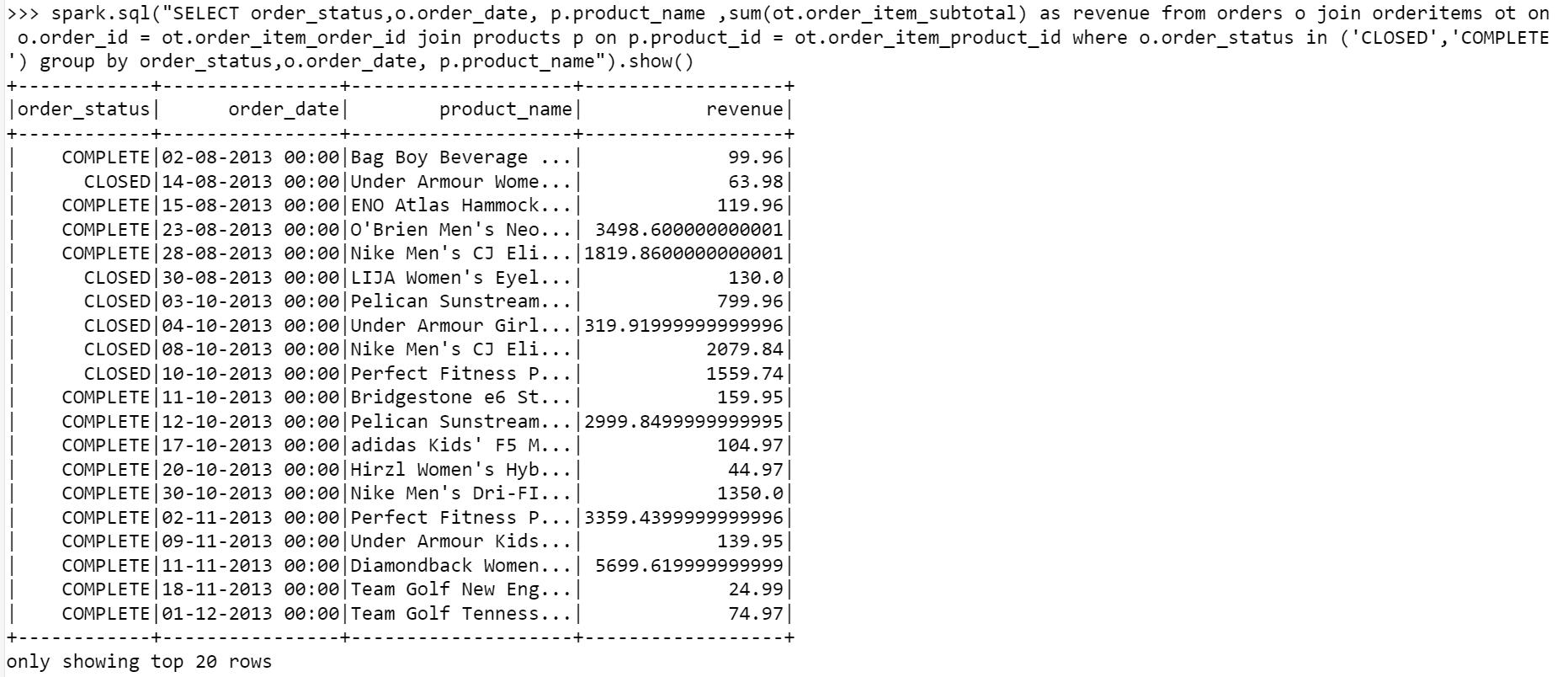
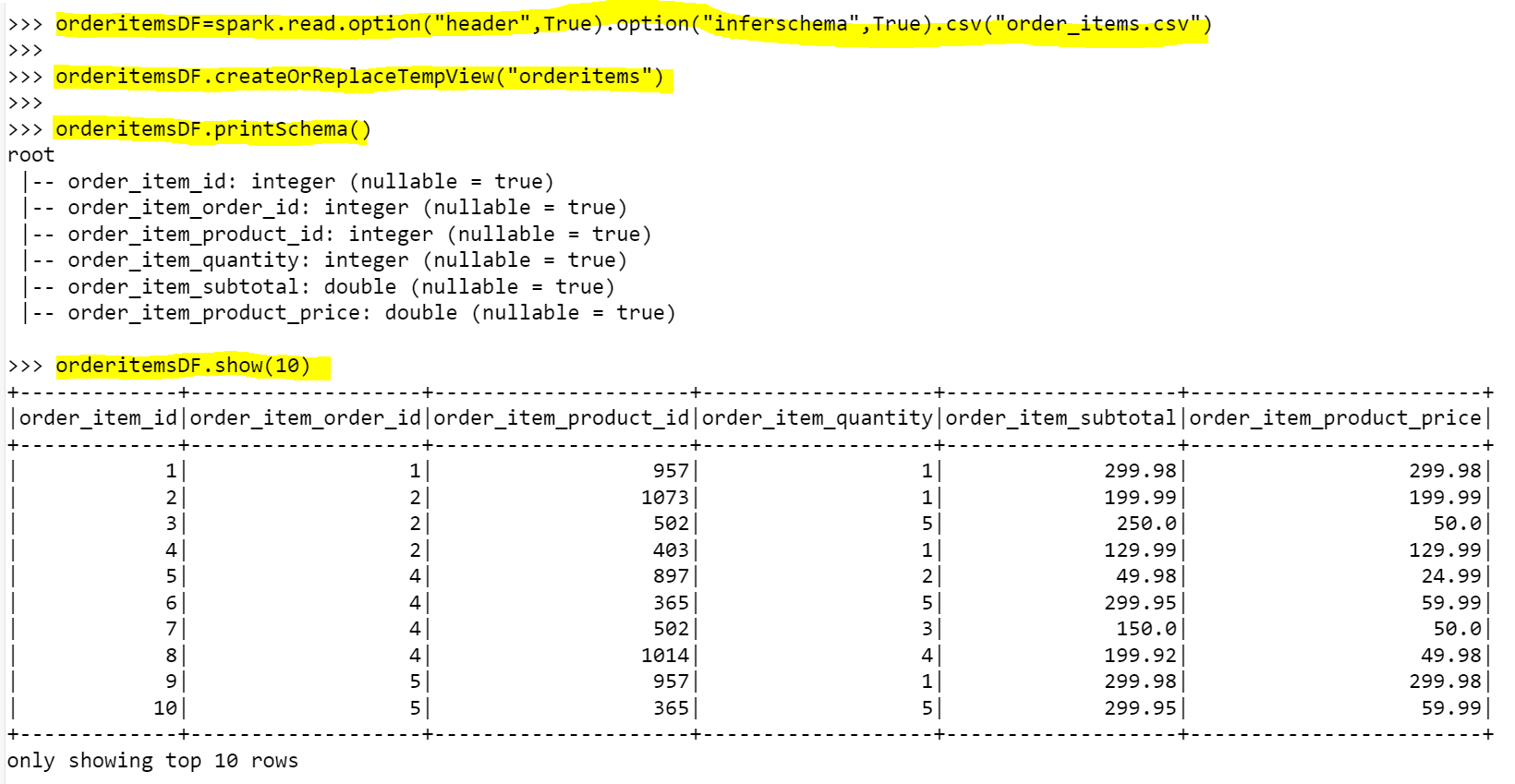
**Pyspark Assignment**

1. **Problem Statement: What is the daily product revenue for CLOSED or COMPLETE orders.**

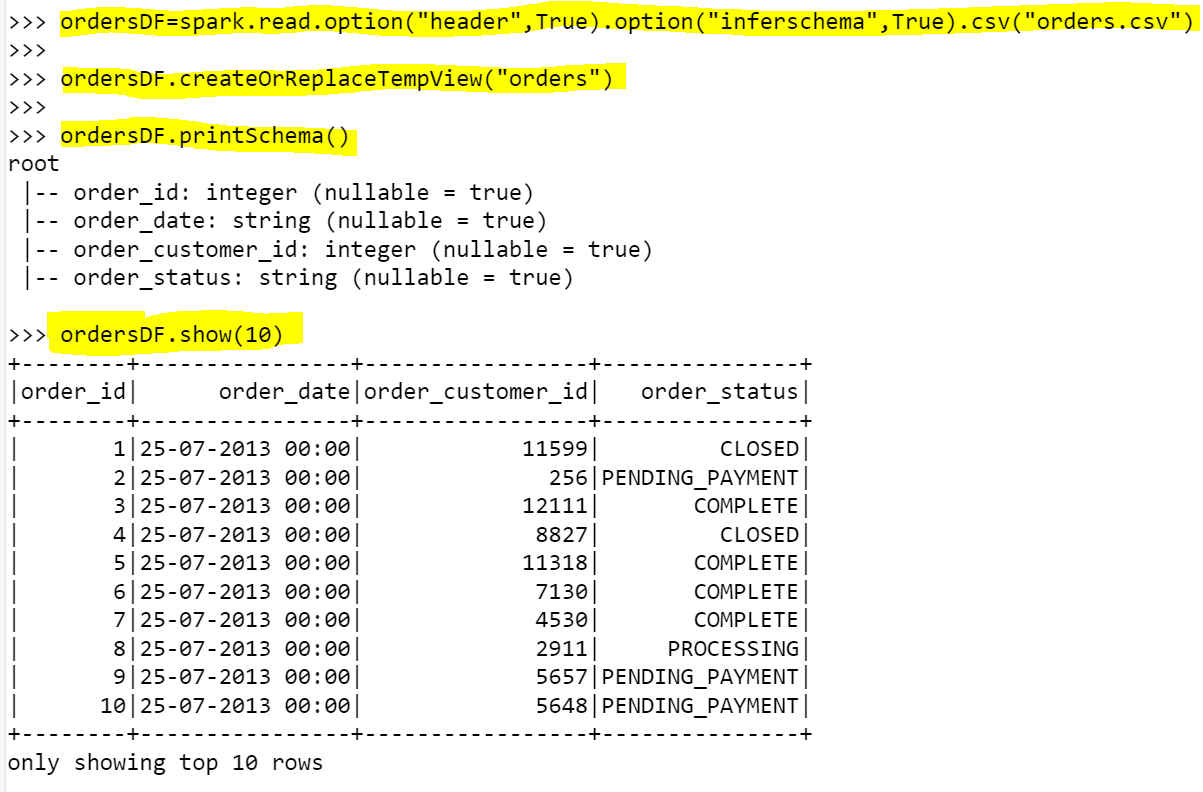


1. Load the required data in to DF like categories, customer,departments,order\_items,orders and products.

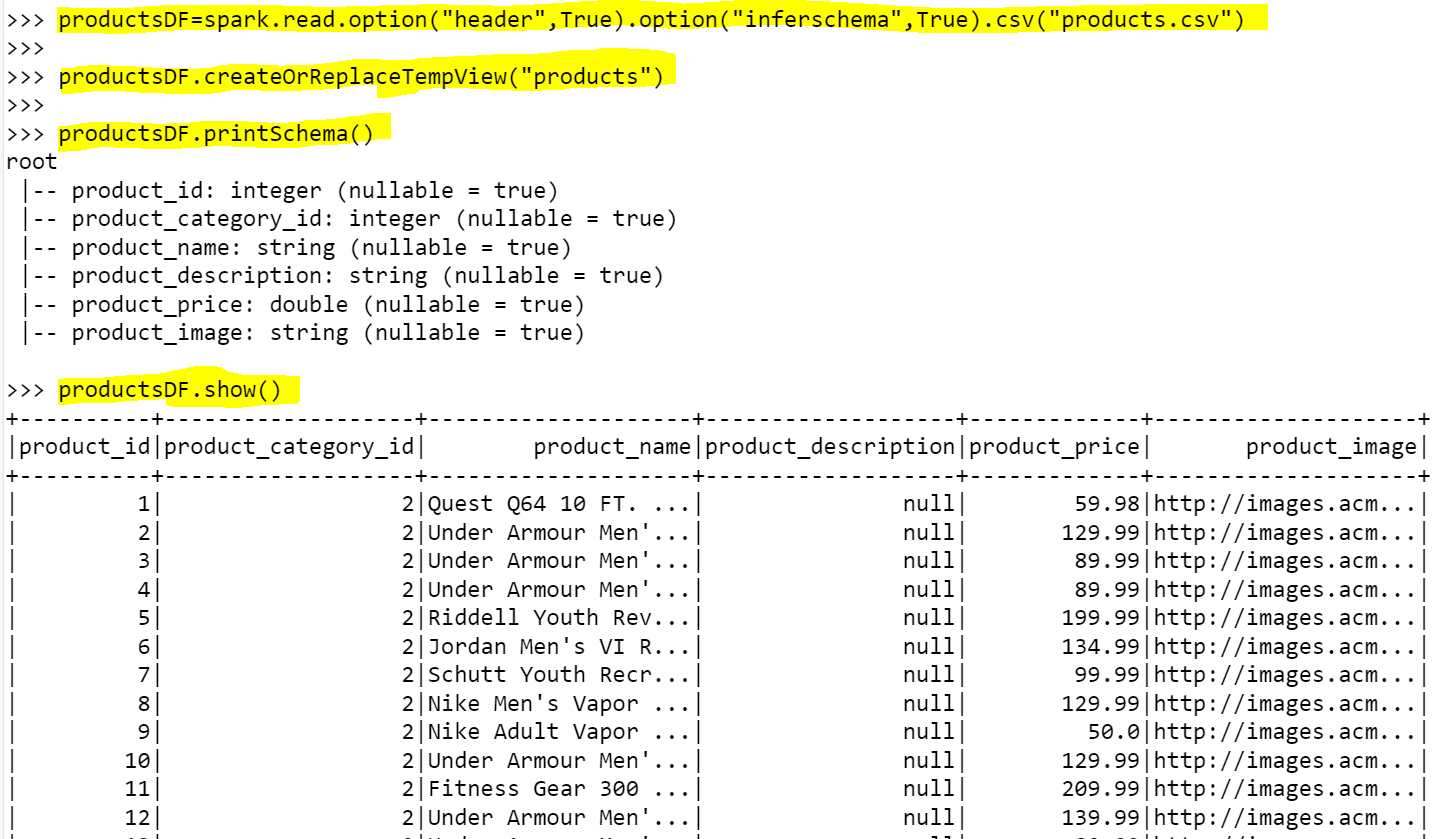
**Order\_item**



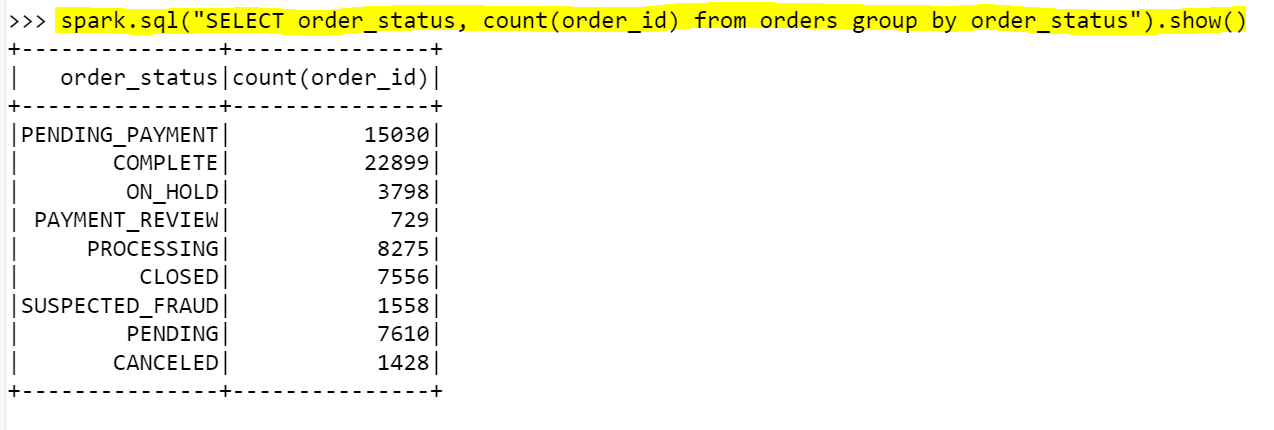
**Orders**



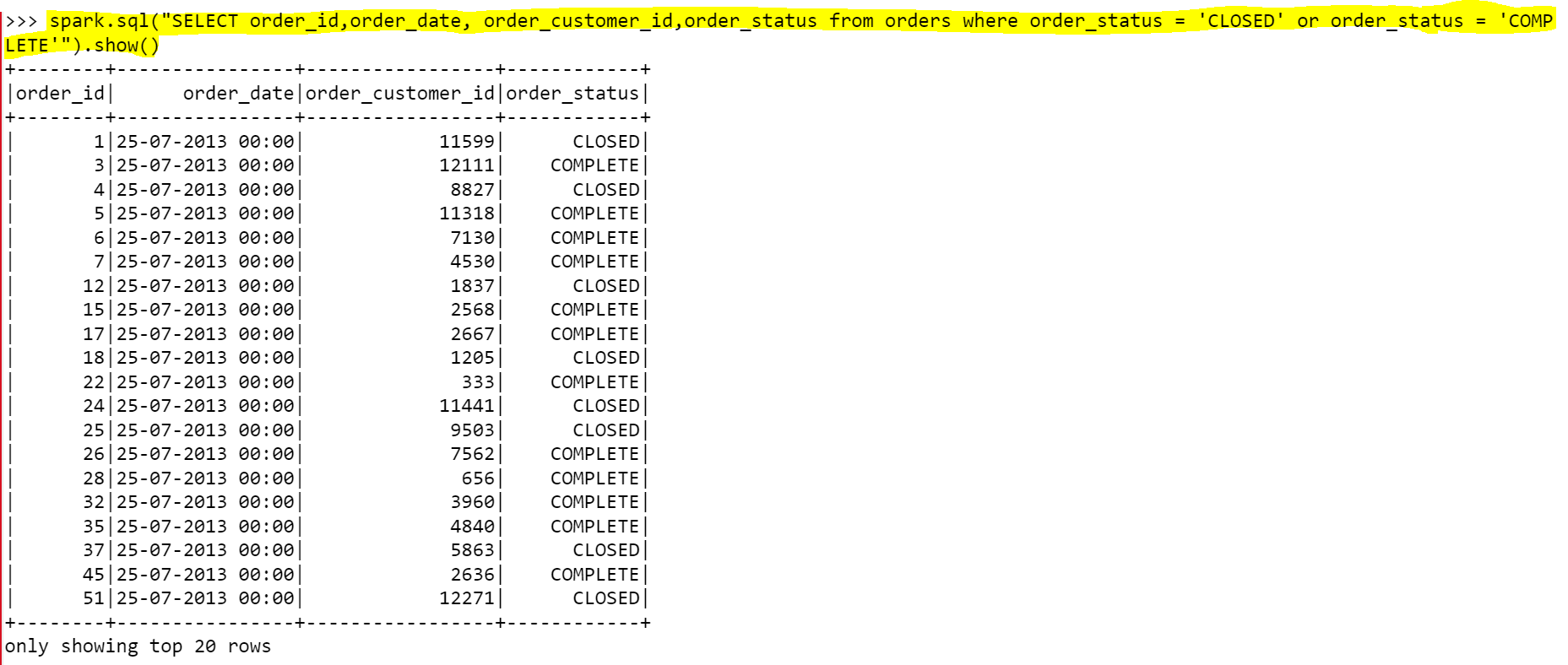
**Products**



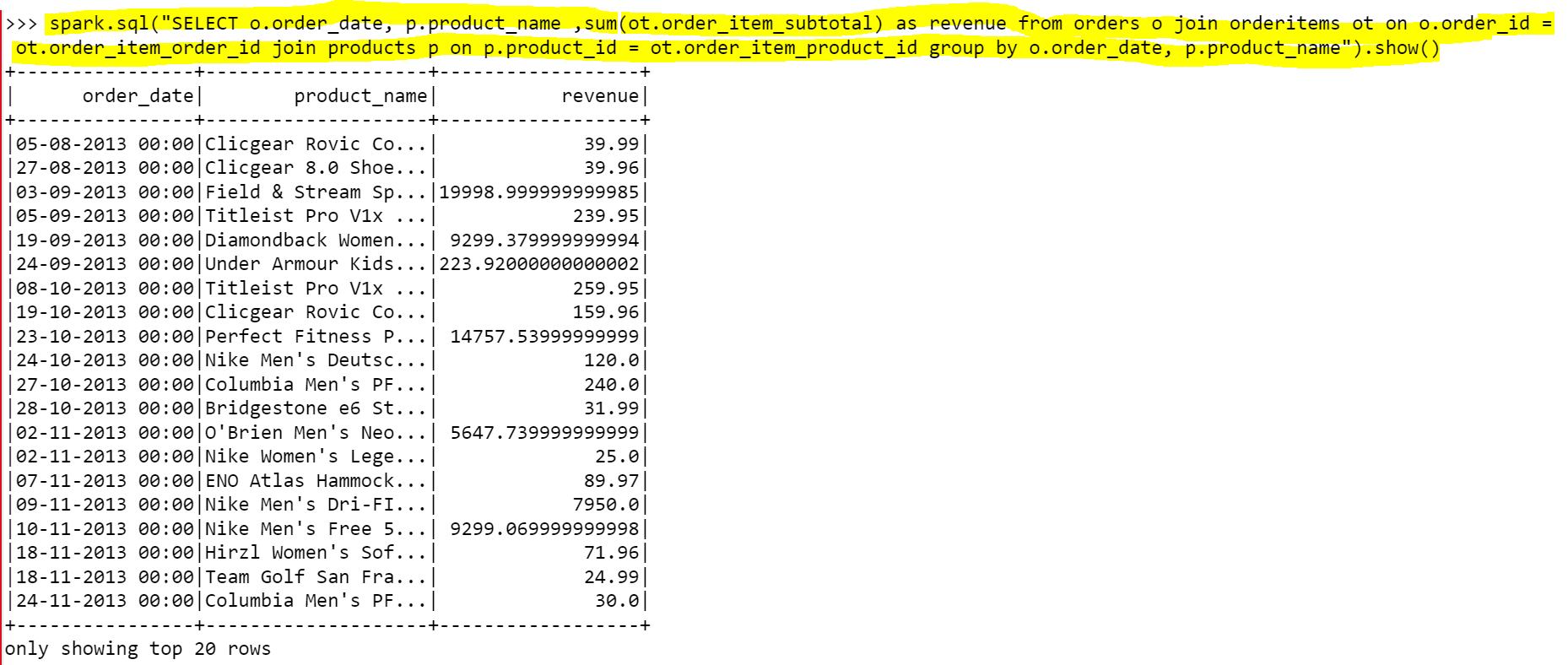
1. **Get the count for each order status**



1. **Filter only COMPLETE or CLOSED orders.**



1. **Join the products , order\_items and orders tables and calculate daily product revenue.**



1. **Write the data in to the table Daily product revenue in Hive.**

