

**Online Retail data case study**

The dataset contains information about the purchase orders in various countries.

It contains 8 columns as shown below:

Column 1-Invoice No

Column 2-StockCode

Column 3-Description of the item purchased

Column 4-Quantity purchased

Column 5-Invoice Date-MM/DD/YYYY HH:MM:SS AM/PM

Column 6-Unit Price

Column 7-CustomerID

Column 8-Country

**Problem statements:**

- 1) Which item was purchased the highest on 1st December 2010?
- 2) Which item was purchased the highest from United Kingdom?
- 3) What are the 3 least purchased items in France?

**1) Which place has the most zika suspected cases?**

**Answer :**

*retail\_mapper.py*

`#!/usr/bin/python`

*retail\_reducer.py*

`#!/usr/bin/python`

**Command**

`cat online_retail.csv | sort | ./retail_mapper.py | ./retail_reducer.py`

**Cloudxlab command**

`hadoop jar /usr/hdp/2.6.5.0-292/hadoop-mapreduce/hadoopstreaming.jar -input`

`/user/support1161/online_retail.csv -output`

`/user/support1161/war_cloudxlab_output -file`

`/home/support1161/retail_mapper.py -file /home/support1161/retail_reducer.py`

`-mapper /home/support1161/retail_mapper.py -reducer`

`/home/support1161/retail_reducer.py`