Sales Forecasting Analysis Project - Power BI Documentation

1. Project Overview

The objective of this project is to analyze and visualize sales data to uncover key insights, trends, and performance across various business dimensions. The dataset used contains sales information, and the analysis focuses on metrics such as total sales, sales by category, region, and customer segmentation. Additionally, forecasting models will be developed to project future sales performance.

2. Dataset Description

The dataset used for this analysis is from Kaggle[Sales Forecasting Dataset] https://www.kaggle.com/datasets/rohitsahoo/sales-forecasting).
It contains the following fields:
- Row ID: Unique identifier for each row
- Order ID: Unique identifier for each order
- Order Date: Date the order was placed
- Ship Date: Date the order was shipped
- Ship Mode: Mode of shipment
- Customer ID: Unique customer identifier
- Customer Name: Name of the customer
- Segment: Customer segment (e.g., Consumer, Corporate, Home Office)
- Country: Country of the customer
- City: City of the customer
- State: State of the customer
- Postal Code: Postal code of the customer
- Region: Sales region (e.g., East, West, Central)
- Product ID: Unique product identifier
- Category: Product category (e.g., Furniture, Office Supplies, Technology)
- Sub-Category: Product sub-category (e.g., Chairs, Phones)
- Product Name: Name of the product
- Sales: Total sales value for the product

3. Data Cleaning and Preparation

3.1 Steps Performed in Power Query Editor

- **Data Type Verification:** Ensured that all columns have appropriate data types (e.g., `Order Date` set to Date, `Sales` set to Decimal).
- **Handling Missing Values:** Checked for and removed any null values in important columns such as `Order Date`, `Sales`, and `Product ID`.

- **Removing Duplicates:** Identified and removed duplicate records based on `Order ID` to avoid double-counting of sales data.
- **Creating a Year Column:** Extracted the year from the `Order Date` field for time-based analysis using the formula:

DAX Formula:
Year = Date.Year([Order Date])

3.2 Date Issues Resolved

Addressed date formatting issues by converting `Order Date` to the proper Date type in Power Query and re-checking for valid entries.

4. Analysis and Calculations

4.1 Key Metrics

The following metrics were calculated to provide actionable insights:

- **Total Sales:**

DAX Formula:
Total Sales = SUM('SalesData'[Sales])

- **Sales by Category:**
  - A bar chart was created with `Category` as the X-axis and `Sales` as the Y-axis to visualize the sales distribution across product categories.
  - Alternatively, a measure was used for advanced calculations:

DAX Formula:
Sales by Category = CALCULATE(SUM('SalesData'[Sales]), ALLEXCEPT('SalesData', 'SalesData'[Category]))

- **Sales by STATE:**
  - Visualized using a bar chart and map, showing total sales distributed across different States.

- **Sales by Customer Segment:**
  - Analyzed how sales differ by customer segments such as Corporate, Consumer, and Home Office, using a pie chart and summary table.

4.2 Time-based Analysis

- **Sales Over Time:**

- Created a line chart to display total sales across different years and months.
- Used the extracted `*Year*` and `*Month*` columns to analyze sales trends over time.

4.3 Forecasting

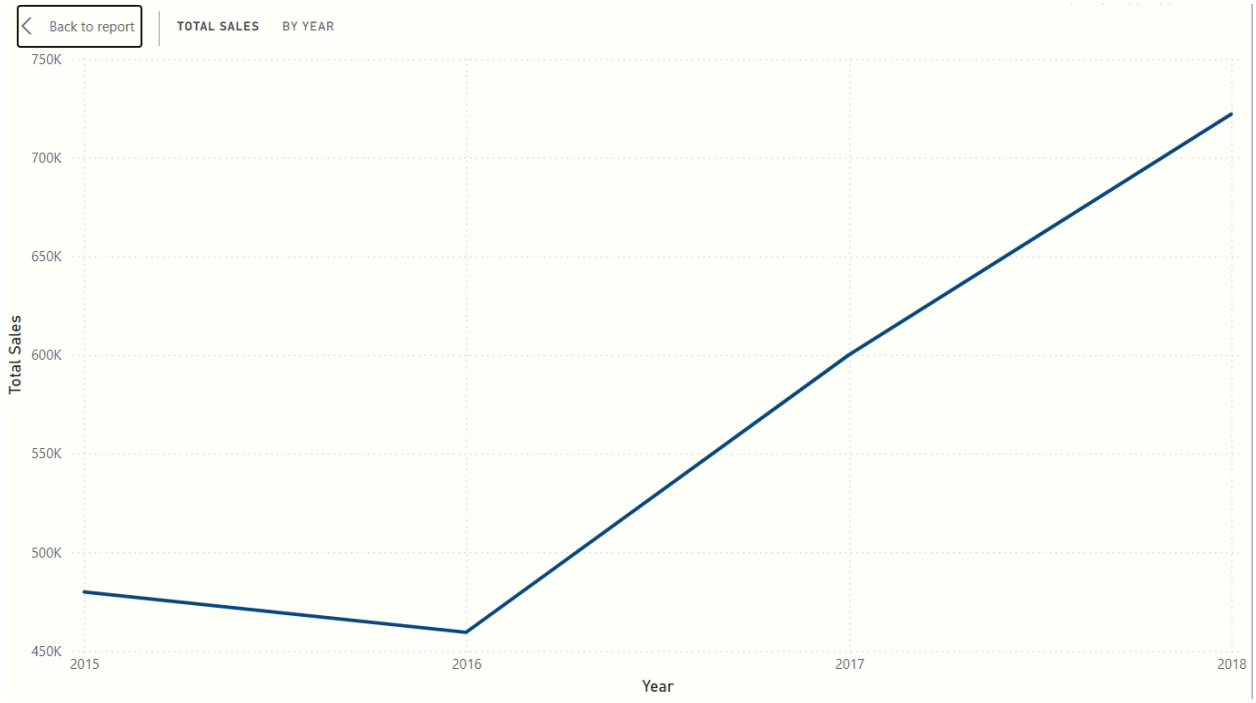
A forecasting model was created to predict future sales trends based on historical data:

- **Forecasting Visual:** Used the built-in forecasting feature in Power BI on the `*Order Date*` and `*Sales*` data to predict future trends.
- **Settings for forecasting:**
  - Confidence Interval: 95%
  - Forecast Length: 12 months

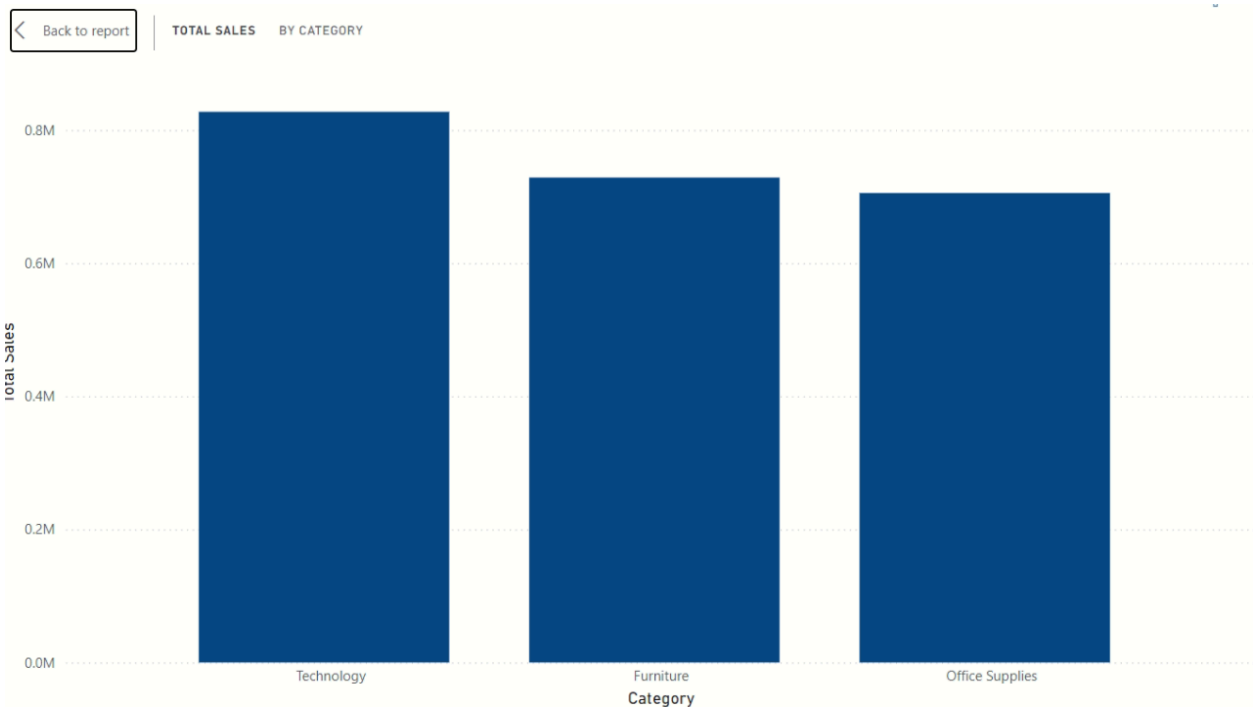
5. Visualizations

The following visuals were created to present the data effectively:

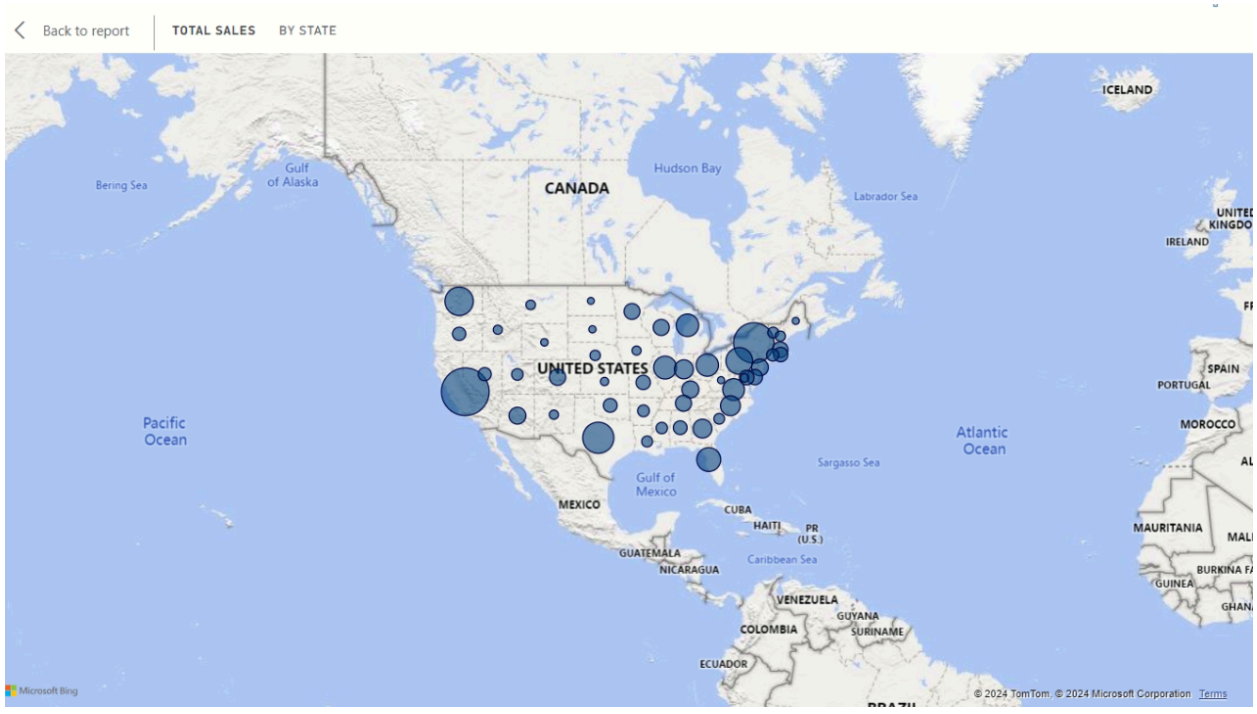
- **Total Sales Dashboard:**
  A card visual showing overall sales.



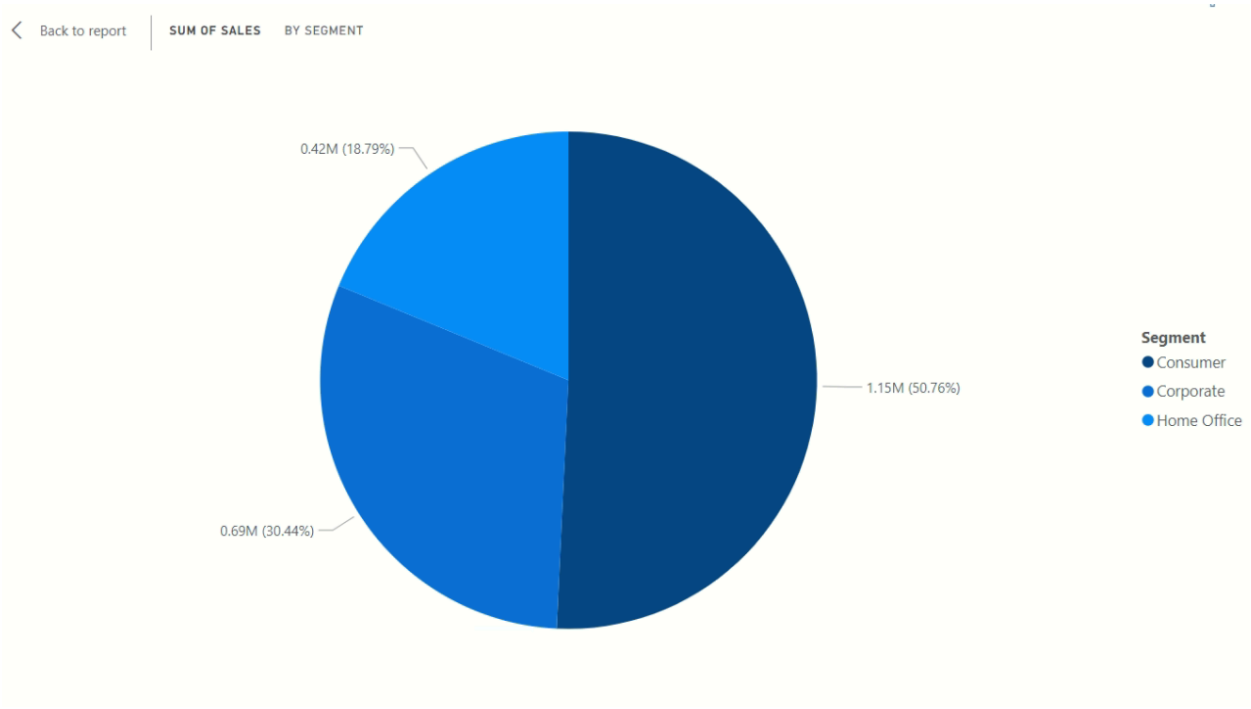
- **Sales by Category (Bar Chart):**
  Displaying total sales across each product category (Furniture, Office Supplies, Technology).



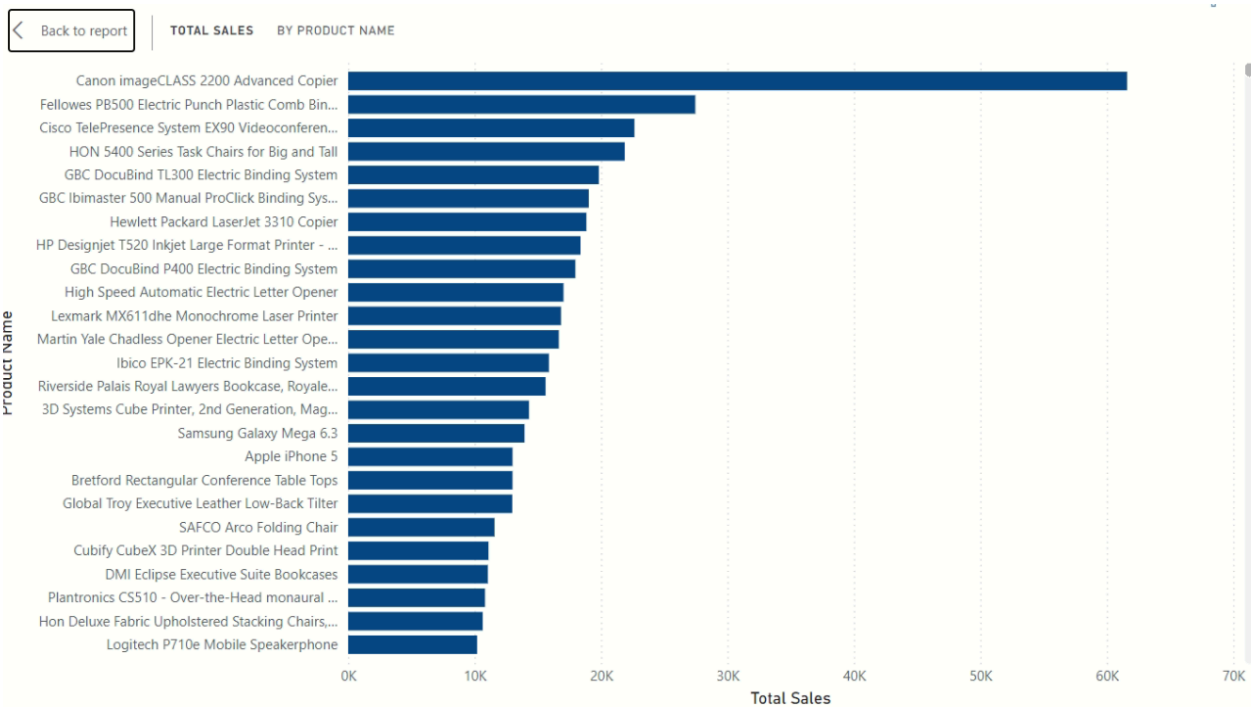
- **Sales by Region (Map and Bar Chart):**
  A map visual showing geographic sales distribution.



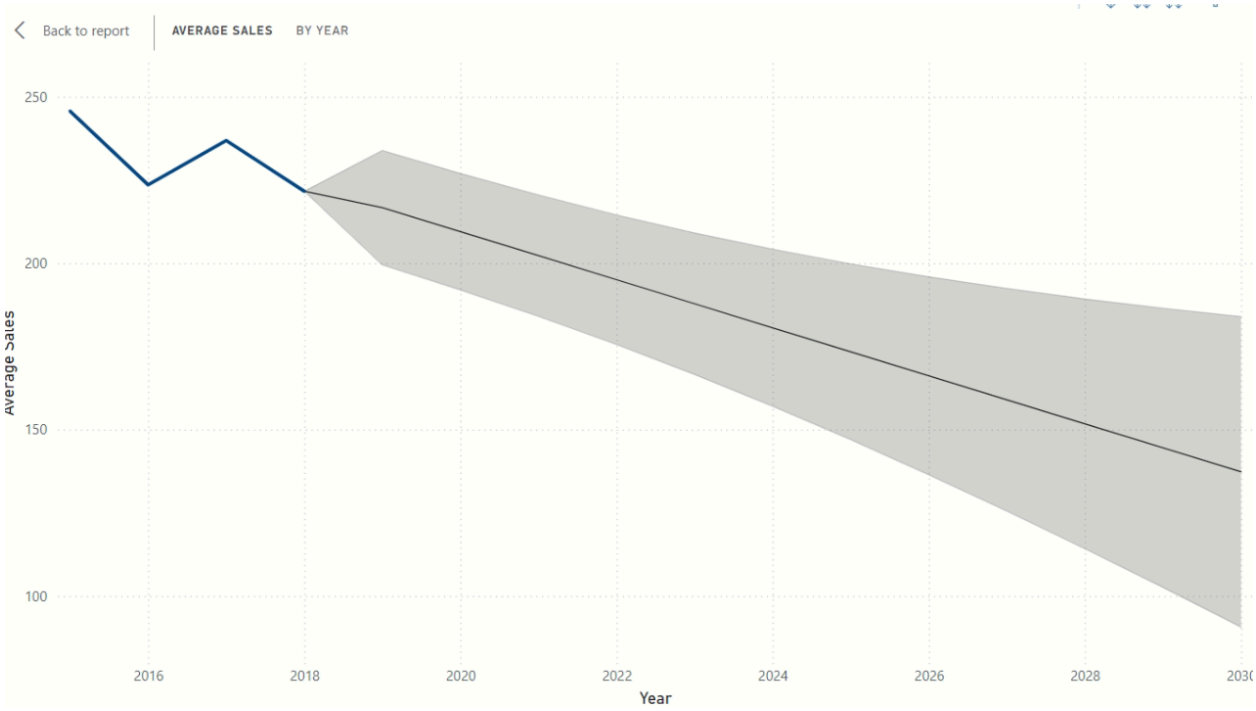
- **Sales by Segment (Pie Chart):**
  Segmenting sales based on customer types (Consumer, Corporate, Home Office).



- **Total Sales by Product Name (Clustered Bar Chart):**
  Total sales by Product Name.

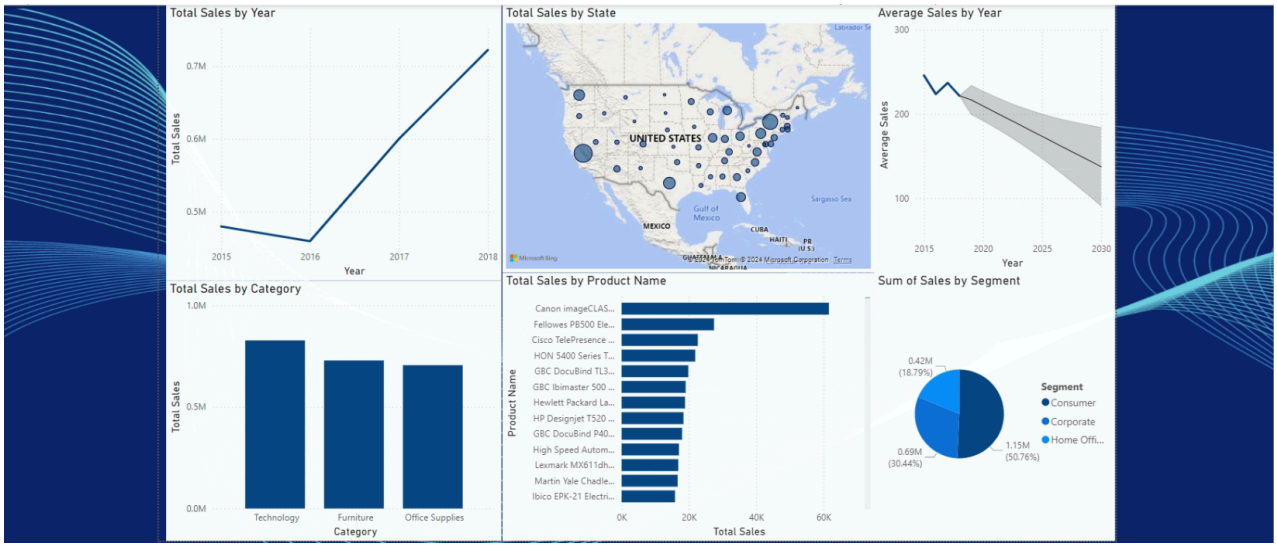


- **Sales Forecasting (Line Chart with Forecast):**
  A forecasting chart showing predicted sales for the next 12 months.



6. Conclusion

This analysis provided key insights into sales performance across categories, regions, and customer segments. The visualization and forecasting tools in Power BI allowed for both historical analysis and future projections, helping stakeholders make informed decisions.



Key Findings:

- *<u>Top-Performing Category</u>: Technology had the highest sales overall.*
- *<u>Regional Sales Trends</u>: The West region consistently had the highest sales.*

- *Sales Forecast: Projected growth in sales for the next 12 months, with a peak expected during the holiday season.*

7. Recommendations for Future Work

- Data Enrichment: Including additional data such as customer demographics and marketing spend could improve the quality of the analysis.
- Predictive Analytics: Implement more advanced predictive models using R or Python to enhance forecasting accuracy.
- Automation: Set up automatic data refresh in Power BI for real-time dashboard updates.

8. Tools Used

- Power BI: For data visualization and reporting.
- Power Query: For data cleaning and transformation.
- DAX (Data Analysis Expressions): For calculated columns and measures.