Northeastern University
College *of* Engineering

Data Warehousing & Integration
IE 6750
FALL 2024

[Restaurant & Food Delivery]

On prem Project Report

Group 11

Student 1- Jahnavi Redyy Ganesina

Student 2- Manasi Bondalapati

Student 1 ganesina.j@northeastern.edu

Student 2 bondalapati.m@northeastern.edu

Submission Date:_____12/07/24_____

Table Of Contents

Problem Statement

In the dynamic restaurant and food delivery sector, businesses face challenges in managing extensive orders, delivery logistics, inventory, and sales across various platforms. Data fragmentation from multiple sources like POS systems, delivery platforms, customer databases, and supply chain management can lead to operational inefficiencies such as delayed deliveries, inaccurate inventory forecasts, and diminished customer satisfaction.

Business owners often lack a unified view of their operations, making it difficult to make data-driven decisions to enhance profitability and service quality. Implementing a centralized data integration and analytics solution can address these challenges by consolidating data from various sources, enabling restaurants and delivery services to analyze key metrics and derive actionable insights. This approach can streamline operations, improve customer experiences, and enhance overall business performance.

Project Objective

The primary goal of this project is to implement a data integration and analytics solution using Talend to process and analyze data from the restaurant and food delivery sector. The project involves:

1. Extracting data from a PostgreSQL OLTP system.
2. Transforming the data using various Talend components to ensure consistency, accuracy, and usability.
3. Loading the transformed data into a PostgreSQL OLAP database for advanced analytics.
4. Implementing Slowly Changing Dimension (SCD) Type 2 to track historical changes in restaurant ratings, enabling detailed analysis of trends over time.

This Talend-based ETL pipeline ensures efficient data processing and provides a centralized platform for analyzing sales, customer behavior, and operational efficiency, supporting data-driven decision-making.

Key KPIs and Metrics Analyzed

Sales and Order Metrics

- Order Value and Service Fee Correlation: Identifying the relationship between order values and associated service fees.
- Order Frequency by Day of Week: Analyzing peak order days to optimize staffing and resources.

Delivery Partner Performance Metrics

- Delivery Partner Performance: Evaluating efficiency based on delivery time, ratings, and success rates.

- Average Delivery Time by Region: Identifying areas with delays and improving logistics.

Customer Behavior Metrics

- Top Restaurants Based on Ratings: Highlighting restaurants with the best customer feedback.
- Popular Payment Types: Analyzing preferred payment methods to enhance payment gateway efficiency.

Operational Metrics

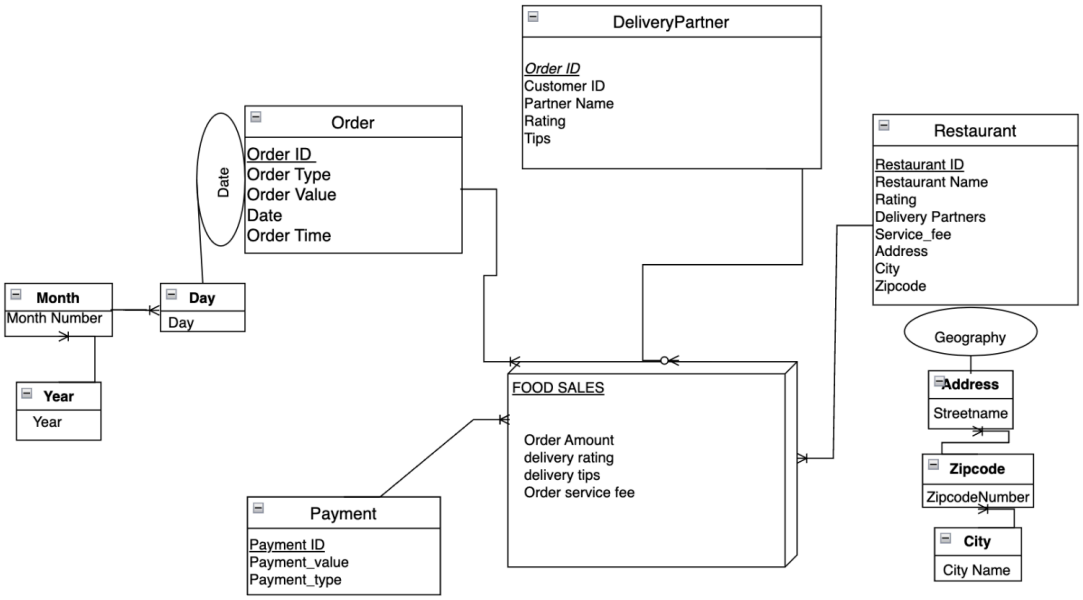
- Inventory Forecasting Accuracy: Leveraging historical order data to predict future inventory needs.
- Delivery Route Optimization: Improving delivery times through data-driven route planning.

Restaurant Performance Metrics

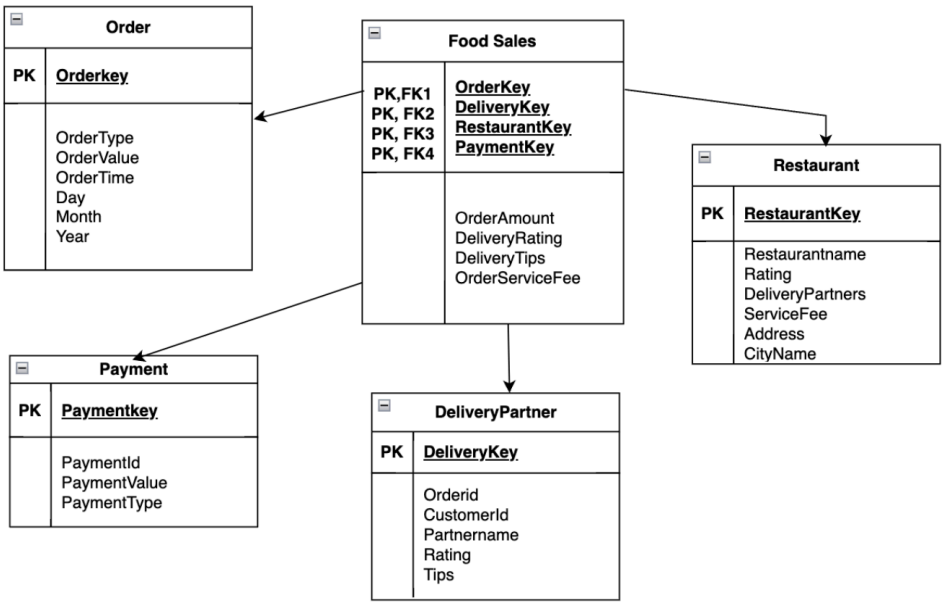
- Revenue per Restaurant: Identifying top-performing restaurants.
- Order Cancellation Rates: Addressing the root causes of cancellations to improve service reliability.

This project demonstrates how integrating data using Talend and maintaining historical trends with SCD Type 2 can deliver actionable insights, ensuring higher operational efficiency, improved customer satisfaction, and enhanced business growth.

CONCEPTUAL DIAGRAM



LOGICAL DIAGRAM

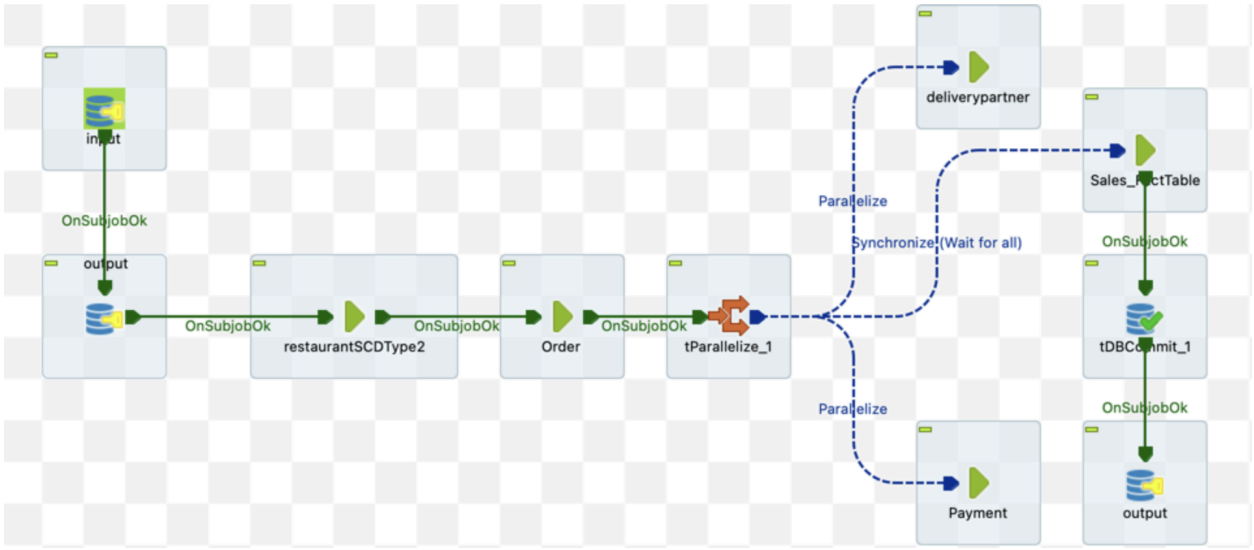


Data Sources

The pipeline processes multiple data sources:
1)Restaurant data was sourced from the CSV file "Food Establishment Inspections in the City of Boston" by @davidtalby.
2) Additional data was generated using ChatGPT and python scripts for the project.
3) https://mockaroo.com/

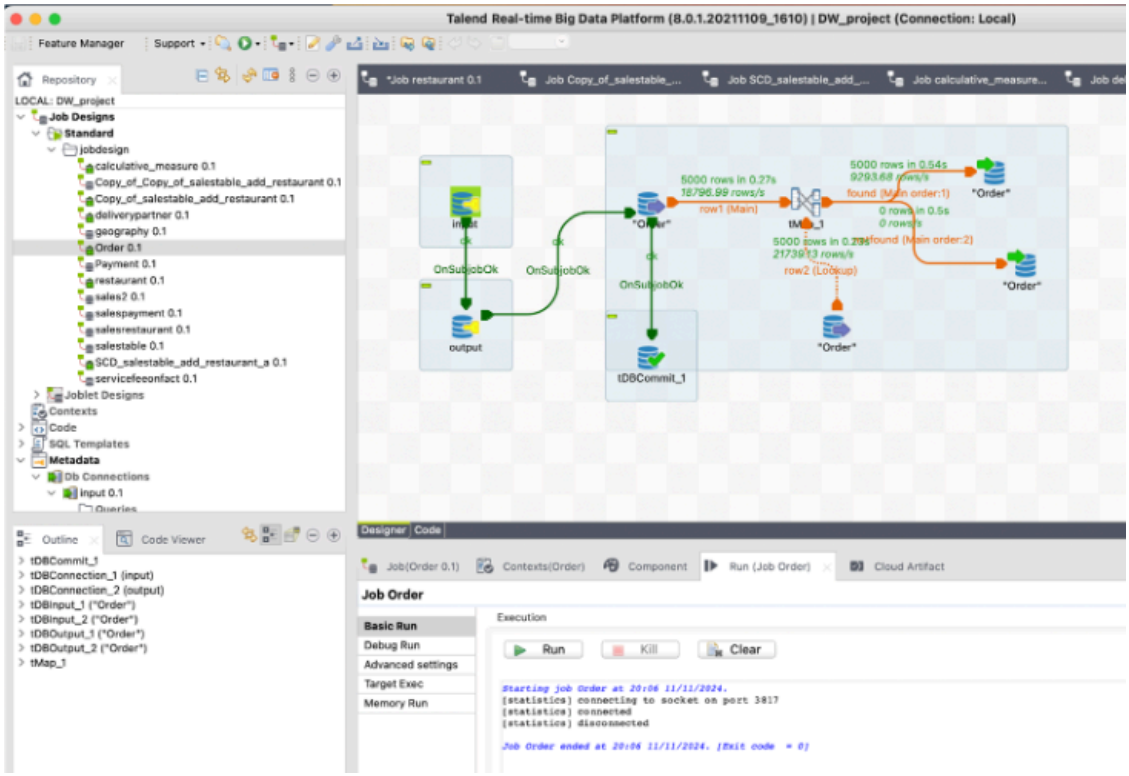
Data Warehouse:
- Order (Postgres)
- Payment (Postgres)
- Order Reviews (Postgres)
- Delivery Partner (Postgres)
- Restaurant(Postgres)
- Geography(CSV format)

ETL DIAGRAM

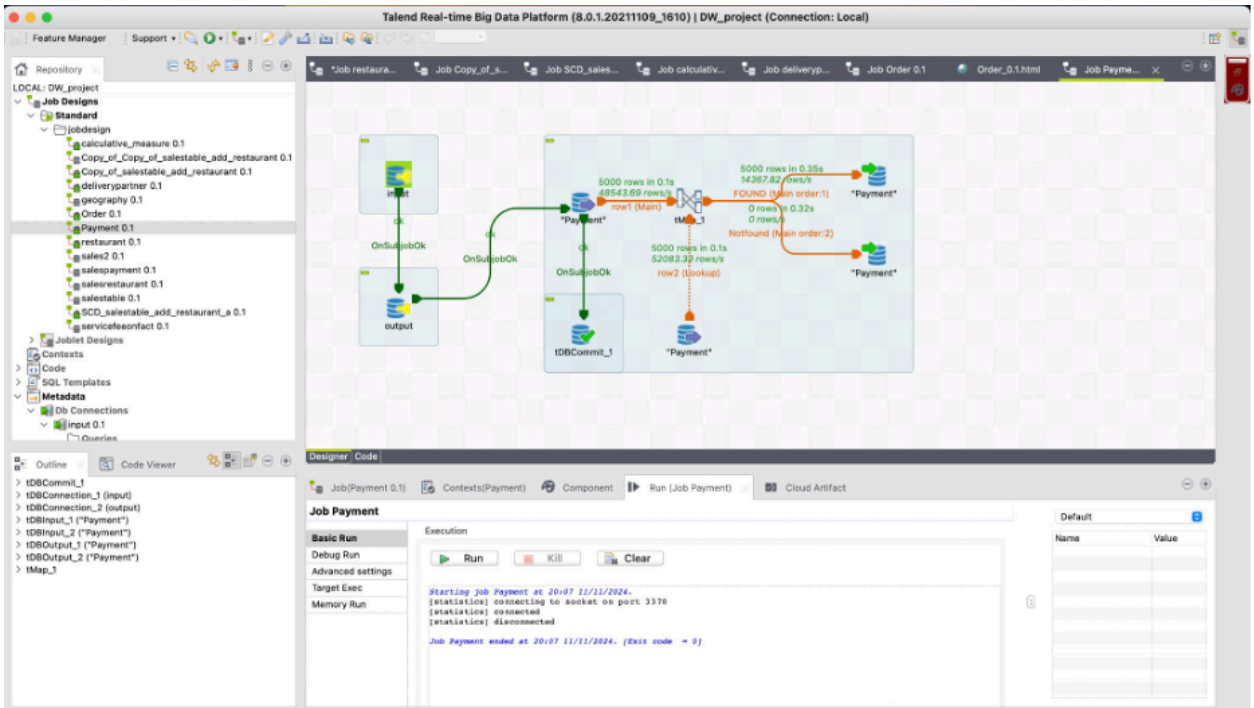


Data Loading

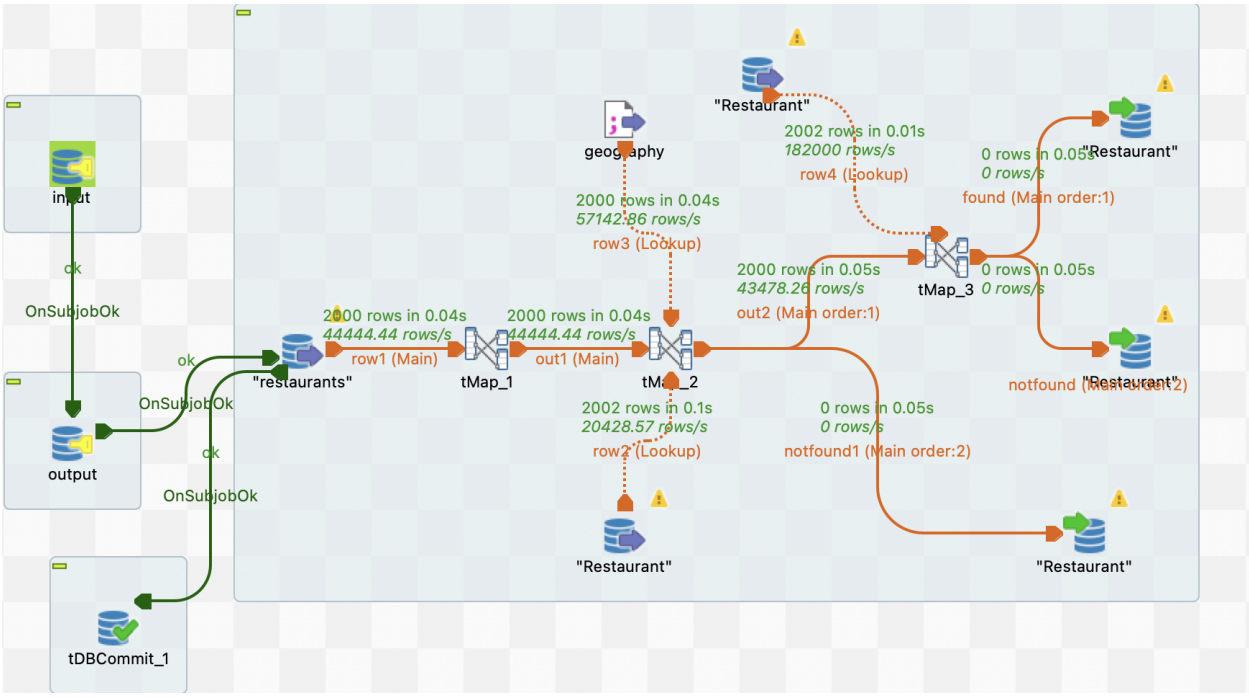
1) Order(Dimension)



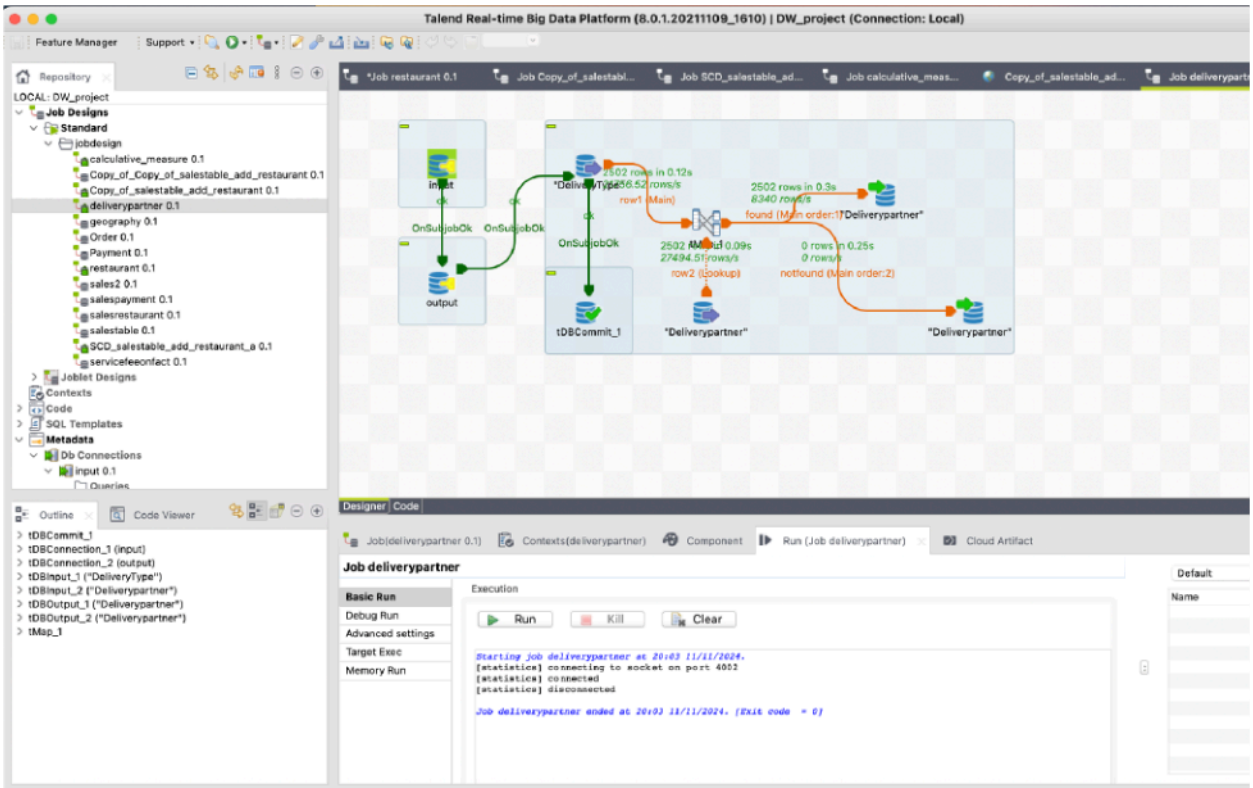
2) Payment(Dimension)



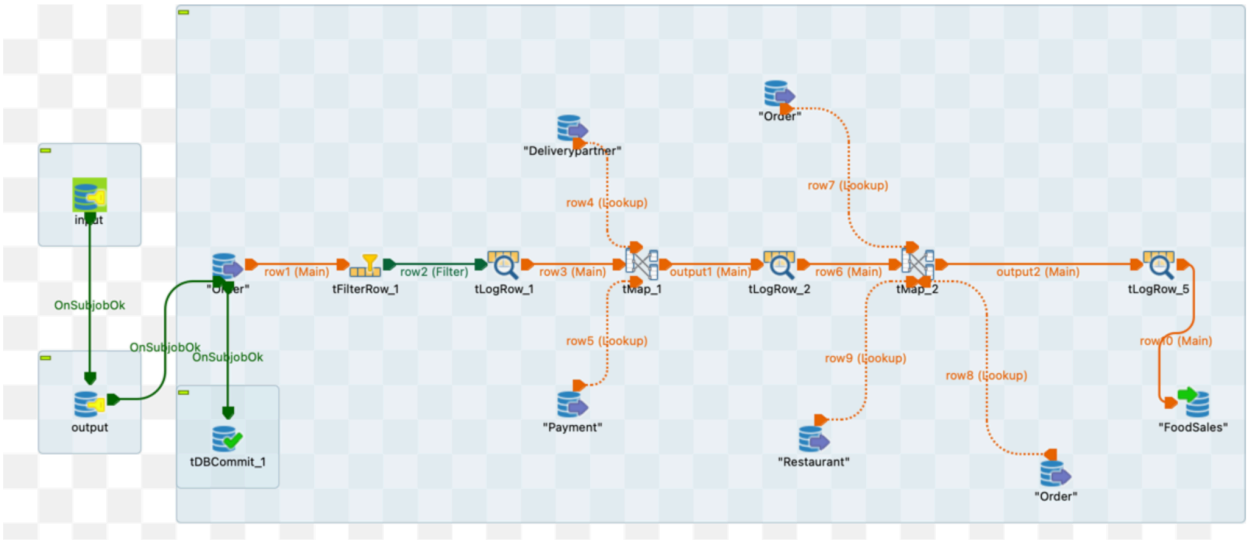
3) Restaurant(Dimension)
Loading the street address, city and zipcode from a csv file.



4) Delivery Partner(Dimension)

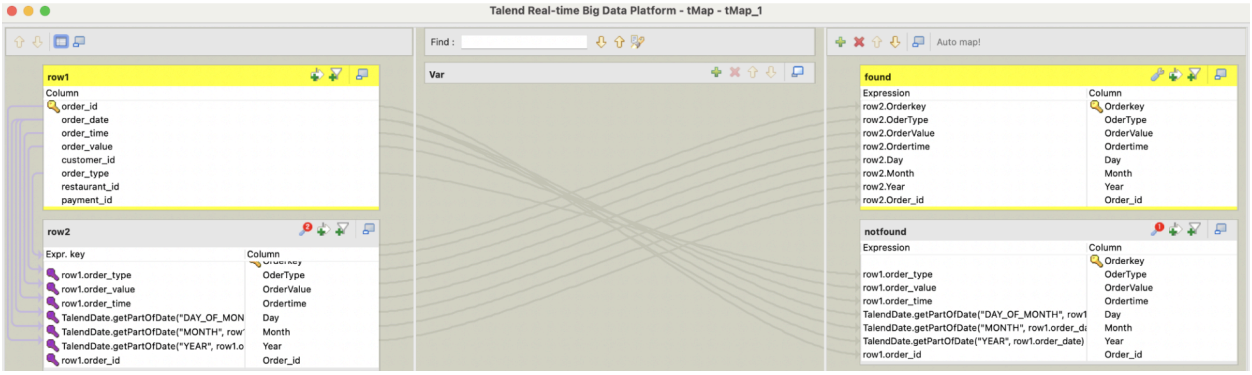


5) Food Sales (Fact)



Data Transformation Phase

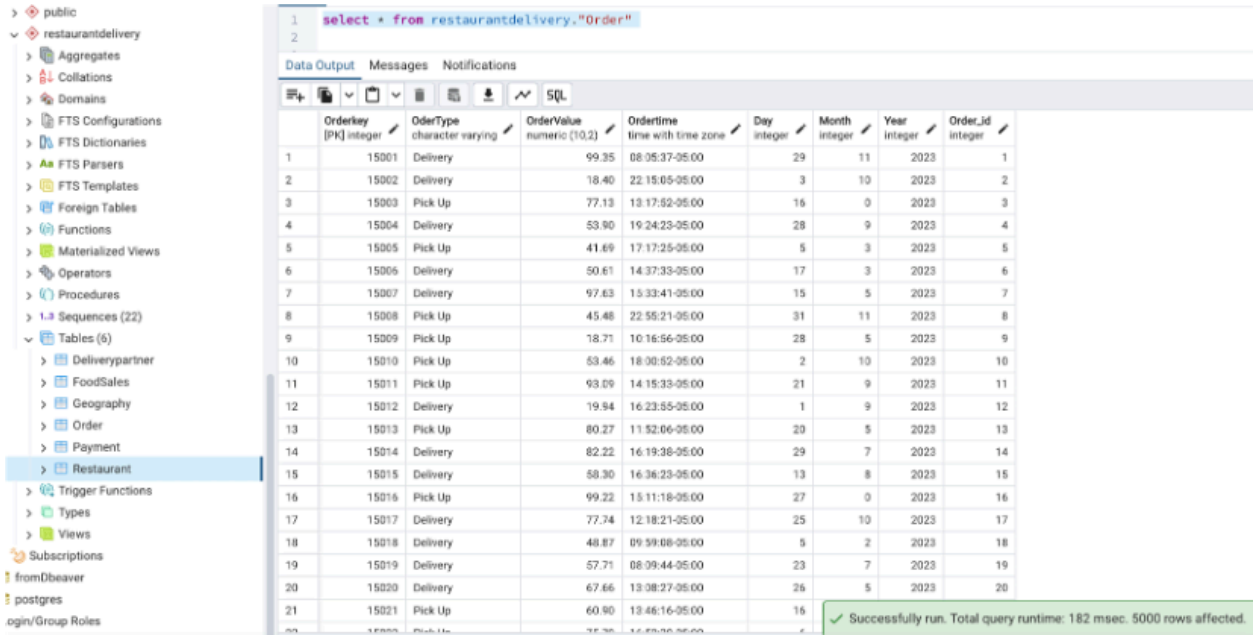
Transformate order_date column in OLTP to extract day number, month and year in Order table in the data warehouse.



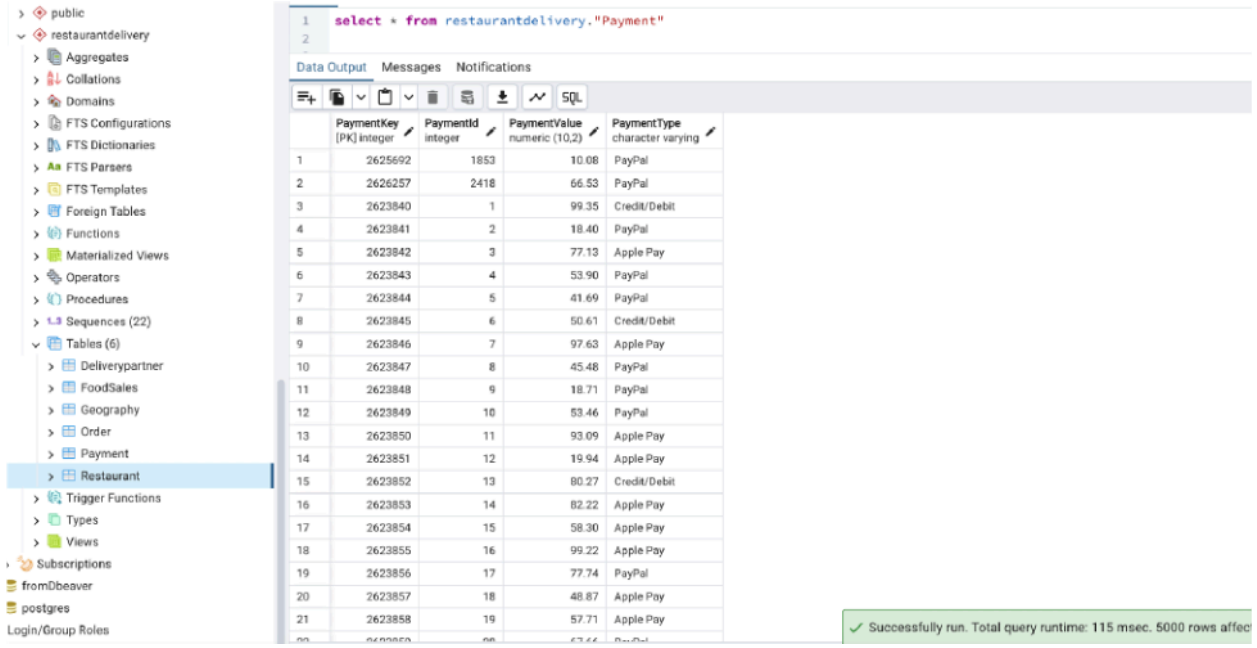
Data Analysis Phase

- Executes SQL queries on the transformed data

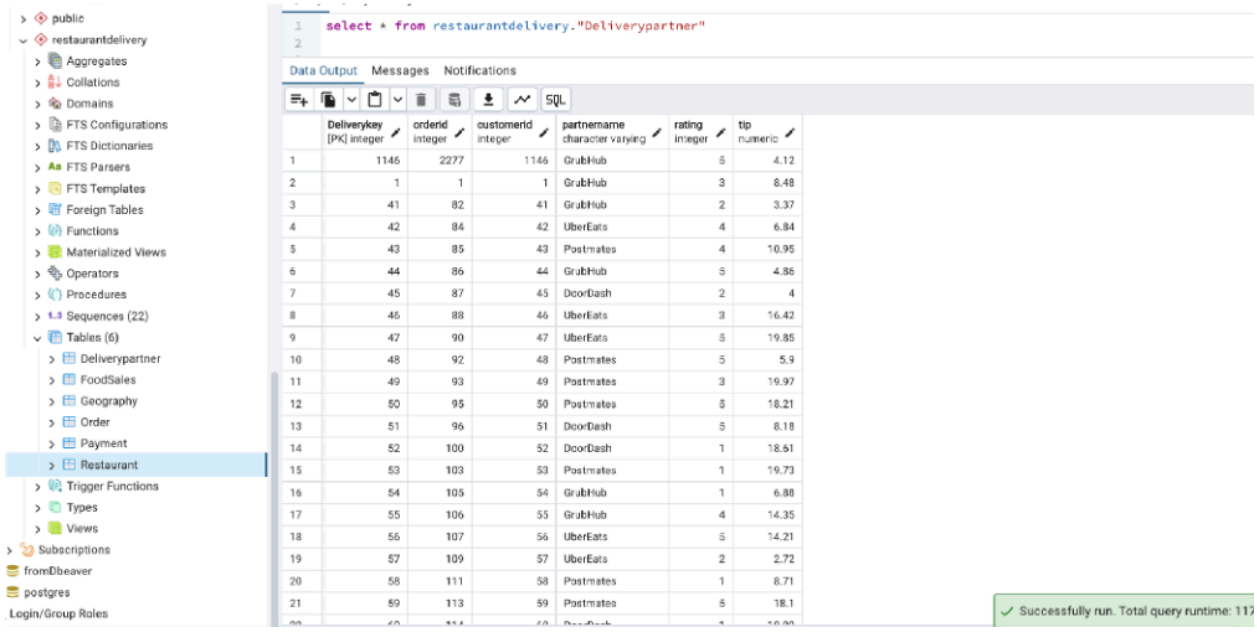
1) Order



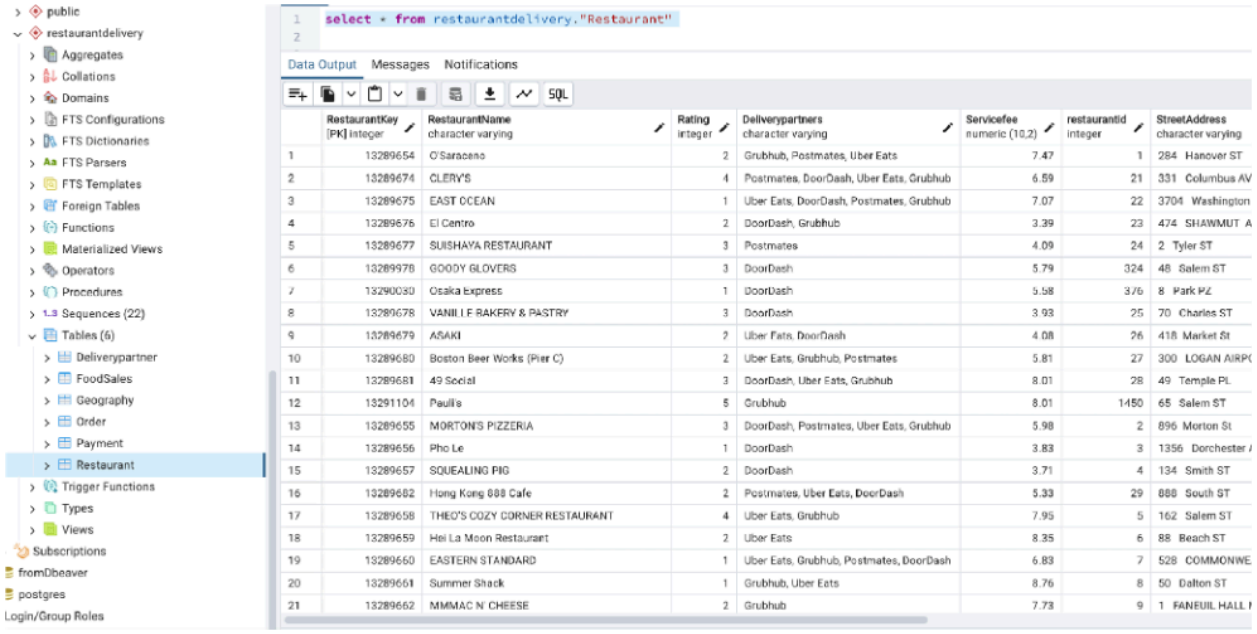
2) Payment



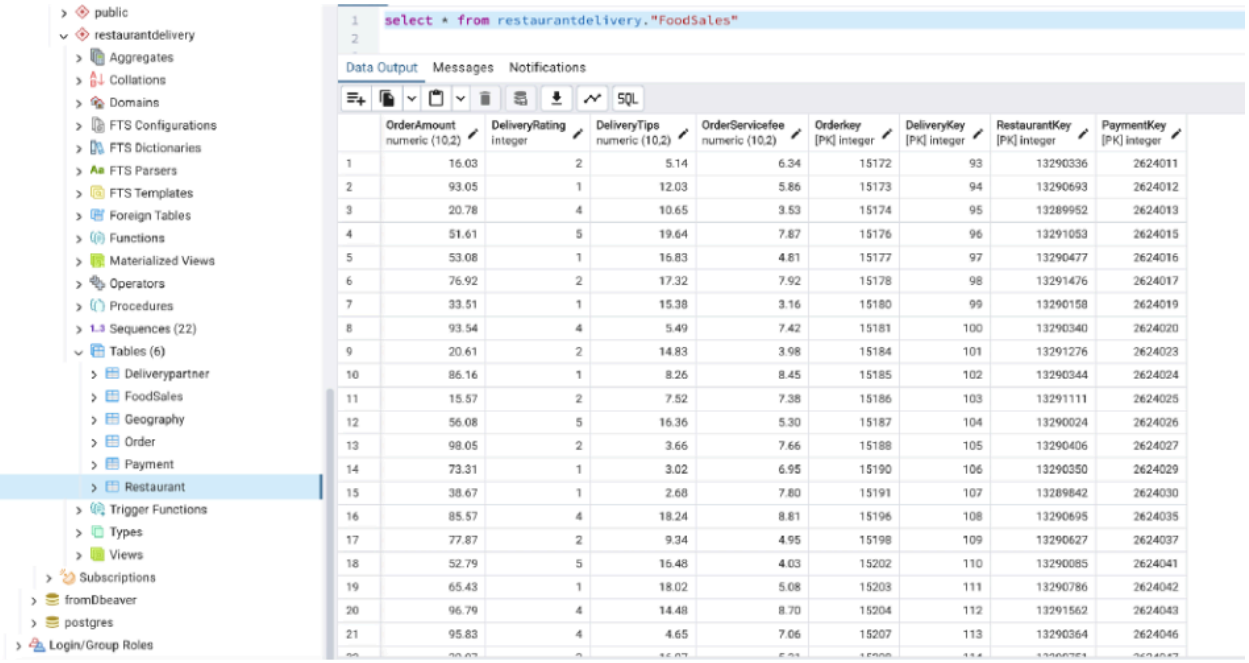
3) Delivery Partner



4) Restaurant



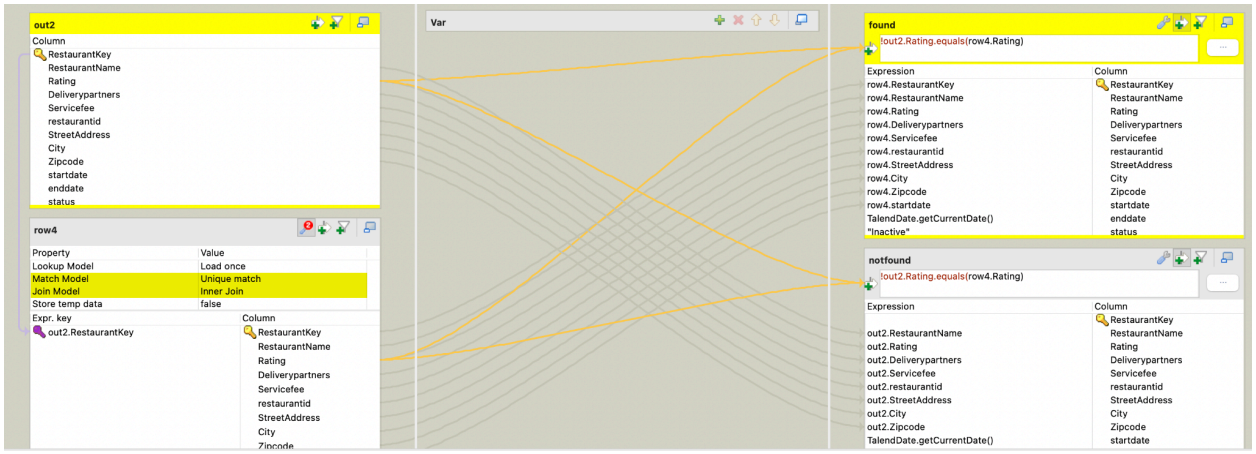
5) Food Sales



	OrderAmount numeric (10,2)	DeliveryRating integer	DeliveryTips numeric (10,2)	OrderServicefee numeric (10,2)	Orderkey [PK] integer	DeliveryKey [PK] integer	RestaurantKey [PK] integer	PaymentKey [PK] integer
1	16.03	2	5.14	6.34	15172	93	13290336	2624011
2	93.05	1	12.03	5.86	15173	94	13290693	2624012
3	20.78	4	10.65	3.53	15174	95	13289952	2624013
4	51.61	5	19.64	7.87	15176	96	13291053	2624015
5	53.08	1	16.83	4.81	15177	97	13290477	2624016
6	76.92	2	17.32	7.92	15178	98	13291476	2624017
7	33.51	1	15.38	3.16	15180	99	13290158	2624019
8	93.54	4	5.49	7.42	15181	100	13290340	2624020
9	20.61	2	14.83	3.98	15184	101	13291276	2624023
10	86.16	1	8.26	8.45	15185	102	13290344	2624024
11	15.57	2	7.52	7.38	15186	103	13291111	2624025
12	56.08	5	16.36	5.30	15187	104	13290024	2624026
13	98.05	2	3.66	7.66	15188	105	13290406	2624027
14	73.31	1	3.02	6.95	15190	106	13290350	2624029
15	38.67	1	2.68	7.80	15191	107	13289842	2624030
16	85.57	4	18.24	8.81	15196	108	13290695	2624035
17	77.87	2	9.34	4.95	15198	109	13290627	2624037
18	52.79	5	16.48	4.03	15202	110	13290085	2624041
19	65.43	1	18.02	5.08	15203	111	13290786	2624042
20	96.79	4	14.48	8.70	15204	112	13291562	2624043
21	95.83	4	4.65	7.06	15207	113	13290364	2624046

Slowing Changing Dimension

Implemented SCD Type 2 for restaurant rating in the restaurant dimension



For example, If the restaurant rating of restaurant named “riverway cafe” changed twice. It was changed from rating 5 to 0, the start date, enddate and status changed. Again the rating changed from 0 to 2, and which is the latest rating, the status remains as ‘Active’.

```
1  SELECT "RestaurantKey", "RestaurantName", "Rating", "startdate", "enddate", "status"
2  FROM "restaurantdelivery"."Restaurant"
3  WHERE "RestaurantName" = 'riverway cafe';
4
```

Data Output | Messages | Notifications

	RestaurantKey [PK] integer	RestaurantName character varying	Rating integer	startdate date	enddate date	status character varying
1	6	riverway cafe	5	[null]	2024-11-15	Inactive
2	2001	riverway cafe	0	2024-11-15	2024-11-15	Inactive
3	2002	riverway cafe	2	2024-11-15	[null]	Active

Analysis Using Power BI

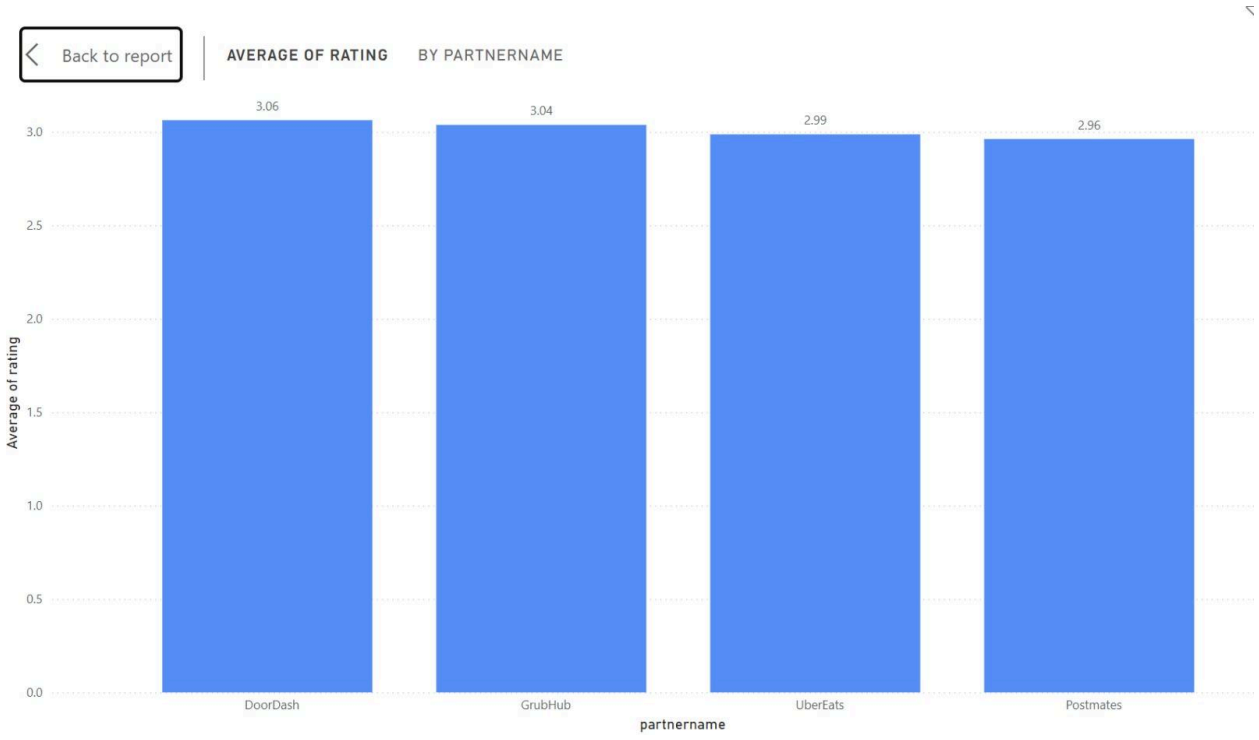
1) Order value and service fee correlation

Scatter Plot (Service Fee vs Net Order Amount)

- Shows a positive correlation between order amount and service fees, with most service fees ranging between 4-8 dollars

- Higher order amounts tend to have slightly higher service fees, though the relationship isn't strongly linear
- The majority of orders fall between $20-$80 in net amount
- Service fees appear to have some standard tiers, as shown by the horizontal clustering of points

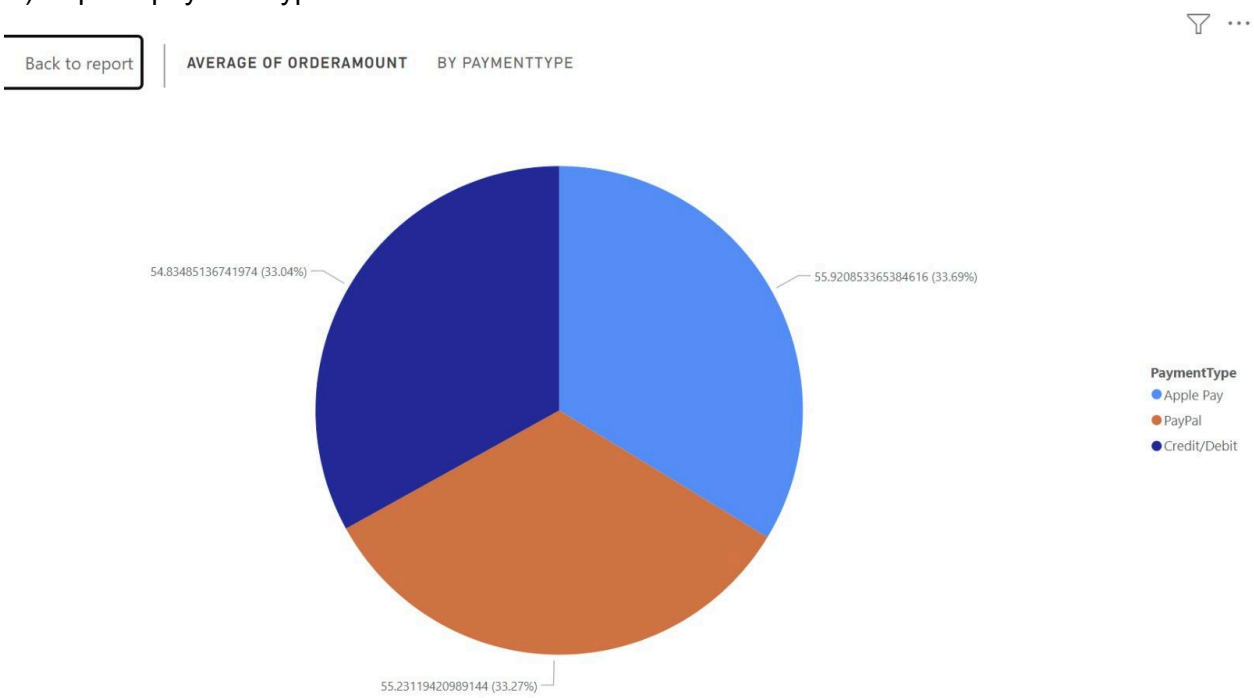
2) Delivery partner performance



Delivery Partner Ratings

- DoorDash leads with the highest average rating of 3.06, followed closely by GrubHub at 3.04
- UberEats and Postmates trail slightly with ratings of 2.99 and 2.96 respectively
- The difference in ratings between all partners is relatively small (within 0.1 points)
- All delivery partners maintain average ratings above 2.9, indicating generally satisfactory service

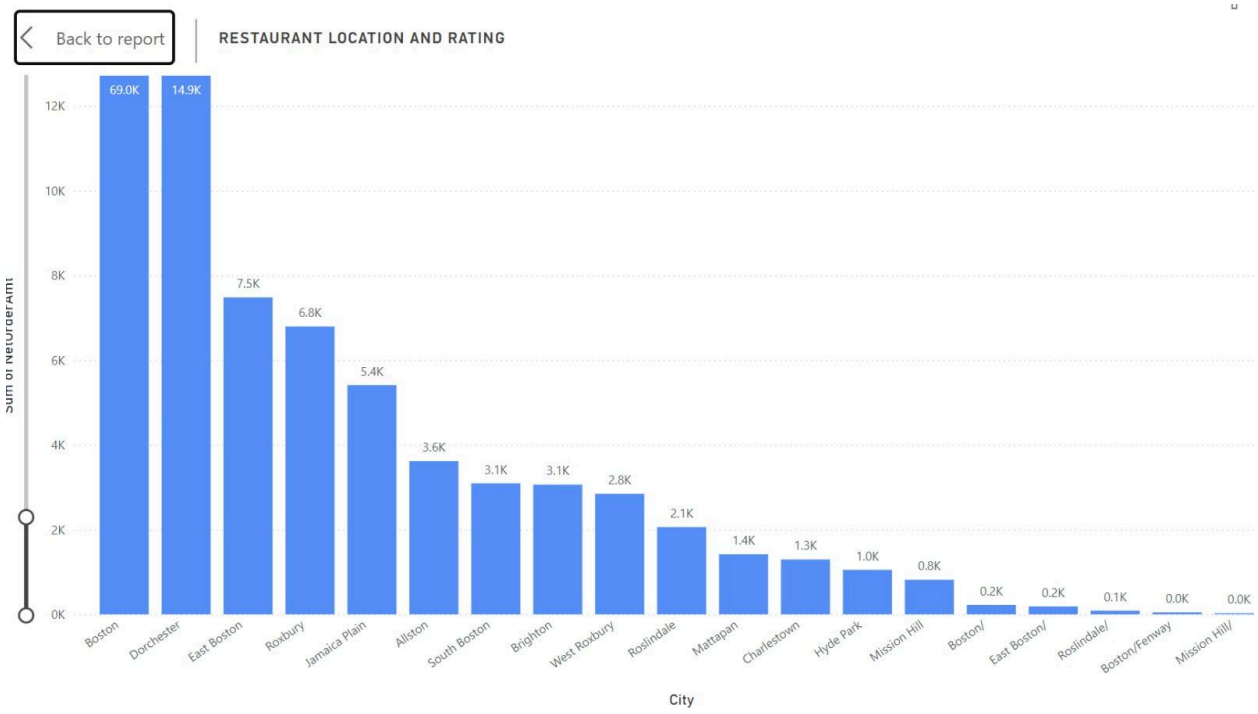
3) Popular payment types



Payment Type Distribution

- Credit/Debit cards dominate with approximately 55% of all transactions
- PayPal accounts for about 33% of payments
- Apple Pay represents roughly 12% of transactions
- Digital payment methods collectively handle 100% of orders, indicating full digitization

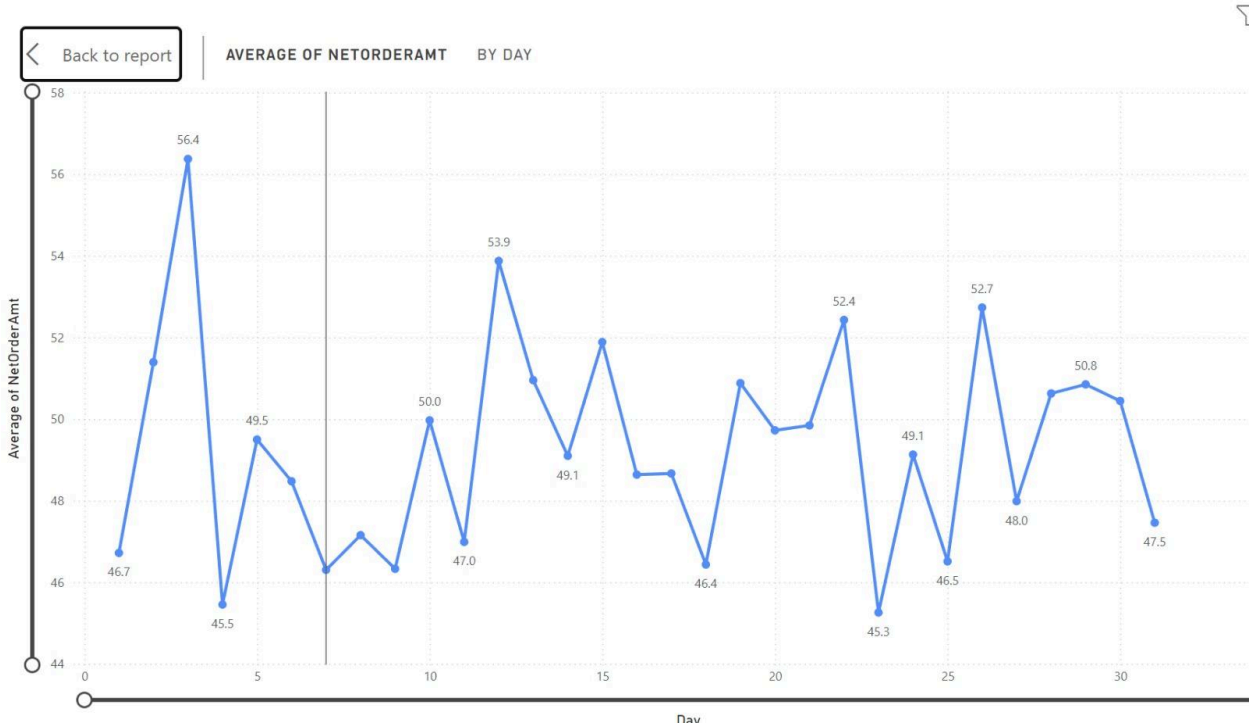
4) Top restaurants based on ratings



Restaurant Location and Net Order Amount

- Boston leads in total net order value at approximately $10K
- Dorchester follows closely with similar order values
- There's a significant drop in order values for locations after the top 5
- The distribution shows a clear concentration of business in core urban areas

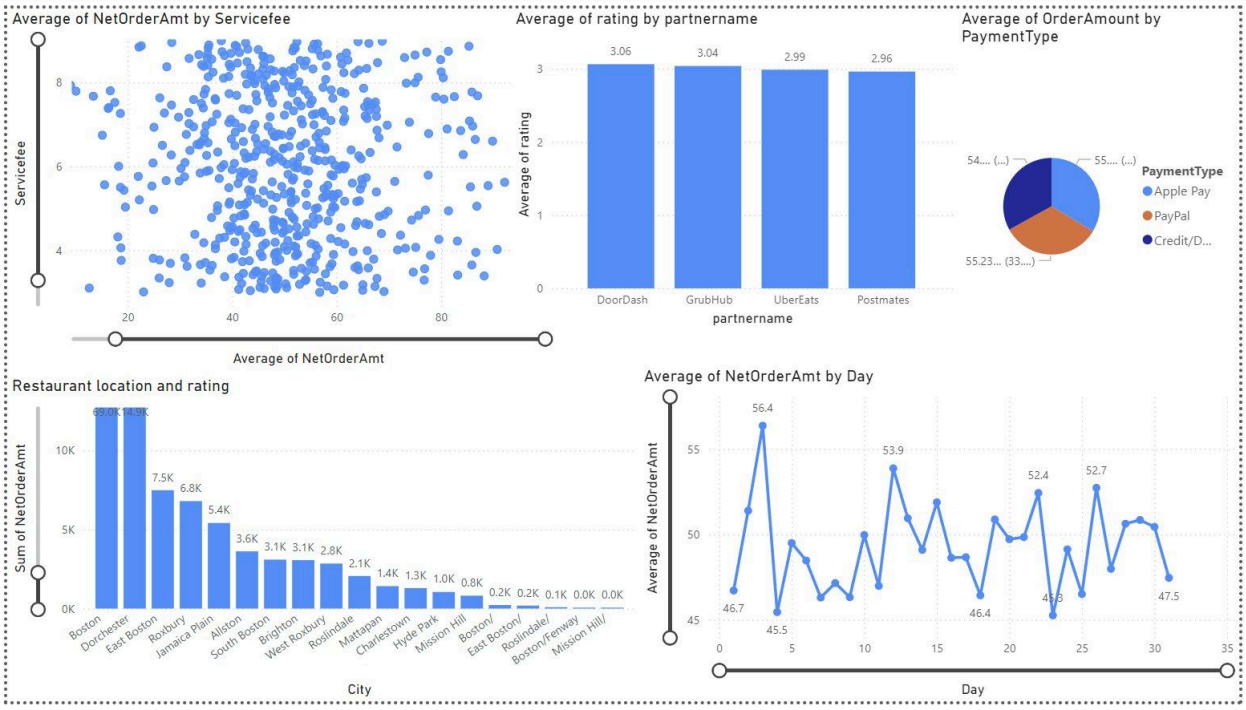
5) Order Frequency by Day of Week



Average Net Order Amount by Day

- Shows significant daily fluctuations in average order values
- Peak average order value reaches approximately $56.4
- Lowest points hover around $45.5
- The trend line suggests slight variability but maintains a relatively stable pattern throughout the month

Dashboard(Power BI)



Conclusion

This data warehouse project analyzing food delivery operations reveals several critical insights about customer behavior, delivery partner performance, and order patterns across different regions.

The analysis demonstrates strong market dynamics in the food delivery ecosystem, with DoorDash leading customer satisfaction ratings at 3.06, followed closely by GrubHub at 3.04. The payment method distribution shows a clear preference for traditional credit/debit cards (55%) over digital wallets like PayPal (33%) and Apple Pay (12%), indicating room for digital payment adoption growth. The correlation between order amounts and service fees shows a consistent pricing structure, with most service fees ranging between $4-8 regardless of order size, suggesting a standardized fee model.

The geographical analysis reveals that Boston and Dorchester are the primary revenue-generating locations, with significantly higher net order amounts compared to other areas. The daily order value trends show notable fluctuations, with peak average orders reaching $56.4 and maintaining a relatively stable baseline around $45.5, indicating consistent consumer demand with predictable peak periods. The service fee structure and its relationship with net order amounts provides valuable insights for optimizing pricing strategies and improving operational efficiency across different market segments.

References

1. "Food Establishment Inspections in the City of Boston" by @davidtalby.
2. ChatGPT
3. https://mockaroo.com/