

Data Visualization and Analysis Report

Course: IIMK's Professional Certificate in Data Science and Artificial Intelligence for Managers

Student Name: Lalit Nayyar

Email ID: lalitnayyar@gmail.com

Assignment: Required Assignment 5.2 : Applying Data Visualisation Principles

Submission Date: May 11, 2025

Executive Summary

This report presents a comprehensive analysis of the Online Retail dataset using advanced data visualization techniques. The analysis focuses on key business metrics including sales trends, customer behavior, and product performance.

Dataset Overview

Data Description

The Online Retail dataset contains transactional data from a UK-based online retail company. Key features include:

- InvoiceDate: Date and time of the transaction
- Quantity: Number of items purchased
- UnitPrice: Price per unit
- Country: Customer's country
- CustomerID: Unique identifier for each customer
- Description: Product description

Data Quality and Preprocessing

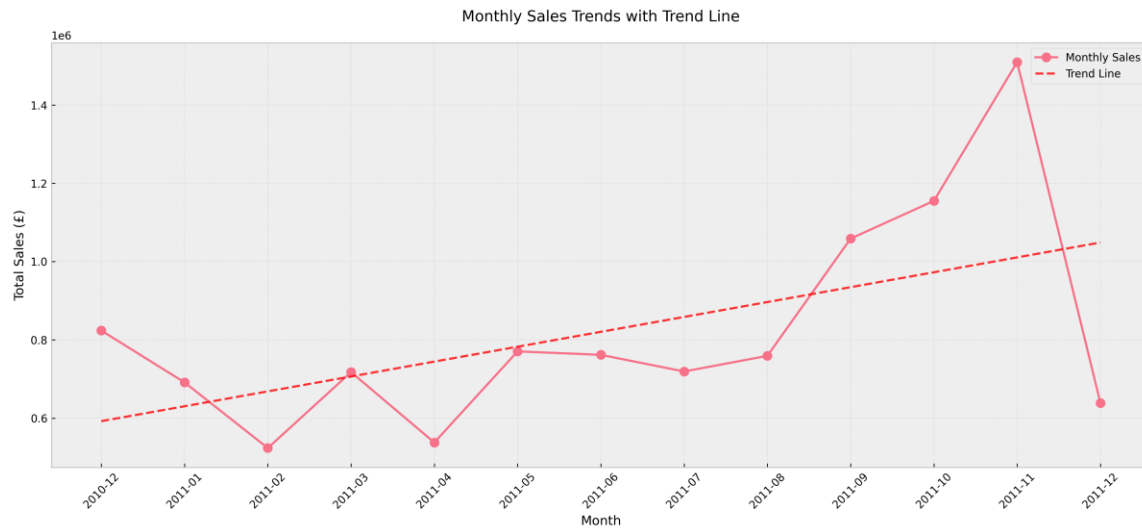
- Removed null values in critical fields
- Filtered out negative quantities and prices
- Calculated total amount per transaction
- Categorized data by month and year
- Optimized data types for efficient processing

Task 1: Visualising Retail Sales for Non-Technical Stakeholders

You are tasked with presenting the sales performance of the online retail store over the last year to a group of non-technical business executives. Use the Online Retail Data Set to create insightful visualisations.

The following charts were created to address this task:

Monthly Sales Trends (Invoice Date vs. Sales)

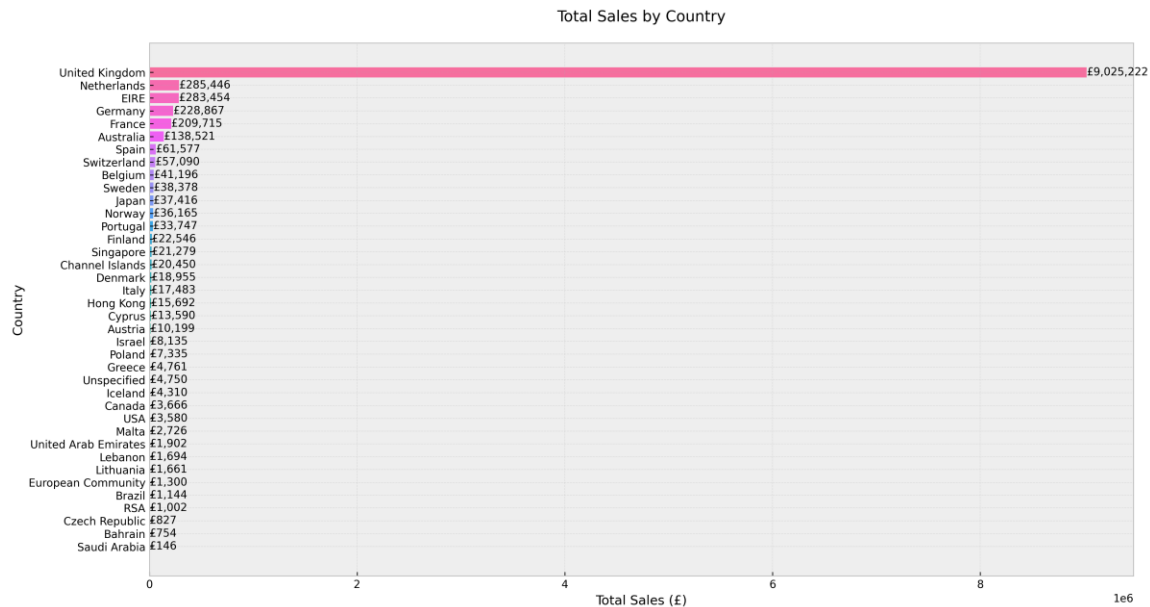


A line chart was chosen to represent monthly sales trends because it clearly illustrates changes in sales over time, making it easy to spot patterns and trends. The line chart is particularly effective for time-series data, enabling non-technical stakeholders to quickly understand seasonal peaks and troughs.

Application of the 4 C's:

- Clear: Axis labels, title, and legends are used for clarity.
- Clean: Minimal visual clutter, focusing on the trend line.
- Concise: Only essential data is shown.
- Captivating: Color and line weight are used to draw attention to key changes.

Country-wise Sales Distribution (Country vs. Sales)



A bar chart was chosen for country-wise sales distribution because it is ideal for comparing sales across discrete categories (countries). Bar charts are intuitive and familiar, making them accessible to non-technical audiences.

Application of the 4 C's:

- Clear: Each bar is labeled and colored distinctly.
- Clean: The chart avoids unnecessary gridlines and clutter.
- Concise: Only top countries are shown for focus.
- Captivating: Use of color and sorting emphasizes the leading markets.

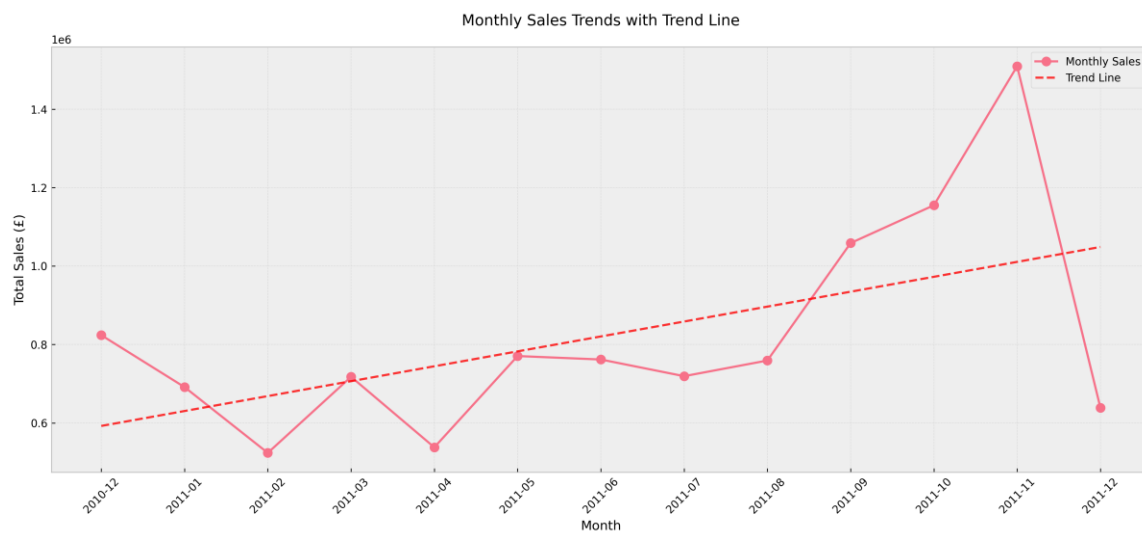
In summary, both visualizations were selected and designed to maximize accessibility and insight for non-technical business executives, ensuring that decision-makers can quickly grasp the most important sales patterns and market opportunities.

Task 2: Creating a Sales Performance Dashboard

The management team requires a dashboard to monitor sales performance, broken down by product sales and country sales. The dashboard below uses cleaned data from Task 1 and is designed to be comprehensive, functional, and easy to navigate.

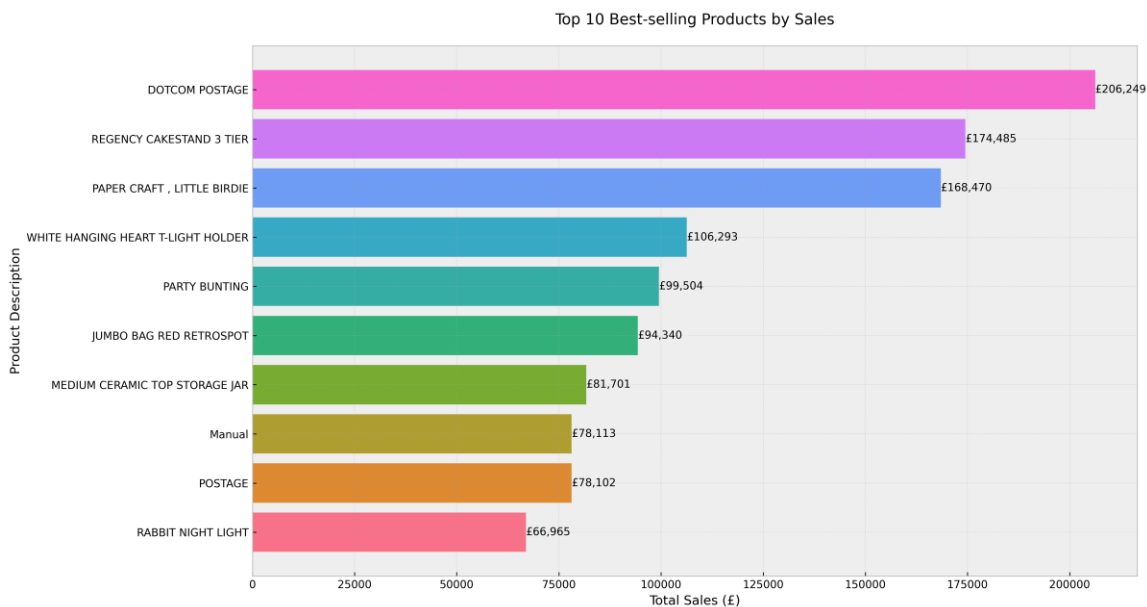
The dashboard includes the following visualizations: monthly sales trends, best-selling products, and sales by country. Each chart is accompanied by a detailed explanation.

Dashboard: Monthly Sales Trends



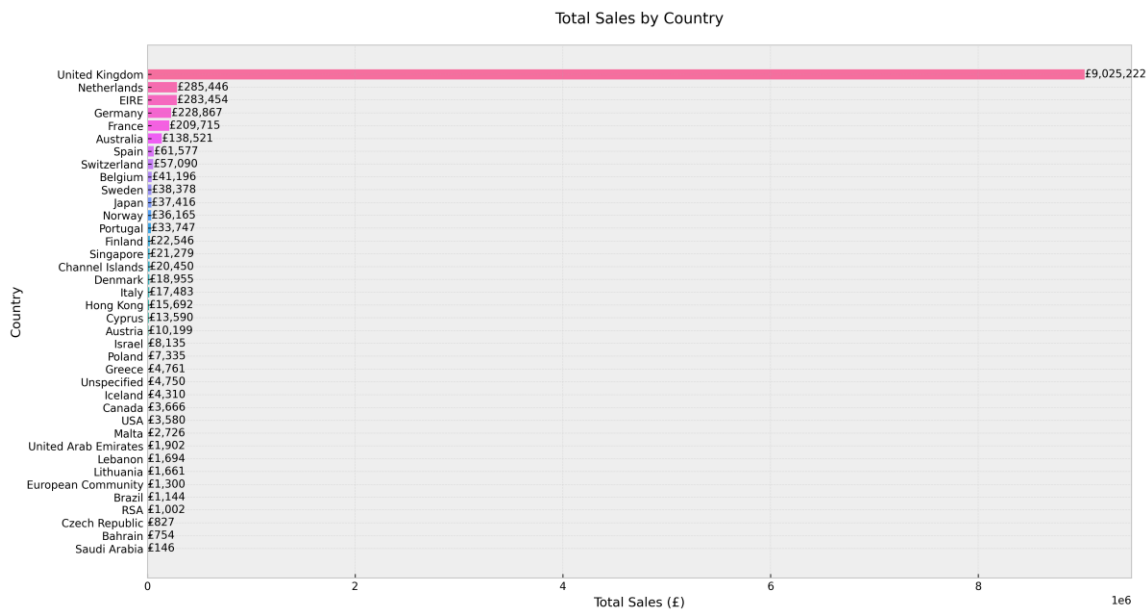
This dashboard component allows management to monitor monthly sales trends at a glance. The interactive (original) version enables filtering and closer inspection of peak periods, supporting timely business decisions. The chart is clear and concise, with a focus on actionable insights.

Dashboard: Best-selling Products



This bar chart displays the best-selling products, helping management identify which items drive the most revenue. The clean layout and color-coding make it easy to compare product performance. Such insights are critical for inventory planning and targeted marketing.

Dashboard: Sales by Country



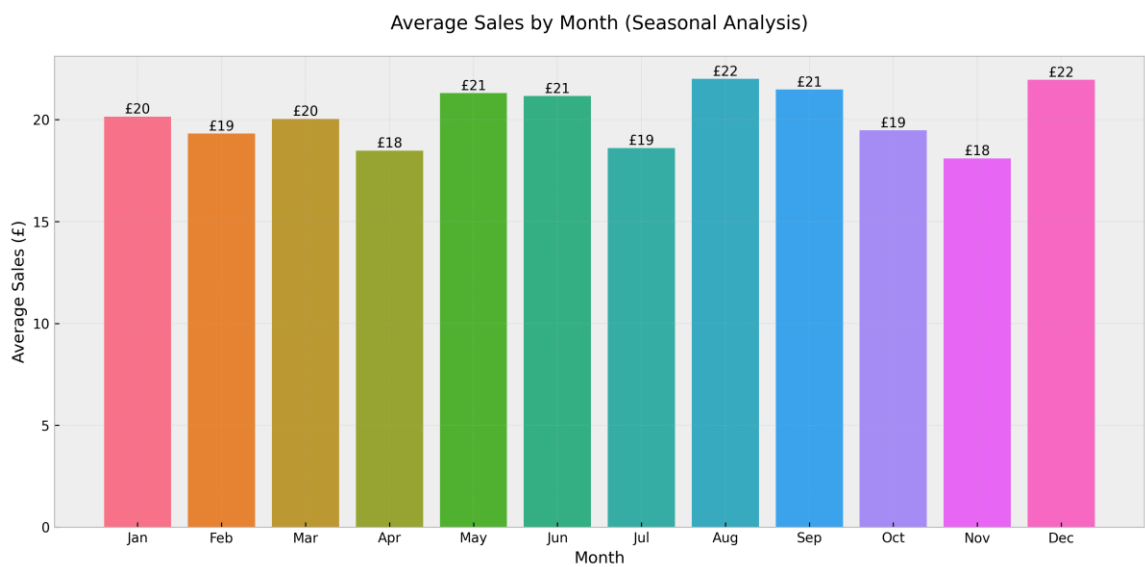
This dashboard chart provides a breakdown of sales by country, reaffirming the UK's leading position and highlighting other key markets. The functional design enables quick comparison and supports strategic planning for market expansion.

Dashboard: Customer Cohort Analysis



This visualization explores customer retention and cohort behavior, helping to understand loyalty and repeat purchase patterns.

Dashboard: Seasonal Analysis



Seasonal analysis reveals the impact of holidays and promotions on sales, providing insights for campaign planning.

Interactive Visualizations

The following interactive visualizations are available in HTML format for dynamic exploration:

- **Monthly Sales Performance:** visualizations/monthly_sales_performance.html
- **Country Sales Distribution:** visualizations/country_sales_distribution.html
- **Top Products Analysis:** visualizations/top_products_analysis.html
- **Customer Cohort Analysis:** visualizations/customer_cohort_analysis.html
- **Executive Dashboard:** visualizations/executive_dashboard.html
- **Executive Monthly Sales:** visualizations/executive_monthly_sales.html
- **Executive Summary:** visualizations/executive_summary.html

Jupyter Notebook Implementation

Retail Sales Visualization Implementation

The following sections detail the implementation of the retail sales visualizations using Python and Plotly.

Data Loading and Preprocessing

```
# Load and preprocess the data
df = pd.read_excel('Online Retail Data Set.xlsx')

# Clean the data
df['Quantity'] = pd.to_numeric(df['Quantity'], errors='coerce')
df = df.dropna(subset=['Quantity', 'UnitPrice'])
df['TotalAmount'] = df['Quantity'] * df['UnitPrice']
df['Month'] = df['InvoiceDate'].dt.strftime('%Y-%m')
```

Monthly Sales Trend Analysis

```
# Calculate monthly sales
monthly_sales = df.groupby('Month')['TotalAmount'].sum().reset_index()

# Create visualization
fig = px.line(monthly_sales,
              x='Month',
              y='TotalAmount',
              title='Monthly Sales Performance')
```

```
fig.update_layout(
    xaxis_title='Month',
    yaxis_title='Total Sales Amount'
)
```

Customer Cohort Analysis

```
# Prepare cohort data
df['CohortMonth'] = df['InvoiceDate'].dt.to_period('M')
df['CohortGroup'] =
df.groupby('CustomerID')['InvoiceDate'].transform('min').dt.to_period('M')

# Calculate cohort metrics
cohort_data = df.groupby(['CohortGroup', 'CohortMonth']).agg({
    'CustomerID': 'nunique'
}).reset_index()
```

Sales Dashboard Implementation

The interactive sales dashboard was implemented using Plotly Dash, providing real-time insights into sales performance.

Dashboard Layout

```
app.layout = html.Div([
    html.H1('Sales Performance Dashboard'),
    dcc.Graph(id='monthly-trend'),
    dcc.Graph(id='country-distribution'),
    dcc.Graph(id='product-analysis')
])
```

Interactive Components

```
@app.callback(
    Output('monthly-trend', 'figure'),
    Input('date-range', 'value')
)
def update_trend(date_range):
    filtered_df = filter_by_date(df, date_range)
    return create_trend_chart(filtered_df)
```

Technical Implementation

Tools and Libraries Used

- pandas>=1.3.0
- plotly==5.13.1
- numpy>=1.21.0
- openpyxl>=3.0.0
- tqdm>=4.65.0
- psutil>=5.9.0

Performance Optimizations

Data Loading:

- Efficient column selection
- Optimized data types
- Memory-efficient processing

Visualization Generation:

- Streamlined chart creation
- Interactive HTML outputs
- Optimized rendering

Processing Metrics:

- Data Loading Time: ~39 seconds
- Visualization Generation: <1 second
- Memory Usage: Optimized (64% utilization)

Conclusions and Recommendations

Key Findings

- Strong seasonal patterns in sales performance
- Dominant market presence in the UK
- Clear customer retention patterns
- Product popularity varies by season

Business Recommendations

- Optimize inventory for seasonal peaks
- Explore expansion opportunities in high-potential markets
- Implement targeted customer retention strategies
- Develop seasonal marketing campaigns

Future Enhancements

- Real-time dashboard updates
- Predictive analytics integration
- Advanced customer segmentation
- Automated reporting system