**Data Analysis Project**

Market Basket Analysis

**Introduction:** This project aims to analyze the order data of an online store using data analysis tools: **seaborn**, **Pandas**, **mlxtend** to understand customer behavior and orders.

**Data used:** Includes several databases and data imported from CSV:

* **aisles.csv**: Contains information about the aisles in the store.
* **departments.csv**: Displays the different sections in the store.
* **order\_products\_prior.csv**: Contains information about previously ordered products.
* **order\_products\_train.csv**: Training application data.
* **orders.csv**: Information about requests.
* **products.csv**: Product details.

**Analysis steps:**

1. *Data collection:*

Import order, product, and department data from CSV files.

1. *Data exploration:*

We used **head()** and **info()** to examine the data and understand its structure.

1. *Data cleaning:*  Remove missing or duplicate values ​​using **drop\_duplicates()** and **fillna().**
2. *Application of the* ***FP- Growth*** *algorithm:*

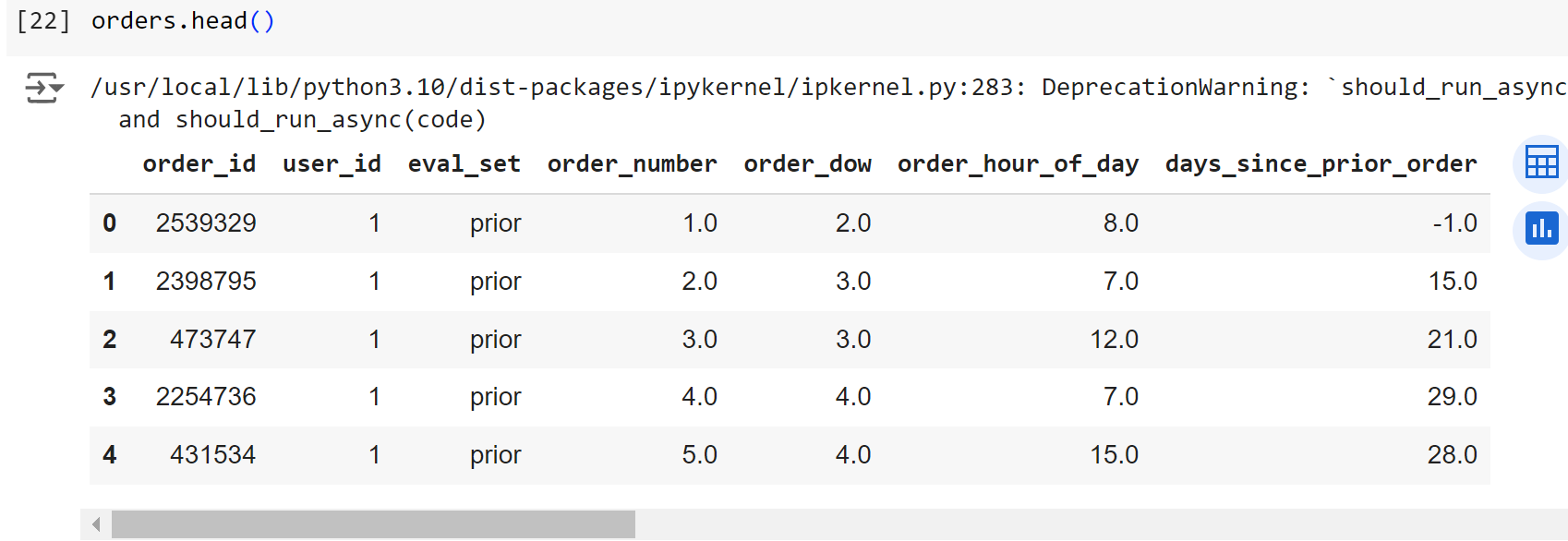
Used to extract recurring patterns from large data sets.

1. *Visualization:*

Use libraries like **Matplotlib** and **Seaborn** to plot graphs to better understand the results.

**WHAT WE DO:**

We set -1 for missing values ​​and the result was:

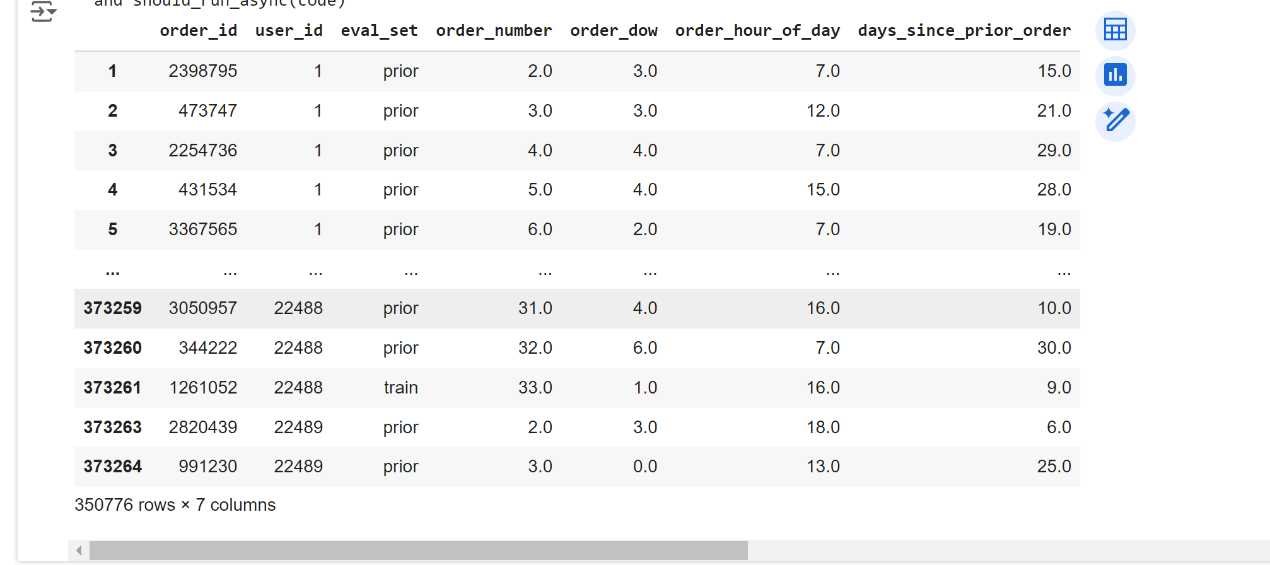


We calculated the number of products in each section to find out which section contains the largest number of products:

صورة تحتوي على نص, لقطة شاشة, تخطيط, خط

تم إنشاء الوصف تلقائياً

We excluded the first customer request to understand customer behavior more accurately:



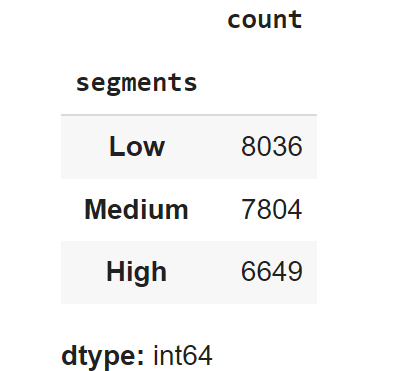
We analyzed the number of customer requests and calculated the mean requests:

صورة تحتوي على نص, الخط, لقطة شاشة, أبيض

تم إنشاء الوصف تلقائياًصورة تحتوي على نص, لقطة شاشة, رقم, الخط

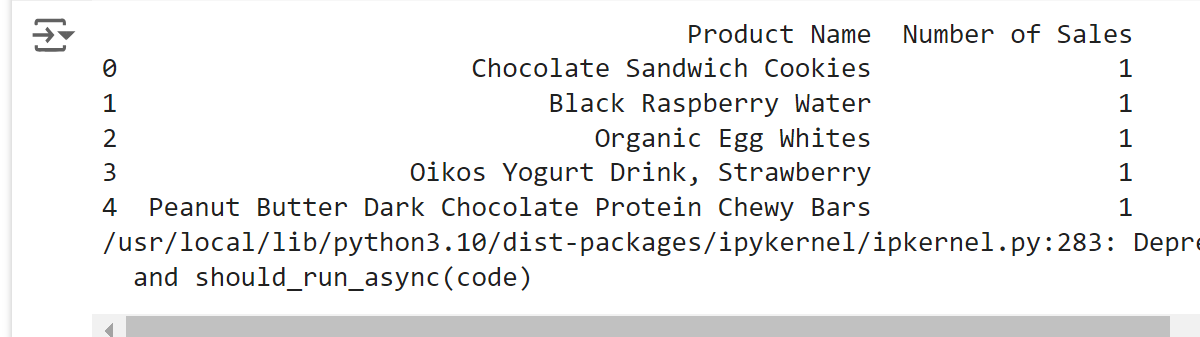
تم إنشاء الوصف تلقائياً

We put the word high if the number of requests is high, medium if the number is medium, and low if it is low, and we calculated the number of each of them:

صورة تحتوي على نص, لقطة شاشة, رقم

تم إنشاء الوصف تلقائياً

We have identified the five best-selling products and the five least-selling products:

صورة تحتوي على نص, الخط, خط, لقطة شاشة

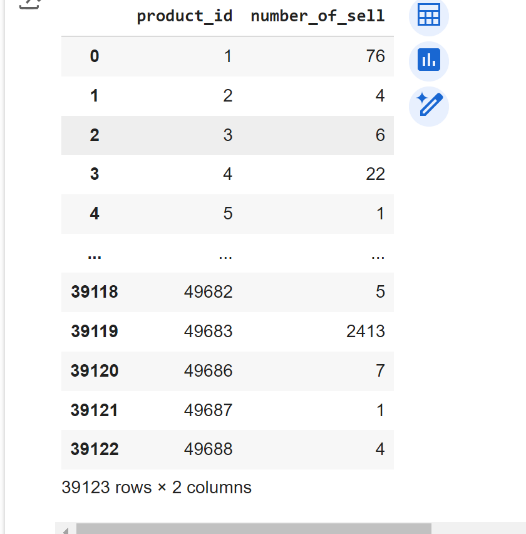
تم إنشاء الوصف تلقائياً

We analyzed the number of requests for each day of the week and represented the number 0 as Sunday and the number 6 as Saturday:

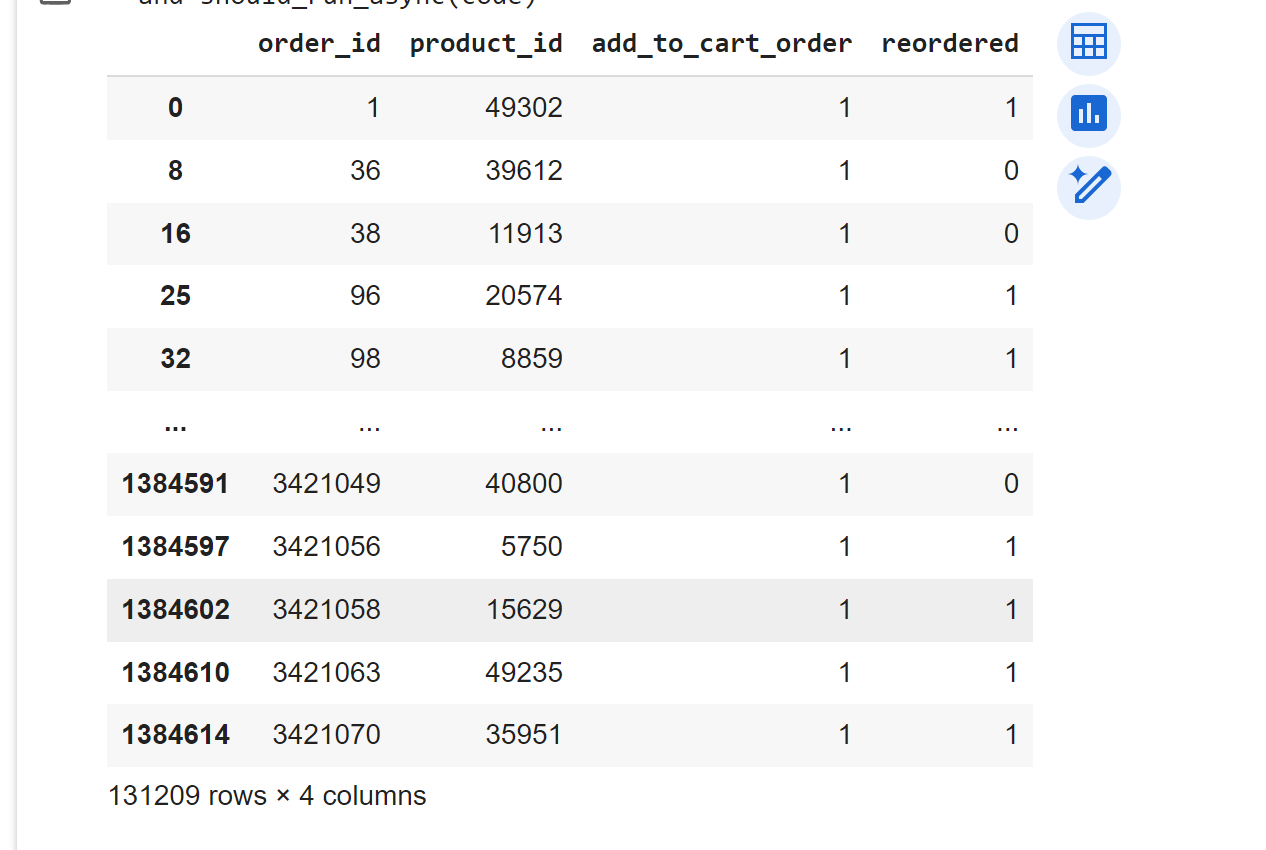
صورة تحتوي على نص, لقطة شاشة, الخط, رقم

تم إنشاء الوصف تلقائياً

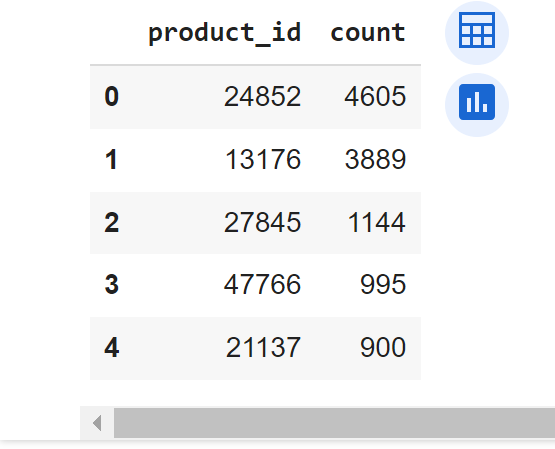
We calculated the number of sales for each product to determine the best-selling products:



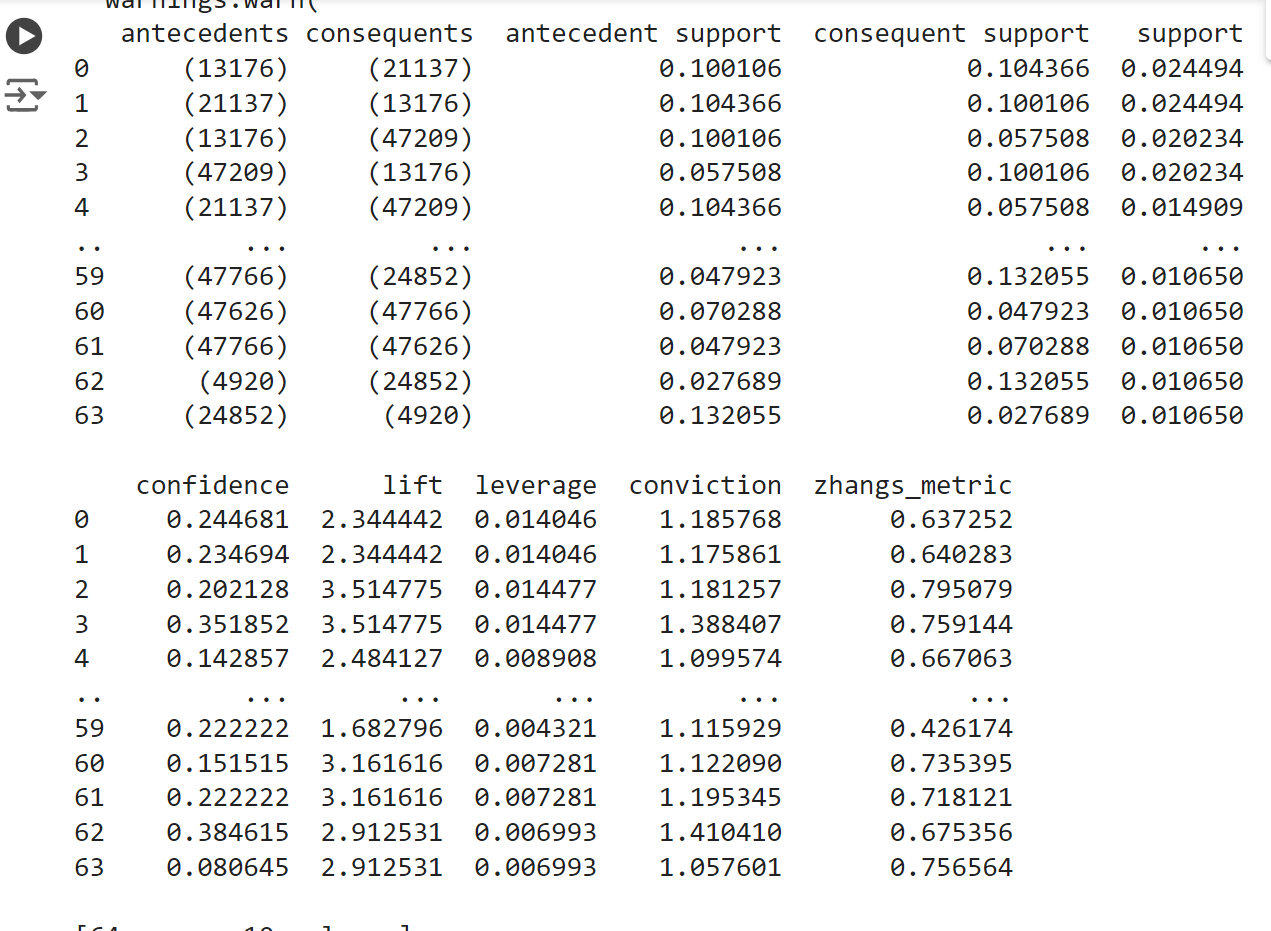
We have identified the first item added to the shopping cart:



We identified the top five products added as the first product in the cart and how many times they were ordered to find out which product is the most popular with customers:



We analyzed the shopping cart data using FP- GROWTH:



**Conclusion**

At the end of this project, we were able to analyze the order data in the online store using different data analytics tools. By exploring and cleaning the data, we were able to understand customer behavior and identify recurring patterns in purchases. The FP-Growth algorithm helped us extract popular product groups, providing valuable insights to improve marketing strategies and enhance customer experience. We hope that these results will contribute to enhancing store performance and increasing customer satisfaction in the future.