DATA & AI LABORATORY

(Big) Data and Artificial Intelligence

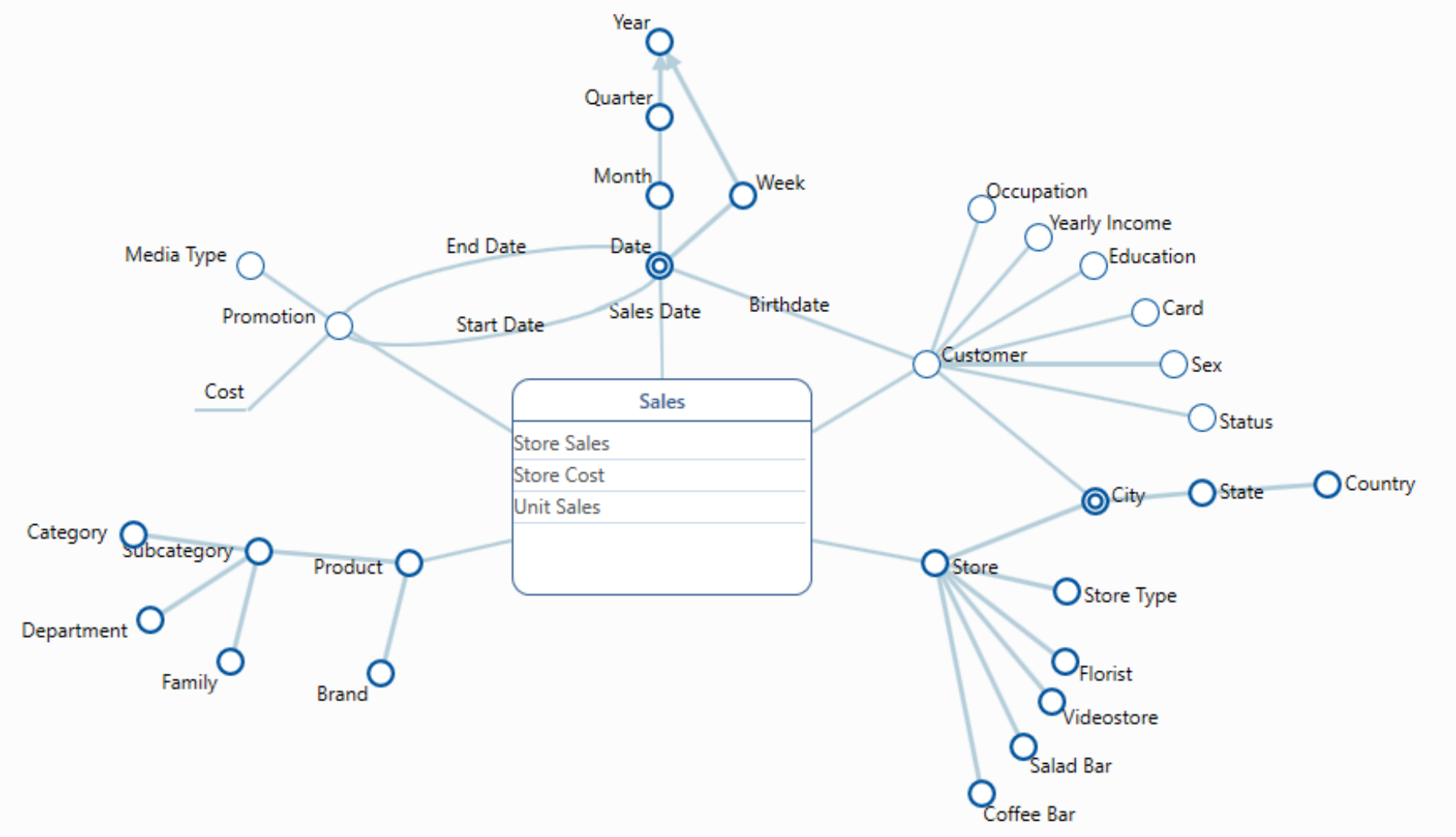
Enrico Gallinucci

28/09/2024

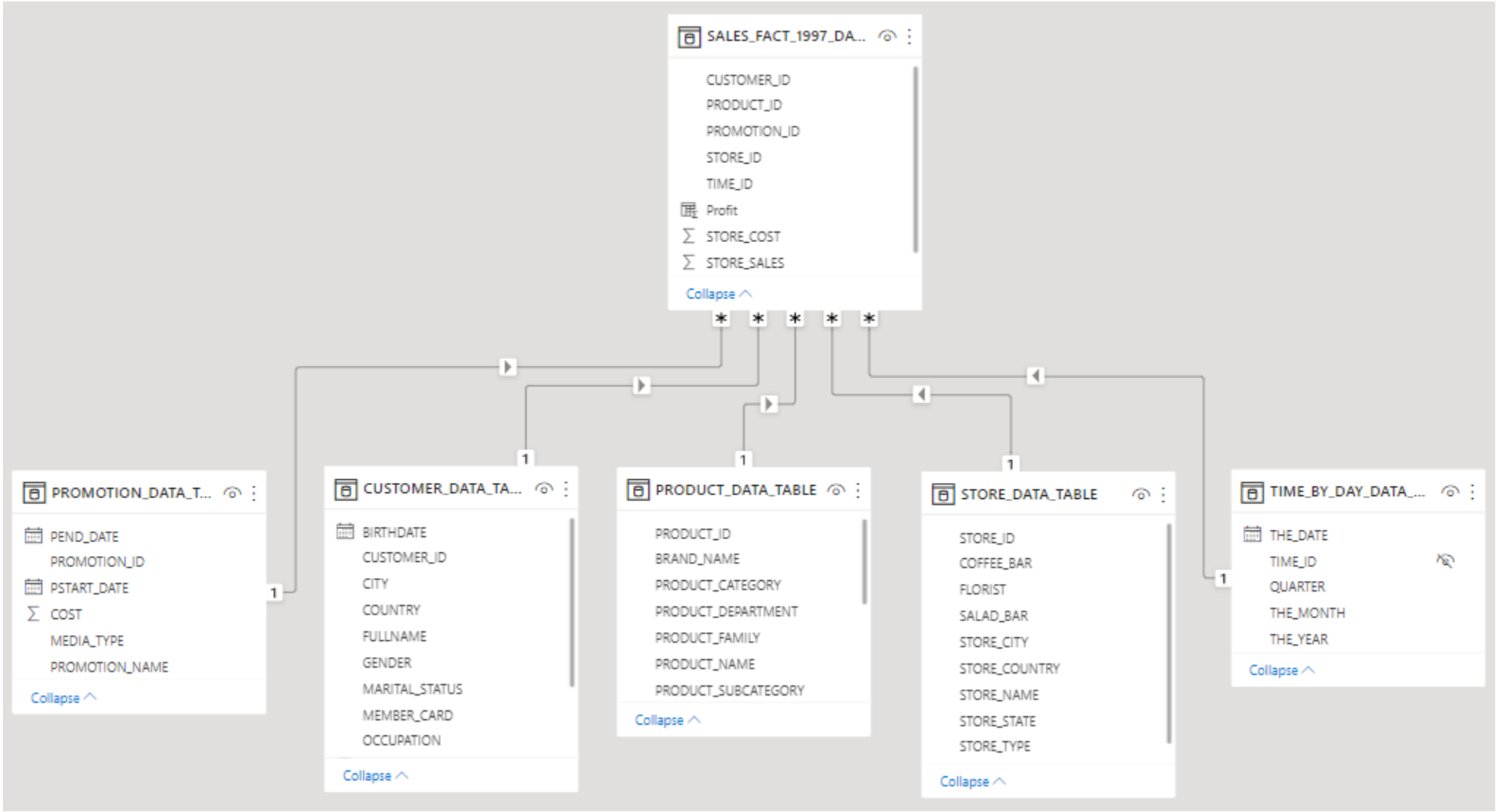
BOLOGNA BUSINESS SCHOOL
Alma Mater Studiorum Università di Bologna

A Brief Recap

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	6897	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

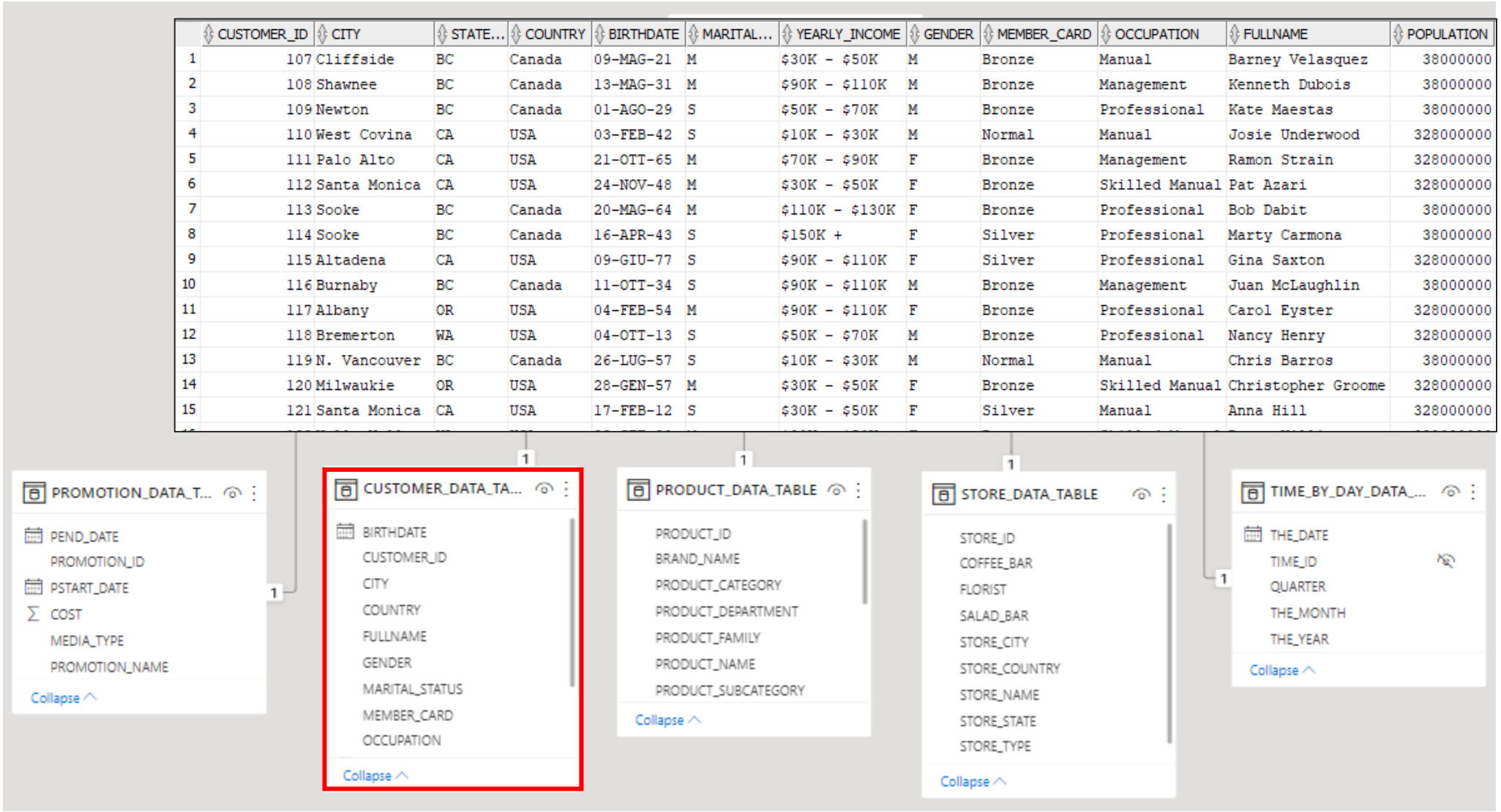
STORE_DATA_TABLE
- STORE_ID
- COFFEE_BAR
- FLORIST
- SALAD_BAR
- STORE_CITY
- STORE_COUNTRY
- STORE_NAME
- STORE_STATE
- STORE_TYPE

Collapse

TIME_BY_DAY_DATA_...
- THE_DATE
- TIME_ID
- QUARTER
- THE_MONTH
- THE_YEAR

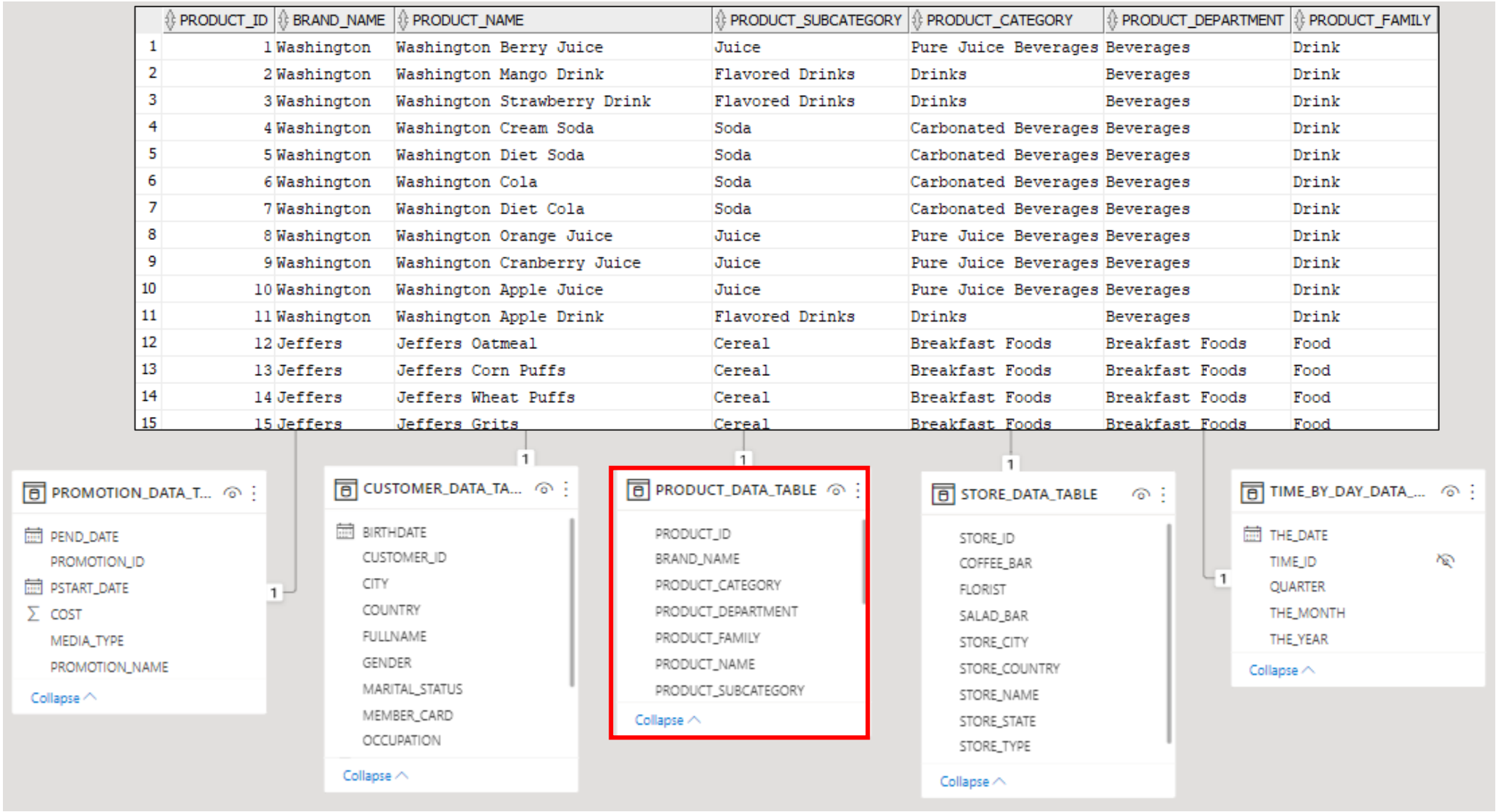
Collapse

Database Structure – Foodmart (Sales)



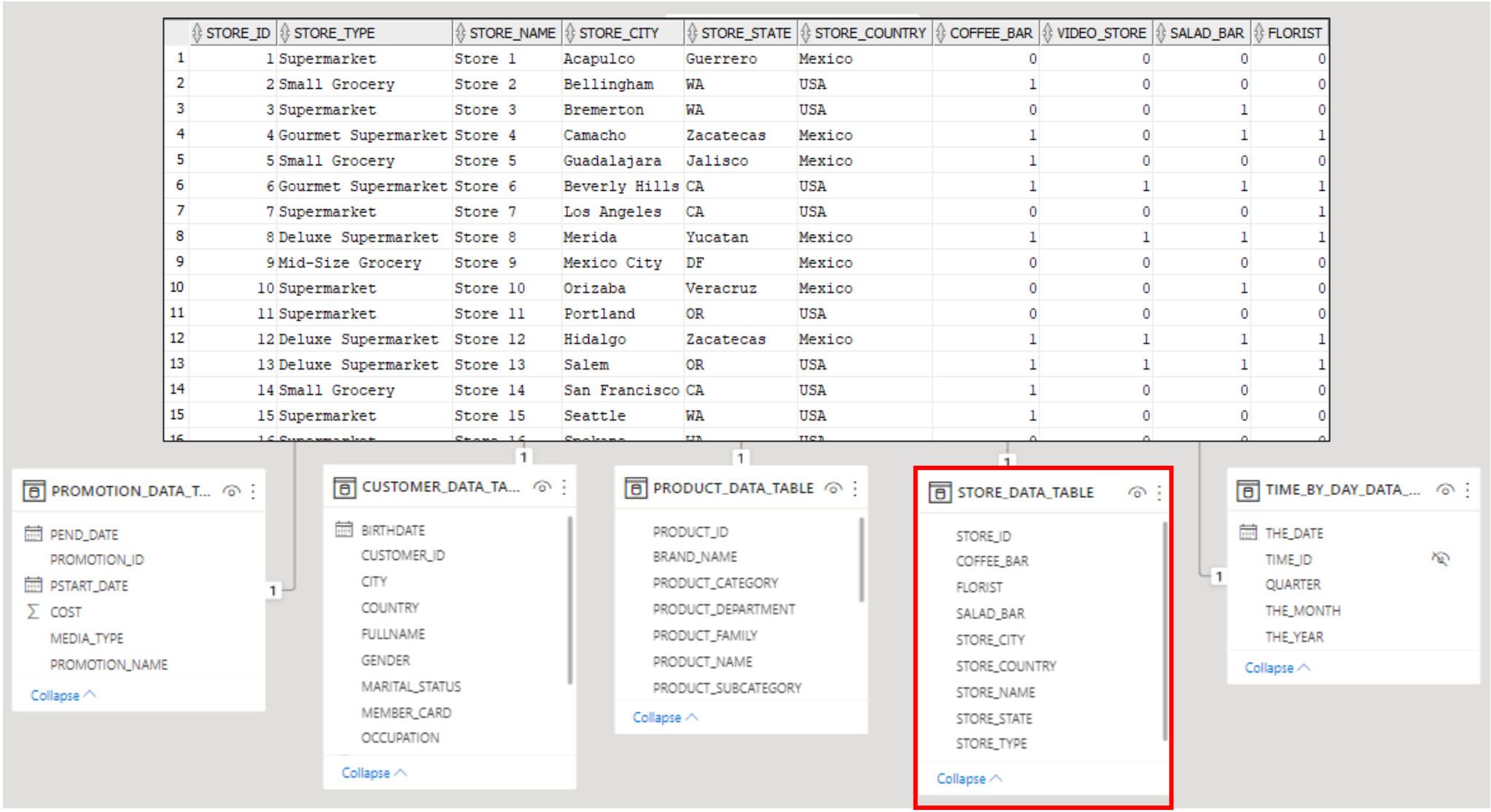
	CUSTOMER_ID	CITY	STATE...	COUNTRY	BIRTHDATE	MARITAL...	YEARLY_INCOME	GENDER	MEMBER_CARD	OCCUPATION	FULLNAME	POPULATION
1	107	Cliffside	BC	Canada	09-MAG-21	M	$30K - $50K	M	Bronze	Manual	Barney Velasquez	38000000
2	108	Shawnee	BC	Canada	13-MAG-31	M	$90K - $110K	M	Bronze	Management	Kenneth Dubois	38000000
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

Database Structure – Foodmart (Sales)



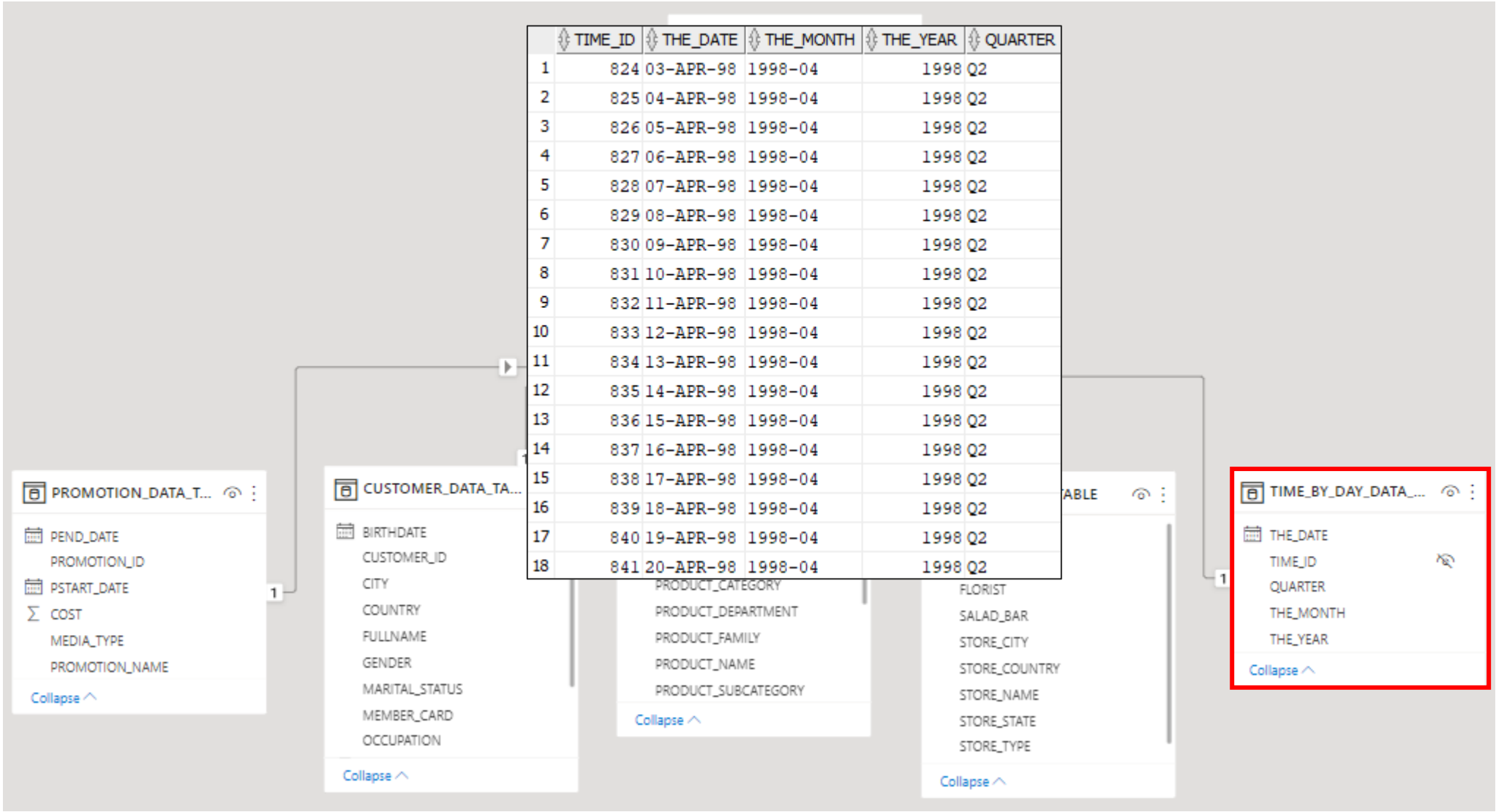
	PRODUCT_ID	BRAND_NAME	PRODUCT_NAME	PRODUCT_SUBCATEGORY	PRODUCT_CATEGORY	PRODUCT_DEPARTMENT	PRODUCT_FAMILY
1	1	Washington	Washington Berry Juice	Juice	Pure Juice Beverages	Beverages	Drink
2	2	Washington	Washington Mango Drink	Flavored Drinks	Drinks	Beverages	Drink
3	3	Washington	Washington Strawberry 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

Database Structure – Foodmart (Sales)

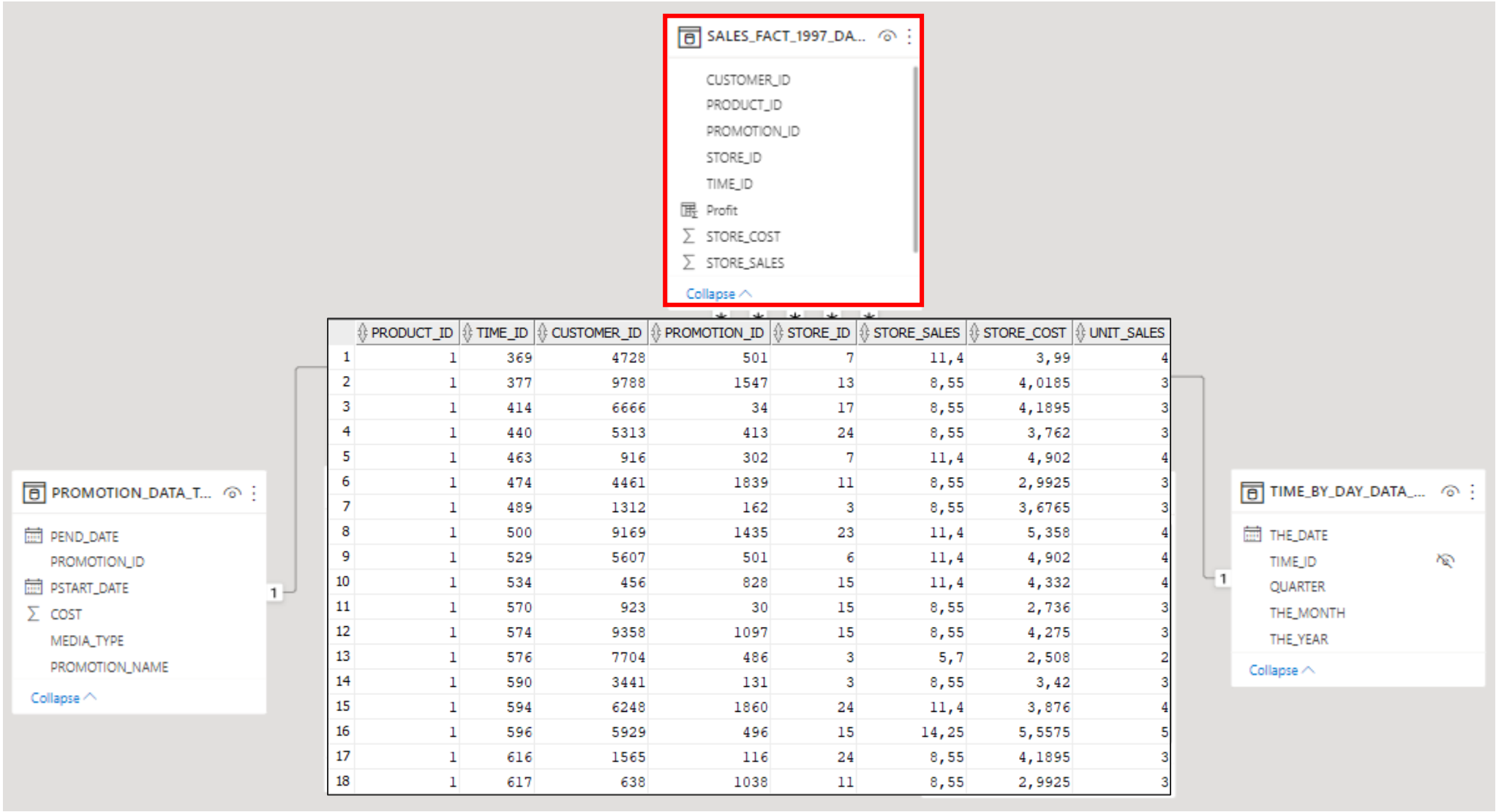


	STORE_ID	STORE_TYPE	STORE_NAME	STORE_CITY	STORE_STATE	STORE_COUNTRY	COFFEE_BAR	VIDEO_STORE	SALAD_BAR	FLORIST
1	1	Supermarket	Store 1	Acapulco	Guerrero	Mexico	0	0	0	0
2	2	Small Grocery	Store 2	Bellingham	WA	USA	1	0	0	0
3	3	Supermarket	Store 3	Bremerton	WA	USA	0	0	1	0
4	4	Gourmet Supermarket	Store 4	Camacho	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

Database Structure – Foodmart (Sales)

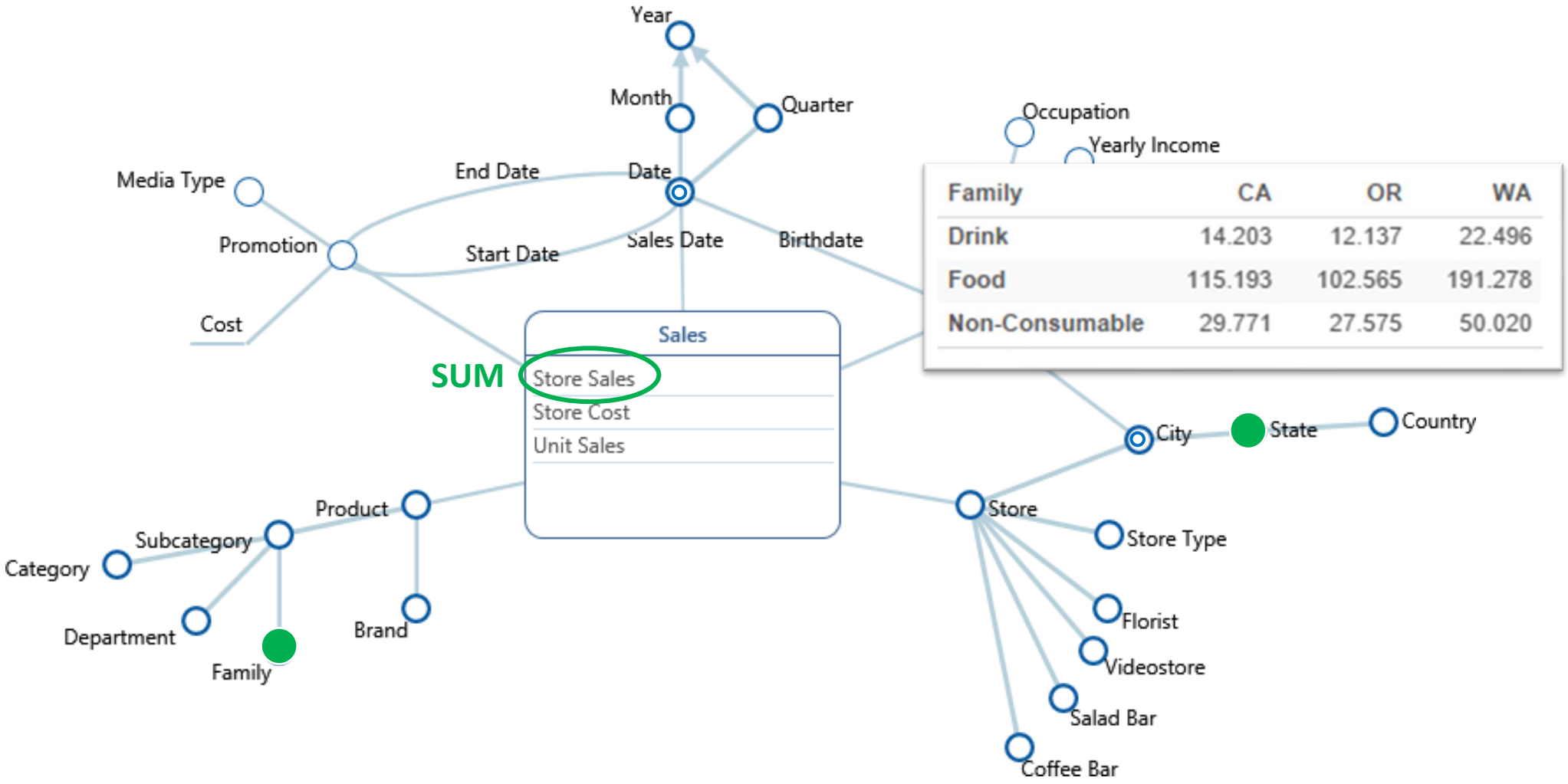


Database Structure – Foodmart (Sales)



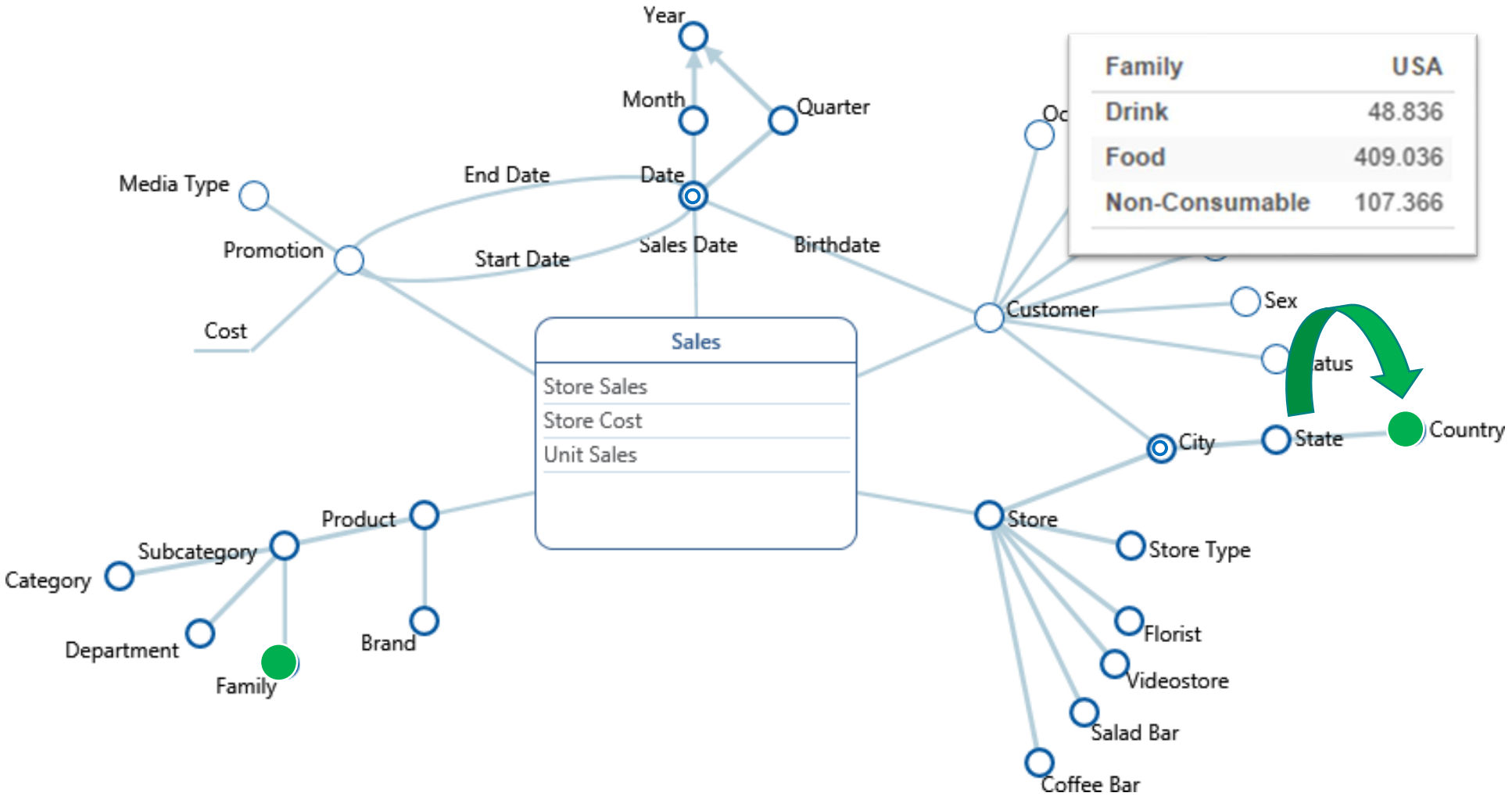
	PRODUCT_ID	TIME_ID	CUSTOMER_ID	PROMOTION_ID	STORE_ID	STORE_SALES	STORE_COST	UNIT_SALES
1	1	369	4728	501	7	11,4	3,99	4
2	1	377	9788	1547	13	8,55	4,0185	3
3	1	414	6666	34	17	8,55	4,1895	3
4	1	440	5313	413	24	8,55	3,762	3
5	1	463	916	302	7	11,4	4,902	4
6	1	474	4461	1839	11	8,55	2,9925	3
7	1	489	1312	162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

OLAP query

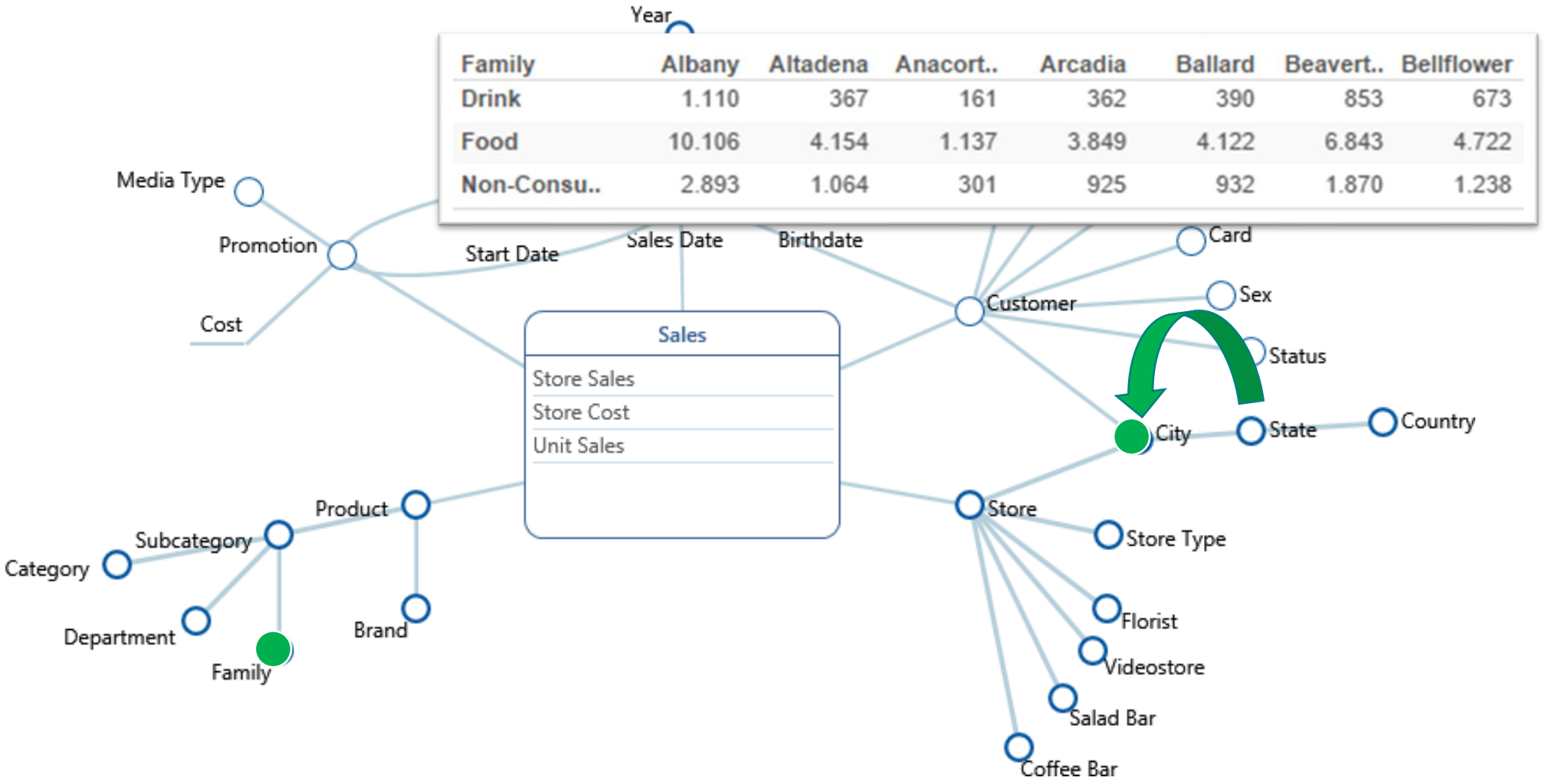


Family	CA	OR	WA
Drink	14.203	12.137	22.496
Food	115.193	102.565	191.278
Non-Consumable	29.771	27.575	50.020

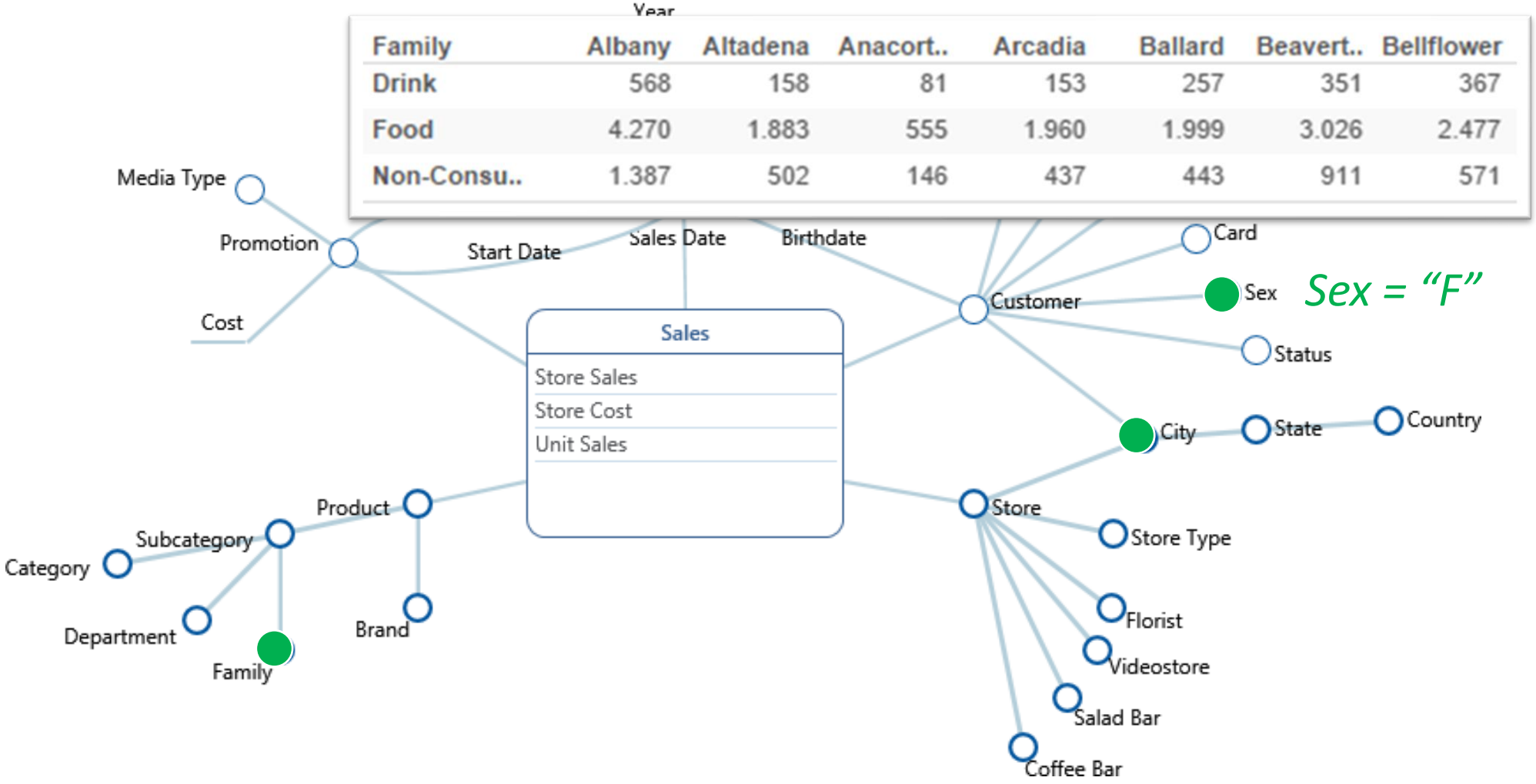
OLAP Operators: Roll-Up



OLAP Operators: Drill-down



OLAP Operators: Slice & Dice



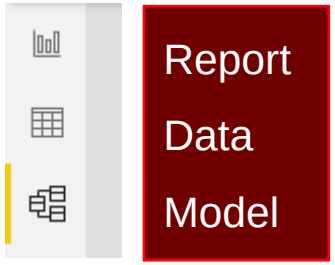
What is Power BI

- Power BI is a commercial tool for data visualization and analysis that is part of the Microsoft Power Platform

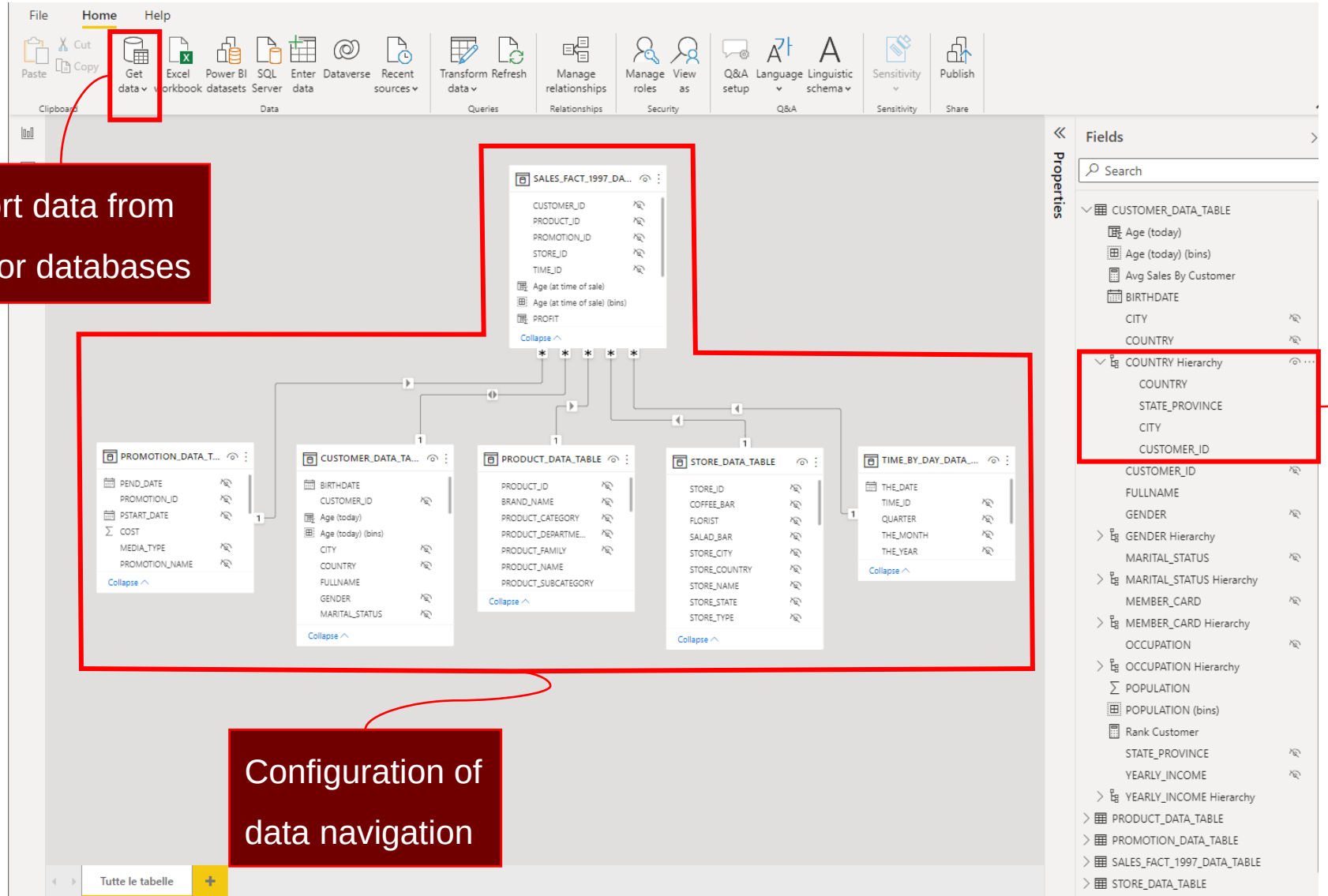
- It provides a simple GUI to formulate queries on any kind of data source
  - Not exactly an OLAP tool
  - Besides advanced BI software (e.g., Oracle, SAP), similar commercial alternatives are Tableau and Qlik
  - Open-source alternatives
    - Saiku
    - Apache Superset
  - Power BI is easy(-ish) to use and provides appealing visualizations
    - Desktop version is free

Power BI main windows

- Report
  - Create charts and dashboards
- Data
  - Change data types
- Model
  - Setup connection to data source
  - Create hierarchies



Power BI main windows: Model

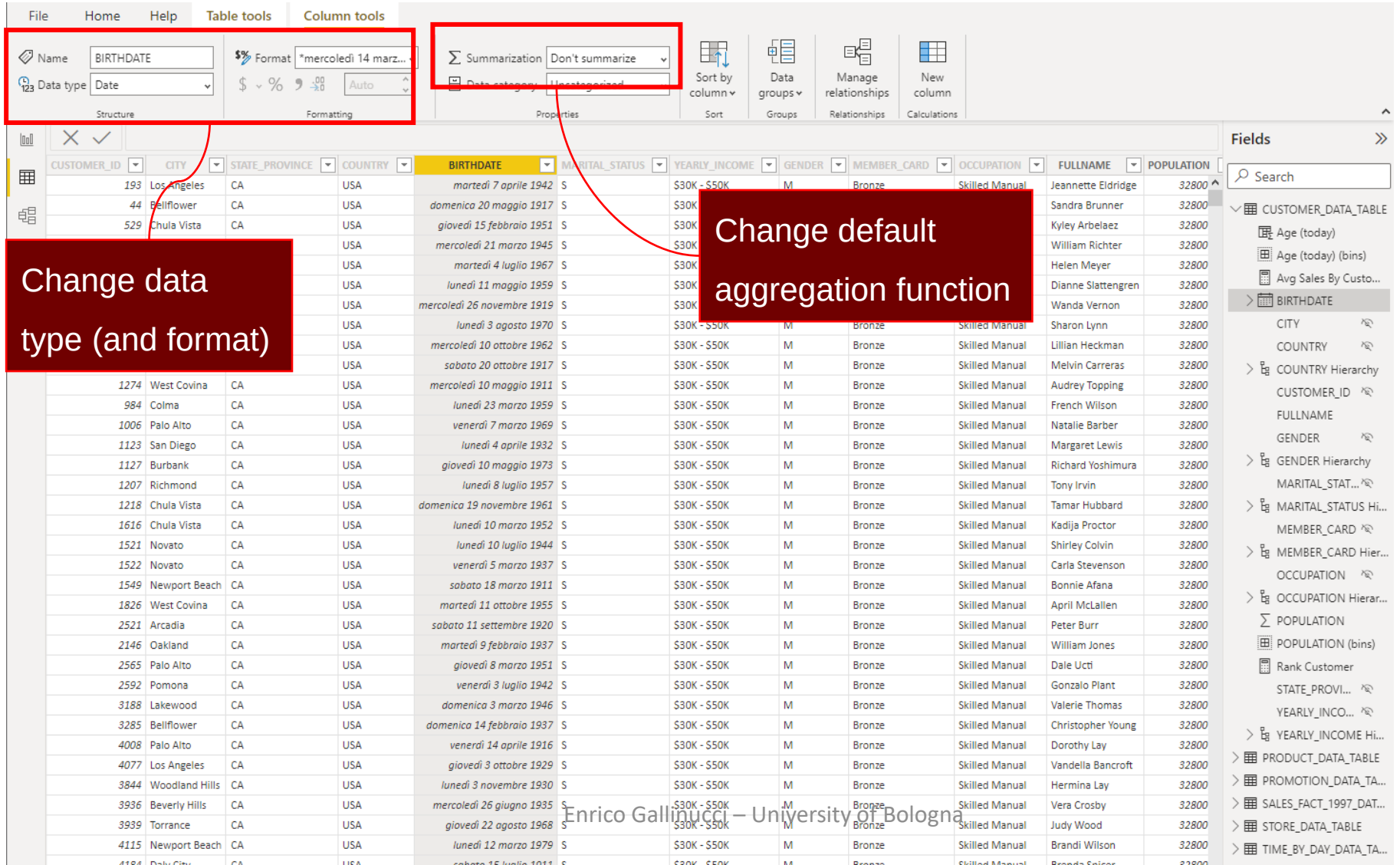


Import data from files or databases

Configuration of data navigation

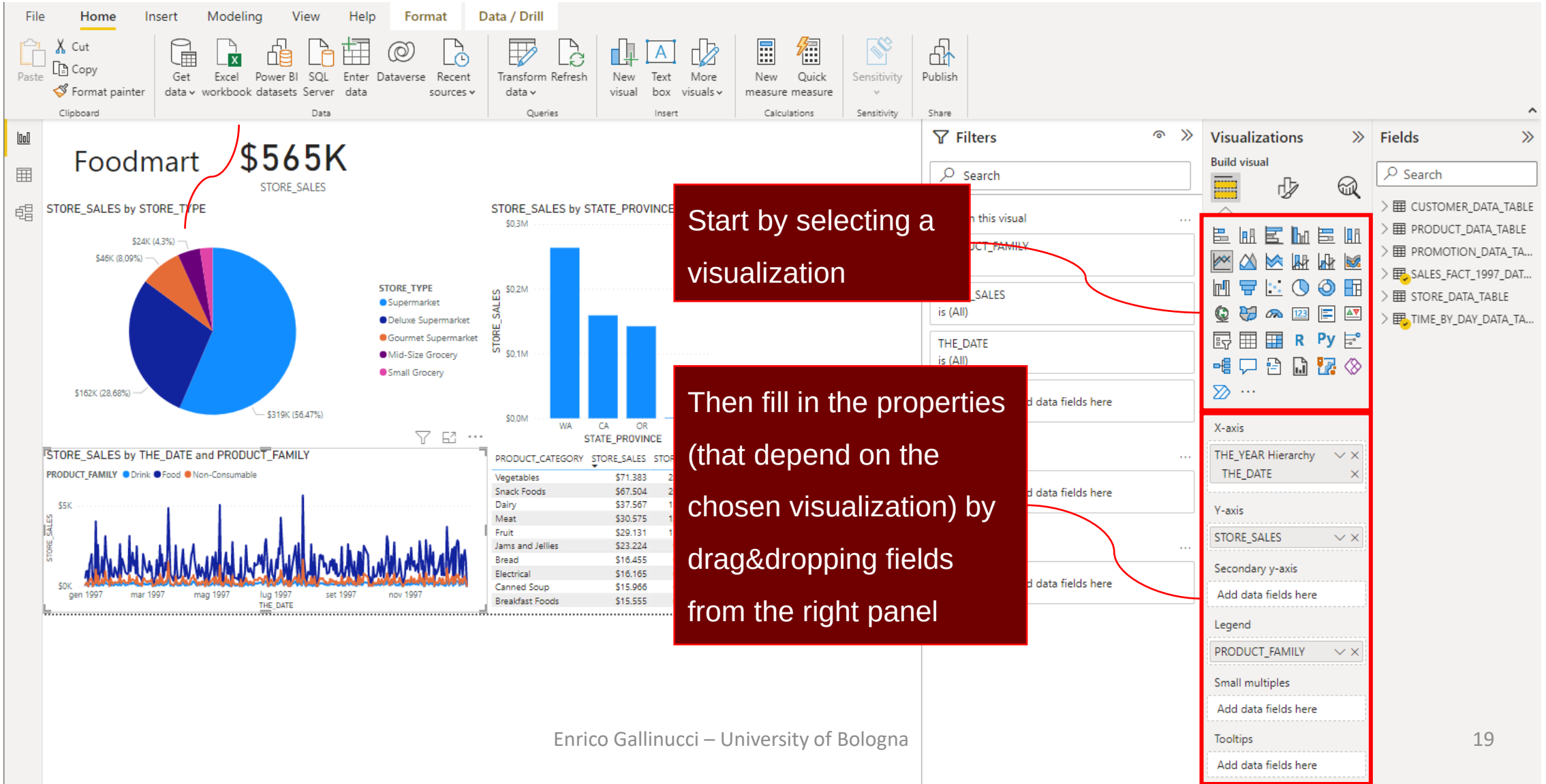
Hierarchy

Power BI main windows: Data



CUSTOMER_ID	CITY	STATE_PROVINCE	COUNTRY	BIRTHDATE	MARITAL_STATUS	YEARLY_INCOME	GENDER	MEMBER_CARD	OCCUPATION	FULLNAME	POPULATION
193	Los Angeles	CA	USA	martedì 7 aprile 1942	S	$30K - $50K	M	Bronze	Skilled Manual	Jeannette Eldridge	32800
44	Bellflower	CA	USA	domenica 20 maggio 1917	S	$30K				Sandra Brunner	32800
529	Chula Vista	CA	USA	giovedì 15 febbraio 1951	S	$30K				Kyley Arbelaez	32800
			USA	mercoledì 21 marzo 1945	S	$30K				William Richter	32800
			USA	martedì 4 luglio 1967	S	$30K				Helen Meyer	32800
			USA	lunedì 11 maggio 1959	S	$30K				Dianne Slattengren	32800
			USA	mercoledì 26 novembre 1919	S	$30K				Wanda Vernon	32800
			USA	lunedì 3 agosto 1970	S	$30K - $50K	M	Bronze	Skilled Manual	Sharon Lynn	32800
			USA	mercoledì 10 ottobre 1962	S	$30K - $50K	M	Bronze	Skilled Manual	Lillian Heckman	32800
			USA	sabato 20 ottobre 1917	S	$30K - $50K	M	Bronze	Skilled Manual	Melvin Carreras	32800
1274	West Covina	CA	USA	mercoledì 10 maggio 1911	S	$30K - $50K	M	Bronze	Skilled Manual	Audrey Topping	32800
984	Colma	CA	USA	lunedì 23 marzo 1959	S	$30K - $50K	M	Bronze	Skilled Manual	French Wilson	32800
1006	Palo Alto	CA	USA	venerdì 7 marzo 1969	S	$30K - $50K	M	Bronze	Skilled Manual	Natalie Barber	32800
1123	San Diego	CA	USA	lunedì 4 aprile 1932	S	$30K - $50K	M	Bronze	Skilled Manual	Margaret Lewis	32800
1127	Burbank	CA	USA	giovedì 10 maggio 1973	S	$30K - $50K	M	Bronze	Skilled Manual	Richard Yoshimura	32800
1207	Richmond	CA	USA	lunedì 8 luglio 1957	S	$30K - $50K	M	Bronze	Skilled Manual	Tony Irvin	32800
1218	Chula Vista	CA	USA	domenica 19 novembre 1961	S	$30K - $50K	M	Bronze	Skilled Manual	Tamar Hubbard	32800
1616	Chula Vista	CA	USA	lunedì 10 marzo 1952	S	$30K - $50K	M	Bronze	Skilled Manual	Kadija Proctor	32800
1521	Novato	CA	USA	lunedì 10 luglio 1944	S	$30K - $50K	M	Bronze	Skilled Manual	Shirley Colvin	32800
1522	Novato	CA	USA	venerdì 5 marzo 1937	S	$30K - $50K	M	Bronze	Skilled Manual	Carla Stevenson	32800
1549	Newport Beach	CA	USA	sabato 18 marzo 1911	S	$30K - $50K	M	Bronze	Skilled Manual	Bonnie Afana	32800
1826	West Covina	CA	USA	martedì 11 ottobre 1955	S	$30K - $50K	M	Bronze	Skilled Manual	April McLallen	32800
2521	Arcadia	CA	USA	sabato 11 settembre 1920	S	$30K - $50K	M	Bronze	Skilled Manual	Peter Burr	32800
2146	Oakland	CA	USA	martedì 9 febbraio 1937	S	$30K - $50K	M	Bronze	Skilled Manual	William Jones	32800
2565	Palo Alto	CA	USA	giovedì 8 marzo 1951	S	$30K - $50K	M	Bronze	Skilled Manual	Dale Ucti	32800
2592	Pomona	CA	USA	venerdì 3 luglio 1942	S	$30K - $50K	M	Bronze	Skilled Manual	Gonzalo Plant	32800
3188	Lakewood	CA	USA	domenica 3 marzo 1946	S	$30K - $50K	M	Bronze	Skilled Manual	Valerie Thomas	32800
3285	Bellflower	CA	USA	domenica 14 febbraio 1937	S	$30K - $50K	M	Bronze	Skilled Manual	Christopher Young	32800
4008	Palo Alto	CA	USA	venerdì 14 aprile 1916	S	$30K - $50K	M	Bronze	Skilled Manual	Dorothy Lay	32800
4077	Los Angeles	CA	USA	giovedì 3 ottobre 1929	S	$30K - $50K	M	Bronze	Skilled Manual	Vandella Bancroft	32800
3844	Woodland Hills	CA	USA	lunedì 3 novembre 1930	S	$30K - $50K	M	Bronze	Skilled Manual	Hermina Lay	32800
3936	Beverly Hills	CA	USA	mercoledì 26 giugno 1935	S	$30K - $50K	M	Bronze	Skilled Manual	Vera Crosby	32800
3939	Torrance	CA	USA	giovedì 22 agosto 1968	S	$30K - $50K	M	Bronze	Skilled Manual	Judy Wood	32800
4115	Newport Beach	CA	USA	lunedì 12 marzo 1979	S	$30K - $50K	M	Bronze	Skilled Manual	Brandi Wilson	32800

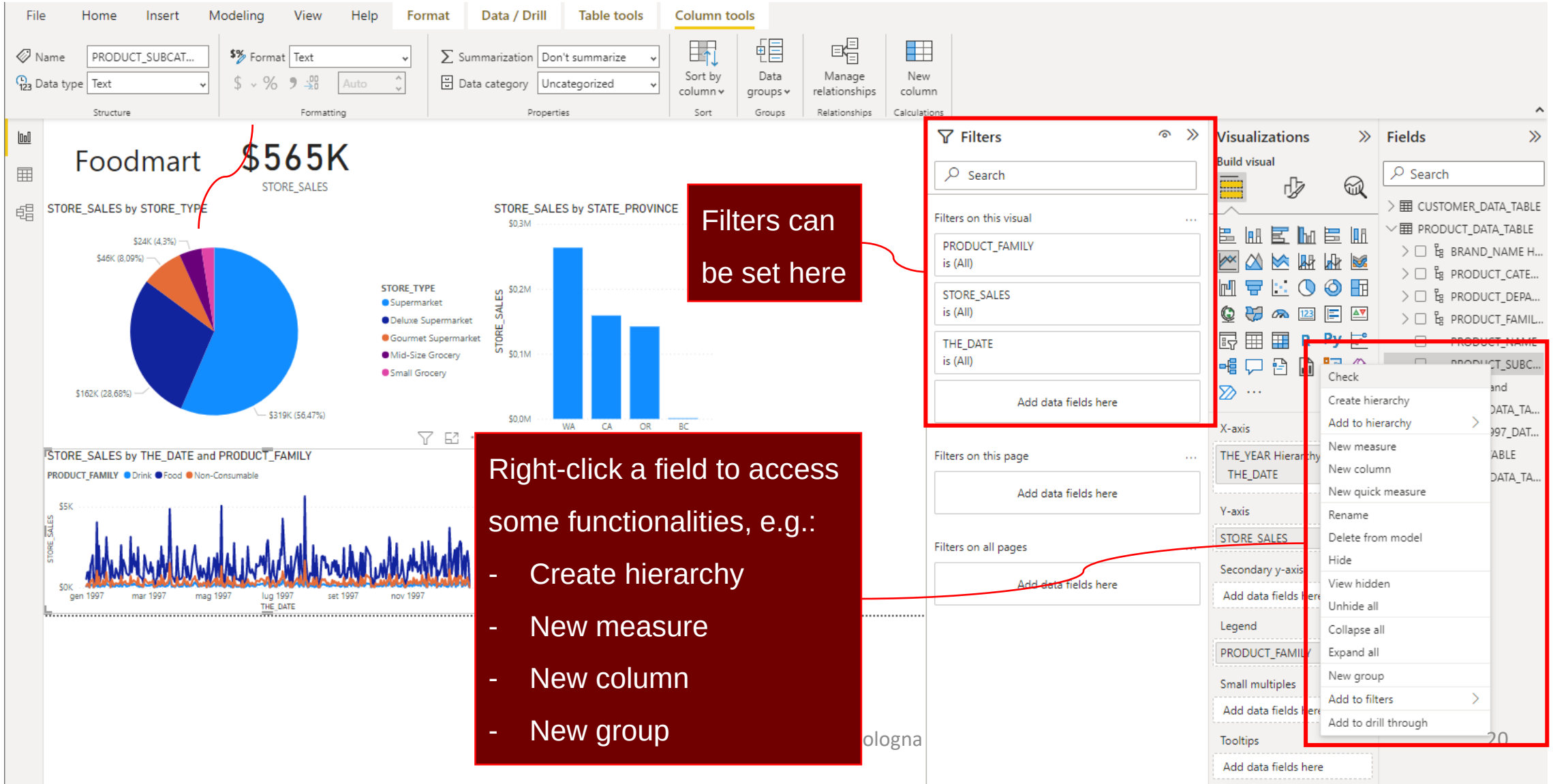
Power BI main windows: Report



Start by selecting a visualization

Then fill in the properties (that depend on the chosen visualization) by drag&dropping fields from the right panel

Power BI main windows: Report



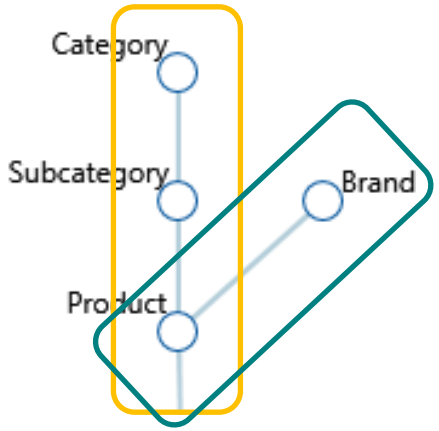
Filters can be set here

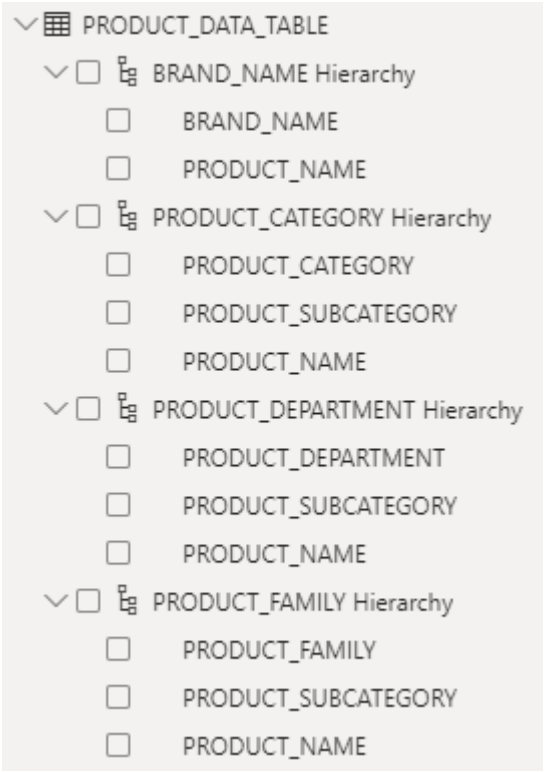
Right-click a field to access some functionalities, e.g.:

- Create hierarchy
- New measure
- New column
- New group

Hierarchies

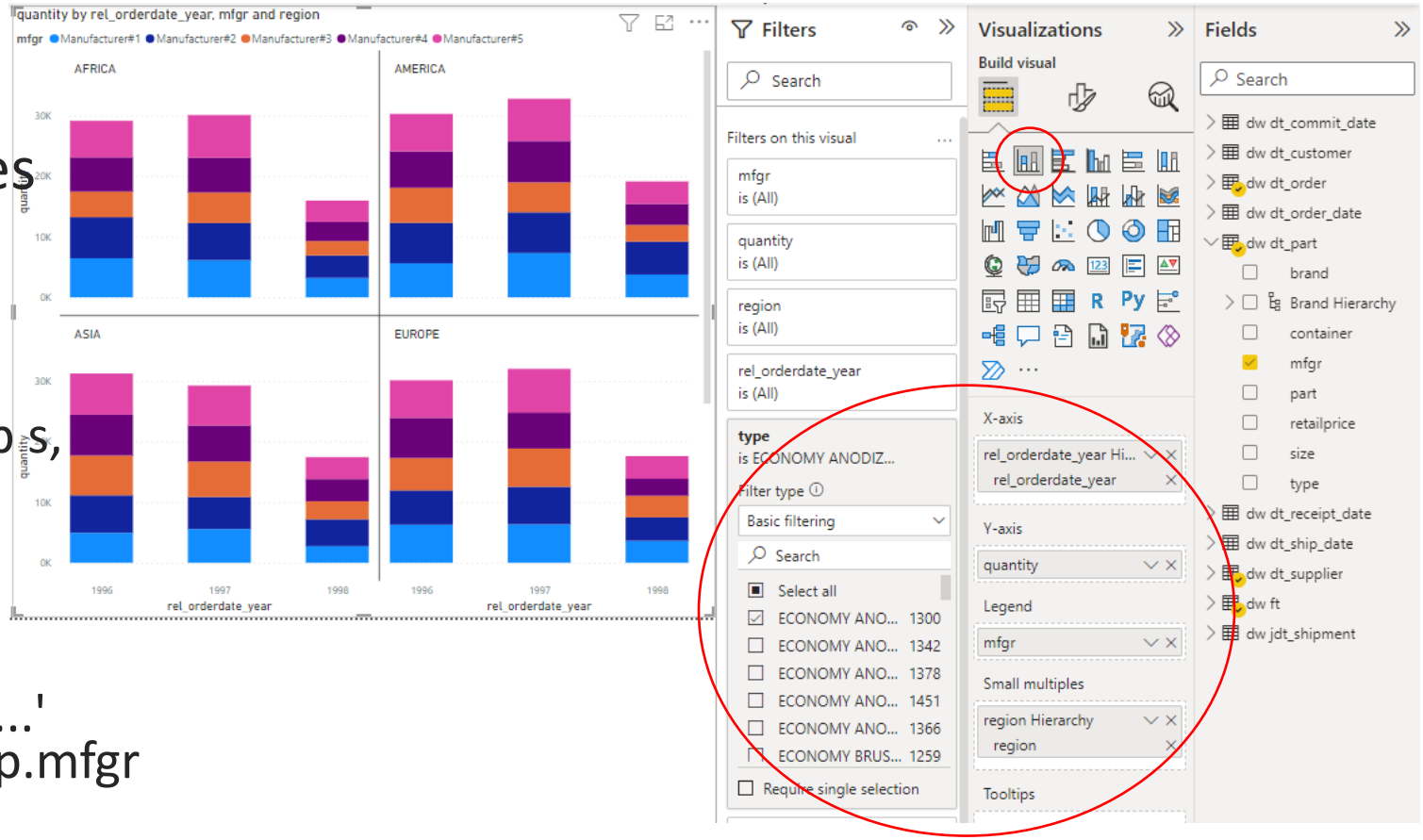
- Each path from leaves to root becomes a different hierarchy; shared attributes are duplicated





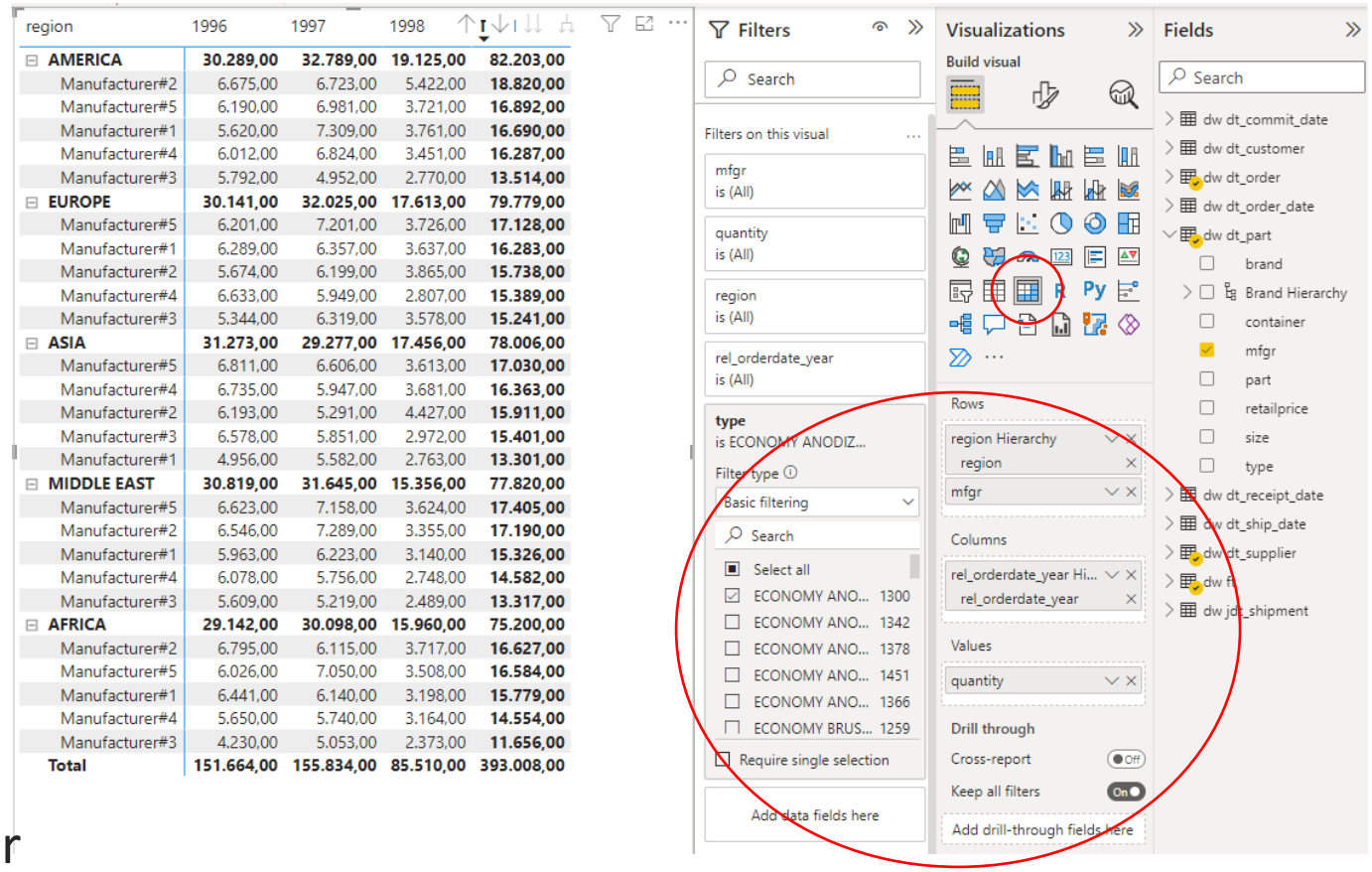
OLAP basics

- Queries are formulated by
  1) choosing a visualization
  2) drag&dropping attributes into the visualization properties
     - SELECT d.year, s.region, p.mfgr, sum(ft.quantity)
       FROM ft, dt_part p, dt_supp s, dt_order o, dt_date d
       WHERE ft.idpart = p.id
         AND ft.idsupp = s.id
         AND ft.idorder = o.id
         AND o.iddate = d.id
         AND p.type = 'ECONOMY ...'
       GROUP BY d.year, s.region, p.mfgr



OLAP basics

- The same query can be issued in different ways, where only the visualization changes (but the data is the same)
  - SELECT d.year, s.region, p.mfgr, sum(ft.quantity)
    FROM ft, dt_part p, dt_supp s, dt_order o, dt_date d
    WHERE ft.idpart = p.id
    AND ft.idsupp = s.id
    AND ft.idorder = o.id
    AND o.iddate = d.id
    AND p.type = 'ECONOMY ...'
    GROUP BY d.year, s.region, p.mfgr

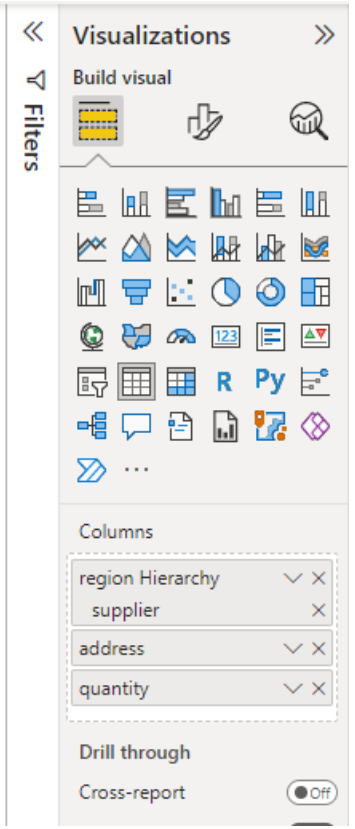


region	1996	1997	1998	
AMERICA	**30.289,00**	**32.789,00**	**19.125,00**	**82.203,00**
Manufacturer#2	6.675,00	6.723,00	5.422,00	**18.820,00**
Manufacturer#5	6.190,00	6.981,00	3.721,00	**16.892,00**
Manufacturer#1	5.620,00	7.309,00	3.761,00	**16.690,00**
Manufacturer#4	6.012,00	6.824,00	3.451,00	**16.287,00**
Manufacturer#3	5.792,00	4.952,00	2.770,00	**13.514,00**
EUROPE	**30.141,00**	**32.025,00**	**17.613,00**	**79.779,00**
Manufacturer#5	6.201,00	7.201,00	3.726,00	**17.128,00**
Manufacturer#1	6.289,00	6.357,00	3.637,00	**16.283,00**
Manufacturer#2	5.674,00	6.199,00	3.865,00	**15.738,00**
Manufacturer#4	6.633,00	5.949,00	2.807,00	**15.389,00**
Manufacturer#3	5.344,00	6.319,00	3.578,00	**15.241,00**
ASIA	**31.273,00**	**29.277,00**	**17.456,00**	**78.006,00**
Manufacturer#5	6.811,00	6.606,00	3.613,00	**17.030,00**
Manufacturer#4	6.735,00	5.947,00	3.681,00	**16.363,00**
Manufacturer#2	6.193,00	5.291,00	4.427,00	**15.911,00**
Manufacturer#3	6.578,00	5.851,00	2.972,00	**15.401,00**
Manufacturer#1	4.956,00	5.582,00	2.763,00	**13.301,00**
MIDDLE EAST	**30.819,00**	**31.645,00**	**15.356,00**	**77.820,00**
Manufacturer#5	6.623,00	7.158,00	3.624,00	**17.405,00**
Manufacturer#2	6.546,00	7.289,00	3.355,00	**17.190,00**
Manufacturer#1	5.963,00	6.223,00	3.140,00	**15.326,00**
Manufacturer#4	6.078,00	5.756,00	2.748,00	**14.582,00**
Manufacturer#3	5.609,00	5.219,00	2.489,00	**13.317,00**
AFRICA	**29.142,00**	**30.098,00**	**15.960,00**	**75.200,00**
Manufacturer#2	6.795,00	6.115,00	3.717,00	**16.627,00**
Manufacturer#5	6.026,00	7.050,00	3.508,00	**16.584,00**
Manufacturer#1	6.441,00	6.140,00	3.198,00	**15.779,00**
Manufacturer#4	5.650,00	5.740,00	3.164,00	**14.554,00**
Manufacturer#3	4.230,00	5.053,00	2.373,00	**11.656,00**
Total	**151.664,00**	**155.834,00**	**85.510,00**	**393.008,00**

OLAP basics

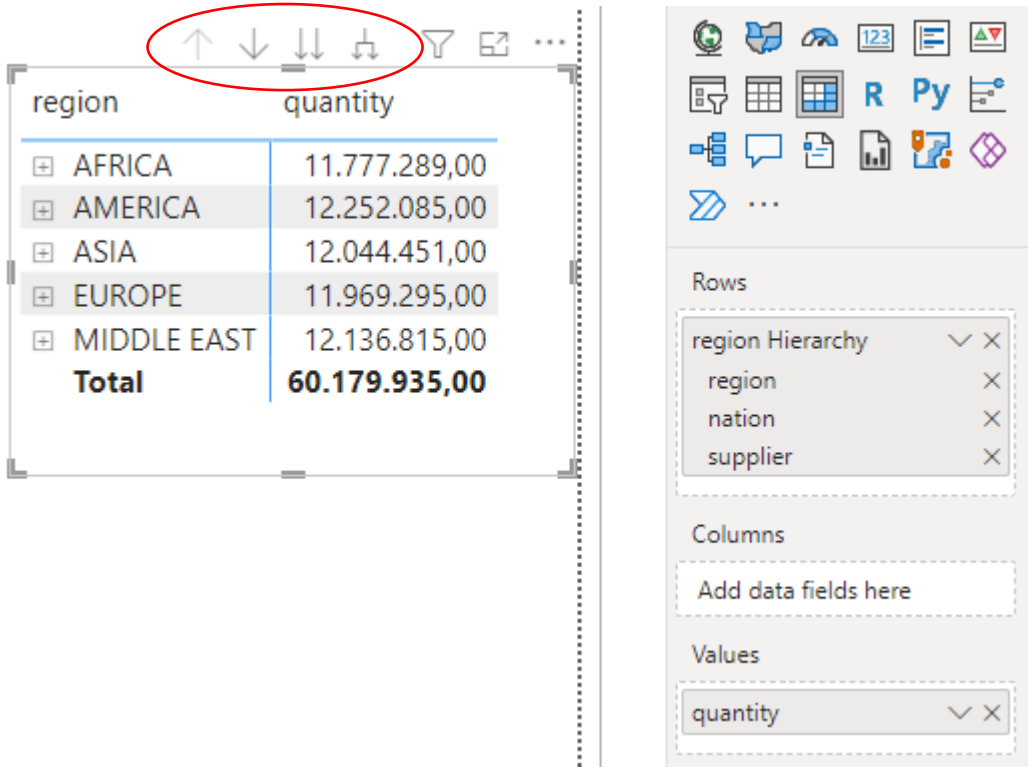
- Descriptive attributes can be used only in combination with the attribute that they describe

supplier	address	quantity
Supplier#000000001	N kw4gn1OM Ahw3Sg70BBgQw57lgjzj55R	6.106,00
Supplier#000000002	ji3yh016B5	5.975,00
Supplier#000000003	mxBQBnxO3CSwl7	5.767,00
Supplier#000000004	7zR323R73NMB77wi1	5.445,00
Supplier#000000005	AmMQ7Mg 10ByLCP52M13xN31jh5hzOgnm00B	5.239,00
Supplier#000000006	QQL6hxmnMkkgMwgm7CB5B 30Llz	5.180,00
Supplier#000000007	z45m2jBRzI5ilLNz4	5.973,00
Supplier#000000008	xz5m4C A4AAj0kANQ	4.681,00
Supplier#000000009	m7k7CnC3wiP	5.564,00
Supplier#000000010	wN1S4mQ0g7Px5Lj34xw6kS4Li4NzB4mO	5.556,00
Supplier#000000011	ikz2MhBMQAg	6.294,00
Supplier#000000012	Nym5z1Si43B 2yQxm2yOg4Q45kRLxg6ymgz4w0	5.764,00
Supplier#000000013	M6y461ChMgR6gwM3m275kQm6kQ	6.448,00
Supplier#000000014	l4iQ47yygAilnQiS2yg2	5.444,00
Supplier#000000015	Mh72i PQAxzwO5MAiONw7C0LIA	5.527,00
Supplier#000000016	B7wLkSLRjNS MS1C	5.855,00
Supplier#000000017	wPgjxnQPOnz	5.845,00
Supplier#000000018	2LAO i0 Q5	6.370,00
Supplier#000000019	2i024kik5LxzOiPxCOwBPz3jx C71 QyOgiROOP0	6.237,00
Total		**60.179.935,00**



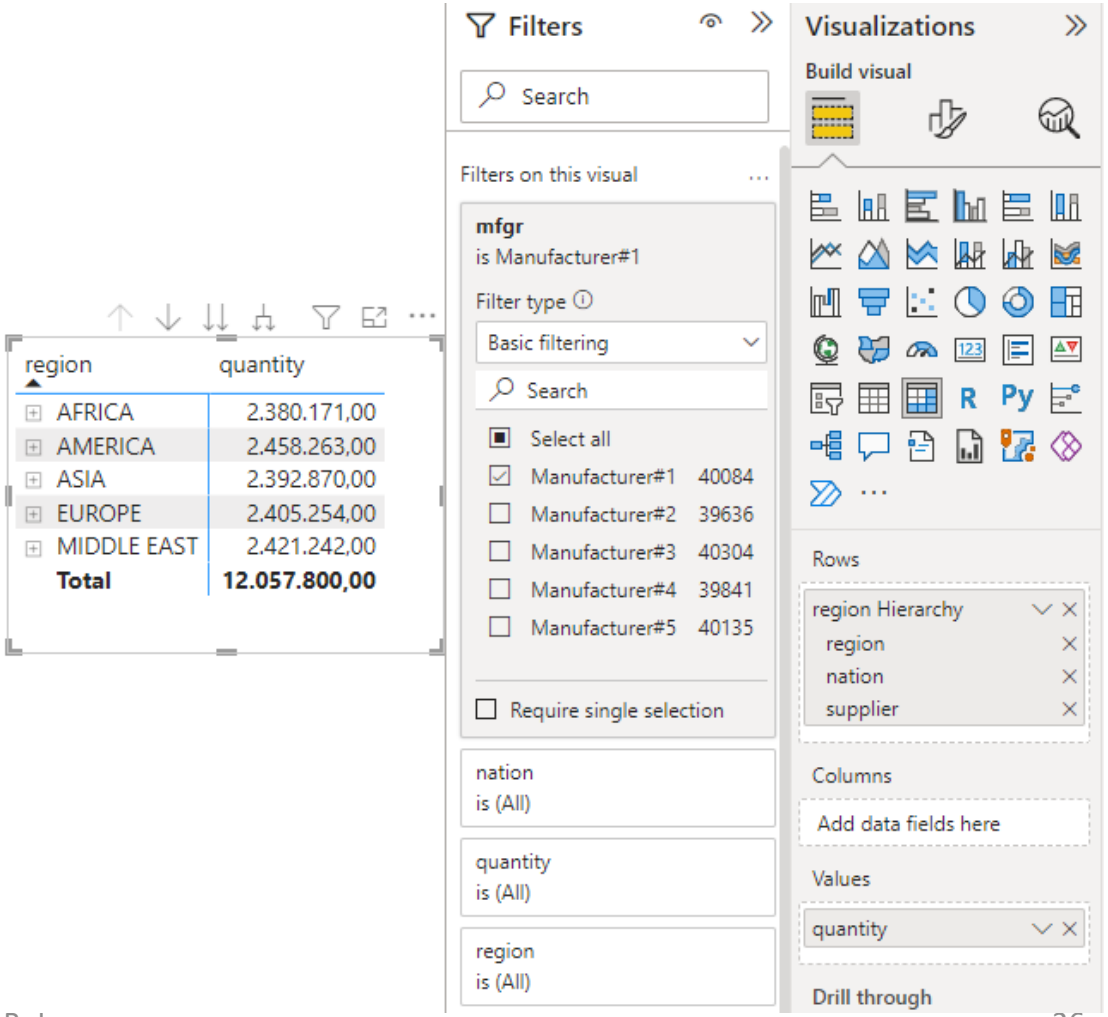
OLAP basics

- Navigation of hierarchies is enabled when hierarchies are used in some (not all) visualizations
  - Drill-up: corresponds to roll-up
  - Drill-down: interactive mode to slice-and-drill
  - Next-level: drills-down and replaces the current level with the finer one in the hierarchy
  - Expand-all: drills-down and keeps the current level



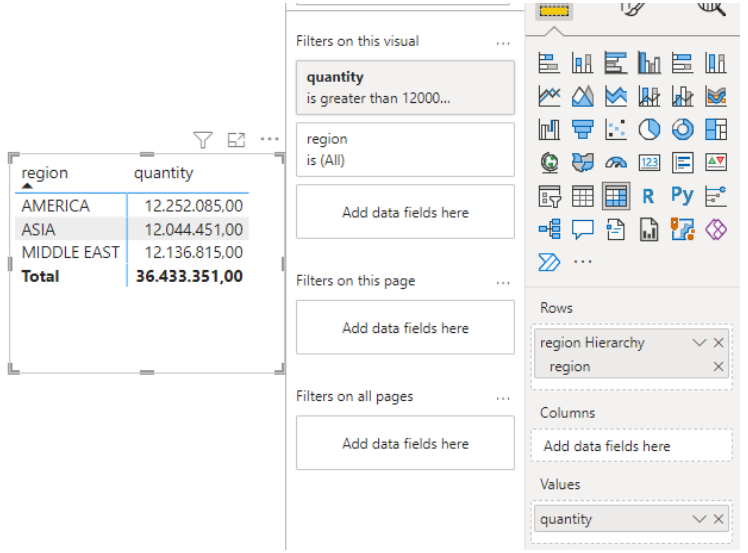
OLAP basics

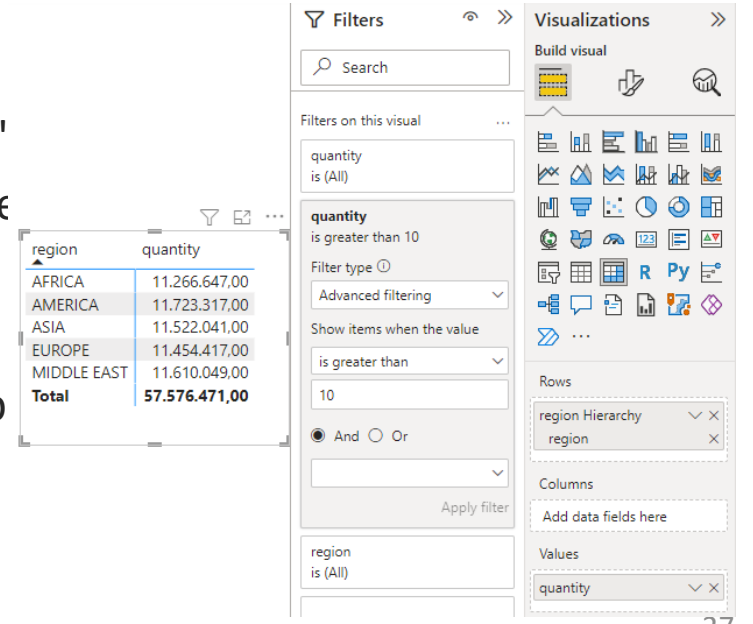
- Slice: open the filters panel
  - Either add an attribute/measure (e.g., *mfgr*)
  - Or expand an attribute/measure already used in the query (e.g., *region*)



OLAP basics

- Slice: open the filters panel
  - On attributes: select one or more categorical values
  - On measures: select a range of values, either pre- or post-aggreagtion
    - Post-aggregation (top-right figure)
      - "SELECT [..] GROUP BY [..] *HAVING SUM(quantity) > 12M*'
      - Done when filtering on the attribute that is already in the
    - Pre-aggregation (bottom-right figure)
      - "SELECT [..] *WHERE quantity > 10* [..]«
      - Requires drag&dropping *quantity* again from the Fields p
  - Filter type "Top N" al

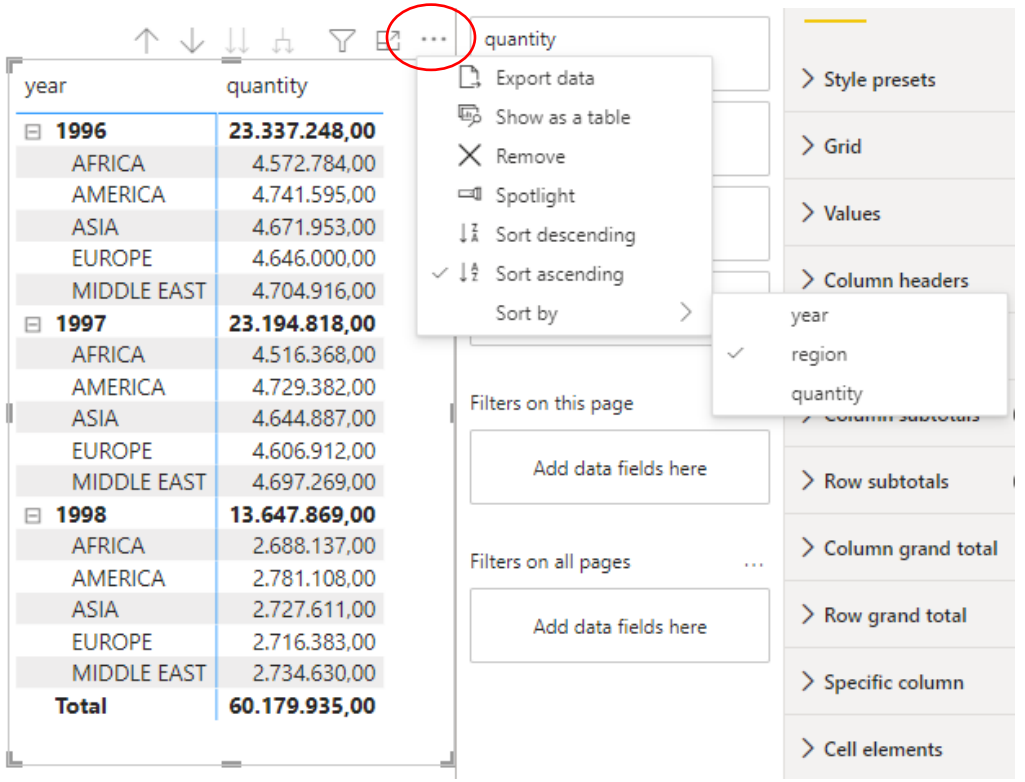




OLAP basics

- Ordering
  - Click on an attribute in the visualization
  - Or click the three dots in the top-right corner

region	quantity
AMERICA	12.252.085,00
MIDDLE EAST	12.136.815,00
ASIA	12.044.451,00
EUROPE	11.969.295,00
AFRICA	11.777.289,00
Total	**60.179.935,00**



More functionalities: visualization

- Visualizations > Legend
  - Use to break the marks (e.g., bars, trend lines, pie slices) into multiple sub-marks
  - [see OLAP basics]
- Visualizations > Small multiples
  - Use to break the chart into multiple sub-charts
  - [see OLAP basics]
- Data/Drill > Visual table
  - Use to get a table-like visualization of the chart
- Data/Drill > Data point table
  - Useful to export to check the original values before the aggregation

More functionalities: new fields

- Right-click on an attribute > Create > ...
  - > Group (discrete attribute)
    - Manually take some members from an attribute and put them in a new attribute
  - > Group (numeric attribute)
    - Discretize numerical into bins (e.g., the account balance of customers)
    - Support only for equi-width binning
- Table tools > …
  - New column
    - Create a new (virtual column) based on a custom formula
  - New measure
    - Define a function to aggregate data (to be used in place of SUM(), AVG(), etc.)

More functionalities: columns

- Create a new (virtual column) based on a custom formula
- Examples in FT
  - *profit = 'dw ft'[extendedprice] * 'dw ft'[quantity]*
  - *taxed = IF('dw ft'[tax]>0,"taxed","not taxed")*

- In DT_CUSTOMER
  - *order_age = DATEDIFF('dw dt_order'[rel_orderdate_date], TODAY(), YEAR)*

More functionalities: measures

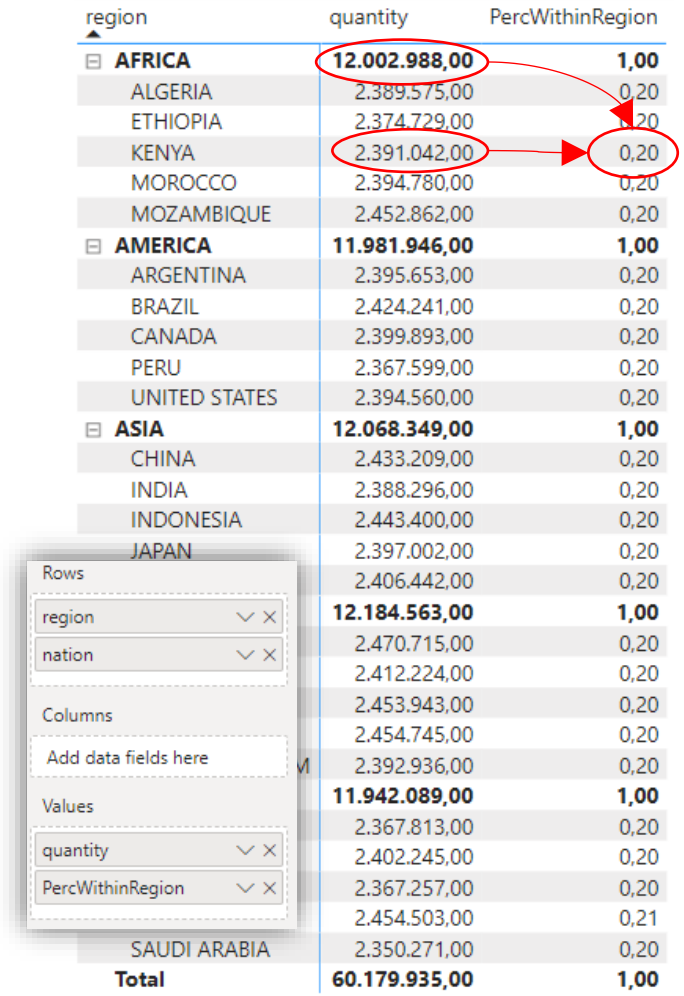
- Define a function to aggregate data (to be used in place of SUM(), AVG(), etc.)
- Examples in FT:
  - *sum_of_profit = SUM('dw ft'[profit])*
  - *avg_profit_by_customer = SUM('dw ft'[profit]) / DISTINCTCOUNT('dw dt_customer'[custkey])*

More functionalities: advanced measures

- Use coarser data at a finer aggregation level
  - For instance:
    - Calculate the sum(Quantity) by Nation (and Region)
    - Divide the result by the sum(Quantity) by Region
    - *PercWithinRegion = SUM('dw ft'[quantity]) /*
      CALCULATE(
      SUM('dw ft'[quantity]),
      REMOVEFILTERS('dw dt_customer'[nation])
)

CALCULATE computes the *SUM(qty)* by removing *nation* from the group by defined in the visualization properties

CALCULATE returns just a value, whereas SUMMARIZE returns a table



region	quantity	PercWithinRegion
AFRICA	**12.002.988,00**	**1,00**
ALGERIA	2.389.575,00	0,20
ETHIOPIA	2.374.729,00	0,20
KENYA	2.391.042,00	0,20
MOROCCO	2.394.780,00	0,20
MOZAMBIQUE	2.452.862,00	0,20
AMERICA	**11.981.946,00**	**1,00**
ARGENTINA	2.395.653,00	0,20
BRAZIL	2.424.241,00	0,20
CANADA	2.399.893,00	0,20
PERU	2.367.599,00	0,20
UNITED STATES	2.394.560,00	0,20
ASIA	**12.068.349,00**	**1,00**
CHINA	2.433.209,00	0,20
INDIA	2.388.296,00	0,20
INDONESIA	2.443.400,00	0,20
JAPAN	2.397.002,00	0,20
	2.406.442,00	0,20
	12.184.563,00	**1,00**
	2.470.715,00	0,20
	2.412.224,00	0,20
	2.453.943,00	0,20
	2.454.745,00	0,20
	2.392.936,00	0,20
	11.942.089,00	**1,00**
	2.367.813,00	0,20
	2.402.245,00	0,20
	2.367.257,00	0,20
	2.454.503,00	0,21
SAUDI ARABIA	2.350.271,00	0,20
Total	**60.179.935,00**	**1,00**

References

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

A *lot* of YouTube videos