Case Study on Superstore Sales using MapReduce Streaming

Dataset Description

Column Name	Description
Invoice ID	Unique identifier for each transaction.
Branch	The branch where the sale occurred.
City	The city where the branch is located.
Customer type	Indicates whether the customer is a member or a normal customer.
Gender	Gender of the customer.
Product line	Category of the product purchased.
Unit price	Price per unit of the product.
Quantity	Number of units purchased.
Tax 5%	Tax applied to the purchase.
Total	Total price including tax.
Date	Date of the transaction.
Time	Time of the transaction.
Payment	Mode of payment.
COGS	Cost of goods sold.
Gross margin percentage	Gross margin percentage for the sale.
Gross income	Gross income from the sale.
Rating	Customer's rating of the purchase.

You are been given a dataset of supermarket sales. Write a MapReduce Streaming job to compute the below questions.

1. You are given a dataset of supermarket sales. Write a MapReduce Streaming job to compute the total sales revenue generated by each branch.
2. Write a MapReduce Streaming job to find the average customer rating per product line.
3. Create a MapReduce Streaming job to determine the most popular payment method in each city.
4. Write a MapReduce Streaming job to calculate the total quantity of products sold per product line across all branches.
5. Design a MapReduce Streaming job to determine the most frequently purchased product line by each customer type.
6. Write a MapReduce Streaming job to find out which day of the week has the highest sales revenue.
7. Create a MapReduce Streaming job to find the product line with the highest sales revenue over the entire dataset.
8. Design a MapReduce Streaming job to identify trending products by calculating the increase in sales quantity week-over-week for each product line.