ABC Retail Data-Driven Analysis Documentation

---

1. Project Overview

- **Project Title**: *Data-Driven Insights for ABC Retail*
- **Objective**: Use data to identify underperforming products, peak sales days, and key customer behavior to support strategic decision-making.

---

2. Data Sources and Preprocessing

Data Files:

- `customer_data.csv`: Contains customer information.
- `product_data.csv`: Contains product details.
- `sales_data.csv`: Sales transaction records.

Data Cleaning and Preparation in R: The data was cleaned using R (ABC_Retail_Data_Exploration_and_Cleaning_script) to ensure consistency, remove duplicates, address missing values, and prepare data for analysis. Below is a summary of each cleaning step.

Customer Data Cleaning (`customer_data`)

- **Duplicate Check**: Verified `CustomerID` uniqueness and removed duplicates.
- **Missing Values**:
  - Checked for missing values in `CustomerID` and `CustomerName`.
  - Replaced specific `CustomerID`s with missing names.
- **Address Formatting**: Combined separate address fields into a single `FullAddress_Combined` column.
- **Phone Number Formatting**: Formatted phone numbers to `(###)-###-####`.
- **Add Anonymous Customer**: Added an entry with `CustomerID = 0` for anonymous purchases.
- **Saved Cleaned File**: Saved as `abc_customer_data.csv`.

Product Data Cleaning (`product_data`)

- **Duplicate and Missing Values Check**: Verified `ProductID` uniqueness, filled missing values in `Price` and `StockQuantity` with `0`.

- **Category Updates**: Manually updated missing `Category` values for certain products.
- **Inventory Calculation**: Added `StockValue` column by multiplying `Price` by `StockQuantity`.
- **Add Unnamed Product**: Included a `ProductID = 0` entry for uncategorized sales.
- **Saved Cleaned File**: Saved as `inventory_data.csv`.

Sales Data Cleaning (`sales_data`)

- **Duplicate and Missing Values Check**: Removed rows with missing or blank `SalesID`.
- **Replace Missing Values**: Filled missing values in `CustomerID`, `ProductID`, and `SaleAmount` with `0`.
- **Date Handling**: Converted `SaleDate` to date format, added `SaleDay` column for day of the week.
- **Saved Cleaned File**: Saved as `customer_sales_data.csv`.

Merged Dataset (`customer_sales_data`)

- **Join Operation**: Merged `sales_data`, `customer_data`, and `product_data` based on `CustomerID` and `ProductID`.
- **Reordering Columns**: Placed key fields like `SaleDate`, `SaleDay`, and `SaleAmount` prominently for easier analysis.

---

3. Analysis Tools and Techniques

This project employed a range of tools to analyze and visualize the data:

Python Analysis

- **Purpose**: Initial exploratory data analysis, visualization, and preliminary insights.
- **Files/Code**: Saved as ABC_Retail.ipynb

SQL Analysis in SQLite

- **Purpose**: Perform targeted queries for data aggregation.
- **Files/Code**: SQL queries saved as ABC_Retail_SQL_Queries.sql

Excel Pivot Tables and Macros

- **Purpose**: To dynamically explore data and automate repetitive tasks.
- **Files**: Excel workbook with pivot tables and macros saved as ABC_Retail.xlsm
- **Macros Used**:
    - `TopCustomers`: Displays top customers based on purchase amount.
    - `TopProducts`: Lists best-selling products.
    - `BestSellingDays`: Shows sales distribution by day of the week.
    - `TopSellingMonth`: Highlights peak months for sales.

---

4. Key Analyses and Insights

4.1 Product Performance

- **Objective**: Identify high and low-performing products.
- **Excel Pivot Table**:
  - **Pivot Fields**: `ProductName` as Row field, `SaleAmount` as Values.
  - **Macro Used**: `TopProducts`.
- **Actionable Recommendation**: Develop promotional strategies for underperforming products and stock more high-demand products.

4.2 Customer Behavior

- **Objective**: Categorized customers based on spending and purchase frequency.
- **Excel Pivot Table**:
  - **Pivot Fields**: `CustomerName` as Row field, `SaleAmount` as Values.
  - **Macro Used**: `TopCustomers`.
- **Actionable Recommendation**: Implement loyalty programs to incentivize high-value and frequent customers.

4.3 Sales Timing Analysis

- **Objective**: Determine peak sales days and seasonal trends.
- **Excel Pivot Table**:
  - **Best Day Analysis**:
    - **Pivot Fields**: `SaleDay` as Row field, `SaleAmount` as Values.
    - **Macro Used**: `BestSellingDays`.
  - **Best Month Analysis**:
    - **Pivot Fields**: `SaleDate` (grouped by month) as Row field, `SaleAmount` as Values.
    - **Macro Used**: `TopSellingMonth`.
- **Actionable Recommendation**: Align marketing campaigns with peak days and months to maximize sales.

4.4 Inventory Insights

- **Objective**: Evaluate stock levels and identify inventory adjustments.
- **Excel Pivot Table**:
  - **Pivot Fields**: `ProductName` as Row field, `StockQuantity` and `StockValue` as Values.
- **Actionable Recommendation**: Identify slow-moving products and prioritize stock for high-demand items.

---

5. Implementation and Action Plan

- **Marketing Initiatives**: Targeted campaigns aligned with peak sales days and months.
- **Inventory Management**: Review stock levels regularly and adjust based on turnover rates.
- **Customer Retention Programs**: Introduce loyalty rewards for valuable and frequent customers.

---

6. Technical Details

Database Schema

- **customer_sales_data Table**: Merged sales data including customer and product information.

R Code: Data Cleaning Script

- The R script for data cleaning outlines steps such as removing duplicates, handling missing values, and formatting columns.

Excel Macros

- **File**: `Data_Analysis.xlsx`.
- **Macro Functions**:
  - `TopCustomers`: Summarizes top customers.
  - `TopProducts`: Lists best-selling products.
  - `BestSellingDays`: Shows high-performing days.
  - `TopSellingMonth`: Highlights peak sales months.
- **Code**: VBA code snippets are embedded within the Excel workbook.

---

7. Conclusion

The combination of R, Python, SQL, and Excel pivot tables provided a thorough analysis, enabling data-driven decisions for targeted marketing, improved inventory management, and enhanced customer engagement.
