**📥 Step 1: Access and Understand the Dataset**

You can download the dataset from Maven Analytics' Data Playground:

🔗 Global Electronics Retailer Dataset

This dataset includes multiple CSV files representing different aspects of the business:

* **Customers**: Customer demographics and details
* **Products**: Information about electronic products
* **Stores**: Store locations and details
* **Sales Transactions**: Records of sales made
* **Currency Exchange Rates**: Exchange rates for different currencies

**🛠️ Step 2: Set Up Your Python Environment**

Ensure you have the necessary Python libraries installed:

pip install pandas matplotlib seaborn

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

**📂 Step 3: Load the Data**

Load each CSV file into a Pandas DataFrame:

customers = pd.read\_csv('customers.csv')

products = pd.read\_csv('products.csv')

stores = pd.read\_csv('stores.csv')

sales = pd.read\_csv('sales.csv')

exchange\_rates = pd.read\_csv('exchange\_rates.csv')

**🔍 Step 4: Explore and Clean the Data**

Perform initial exploration:

print(sales.head())

print(sales.info())

print(sales.describe())

Check for missing values:

print(sales.isnull().sum())

Handle missing values and data types as needed.

**Step 5: Merge DataFrames for Analysis**

Combine relevant DataFrames to create a comprehensive dataset:

# Merge sales with products

sales\_products = pd.merge(sales, products, on='product\_id', how='left')

# Merge with customers

sales\_products\_customers = pd.merge(sales\_products, customers, on='customer\_id', how='left')

# Merge with stores

full\_data = pd.merge(sales\_products\_customers, stores, on='store\_id', how='left')

**📈 Step 6: Perform Exploratory Data Analysis (EDA)**

Analyze total sales over time:

# Convert 'sale\_date' to datetime

full\_data['sale\_date'] = pd.to\_datetime(full\_data['sale\_date'])

# Group by month

monthly\_sales = full\_data.groupby(full\_data['sale\_date'].dt.to\_period('M')).agg({'sales\_amount': 'sum'}).reset\_index()

monthly\_sales['sale\_date'] = monthly\_sales['sale\_date'].dt.to\_timestamp()

# Plot

plt.figure(figsize=(12,6))

sns.lineplot(data=monthly\_sales, x='sale\_date', y='sales\_amount')

plt.title('Monthly Sales Over Time')

plt.xlabel('Month')

plt.ylabel('Total Sales')

plt.xticks(rotation=45)

plt.tight\_layout()

plt.show()

Identify top-selling products:

top\_products = full\_data.groupby('product\_name').agg({'sales\_amount': 'sum'}).sort\_values(by='sales\_amount', ascending=False).head(10)

print(top\_products)

**📊 Step 7: Visualize Key Insights**

Create a bar chart for top-selling products:

top\_products.plot(kind='bar', figsize=(10,6), legend=False)

plt.title('Top 10 Selling Products')

plt.xlabel('Product')

plt.ylabel('Total Sales')

plt.xticks(rotation=45)

plt.tight\_layout()

plt.show()

**💡 Step 8: Derive Business Insights**

Based on your analysis, you might uncover insights such as:

* **Seasonal Trends**: Identifying peak sales periods
* **Product Performance**: Determining which products contribute most to revenue
* **Customer Demographics**: Understanding which customer segments are most profitable
* **Store Performance**: Comparing sales across different store locations[Maven Analytics+2Maven Analytics+2Maven Analytics+2](https://mavenanalytics.io/project/20048?utm_source=chatgpt.com)

**📁 Step 9: Document and Share Your Findings**

Compile your analysis into a Jupyter Notebook or a report, including:

* **Introduction**: Purpose of the analysis
* **Methodology**: Steps taken during analysis
* **Findings**: Key insights with visualizations
* **Recommendations**: Suggestions based on the data

Consider sharing your project on platforms like GitHub or your personal portfolio to showcase your skills.

The **Global Electronics Retailer** dataset offers a wide range of analysis opportunities beyond basic sales tracking. Here’s a breakdown of **what else you can analyze**, categorized by business area:

**🛍️ Sales & Revenue**

* **Monthly/Quarterly Sales Trends**: Understand seasonality or sales cycles.
* **Sales by Product Category or Brand**: Identify which products drive revenue.
* **High vs Low Performing Stores**: Rank stores by total revenue or growth rate.
* **Sales by Region/Country**: Map sales geographically to spot strong/weak markets.
* **Average Order Value (AOV)** over time or by store/customer segment.

**👥 Customer Insights**

* **Customer Lifetime Value (CLV)**: Estimate total revenue per customer.
* **Customer Segmentation**: Group customers by purchase behavior or location.
* **Repeat Purchase Rates**: Understand loyalty and retention.
* **Top 10 Customers**: Rank by total spend or frequency.

**🛒 Product Performance**

* **Best-selling Products**: By volume and revenue.
* **Product Returns (if data available)**: Identify products with high return rates.
* **Inventory Turnover**: Based on product sales vs inventory levels.
* **Cross-sell/Upsell Opportunities**: What products are commonly bought together?

**🏬 Store/Location Analysis**

* **Footfall vs Sales (if available)**: Efficiency per store.
* **Store Growth Trends**: Which stores are improving or declining?
* **Currency Adjustment Impacts**: How do exchange rates affect sales in different countries?

**📉 Operational Efficiency**

* **Order Fulfillment Time** (if data available): Are there delays in shipping?
* **Stockouts and Overstocking Trends**: Optimize inventory management.
* **Profit Margins by Product or Region**: After currency conversion.

**💱 Currency & Pricing Insights**

* **Revenue Before & After FX Conversion**: True performance across markets.
* **Pricing Sensitivity**: If pricing is dynamic, how do changes affect sales?

**📊 Advanced Analyses (Optional if doing deeper work)**

* **Forecasting**: Use ARIMA or Prophet to forecast future sales.
* **Anomaly Detection**: Detect unexpected drops or spikes in sales.
* **Correlation Analysis**: Between discounting, sales, and customer behavior.

**Sales Performance**

* **Total Sales** over time (monthly/yearly).
* **Sales by Product Category/Subcategory**.
* **Sales by Region/Store/Country**.
* **Online vs Physical Store Performance** using square meter (null = online).

**🔸 Customer Analysis**

* **Customer Segmentation** based on:
  + Purchase frequency
  + Recency
  + Total spend
* **Customer Lifetime Value (CLV)**
* **Demographics** (e.g., Gender, Age from Birthday, Region)

**🔸 Product Insights**

* **Top-Selling Products & Brands**
* **Product Profitability**: (Unit Price - Unit Cost) \* Quantity
* **Color/Brand Preferences by Region**

**🔸 Store Performance**

* **Sales per Square Meter**
* **Store Age Impact**: Time since Open Date
* **Store Rankings by Revenue and Growth**

**🔸 Order Insights**

* **Average Order Value (AOV)** = Total Sales / Number of Orders
* **Order Delivery Time** (excluding NULL Delivery Date)
* **Trends in Quantity Ordered**

**📊 Key Findings: Monthly Sales Trend**

1. **Strong Growth Period (2018–2019):**
   * There was a consistent and significant upward trend in total sales from early 2018 through late 2019.
   * Monthly sales peaked in late 2019, reaching over **$2.5M**, indicating strong business momentum during this period.
2. **Seasonal and Periodic Spikes:**
   * Recurring spikes, particularly in **November and December**, suggest **seasonal sales surges**, possibly driven by year-end promotions or holiday demand.
3. **Sudden Declines:**
   * Sharp drops appear consistently after peak months, likely reflecting **post-holiday slowdowns** or fulfillment delays (e.g., delivery-related).
   * Notable dips in early 2020 align with potential **external disruptions** (e.g., COVID-19 pandemic), which may have impacted operations or consumer demand.
4. **Downward Trend Post-2020:**
   * From 2020 onward, monthly sales declined significantly and remained relatively low and volatile, indicating a **business contraction** or **reduced market activity**.
5. **Early Volatility (2016–2017):**
   * Sales between 2016 and 2017 show **irregular patterns with frequent fluctuations**, which may reflect an **early-stage business**, **data inconsistencies**, or **supply chain inefficiencies**.

**💡 Business Implications**

* **Seasonal campaigns** should be prioritized in Q4, as historical data shows strong responsiveness.
* Investigate the **2018–2019 growth period** to replicate successful strategies.
* Address volatility and underperformance post-2020 by analyzing external factors, product availability, or marketing effectiveness.
* Consider using **forecasting models** to anticipate future sales patterns and plan resources accordingly.