

## Dataset

[https://docs.google.com/spreadsheets/d/1u3N8Ou2iX-rwmangt4FzGID\\_ryM\\_YJEi5T79mRPIVNY/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1u3N8Ou2iX-rwmangt4FzGID_ryM_YJEi5T79mRPIVNY/edit?usp=sharing)

### 1. Analyzing the Relationship Between Product Categories and Sales

#### Question:

- How can you calculate the total sales for each **product category** using **Pandas** and **NumPy**? What insights can you draw from the distribution of sales across categories?

#### Steps to Solve:

- Use **Pandas** to group data by the **category\_name** column and calculate the total sales by multiplying **quantity** and **price** columns.
- Use **NumPy** to perform aggregate calculations (like sum, mean) for each category.
- Use **Matplotlib** to visualize the total sales for each category.

### 2. Identifying Customers with High Lifetime Value (CLV)

#### Question:

- How can you identify and segment customers with the highest **Lifetime Value (CLV)**? Which demographic factors (e.g., age, gender) correlate most strongly with CLV?

#### Steps to Solve:

- **Pandas** to group customers by **customer\_id** and calculate **CLV** (sum of **total\_sales** per customer).
- Use **NumPy** to calculate summary statistics of CLV.
- Visualize customer CLV distribution and analyze correlations with demographic factors (**age**, **gender**).

### 3. Analyzing the Impact of Review Scores on Purchase Frequency

#### Question:

- Is there any relationship between the **review score** given by customers and the **frequency of purchases**? Can you visualize any trends or patterns using **Pandas**, **NumPy**, and **Matplotlib**?

#### Steps to Solve:

- Use **Pandas** to calculate the number of purchases per review score.
- Use **NumPy** to analyze any trends or correlations between review score and the number of purchases.
- Visualize the relationship using a scatter plot or line plot.

### Investigating the Relationship Between Age and Total Spending

#### Question:

- How can you calculate the **total spending** of customers based on their **age**? Is there a trend or pattern in spending behavior among different age groups?

#### Steps to Solve:

- Use **Pandas** to create age bins and group the data by these bins.
- Calculate total spending (**quantity \* price**) for each age group.
- Use **NumPy** for aggregate statistics and **Matplotlib** to visualize the results.

### Identifying Patterns in Payment Methods by Gender

#### Question:

- Do different genders use **payment methods** in different proportions? Can you visualize this data using **Pandas** and **Matplotlib** to identify any trends or patterns?

### Steps to Solve:

- Use **Pandas** to group data by **payment\_method** and **gender**, calculating the count of each payment method per gender.
- Use **Matplotlib** to visualize the results with a stacked bar chart or pie chart.