

PySpark Assignment-2

Objective: The purpose of this project is to analyze sales data from TATA NEU using pyspark. You are provided with two datasets, TATA_TB1 and TATA_TB2, which contain detailed sales transactions. Your task is to write pyspark queries to extract meaningful insights, perform data aggregation, and generate reports based on the datasets.

Questions to be answered:

- 1) Write a query to calculate the total records in two tables?
- 2) Write a query to calculate the total unique count of customers?
- 3) Write a query to fetch the oldest order date and latest order date?
- 4) Write query to get unique years?
- 5) Write a query to get the no. of regions and display the region names?
- 6) Write a query to get the no. of countries and display the country names?
- 7) Write a query to get the no. of states and display the state names?
- 8) Write a query to get the no. of cities and display the city names?
- 9) Write a query to calculate the total count of products?
- 10) Write a query to calculate total sales, total profit and total order quantity?
- 11) Write a query to calculate the total sales amount, total order quantity for each category.
Display the category, total sales, and total order qty and order by total sales from highest to lowest?
- 12) Write a query to calculate the total profit amount for each category. Display the category, total profit, and total order qty and order by total profit from highest to lowest?
- 13) Write a query to fetch the subcategories where total sales are greater than 100000?
- 14) Write a query to fetch the products where total profit is greater than 2500 and sort it based on profit from highest to lowest??
- 15) Write a query to get the total sales and total profit for Office Supplies category?
- 16) Write a query to get the total sales and total profit for Furniture category and Tables, Bookcases sub-categories ?
- 17) Write a query to get the total sales and total profit for Technology category and the Accessories, Copiers, Phones sub-categories ?

- 18) Write a query to get total sales and total profit by Region, Segment and sort the sales from highest to lowest??
- 19) Write a query to get total sales and total profit by Country, State and city and sort the sales from highest to lowest??
- 20) Write a query to get total sales and total orderqty by CustomerName sort it by total sales from highest to lowest?
- 21) Identify the top 5 products with the highest sales (by sales amount). Show the product name, total sales, and total qty?
- 22) Write a query to get total sales by City having sales greater than 35000?
- 23) Write a query to get total sales by CustomerName having sales greater than 10000?
- 24) Write a query to get total sales and total profit by shipmode and sort the sales in ascending order and profit in descending order?
- 25) Write a query to get total sales for North and central region?
- 26) Write a query to get total sales and total profit for Italy and Spain countries?
- 27) Write a query to get the total sales and total profit for each year sort the sales from highest to lowest?
- 28) Find the top 10 customers who spent the most across all transactions. Display the customer name, total amount spent, and number of orders placed?
- 29) Write a query to find which products are most preferred by customers based on the total sales. Display customer name, favorite product9 (top 3 products per each customer), and total sales on that product?
- 30) Write a query to get 7th rank customer name based on total sales? Display customer name, sales amount and rank.
- 31) Write query to get total sales and total profit in years 2011 and 2013 Display year wise total sales and total profit?
- 32) Write a query to get total sales, total profit and total order qty by country, state, category and sub-category and sort it from highest to lowest based on sales column?

- 33) write a function to get the region sales dynamically, if we pass any region, that region sales should be displayed?
- 34) write a function to get the country sales, profit dynamically, if we pass any country, that country sales, profit should be displayed?
- 35) write a function to get the Category orderquantity dynamically, if we pass any Category, that Category orderquantity should be displayed.?
- 36) write a function to get the City sales and profit dynamically, if we pass any city, that City sales and profit should be displayed?