Global Electronics Retailer – Requirements Document

1. Project Overview

Global **Electronics Retailer Analysis** is a Python-based data pipeline designed to clean, integrate, analyze, and visualize data from a multinational electronics retail business. The final output is an automated PDF report (Global_Electronics_Retailer_Report.pdf) that includes visualizations for decision-makers.

---

2. Objectives

- Clean and preprocess customer, product, sales, store, and currency exchange datasets.
- Merge datasets into a unified view of operations.
- Perform sales, store, product, and customer analysis.
- Generate PDF reports containing charts, metrics, and insights.
- Enable business stakeholders to make data-driven decisions.

---

3. Data Sources

The data files are stored in the data/ directory:

File Name	Description
Customers.csv	Customer profiles and demographics
Products.csv	Product catalog, cost, and price information
Stores.csv	Store metadata including location and size
Sales.csv	Transaction-level sales data
Exchange_Rates.csv	Exchange rates for currencies

---

4. Tools & Technologies

- **Language:** Python
- **Libraries:**

- pandas – data manipulation
- matplotlib, seaborn – visualizations
- plotly – interactive (not included in PDF yet)
- PdfPages – PDF report generation
- numpy, datetime, scipy – calculations and formatting

---

5. Functional Requirements

5.1 Data Cleaning

- Parse dates (Order Date, Delivery Date, etc.)
- Convert strings with currency symbols to numeric
- Handle missing values (e.g., fill, warn, or drop)
- Replace incorrect state codes (e.g., Napoli → NA)
- Drop duplicate rows (keep the first)

5.2 Data Integration

- Merge all datasets using keys (ProductKey, CustomerKey, etc.)
- Calculate derived fields:
  - Sales Amount USD, Profit, Profit Margin
  - Age, Store Age, Repeat Purchase Count

5.3 Analysis

- **Sales Analysis:** Monthly profit vs revenue trends, Sales by geography
- **Store Performance:** Top/bottom stores by revenue & AOV, Physical vs Online stores trends, Store age impact on performance
- **Product Performance:** Best/worst selling products, Product category distribution, category and subcategory trends, Brand performance by country
- **Customer Analysis:** Repeat purchase rate, gender distribution, sales by age group and gender, Top repeat customers
- **Delivery Analysis:** Delivery time distribution

5.4 Report Generation

- Generate PDF with:
    - First page (cover)
    - Sectioned charts with titles

---

6. Output

- Global_Electronics_Retailer_Report.pdf containing Visual analytics.

---

7. Assumptions

- All CSVs are stored under data/ relative to the script location.
- Files are consistently encoded (unicode_escape).
- All data files (Customers, Products, Stores, Sales, Exchange Rates) have unique primary keys. There are no duplicate primary key values within these files, ensuring consistent joins without duplication.
- The user has write permission to export the PDF.
- Dates are in MM/DD/YYYY format.
- Square meter can be NULL for online stores as it does not require physical space.
- Delivery dates can be NULL for Physical stores as customer directly buy from the store.
- Currency conversion is not required as unit price and costs are already in USD in product table.
- Currently, default values are not populated for all columns which have NULLs.

---

8. Limitations / Future Improvements

- Plotly charts are generated but not added to the PDF (future enhancement).
- Real-time dashboards are not included in this version.
- Schedule to generate PDF daily/weekly (future enhancement).
- Share PDF with stakeholders either via email or shared folders (future enhancement).