Natalie Daughety

Data Analytics Bootcamp

Superstore Database

This project is an analysis using SQL and Power Bi of a sample of sales data from a global superstore. The dataset was obtained from Kaggle by searching “sales data”. The first step in preparing the data for analysis was previewing the Excel file the data was located in. Column names were formatted to be similar and spaces were replaced with underscores to assist with SQL formatting. Dates were set to a standard of dd/MM/YYYY at this time as well.

The second step was generating the tables in MySQL workbench. Three dimension tables were created- customer, product and order. The customer table included fields for customer ID, name and locational information. The product table included fields for product ID, name, category and quantity sold. The order table included fields for order ID, shipping mode and date. The fact table linked the primary keys of the dimension tables. Row ID was used along with order ID as these were the only unique identifiers of the data set. Once dimensional tables and the fact table were created, they were sent to Power Bi.

In Power Bi, the tables were previewed and sent through the Power Query Editor. The relationships between the fact and dimension tables were also checked for correct formatting. Measures for total order, order count, average and total sales were also created. In the report, several visualizations are used.

Total sales, average sales and total orders are illustrated by data cards. These calculations were completed through measures. Sales by category are shown in a horizontal bar chart and are listed by product category and sum of sales. The top ten selling products by name are also shown in a horizontal bar chart with a count of quantity sold. Profit by order date is shown as an area chart. In order to summarize the data, the top 10 order dates were used. Sales by region is a vertical bar chart and shows sales by region alongside the sum of sales. The final elongated chart is a line chart showing average and sum of sales over time. Slicers for region and product category are used to further analyze differences by product type and regional sales.

The predicted outcome of the data analysis was to show the trends in sales amounts over time in order to highlight the busiest sales seasons as well as product trends by category. The final report showed that the most profitable time of year to sell is late Spring to Early Summer when looking at the month most sales occurred in. Furniture had the most dynamic level in sales while office supplies stayed consistent. Regional differences were minimal and mostly highlighted by different sale volume high points.