

Use case: Sales Data Processing for Retail Optimization

Problem Statement: XYZ Retail Inc., a prominent player in the retail industry, faces the challenge of efficiently managing and analysing its sales data. The company operates through multiple channels like online platforms, physical stores, and mobile apps. To enhance decision-making, XYZ Retail aims to create an end-to-end data pipeline that ingests, processes, and visualizes sales data.

Objective: Optimize retail operations by designing scalable data pipelines, integrating and cleaning diverse data sources, and enabling advanced analytics through robust storage, transformation, and governance frameworks.

SHREYANS JAIN

Christ reg. no. : 2347155

[Github](#)

Sales Data Analysis Workflow

1. Data Ingestion and Storage

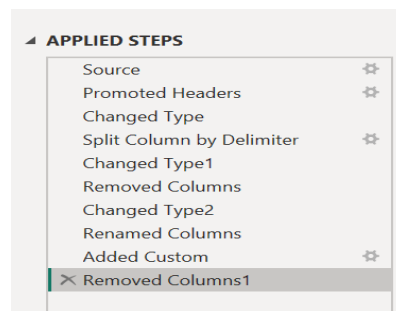
- Created two folders in Azure Storage within the **Sales Containers**:
 - **Raw Data**: For storing unprocessed datasets.
 - **Transformed Data**: For storing cleaned and processed datasets.
- Ingested raw sales data using the **GitHub raw file link**, ensuring a streamlined and scalable data input pipeline.

2. Data Transformation Steps

- **Databricks Transformation:**
 - Loaded raw data into Databricks for initial cleaning and processing:
 - Handled missing values, duplicates, and outliers.
 - Standardized date formats and ensured consistent column naming conventions.
 - **SQL Transformation in Synapse:**
 - Imported transformed data from Databricks into Azure Synapse.
 - Performed further transformations using SQL queries:
 - Derived key metrics like **Total Sales** and **Average Order Value (AOV)**.
 - Conducted channel-wise performance analysis for **online, in-store, and mobile app sales**.
-

3. Power BI Visualization

- Exported the final transformed dataset from Azure Synapse to Power BI.
- Conducted additional data transformations in Power BI to prepare the data for visualization.
- Designed an interactive dashboard with the following features:
 - **Slicers** for filtering by sales channels, regions, and product categories.
 - **Key Performance Indicators (KPIs)** displayed prominently:
 - Total Sales
 - Average Order Value (AOV)
 - Customer Retention Rate
 - **Trend Analysis:** Monthly, weekly, and daily sales patterns were highlighted, identifying peak sales periods and seasonal trends.



4. Actionable Insights

- **Operational Optimization:** Suggested reallocating resources to high-performing channels.
- **Inventory Management:** Recommended stocking up on best-sellers during peak periods.
- **Customer Retention:** Advocated targeted offers and loyalty programs for high-value customers.

Conclusion

The solution leverages Azure services, Databricks, Synapse, and Power BI to create a robust, end-to-end data processing and visualization pipeline. This setup enables real-time, data-driven decision-making by providing insights into customer behavior, product performance, and operational efficiency.

Outputs

Dashboard

The screenshot shows the Microsoft Azure dashboard. At the top, there's a navigation bar with 'Microsoft Azure', an 'Upgrade' button, a search bar, and a 'Copilot' button. Below this, the 'Azure services' section displays icons for various services: Create a resource, Azure Synapse Analytics, Key vaults, App registrations, Data factories, Storage accounts, Logic apps, Quickstart Center, Azure AI services, and More services. The 'Resources' section has tabs for 'Recent' and 'Favorite'. A table lists recent resources:

Name	Type	Last Viewed
synapse-shrey-sales	Synapse workspace	an hour ago
storagesalesdashshrey	Storage account	2 hours ago
storage-account-neostat	Resource group	2 hours ago
sales-kv-shrey	Key vault	2 hours ago

A 'See all' link is available at the bottom of the resources list.

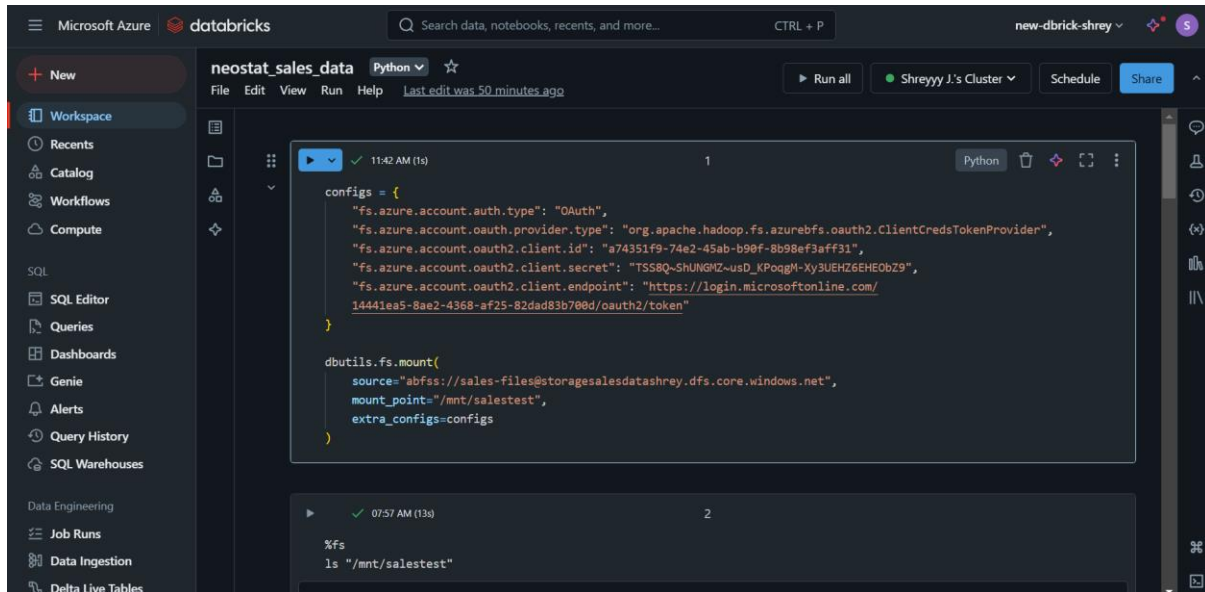
Synapse

The screenshot shows the Microsoft Azure Synapse Analytics interface. The left sidebar contains navigation options: Home, Data, Develop, Integrate, Monitor, and Manage. The main area is divided into 'Data' and 'Workspace' sections. The 'Data' section shows a tree view of resources: Lake database, salesDataNeostat, Tables, sales, and Columns. The 'Columns' section lists the following data types: OrderID (string), CustomerName (string), Location (string), Country (string), StoreCode (string), Product (string), Quantity (long), Price (double), and Date (date). The 'Workspace' section displays a SQL script in a text editor. The script is as follows:

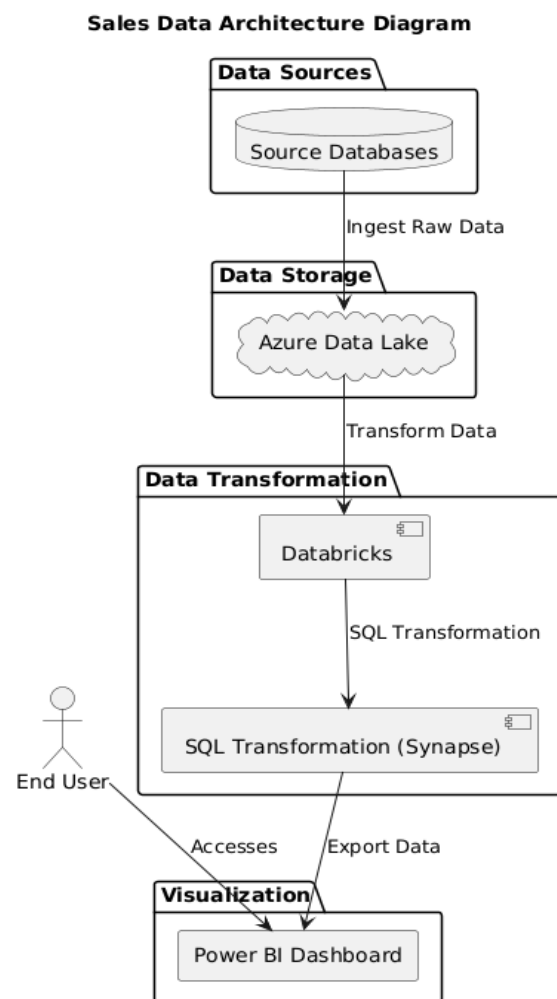
```
1 -- Querying the existing 'sales' table
2 SELECT
3     CustomerName,
4     Location,
5     Country,
6     Product,
7     SUM(Quantity) AS TotalQuantity,
8     AVG(Price) AS AvgPrice,
9     SUM(TotalPrice) AS TotalSales,
10    Year,
11    Month
12 FROM
13     sales
14 GROUP BY
15     CustomerName, Location, Country, Product, Year, Month
16 ORDER BY
17     TotalSales DESC, Year, Month;
18
```

Below the script, there are tabs for 'Results' and 'Messages'. The 'Results' tab shows a status message: '00:00:06 Query executed successfully.'

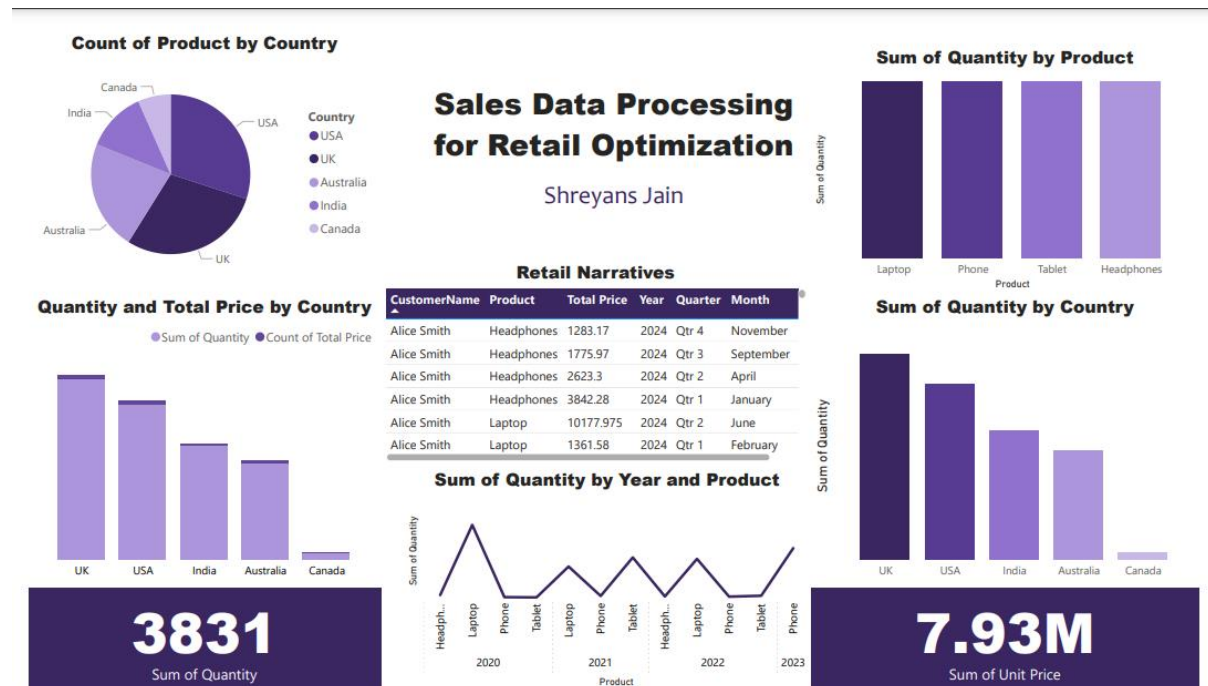
Databricks



Architecture Diagram



PowerBI Dashboard



1. Count of Product by Country (Top Left – Pie Chart)

- **Purpose:** Displays the distribution of products sold across different countries.
- **Insight:**
 - The largest share of product sales is attributed to specific countries, with noticeable variations in contributions from regions like the USA, UK, India, Australia, and Canada.
 - This helps identify the top-performing regions for product distribution.

2. Quantity and Total Price by Country (Bottom Left – Bar Chart)

- **Purpose:** Compares the sum of quantities sold and total price by country.
- **Insight:**
 - The **UK** leads in both quantity and total price, followed by the USA.
 - Canada has significantly lower contributions compared to other countries.
 - This can guide strategies for inventory allocation and pricing.

3. Retail Narratives (Center – Table)

- **Purpose:** Provides detailed transactional data for better granularity.
- **Columns:**

- **Customer Name:** Tracks sales per customer (e.g., Alice Smith).
- **Product:** Lists specific products purchased.
- **Total Price:** Highlights the revenue per transaction.
- **Year, Quarter, and Month:** Enables trend analysis over time.
- **Insight:** Helps drill down into individual customer behavior and track performance trends by product and time.

4. Sum of Quantity by Product (Top Right – Bar Chart)

- **Purpose:** Analyzes product performance based on the quantities sold.
- **Insight:**
 - Products like **Laptops** and **Phones** have high sales volume, indicating strong demand.
 - The performance of products like **Tablets** and **Headphones** can also be compared for inventory and marketing decisions.

5. Sum of Quantity by Country (Middle Right – Bar Chart)

- **Purpose:** Highlights the total quantities sold in each country.
- **Insight:**
 - The **UK and USA** show dominant sales in terms of quantity, reinforcing the trend seen in other charts.

6. Sum of Quantity by Year and Product (Bottom Center – Line Chart)

- **Purpose:** Tracks yearly trends in quantities sold for each product.
- **Insight:**
 - Product demand varies significantly year over year.
 - Peaks in sales can be observed, which might align with seasonal demand or promotional periods.

7. KPI Cards (Bottom)

- **Sum of Quantity (Left):**
 - **Value:** 3831.
 - Indicates the total number of items sold across all products and regions.
- **Sum of Unit Price (Right):**
 - **Value:** 7.93M.
 - Represents the aggregate revenue generated from sales.

8. Geographical Heatmap (Not Visible in Screenshot)

- If included elsewhere, this may display regional sales distribution, providing spatial insights into sales performance.
-

Conclusion

- The dashboard integrates data transformation and visualization effectively to derive actionable insights.
- Key takeaways include:
 - **Top Performing Regions and Products:** The UK and USA dominate sales, while laptops and phones perform best.
 - **Customer and Seasonal Insights:** Detailed breakdowns in retail narratives and sales trends aid targeted strategies.
- This dashboard is a valuable tool for driving data-driven decisions in sales optimization and resource allocation.

References

Power BI Overview: <https://powerbi.microsoft.com/en-us/>

Power BI Documentation: <https://learn.microsoft.com/en-us/power-bi/>

Azure Synapse Analytics Overview: <https://azure.microsoft.com/en-us/products/synapse-analytics/>

Azure Synapse Analytics Documentation: <https://learn.microsoft.com/en-us/azure/synapse-analytics/>

Connecting Power BI to Azure Synapse: <https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-connect-azure-synapse>

SQL Server Documentation: <https://learn.microsoft.com/en-us/sql/sql-server/>

T-SQL Reference: <https://learn.microsoft.com/en-us/sql/t-sql/>

Data Modeling in Power BI: <https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-modeling>

Power BI Best Practices: <https://learn.microsoft.com/en-us/power-bi/guidance/overview>

Azure Synapse Training: <https://learn.microsoft.com/en-us/training/paths/implement-azure-synapse-analytics/>