PHARMACEUTICAL COMPANY DATABASE MANAGEMENT AND ANALYSIS

GROUP 04
ARJUN JANARDHAN
ADITI NAMDEO

Problem Setting :

Pharmacy is one of the main components of a thriving human civilization and is extremely important to the standard of living and defines the health and sanitation of the country or city. Hence it is extremely important that medicines are in the right hands and is distributed extremely efficiently across a vast range of networks. A database for a Pharmacy is an extremely efficient and an important tool in maintaining is distribution network. In this particular problem, we will see how to create a database and a multidimensional schema for a company called Glenn Pharma responsible for supplying medicines across Massachussetts.

Problem Definition :

This project intends to build the database design of a business model similar to a Pharmaceutical database. Keeping extensibility and scalability in mind, we will build a module that can be converted to a microservice architecture or transferred to a data warehouse to perform data analysis for the prediction of future trends in technologies. The transformed data is loaded in a data warehouse where analysis is done. A few of the analysis topics are mentioned below:

1. What drug generates the maximum revenue
2. Who are the top performing salesmen?
3. Who are the biggest customers?
4. What is the monthly sales analysis.

Data :

The data is collected from the database of a Pharmaceutical company called Glenn Pharma and it is found on dataworld.org. It contains the details about the meetings held between salesmen and customers, the salesmen, the sutomers and the products which the company is currently dealing with.

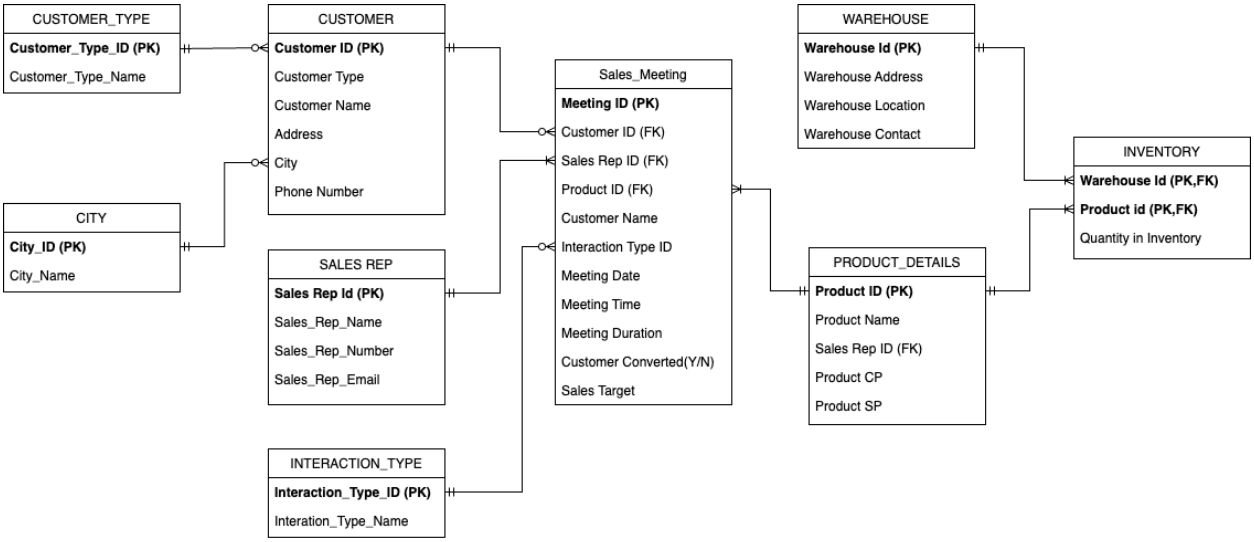
Data Description :

The meeting table contains the record of 2585 meetings held between customers and the sales representatives and also gives and information of the amount of sales intended for that meeting and whether the sale was converted or not. The product table contains a list of 30 products with each sales rep responsible for one product respectively. The customer table contains a list of all the customers and their contact information which will be useful to sales reps. We also have the inventory table which gives a list of all the products and their quantities which are stored in the data warehouse.

END GOAL :

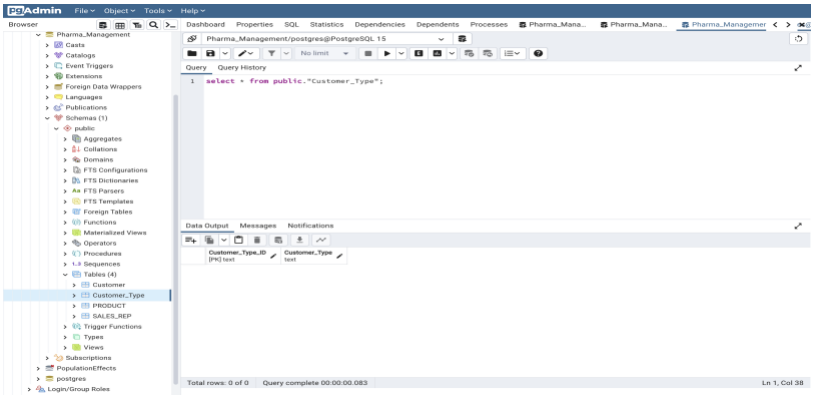
Our end goal in this project is to create a multidimensional model of the pharmaceutical database which will be useful for further analysis.

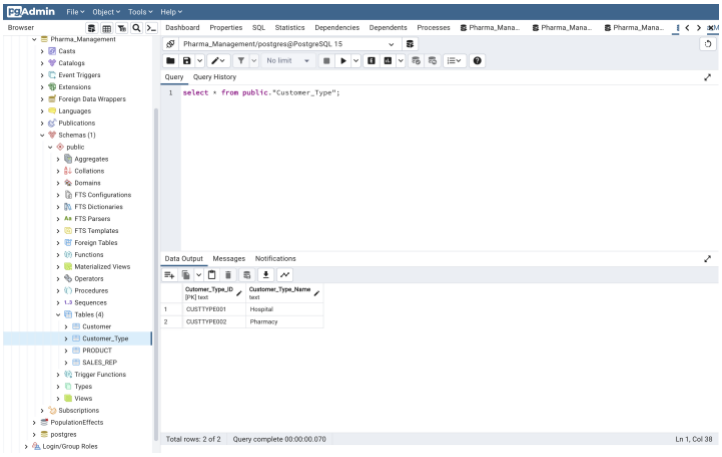
ENTITY RELATIONSHIP DIAGRAM :



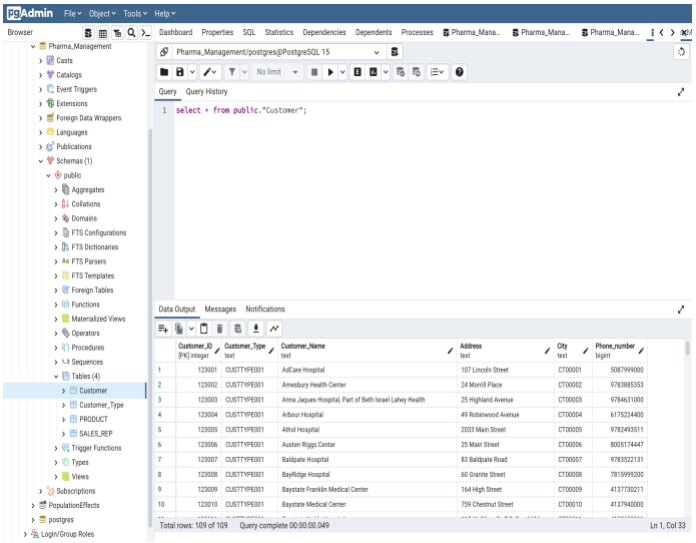
ERD SCHEMA AND DATA INSERTION INTO SCHEMA

CUSTOMER TYPE :

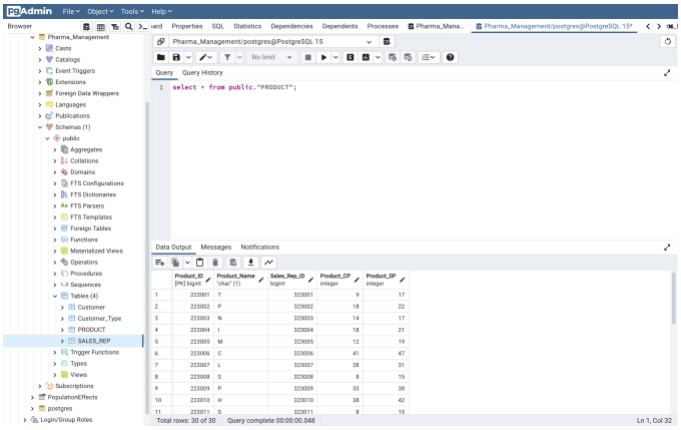




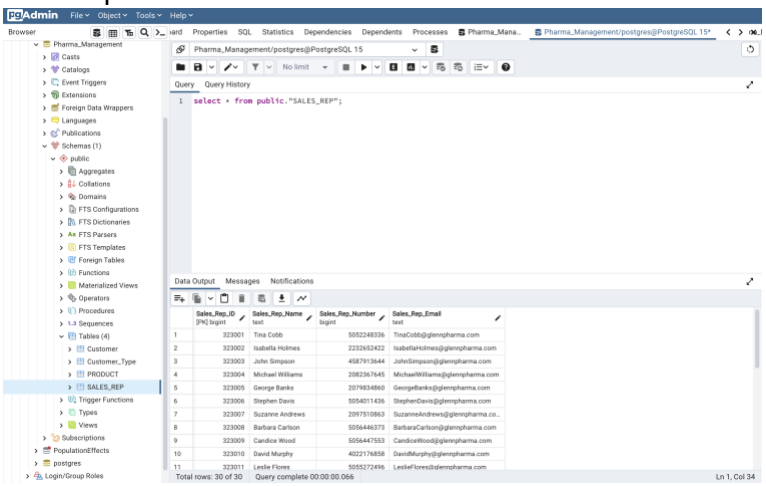
CUSTOMER :



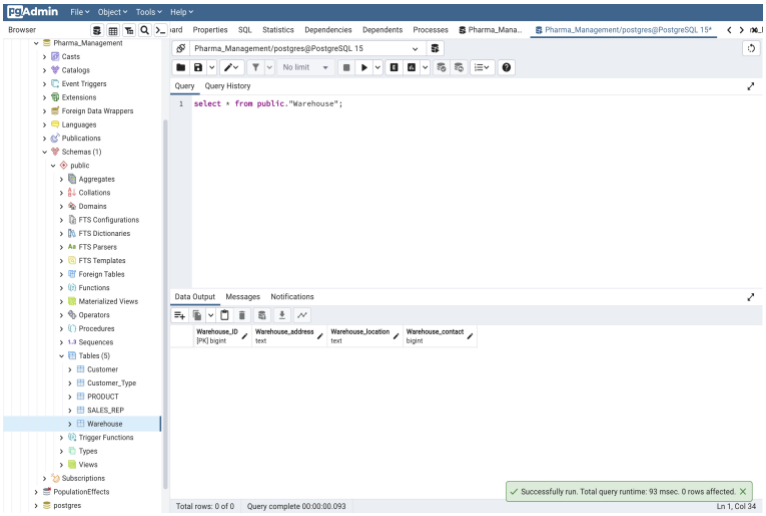
PRODUCT :

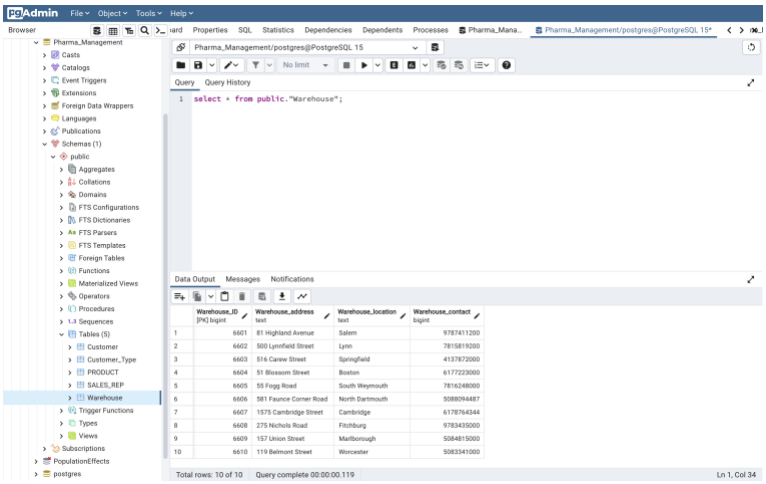


Sales rep

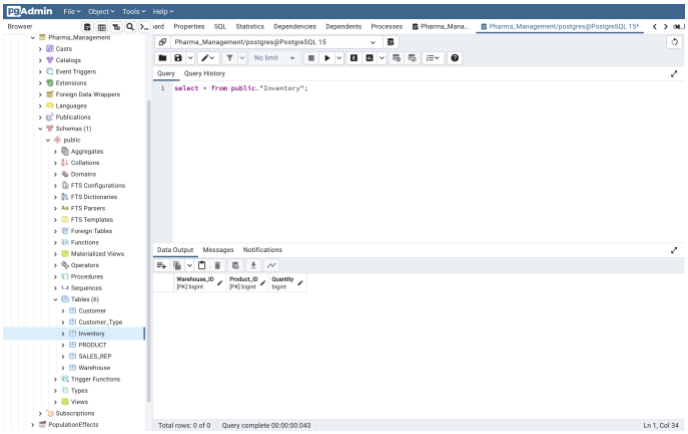


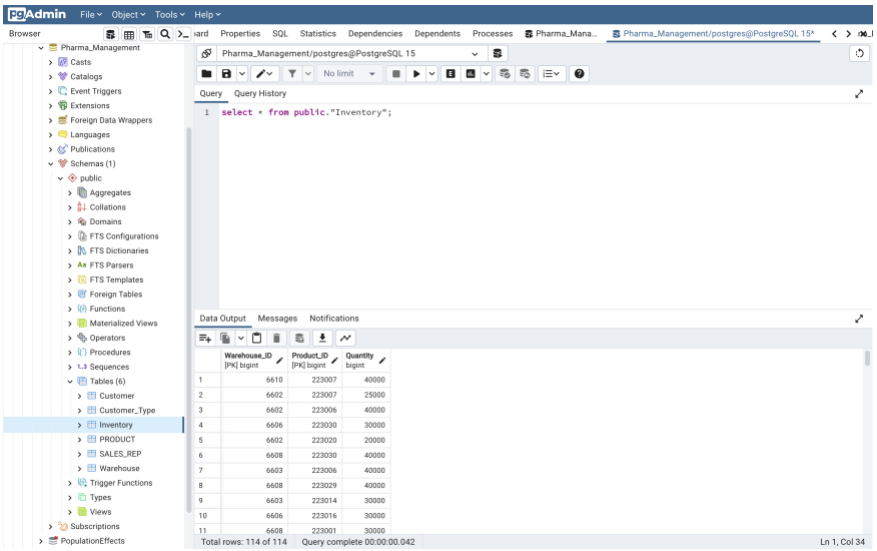
Warehouse :



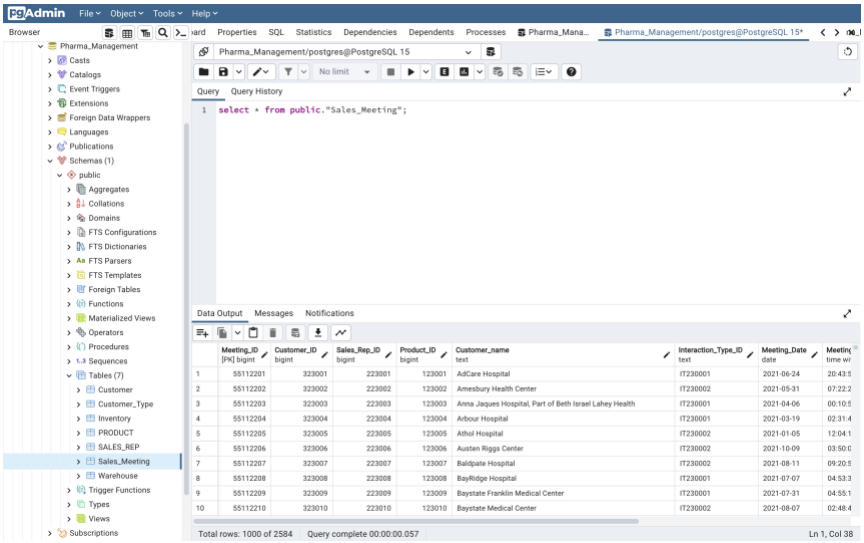


INVENTORY:

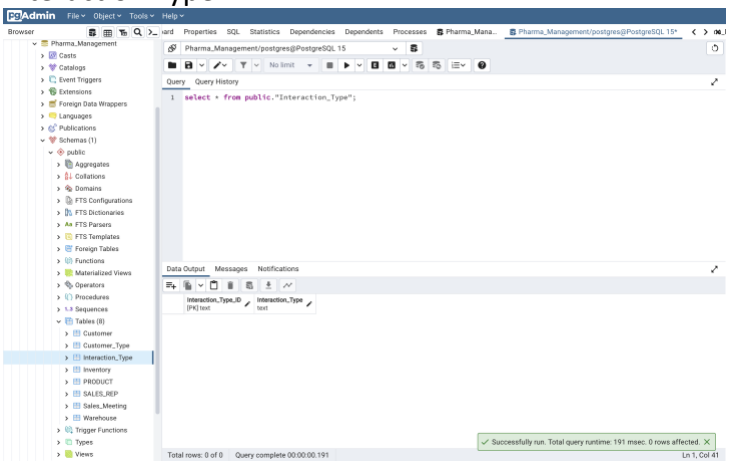


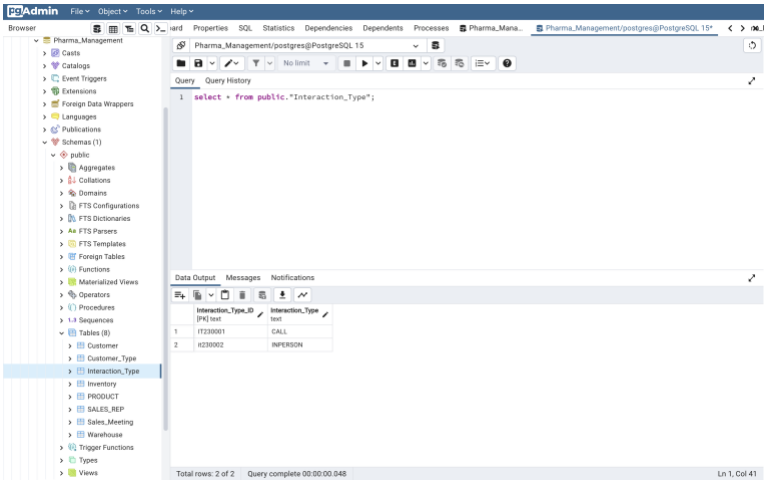


Sales_Meeting:

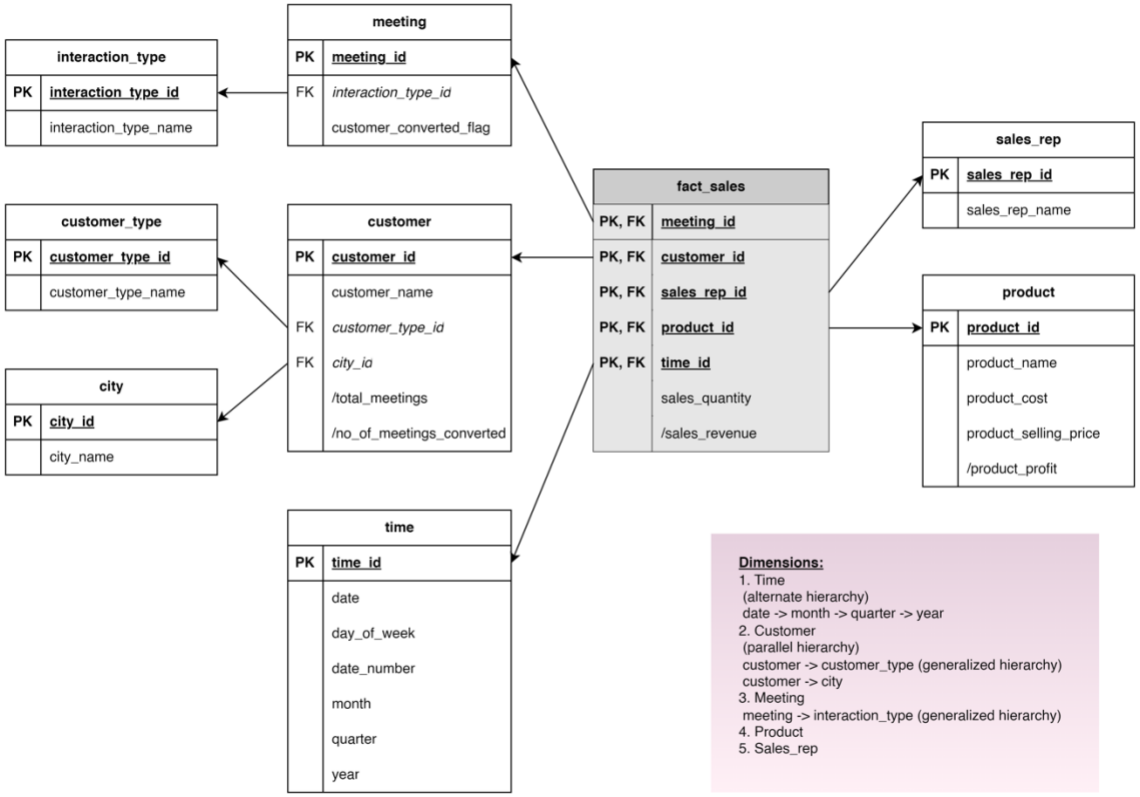


Interaction Type :





MULTIDIMESIONAL MODEL :



Dimension Tables :

Interaction_Type, Customer_type, city, time, customer, meeting, product, sales_rep

Fact Table : fact_sales

HIERARCHIES IMPLEMENTED :

1. Time (alternate hierarchy)
   Date → month → quarter → year

2. Customer (parallel hierarchy)
   Customer → customer_type(generalized hierarchy)
   Customer → city

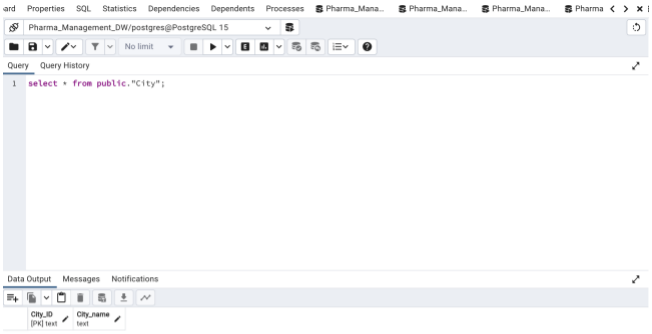
3. Meeting
   Meeting → interaction_type (generalized hierarchy)

4. Product

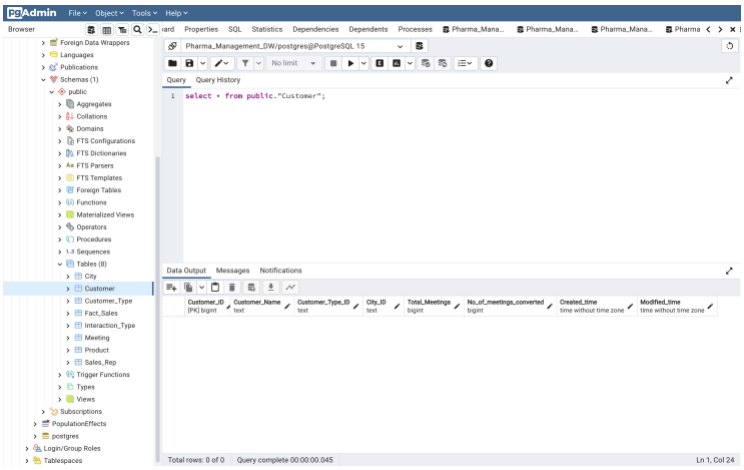
5. Sales_rep

ROLAP IMPLEMENTATION

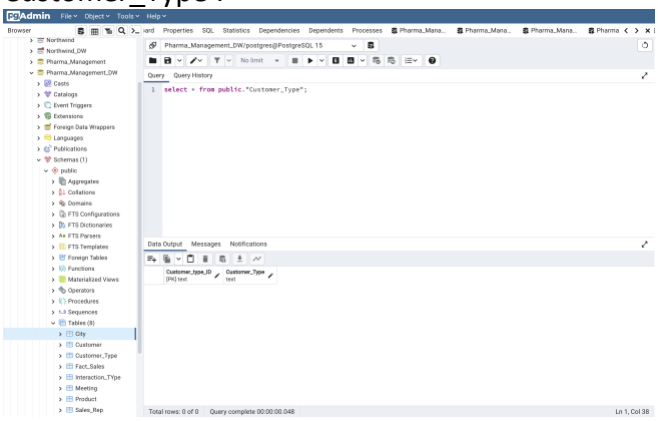
MULTIDIM SCHEMA :



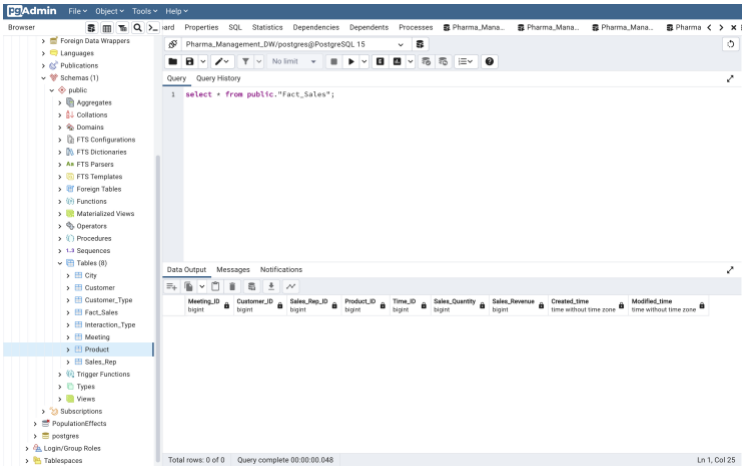
Customer :



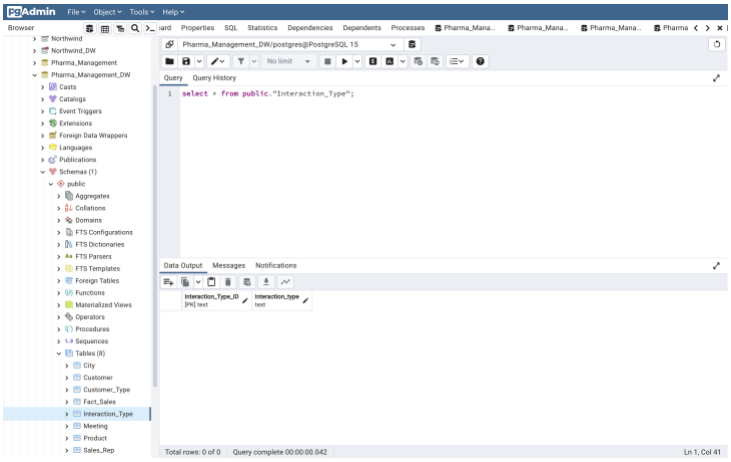
Customer_Type :



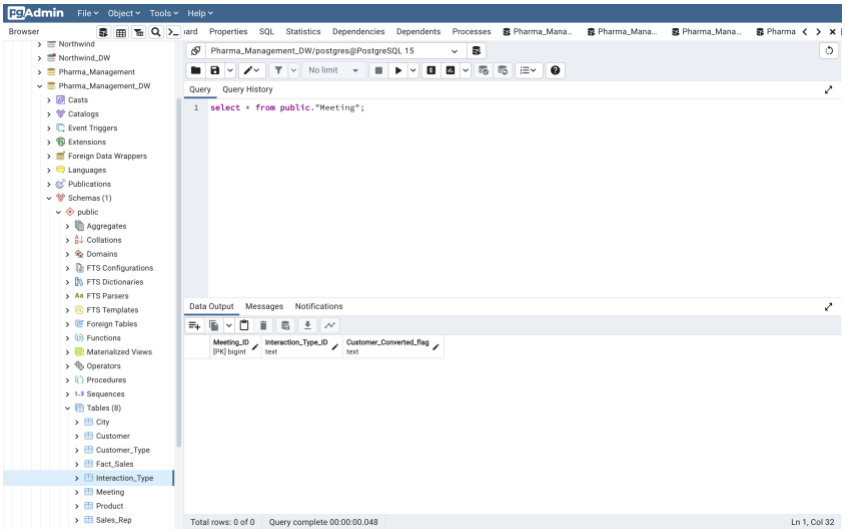
Sales fact Table :



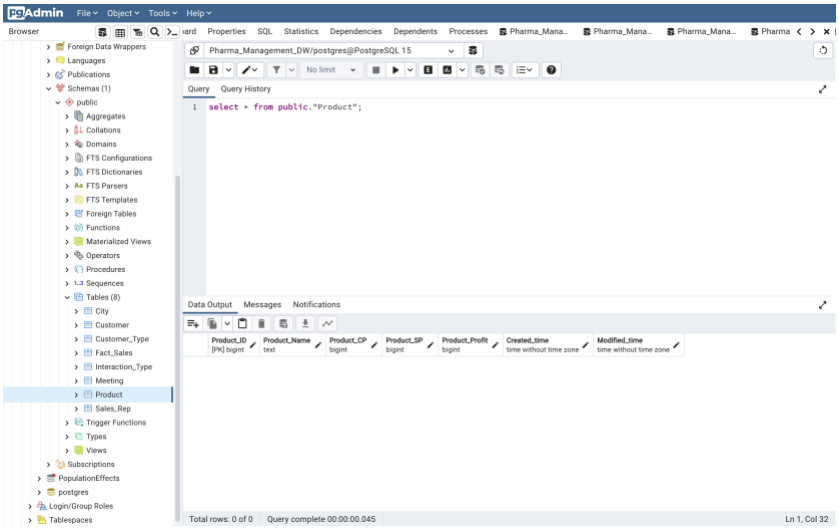
Interaction_Type :



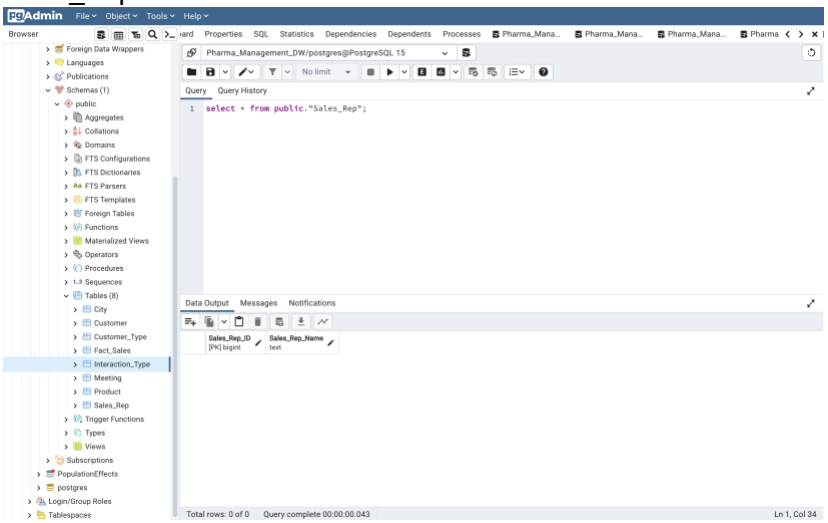
Meeting :



PRODUCT :



Sales_Rep :



OLAP Operations

1. What is the total sales target for all sales reps?

   SELECT SUM(Sales_target) AS Total_Sales_Target FROM Meetings;

   OLAP Operation: Roll-up: Total_Sales_Target by sales_rep_id, customer_id, and product_id

2. What is the total sales target for each sales rep?

   SELECT sales_rep_id, SUM(Sales_target) AS Sales_Target FROM Meetings GROUP BY sales_rep_id;

OLAP Operation: Roll-up: Sales_Target by sales_rep_id and customer_id

3. What is the average meeting duration by product?

   SELECT product_name, AVG(meeting_duration) AS Avg_Meeting_Duration FROM Meetings JOIN Products ON Meetings.product_id = Products.product_id GROUP BY product_name;

   OLAP Operation: Slice: Avg_Meeting_Duration by product_name

4. What is the total number of meetings for each customer who has converted?

   SELECT customer_id, COUNT(Meeting) AS Total_Meetings FROM Meetings WHERE Customer_converted_Y_N = 'Y' GROUP BY customer_id;

   OLAP Operation: Slice: Total_Meetings by customer_id, where Customer_converted_Y_N = 'Y'

5. What is the total revenue generated by each sales rep, including the cost of products?

   SELECT sales_rep_id, SUM(product_sellingprice - product_costprice) AS Total_Revenue FROM Meetings JOIN Products ON Meetings.product_id = Products.product_id GROUP BY sales_rep_id;

   OLAP Operation: Drill-down: Total_Revenue by sales_rep_id and product_id, including product_costprice and product_sellingprice

6. How many meetings have occurred for each product?
   SELECT product_name, COUNT(Meeting) AS Total_Meetings FROM Meetings JOIN Products ON Meetings.product_id = Products.product_id GROUP BY product_name;

   OLAP Operation: Drill-down: Total_Meetings by product_name

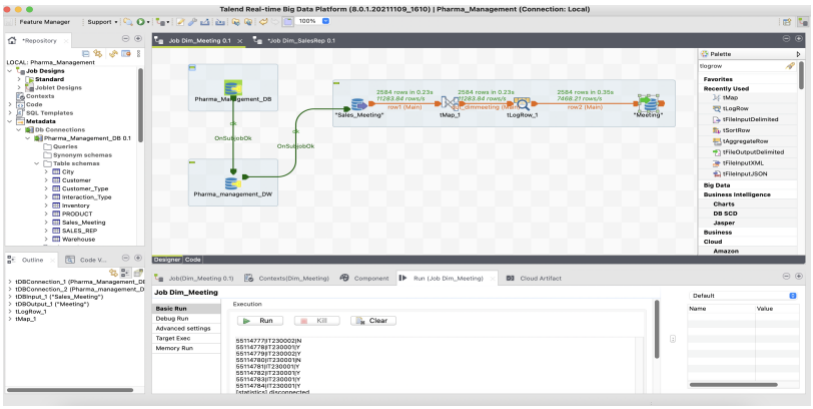
6. What is the total revenue generated by each product?

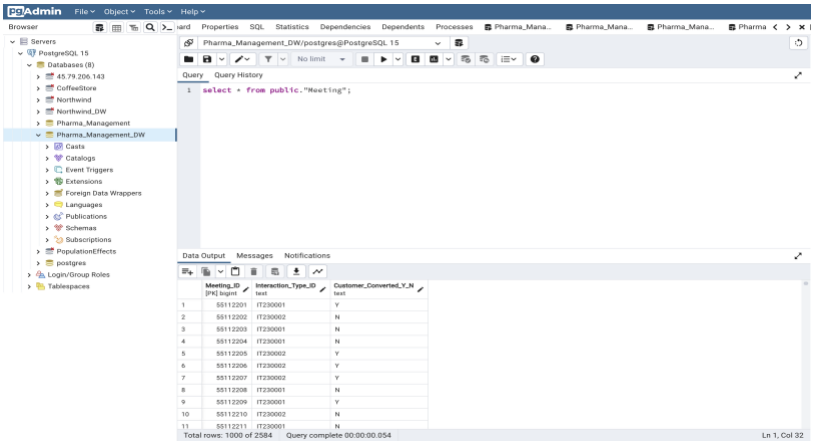
   SELECT product_name, SUM(product_sellingprice - product_costprice) AS Total_Revenue FROM Meetings JOIN Products ON Meetings.product_id = Products.product_id GROUP BY product_name;

   OLAP Operation: Drill-down: Total_Revenue by product_name, including product_costprice and product_sellingprice

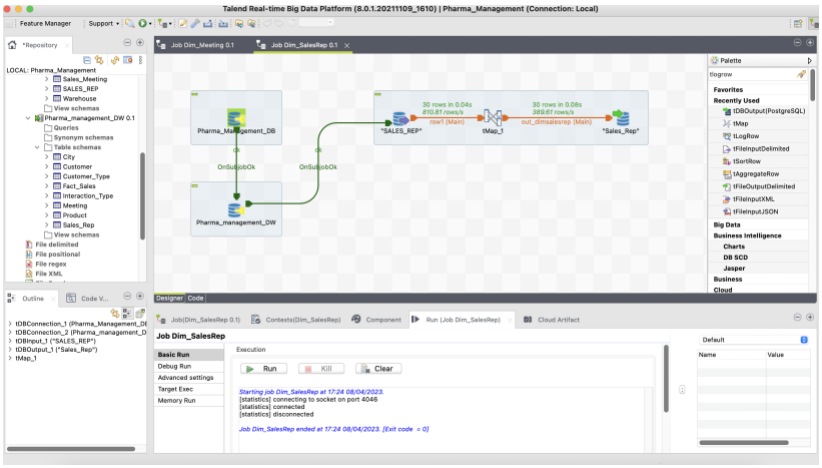
TALEND IMPLEMENTATION

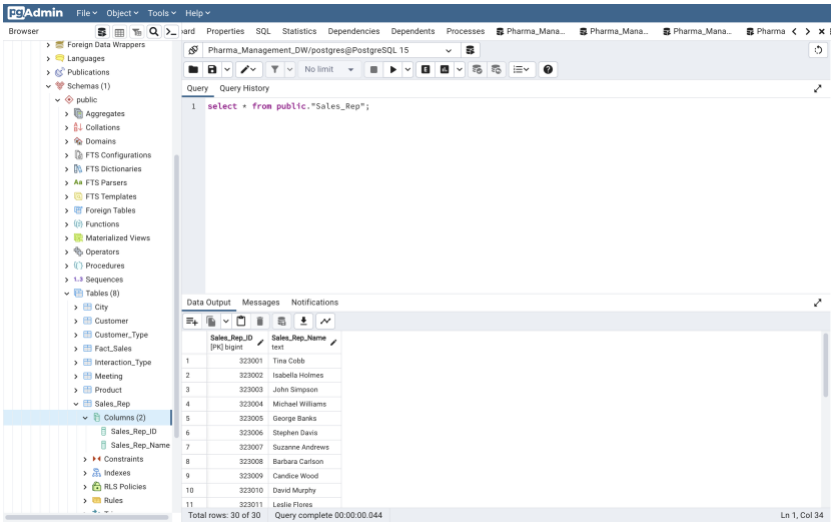
Meeting Dimension Table :



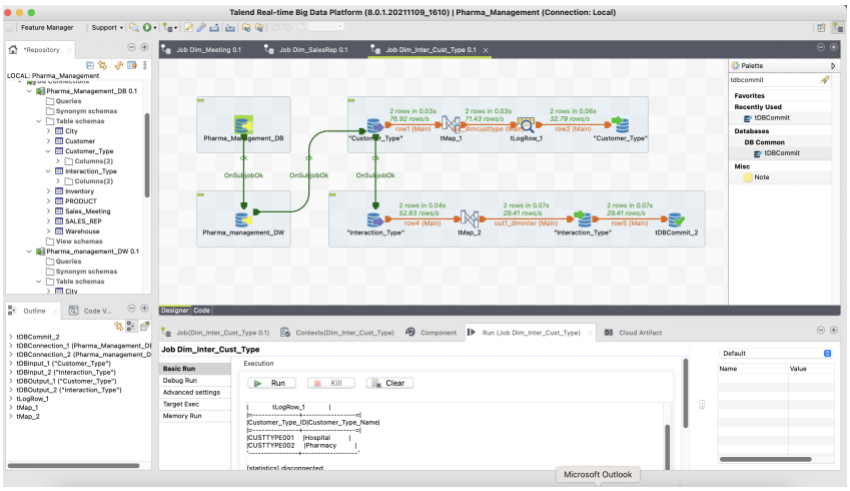


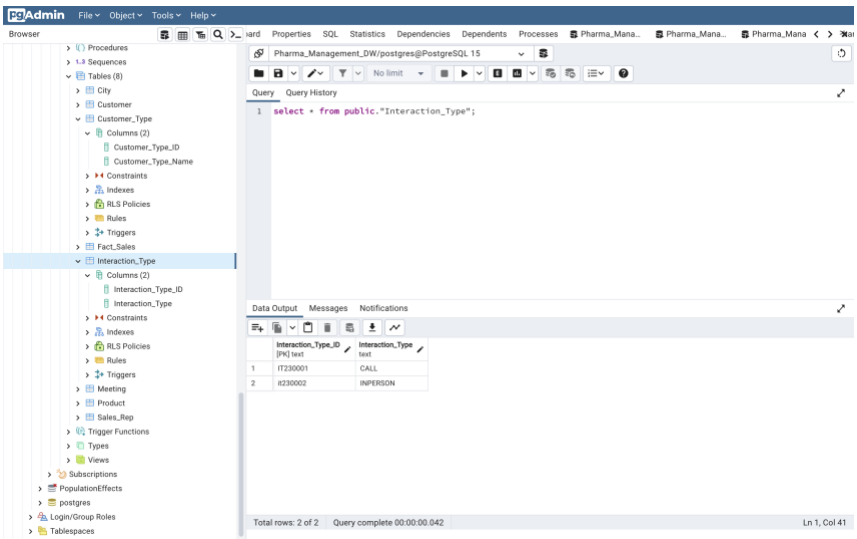
Sales_Rep Dimension Table :

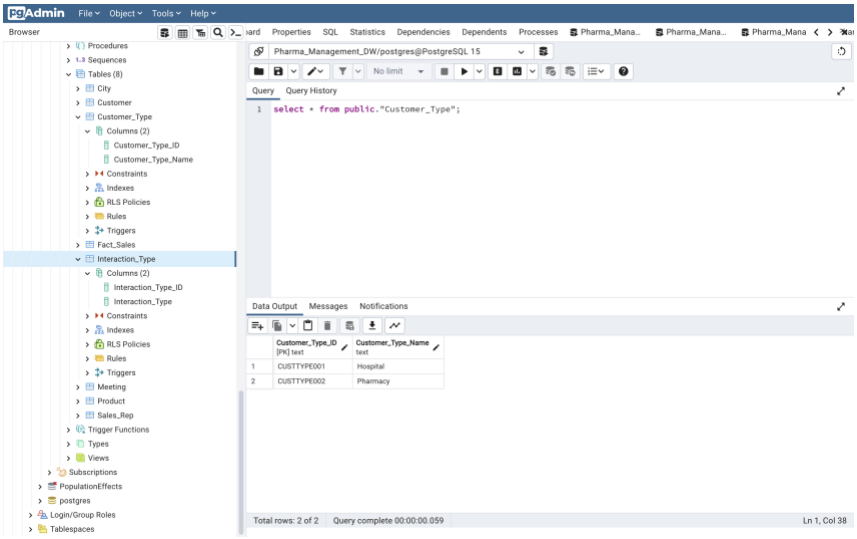




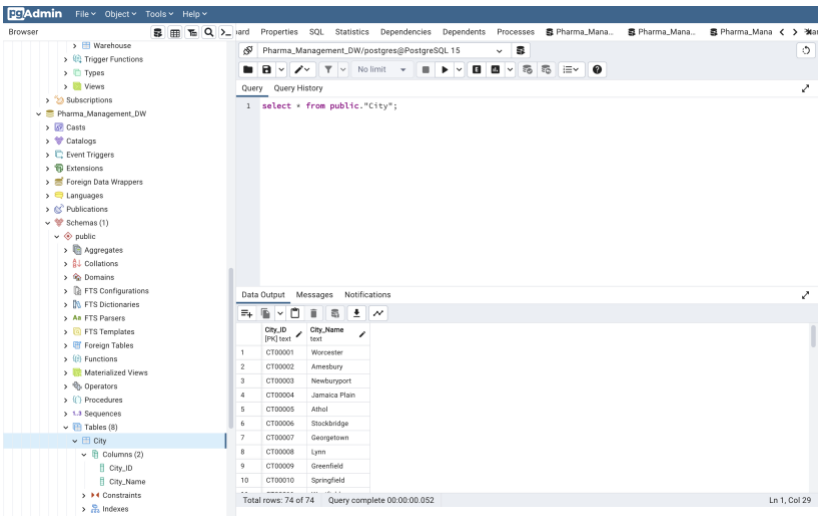
Customer_Type and Interaction_Type Dimension Tables :



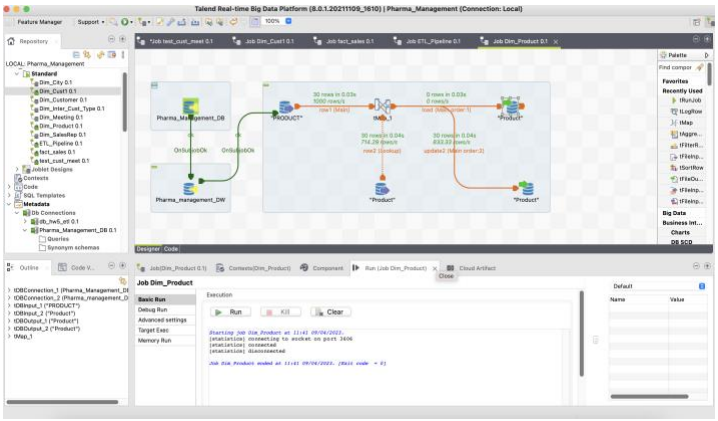


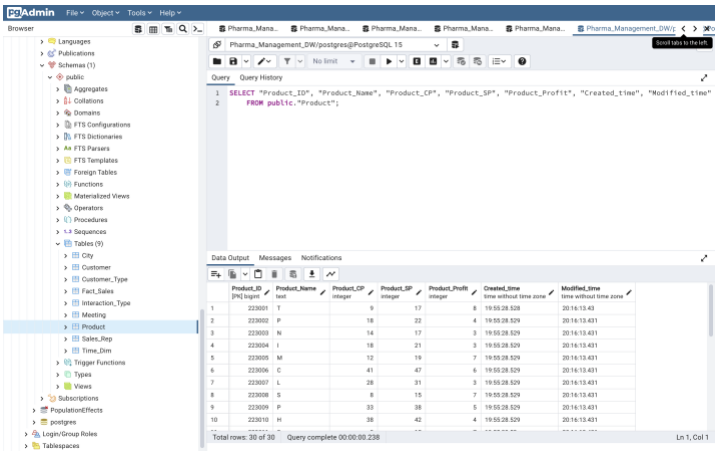


City Dimension table :

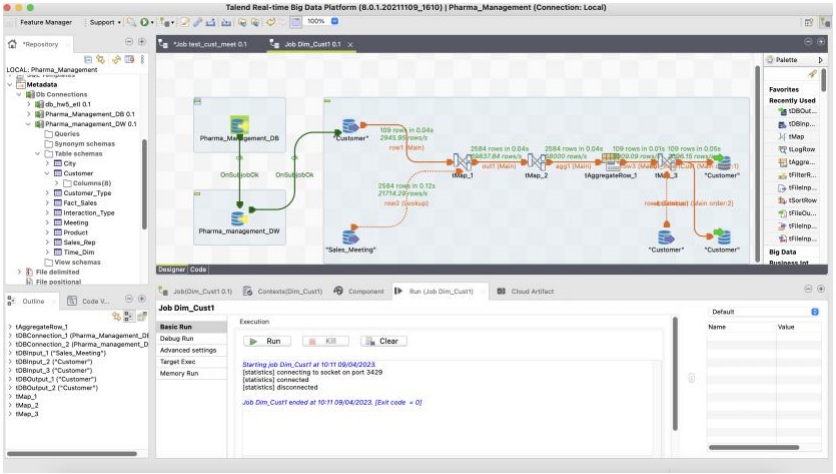


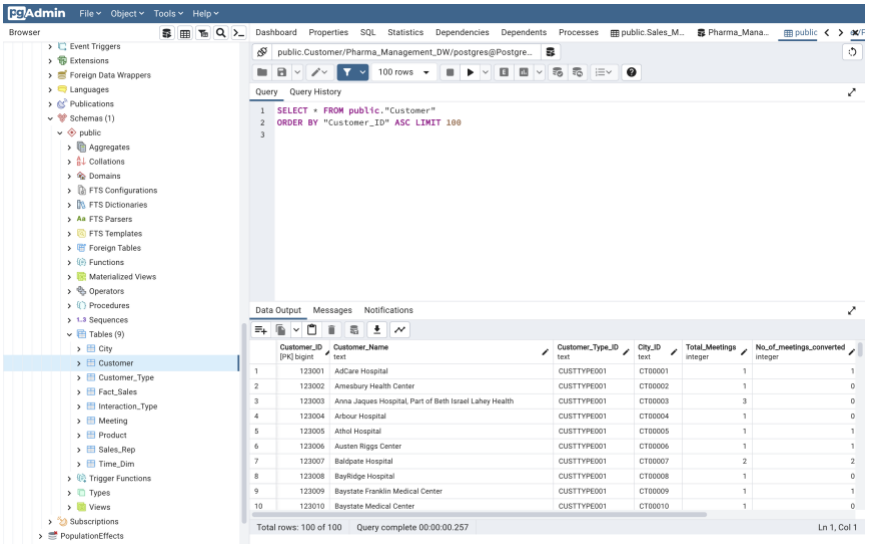
PRODUCT DIMENSION TABLE :

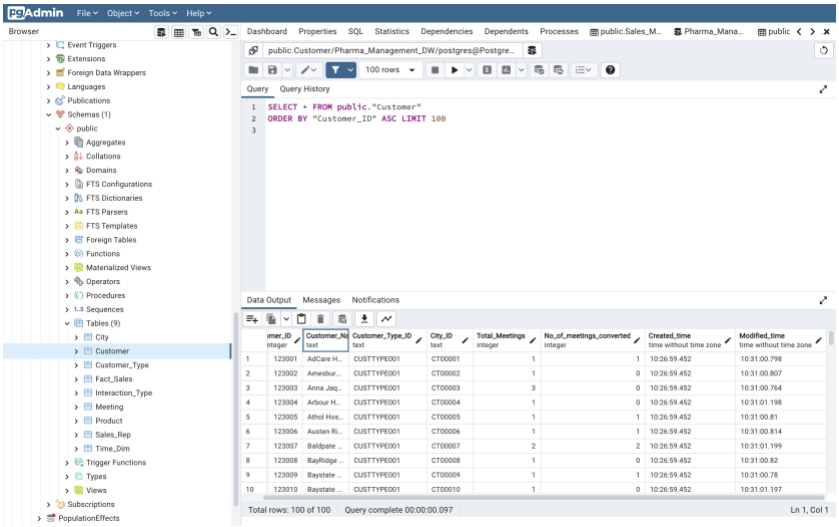




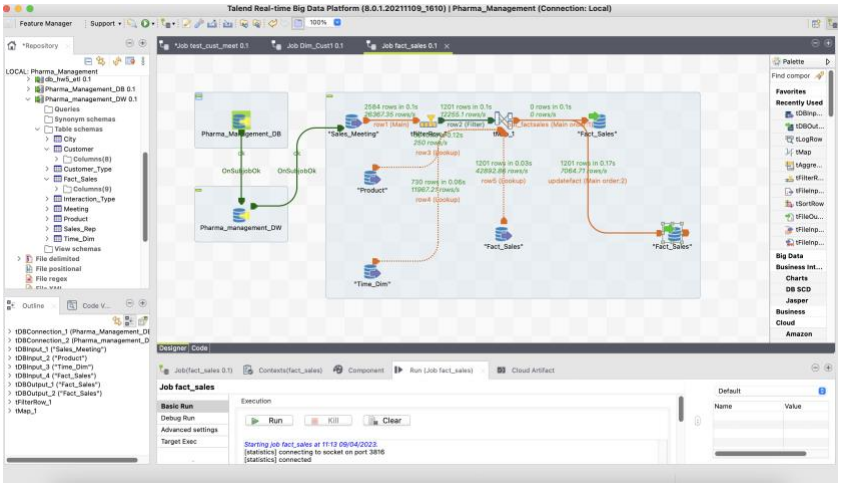
Customer Dimension table :

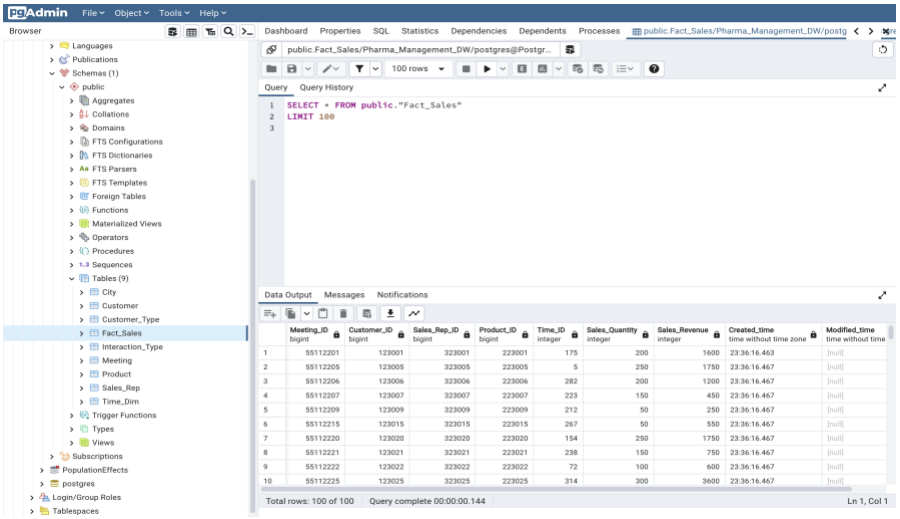


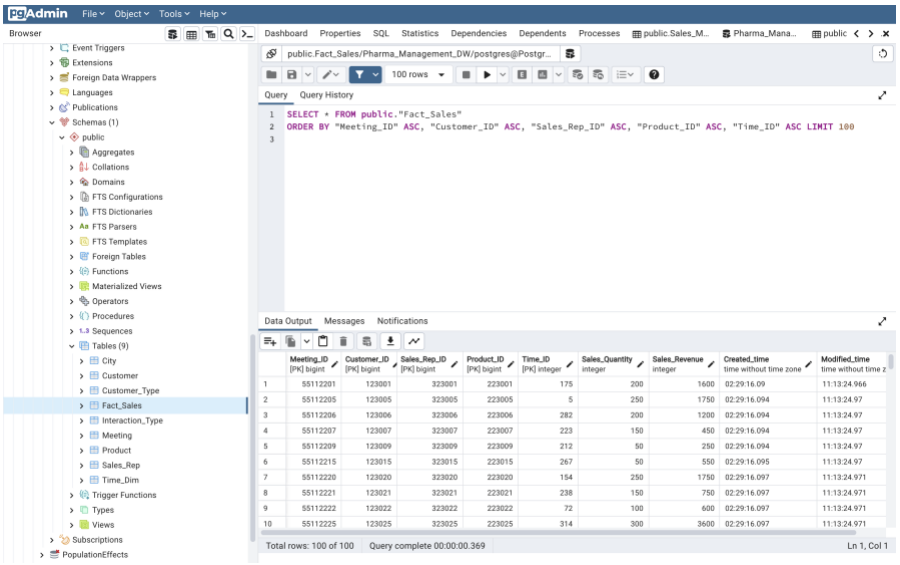




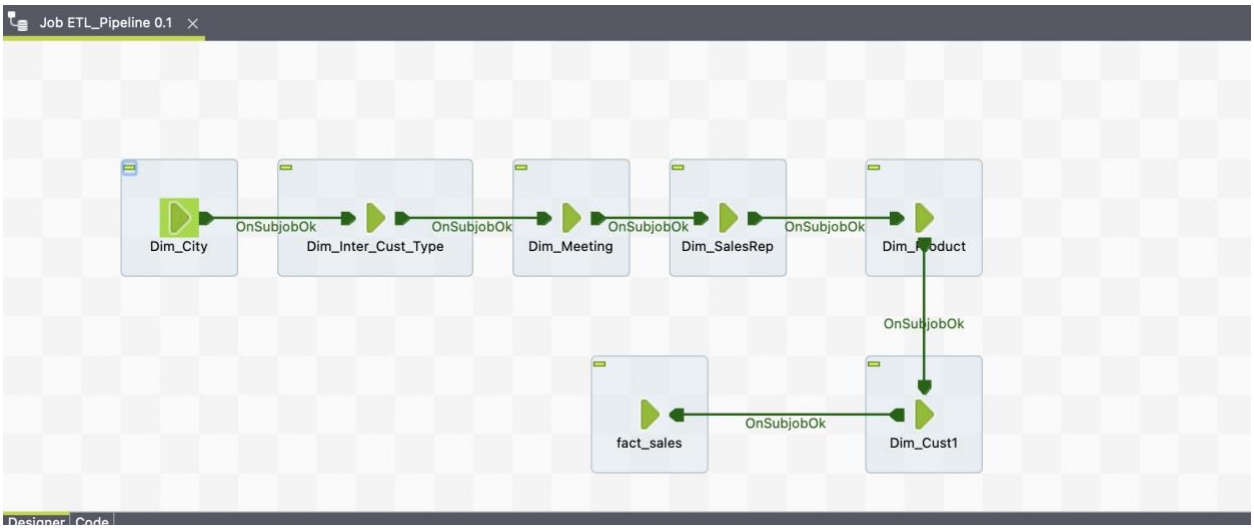
Sales Fact Table :







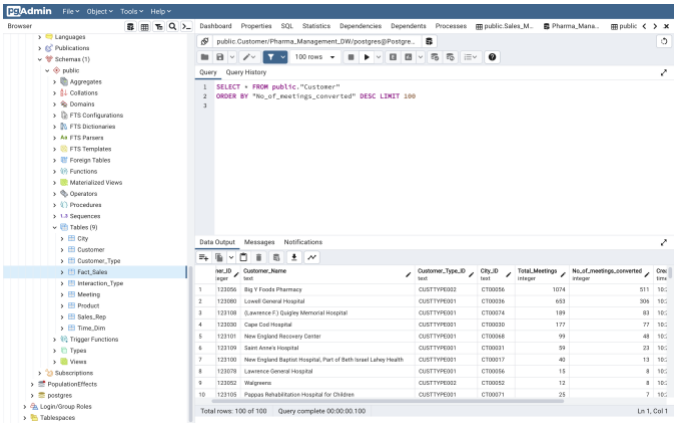
FINAL TALEND ETL PIPELINE :



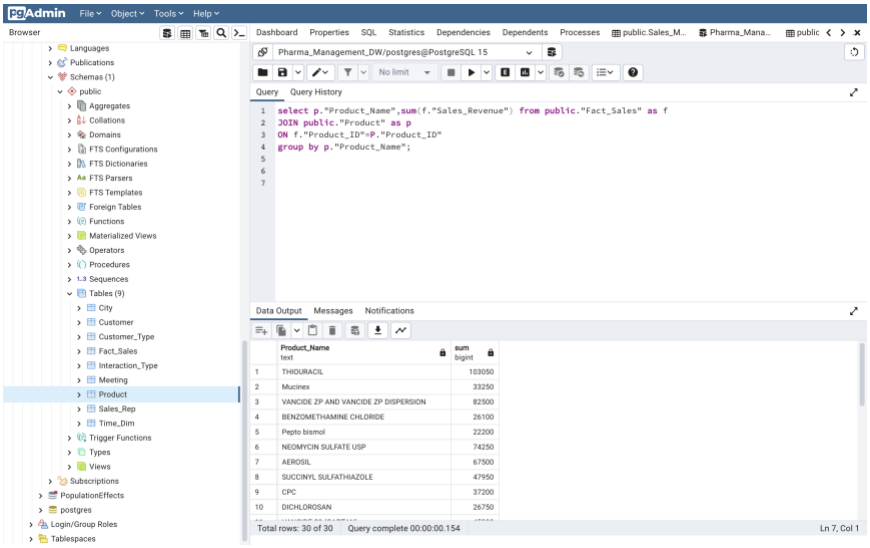
SOME QUERY RESULTS :

Customers who have converted most number of meetings by count

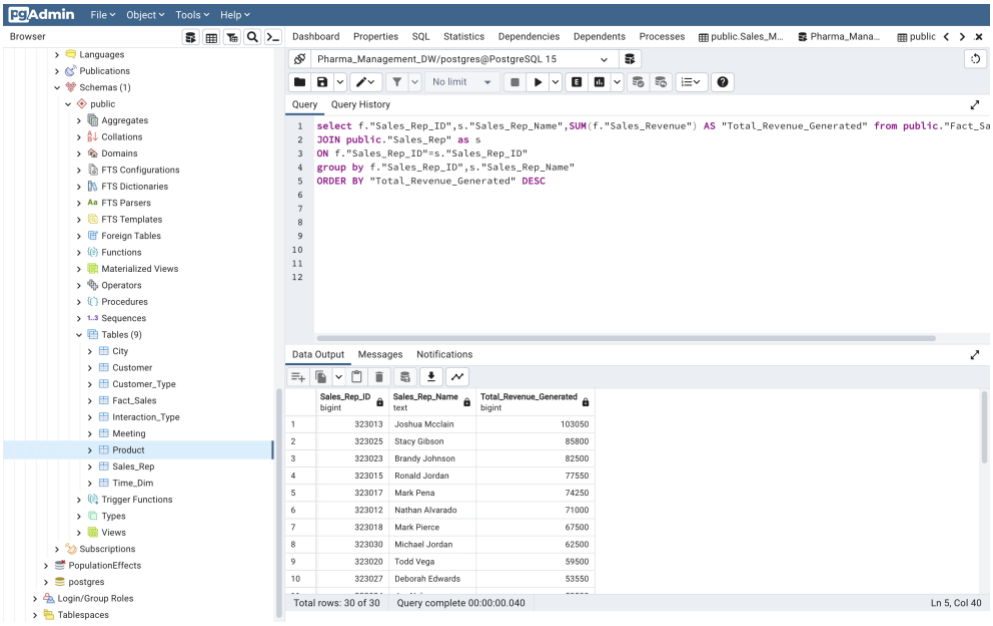
Ans. It can be observed that Big Y Foods Pharmacy is responsible for the highest converted calls with 511 calls



Product generating most revenue : THIOURACIL



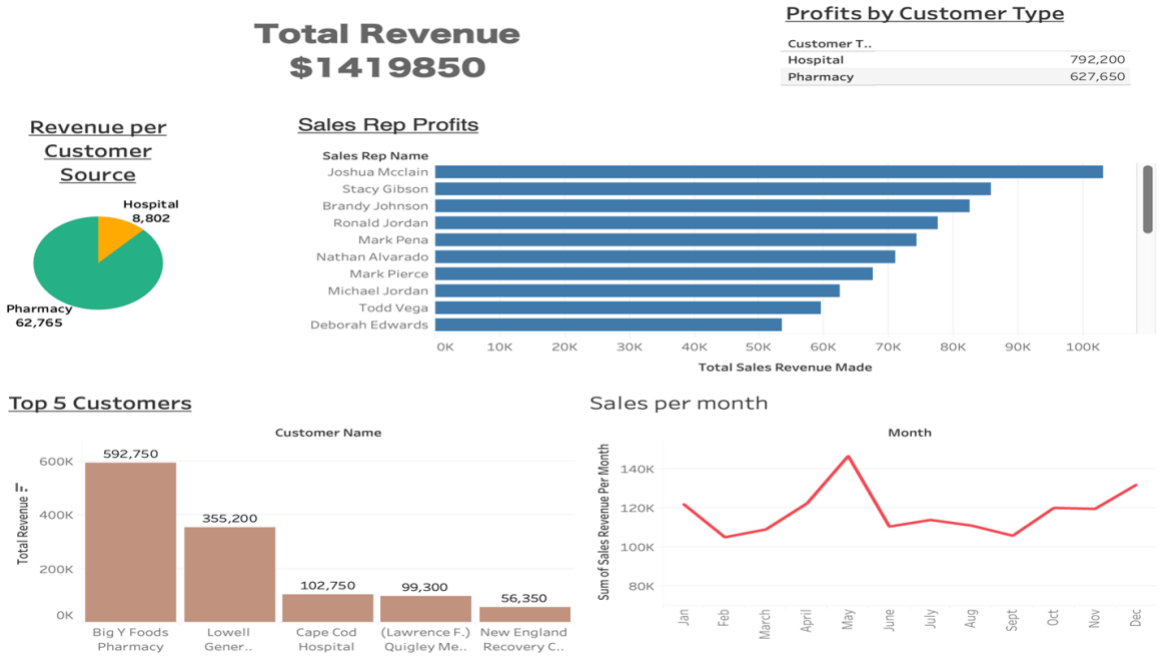
List of sales reps who have generated most revenue


pgAdmin
File
Object
Tools
Help
Browser
Dashboard
Properties
SQL
Statistics
Dependencies
Dependents
Processes
public.Sales_M...
Pharma_Mana...
public
Languages
Publications
Schemas (1)
public
Aggregates
Collations
Domains
FTS Configurations
FTS Dictionaries
FTS Parsers
FTS Templates
Foreign Tables
Functions
Materialized Views
Operators
Procedures
Sequences
Tables (9)
City
Customer
Customer_Type
Fact_Sales
Interaction_Type
Meeting
Product
Sales_Rep
Time_Dim
Trigger Functions
Types
Views
Subscriptions
PopulationEffects
postgres
Login/Group Roles
Tablespaces
Pharma_Management_DW/postgres@PostgreSQL 15
No limit
Query
Query History
1 select f."Sales_Rep_ID",s."Sales_Rep_Name",SUM(f."Sales_Revenue") AS "Total_Revenue_Generated" from public."Fact_Sa
2 JOIN public."Sales_Rep" as s
3 ON f."Sales_Rep_ID"=s."Sales_Rep_ID"
4 group by f."Sales_Rep_ID",s."Sales_Rep_Name"
5 ORDER BY "Total_Revenue_Generated" DESC
6
7
8
9
10
11
12
Data Output
Messages
Notifications
Sales_Rep_ID bigint
Sales_Rep_Name text
Total_Revenue_Generated bigint
1 323013 Joshua Mcclain 103050
2 323025 Stacy Gibson 85800
3 323023 Brandy Johnson 82500
4 323015 Ronald Jordan 77550
5 323017 Mark Pena 74250
6 323012 Nathan Alvarado 71000
7 323018 Mark Pierce 67500
8 323030 Michael Jordan 62500
9 323020 Todd Vega 59500
10 323027 Deborah Edwards 53550
Total rows: 30 of 30
Query complete 00:00:00.040
Ln 5, Col 40

MILESTONE 7

DATA VISUALIZATION ON TABLEAU

DASHBOARD 1 : SALES REVENUE SUMMARY



CONCLUSION FROM ABOVE DASHBOARD:

The following KPI Metrics were defined and certain conclusions were drawn from them

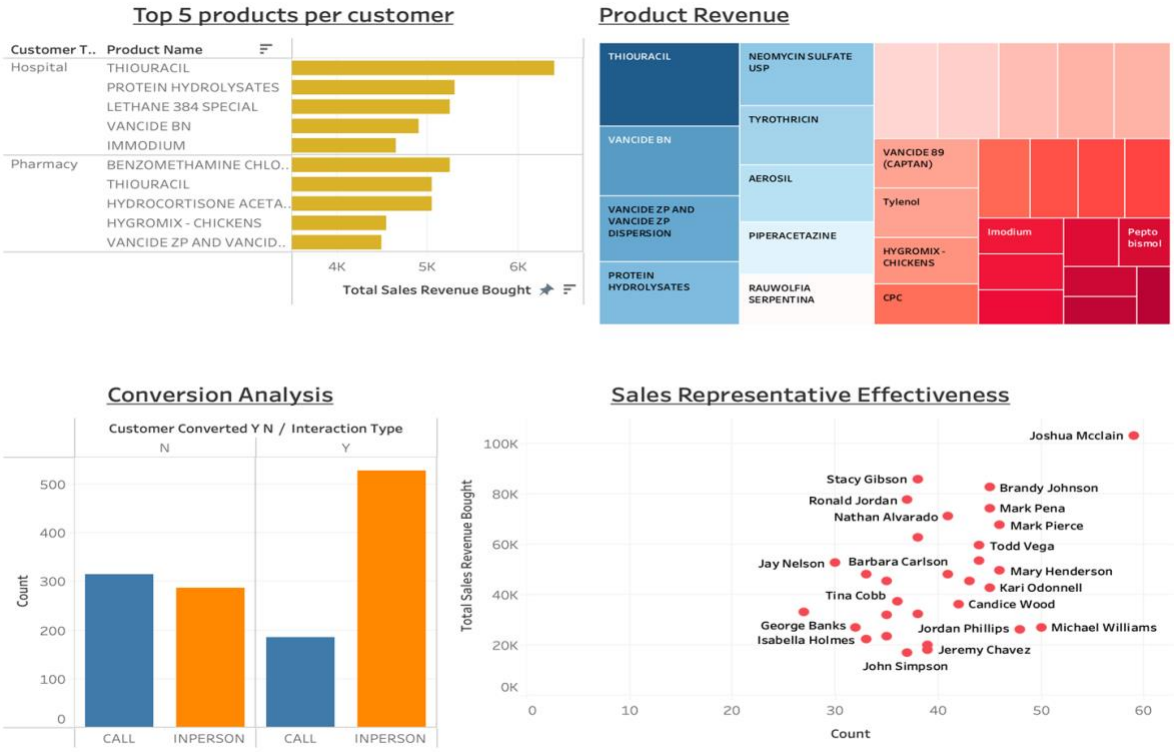
1. **Total Revenue** - $1419850

2. **Profits per Customer Type:**
   We can observe that the revenue generated by Hospitals and Pharmacies were $792200 and $627650 respectively

3. **Revenue per Customer Source :**
   We have about 99 hospitals and 11 pharmacies in the dataset. Though the overall revenue generated by hospitals is more than pharmacies, but who is more effective? Upon calculating the revenue per customer source, we can observe that each Pharmacy generates about $62,765 and each hospital generates about $8,800 of revenue. Thus we can observe that Pharmacies generate a lot more revenue and the focus of the business should be to consider Pharmacies as the primary clients.

4. **Sales per Month**
   The month of May generates the maximum revenue.

DASHBOARD 2 – PERFORMANCE ANALYSIS



CONCLUSIONS DRAWN FROM ABOVE DASHOARD:

1. **Top 5 products per Customer**
   Thouracil and Benzamethamine Chloride generate the maximum Sales revenue for Hospital and Pharmacies respectively.

2. **Product revenue:**
   Thouracil generates the maximum revenue.

3. **Conversion Analytics :**
   We can clearly see that InPerson meetings are much more effective and hence the business should plan more Inperson meetings to have a better conversion rate.

4. **Sales Represenatative Effectiveness :**
   One of the most important metrics, it helps us to understand that Joshua Mcclain was the most hardworking and most profit generating while George banks held the least meetings and generated least revenue. Thus business can improve by training employees who are less effective and reward the higher performing employees.