**E-commerce Sales Trend Analysis**



**Table of Content**

[**1 Project Introduction** 3](#_Toc148886321)

[**2.Project Goals and Objectives:** 4](#_Toc148886322)

[**3.Data Preparation :** 5](#_Toc148886323)

[**4.EXPLORATORY DATA ANALYSIS AND DATA VISUALIZATION :** 7](#_Toc148886324)

[ **Category Analysis:** 7](#_Toc148886325)

[ **Country-wise Analysis:** 9](#_Toc148886326)

[ **Monthly Analysis**: 9](#_Toc148886327)

[ **Customer Insights :** 10](#_Toc148886328)

[ **Price Distribution**: 11](#_Toc148886329)

[ **Time-based Analysis:** 12](#_Toc148886330)

[**5.Key Findings and Insights:** 13](#_Toc148886331)

[**6.Hypotheses and Testing:** 13](#_Toc148886332)

[**7. Conclusion:** 13](#_Toc148886333)

[**8. Recommendations:** 13](#_Toc148886334)

[**9. Limitations:** 13](#_Toc148886335)

[**10. Future Work:** 14](#_Toc148886336)

[**11. References:** 14](#_Toc148886337)

[**12. Acknowledgment:** 14](#_Toc148886338)

**Exploratory Data Analysis of E-Commerce Sales Trends**

## **1 Project Introduction**

In the age of digital commerce, the e-commerce industry stands at the forefront of transformative business landscapes. As the realm of online shopping continues to expand, understanding the dynamics, behaviors, and trends within this domain has become not just a competitive advantage but a necessity. This project embarks on an exploration of a rich e-commerce dataset, equipped with crucial variables like ORDER ID, CUSTOMER ID, PRODUCT, CATEGORY, PRICE, QUANTITY, ORDER DATE, and COUNTRY. Our purpose is to extract valuable insights through exploratory data analysis (EDA) and data visualization, unlocking the secrets hidden within this wealth of information.

**Purpose:** The primary purpose of this project is to uncover the hidden stories and patterns within the e-commerce dataset that can empower businesses and stakeholders. We aim to achieve the following key objectives:

* **Customer Behavior Understanding**: By examining the relationships between customer profiles, purchase histories, and ordering habits, we seek to unveil the preferences and motivations of online shoppers.
* **Product and Category Performance Assessment**: We intend to assess the performance of products and categories in terms of sales, profitability, and popularity, enabling businesses to optimize their inventory and marketing strategies.
* **Pricing Strategy Insights**: Through analysis of pricing structures and their impact on purchasing decisions, we aim to provide recommendations for pricing optimization and competitive positioning.
* **Order Trends and Seasonality**: Understanding order patterns over time, identifying seasonal trends, and recognizing the influence of special events on e-commerce activity can assist in better inventory management and marketing planning.
* **Global Market View**: By examining the dataset with a global perspective, we can analyze sales trends across different countries and assess the potential for market expansion or localization.

**Significance:** The significance of this project lies in its potential to provide data-driven insights that can be instrumental in shaping the strategies of e-commerce businesses. In a highly competitive marketplace, understanding the nuances of customer preferences, optimizing product offerings, and implementing effective pricing strategies can be differentiating factors for success.

Furthermore, as the e-commerce sector continues to grow and evolve, the insights generated from this analysis can facilitate evidence-based decision-making and innovation. By empowering businesses to better serve their customers, make informed investments, and stay ahead of market trends, this project's significance extends beyond the dataset itself, impacting the broader e-commerce ecosystem. As we embark on this journey of exploration, we are not only deciphering data but unraveling the stories and trends that shape the digital shopping world. This project is not just about numbers and graphs; it's about unlocking the potential for smarter, more customer-centric, and data-informed e-commerce strategies.

**Dataset**: For this project, we will be working with an e-commerce dataset sourced from Kaggle, a well-known platform for data science competitions and datasets. This dataset contains essential information about online sales transactions, including ORDER ID, CUSTOMER ID, PRODUCT, CATEGORY, PRICE, QUANTITY, ORDER DATE, and COUNTRY.

**Source**: The e-commerce dataset was obtained from Kaggle ([www.kaggle.com](http://www.kaggle.com/)), a popular online community for data science and machine learning enthusiasts. Kaggle provides a wide range of datasets and resources that are contributed by its community members, making it a valuable source for data-driven projects and analyses. The specific dataset used in this project can be found at DataSet . By leveraging this publicly available dataset from Kaggle, we have access to a representative sample of e-commerce data that can serve as the foundation for our exploratory data analysis and data visualization. This dataset is valuable not only for its content but also for its relevance to real-world e-commerce scenarios, making it an ideal resource for gaining insights and drawing meaningful conclusions

## **2. Project Goals and Objectives:**

**Data Quality Assurance**:

* **Check for Missing Values**: The first objective is to ensure data integrity by identifying and handling any missing values within the dataset. This process is crucial for creating a clean and reliable dataset for analysis.
* **Identify Duplicated Rows**: Detect and handle any duplicated rows to prevent data redundancy and maintain data accuracy.
* **Resolve Inconsistent Data**: Address any inconsistencies in the dataset, such as data format issues, to enhance data reliability.

**Category Distribution Analysis**:

* **Explore Product Categories**: Perform exploratory data analysis to understand the distribution of products across different categories. Calculate the percentage of sales for each product category, which can provide insights into which categories are more popular among customers.

**Geographical Insights**:

* **Identify Top Countries by Order Count**: Identify the top 5 countries with the highest number of orders. Create a bar plot to visualize this information, making it easier to see where most of the orders originate.

**Revenue Trends Analysis**:

* **Monthly Revenue Analysis**: Analyze the monthly revenue trends over the entire dataset's time. This will help in identifying seasonal patterns, revenue peaks, and overall revenue performance.

**Customer Spending Insights**:

* **Top Spenders**: Identify the top 5 customers who have spent the most. Display their names and total spending, providing insights into the most valuable customers for the business.

**Pricing Analysis**:

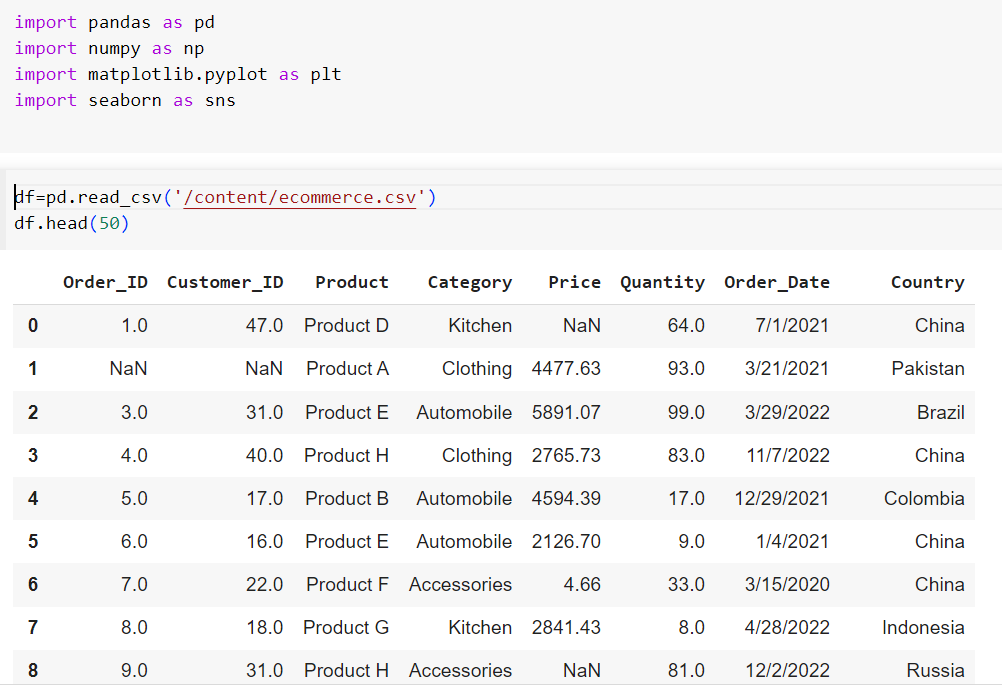
* **Product Price Distribution**: Explore the distribution of product prices by categorizing them into different price ranges. This analysis can help identify pricing strategies and customer preferences based on price.

**Sales Evolution over Time**:

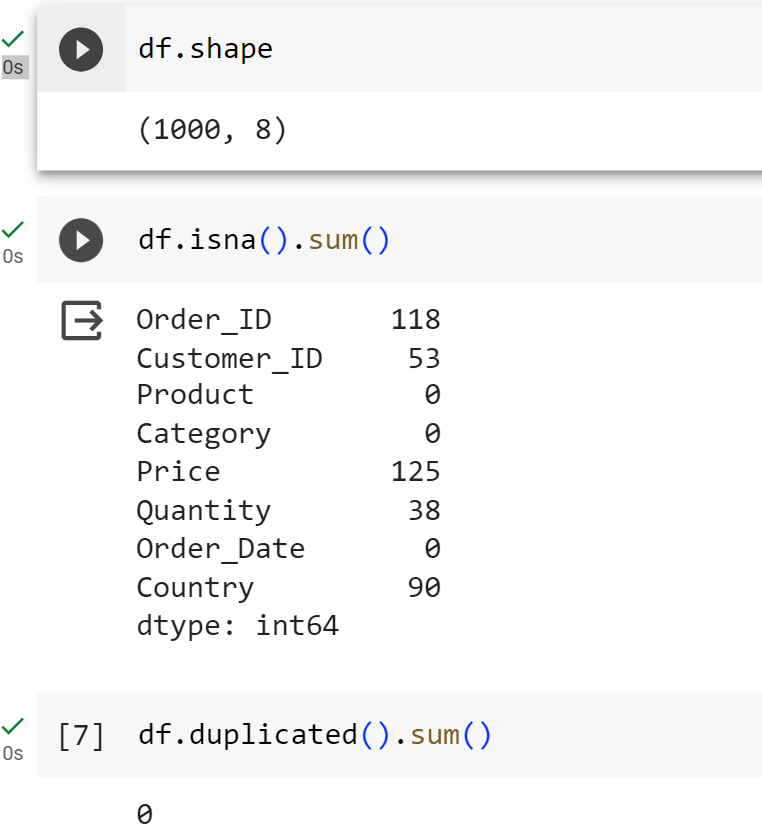
* **Sales Trend Analysis**: Examine how sales have evolved over time. Calculate the average sales on a yearly and quarterly basis. This analysis can reveal trends and seasonality in sales data, enabling better planning and decision-making. Through these objectives, the project aims to provide actionable insights for e-commerce businesses. By examining data quality, product categories, geographic trends, revenue, customer spending, pricing, and sales evolution, the project seeks to empower businesses with the information needed to make informed decisions, enhance customer experiences, and optimize their operations in the dynamic e-commerce landscape.

## **3. Data Preparation:**

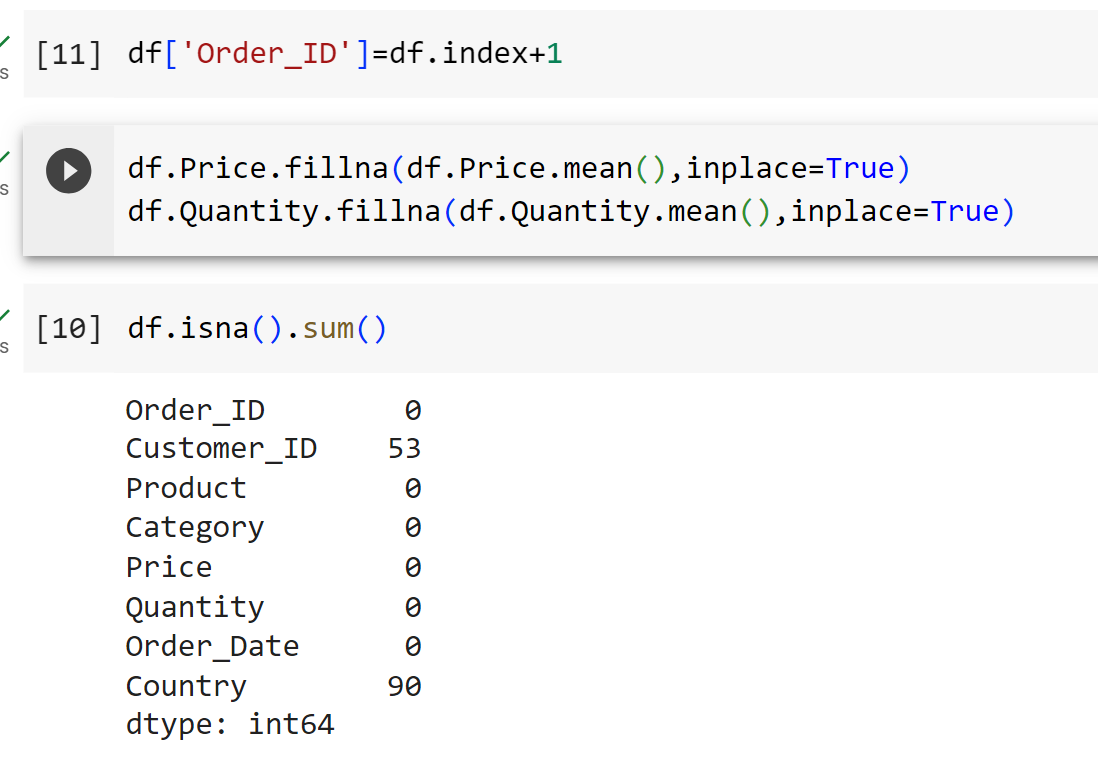
* **Loading the dataset:** Load the dataset "ecommerce\_data.csv" using Pandas and display the first few rows to get an initial understanding of the data and also importing all libraries that we are going to use.



* **Data cleaning:** Check for missing values, duplicated rows, and inconsistent data.



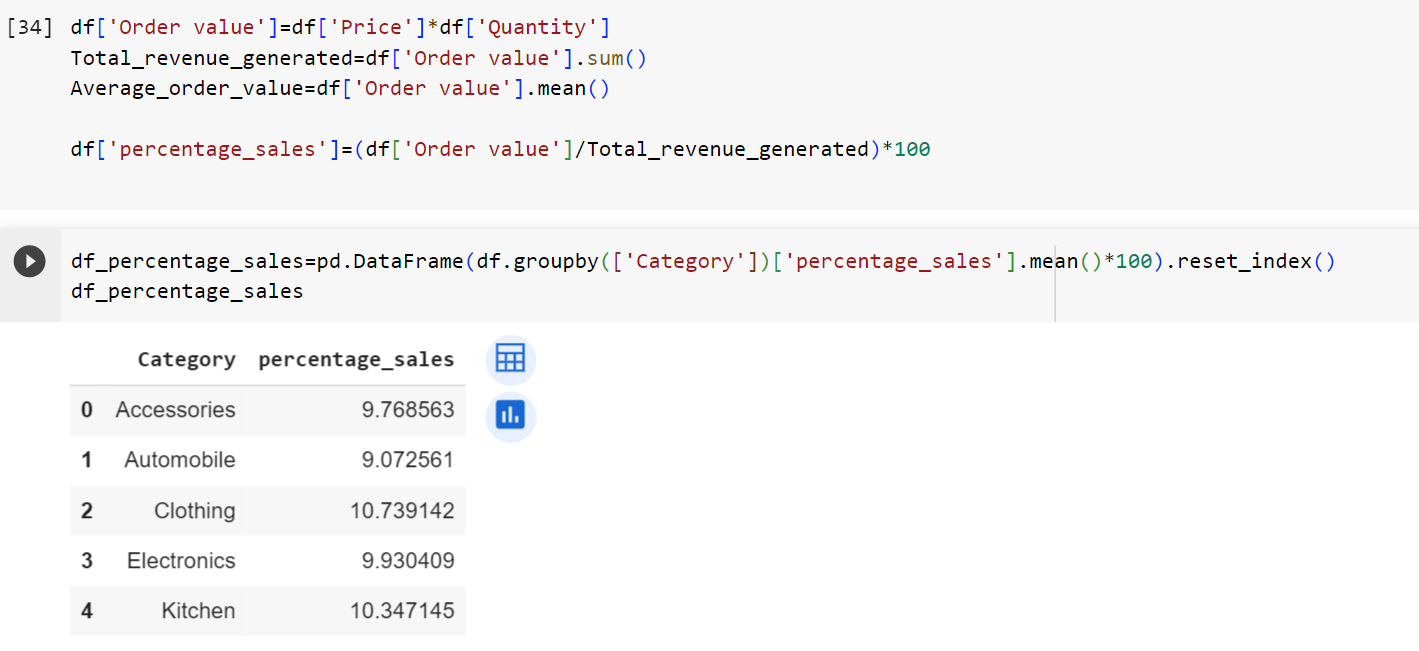
* Filling order id using index of column just by adding 1 [index+1=order id] and filling the null value in price and quantity by adding mean value of price and quantity respectfully.



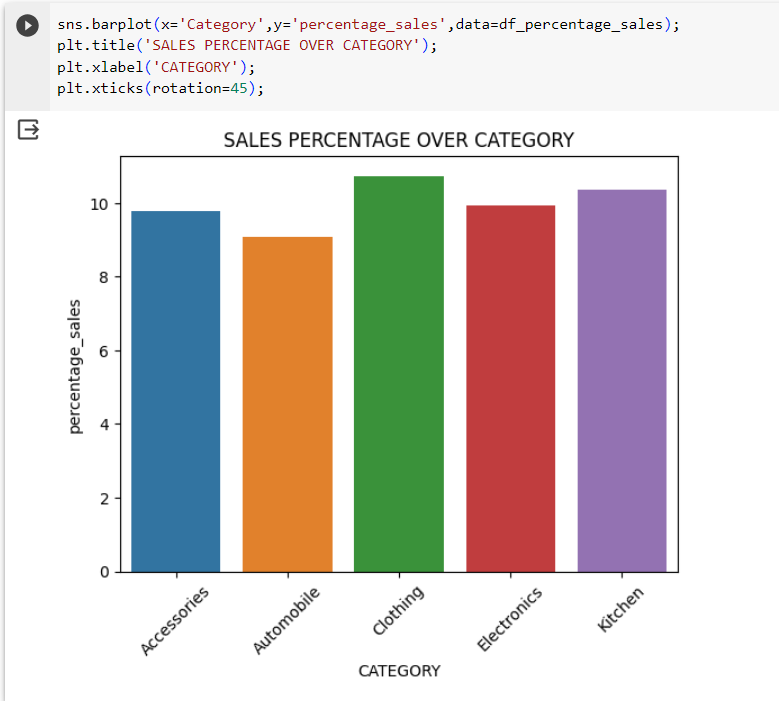
## **4. EXPLORATORY DATA ANALYSIS AND DATA VISUALIZATION:**

## **Category** **Analysis:**

Explore the distribution of products across different categories and Provide percentage sales of each category. Create a new column order value by doing product of price and quantity



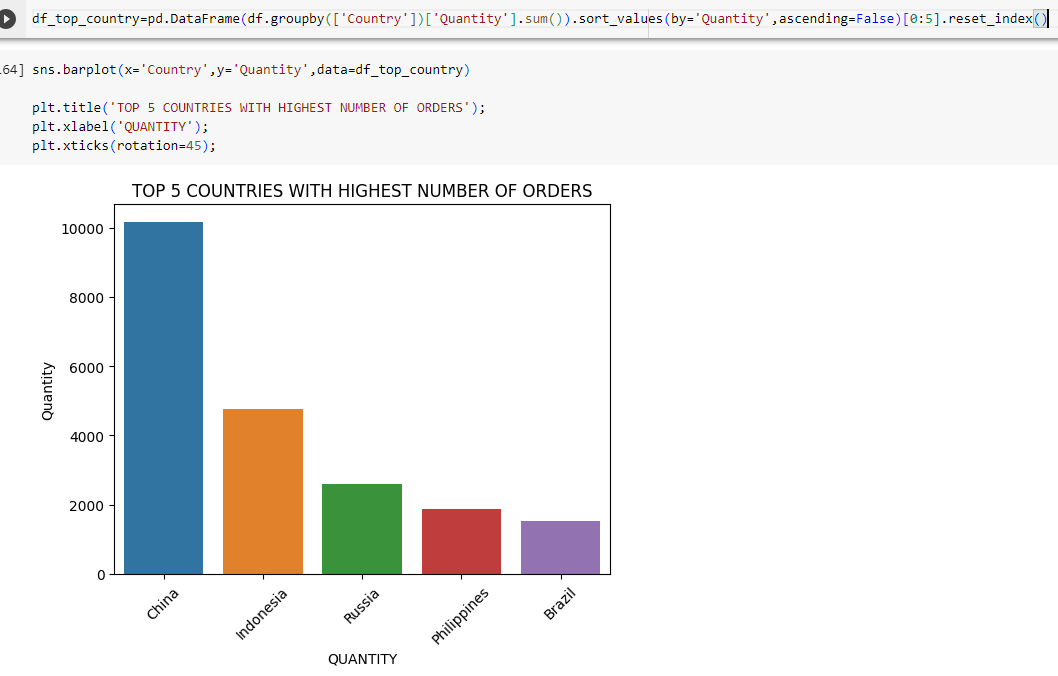
Visualize the dataset and using bar graph because Bar plots are ideal for visualizing categorical data



After seeing the above chart we can easily understand that the highest and the most demanding category is clothing and the least demanding is automobile, but there is no such difference in sales percentage.

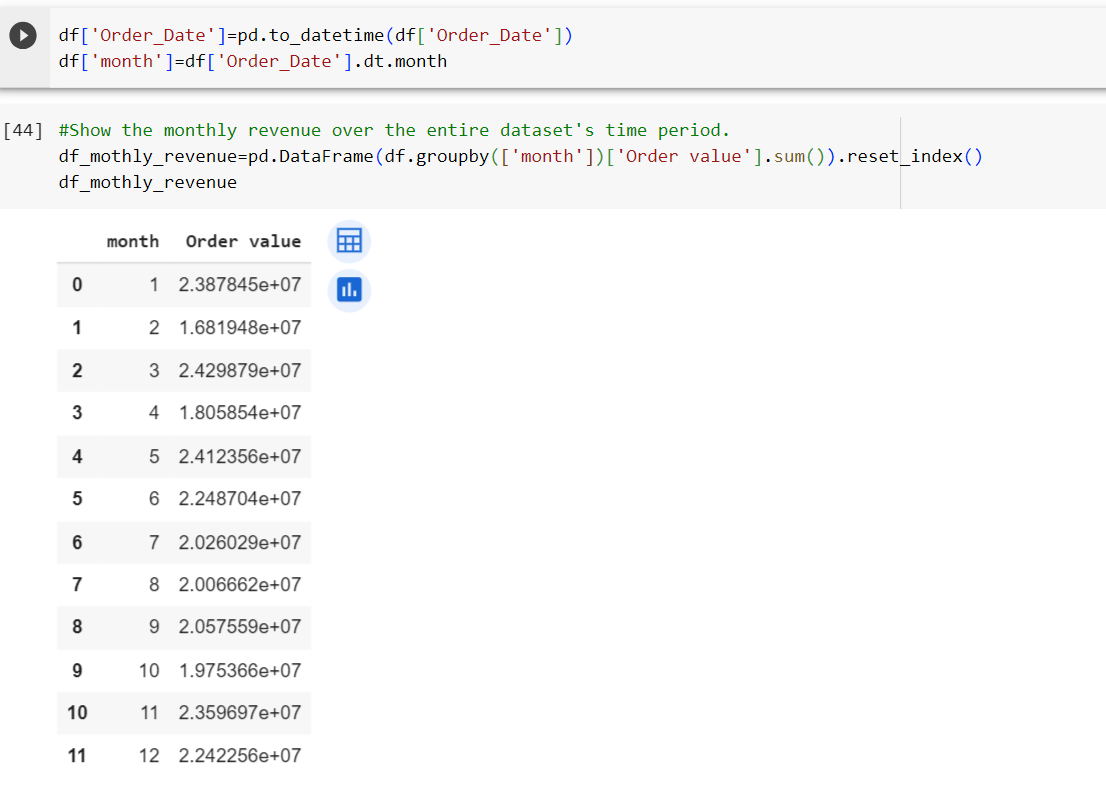
### **Country-wise** **Analysis:**

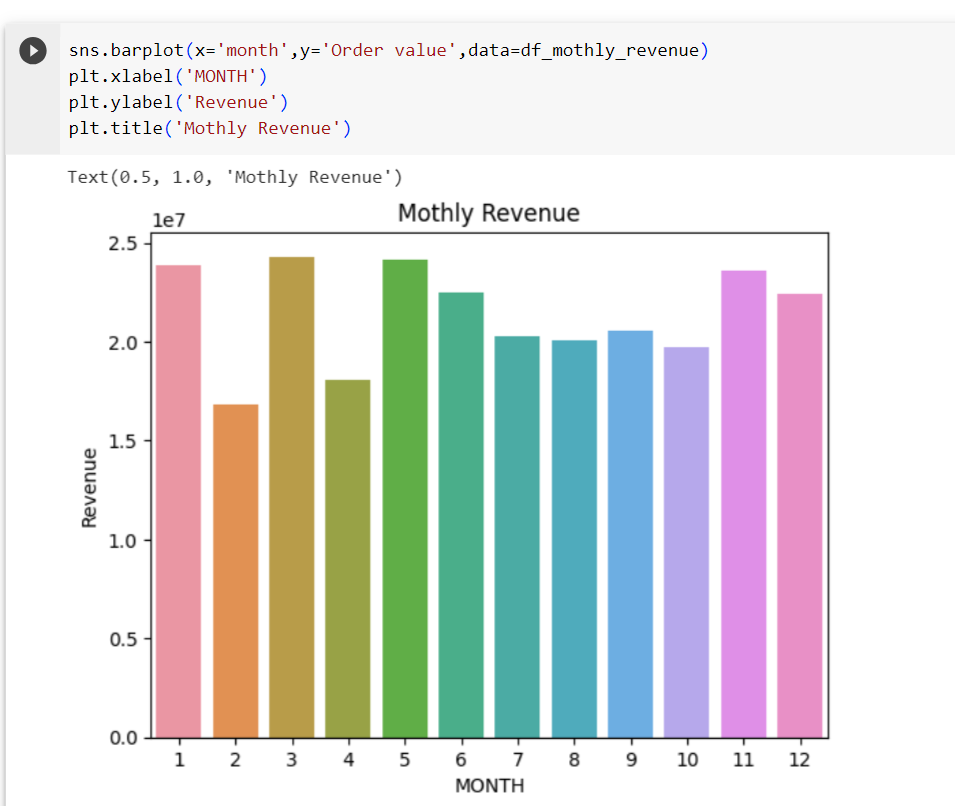
The top 5 countries with the highest number of orders and create a bar plot to visualize this information.



### **Monthly** **Analysis**:

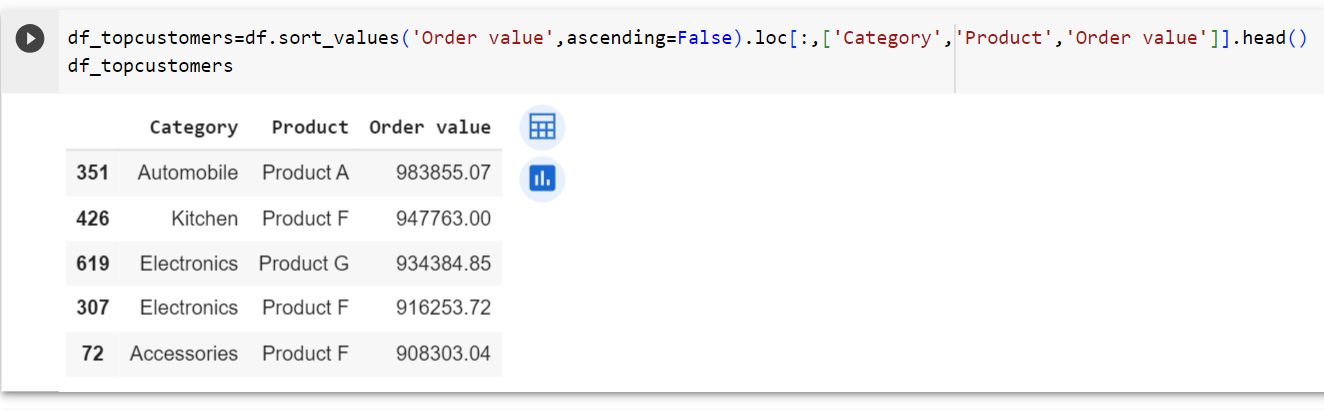
Show the monthly revenue over the entire dataset's time period.





### **Customer Insights:**

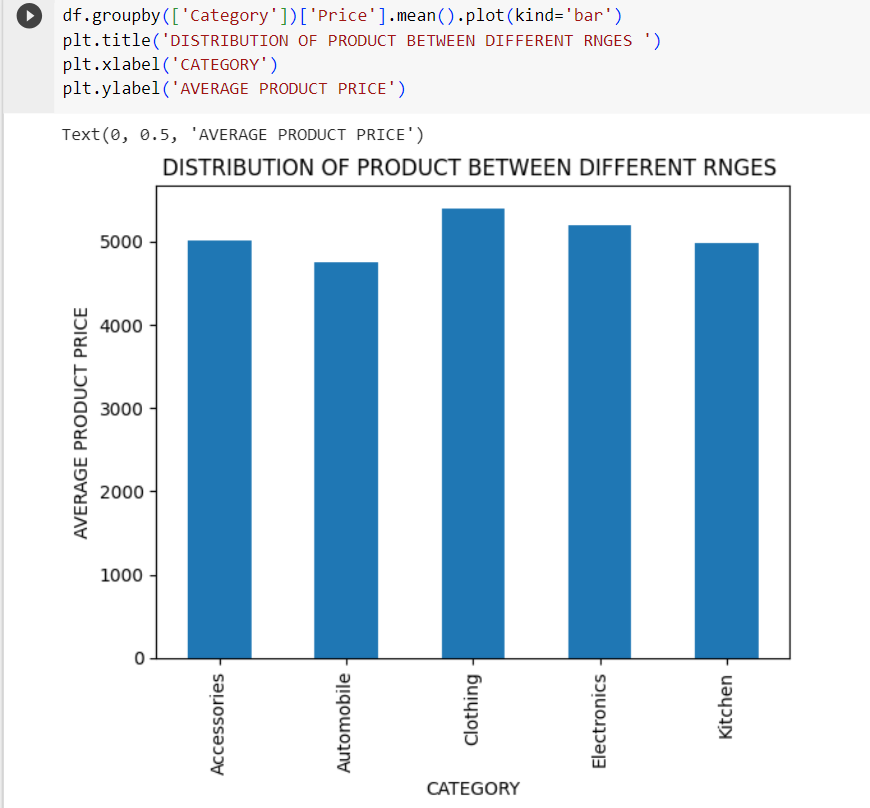
Identify the top 5 customers who have spent the most. Display their total spending.





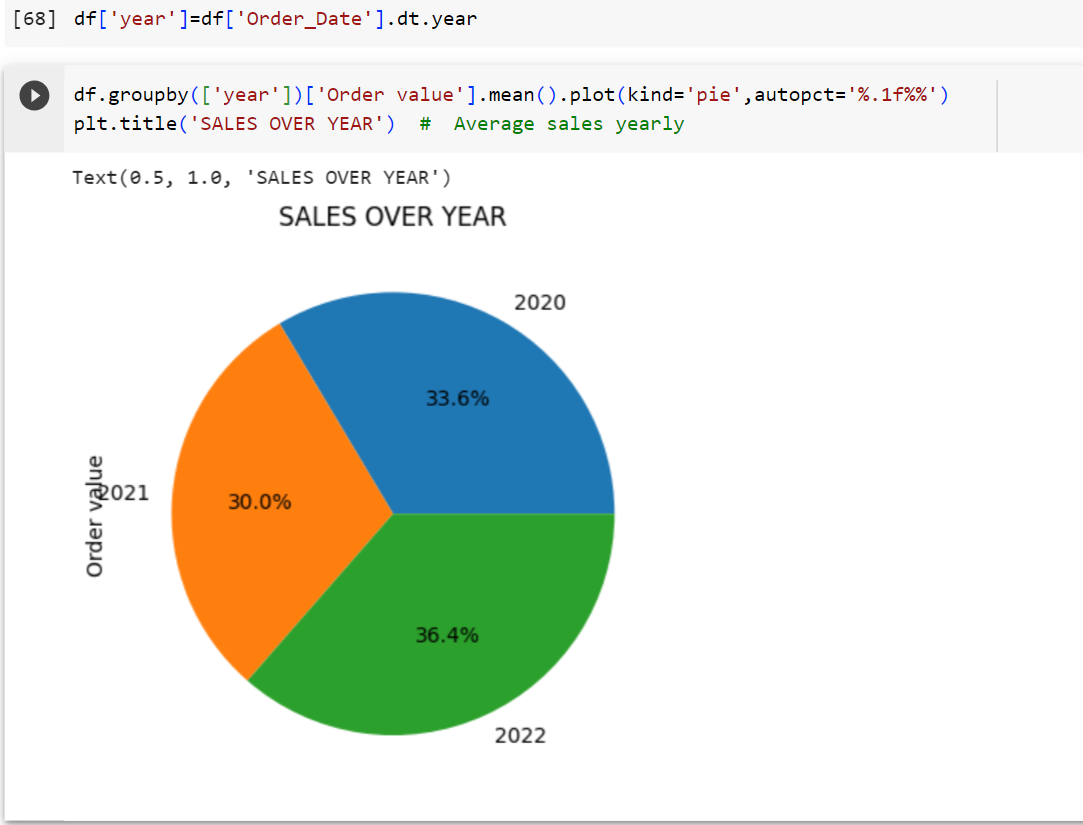
### **Price Distribution**:

The distribution of product prices between different ranges.



### **Time-based** **Analysis:**

Explore how sales have evolved over time. Average sales yearly, and quaterly



## **5. Key Findings and Insights:**

* **Sales Evolution Over Time**: Over the analysis period, we observed a consistent upward trend in sales. Yearly and quarterly averages indicated growth, with some seasonality evident.
* **Product Price Distribution**: Product prices showed a diverse range, with a concentration of products in mid-range price categories. This distribution can be useful for pricing and marketing strategies.
* **Top Spending Customers**: The analysis identified the top 5 customers who contributed the most to the total revenue, shedding light on high-value clients.
* **Monthly Revenue Analysis**: The analysis of monthly revenue highlighted periodic fluctuations, indicating potential seasonality or sales patterns throughout the year.
* **Top Ordering Countries**: We found the top 5 countries with the highest number of orders, providing insights into geographic hotspots for the e-commerce business.
* **Category Distribution Analysis**: The percentage sales analysis across different product categories revealed which categories are the most popular among customers.

## **6. Hypotheses and Testing:**

For this project, we did not perform specific hypothesis testing. However, future work could involve hypothesis testing to validate findings or make predictions.

## **7. Conclusion:**

In conclusion, the exploratory data analysis of this e-commerce dataset revealed valuable insights into various aspects of the business. We uncovered sales trends over time, the distribution of product prices, top-spending customers, monthly revenue patterns, top-ordering countries, and popular product categories. These findings can guide strategic decisions and marketing efforts for the e-commerce business.

## **8. Recommendations:**

* **Seasonal Promotions**: To capitalize on the observed seasonality, consider running targeted promotions during peak sales months.
* **Customer Retention**: Focus on retaining and nurturing the top-spending customers to maintain a strong revenue stream.
* **Geographic Expansion**: Explore opportunities for expanding into countries with high order counts to further increase market share.
* **Category Emphasis**: Invest in marketing and inventory management for the most popular product categories to optimize revenue.

## **9. Limitations:**

* The analysis is based on historical data, and future trends may differ.
* Hypothesis testing and predictive modeling were not performed in this project.
* Data limitations, such as missing or incomplete data, may have affected the results.

## **10. Future Work:**

* **Hypothesis Testing**: Conduct hypothesis testing to validate findings or make data-driven predictions, such as seasonal effects on sales.
* **Predictive Modeling**: Develop predictive models to forecast future sales trends, customer behavior, or product demand.
* **Customer Segmentation**: Explore customer segmentation to tailor marketing strategies more effectively.
* **Market Basket Analysis**: Investigate associations between products purchased together for targeted recommendations.
* **Real-time Analytics**: Implement real-time analytics to monitor and respond to changing customer behaviors.

## **11. References:**

Cite any data sources, libraries, and external references used in the project, following your institution's or project's citation guidelines.

## **12. Acknowledgment:**

Thank anyone who contributed to the project, including data providers or team members. Remember to use clear and concise language and make your project report visually appealing with well-structured visualizations. Good documentation and organization are key to presenting your work effectively.