# Exercise

The dataset is for a retail store. It includes columns for date, product, quantity sold, unit price, and a region of sale. There might be some intentional 'dirty' data for cleaning purposes.

<https://github.com/AtulKadlag/Marathi_DataAnalyst/blob/main/retail_sales_data.csv>

download this dataset and save in your local python project directory.

**Exercises:**

This exercise is going to test your knowledge on following things in Python. This is good practice for your understanding of Python Data analyst, of whatever we learned till now.

**1. Data Cleaning**

* Load the dataset using Pandas.
* Identify and handle missing values.
* Find any inconsistencies in the data (e.g., negative prices or quantities) and correct them.
* Standardize the format of the date column.

**2. Data Manipulation**

* Create a new column 'Total Sales' which is the product of 'Quantity Sold' and 'Unit Price'.
* Group the data by 'Product' and calculate the total quantity sold for each product.
* Add a column 'Month' extracted from the 'Date' column.

**3. Data Visualization**

* Create a line plot showing the total sales over time.
* Generate a bar chart displaying total sales for each product.
* Create a scatter plot showing the relationship between quantity sold and unit price.

# Detailed Instructions to complete this exercise:

**1. Data Cleaning**

* **Load the Dataset**: Use **pandas.read\_csv()** to load the dataset into a DataFrame.
* **Handle Missing Values**: Check for missing values using **df.isnull().sum()**. You can fill missing values with an appropriate strategy, like mean or median for numerical columns and mode for categorical columns.
* **Correct Inconsistencies**: Look for negative values in 'Quantity Sold' and 'Unit Price'. These can be treated as data entry errors and converted to positive values.
* **Standardize Date Format**: Ensure all dates are in a consistent format. Use **pd.to\_datetime(df['Date'])** to convert the 'Date' column to a datetime format.

**2. Data Manipulation**

* **Create 'Total Sales' Column**: **df['Total Sales'] = df['Quantity Sold'] \* df['Unit Price']**.
* **Group by 'Product'**: Use **df.groupby('Product')['Quantity Sold'].sum()** to find the total quantity sold per product.
* **Extract 'Month' from 'Date'**: **df['Month'] = df['Date'].dt.month**.

**3. Data Visualization**

* **Line Plot for Total Sales Over Time**: Sort the data by date and use Matplotlib to plot 'Date' against 'Total Sales'.
* **Bar Chart for Total Sales by Product**: Group data by 'Product' and plot a bar chart of the total sales per product.
* **Scatter Plot for Quantity vs. Unit Price**: Use Matplotlib to create a scatter plot with 'Quantity Sold' on one axis and 'Unit Price' on the other.