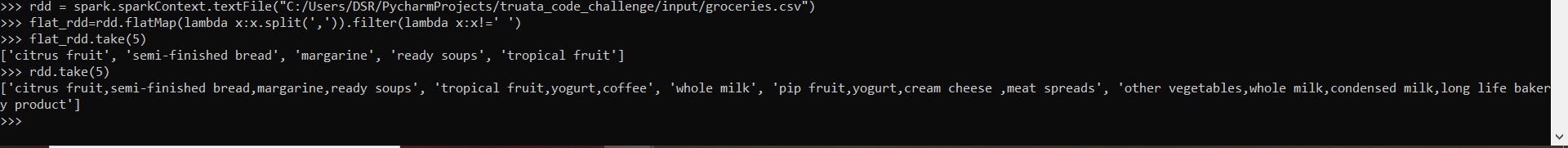
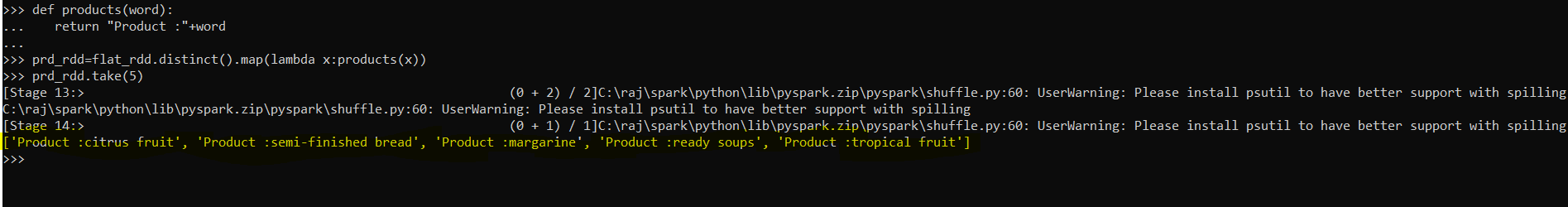
**Pyspark-Shell Output details:**

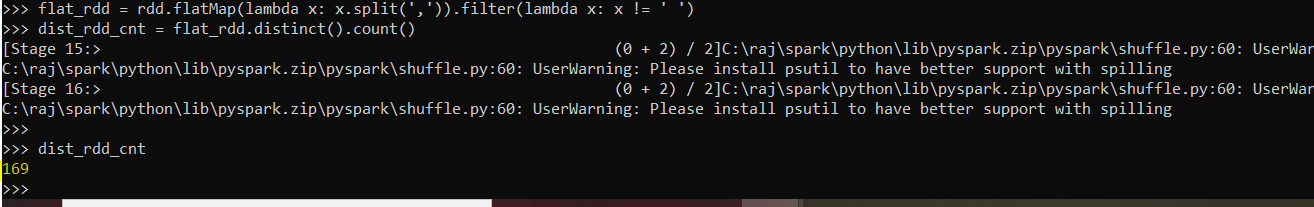
**PART\_1:**  
Task\_1\_1: Loading Data and take 5 records



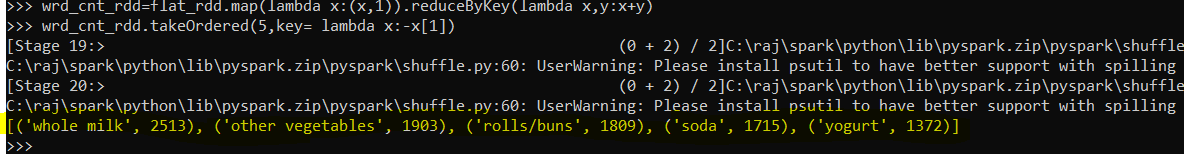
Task\_1\_2:



Task\_1\_2a:

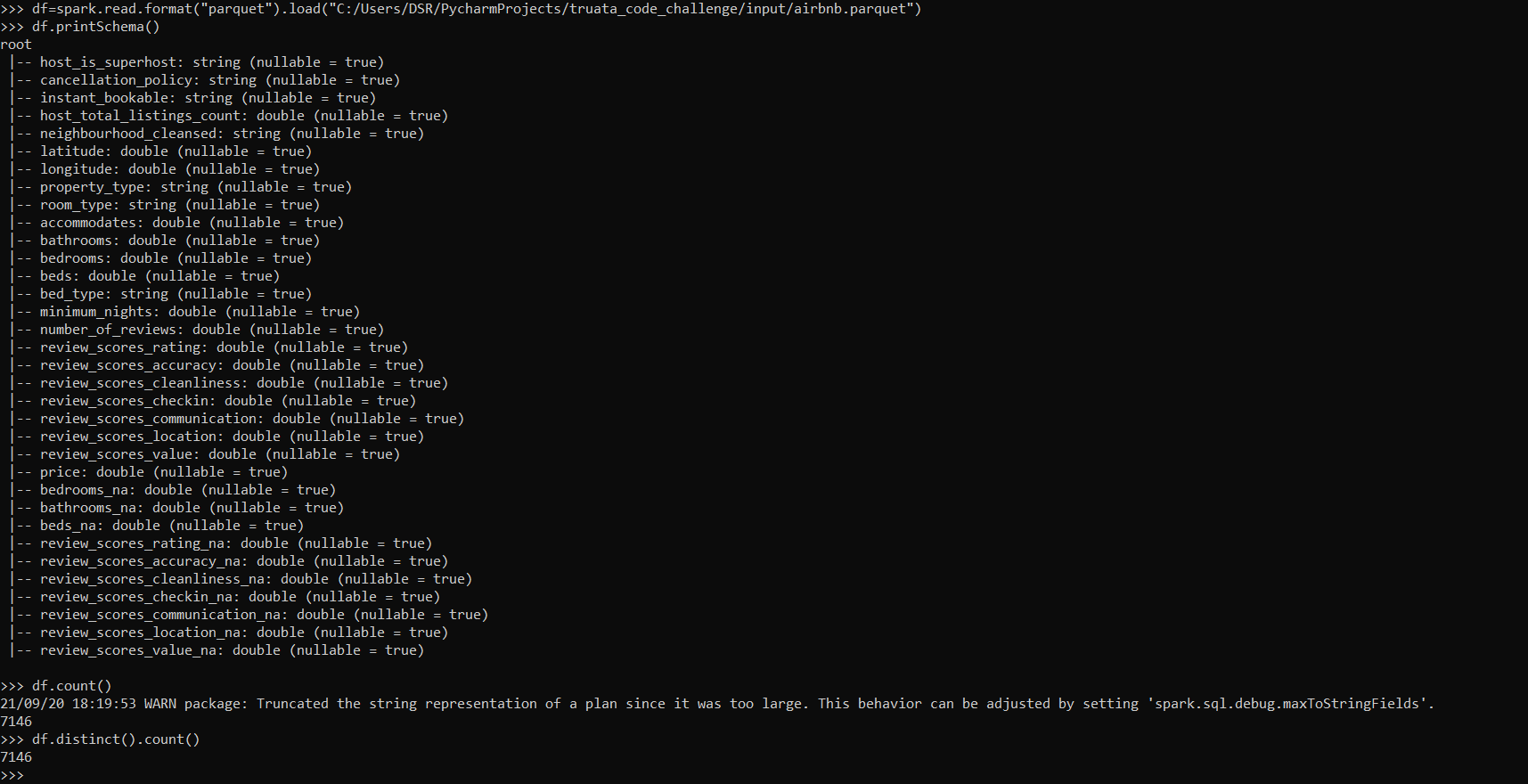


Task\_1\_3:

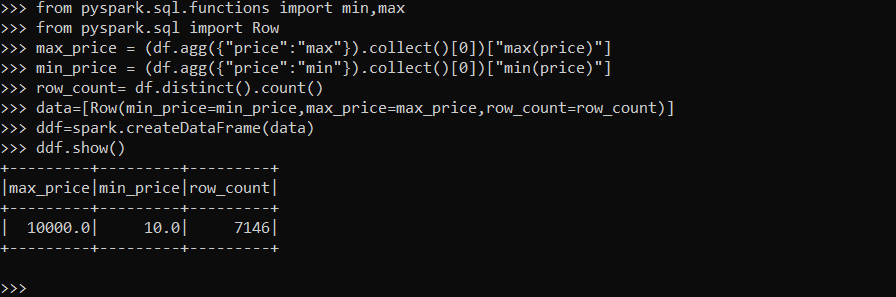


**PART\_2:**

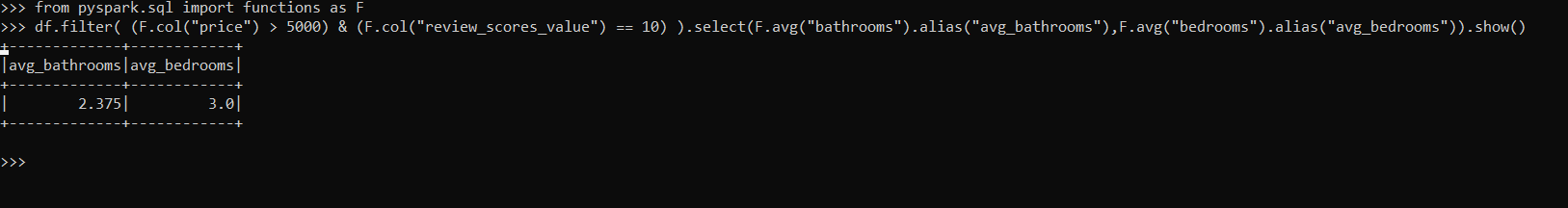
Task\_2\_1:



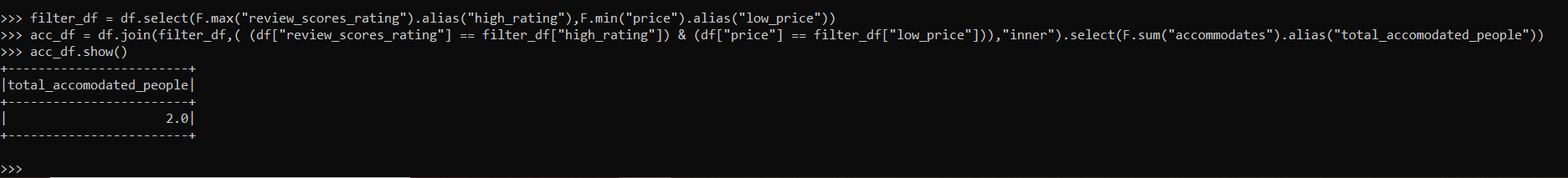
Task\_2\_2: Multiple Ways



Task\_2\_3:



Task\_2\_4:



Task\_2\_5:

Added code snippet to generate the required Airflow DAG

PART3:

Task\_3\_1 & Task 3\_2:

Added the code snippet to work with Spark Mlib as it wasn’t showing the required output.

Note: This needs some external libraries but my windows system was not supporting for that.