Data, Inc.

May 18, 2025

Regional Sales

Product, Store, & Team Analysis

This report outlines the end-to-end process of preparing, analyzing, and visualizing a US Regional Sales dataset.

All data used in this analysis is sourced from the U.S. Regional Sales practice dataset, available at: https://www.kaggle.com/datasets/talhabu/us-regional-sales-data/data.

The report is organized into three main components:

1) **Data Structuring** - The project began with structuring the data in MySQL using a star schema design to support efficient querying and future scalability.

2) **Data Visualization** - Cleaned and transformed data was the imported into Power BI, where a multi-page interactive dashboard was developed to explore key metrics across sales channels, products, and store performance.

3) **Insights and Findings** - Through data modeling and visualization, several actionable insights were uncovered regarding sales trends, product performance, team contributions, and seasonal patterns.

The resulting analysis provides a strong foundation for strategic data-driven decision-making and scalable integration of future datasets.

Regional Sales

Data Analysis

MySQL

To begin cleaning and structuring the data, a new MySQL database was created for the US Regional Sales Dataset. The dataset was imported from a .csv file into a new table containing the original attributes: OrderNumber, Sales Channel, WarehouseCode, Procured Date, Order Date, Ship Date, Delivery Date, CurrencyCode, _SalesTeamID, _CustomerID, _StoreID, _ProductID, Order Quantity, Discount Applied, Unit Cost, and Unit Price.

To streamline the data for sales analysis, only relevant columns were retained, and a star schema was designed. Irrelevant fields like WarehouseCode were excluded. Dimension tables were created for sales type, sales team, store, product, date, and order information. A central fact table was built to reference these dimensions via their respective ID columns. Although the original dataset contained limited detail for each dimension, the schema was designed to allow future expansion and optimized querying.

Each column’s data type was assigned appropriately, and primary/foreign keys were established to define relationships. Data from the original table was then inserted into the new schema, ensuring that ID columns held only distinct values. While creating the Date dimension table, non-standard Order Date formats were found and corrected to comply with SQL standards. A view was created to display distinct dates, simplifying data entry and query readability. Similarly, Unit Cost and Unit Price columns were reformatted to be compatible with the Decimal data type.

Finally, the fact table was populated by joining the cleaned data with the dimension tables. The resulting row count was verified against the original table to ensure completeness. With the schema finalized, the dataset was ready for import into Microsoft Power BI for further transformation and analysis.

Regional Sales

Data Analysis

Power BI

The Regional Sales schema was imported into Power BI using the MySQL Database connector. In Power Query Editor, data types were verified, and table/column names were renamed for clarity and usability. The star schema was then reviewed in Model View to confirm correct relationships, cardinality, and cross-filtering directions between fact and dimension tables.

Three main areas of analysis were identified: sales, products, and stores. Accordingly, a three-page dashboard was created, each page featuring five key metric cards and two charts tailored to its focus. DAX measures were developed to calculate metrics not directly available in the dataset, such as average sale price per order, total profit per order, and top-selling product. Slicers were added to allow filtering by relevant category and time frame.

The dataset included quantitative order data (e.g., Order Quantity, Unit Cost, Sale Price), but lacked qualitative details in the dimension tables, which limited both analytical depth and visual appeal. To address this, four fictional product categories were created and mapped to Product IDs. Additionally, because store locations were not provided, random ZIP codes were sourced from a separate USPS Excel file and imported as a new table. This table was joined to the store dimension, enabling geographic visualization of store sales via a bubble map.

Chart selection was tailored for clarity and insight:

- **Stacked bar charts**: Used to compare sales and profit by sale type or product category.
- **Stacked column charts**: Replaced initial line/area charts to better represent revenue and profit trends over time (by year, quarter, and month).
- **100% stacked area chart**: Illustrated shifts in product types sold over time.
- **Stacked area chart**: Showcased store sales over time, filtered by default to the top 10 most profitable stores, alongside the geographic map.

Additional qualitative attributes (e.g., product names, sales team names, and categories) would have further enriched the dashboard's functionality and user experience.

Regional Sales

Data Analysis

Insights

Through structuring, analysis, and visualization of the US Regional Sales data, the following insights were uncovered:

1. **In-store sales** accounted for the highest number of orders and total revenue, outperforming online, distributor, and wholesale channels.
2. **Quarters 3 and 4** consistently generated the highest sales, indicating strong seasonal trends.
3. **The year 2020** was the most profitable to date.
4. **Electronics** emerged as the top-performing product category in terms of sales volume.
5. **Product #23** had the highest quantity sold among all 47 products.
6. **Product #4** generated the highest total revenue.
7. Sales across product categories were **evenly distributed over time**, with no major fluctuations.
8. Of the 367 stores, **Store #284** was the most profitable on average.
9. **Sales Team #13** led in average earnings across all teams.
10. The **average total sales per store** was approximately **$49,740**.

Conclusion

This project demonstrated the value of integrating structured data modeling with dynamic visualization tools. By designing a robust star schema in MySQL and building an interactive Power BI dashboard, critical business insights were made clear and accessible. These findings can inform strategic decisions around inventory planning, team performance evaluation, seasonal marketing, and geographic expansion. With the current framework, future datasets can be easily incorporated for ongoing analysis and improved decision-making.