**Cargar dos fuentes de datos y realizar transformaciones + acciones y queries**

1.- cargar dos fuentes de datos que al menos tengan un campo en común para poder hacer JOINS. el proceso más sencillo en cargar un par de CSVs utilizando el método visto en clase.

1.1 Generar CSV

Lo primero que se hizo fue de [1] ejecutar los querys para Customer y Order para de ahí hacer copy paste en algún editor de text y dar formato de CSV.

1.2 Subir CSV a Databrick y visulizar

Nos movemos a Data -> Add Data

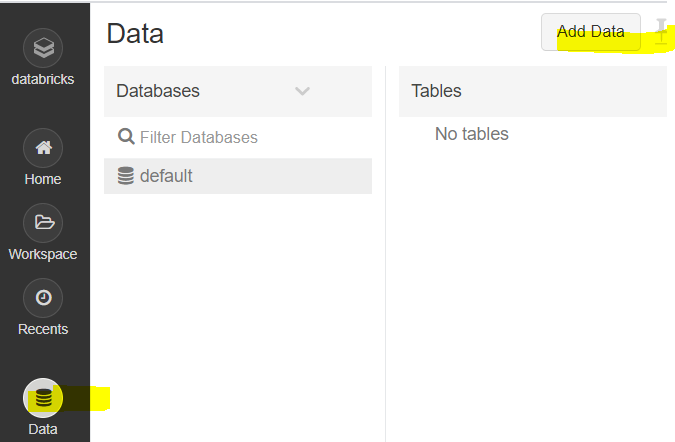


Figure 1Agregar Data

En la Ventana que se abre avente los CSV creados.

En el workspace creado se visualizaron dichos datasets.

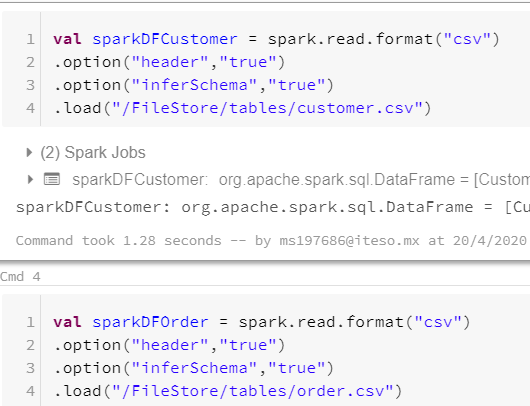


Figure formato a datasets

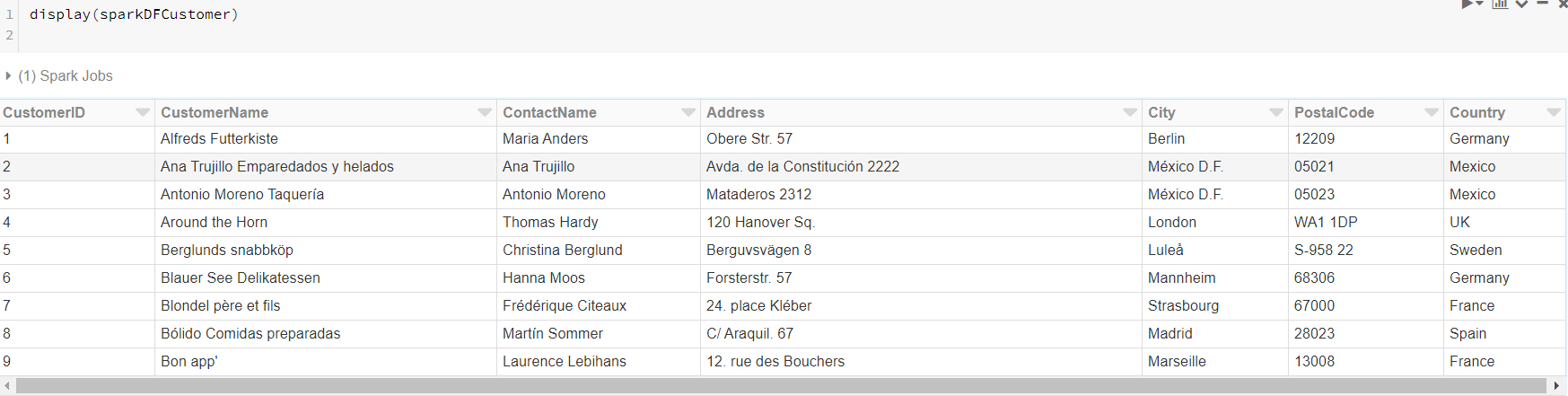


Figure Customers dataset desplegado

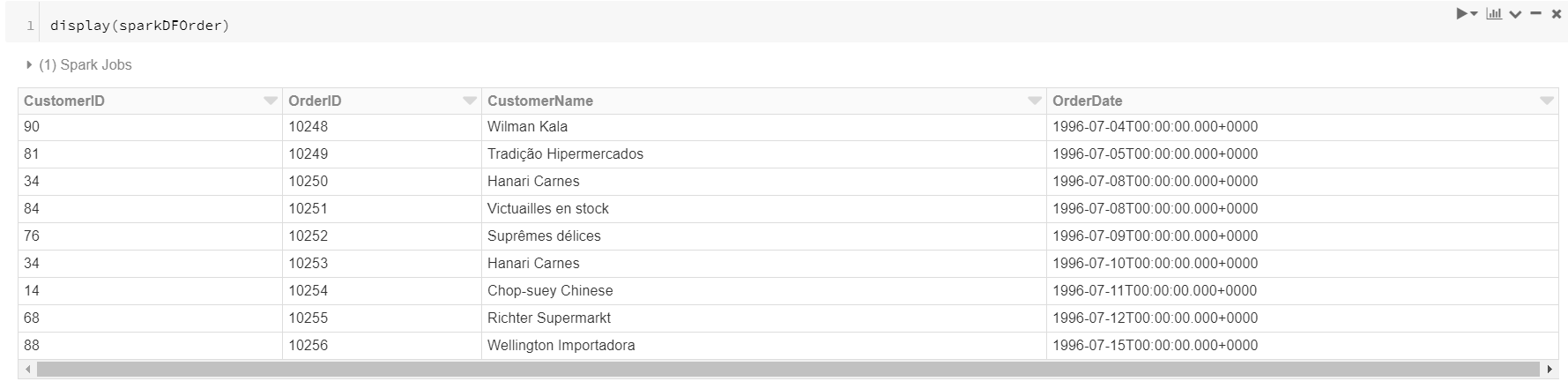


Figure Order dataset desplegado

2.- utilizar dataframes para realizar un JOIN entre ambos datasets y aplicar al menos las transformaciones filter, groupby y aggr.

Se crearon primero unas tablas temporales para los dataset de Customer y Order para poder aplicar un query de join.

Los dos Datasets comparten el campo CustomerID por lo tanto use un join de SQL para crear un dataframe resultado de la unión de los dos anteriores dataframes.

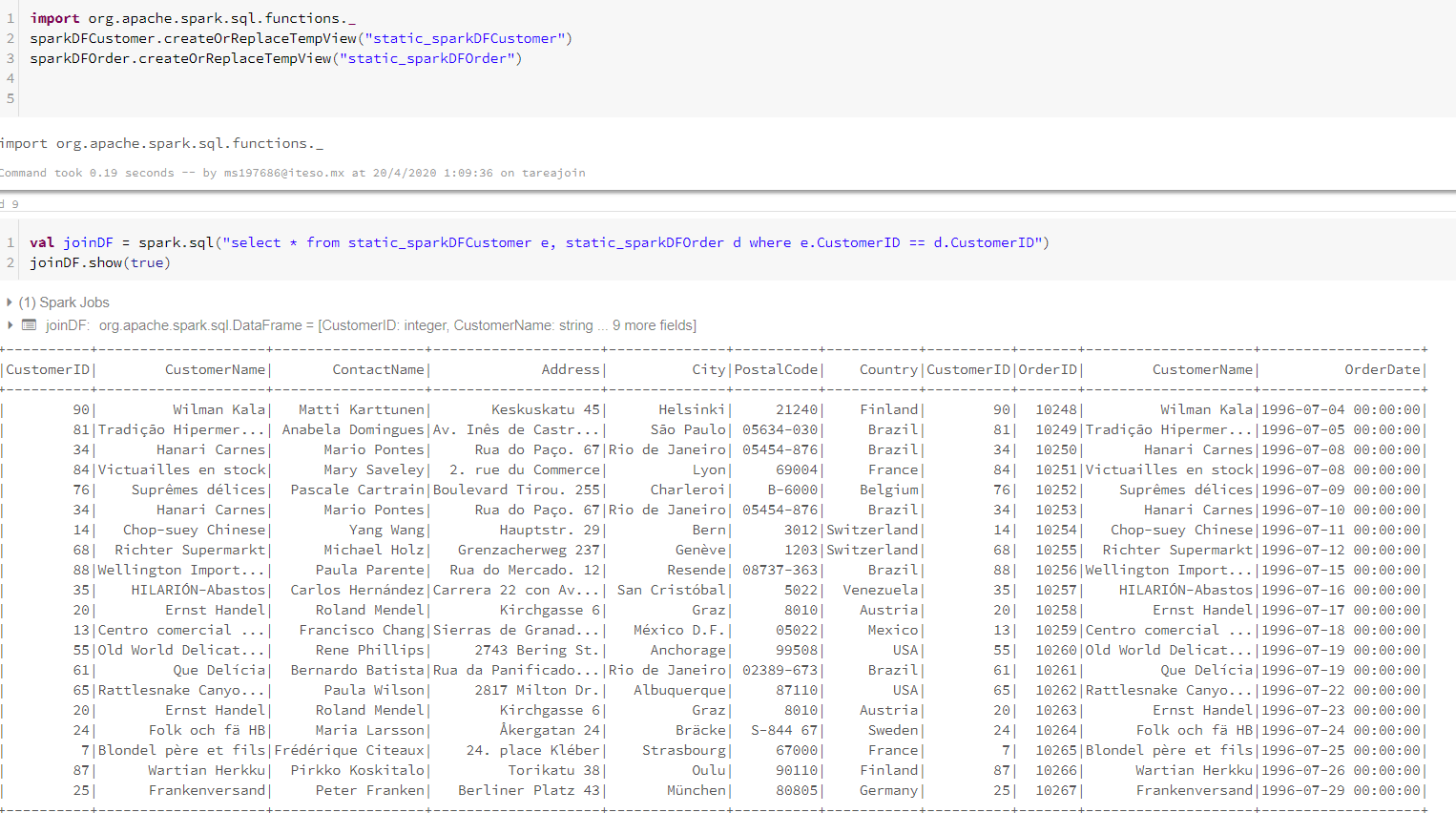


Figure 5Join de dataframes desplegado

Al dataframe creado con el join aplique un filter para optener las ordenes que se hicieron en Mexico, un groupBy para agrupar las ordenes por ciudad y ahí mismo use un agg para agregar el OrderID.

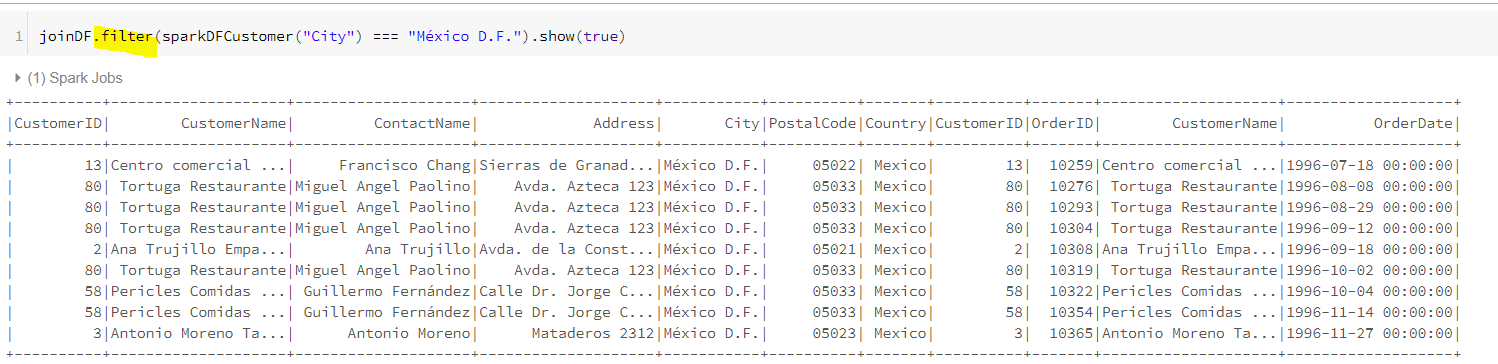


Figure 6Filtrado por país de dataframe

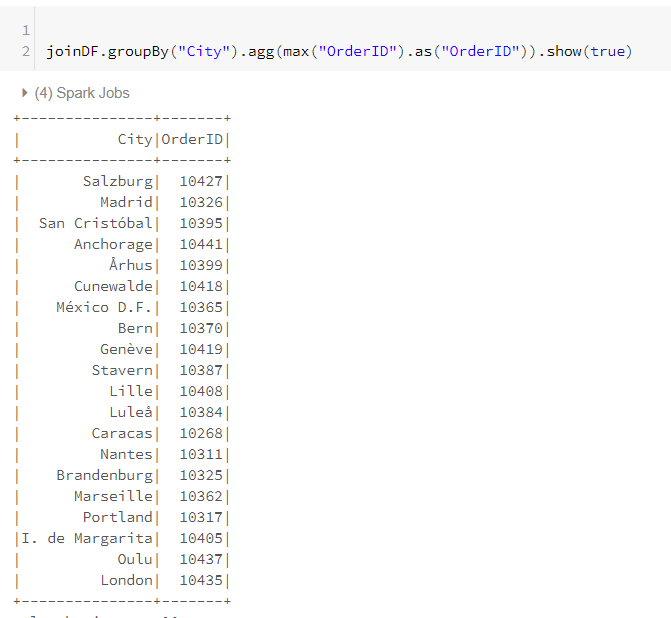


Figure 7Agrupado por ciudad de datagrame y agregación de OrderID

3.- crear una tabla temporal (como la vista en clase) y generar al menos 3 queries sobre el JOIN de los dato

Por ultimo se creo una tabla temporal del dataframe que tiene los dos dataset unidos y se aplicaron tres querys en su formato original para hacer la cuenta de los elementos y para ordenar por OrderID y por País.

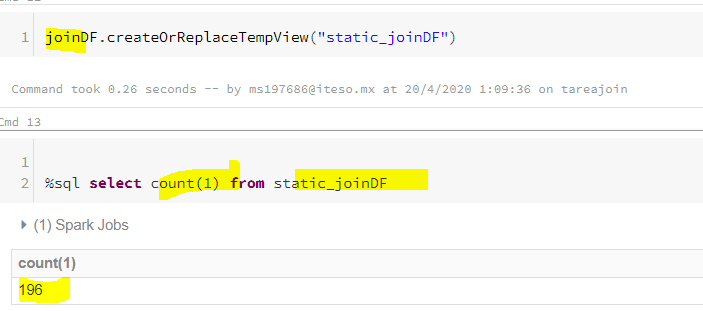


Figure 8query para Cuenta de elementos

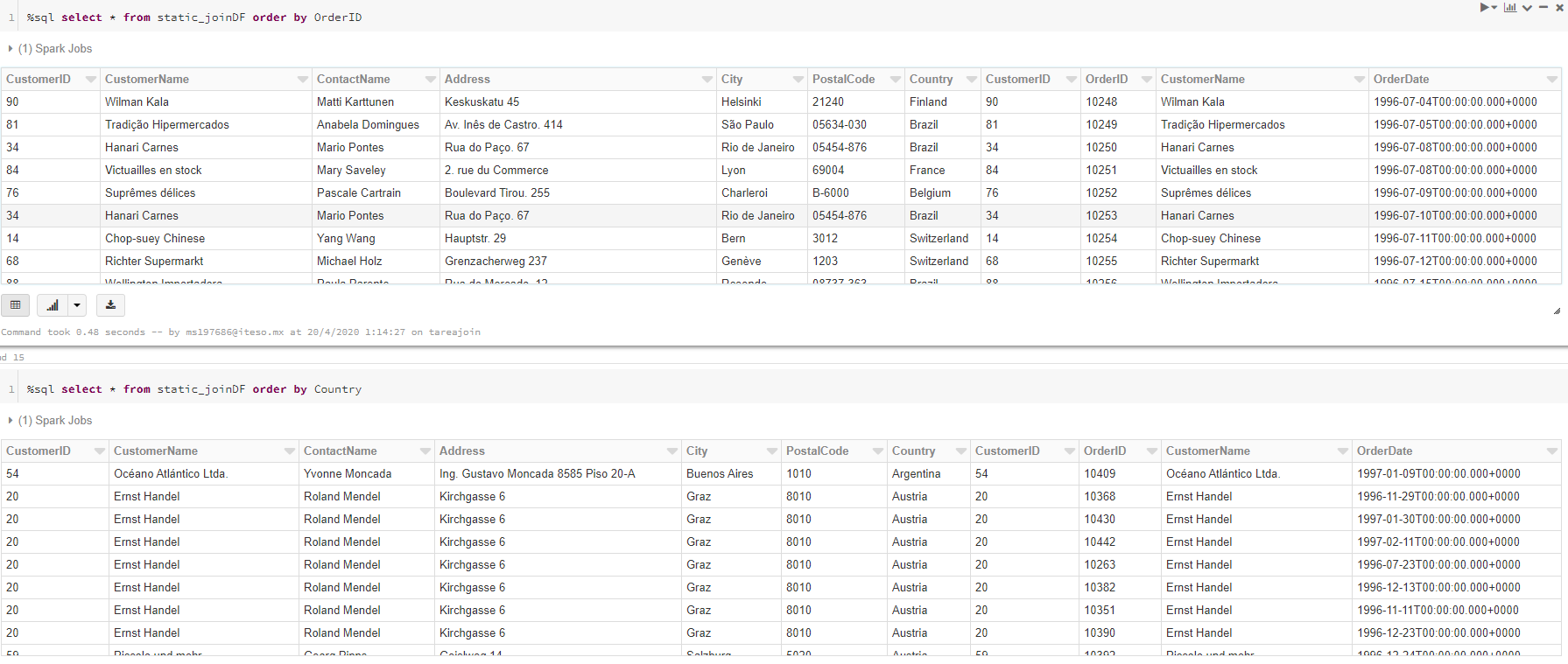


Figure query para ordenado por OrderId y país

**Referencias**

[1] «trysql\_select\_join» [En línea]. Disponible en:

<https://www.w3schools.com/sql/trysql.asp?filename=trysql_select_join> [Accedido: 20-Abl-2020].

[2] « Spark by {Examples}» [En línea]. Disponible en:

<https://sparkbyexamples.com/spark/spark-sql-dataframe-join/#jointype-object> [Accedido: 20-Abl-2020].