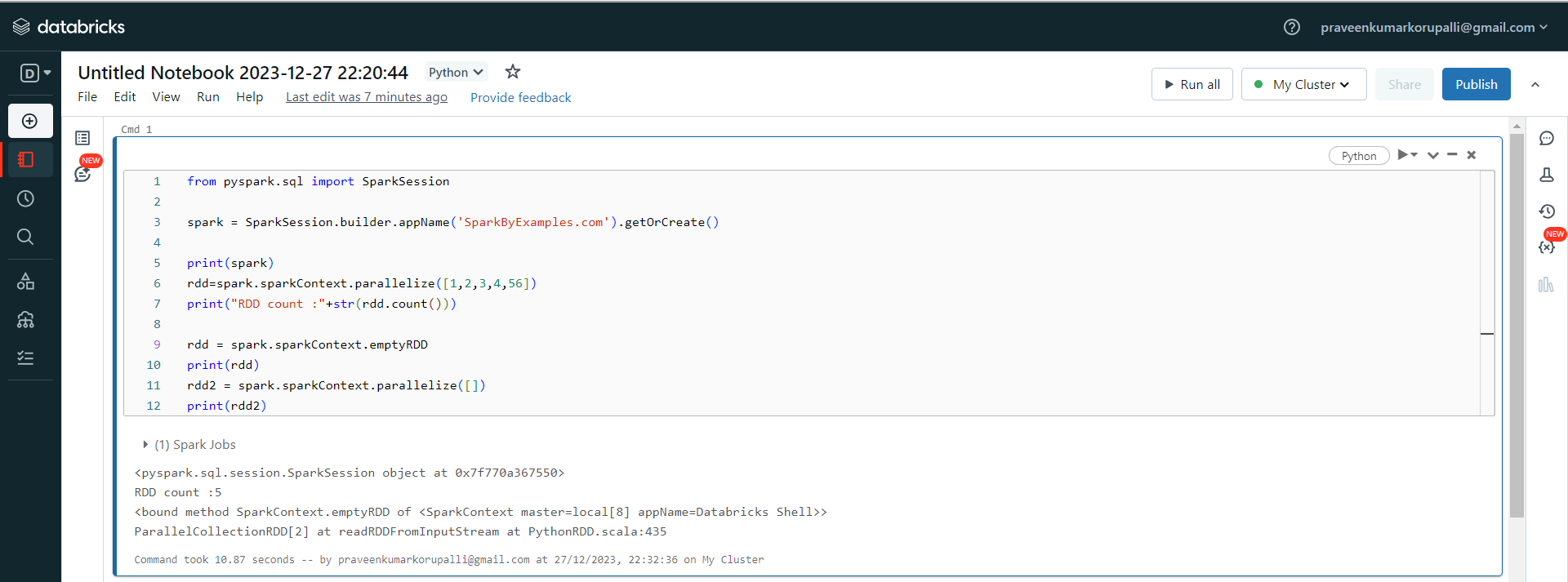
K. Praveen Kumar

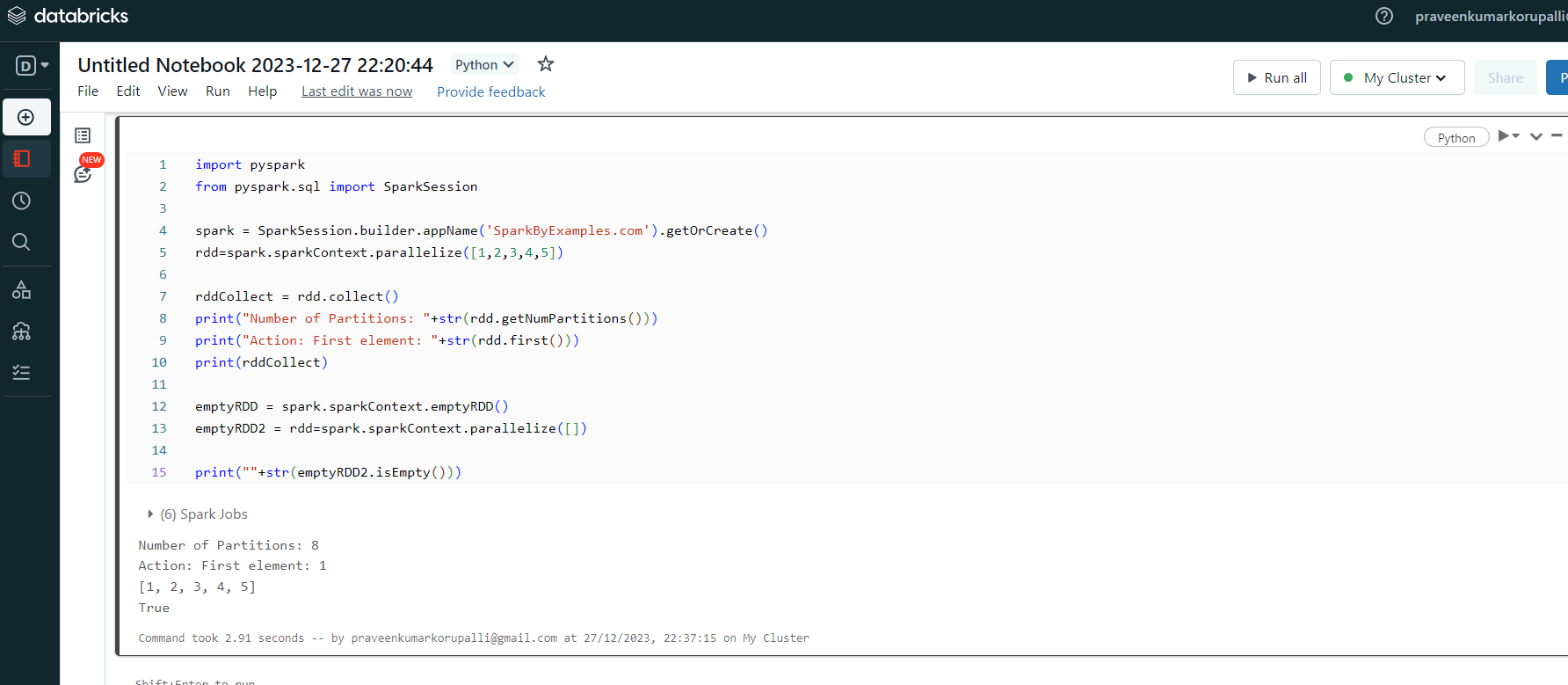
PySpark Assignment-5

Pyspark Examples:

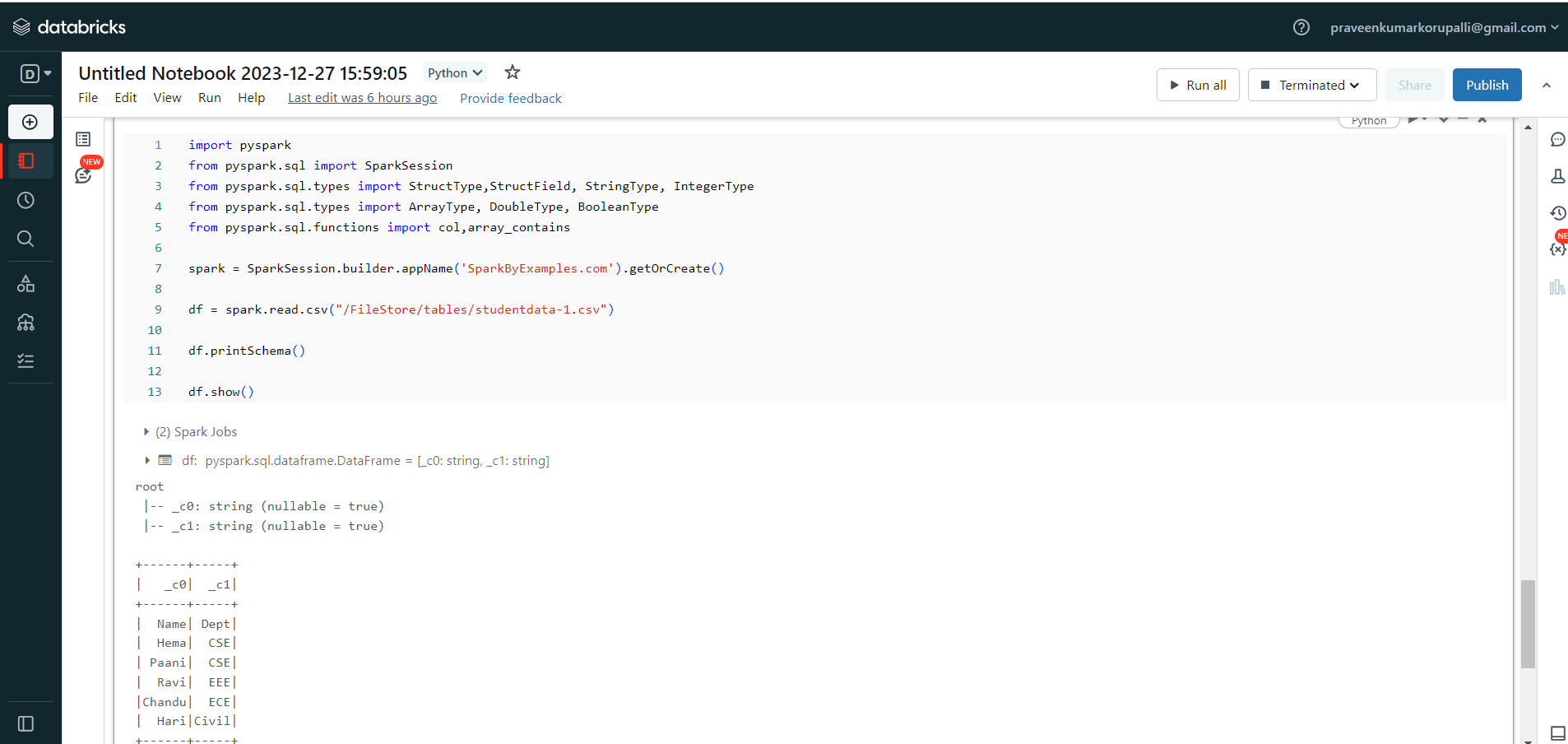
Creation of RDD:



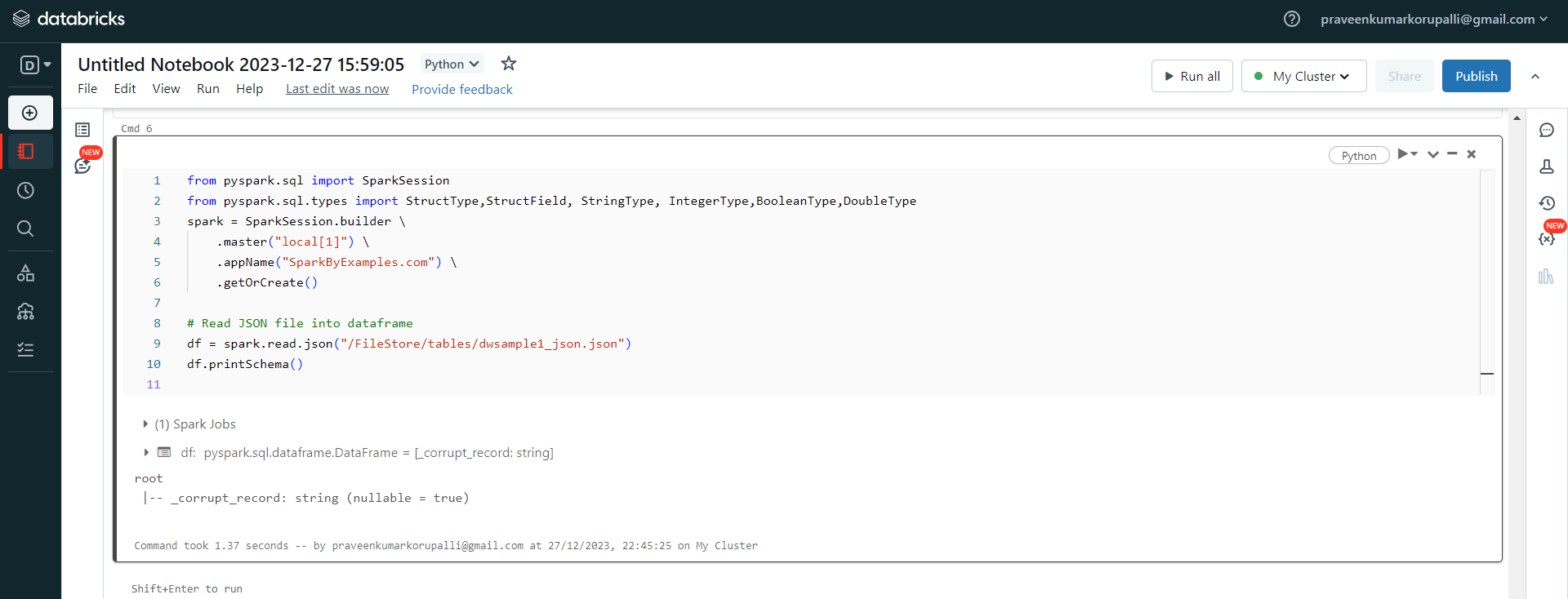
Create a dataframe through parallelize():



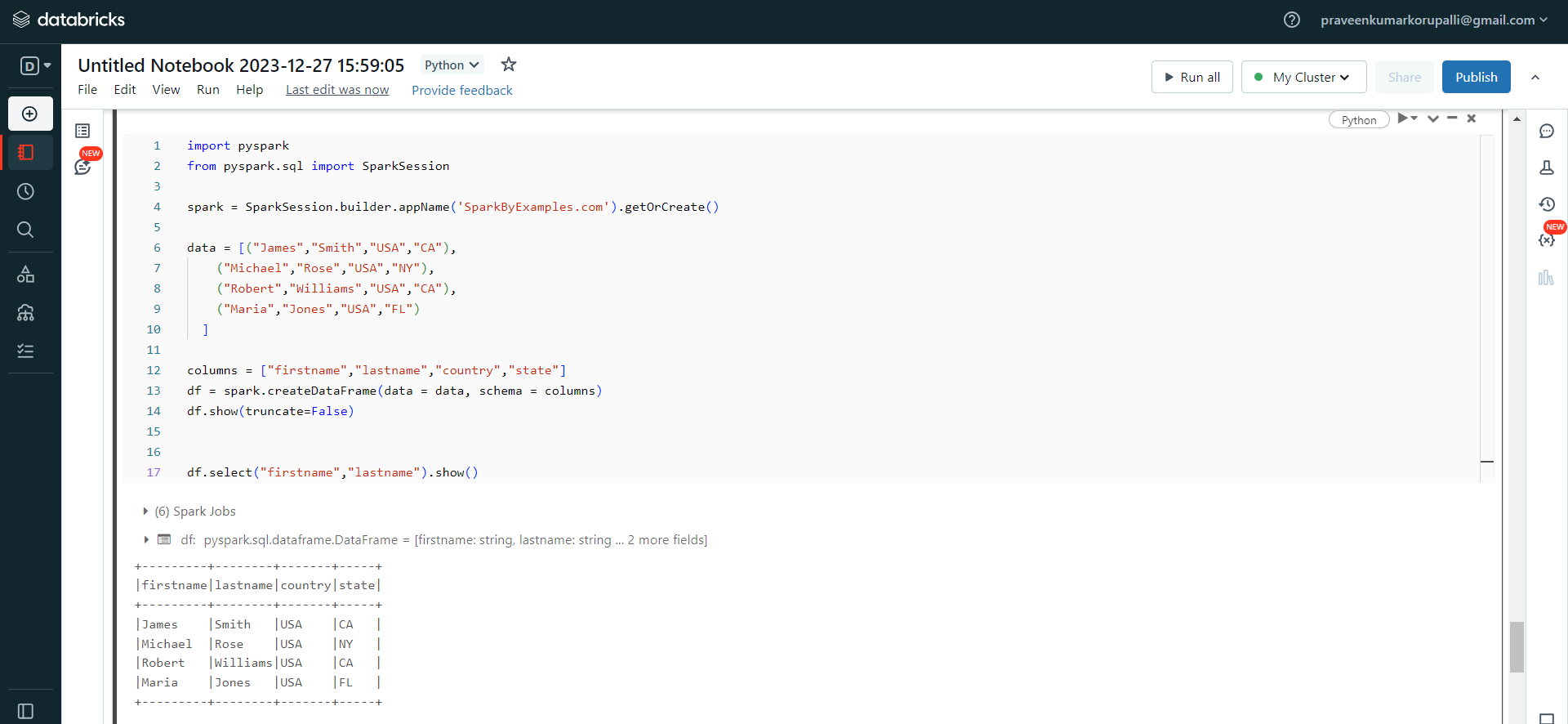
Reading a CSV file:

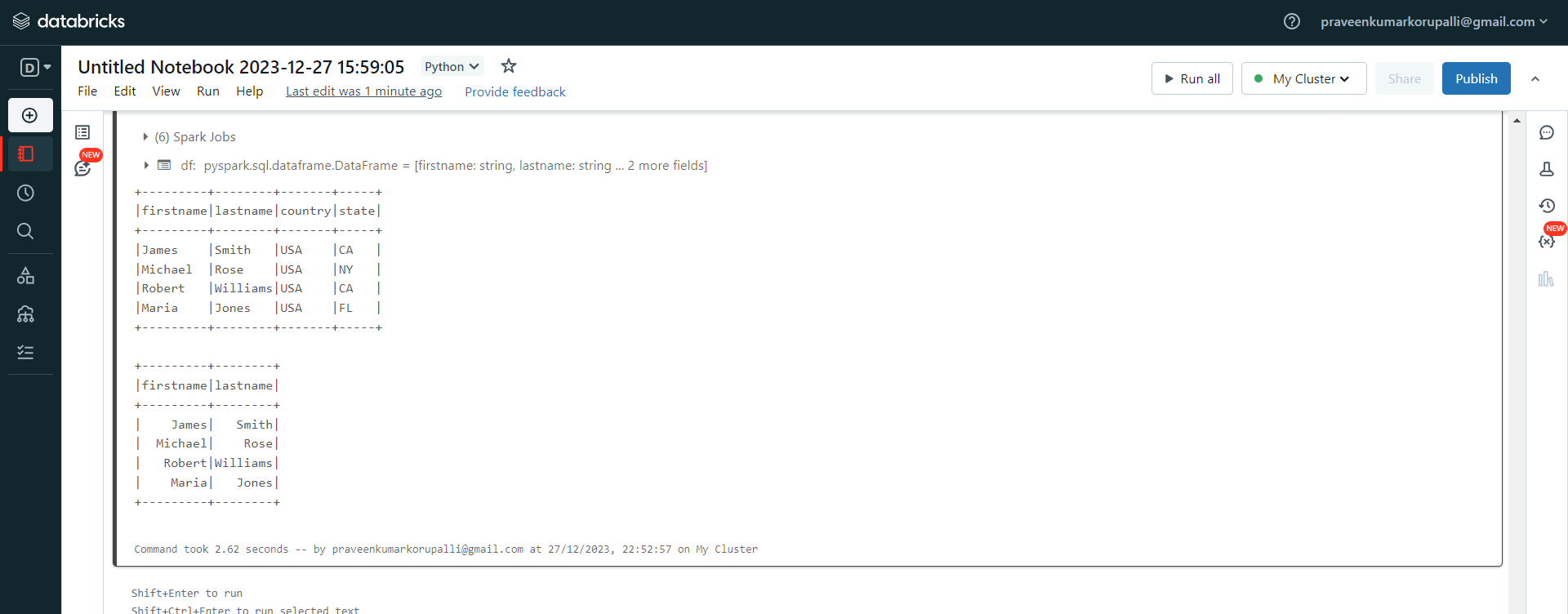


Reading a json file:

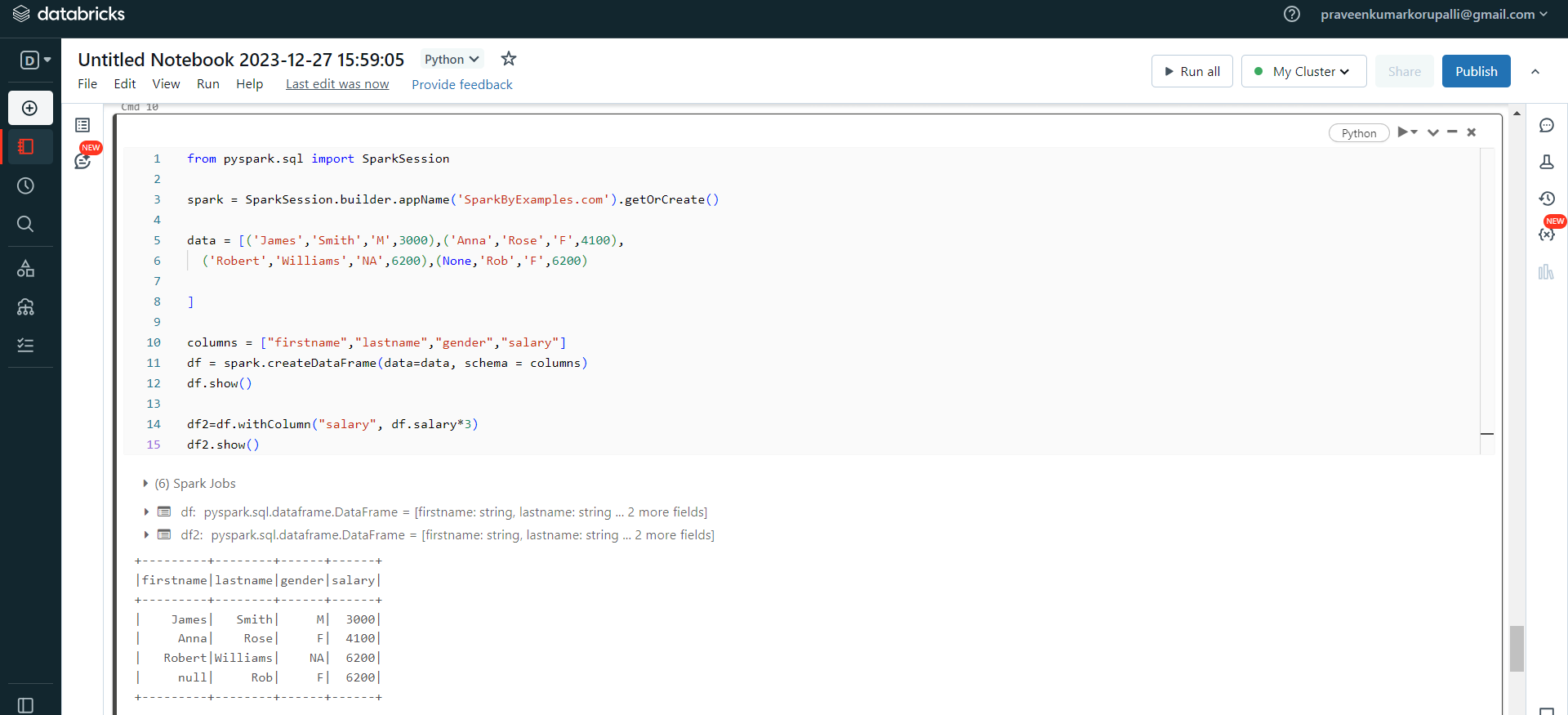


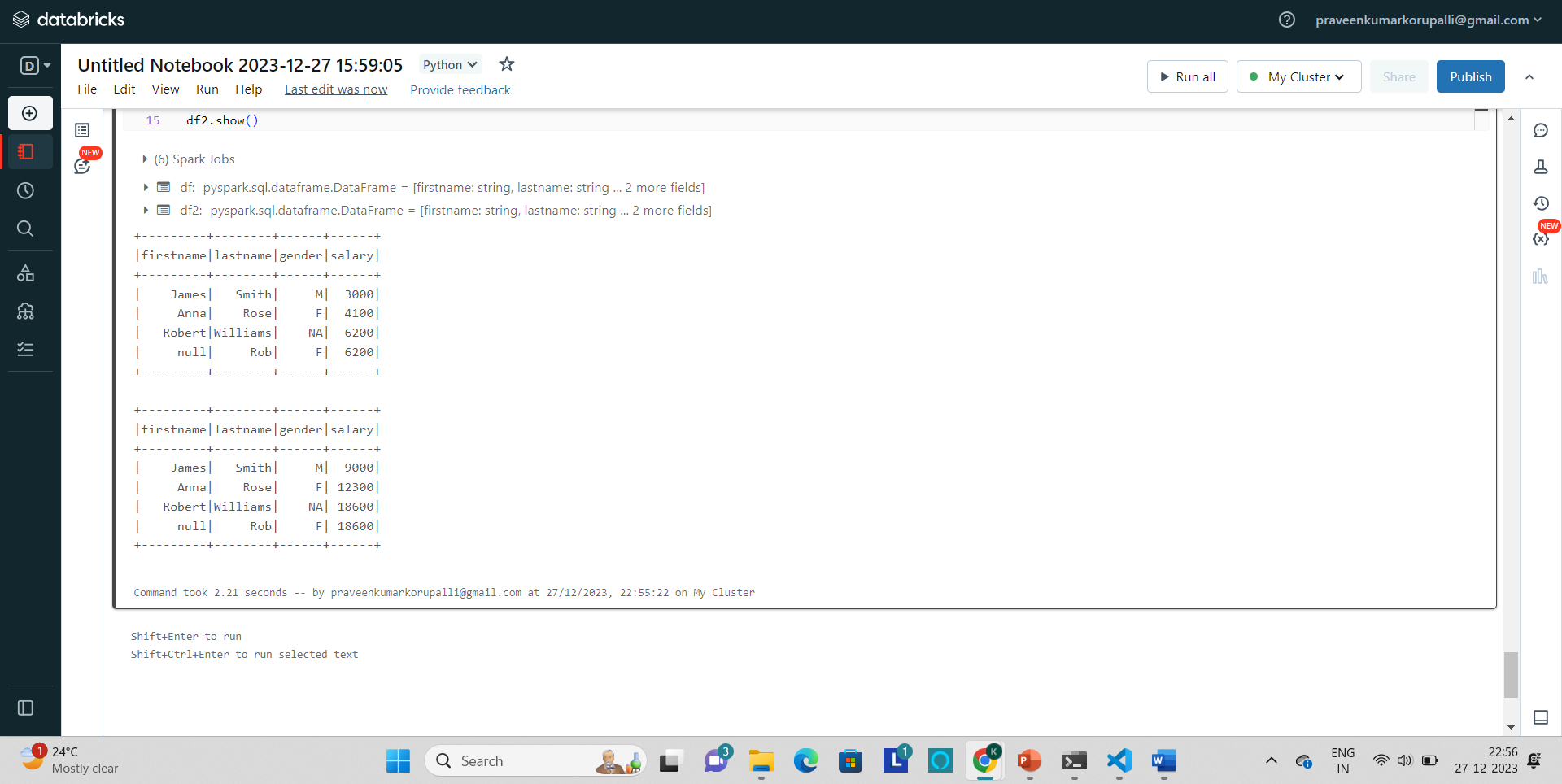
Select Columns:



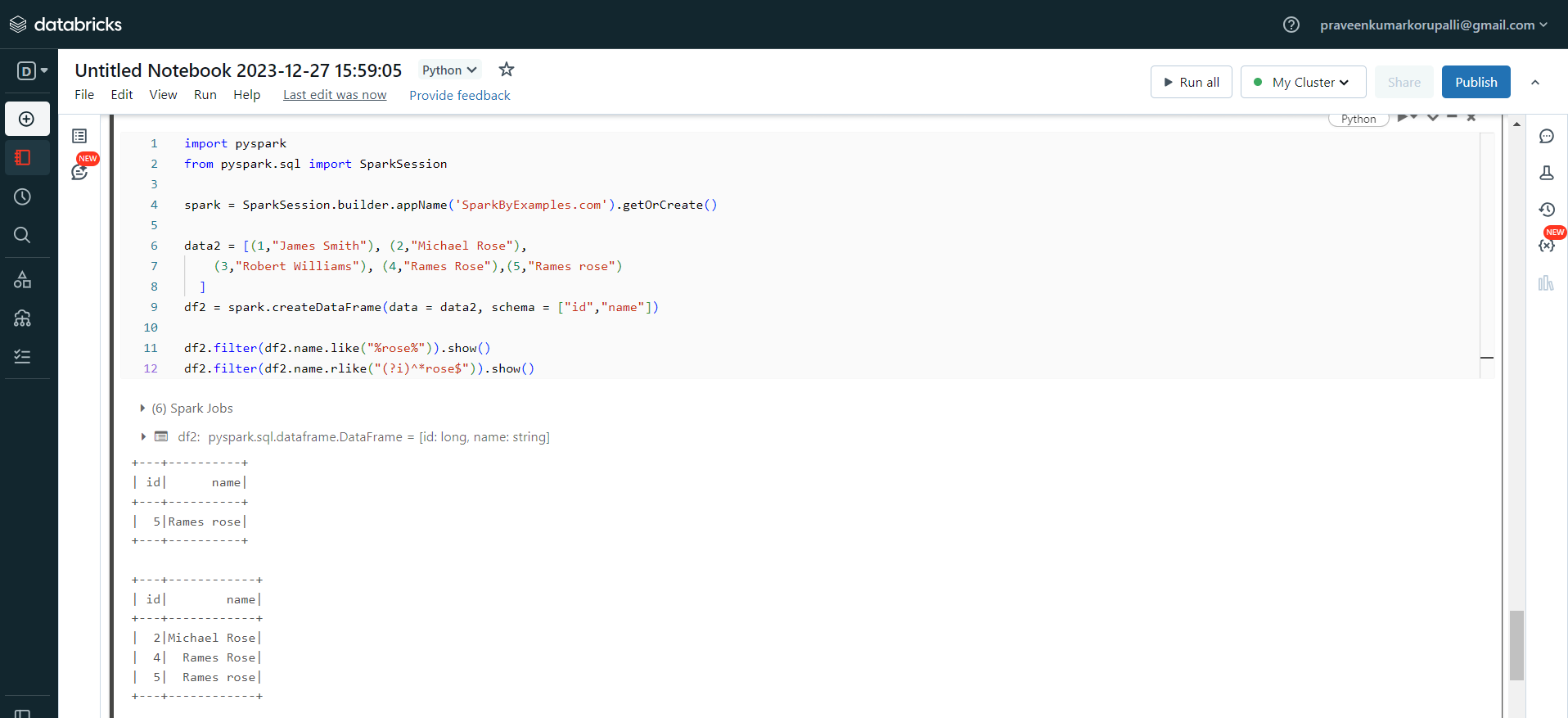


Update Columns:

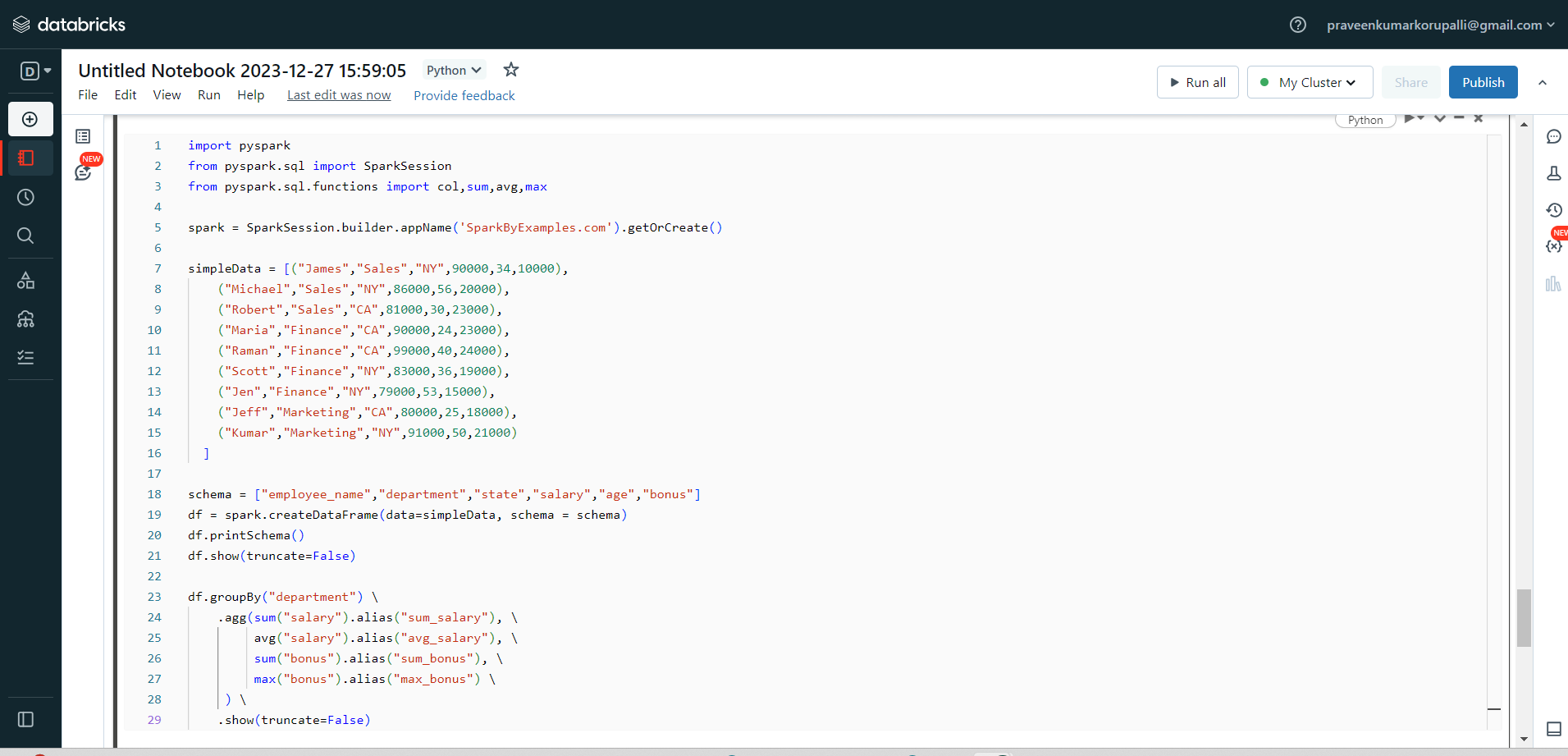




Filter:

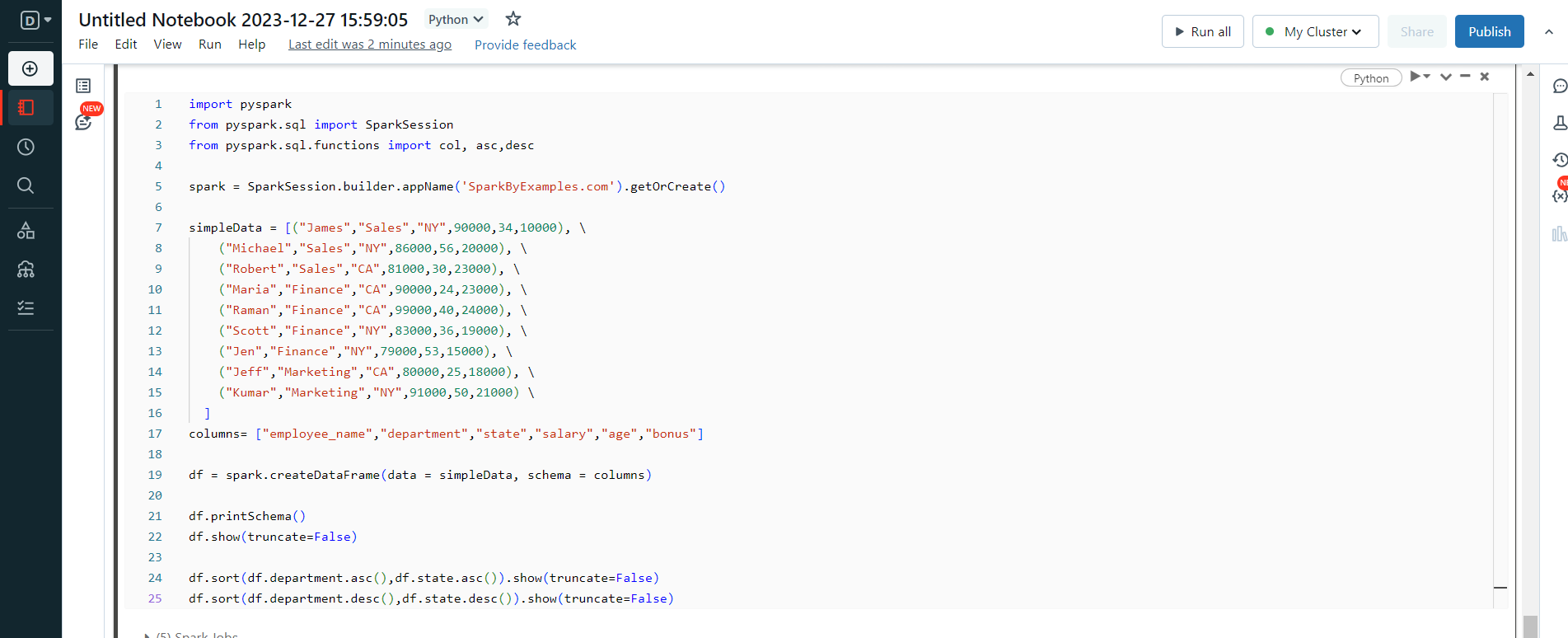


GroupBy:

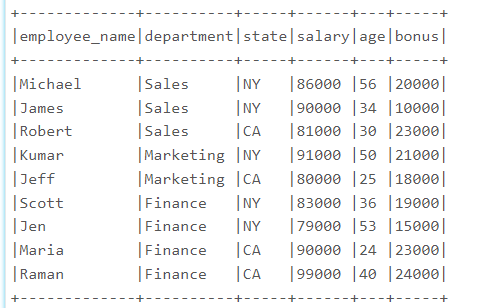
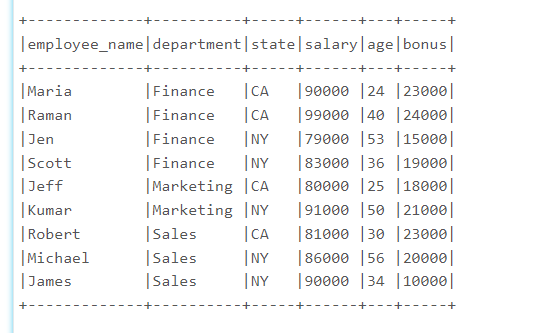




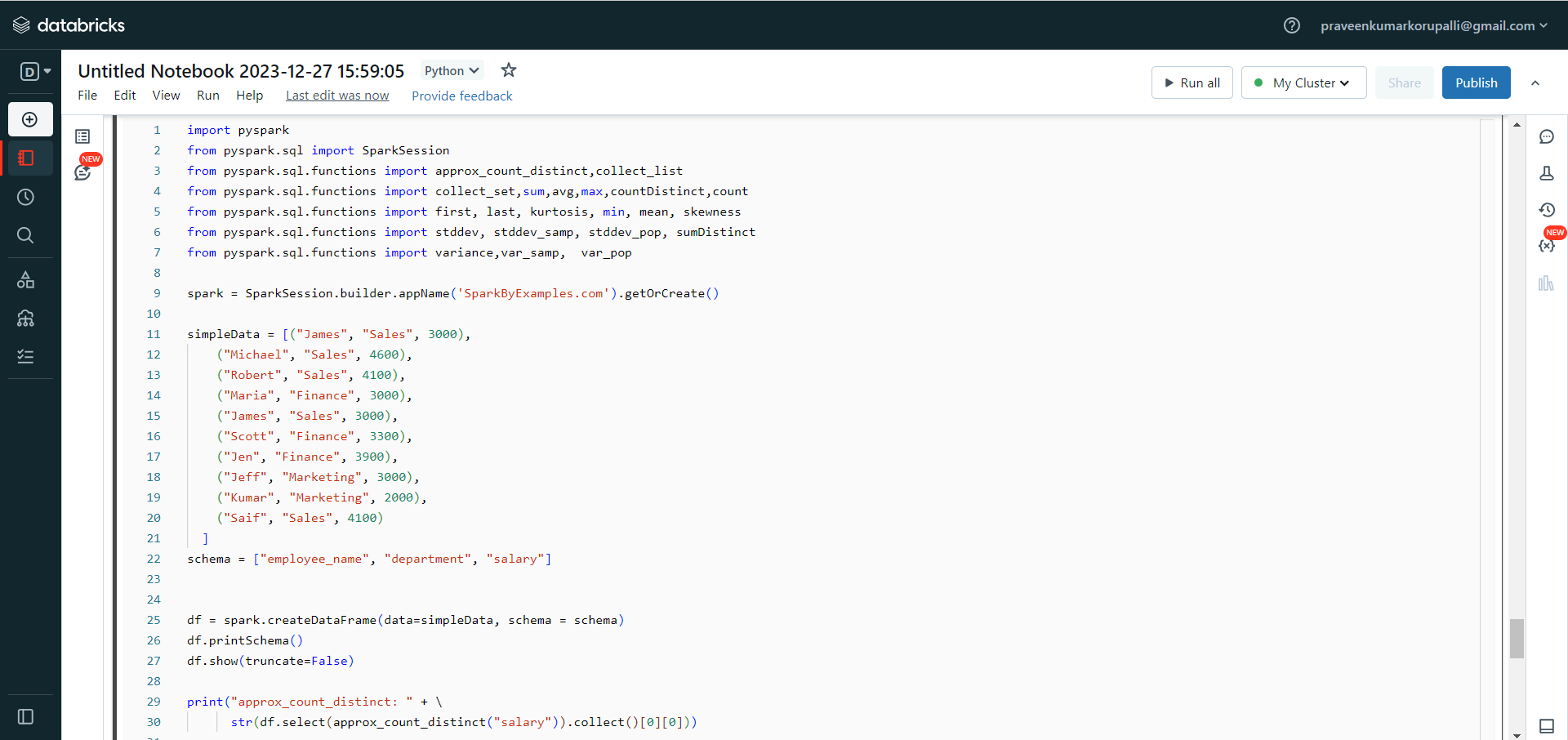
OrderBy:

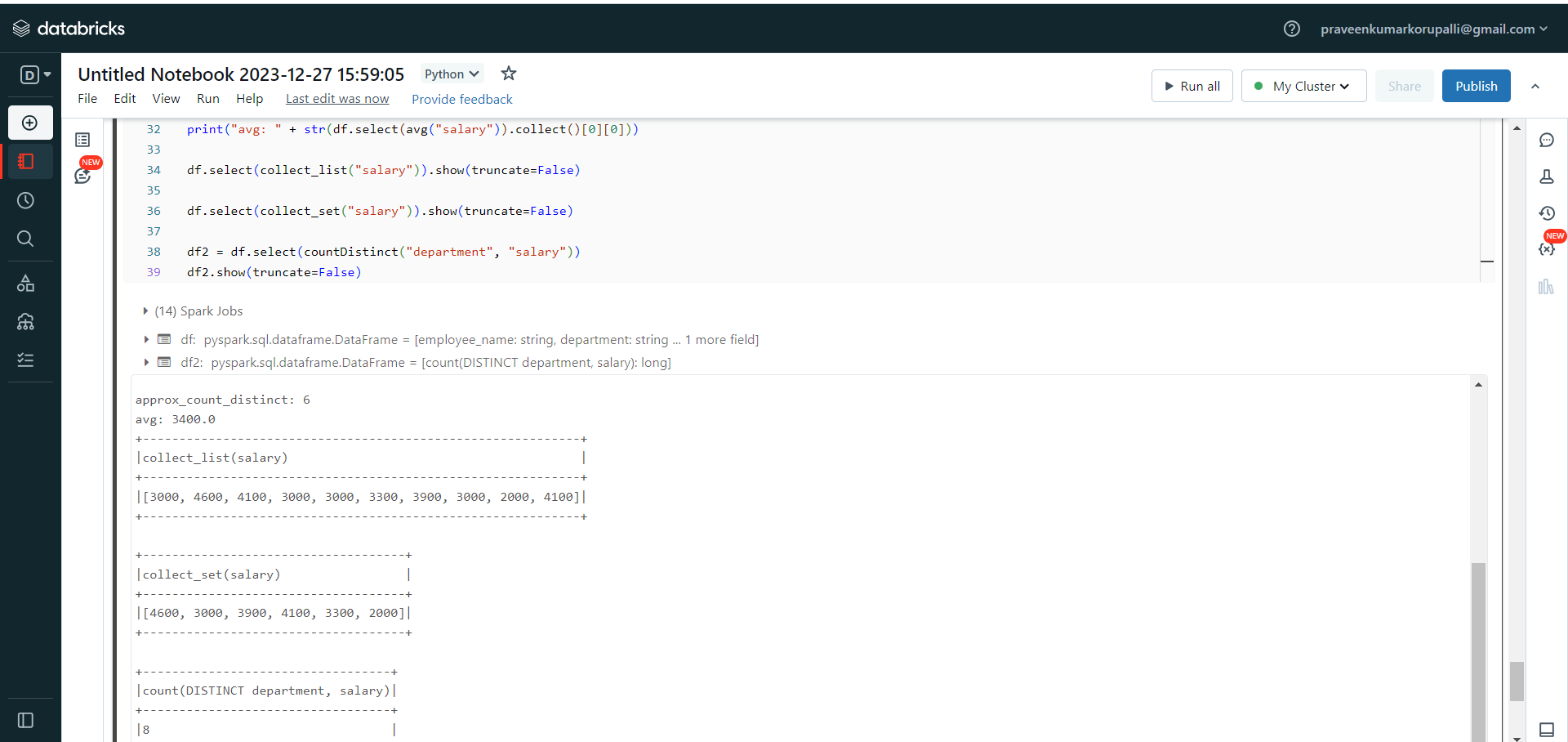




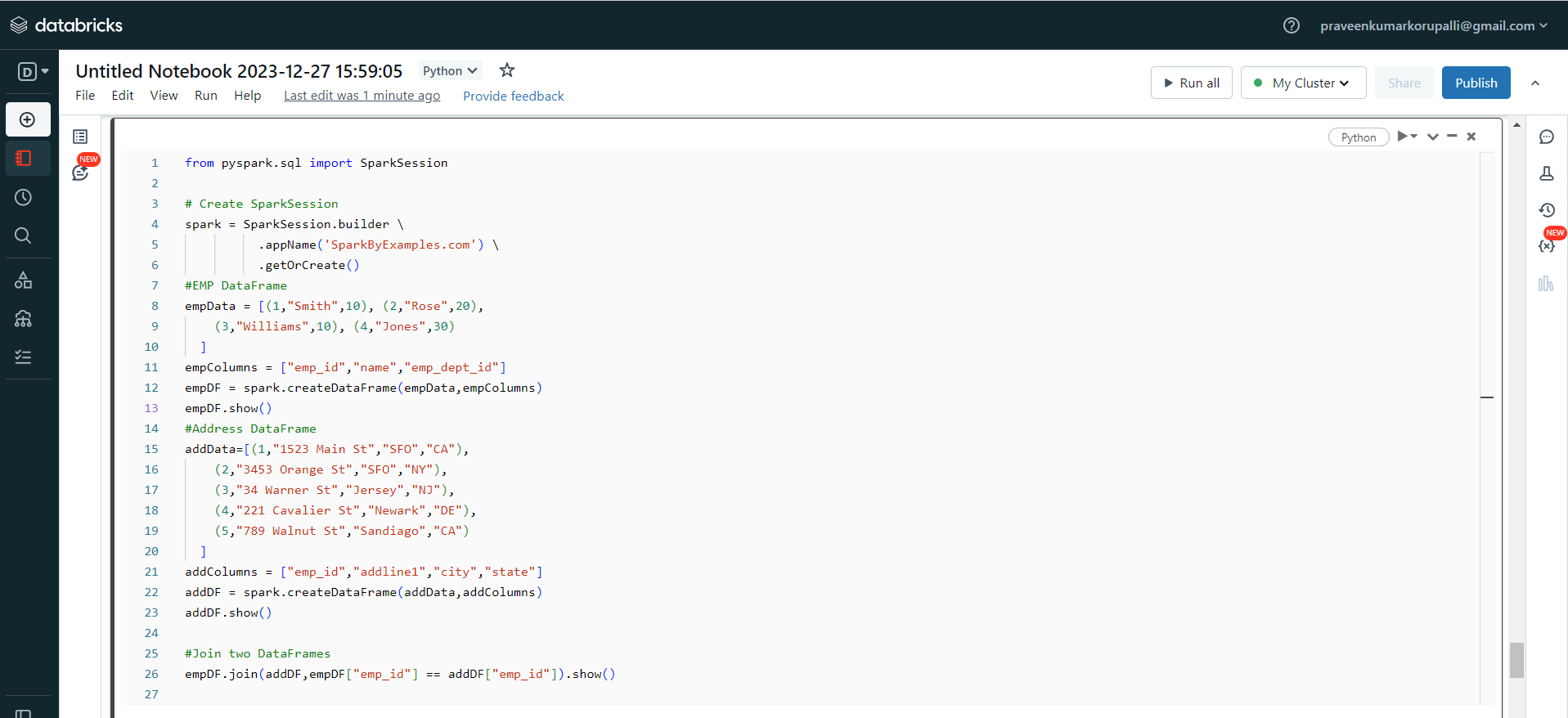


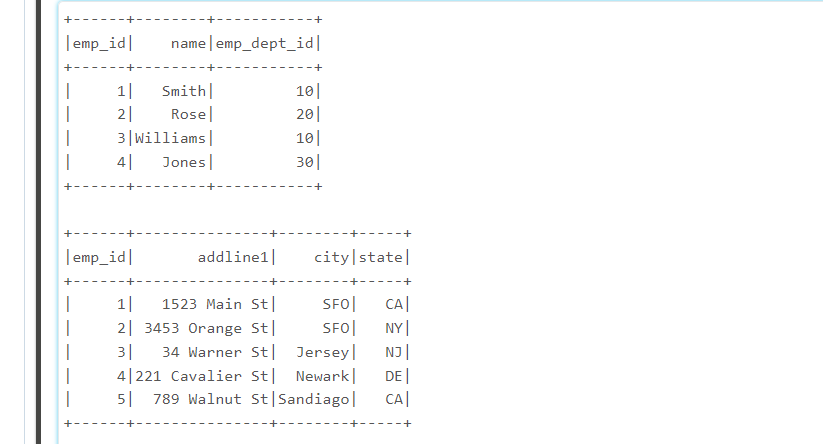
Aggregate Functions:

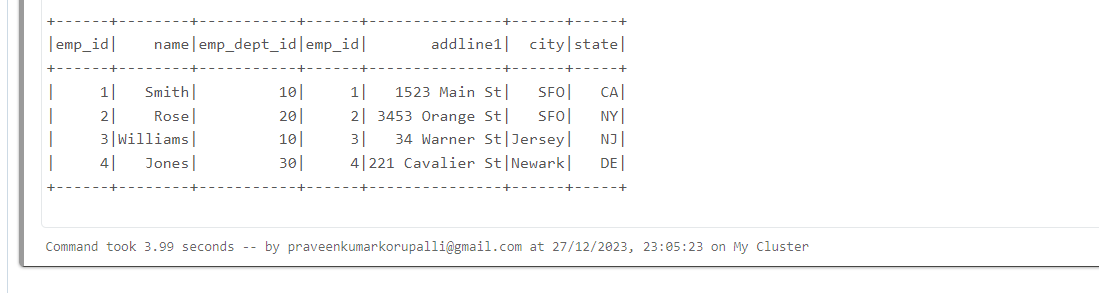




Joining two dataframes:

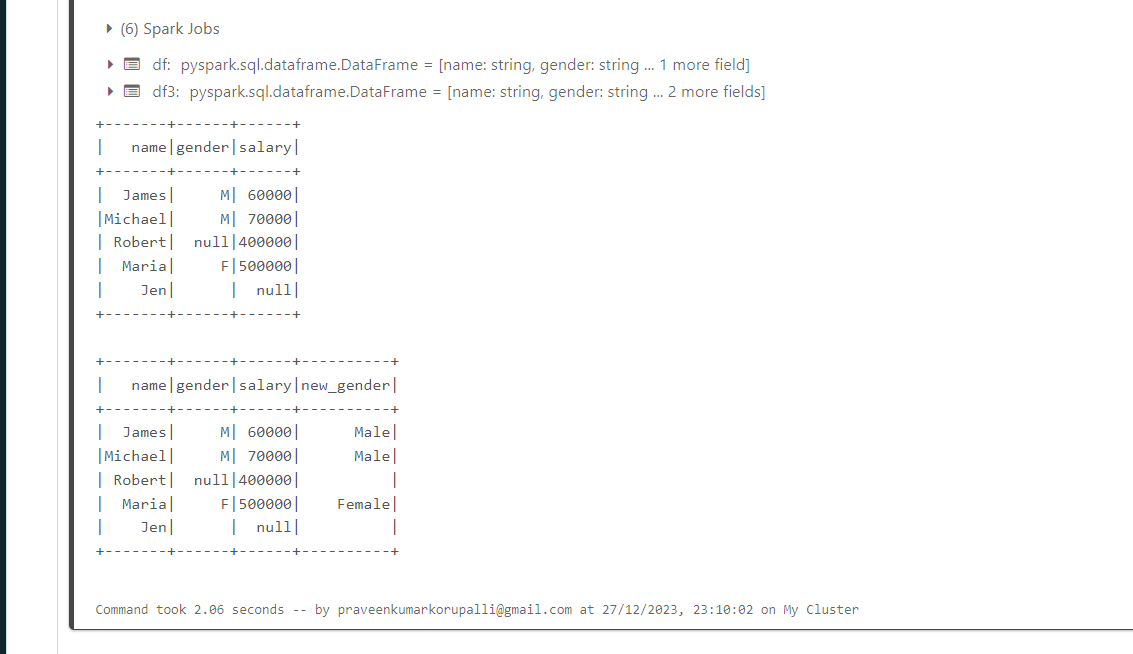






Sql case When:





Creating the Pandas DataFrame:

