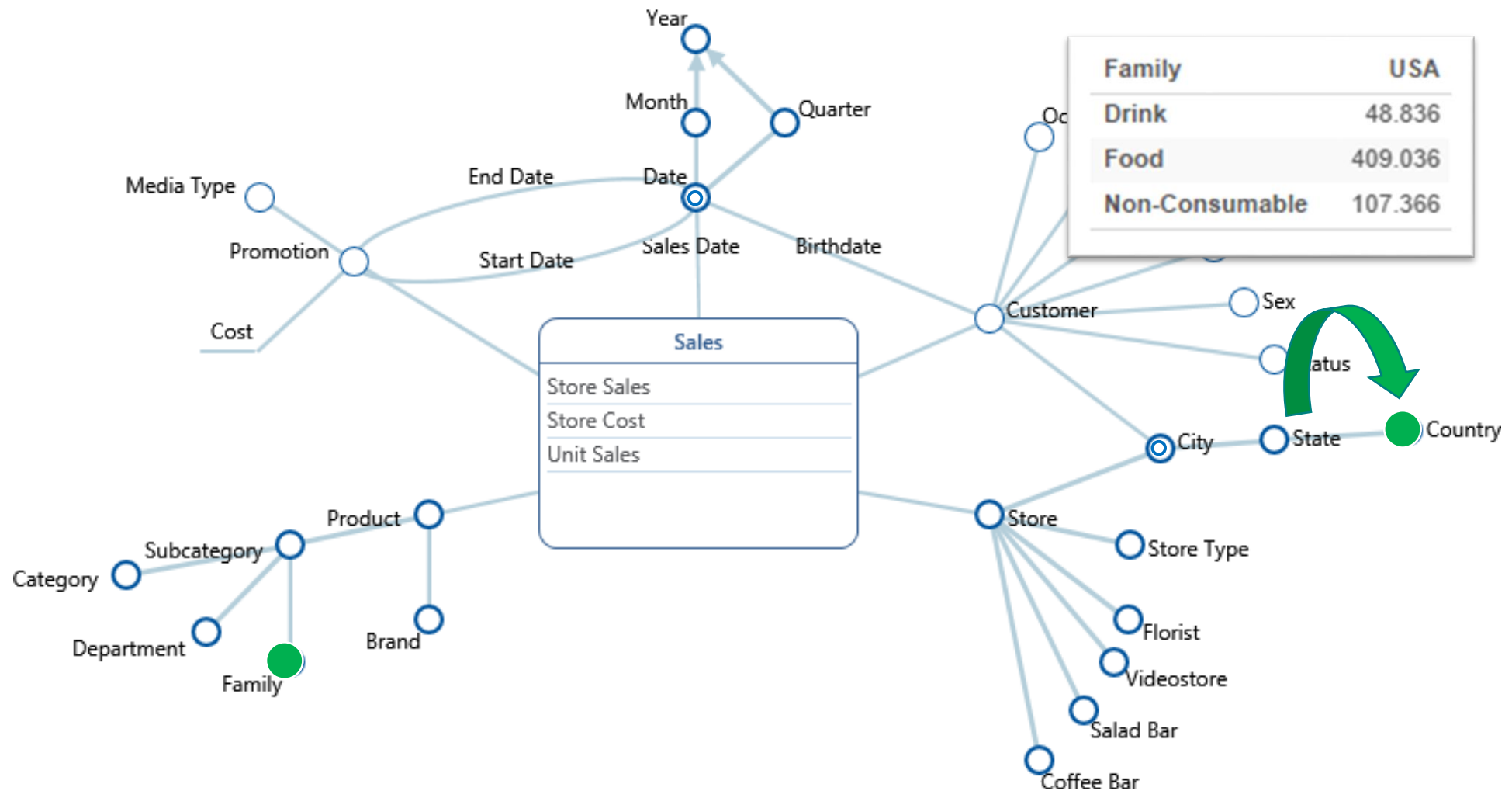
**TIME\_BY\_DAY\_DATA** (Dimension Table):

- THE\_DATE
- TIME\_ID
- QUARTER
- THE\_MONTH
- THE\_YEAR

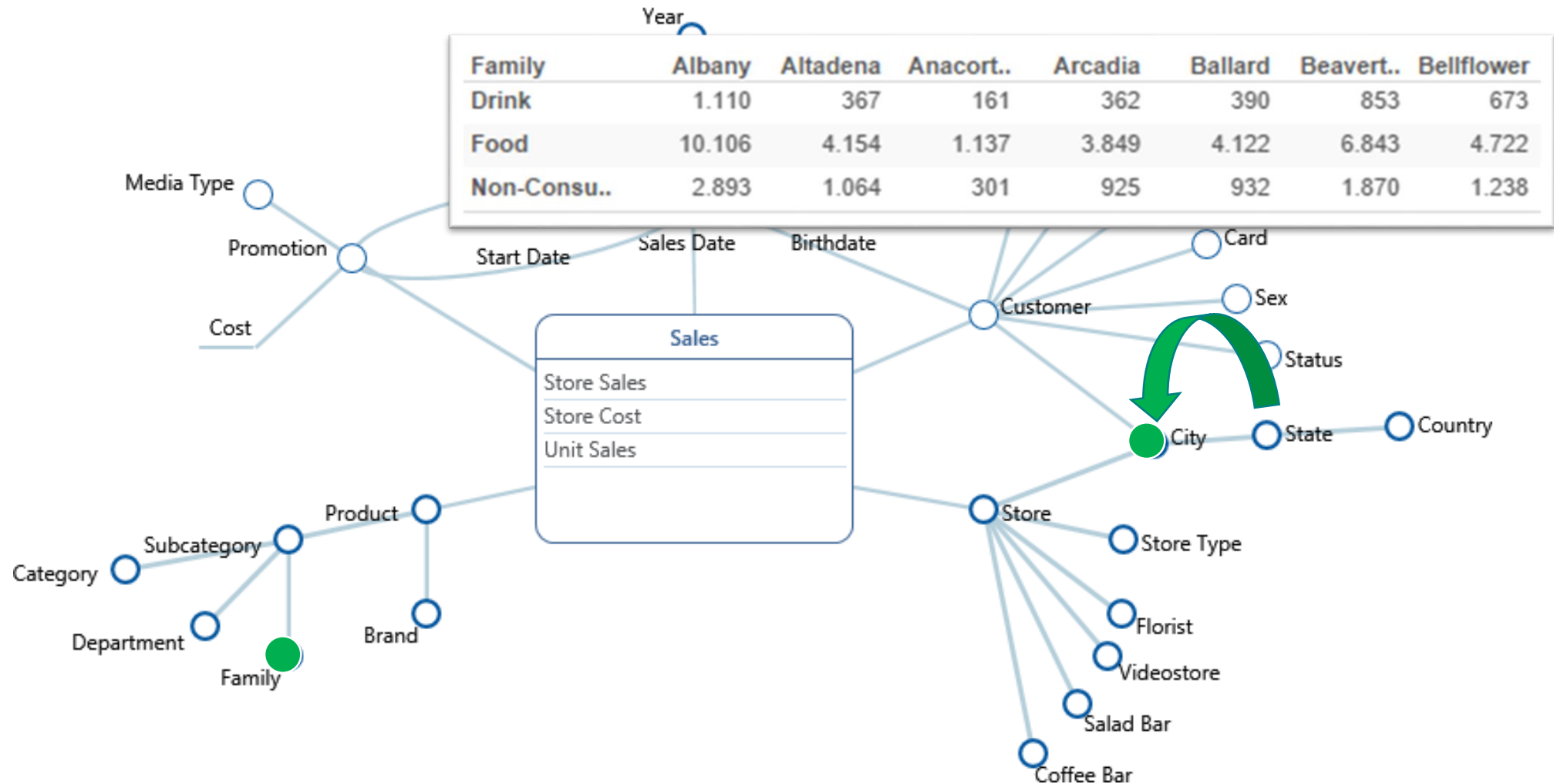
# OLAP query



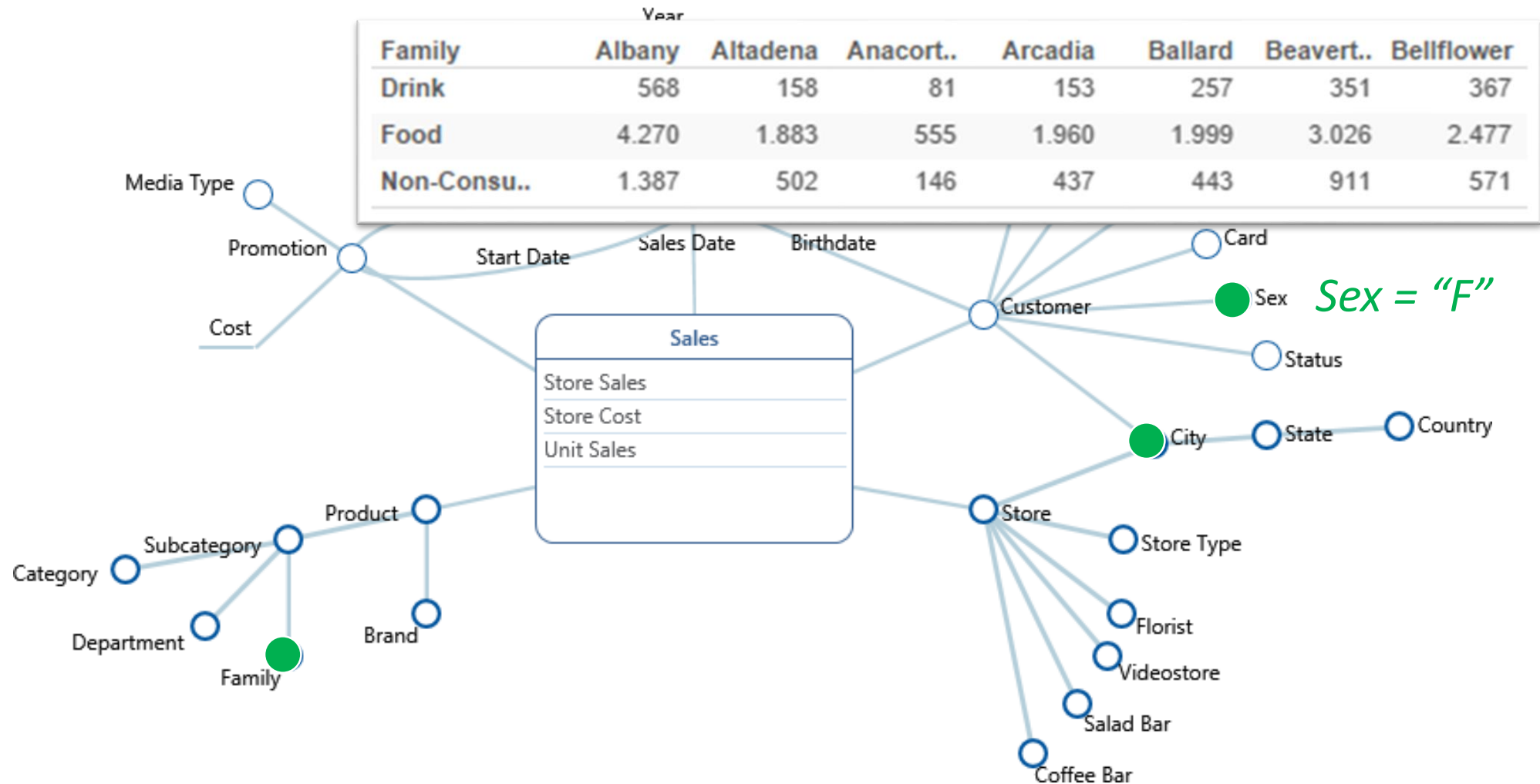
# OLAP Operators: Roll-Up



# OLAP Operators: Drill-down



# OLAP Operators: Slice & Dice



# Exercises - Part I

Setup the data on PowerBI

# Exercise 0

- Get the data from <https://bit.ly/bbs-scm>
- Unzip the Foodmart.zip file
- Open the CSV files on PowerBI
- Setup the relationships between the files
- Clean the data
  - Check the month
  - Remove wrong sales
  - IDs or names?
  - Hide unused fields
- Create hierarchies



# Exercises - Part II

Familiarize with PowerBI querying functionalities

# Exercise 1

- Use a bar chart to plot the total sum of *STORE\_SALES* for each *STORE\_STATE*
  - Which one is the state with the highest sales?
- Apply a drill-down operation to show the sales at the *STORE\_CITY* level
  - Are there cities whose sales are much lower than the others'?
- How many stores are there in each *STORE\_STATE*? In each *STORE\_CITY*?
  - Color the bars based on the *Count(Distinct)* summarization function over the *STORE\_NAME* attribute
  - Would it be reasonable to say that cities with fewer stores also have lower total sales?

# Exercise 2

- Use a bar chart to plot the total sum of *STORE\_SALES* by *STORE\_CITY* and assign the *STORE\_TYPE* to the Legend property
  - Can you notice any interesting pattern?
- Use a bar chart to plot the total sum of *STORE\_SALES* by *STORE\_TYPE*
  - Assign the number of stores to the color property
  - Is the result surprising/expected?

# Exercise 3

- Use a line chart to plot the monthly sales trend
  - Any interesting pattern?
- Split the previous chart by *STORE\_STATE*
  - Put the *STORE\_STATE* in the Legend
  - Does the previous pattern hold for each state?
- Visualise the impact of each *STORE\_FAMILY* on the total sales while still showing the monthly trends
  - Use a Stacked area chart, where the *STORE\_STATE* is in the Small multiples and the *PRODUCT\_FAMILY* in the Legend

# Exercise 4

- Analyze sales by *STORE\_TYPE* (sorted by descending order)
- Drill-down to the stores
- Add the number of customers
  - Use the *Count(Distinct)* summarization function
  - In case of wrong calculation (i.e., if you get the same value in all rows):
    - Go back to the Model
    - Double-click the relationships between CUSTOMER and SALES
    - Set the *Cross filter direction* to *Both*
- Add the average sales per customer
  - Create a new measure, calculated by dividing the sum of store sales by the count of distinct customers

# Exercise 5

- Create a table to visualize the sales for each *OCCUPATION* (*Customer* dimension)
- Exclude (i.e., filter out) the tuples where the value of *STORE\_SALES* is lower than 5
- Apply another filter (in addition to the previous one) to exclude all occupations where the total sales is lower than 80K

# Exercise 6

- Create a table to visualize with the top ten customers by total sales
  - Show both *CUSTOMER\_ID* and *FULLNAME*
- Add the *Occupation* field
- Turn it into a matrix (without the *FULLNAME*)
- Add a measure on the Customer table calculating a ranking of customers
  - First, declare a new measure simply calculating the sum of *STORE\_SALES*
  - Then, declare a new measure calculating the RANKX, where
    - The 1st parameter is the attribute that we want to order, i.e., the *CUSTOMER\_ID*
    - The 2nd parameter is the measure to be used for ordering, i.e., the one declared above
- Take the first ten customers for each occupation by filtering on the rank

# Exercise 7

- Create a histogram of StoreSales
  - Right-click on *STORE\_SALES* > New group > Create bins of size 2
  - Create a bar chart showing the count of records for each bin
- Use the same binning to plot a bar chart with the average *STORE\_COST* for each bin
  - Do you see a correlation in the data?
- Plot the same result as a scatter chart
  - Find the chart in the list of visuals
  - Put *STORE\_COST* and *STORE\_SALES* on X and Y axis, respectively (without summarizing)



# Exercise 8

- Create a new column calculating the profits
  - $PROFIT = STORE\_SALES - STORE\_COST$
- Create a line chart showing the monthly trend of profits, sales, and costs

# Part III

Open exploration of the cube

# Exercise 9

- Goal: describe sales from the perspective of customers
- Some hints:
  - Check distribution of sales (or profits) with respect to different attributes
  - Are there correlations between sales and number of customers?
  - Create bins where necessary (e.g., population)
  - Try some combinations of attributes (e.g., with the stacked bar chart)
  - Calculate the age of customers from their birthdate
    - A new column must be defined
    - `<newColumn> = DATEDIFF(<date1>, <date2>, YEAR)`

# Exercise 10

- Goal: correlate sales between customer and product attributes
- Some hints:
  - Are sales of the different product families equally distributed among genders?
    - Explore the whole hierarchy using drill-down and roll-up functionalities
  - Rank brands by yearly income of customers
    - Show the top-3 brands for each yearly income
    - Plot ranks in a matrix
      - Possibly add "Background color" to "Cell elements" to make it a heatmap

# References

Doc: <https://docs.microsoft.com/en-us/power-bi/fundamentals/>

A *lot* of YouTube videos

