1. Write **Pig Latin scripts** for the following on **Online Retail Dataset**.
2. Total number of unique customers in the "United Kingdom".
3. Find total invoices generated in the year 2009.
4. Generate the total revenue collected from Country="France".
5. Find the country in which maximum number of items were sold.

raw\_data = LOAD 'online-retail.csv' USING PigStorage(',')

AS (InvoiceNo:chararray, StockCode:chararray, Description:chararray,

Quantity:int, InvoiceDate:chararray, UnitPrice:float,

CustomerID:chararray, Country:chararray);

* 1. **Total number of unique customers in the "United Kingdom".**

filtered\_data = FILTER raw\_data BY Country == 'United Kingdom';

unique\_customers = DISTINCT (FOREACH filtered\_data GENERATE CustomerID);

total\_unique\_customers = FOREACH (GROUP unique\_customers ALL) GENERATE

COUNT(unique\_customers) AS TotalUniqueCustomers;

DUMP total\_unique\_customers;

Output: 15

1. **Find total invoices generated in the year 2009.**

filtered\_data = FILTER raw\_data BY InvoiceDate == '2009';

grouped\_by\_invoice = GROUP filtered\_data BY InvoiceNo;

total\_invoices = FOREACH (GROUP grouped\_by\_invoice ALL) GENERATE

COUNT(grouped\_by\_invoice) AS TotalInvoices;

DUMP total\_invoices;

Output:3

1. **Generate the total revenue collected from Country="France".**

france\_data = FILTER raw\_data BY Country == 'France';

france\_revenue = FOREACH france\_data GENERATE (Quantity \* UnitPrice) AS Revenue;

total\_revenue = FOREACH (GROUP france\_revenue ALL) GENERATE SUM(france\_revenue.Revenue) AS TotalRevenue;

DUMP total\_revenue;

Output: 177.47999954223633

1. **Find the country in which maximum number of items were sold.**

grouped\_by\_country = GROUP raw\_data BY Country;

country\_totals = FOREACH grouped\_by\_country GENERATE group AS Country,

SUM(raw\_data.Quantity) AS TotalQuantity;

sorted\_totals = ORDER country\_totals BY TotalQuantity DESC;

max\_country = LIMIT sorted\_totals 1;

DUMP max\_country;

Output: United Kingdom,1815