Data Warehouse Design

Sales Domain

"Coffee Shop"

Hijir Della Wirasti

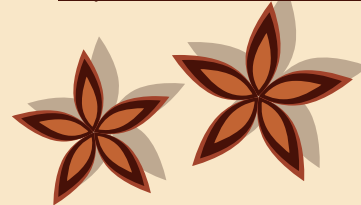17 November 2024

https://github.com/hijirdella/Data-Warehouse-Design.git

https://www.linkedin.com/in/hijirdella/

Table of **contents**

1. Objectives

Objective

1. Design a Data Warehouse (DWH) schema for the Sales domain.
2. Create an Entity-Relationship Diagram (ERD) to represent the schema.
3. Use either the Star or Snowflake schema to describe the design.
4. Generate three Data Mart tables for analytical and reporting purposes.

The goal

1. ERD Diagram
2. Star Schema Description
3. Sample Queries (Data Mart Tables)

Whatare we working on

2. Dataset Selection (1)


```sql
1  SELECT * FROM public.dim_product
2  ORDER BY product_id ASC
```

Data Output   Messages   Notifications

	product_id [PK] integer	product_name character varying (100)	category character varying (50)
1	1	Caffè Americano	Coffee
2	2	Caramel Macchiato	Coffee
3	3	Iced Green Tea Latte	Tea
4	4	Java Chip Frappuccino	Coffee
5	5	Croissant	Bakery
6	6	Pumpkin Spice Latte	Coffee
7	7	Blueberry Muffin	Bakery
8	8	Chai Tea Latte	Tea
9	9	Flat White	Coffee
10	10	Mocha	Coffee

```sql
1  SELECT * FROM public.dim_sales_name
2  ORDER BY sales_name_id ASC
```

Data Output   Messages   Notifications

	sales_name_id [PK] integer	sales_name character varying (100)	sales_age integer	sales_gender character varying (10)
1	1	John Doe	35	Male
2	2	Jane Smith	28	Female
3	3	Michael Johnson	45	Male
4	4	Emily Brown	32	Female
5	5	Daniel Williams	40	Male
6	6	Jessica Davis	26	Female
7	7	David Martinez	38	Male
8	8	Sophia Wilson	29	Female
9	9	James Anderson	41	Male
10	10	Olivia Taylor	34	Female

Selected Dataset:

Sales SQL Dataset

Domain Overview:

The Sales SQL Dataset provides a structured dataset designed for analyzing sales performance, trends, and insights. This dataset includes one fact table and several dimension tables to support business intelligence and reporting tasks.

2. Dataset Selection (2)

```sql
1  SELECT * FROM public.dim_store
2  ORDER BY store_id ASC
```

Data Output | Messages | Notifications

store_id [PK] integer	store_name character varying (100)	city character varying (50)	state character varying (50)	country character varying (50)
1	Starbucks Central	New York	NY	USA
2	Downtown Starbucks	Los Angeles	CA	USA
3	Starbucks Corner	Chicago	IL	USA

```sql
1  SELECT * FROM public.dim_time
2  ORDER BY time_id ASC
```

Data Output | Messages | Notifications

time_id [PK] integer	date date	day_of_week character varying (10)	month character varying (20)	year integer
1	2024-01-01	Monday	January	2024
2	2024-01-02	Tuesday	January	2024
3	2024-01-03	Wednesday	January	2024
4	2024-01-04	Thursday	January	2024
5	2024-01-05	Friday	January	2024
6	2024-01-06	Saturday	January	2024
7	2024-01-07	Sunday	January	2024
8	2024-01-08	Monday	January	2024
9	2024-01-09	Tuesday	January	2024
10	2024-01-10	Wednesday	January	2024
11	2024-01-11	Thursday	January	2024
12	2024-01-12	Friday	January	2024
13	2024-01-13	Saturday	January	2024
14	2024-01-14	Sunday	January	2024

Total rows: 31 of 31   Query complete 00:00:00.355

```sql
1  SELECT * FROM public.fact_sales
2
```

Data Output | Messages | Notifications

sale_id integer	store_id integer	sales_name_id integer	time_id integer	product_id integer	quantity integer	price integer
1	13	3	7	10	4	6
2	12	2	31	5	3	9
3	1	2	4	3	2	6
4	2	3	22	7	2	6
5	3	1	30	9	4	5
6	4	1	18	6	5	11
7	5	2	27	6	1	12
8	6	2	22	8	1	15
9	7	2	25	8	5	11
10	8	1	27	2	5	7
11	9	2	5	10	2	5
12	10	1	19	2	3	12
13	11	1	17	10	5	6
14	14	3	26	4	5	11
15	15	1	23	1	2	11

Total rows: 1000 of 1000   Query complete 00:00:01.285

Selected Dataset:

Sales SQL Dataset

Domain Overview:

The Sales SQL Dataset provides a structured dataset designed for analyzing sales performance, trends, and insights. This dataset includes one fact table and several dimension tables to support business intelligence and reporting tasks.

3. ERD Diagram (1)

**ERD Diagram Explanation - Sales Star Schema**

The diagram represents a **Star Schema** designed for a **Sales Data Warehouse**, structured to optimize analytical queries. It consists of one **Fact Table** (fact_sales) and four **Dimension Tables** (dim_product, dim_store, dim_time, dim_sales_name).
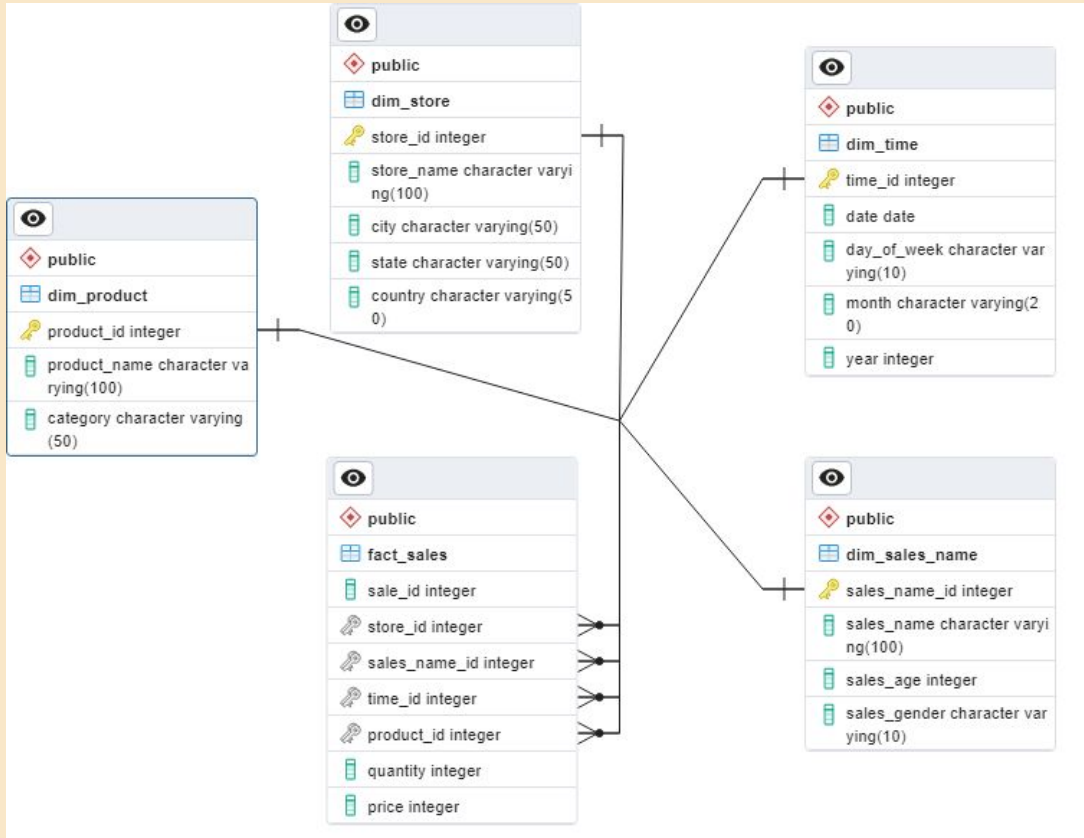
3. ERD Diagram (2)

1. Fact Table: `fact_sales`

- **Purpose:** Captures sales transactions and serves as the central table connecting all dimensions.
- **Columns:**
- `sale_id` (PK): Unique identifier for each transaction.
- Foreign Keys:
  - `store_id`: Links to the store of the sale.
  - `sales_name_id`: Links to the salesperson.
  - `time_id`: Links to the date of the sale.
  - `product_id`: Links to the product sold.
- Attributes:
  - `quantity`: Number of items sold.
  - `price`: Price per unit.
- **Use Case:** Enables aggregation and detailed analysis of sales data, such as total revenue and quantity sold.
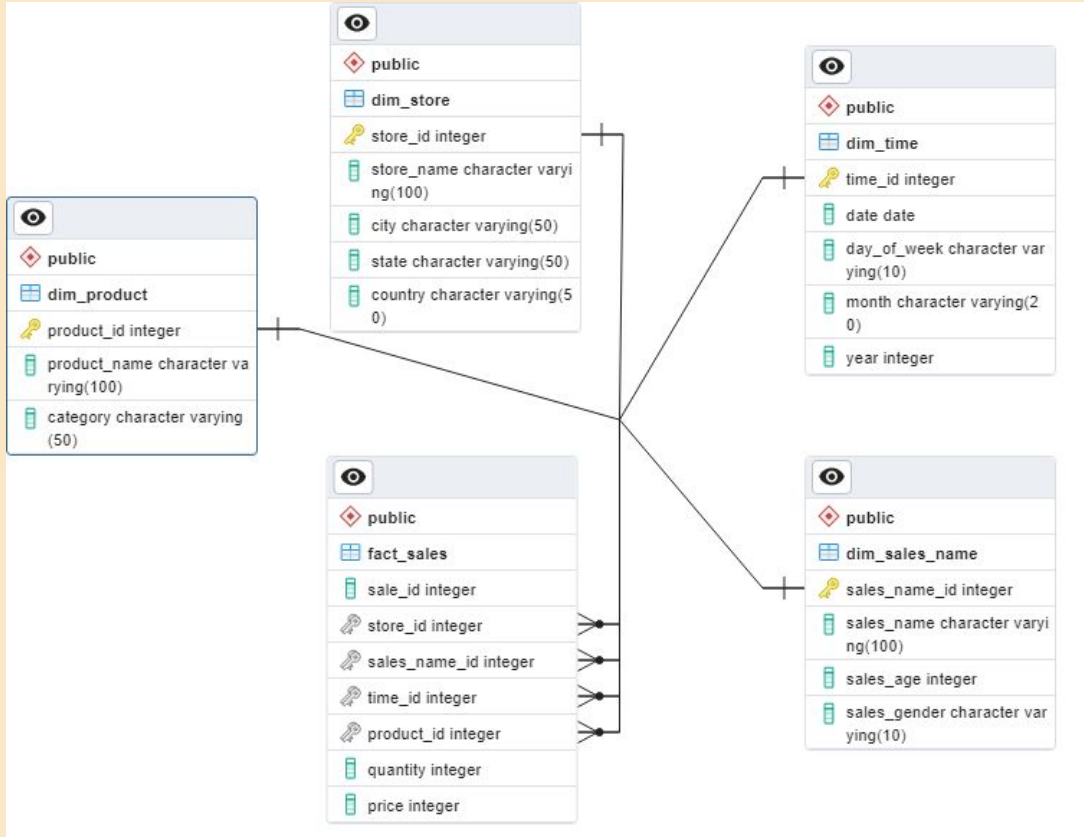
2. Dimension Tables

These provide descriptive attributes, allowing deeper analysis of sales data:

1. **dim_product** (Product Details):
    - **Columns:**
        - **product_id** (PK), **product_name**, **category**.
    - **Purpose:** Analyze sales by product and category (e.g., most sold products).
2. **dim_store** (Store Information):
    - **Columns:**
        - **store_id** (PK), **store_name**, **city**, **state**, **country**.
    - **Purpose:** Compare sales performance across regions (city, state, or country).
3. **dim_time** (Time Details):
    - **Columns:**
        - **time_id** (PK), **date**, **day_of_week**, **month**, **year**.
    - **Purpose:** Identify sales trends over time (daily, monthly, yearly).
4. **dim_sales_name** (Salesperson Information):
    - **Columns:**
        - **sales_name_id** (PK), **sales_name**, **sales_age**, **sales_gender**.
    - **Purpose:** Evaluate sales performance by individual sales staff.
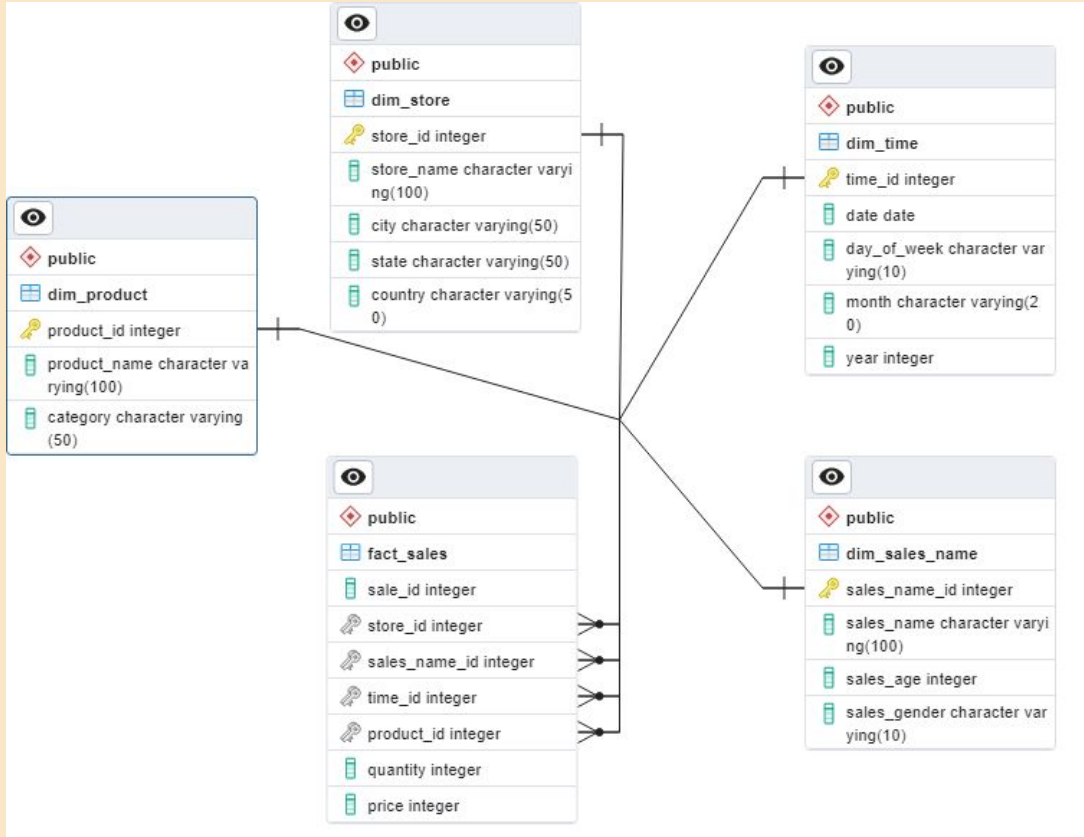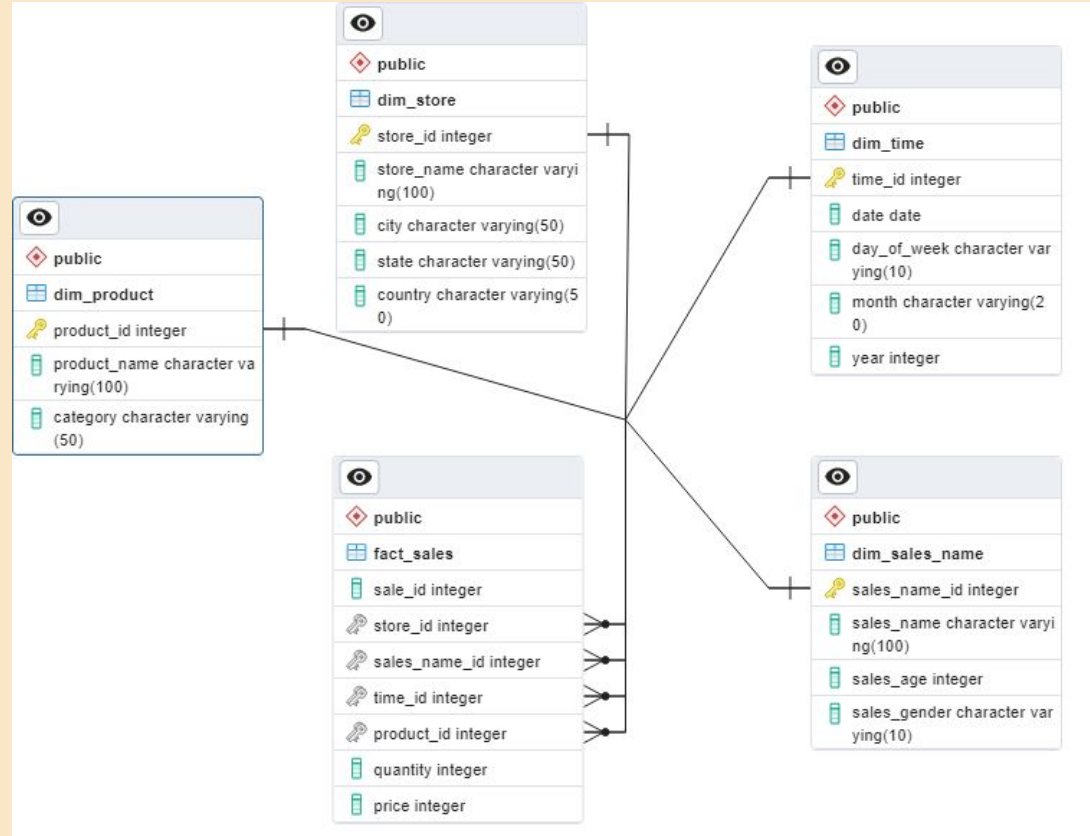
4. Schema Description

Schema Design Highlights

- **Star Schema Structure:** Simple and optimized for querying, with denormalized dimensions for faster performance.
- **Relationships:**
  - Each foreign key in fact_sales links to the primary key in a dimension table.
- **Scalability:** Designed for high performance in analytical workloads by minimizing joins.
- **Integrity:** Primary and foreign key constraints ensure consistent relationships.

Analytical Capabilities

- **Product Trends:** Analyze best-performing products and categories.
- **Regional Sales:** Evaluate sales by location (e.g., country, city).
- **Time-Based Insights:** Track sales patterns by day, week, month, or year.
- **Sales Team Analysis:** Assess individual and group performance of sales staff.

This schema is ideal for supporting comprehensive business intelligence and decision-making.

public

dim_store
- store_id integer
- store_name character varying(100)
- city character varying(50)
- state character varying(50)
- country character varying(50)

public

dim_time
- time_id integer
- date date
- day_of_week character varying(10)
- month character varying(20)
- year integer

public

dim_product
- product_id integer
- product_name character varying(100)
- category character varying(50)

public

fact_sales
- sale_id integer
- store_id integer
- sales_name_id integer
- time_id integer
- product_id integer
- quantity integer
- price integer

public

dim_sales_name
- sales_name_id integer
- sales_name character varying(100)
- sales_age integer
- sales_gender character varying(10)

5. Data Mart

a. Top Stores by Revenue

```sql
--a. Top Stores by Revenue
SELECT *
FROM data_mart.dm_store_performance
ORDER BY total_revenue DESC;
```

Data Output   Messages   Notifications

	store_id integer	store_name character varying (100)	city character varying (50)	state character varying (50)	country character varying (50)	total_sales_transactions bigint	total_quantity_sold bigint	total_revenue bigint	avg_price_per_unit numeric
1	3	Starbucks Corner	Chicago	IL	USA	346	1080	10474	9.7803468208092
2	2	Downtown Starbucks	Los Angeles	CA	USA	325	969	9767	10.0584615384615
3	1	Starbucks Central	New York	NY	USA	329	952	9413	9.8662613981762

Key Takeaways:

1. **Starbucks Corner (Chicago)** is the highest-performing store in terms of both total revenue and quantity sold.
2. Pricing strategies at **Downtown Starbucks (Los Angeles)** are likely more effective for maximizing revenue per transaction.
3. Performance between the top stores is relatively close, suggesting opportunities to replicate high-performing strategies across locations.

Recommendations:

1. **Evaluate Product Mix and Pricing:** Investigate why the average price per unit is higher at Downtown Starbucks. This could inform strategies for increasing per-unit revenue across other locations.
2. **Analyze Customer Behavior:** Look into customer demographics and preferences at each store to understand the drivers behind higher transaction volume and revenue.
3. **Expand High-Performing Strategies:** Replicate successful strategies from Starbucks Corner in other locations to boost overall sales volume and revenue.

5. Data Mart

b. Best-Selling Products

```
103  --b. Best-Selling Products
104  SELECT *
105  FROM data_mart.dm_product_sales
106  ORDER BY total_quantity_sold DESC;
```

Data Output   Messages   Notifications

	product_id integer	product_name character varying (100)	category character varying (50)	total_sales_transactions bigint	total_quantity_sold bigint	total_revenue bigint	avg_price_per_unit numeric
1	3	Iced Green Tea Latte	Tea	106	332	3353	10.2358490566037736
2	5	Croissant	Bakery	111	316	3159	10.0270270270270270
3	9	Flat White	Coffee	103	315	3181	9.9417475728155340
4	7	Blueberry Muffin	Bakery	104	314	3017	9.6442307692307692
5	6	Pumpkin Spice Latte	Coffee	103	308	3058	9.8155339805825243
6	2	Caramel Macchiato	Coffee	101	296	2944	9.8514851485148515
7	4	Java Chip Frappuccino	Coffee	98	294	2869	9.7959183673469388
8	1	Caffè Americano	Coffee	90	289	2926	10.3666666666666667
9	10	Mocha	Coffee	92	271	2579	9.4565217391304348
10	8	Chai Tea Latte	Tea	92	266	2568	9.8369565217391304

Key Takeaways:

1. **Iced Green Tea Latte** and **Croissant** are the best-performing products in their respective categories.
2. **Coffee** consistently drives the highest sales across multiple products, making it a key category for revenue generation.
3. Products with a slightly higher price per unit, such as **Caffè Americano** and **Flat White**, balance lower quantities with high revenue efficiency.

Recommendations:

1. **Focus on Coffee Category:**
   - Invest in marketing or promotions for high-performing coffee items (e.g., Flat White, Caramel Macchiato).
   - Experiment with introducing similar premium-priced coffee products.

2. **Leverage Popular Tea and Bakery Products:**
   - Expand options in the **Tea** and **Bakery** categories, leveraging the popularity of items like **Iced Green Tea Latte** and **Croissant**.
   - Offer bundles that combine Tea and Bakery products to boost cross-category sales.

3. **Optimize Pricing Strategies:**
   - Explore minor price adjustments for high-demand items like **Iced Green Tea Latte** to maximize revenue.
   - Maintain premium pricing for items like **Caffè Americano** to preserve revenue efficiency.

4. **Track Seasonal Trends:**
   - Monitor sales for seasonal items like **Pumpkin Spice Latte** and adjust supply and marketing accordingly.

5. Data **Mart**

c. Best Sales Person

```sql
108  --c. Best Sales Person
109  SELECT *
110  FROM data_mart.dm_salesperson_performance
111  ORDER BY total_revenue DESC;
```

Data Output   Messages   Notifications

	sales_name_id integer	sales_name character varying (100)	total_sales_transactions bigint	total_quantity_sold bigint	total_revenue bigint	avg_price_per_unit numeric
1	1	John Doe	120	367	3739	10.3583333333333333
2	2	Jane Smith	107	322	3389	10.5327102803738318
3	9	James Anderson	110	331	3230	9.7636363636363636
4	4	Emily Brown	106	310	3019	9.6509433962264151
5	5	Daniel Williams	101	299	2912	9.4554455445544554
6	6	Jessica Davis	100	302	2844	9.7600000000000000
7	3	Michael Johnson	93	278	2786	10.0860215053763441
8	7	David Martinez	85	280	2720	9.3647058823529412
9	8	Sophia Wilson	92	280	2698	9.7391304347826087
10	10	Olivia Taylor	86	232	2317	10.1279069767441860

Recommendations:

1. **Leverage Best Practices from John Doe and Jane Smith:**
   - Train other salespersons to balance volume with pricing strategies.
   - Focus on identifying best-selling techniques or product bundles.
2. **Encourage Upselling for Low-Priced Sellers:**
   - For salespersons like David Martinez and Sophia Wilson, encourage upselling premium products to improve their average price per unit.
3. **Expand Recognition Programs:**
   - Reward top performers like John Doe and Jane Smith to incentivize consistent performance across the team.
4. **Analyze Customer Segments by Salesperson:**
   - Determine if performance differences are tied to specific customer demographics or product preferences.

Key Takeaways:

1. **John Doe** is the top-performing salesperson across all metrics, making him a key contributor to sales.
2. **Jane Smith** excels in maximizing revenue through high average pricing, indicating potential strength in premium sales strategies.
3. **James Anderson's strategy** relies on high-volume sales, though slightly lower pricing may indicate room for upselling.
4. **Emily Brown** and others in the mid-tier (Jessica Davis, Michael Johnson) maintain consistency, showing balanced performance across metrics.

5. Data **Mart**

d. Sales Trends Over Time

```
113  --d. Sales Trends Over Time
114  SELECT *
115  FROM data_mart.dm_sales_trends_by_day_of_week
116  ORDER BY total_revenue DESC;
```

Data Output   Messages   Notifications

	sales_day_of_week character varying (10)	total_sales_transactions bigint	total_quantity_sold bigint	total_revenue bigint	avg_price_per_unit numeric
1	Monday	169	490	5098	10.3964497041420118
2	Tuesday	169	523	5096	9.9704142011834320
3	Wednesday	175	514	4958	9.7485714285714286
4	Friday	137	418	4213	9.8467153284671533
5	Saturday	128	388	3770	9.7734375000000000
6	Thursday	119	365	3371	9.4285714285714286
7	Sunday	103	303	3148	9.9902912621359223

Key Takeaways:

1. **Monday and Tuesday are the strongest sales days**, likely due to higher foot traffic or marketing strategies targeting the start of the week.
2. **Sunday sees the lowest revenue and quantity sold**, suggesting reduced demand or store closures on this day.
3. **Wednesday shows strong volume sales** but slightly lower average pricing, indicating a focus on discounts or promotions.

Recommendations:

1. **Enhance Monday and Tuesday Campaigns:**
   - Leverage their strong performance to run targeted promotions or product launches.
   - Focus on upselling premium products to maximize the already high average price per unit.
2. **Boost Weekend Sales:**
   - Introduce weekend-exclusive discounts or promotions to attract more customers.
   - Consider adding bundled offers to increase transaction volume and revenue.
3. **Analyze Wednesday Sales Patterns:**
   - Investigate if promotions or discounts drive the higher quantities sold on Wednesday.
   - Adjust pricing strategies to balance quantity and revenue.
4. **Focus on Customer Behavior:**
   - Examine customer preferences by day to tailor marketing efforts.
   - Introduce loyalty programs to encourage repeat purchases, especially on slower days like Sunday.

6. Conclusion

1. **Schema Design and ERD:**
   - The **Star Schema** provides a scalable and efficient structure for querying sales data.
   - The **Fact Table (fact_sales)** links to four **Dimension Tables**: dim_product, dim_store, dim_time, and dim_sales_name.
   - The schema ensures data integrity with clearly defined primary and foreign key relationships.
2. **Insights from Data Marts:**
   - **Store Performance Data Mart:** Highlights top-performing stores by revenue, with Starbucks Corner leading due to high sales volume and efficient pricing.
   - **Product Sales Data Mart:** Identifies best-selling products like Iced Green Tea Latte, showing category dominance by Coffee and Bakery items.
   - **Salesperson Performance Data Mart:** Demonstrates balanced performance across sales staff, with John Doe excelling in both quantity sold and revenue generation.
   - **Sales Trends by Day of Week Data Mart:** Reveals Monday and Tuesday as peak revenue days, while weekends underperform, suggesting actionable opportunities.
3. **Effectiveness of Design:**
   - The schema and data marts enable deep analysis across multiple dimensions (time, location, product, and sales staff).
   - Insights derived from these data marts support strategic decision-making for pricing, promotions, and operational focus.

Key Takeaway:

The well-structured schema and data marts deliver actionable business intelligence, offering a clear path to optimize sales performance and resource allocation.