**Course:** Algorithmic Digital Marketing

**Professor:** Srikanth Krishnamurthy

**Students:** Pramod Gopal, Jeet Khemani

**Assignment:** Analytics and dashboard on snowflake using TPC- DS dataset

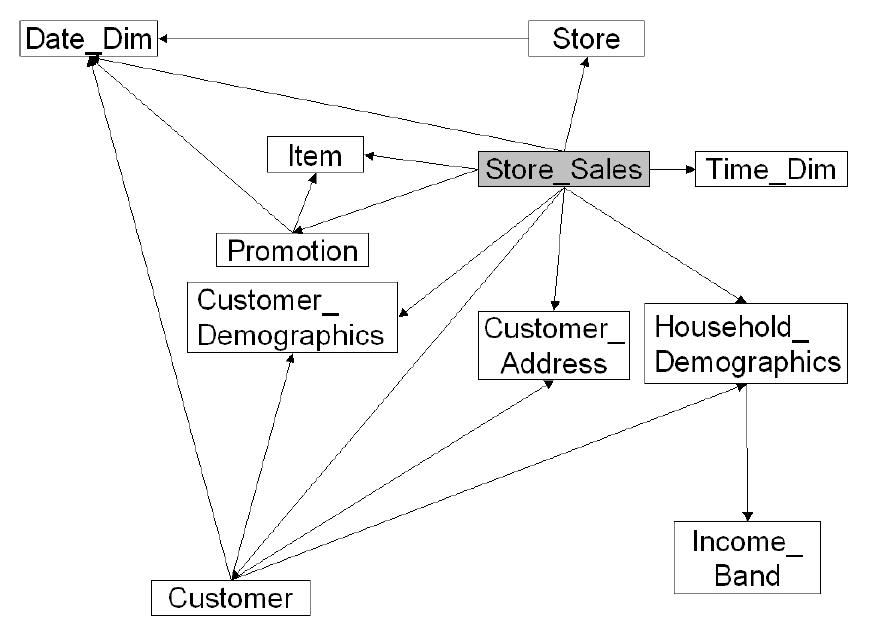
**Introduction**

* 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**

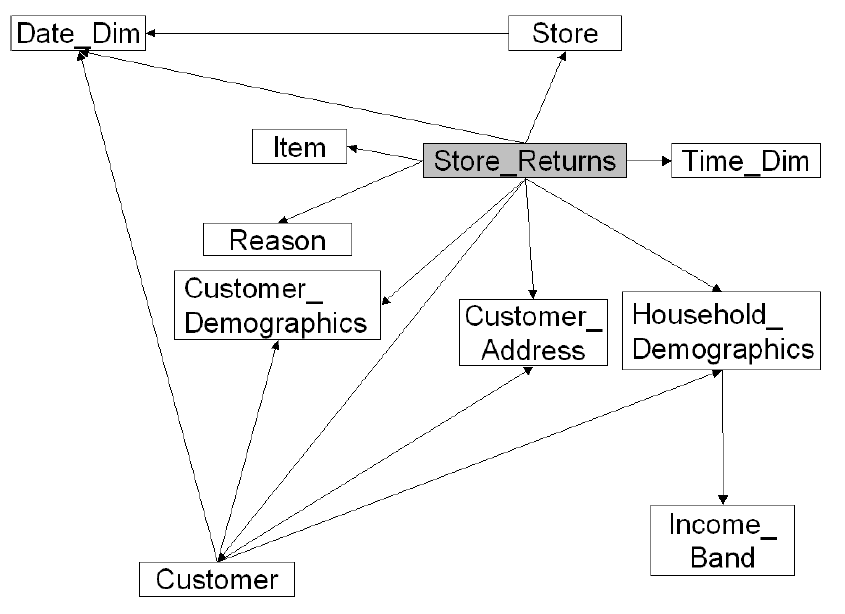
1. **Store Sales**

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.



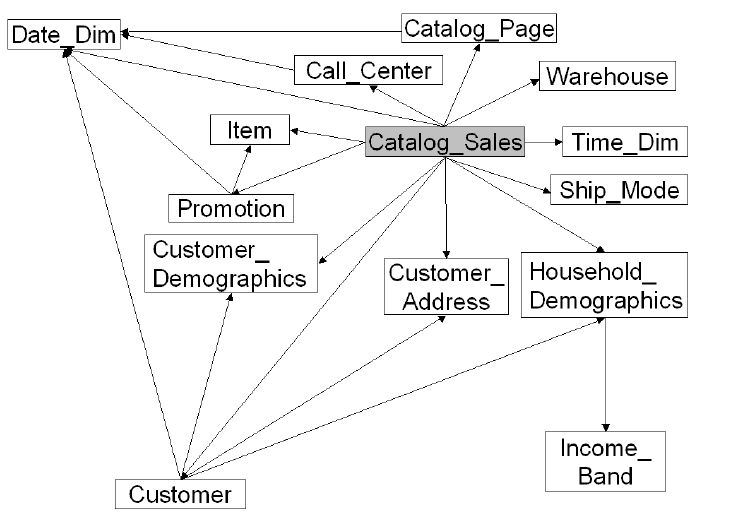
1. **Store Returns (SR)**

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



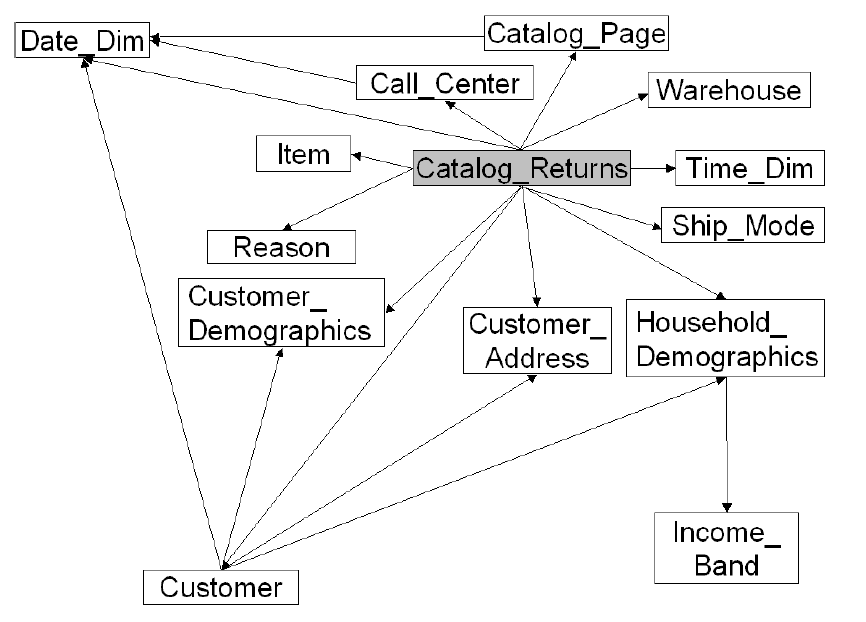
1. **Catalog Sales (CS)**

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



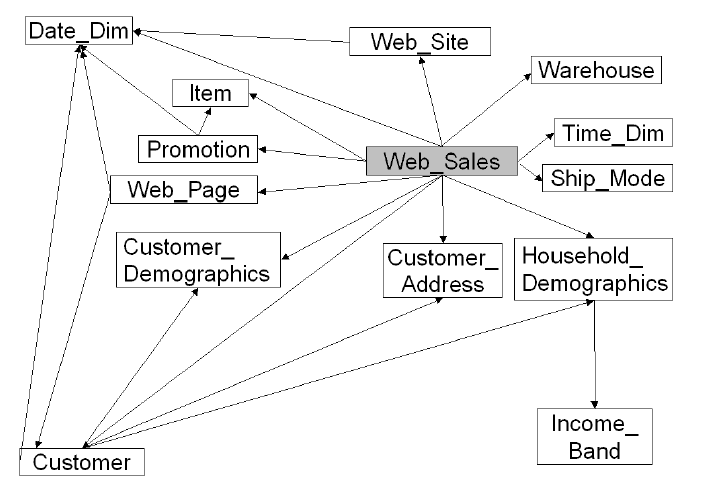
1. **Catalog Returns (CR)**

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



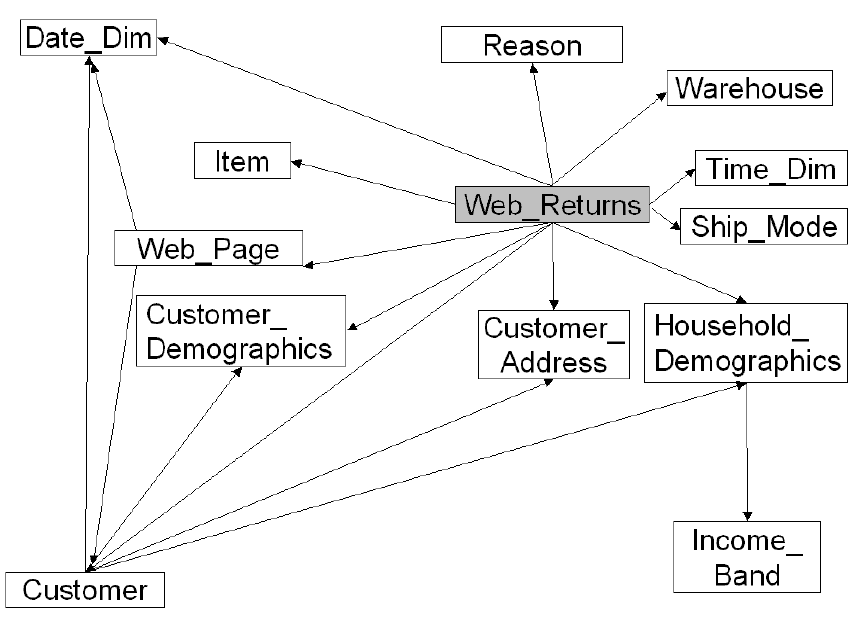
1. **Web Sales (WS)**

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



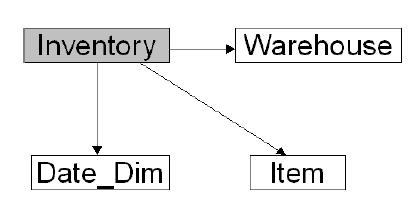
1. **Web Returns (WR)**

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



1. **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

1. **Call Center (CC)**

Each row in this table represents details of a call center

1. **Catalog\_page (CP)**

Each row in this table represents details of a catalog page

1. **Web\_site (WEB)**

Each row in this table represents details of a web site.

1. **Web\_page (WP)**

Each row in this table represents details of a web page within a web site

1. **Warehouse (W)**

Each row in this dimension table represents a warehouse where items are stocked

1. **Customer (C )**

Each row in this dimension table represents a customer

1. **Customer\_address (CA)**

Each row in this table represents a unique customer address (each customer can have more than one address)

1. **Customer\_demographics (CD)**

The customer demographics table contains one row for each unique combination of customer demographic information

1. **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

1. **Household\_demographics (HD)**

Each row of this table defines a household demographic profile

1. **Item (I)**

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

1. **Income\_band (IB)**

Each row in this table represents details of an income range

1. **Promotion (P)**

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

1. **Reason (R )**

Each row in this table represents a reason why an item was returned

1. **Ship\_mode (SM)**

Each row in this table represents a shipping mode

1. **Time\_dim (T)**

Each row in this table represents one second

**Queries**

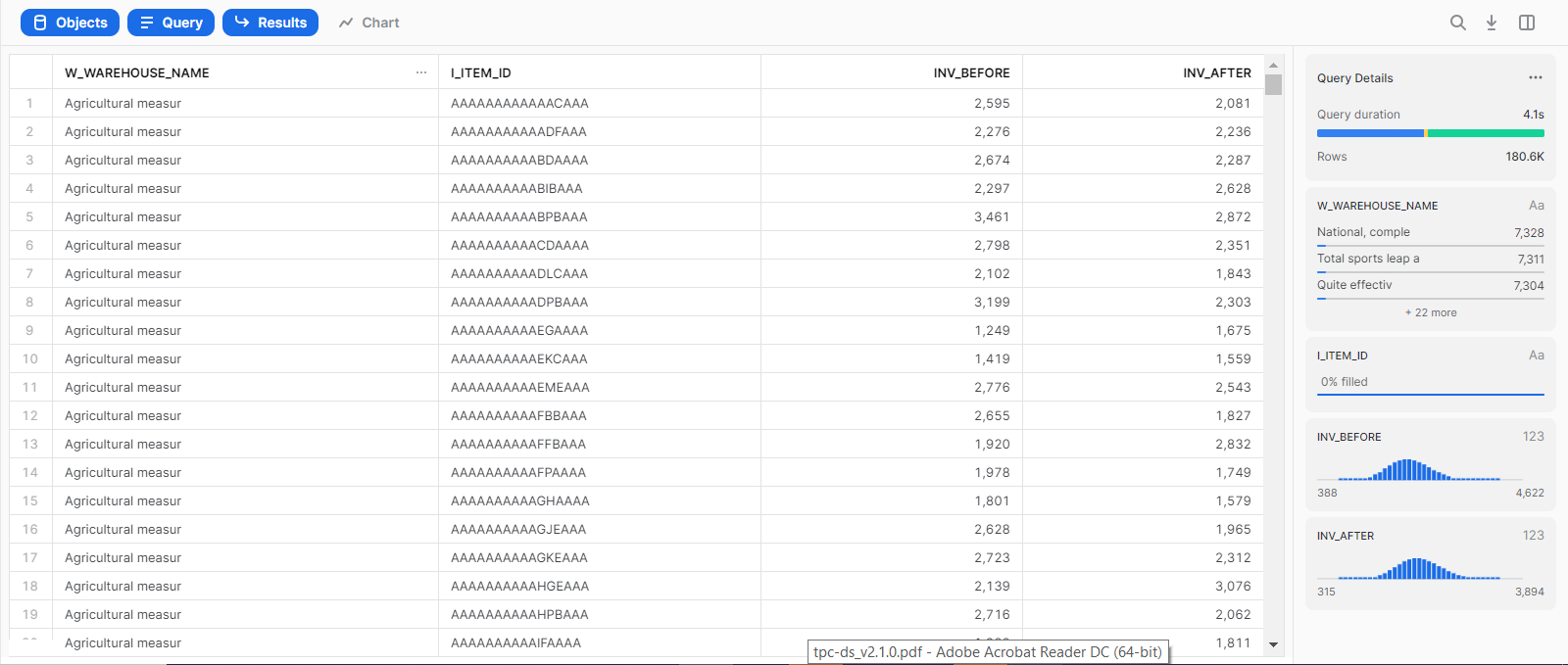
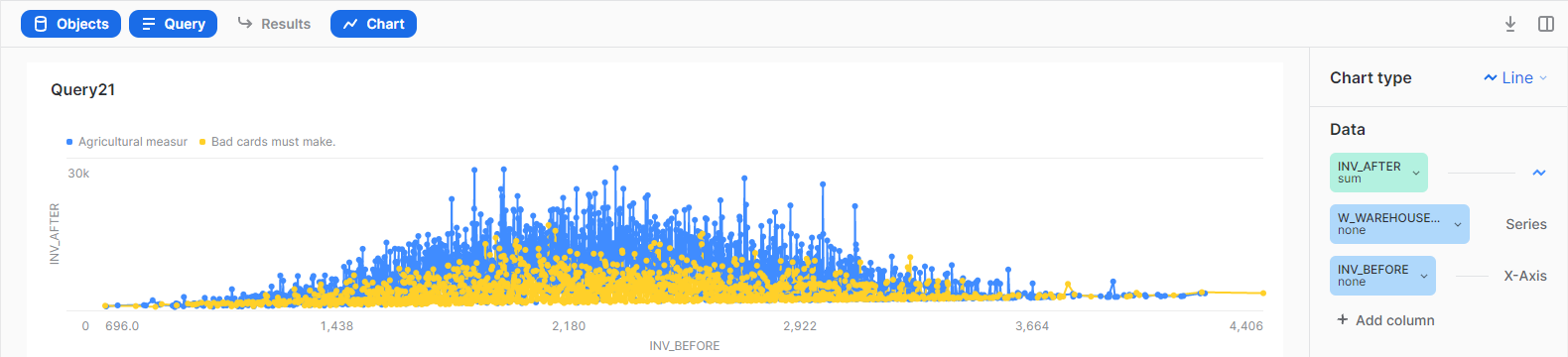
1. *For all items whose price was changed on a given date, compute the percentage change in inventory between the 30-day period BEFORE the price change and the 30-day period AFTER the change. Group this information by warehouse.*

*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.

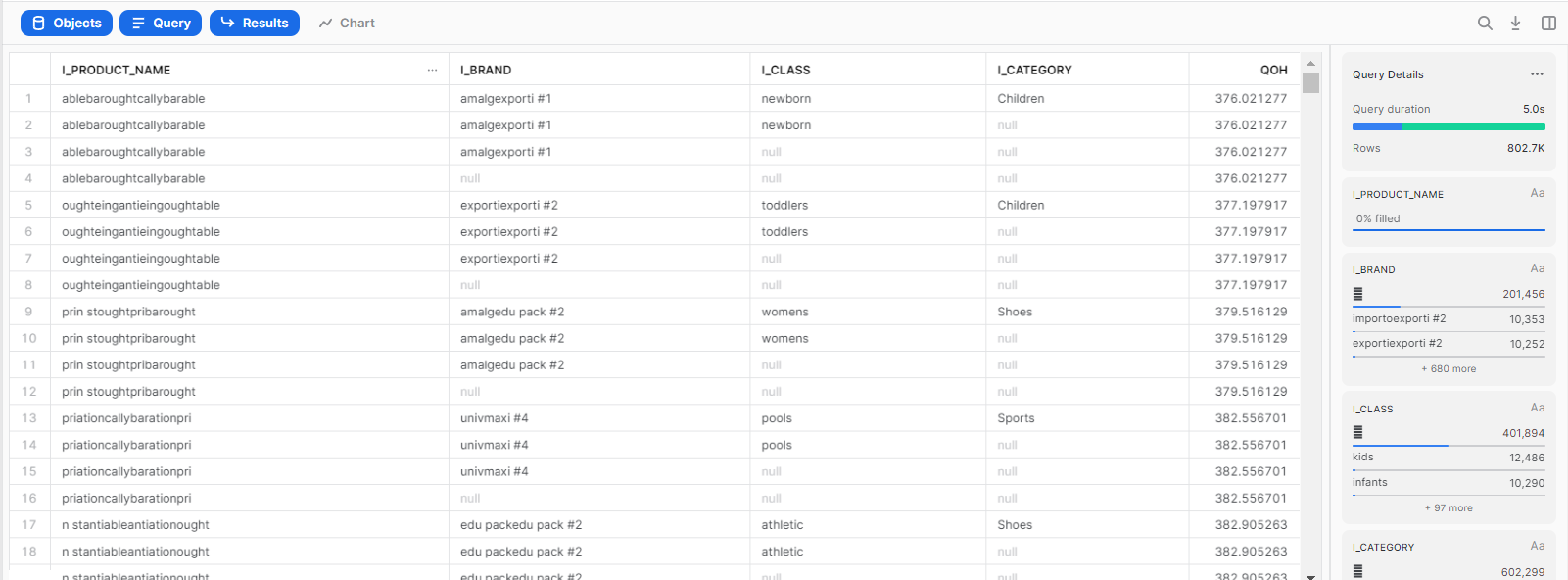
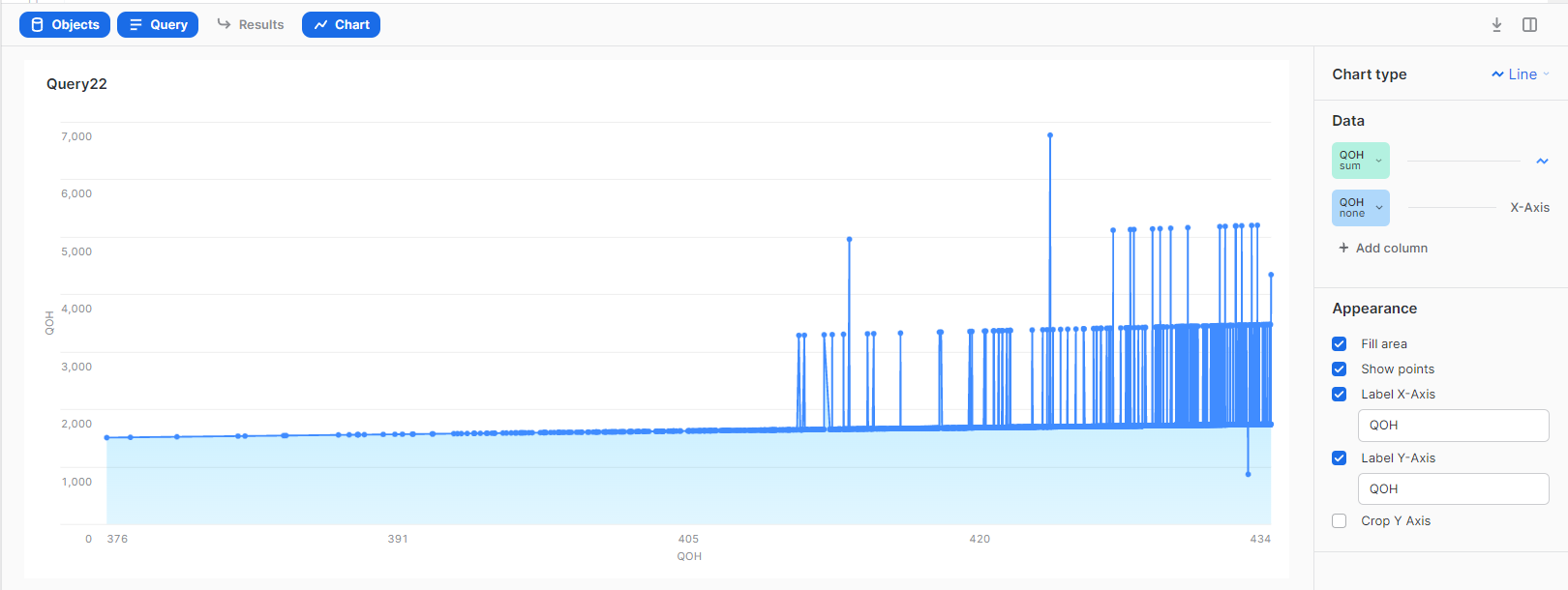


1. *For each product name, brand, class, category, calculate the average quantity on hand. Rollup data by product name, brand, class and category.*

*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.



1. *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*

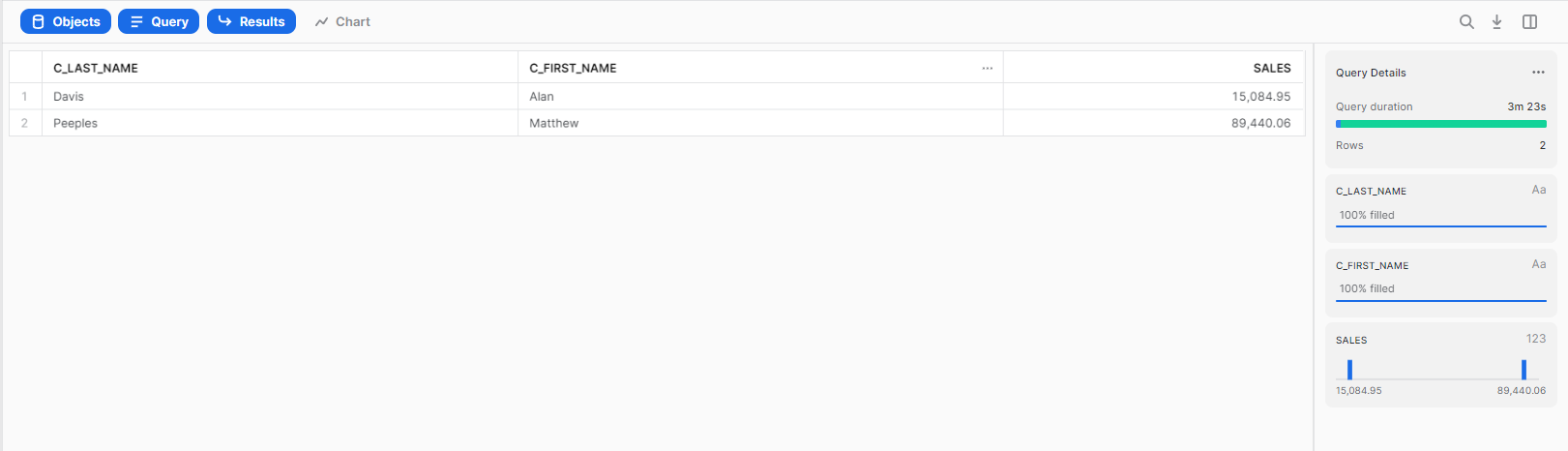
*Qualification Substitution Parameters:*

*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.



1. *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:*

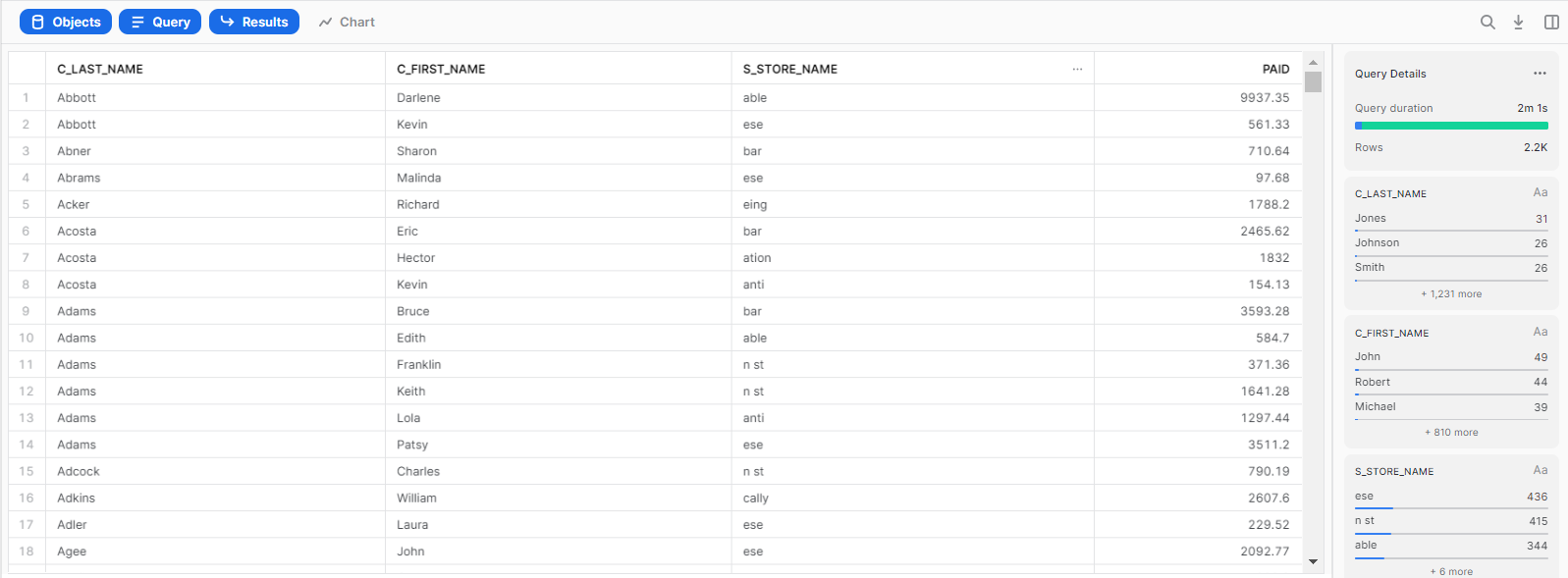
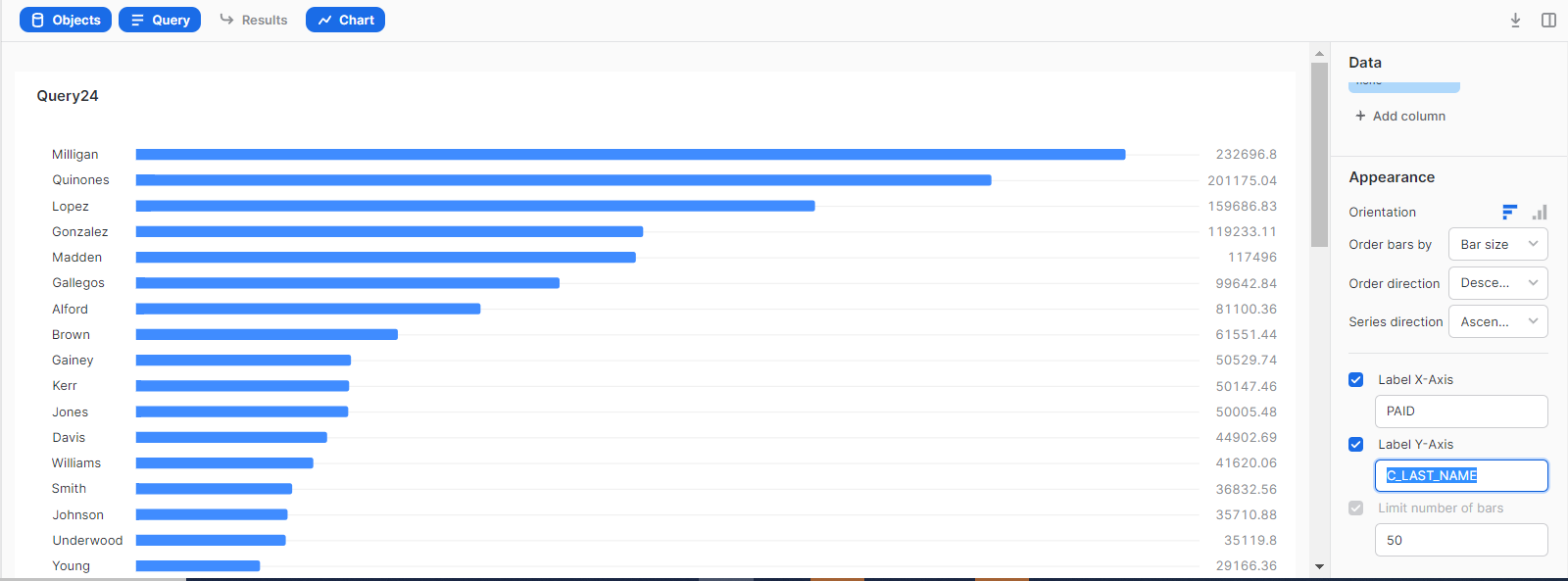
*MARKET = 8*

*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.



1. *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.*

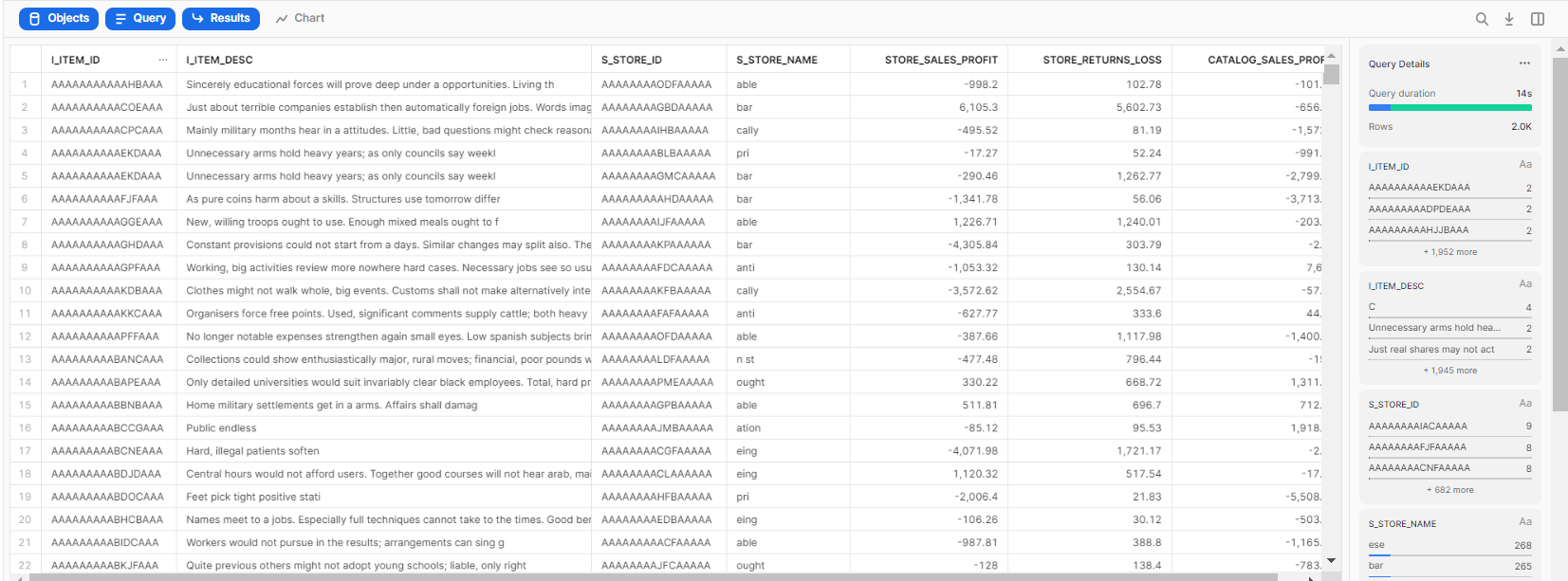
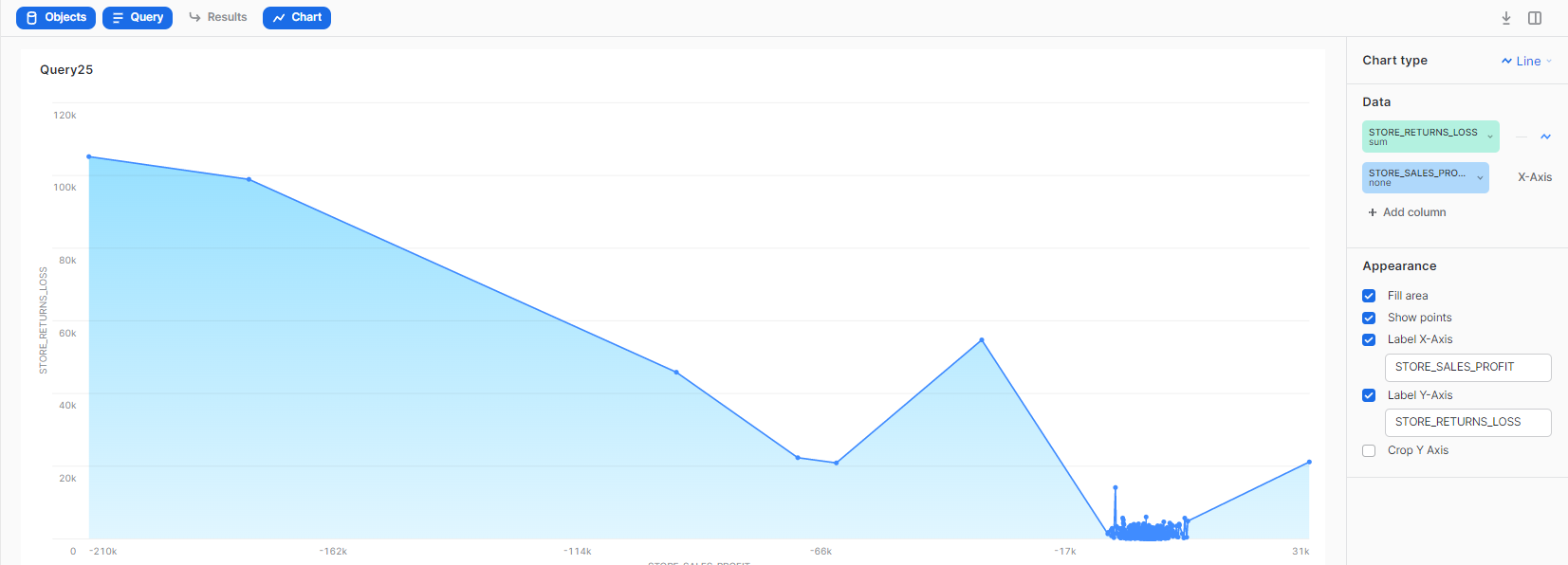
*For these items, compute the sum of net profit of store sales, net loss of store loss and net profit of catalog . Group this information by item and store.*

*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.



1. *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:*

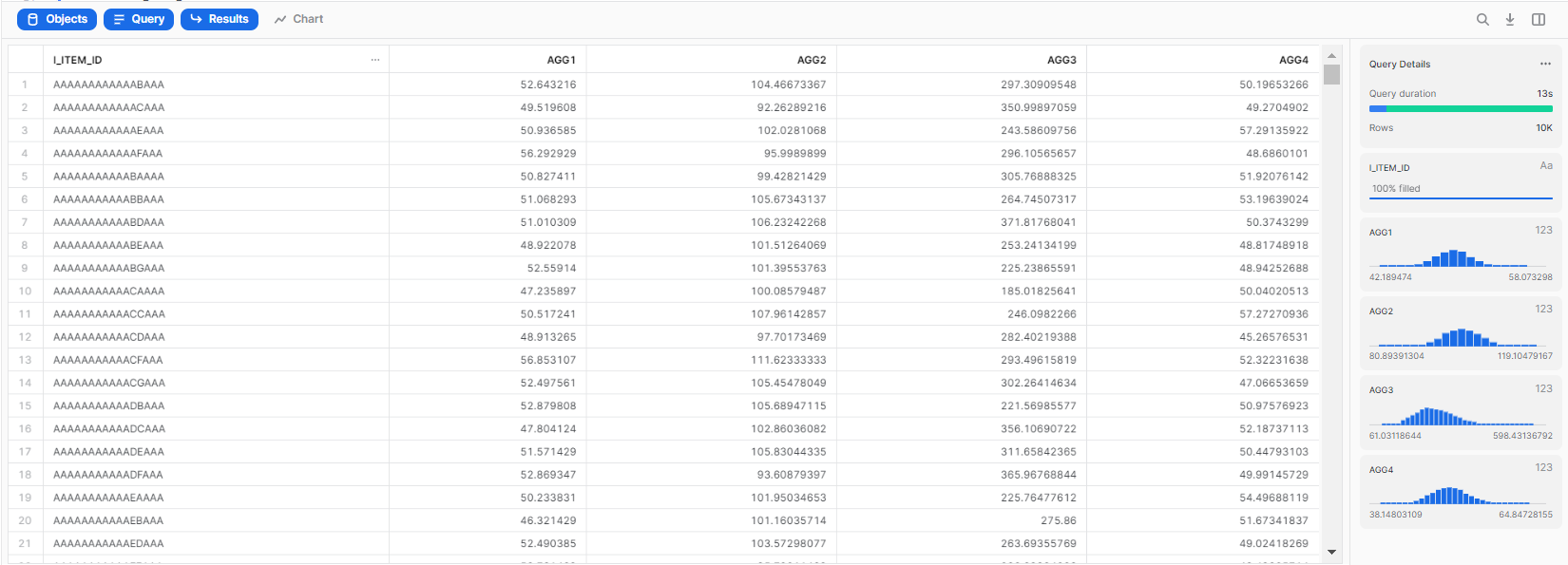
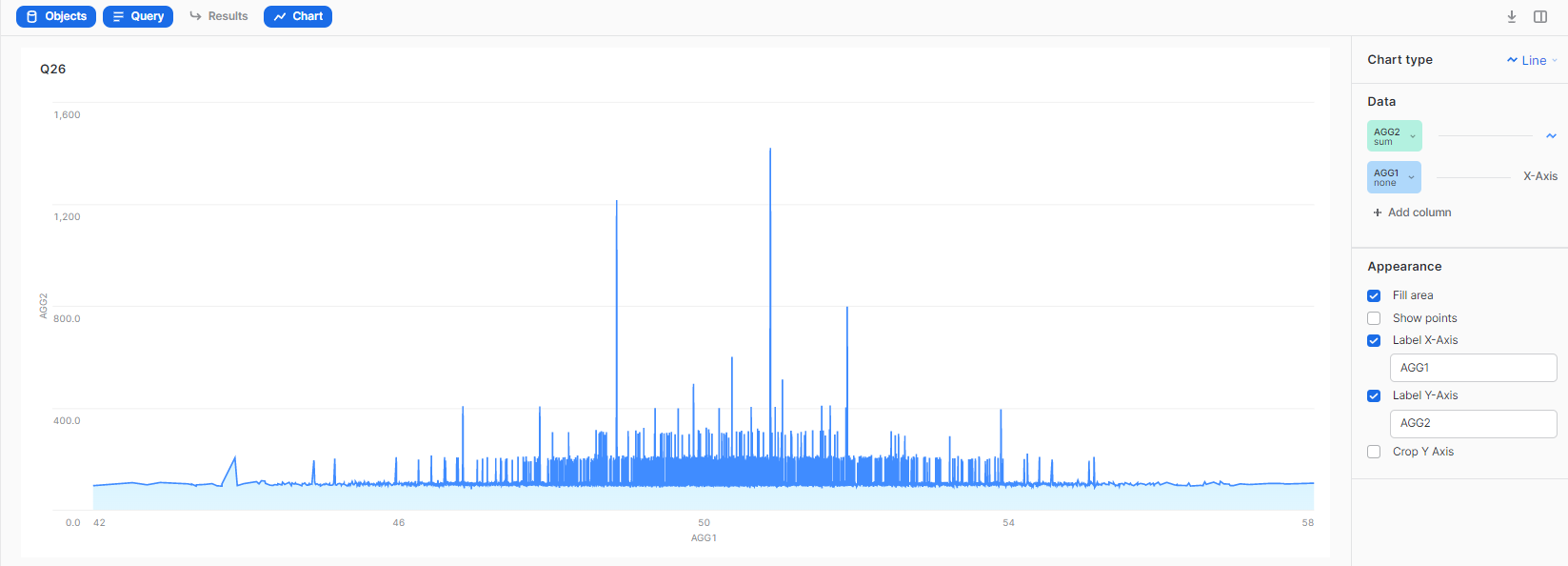
*YEAR.01 = 2000*

*ES.01 = College*

*MS.01 = S*

*GEN.01 = M*

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

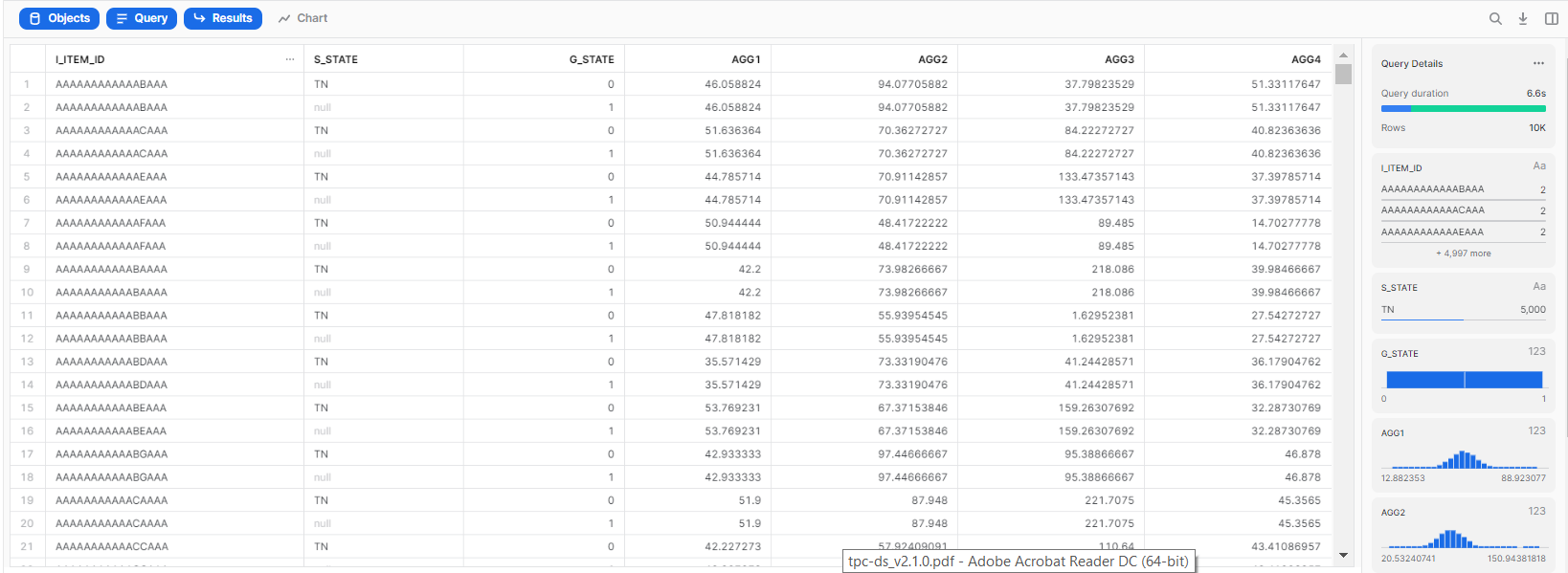


1. *For all items sold in stores located in six states during a given year, find the average quantity, average list price, average list sales price, average coupon amount for a given gender, marital status, education and customer demographic.*

*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.



1. *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.



1. *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.*

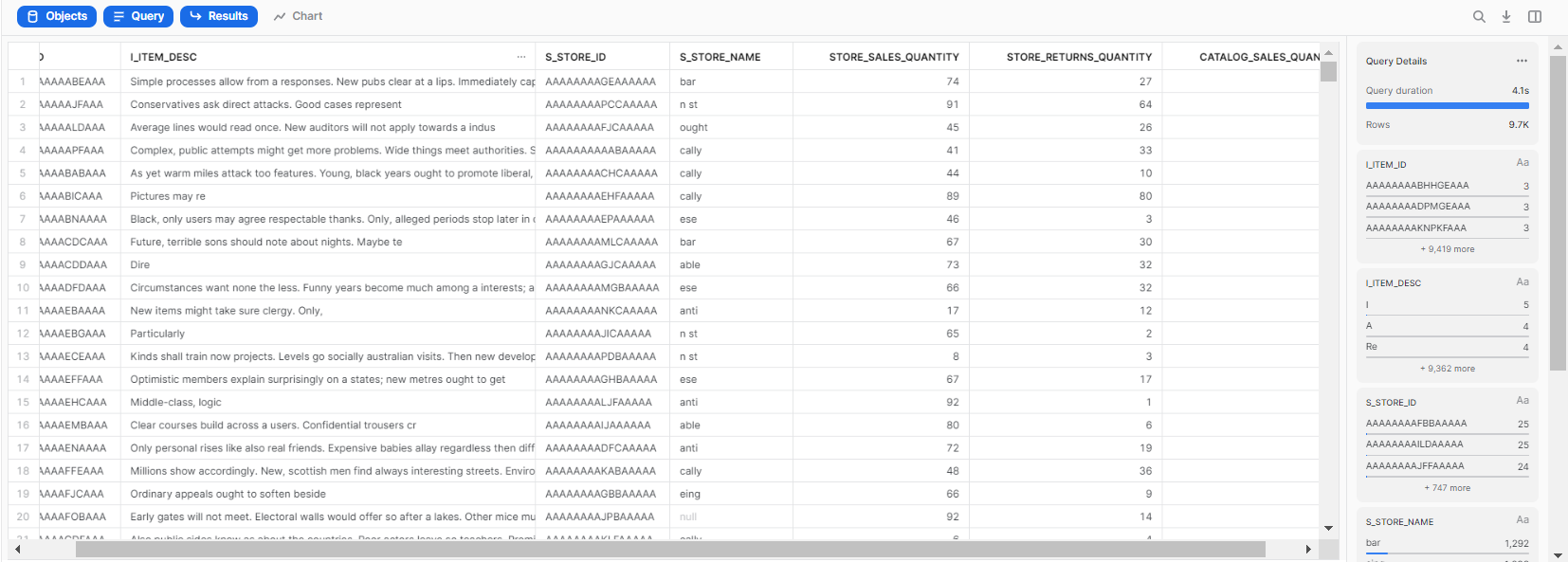
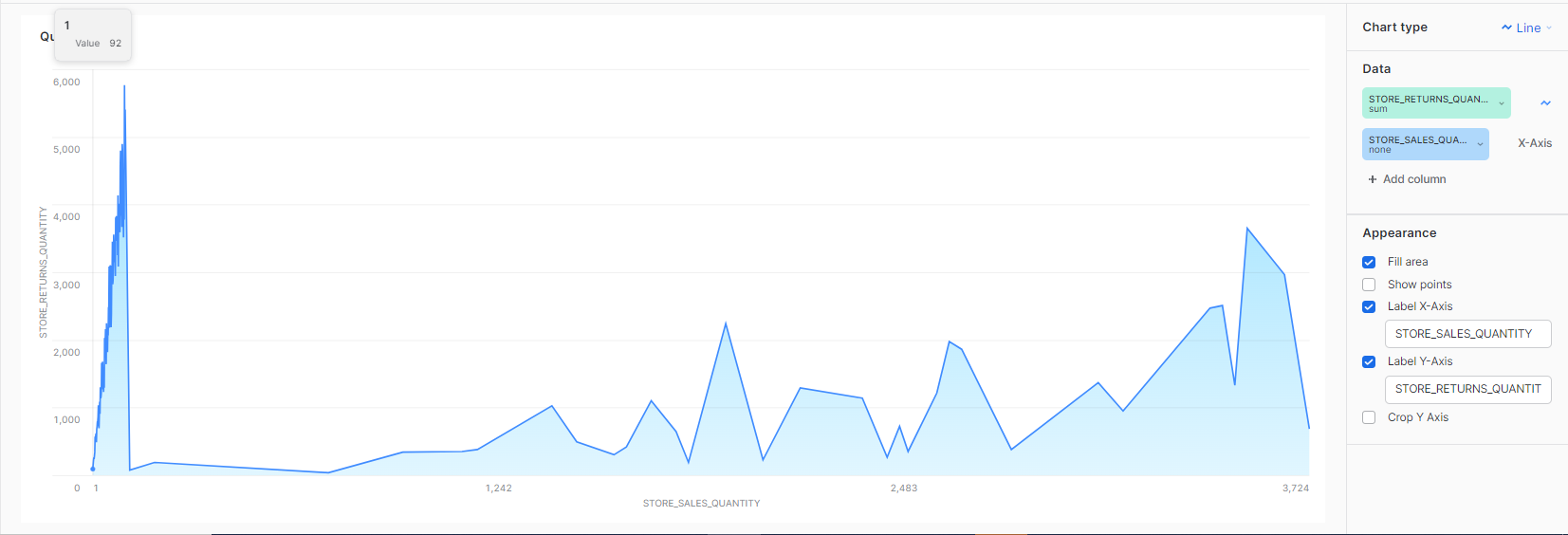
*Qualification Substitution Parameters:*

*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.



1. *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.

