Attached are the global sales and returns data.

The requirement is to create a report:

1. That contains the aggregate (sum) Profit and Quantity, based on Year, Month, Category, Sub Category.
   So your report file should contain:
   Year, Month, Category, Sub Category, Total Quantity Sold, Total Profit

1. This data has to be stored in a partition based on year month. like year=2014/month=11

2. For the total profit and total quantity calculations the returns data should be used to exclude all returns

3. There must be at least one unit test case

4. This program must be runnable with any input directories as parameters to read the sales and returns data

Try to write optimized spark operations.