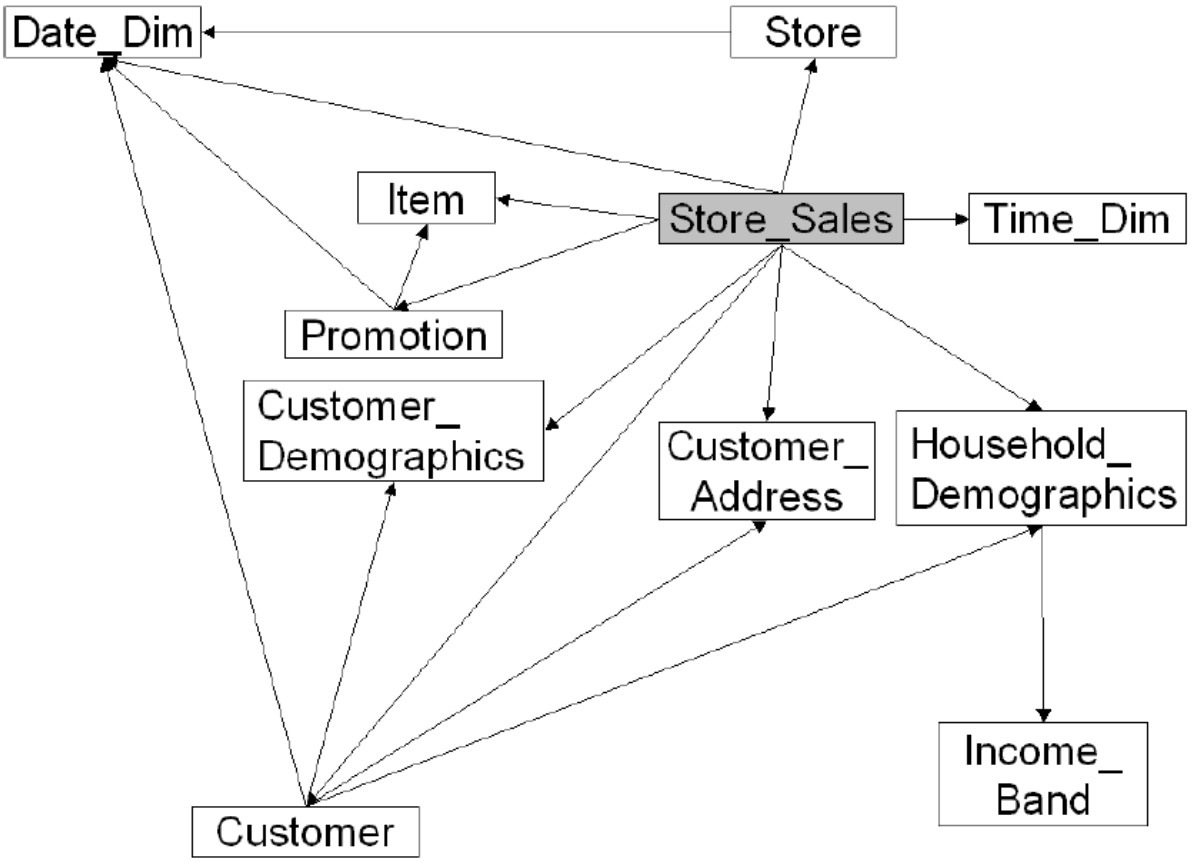
Course: Algorithmic Digital Marketing
Professor: Srikanth Krishnamurthy
Students: Pramod Gopal, Jeet Khemani
Assignment: Analytics and dashboard on snowflake using TPC- DS dataset

Introduction

- TPC-DS models the decision support functions of a retail product supplier.
- The supporting schema contains vital business information, such as customer, order, and product data
- TPC-DS consists of 7 fact tables and 17 dimension tables
- The TPC-DS schema models the sales and sales returns process for an organization that employs three primary sales channels: stores, catalogs, and the Internet

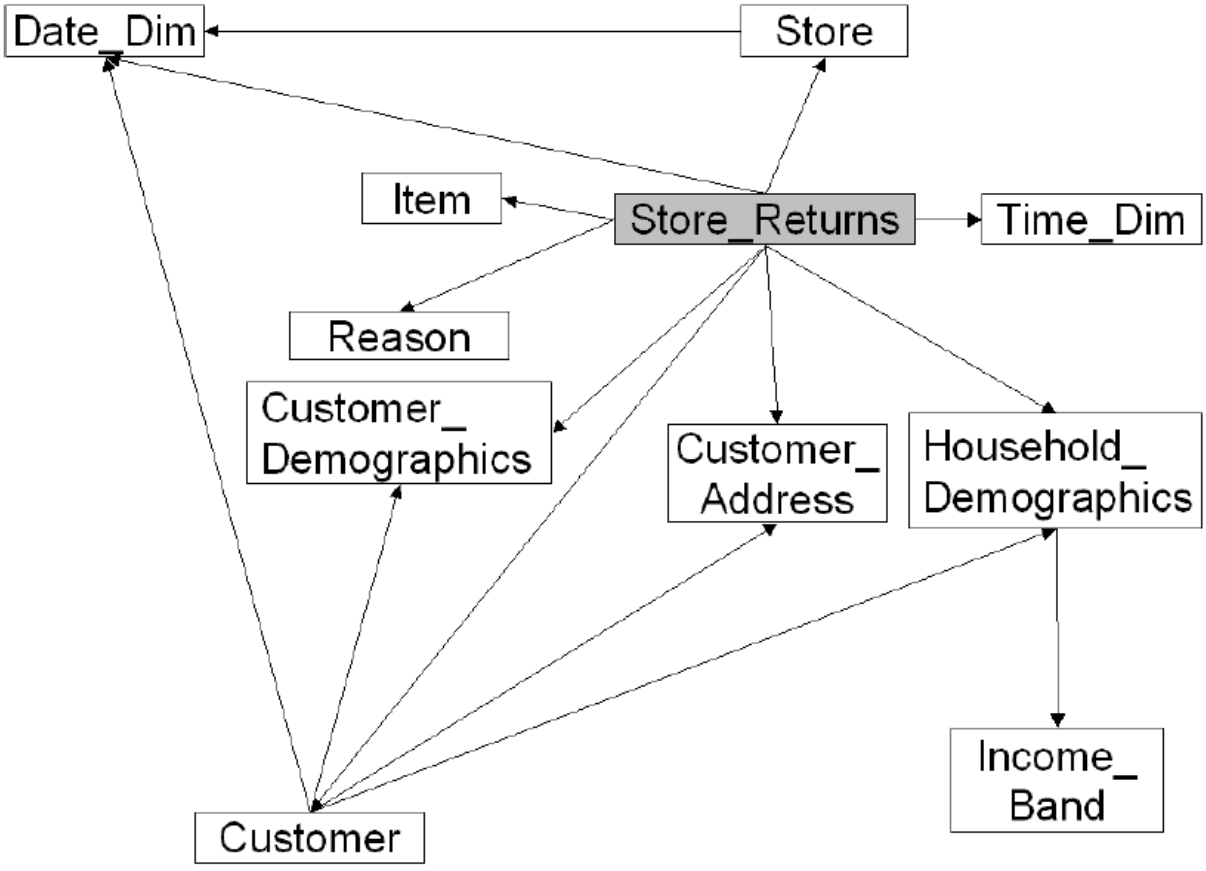
Fact Table Definitions

1. **Store Sales**
   Each row in this table represents a single line item for a sale made through the store channel and recorded in the store_sales fact table.



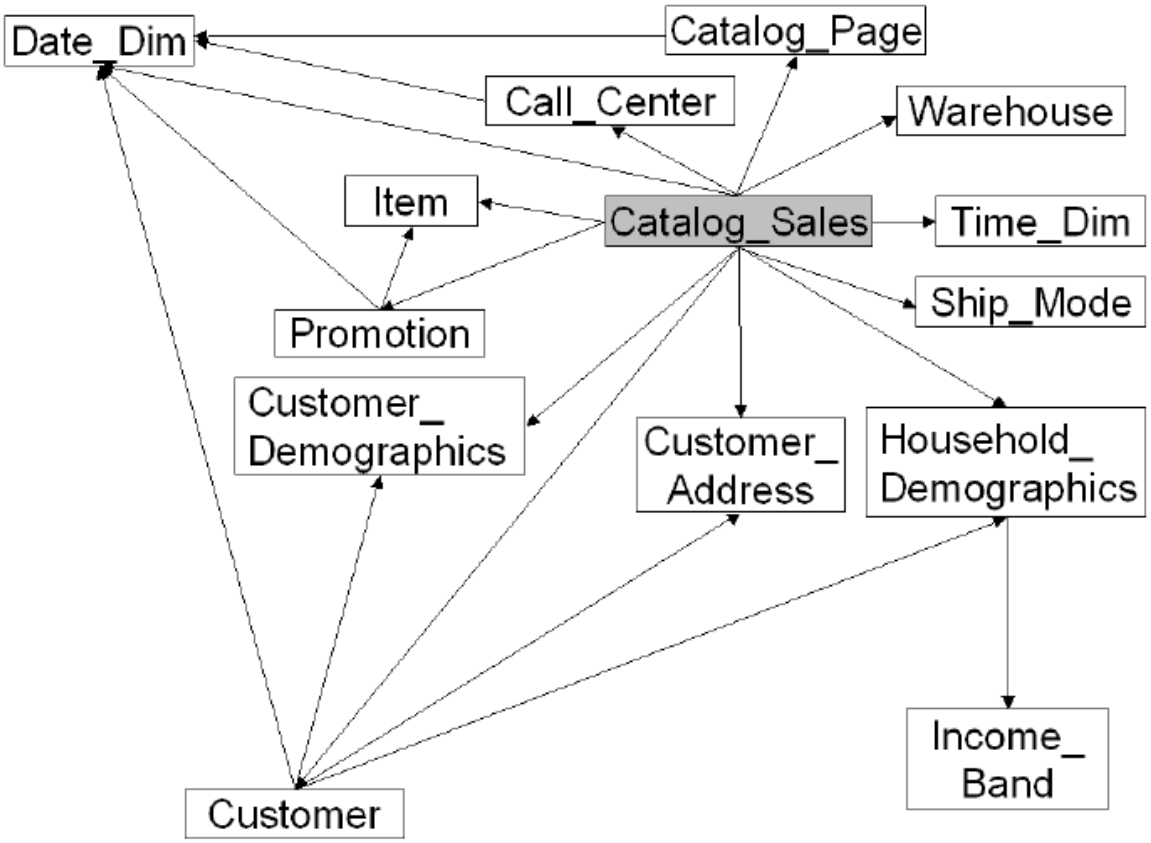
2. **Store Returns (SR)**

   Each row in this table represents a single line item for the return of an item sold through the store channel and recorded in the store_returns fact table



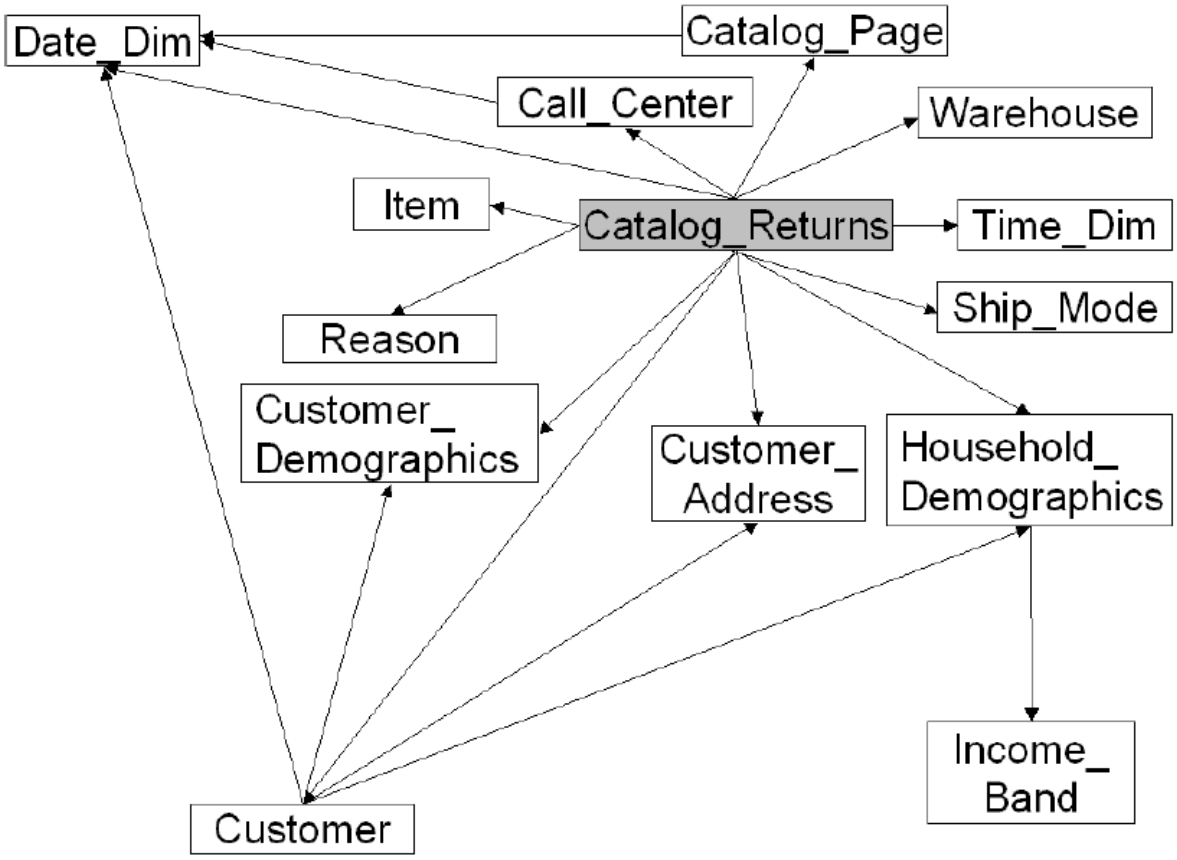
3. **Catalog Sales (CS)**

   Each row in this table represents a single line item for a sale made through the catalog channel and recorded in the catalog_sales fact table



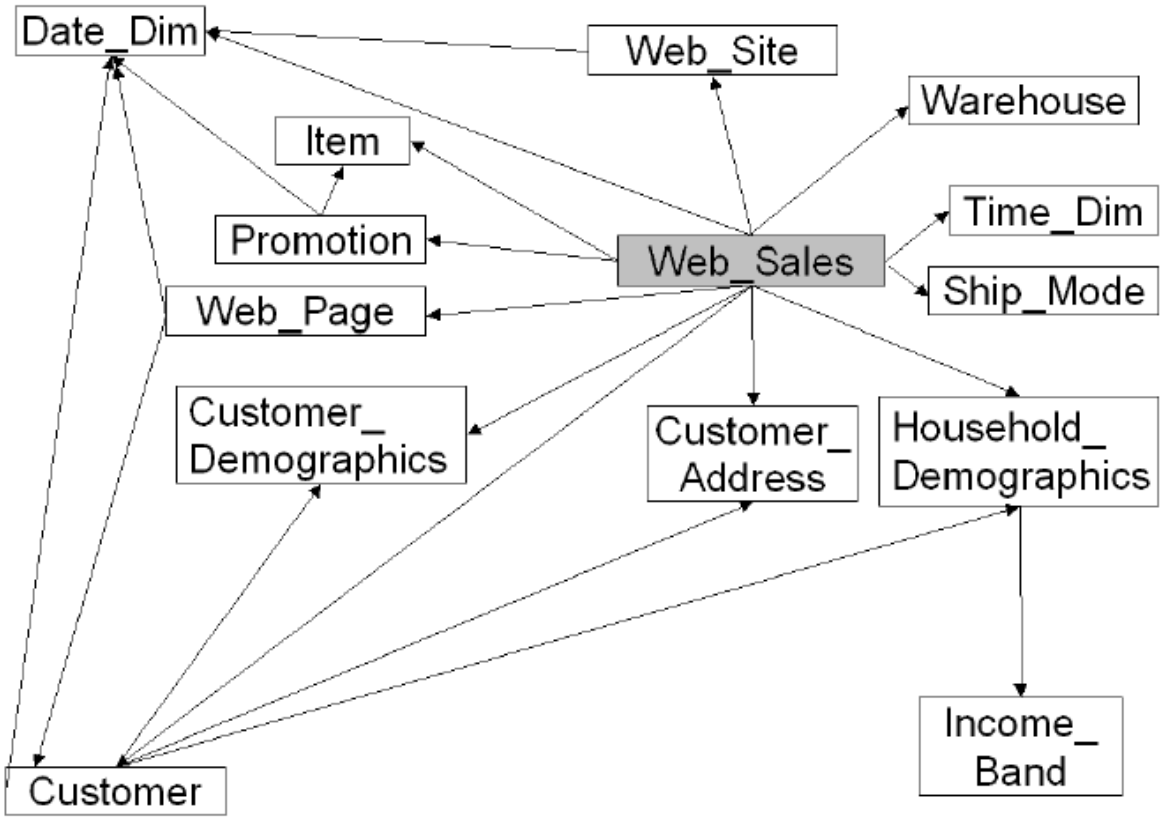
4. **Catalog Returns (CR)**

   Each row in this table represents a single line item for the return of an item sold through the catalog channel and recorded in the catalog_returns table



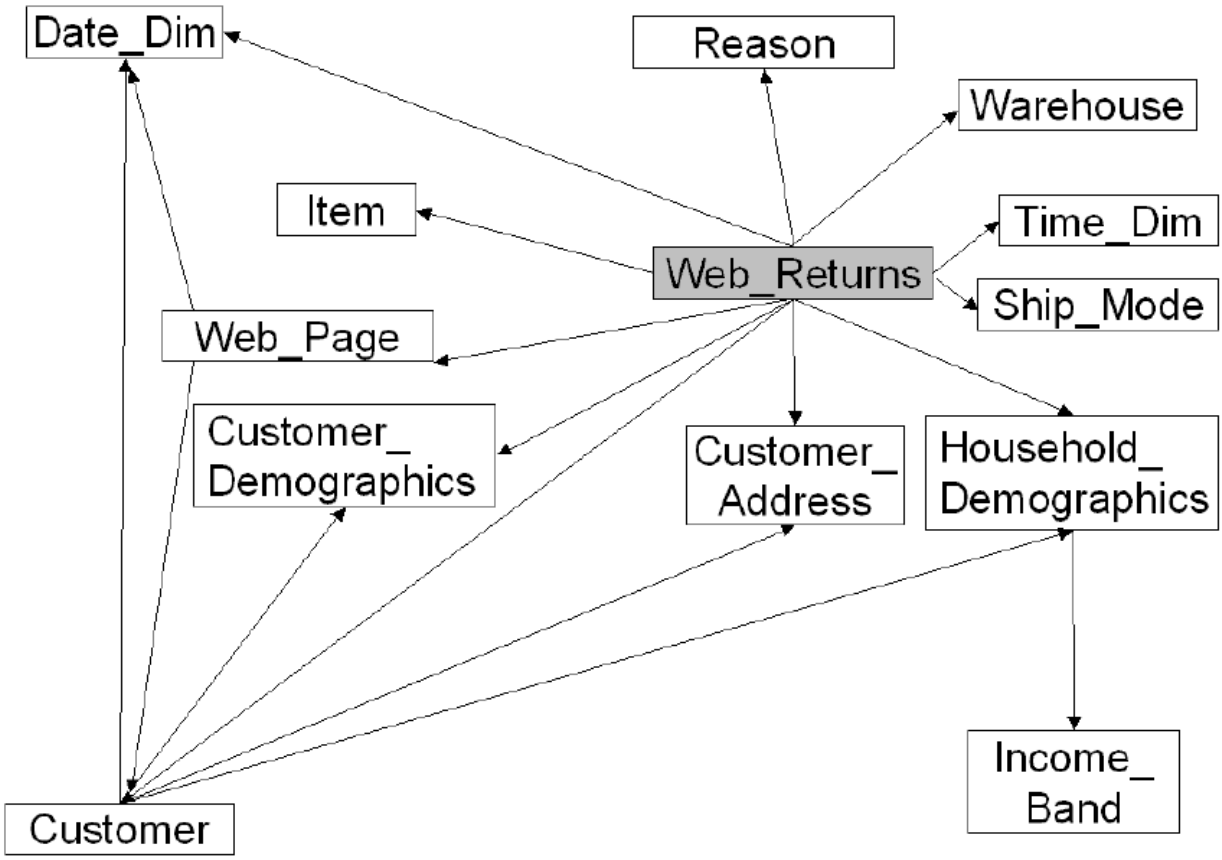
5. **Web Sales (WS)**

   Each row in this table represents a single line item for a sale made through the web channel and recorded in the web_sales fact table



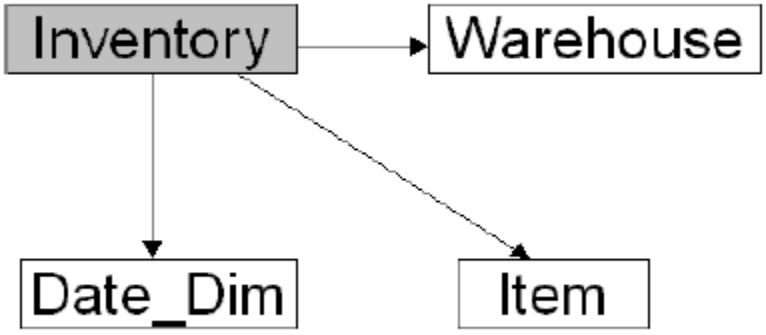
6. Web Returns (WR)

Each row in this table represents a single line item for the return of an item sold through the web sales channel and recorded in the web_returns table



7. **Inventory (INV)**

   Each row in this table represents the quantity of a particular item on-hand at a given warehouse during a specific week



Dimension Table Definitions

1. **Store (S)**
   Each row in this dimension table represents details of a store

2. **Call Center (CC)**
   Each row in this table represents details of a call center

3. **Catalog_page (CP)**

Each row in this table represents details of a catalog page

4. **Web_site (WEB)**
   Each row in this table represents details of a web site.

5. **Web_page (WP)**
   Each row in this table represents details of a web page within a web site

6. **Warehouse (W)**
   Each row in this dimension table represents a warehouse where items are stocked

7. **Customer (C)**
   Each row in this dimension table represents a customer

8. **Customer_address (CA)**
   Each row in this table represents a unique customer address (each customer can have more than one address)

9. **Customer_demographics (CD)**
   The customer demographics table contains one row for each unique combination of customer demographic information

10. **Date_dim (D)**
    Each row in this table represents one calendar day. The surrogate key (d_date_sk) for a given row is derived from the julian date being described by the row

11. **Household_demographics (HD)**
    Each row of this table defines a household demographic profile

12. **Item (I)**
    Each row in this table represents a unique product formulation (e.g., size, color, manufacturer, etc.).

13. **Income_band (IB)**
    Each row in this table represents details of an income range

14. **Promotion (P)**
    Each row in this table represents details of a specific product promotion (e.g., advertising, sales, PR)

15. **Reason (R)**
    Each row in this table represents a reason why an item was returned

16. **Ship_mode (SM)**
    Each row in this table represents a shipping mode

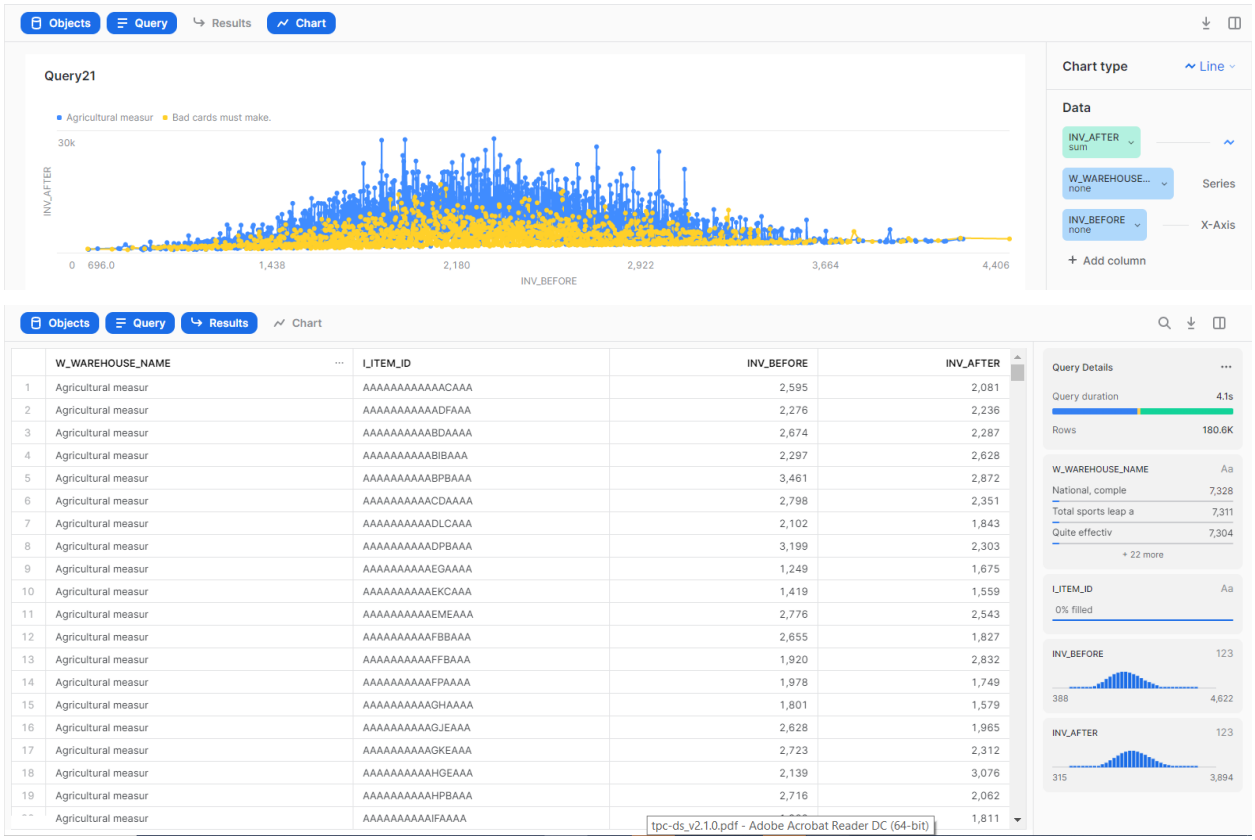
17. **Time_dim (T)**
    Each row in this table represents one second

Queries

1. *For all items whose price was changed on a given date, compute the percentage change in inventory between the 30-day period BEFORE the price change and the 30-day period AFTER the change. Group this information by warehouse.*

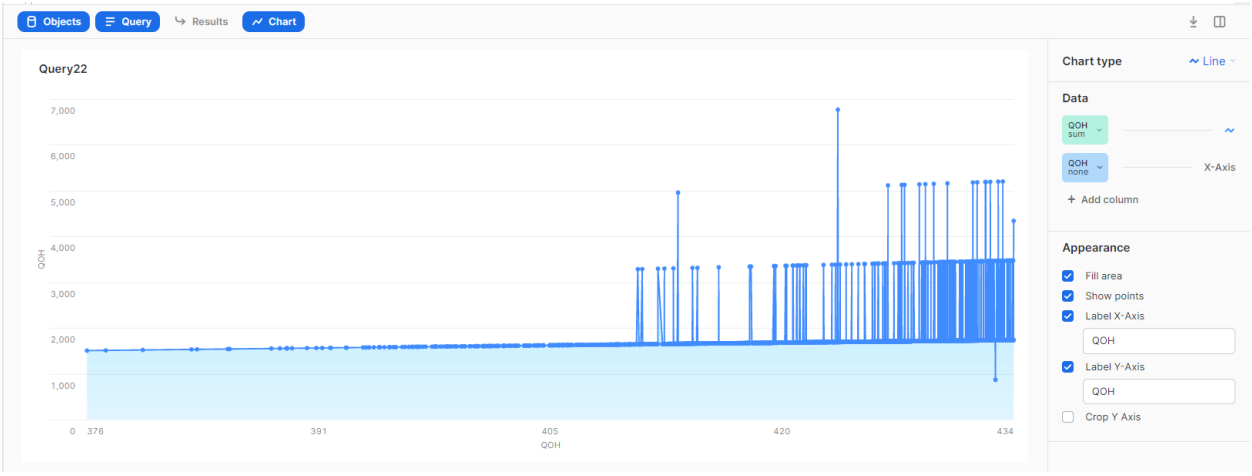
   Qualification Substitution Parameters:
   SALES_DATE.01 = 2000-03-11
   YEAR.01 = 2000

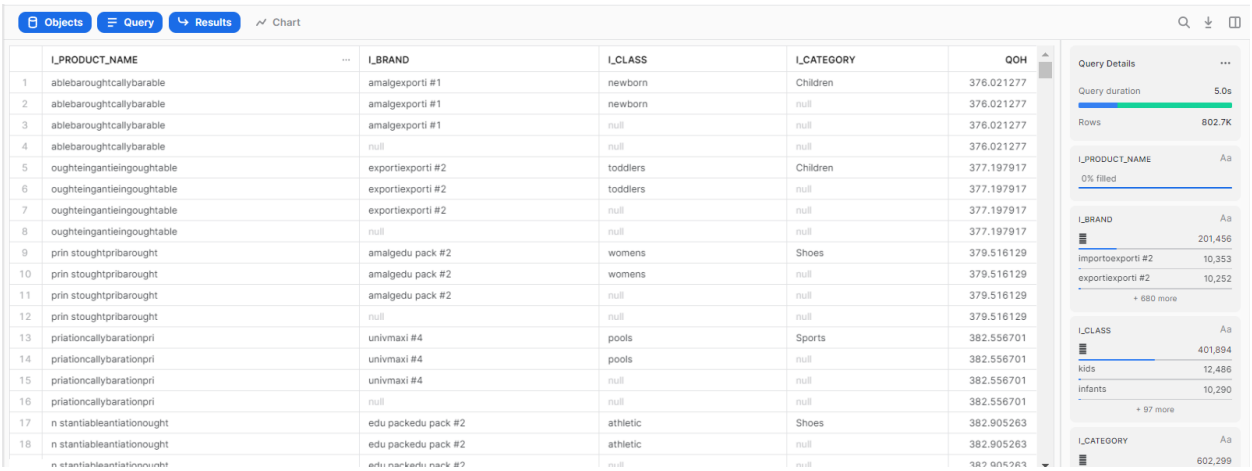
   In this query we aggregated all the inventory before and after the given date and with the condition of a 30 day interval we found the inventory prices change, grouping it with the warehouse.



2. *For each product name, brand, class, category, calculate the average quantity on hand. Rollup data by product name, brand, class and category.*
Qualification Substitution Parameters:
DMS.01 = 1200

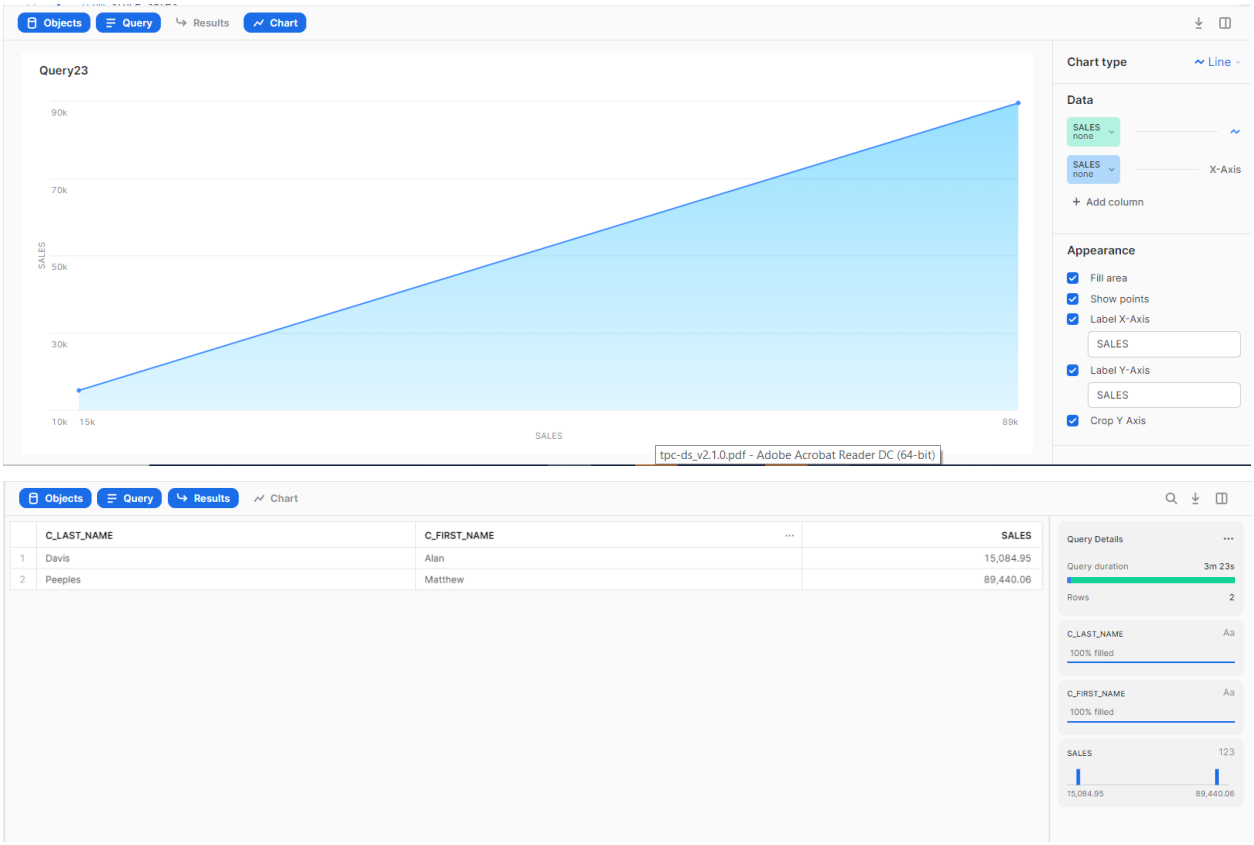
   This is a simple query to find the product name, brand, class, category, average quantity for given parameters. All the parameters are summed up using rollup.





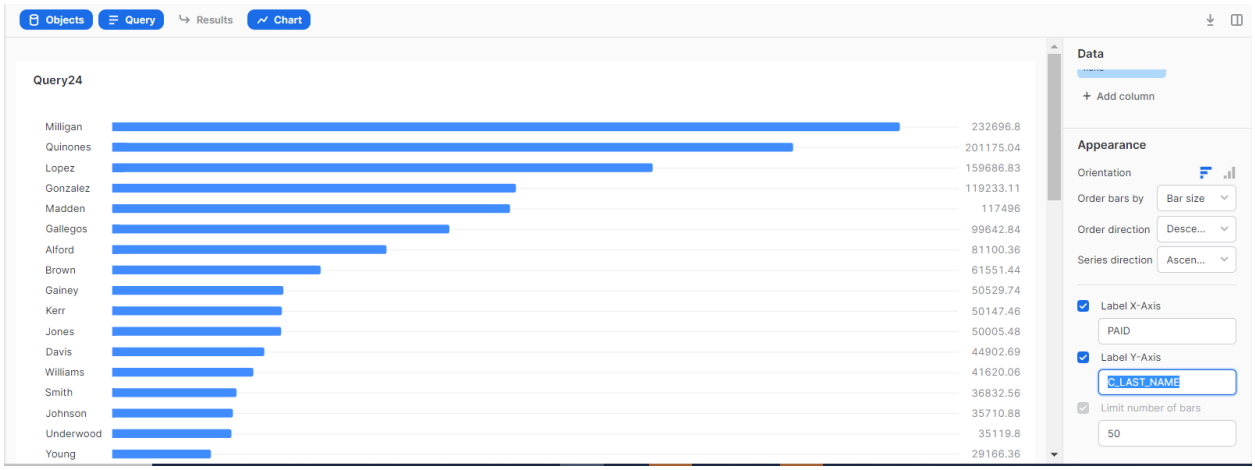
3. *This query contains multiple, related iterations:*
   Find frequently sold items that are items that were sold more than 4 times per day in four consecutive years. Compute the maximum store sales made by any given customer in a period of four consecutive years (same as above). Compute the best store customers as those that are in the 5th percentile of sales. Finally, compute the total sales of sales in March made by our best customers buying our most frequent items
   Qualification Substitution Parameters:
   MONTH.01 = 2
   YEAR.01 = 2000
   TOPPERCENT=50

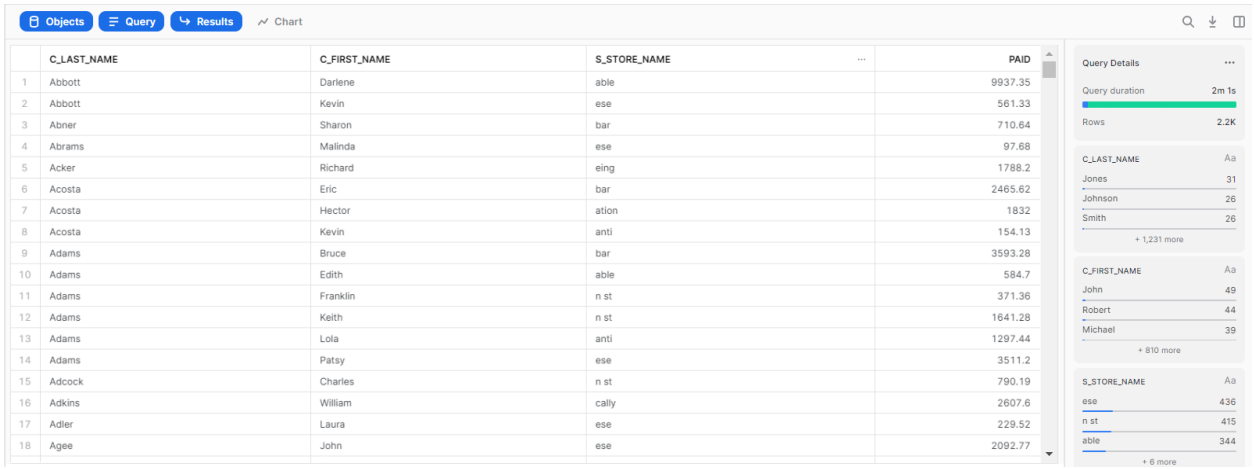
   In this query, sub queries are used to find frequently sold items and maximum store sales for consecutive 4 years, Unioning these found the store for 5th percentile of sales. Finally finding the best customer.



4. *This query contains multiple, related iterations:*
Iteration 1: Calculate the total specified monetary value of items in a specific color for store sales transactions by customer name and store, in a specific market, from customers who currently live in their birth countries and in the neighborhood of the store, and list only those customers for whom the total specified monetary value is greater than 5% of the average value
Iteration 2: Calculate the total specified monetary value of items in a specific color for store sales transactions by customer name and store, in a specific market, from customers who currently live in their birth countries and in the neighborhood of the store, and list only those customers for whom the total specified monetary value is greater than 5% of the average value
Qualification Substitution Parameters:
MARKET = 8
COLOR.1 = pale
COLOR.2 = chiffon
AMOUNTONE = ss_net_paid

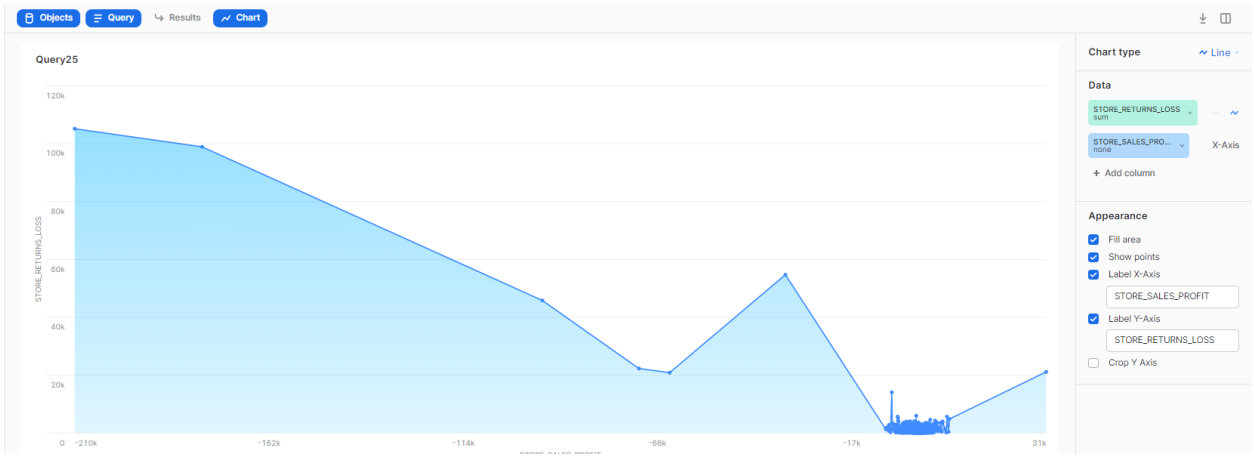
In this query we first fetched the customer and item details and made a temporary table and from that table showed the required data for peach and saddle color.

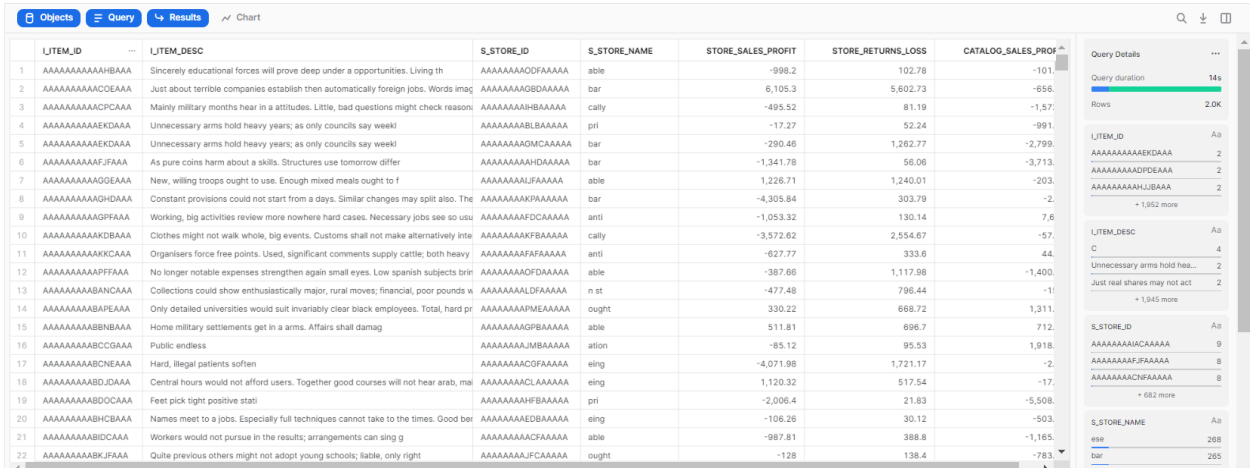




5. *Get all items that were sold in stores in a particular month and year and returned and re-purchased by the customer through the catalog channel in the same month and in the six following months.*
   For these items, compute the sum of net profit of store sales, net loss of store loss and net profit of catalog . Group this information by item and store.
   Qualification Substitution Parameters:
   YEAR.01 = 2001
   AGG.01 = sum

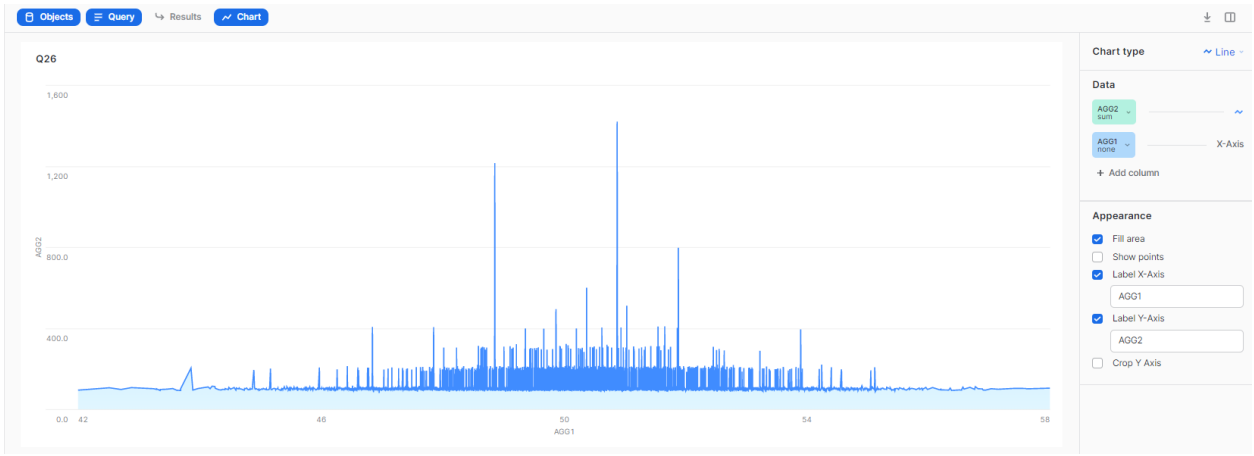
   This is a simple query where all the items that were returned and profit & loss of the company is to be shown. For that, we joined store sales, store returns, catalog sales, store and item tables.

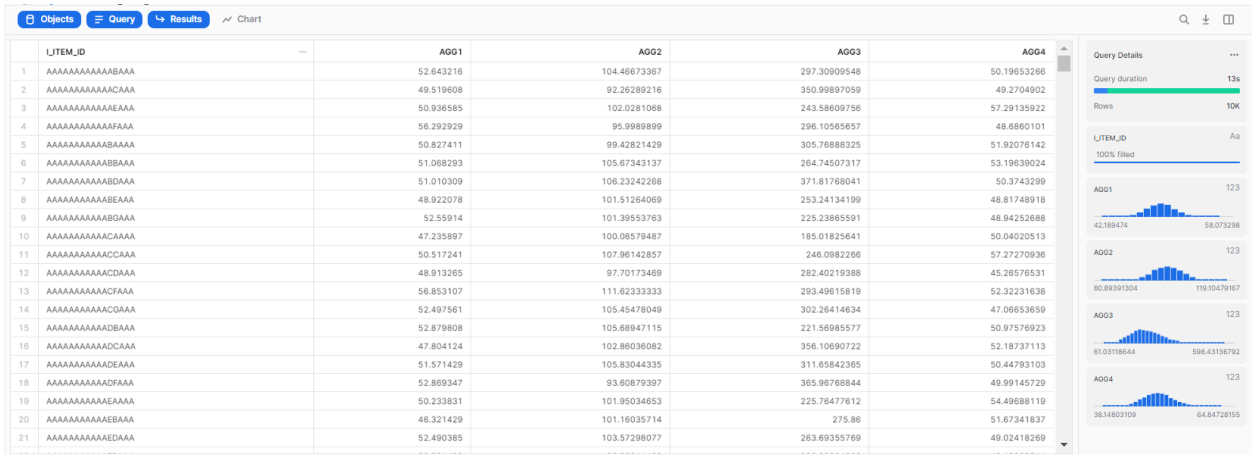




6. *Computes the average quantity, list price, discount, sales price for promotional items sold through the catalog channel where the promotion was not offered by mail or in an event for given gender, marital status and educational status.*
   Qualification Substitution Parameters:
   YEAR.01 = 2000
   ES.01 = College
   MS.01 = S
   GEN.01 = M

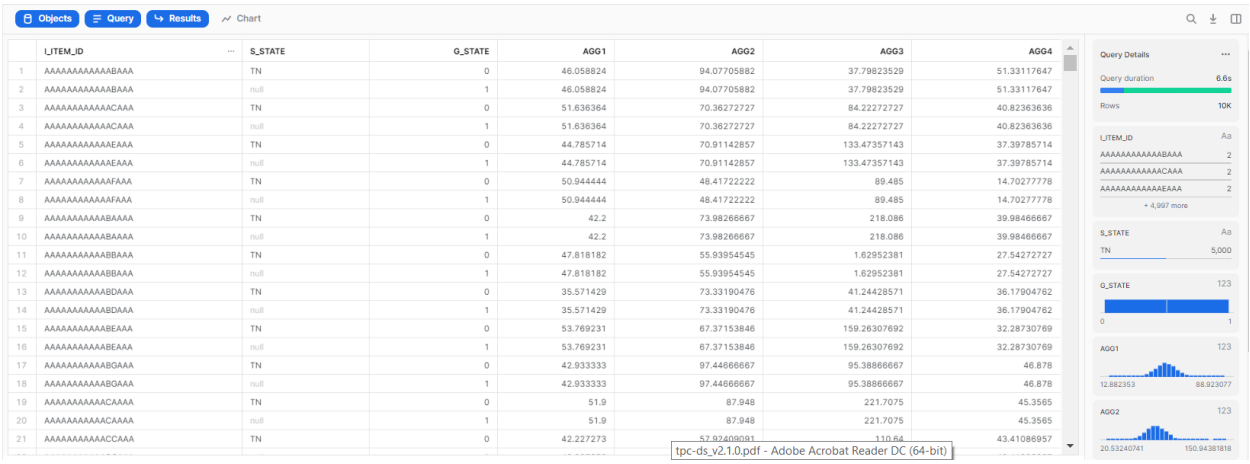
   This is a simple query. Joined catalog sales, customer demographics, item and promotion tables.





7. *For all items sold in stores located in six states during a given year, find the average quantity, average list price, average list sales price, average coupon amount for a given gender, marital status, education and customer demographic.*
   Qualification Substitution Parameters:
   - *STATE_F.01 = TN*
   - *STATE_E.01 = TN*
   - *STATE_D.01 = TN*
   - *STATE_C.01 = TN*
   - *STATE_B.01 = TN*
   - *STATE_A.01 = TN*
   - *ES.01 = College*
   - *MS.01 = S*
   - *GEN.01 = M*
   - *YEAR.01 = 2002*

   This is a simple query. Joined store sales, customer demographics, item and store tables.



8. *Calculate the average list price, number of non empty (null) list prices and number of distinct list prices of six different sales buckets of the store sales channel. Each bucket is defined by a range of distinct items and information about list price, coupon amount and wholesale cost.*
   Qualification Substitution Parameters:
   - *WHOLESALECOST.01=57*
   - *WHOLESALECOST.02=31*
   - *WHOLESALECOST.03=79*
   - *WHOLESALECOST.04=38*
   - *WHOLESALECOST.05=17*
   - *WHOLESALECOST.06=7*
   - *COUPONAMT.01=459*
   - *COUPONAMT.02=2323*
   - *COUPONAMT.03=12214*
   - *COUPONAMT.04=6071*
   - *COUPONAMT.05=836*
   - *COUPONAMT.06=7326*
   - *LISTPRICE.01=8*
   - *LISTPRICE.02=90*
   - *LISTPRICE.03=142*
   - *LISTPRICE.04=135*
   - *LISTPRICE.05=122*
   - *LISTPRICE.06=154*

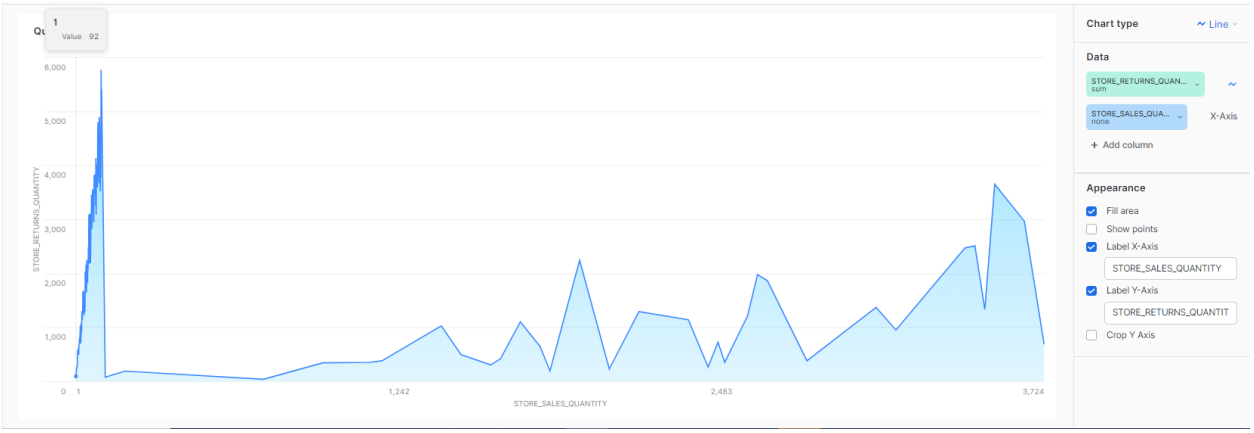
Using sub queries calculated required attributes for given parameters.

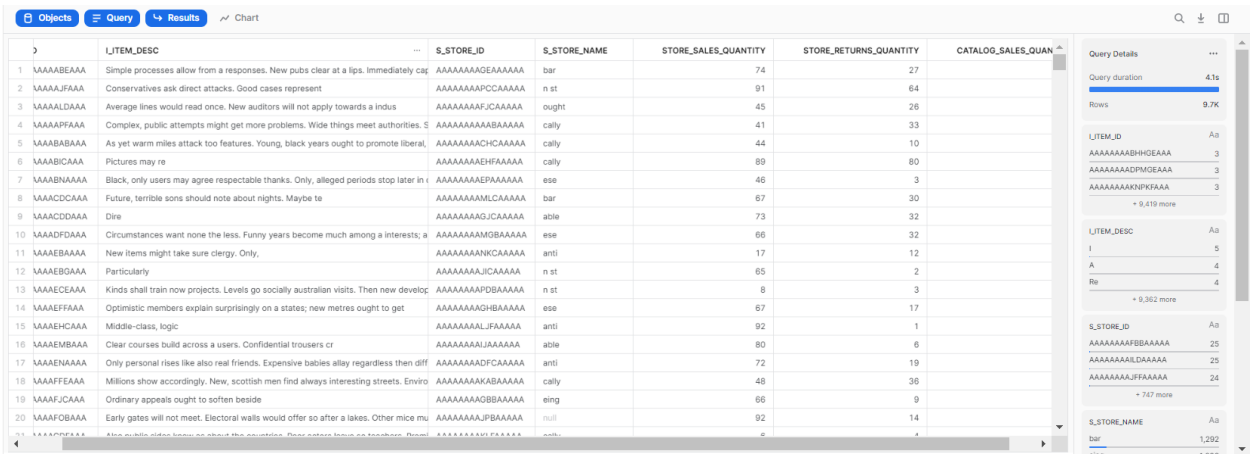
	B1_LP	B1_CNT	B1_CNTD	B2_LP	B2_CNT	B2_CNTD	B3_LP	B3_CNT	B3_CNTD	B4_LP	B4_CNT	B4_CNTD	B5_LP	B5_CNT	B5_CNTD
1	77.9800153	363,306,326	14,946	69.54547306	352,421,116	7,076	134.03773531	279,882,978	11,830	82.34750337	316,397,150	8,695	61.53044452	362,086,461	18,130

9. *Get all items that were sold in stores in a specific month and year and which were returned in the next six months of the same year and re-purchased by the returning customer afterwards through the catalog sales channel in the following three years.*

For those these items, compute the total quantity sold through the store, the quantity returned and the quantity purchased through the catalog. Group this information by item and store.
Qualification Substitution Parameters:
MONTH.01 = 9
YEAR.01 = 1999
AGG.01 = 29

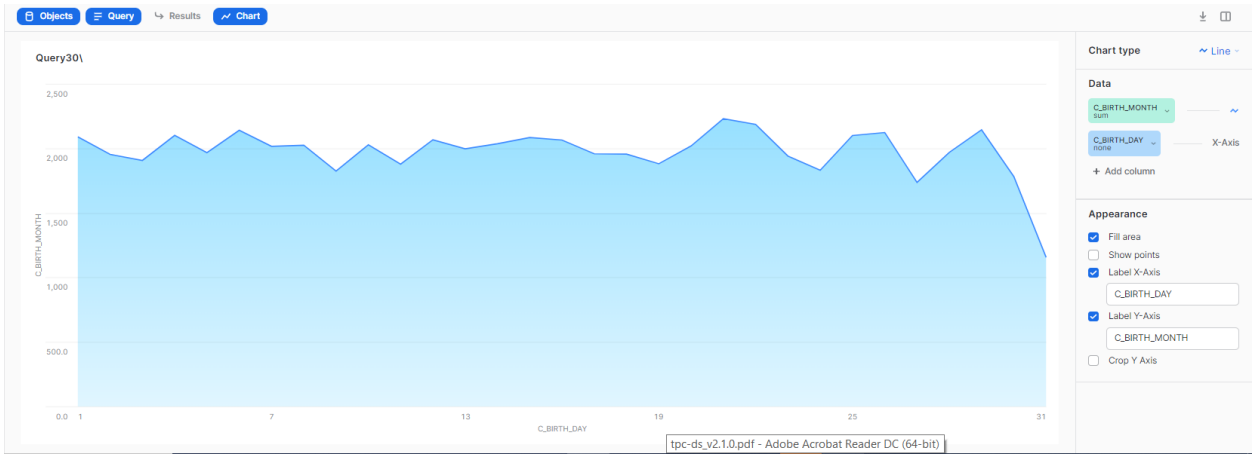
This is a simple query where we found item details, store details and sum of different quantities using store sales, store returns, catalog sales, store and item tables.





10. *Find customers and their detailed customer data who have returned items, which they bought on the web, for an amount that is 20% higher than the average amount a customer returns in a given state in a given time period across all items. Order the output by customer data.*
Qualification Substitution Parameters:
YEAR.01 = 2002
STATE.01 = GA

Made the temporary table of customers who have returned the items and then used it to get the details of customers who have returned more than 20% of the average customer returned items.



C_LOGIN	C_EMAIL_ADDRESS	C_LAST_REVIEW_DATE	CTR_TOTAL_RETURN
null	Lester.Green@kxL.com	2452336	3,963.78
null	Lester.Green@kxL.com	2452336	5,360.61
null	Ignacio.Bishop@Lrse.edu	2452627	6,901.74
null	Edmond.Cook@6.com	2452567	2,405.52
null	Charles.Carpenter@xulfXd.edu	2452611	2,916.27
null	Charles.Carpenter@xulfXd.edu	2452611	4,231.2
null	Charles.Carpenter@xulfXd.edu	2452611	6,439.23
null	Cinda.Holland@h6QL2si8Mv9.com	2452639	3,995.47
null	Cinda.Holland@h6QL2si8Mv9.com	2452639	4,068.55
null	Elliott.Johnson@LZuVpKZ3nXUn0.org	2452510	1,898.04
null	Jason.Price@2165Za6Cx4fU.com	2452446	2,310.28
null	Diana.Williams@h8sfMjfJ.org	2452460	1,702.6
null	David.Arias@1.com	2452388	2,008.98
null	Edith.Chacon@ppZ9YSUhhpzCv9.org	2452436	1,829.03
null	Edith.Chacon@ppZ9YSUhhpzCv9.org	2452436	9,358.01
null	Michael.Knox@iAdjc.org	2452583	1,737.72
null	Russell.Galvan@4puc9gBZFH.com	2452302	7,424.76
null	Larry.Walters@BxuPL5N7udxx1AY3ZK.com	2452338	2,692.8
null	Ana.Forbes@Bi4EZ.org	2452594	2,619.05
null	Michael.Buckner@advFTn.com	null	1,949.48

C_BIRTH_DAY 123
1 31

C_BIRTH_MONTH 123
1 12

C_BIRTH_YEAR 123
1,924 1,992

C_BIRTH_COUNTRY Aa
SWITZERLAND 69
SWAZILAND 67
ALAND ISLANDS 65
+ 208 more

C_LOGIN Aa
0% filled 100% null

C_EMAIL_ADDRESS Aa
Richard.Mcgill@Qq.com 5
Susan.Wood@OKcdGFGh1p... 4

Objects | Query | Results | Chart

	C_CUSTOMER_ID	C_SALUTATION	C_FIRST_NAME	C_LAST_NAME	C_PREFERRED_CUST_FLAG	C_BIRTH_DAY	C_BIRTH_MONTH	C_BIRTH_YEAR	C_BIRTH_COUNTRY
1	AAAAAAAAAAACDBA	Dr.	Lester	Green	Y	5	4	1,924	GUYANA
2	AAAAAAAAAAACDBA	Dr.	Lester	Green	Y	5	4	1,924	GUYANA
3	AAAAAAAAAAAEFCA	Mr.	Ignacio	Bishop	Y	18	5	1,949	HUNGARY
4	AAAAAAAAAAAGCDA	Sir	Edmond	Cook	N	30	6	1,982	POLAND
5	AAAAAAAAAAAHHAA	Mr.	Charles	Carpenter	N	24	1	1,960	FRANCE
6	AAAAAAAAAAAHHAA	Mr.	Charles	Carpenter	N	24	1	1,960	FRANCE
7	AAAAAAAAAAAHHAA	Mr.	Charles	Carpenter	N	24	1	1,960	FRANCE
8	AAAAAAAAAAAHJDA	Mrs.	Cinda	Holland	N	29	7	1,990	CANADA
9	AAAAAAAAAAAHJDA	Mrs.	Cinda	Holland	N	29	7	1,990	CANADA
10	AAAAAAAAAAAIGDA	Sir	Elliott	Johnson	N	13	12	1,930	GUAM
11	AAAAAAAAAAAIJBA	Dr.	Jason	Price	N	7	12	1,967	CÔTE D'IVOIRE
12	AAAAAAAAAAAJGCA	Dr.	Diana	Williams	null	null	null	1,962	PHILIPPINES
13	AAAAAAAAAAALFCA	Sir	David	Arias	Y	21	11	1,971	MONTENEGRO
14	AAAAAAAAAAAMGCA	Miss	Edith	Chacon	N	7	1	1,951	MOROCCO
15	AAAAAAAAAAAMGCA	Miss	Edith	Chacon	N	7	1	1,951	MOROCCO
16	AAAAAAAAAAAMPBA	Dr.	Michael	Knox	N	28	6	1,968	SWITZERLAND
17	AAAAAAAAAAANGCA	Mr.	Russell	Galvan	N	19	1	1,933	SOMALIA
18	AAAAAAAAAAAOADA	Sir	Larry	Walters	Y	18	11	1,932	BERMUDA
19	AAAAAAAAAAAOCDA	Mrs.	Ana	Forbes	N	4	7	1,959	NETHERLANDS ANTILLES
20	AAAAAAAAAAAOECA	null	null	Buckner	null	null	10	null	GREENLAND
21	AAAAAAAAAAAOFCA	null	null	Buckner	null	null	10	null	GREENLAND