Northwind Sales Data Analysis Documentation

This project involves an in-depth analysis of the Northwind Traders sales data, focusing on customer behaviour, product performance, and shipping efficiency. The analysis was done using Python in a Jupyter Notebook and involves data cleaning, data transformation, and exploratory data analysis (EDA), with the aim to uncover key insights about customer behaviour, product popularity, and shipping performance.

Data Preprocessing

Datasets Used:
- **Orders:** Information about orders placed, including dates, freight costs, and customer details.
- **Customers:** Details of customers, including their contact information and location.
- **Products:** Information about products, such as product names, categories, and prices.
- **Order Details:** Specifics about each order, including the products ordered, quantities, and discounts.

Key Preprocessing Steps:
- **Date Conversion:**
  The `orderDate`, `requiredDate`, and `shippedDate` columns were converted to DateTime format to enable time-based analysis.

- **Handling Missing Values:**
  The `shippedDate` column had missing values, which were filled using the `requiredDate` as a proxy, under the assumption that shipments were completed by the required date.

- **Merging Datasets:**
  - The datasets were merged into a single comprehensive dataset, combining customer details, order information, and product data to enable a holistic analysis.

- **Data Transformation:**
  - Days to Ship: A new feature `days_to_ship` was calculated as the difference between `orderDate` and `shippedDate`, representing the time taken to fulfill an order.
  - Customer Segments:Customers were segmented into three categories—Low, Medium, and High—based on their total spend

---

Exploratory Data Analysis (EDA)

- **Revenue Analysis:**

- Revenue distribution was examined across the three customer segments. The analysis revealed that a significant portion of revenue is generated by the "High" segment, highlighting the importance of focusing on high-value customers.

- **Order Frequency:**
  - The monthly order trend was analyzed, showing fluctuations in order volumes over time. A notable increase in orders was observed towards the end of 2014 and the beginning of 2015, possibly indicating seasonal demand or successful marketing efforts.

- **Best-Selling Products:**
  - The top-selling products were identified, with items like "Camembert Pierrot" and "Raclette Courdavault" leading in sales. This insight can guide inventory management and marketing strategies.

- **Shipping Analysis:**
  - The average days to ship each product were calculated, revealing that some products take significantly longer to ship than others.

- **Customer Analysis:**
  - The frequency of orders by customers was analyzed, identifying customers who frequently place orders. The average order value was examined across different customer segments, with "High" segment customers placing larger orders on average.

. **Visualization in Power BI**

The Power BI dashboard visualizes key insights from the analysis:

- **Monthly Order Trend:**
  - This line chart shows the number of orders placed each month from July 2013 to April 2015, highlighting peaks and troughs in sales activity.

- **Revenue Distribution by Customer Segment:**
  - A pie chart illustrates the proportion of revenue contributed by each customer segment, with the "High" segment dominating the revenue share.

- **Best-Selling Products:**
  - A bar chart lists the top products by sales volume, providing a quick overview of the most popular items.

- **Highest-Frequency Customers:**
  - A heatmap identifies the customers who place orders most frequently, allowing the company to target these loyal customers with personalized offers or rewards.

- **Global Sales Distribution:**

- A map visualization shows the geographic distribution of sales, indicating regions where the company is most active.

- **Year-over-Year Sales Comparison:**
    - This line chart compares sales trends across 2013, 2014, and 2015, offering insights into seasonal trends and overall growth.

This analysis of Northwind's sales data provides several actionable insights:

- Focus on High-Value Customers, the "High" revenue segment drives the majority of the company's revenue.

- Optimize Inventory for Best-Selling Products, ensuring adequate stock levels for items like "Camembert Pierrot", to meet customer demand.

- Improve Shipping Times, by addressing delays in shipping for certain products can enhance customer satisfaction and retention, potentially leading to repeat business.

- Capitalize on Seasonal Trends, The observed peaks in orders around the end of the year suggest that seasonal promotions or campaigns could be effective in boosting sales during these periods.