Dataframe Practice

Mridha Md. Nafis Fuad

•

4:09 PM

Due Nov 13

- Load the dataset into a DataFrame and display the first 5 rows.
- Show general info (.info()) and summary statistics (.describe()).
- List all unique values in Region, Country, Item Type, Sales Channel, and Order Priority.
- Count the total number of sales records.
- Filter all orders where Region is “Sub-Saharan Africa” and Sales Channel is “Online”.
- Display only Country, Item Type, and Units Sold for those filtered rows.
- Find all high-priority (Order Priority == 'H') orders with Units Sold greater than 5000.
- Convert Order Date and Ship Date to datetime format.
- Filter all orders where the shipping delay (Ship Date - Order Date) is more than 10 days.
- Create a new column Total Revenue = Units Sold × Unit Price.
- Create another column Total Cost = Units Sold × Unit Cost.
- Create a column Profit = Total Revenue – Total Cost.

- Create a column Profit Margin (%) = (Profit / Total Revenue) × 100, rounded to 2 decimals.
- Compute total revenue per Region.
- Compute average profit per Item Type.
- Find the most sold Item Type (by total Units Sold).
- Find the top 3 Country values by total Profit.
- For each Sales Channel, compute average Units Sold, total Revenue, and total Profit.
- Sort all records by Order Date in ascending order and show the first and last 10 rows.
- Extract Year and Month from Order Date into new columns.
- Group by Year and find total Revenue and total Units Sold per year.
- Find which month (across all years) had the highest average Profit.
- Calculate the average delivery time (Ship Date - Order Date) by Region.
- Identify the most profitable Item Type in each Region.
- Determine which Order Priority has the highest average Profit Margin.

- Compare average delivery delay and profit between Online and Offline channels.
- Export the final processed DataFrame (with new columns) to sales_summary.csv.