**Luís Paulo Lima de Brito**

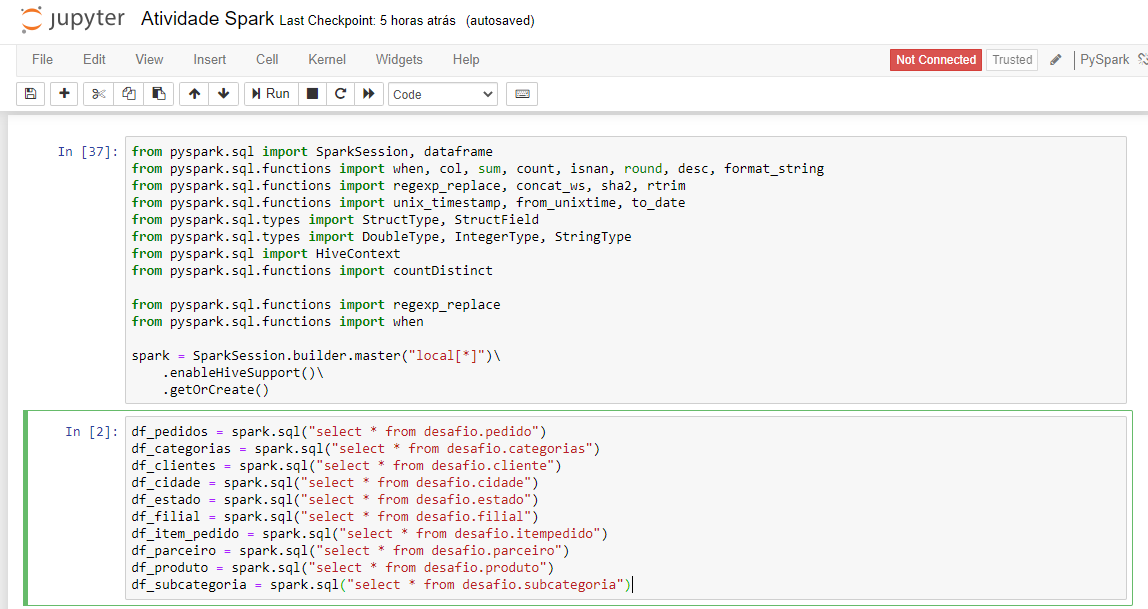
Atividade Spark

- iniciar o cluster

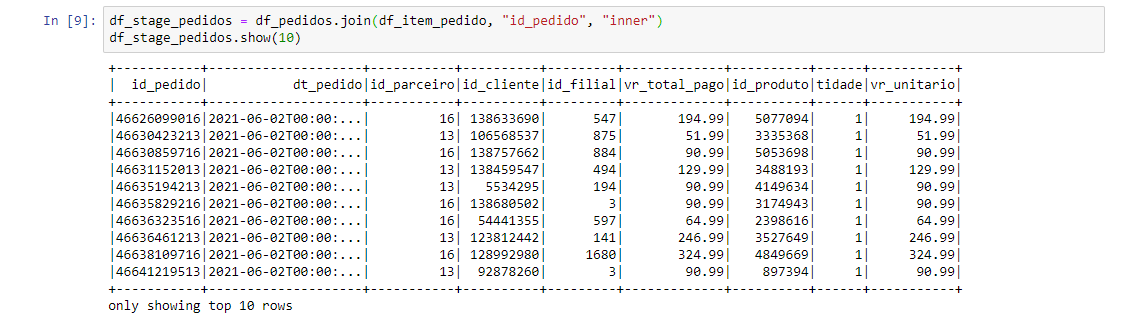
-> docker-compose up –d

Habilitar a porta para ter acesso o jupyter notebook

1 – Criar dataframes das tabelas que estão no hive



2 – Criar um dataframe (df\_pedidos) esta dataframe deve ser a união de pedido e item\_pedido;



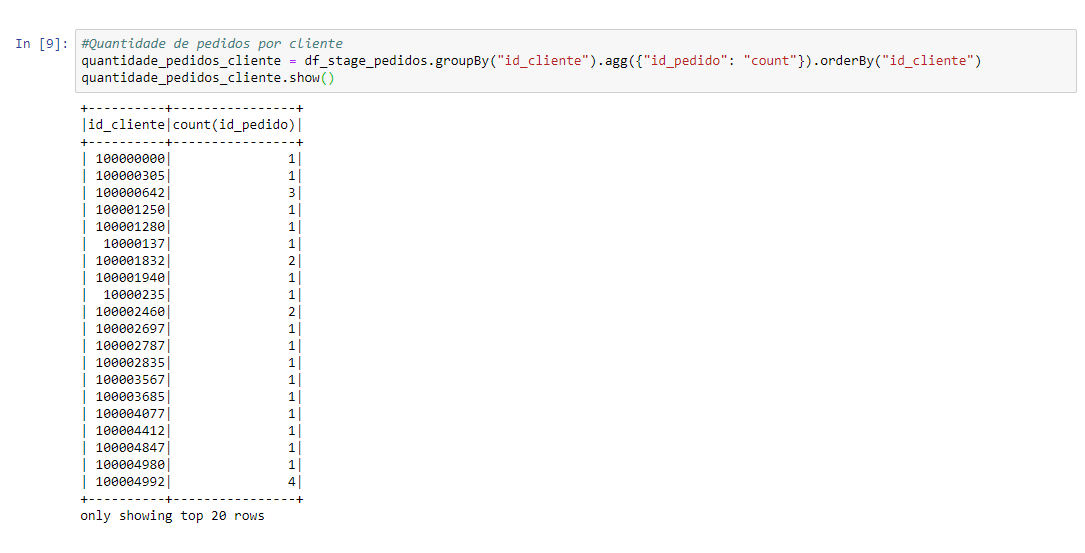
1. Ver a quantidade de pedidos.



1. Quantidade de produtos e agrupa-los por pedido



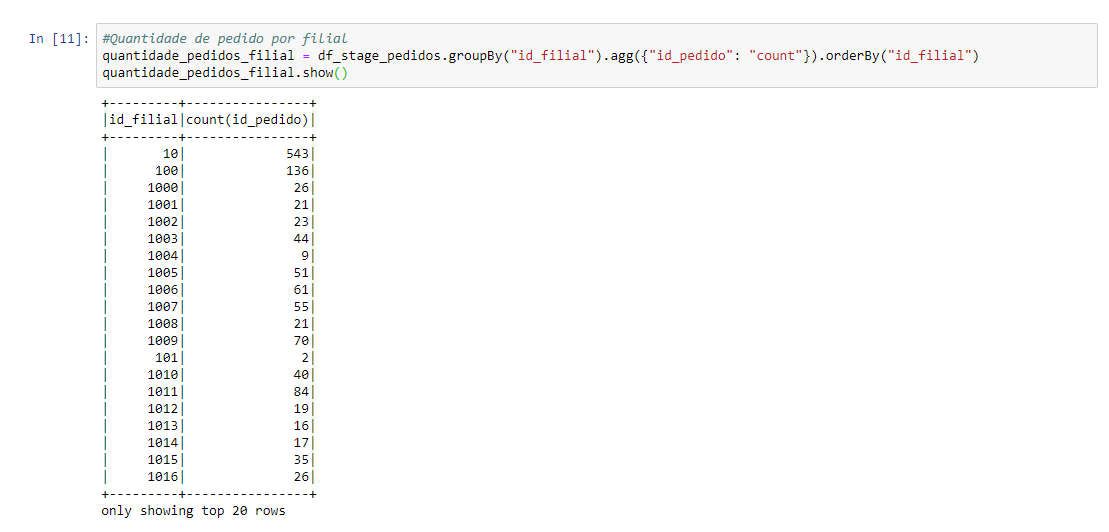
1. Quantidade de pedidos por cliente



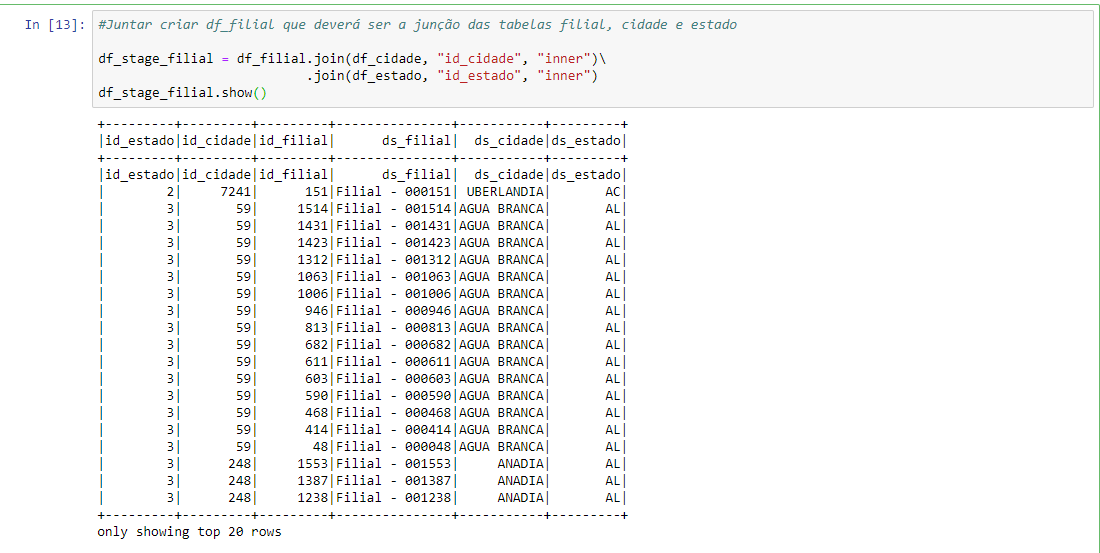
1. Quantidade de pedidos por parceiro



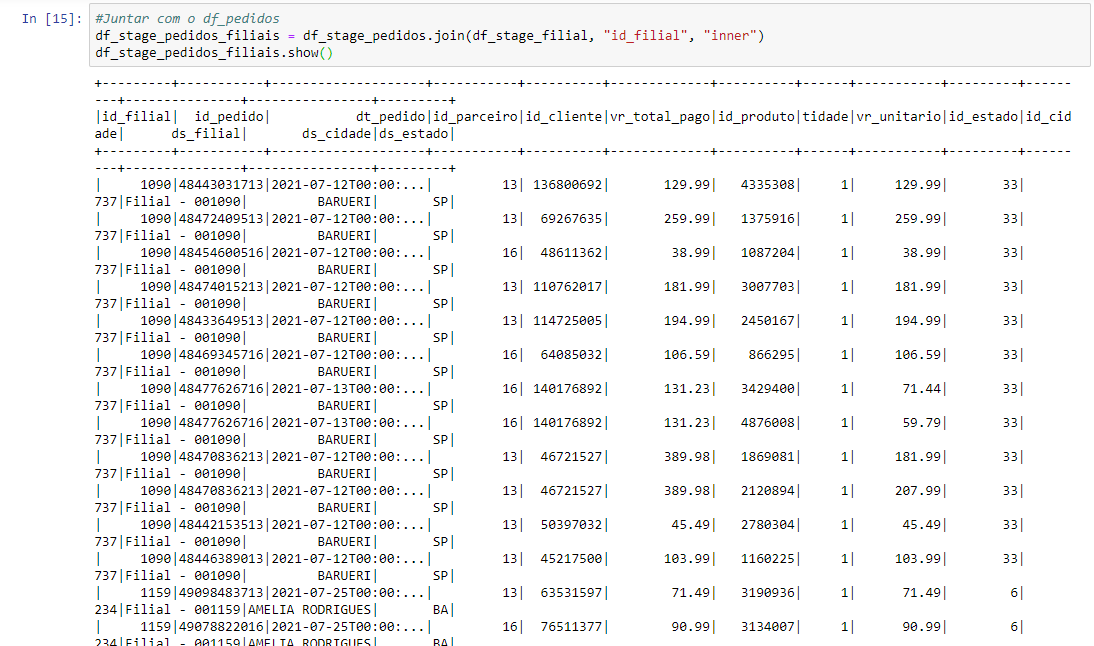
1. Quantidade de pedido por filial



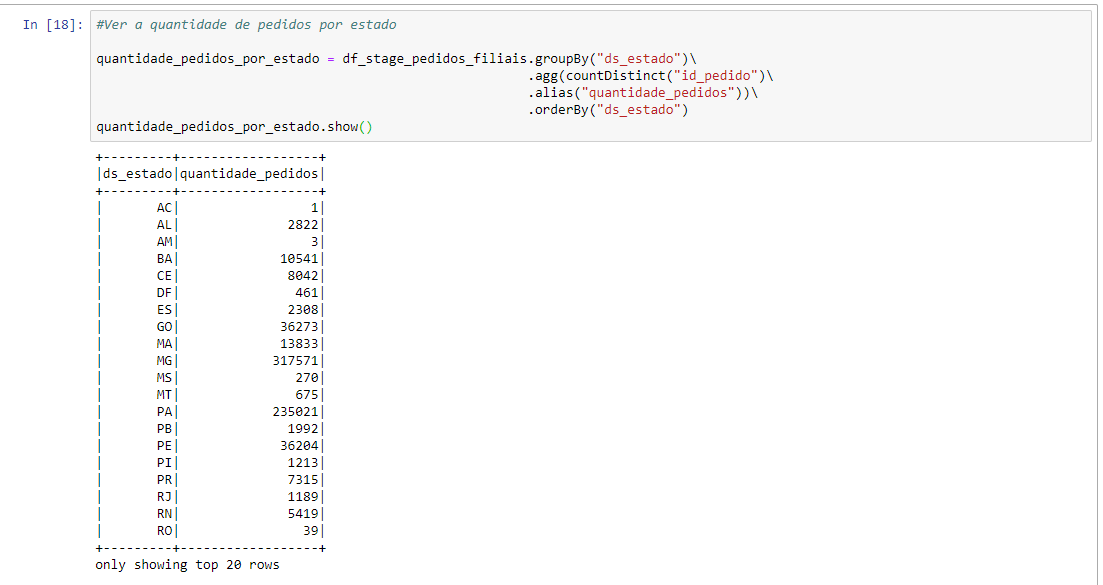
3 – Juntar criar df\_filial que deverá ser a junção das tabelas filial, cidade e estado;



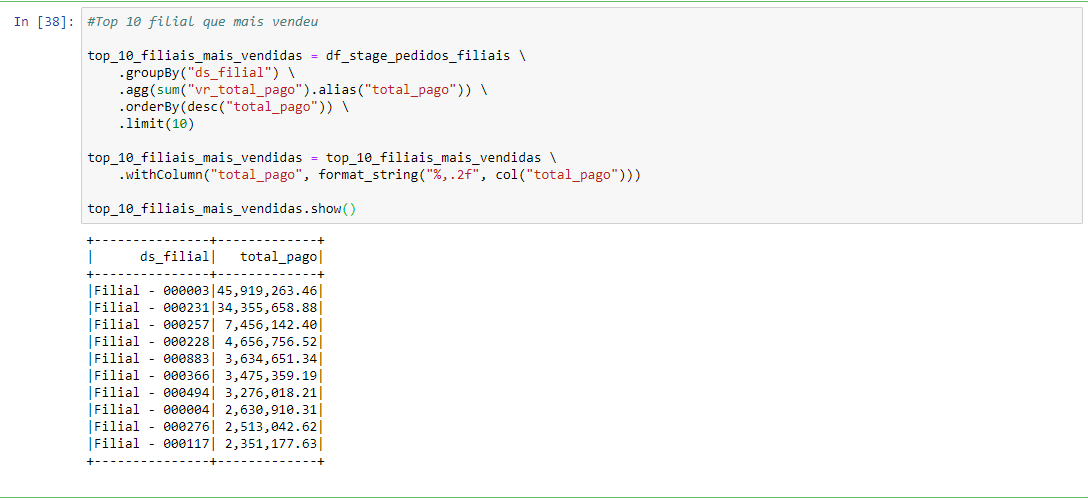
1. Juntar com o df\_pedidos



1. Ver a quantidade de pedidos por estado



1. Top 10 filial que mais vendeu



4 – Criar o dataframe df\_stage juntando todas as bases do nosso modelo relacional

