Comprehensive Data Analysis and Visualization for Retail Business Optimization

# Context

In today's competitive retail landscape, businesses must leverage data to optimize their operations, enhance customer experience, and drive profitability. This assignment places you in the role of a data analyst at a retail company seeking to enhance its business through a comprehensive analysis of sales data. You will employ various programming and data analysis techniques to clean, process, and visualize data, ensuring it aligns with industry best practices and standards.

# Content

As a data analyst, you will work with a dataset containing sales records, customer information, and product details. Your tasks will involve utilizing object-oriented programming principles, working with Python modules and packages, handling files, and scraping data from the web. Additionally, you will perform data manipulation using NumPy and Pandas, and create visualizations with Matplotlib, Plotly, and Seaborn to uncover insights and present your findings effectively.

# Data Description

The dataset for this assignment is publicly available on Kaggle and consists of the following files:  
**1. Sales Data:** Contains transactional sales records, including transaction ID, product ID, customer ID, date of purchase, quantity, and price.  
**2. Customer Data:** Includes customer demographics such as age, gender, location, and loyalty membership status.  
**3. Product Data:** Lists product details like product ID, name, category, and supplier information.  
  
Dataset Link: <https://www.kaggle.com/datasets/kyanyoga/sample-sales-data>

# Objective

The objective of this assignment is to integrate various programming, data handling, and visualization skills to perform a comprehensive analysis of the sales data. By the end of this assignment, you should be able to:  
1. Apply object-oriented programming (OOP) principles in Python.  
2. Utilize Python modules and packages for data handling and web scraping.  
3. Perform data cleaning, manipulation, and analysis using NumPy and Pandas.  
4. Create insightful visualizations using Matplotlib, Plotly, and Seaborn.  
5. Implement best practices and PEP standards in your Python code.

# Tasks

**1. Data Loading and Inspection**  
 - Load the sales, customer, and product datasets into Pandas DataFrames.  
 - Inspect the data for missing values, inconsistencies, and outliers.  
  
**2. Object-Oriented Programming (OOP) Implementation**  
 - Define classes for Customer, Product, and SalesTransaction.  
 - Implement constructors and destructors for these classes.  
 - Demonstrate inheritance by creating a subclass for a specific type of customer (e.g., VIPCustomer) that inherits from the Customer class.  
  
**3. Data Cleaning and Transformation**  
 - Handle missing values and correct data inconsistencies.  
 - Transform data types where necessary (e.g., date formatting).  
 - Merge the datasets to create a unified DataFrame for analysis.  
  
**4. Python Modules and Packages**  
 - Create a custom Python module for data cleaning functions.  
 - Utilize built-in modules like os, time, and sys to manage files and system operations.  
 - Fetch additional data from a REST API using the requests module (e.g., exchange rates for currency conversion).  
  
**5. File Handling**  
 - Write functions to read, append, and handle files in Python.  
 - Save cleaned data to new CSV files.  
  
**6. Web Scraping**  
 - Use BeautifulSoup or Scrapy to scrape additional data (e.g., product reviews) from a relevant e-commerce website.  
  
**7. Data Analysis with NumPy and Pandas**  
 - Perform numerical operations and linear algebra using NumPy.  
 - Manipulate and analyze data using Pandas (e.g., filtering, grouping, and aggregating data).  
 - Load data from various formats (CSV, Excel, JSON) and export analysis results to different formats.  
  
**8. Data Visualization**  
 - Create basic plots (line plot, scatter plot, pie chart, bar plot) using Matplotlib.  
 - Enhance visualizations with annotations and customizations.  
 - Create interactive plots using Plotly (e.g., histograms, bar plots).  
 - Generate advanced visualizations with Seaborn (e.g., countplot, barplot, scatterplot).  
  
**9. Geospatial Analysis**  
 - Utilize latitude and longitude data for creating maps and charts.  
 - Plot customer distribution and sales hotspots on a map.  
  
**10. Programming Best Practices**  
 - Follow PEP standards for code style and documentation.  
 - Write efficient, readable, and reusable code.  
 - Implement error handling and debugging strategies.  
  
**11. Problem-solving and Debugging with ChatGPT**  
 - Use ChatGPT to assist in solving coding problems and debugging your scripts.  
  
**12. Final Report and Presentation**  
 - Summarize your findings and insights in a comprehensive report.  
 - Create a presentation with key visualizations and insights for stakeholders.