27/12/2023

Chakradhar Bhogapurapu, BIBA

Pyspark Daily Assessment

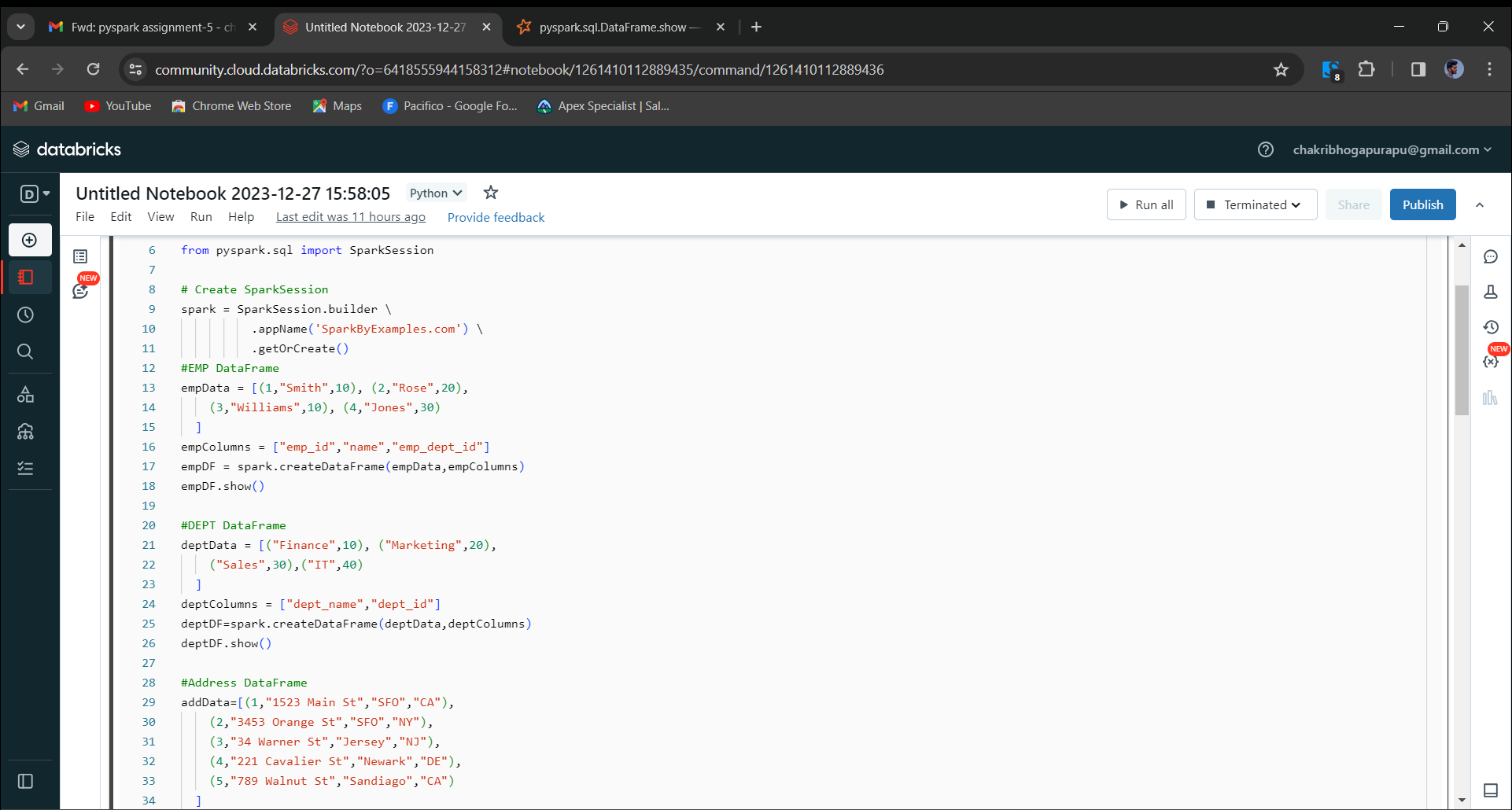
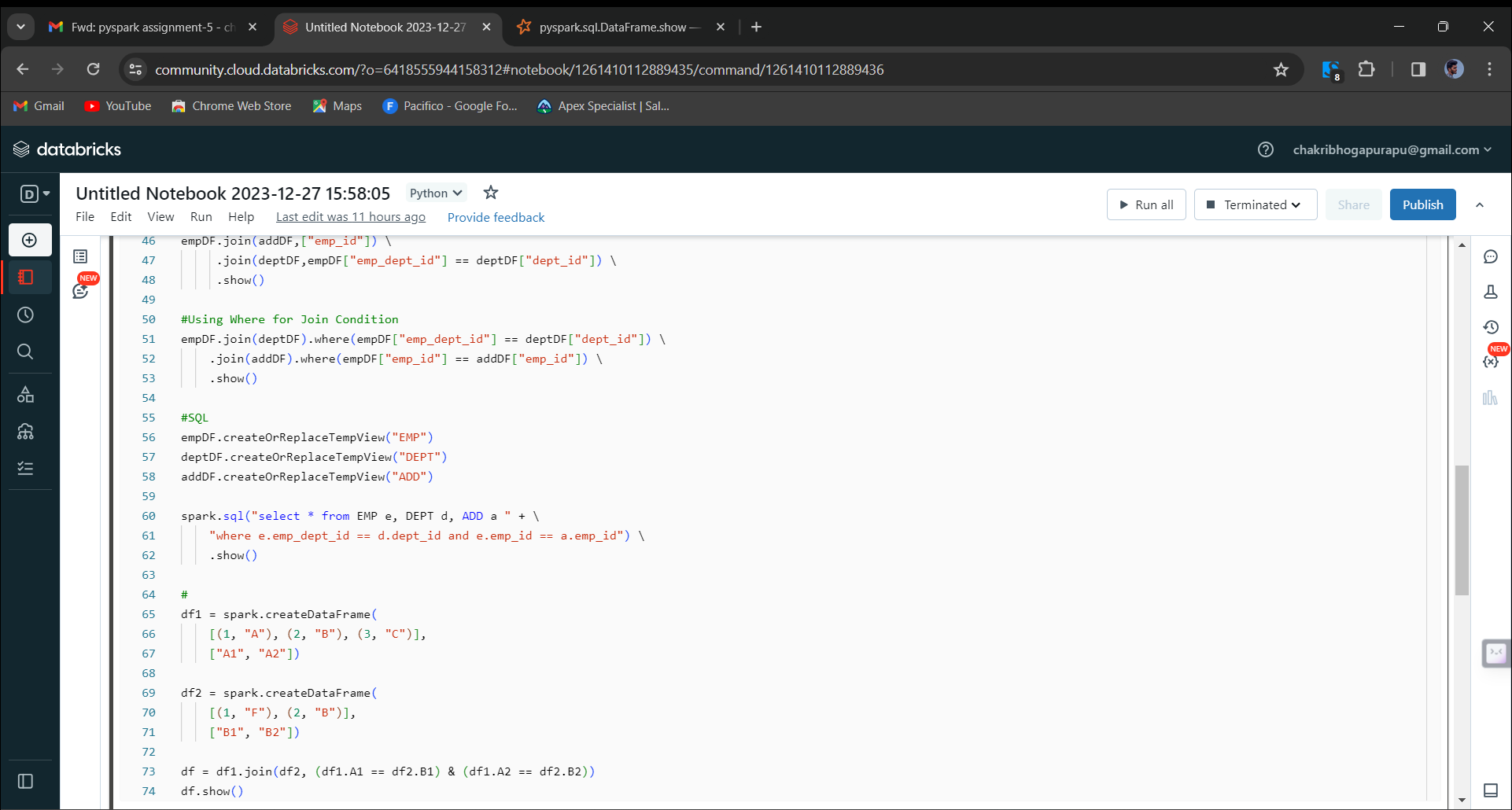
🡪Reading a csv file



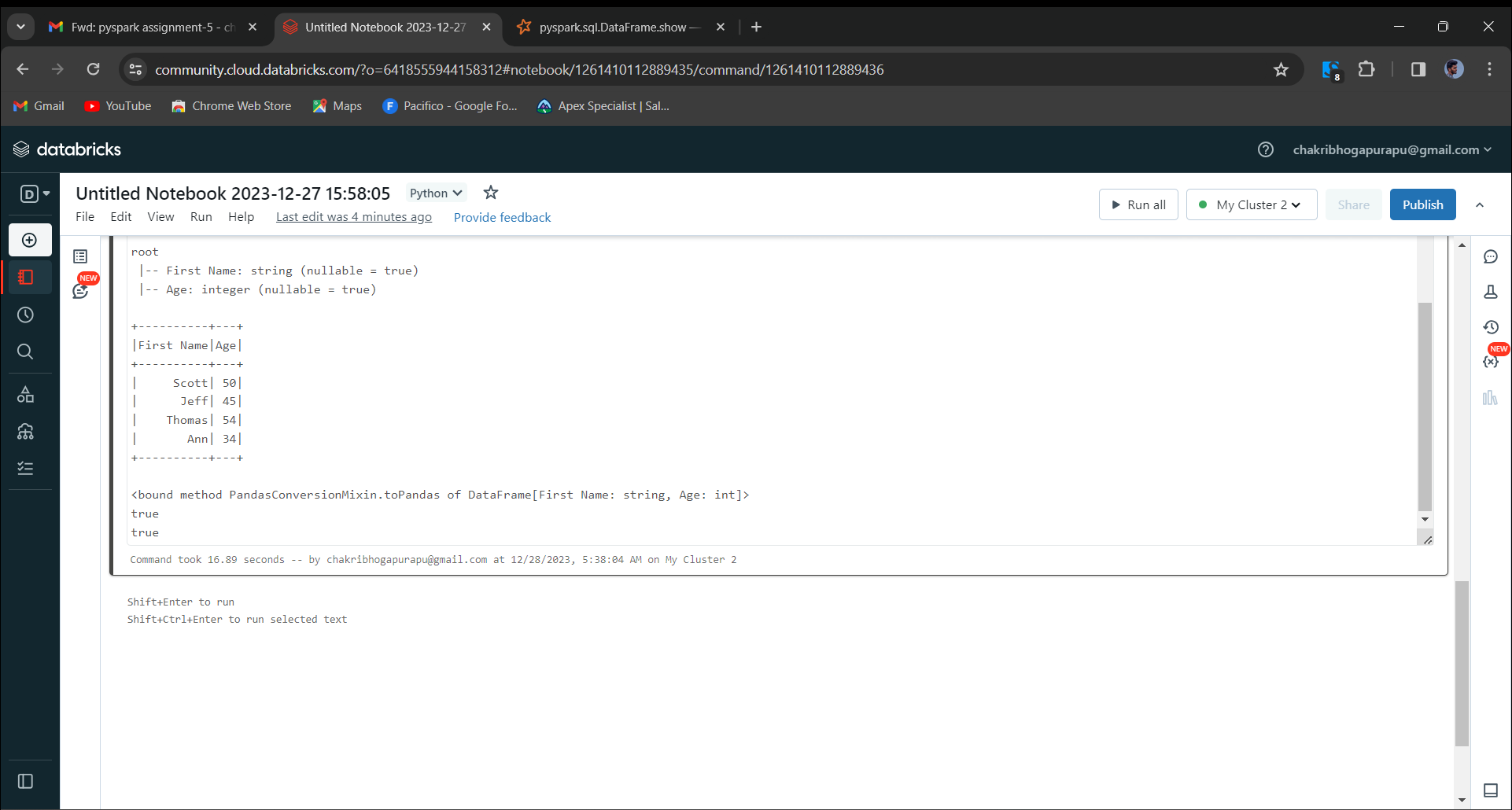
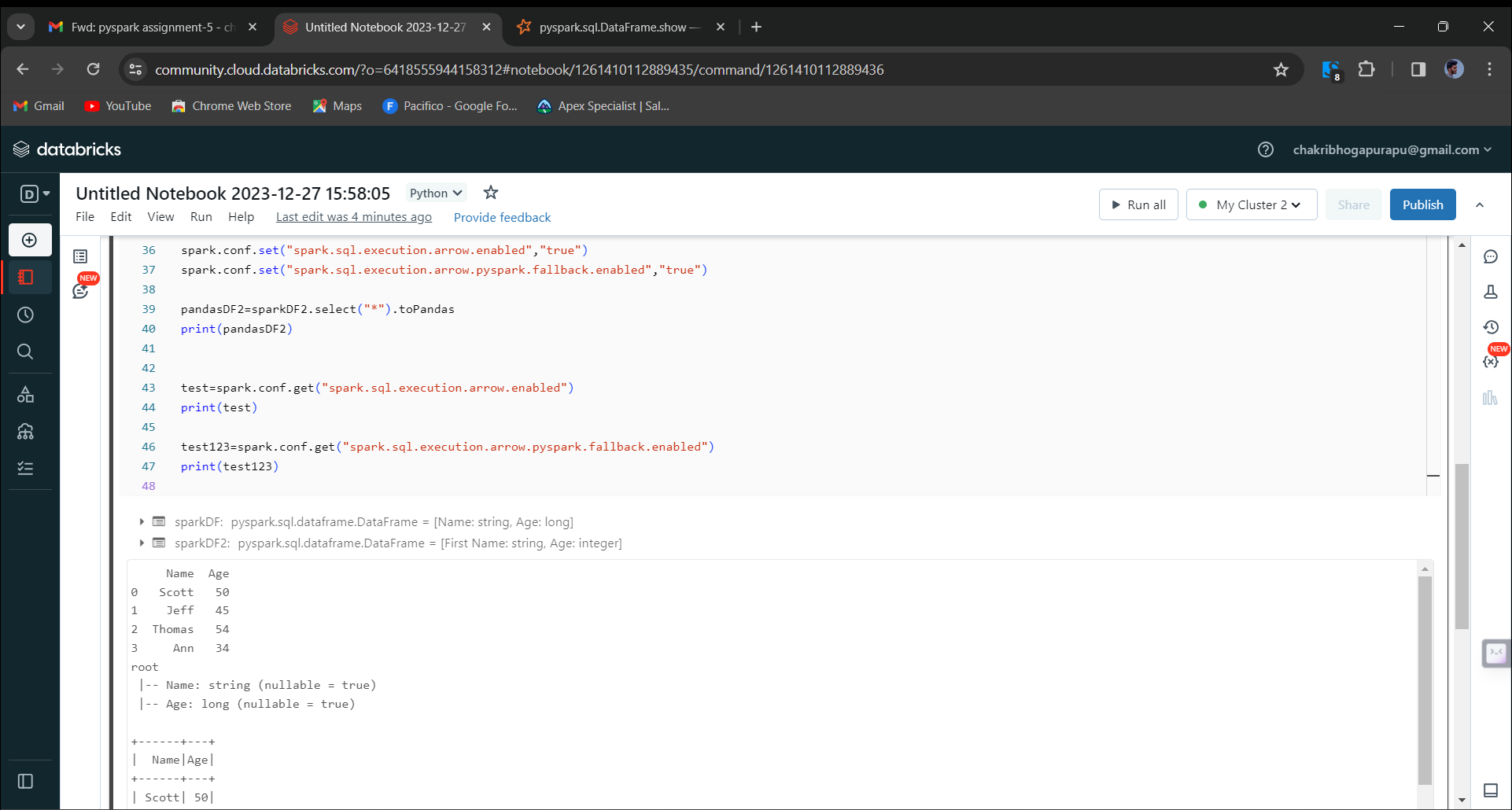
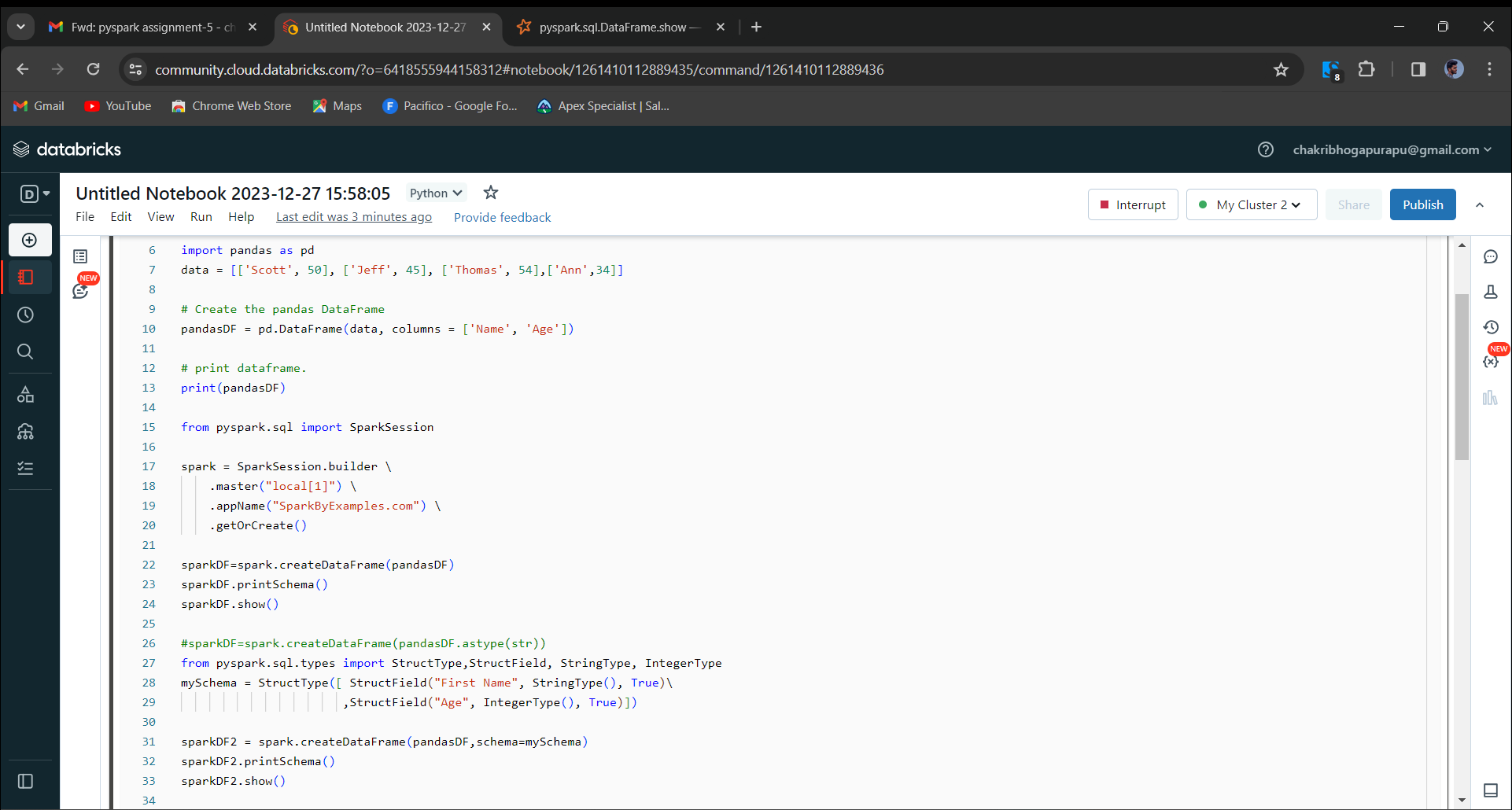
To load csv file into the pyspark session.

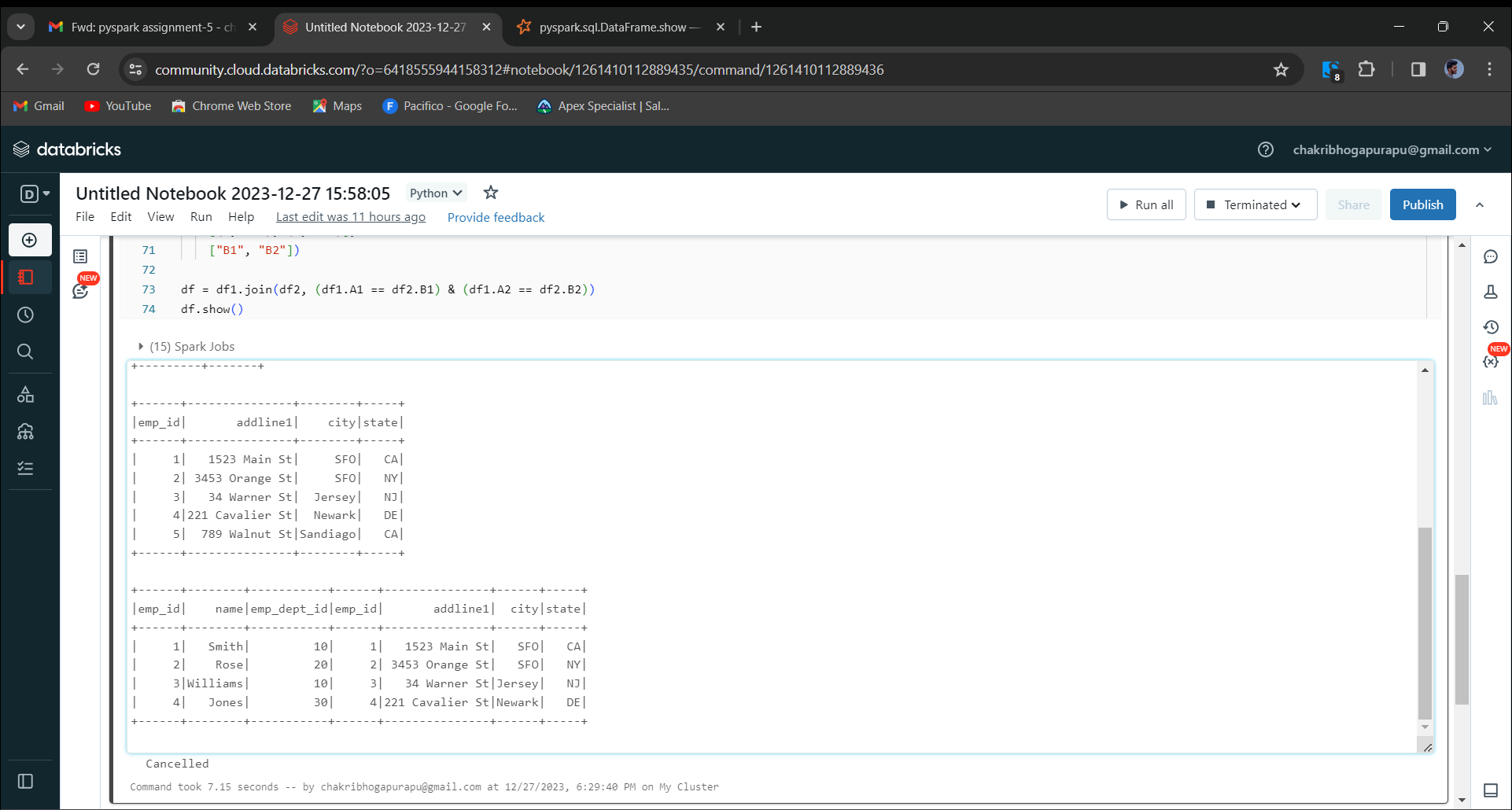
Then the method show() will present the data present in the csv file.

🡪Joins

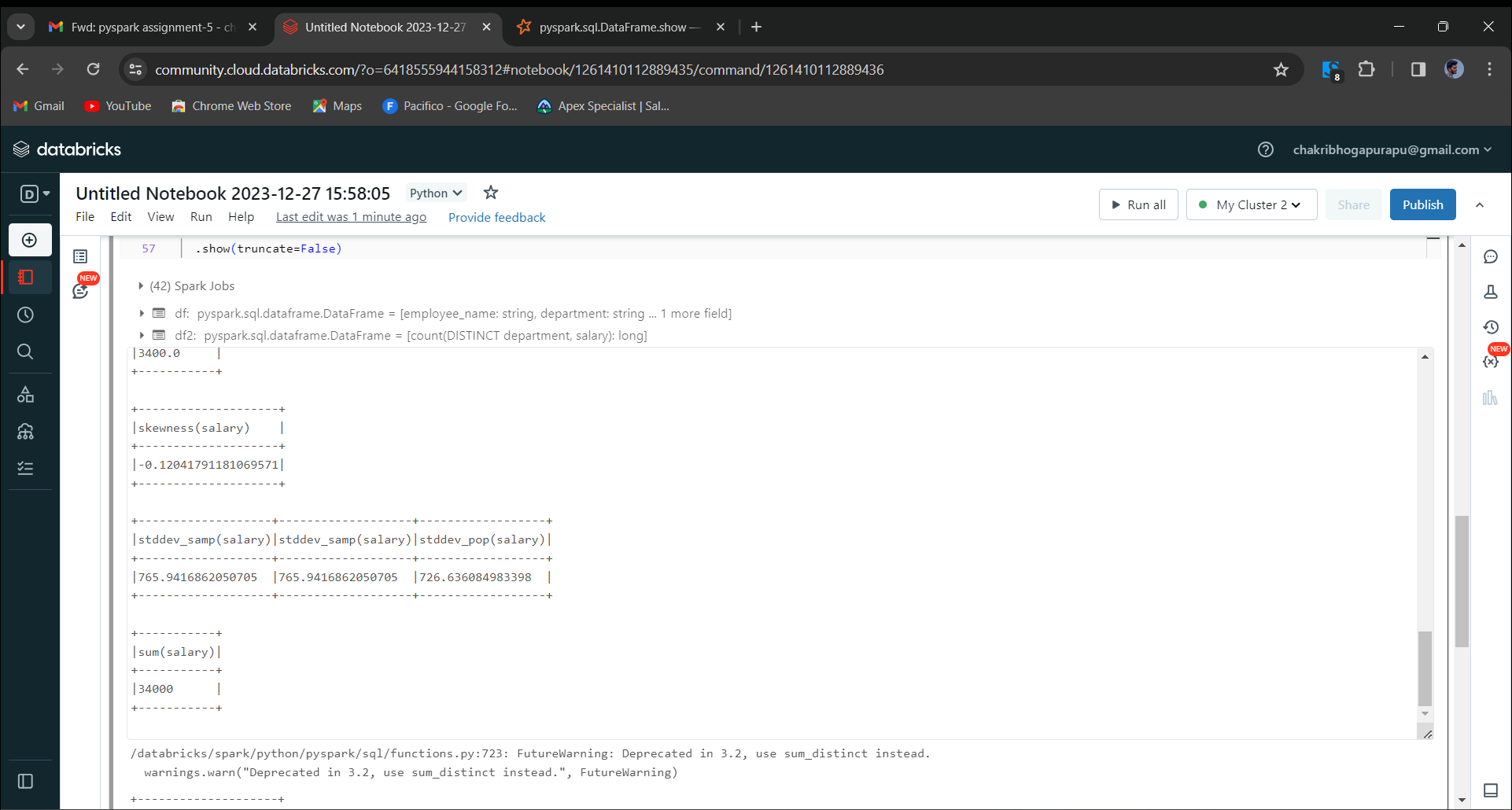
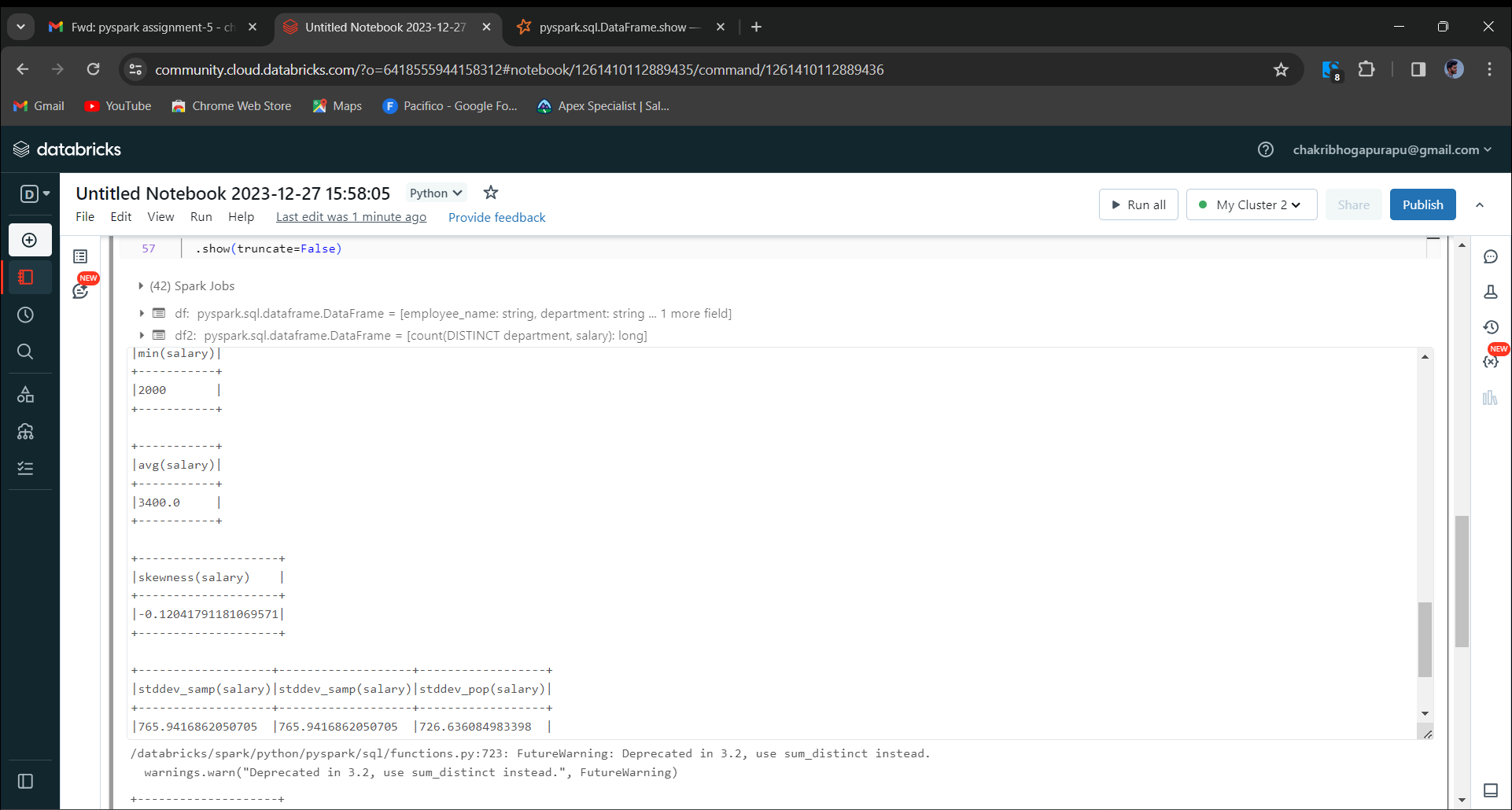
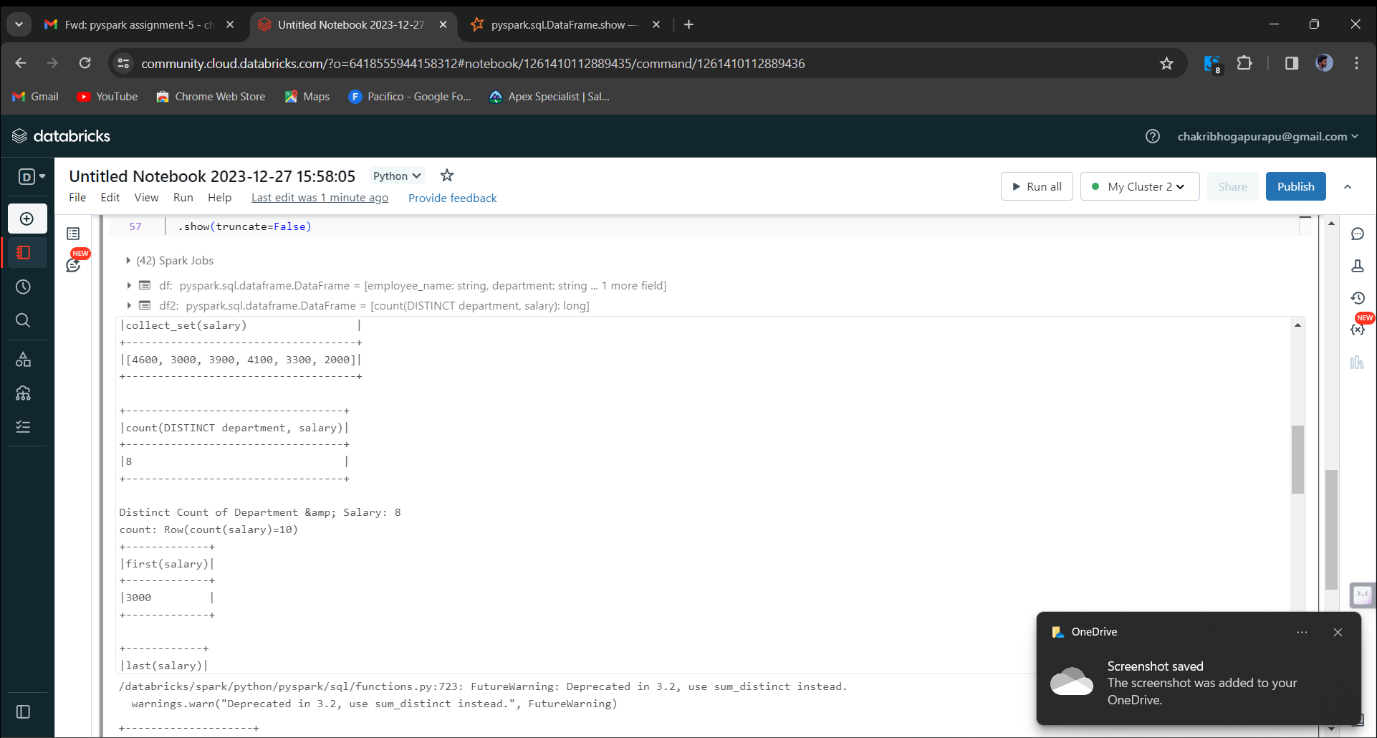
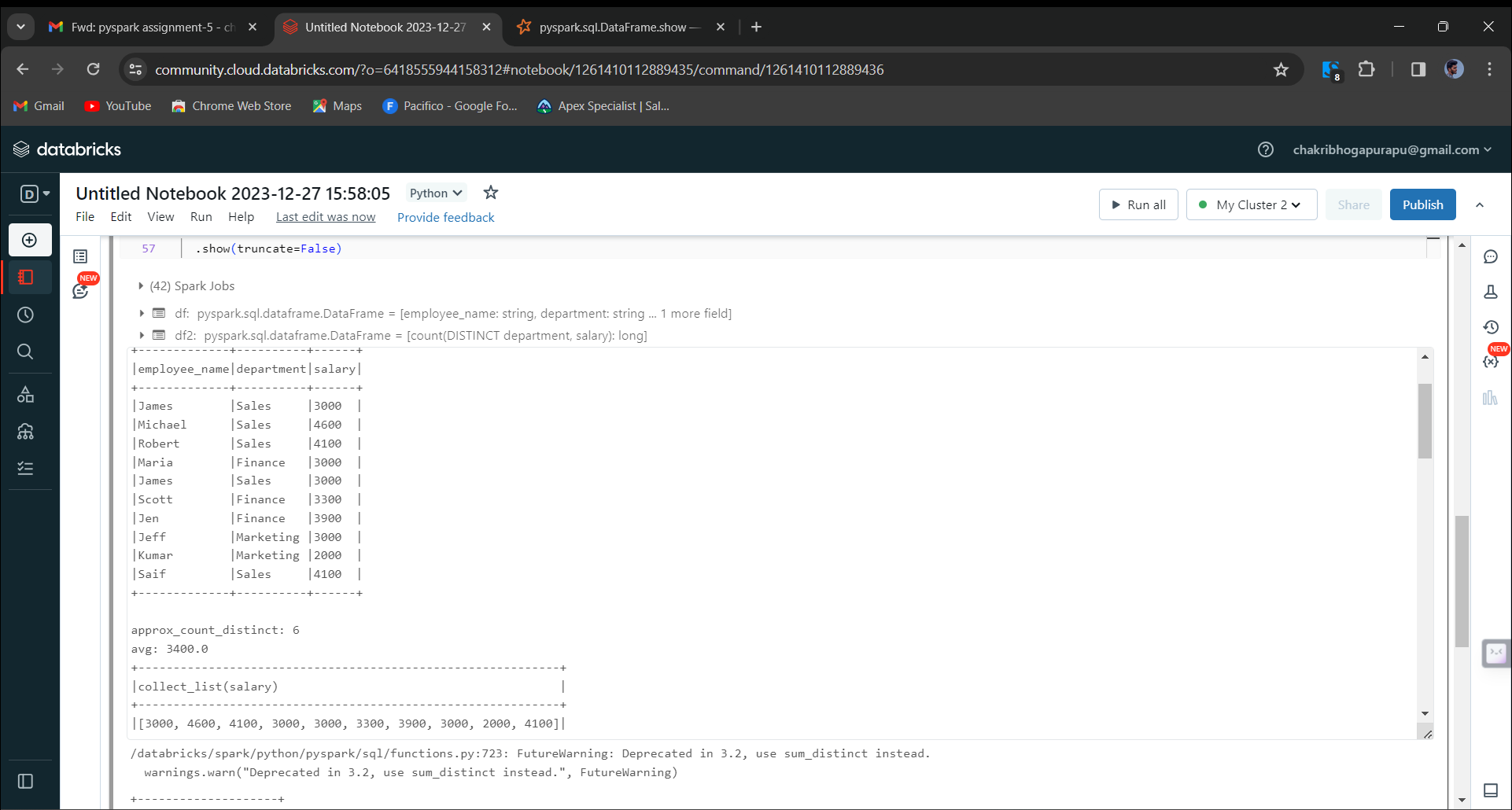
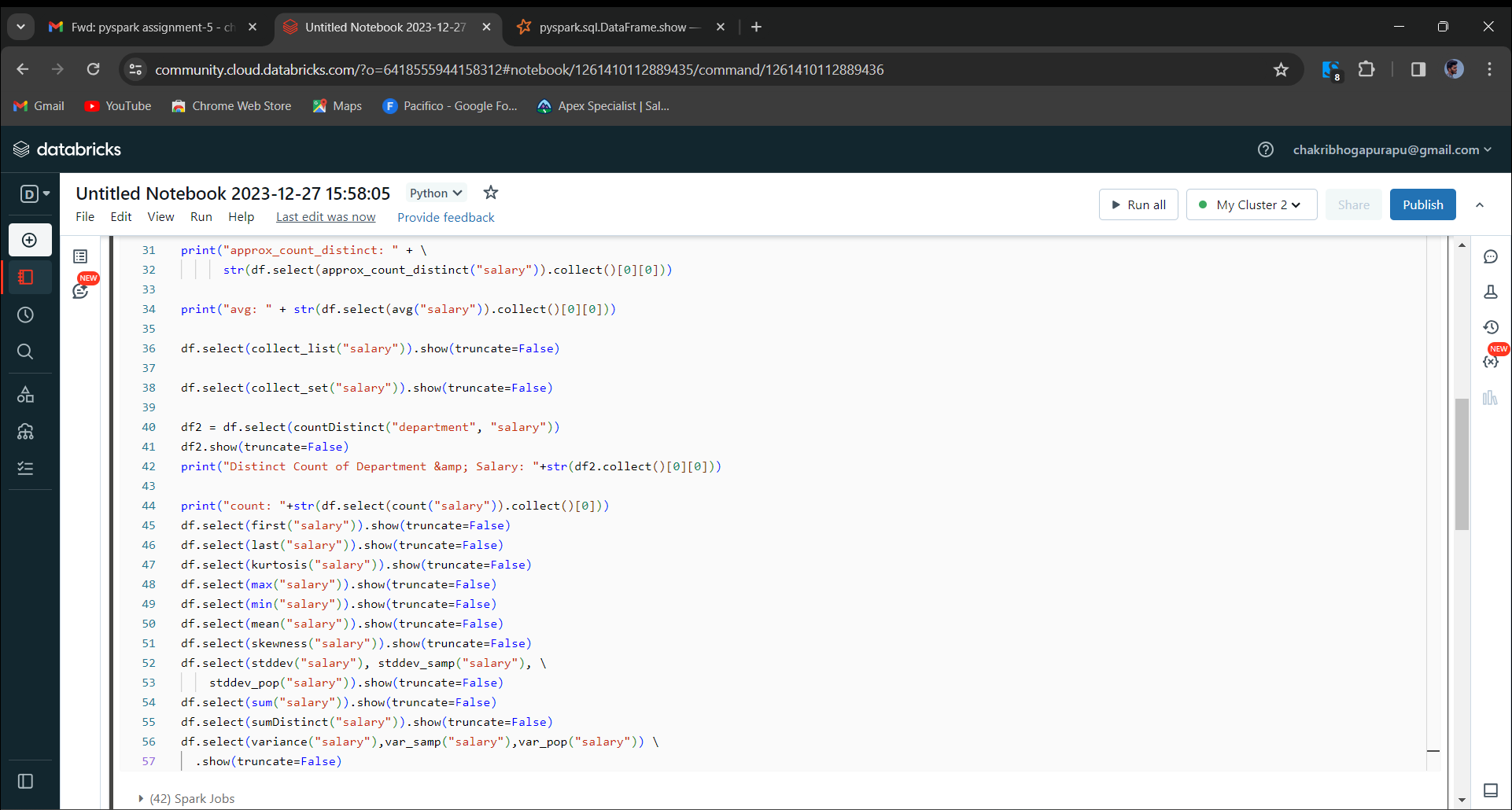
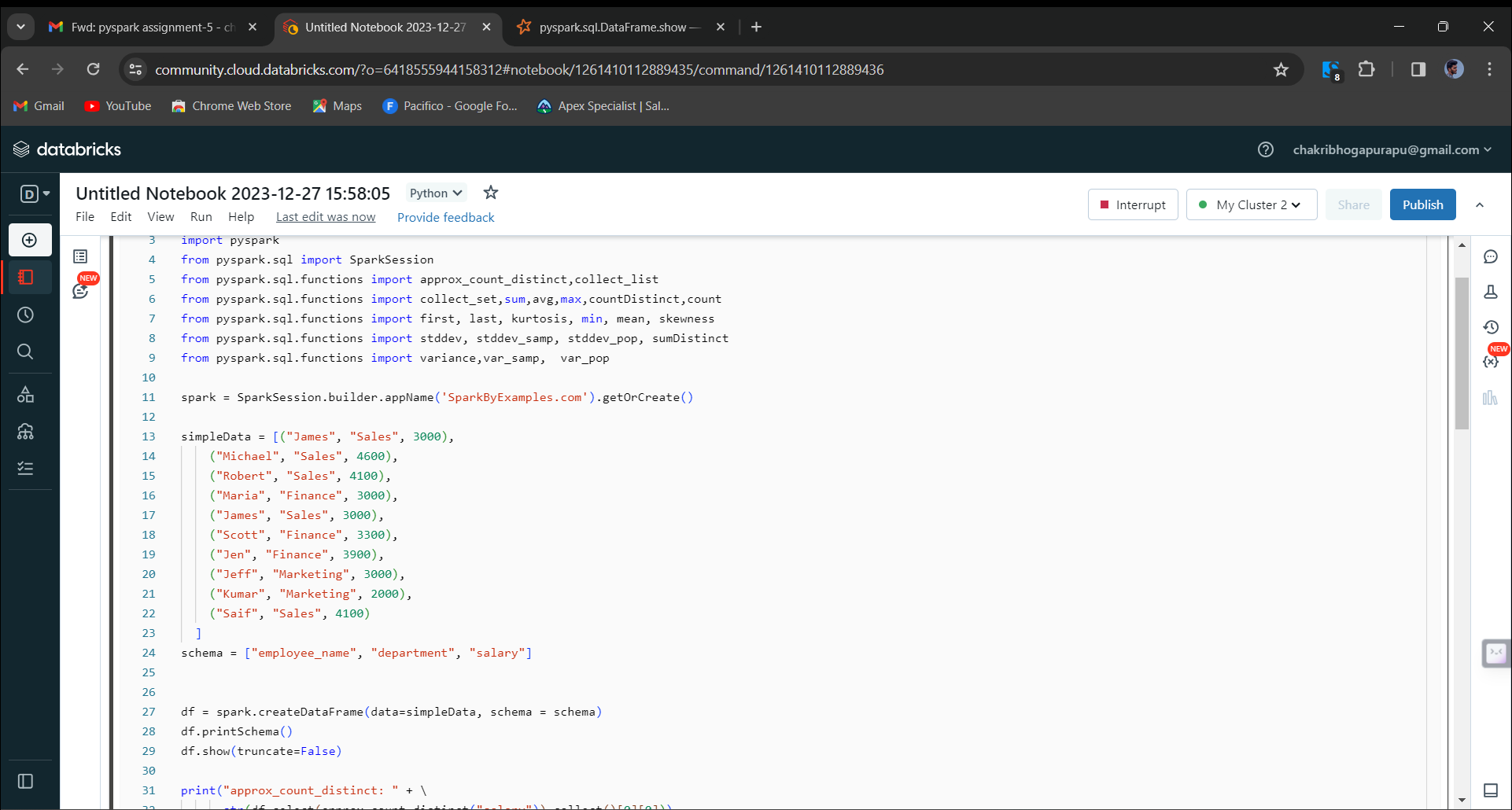


🡪Pandas Pyspark Dataframe

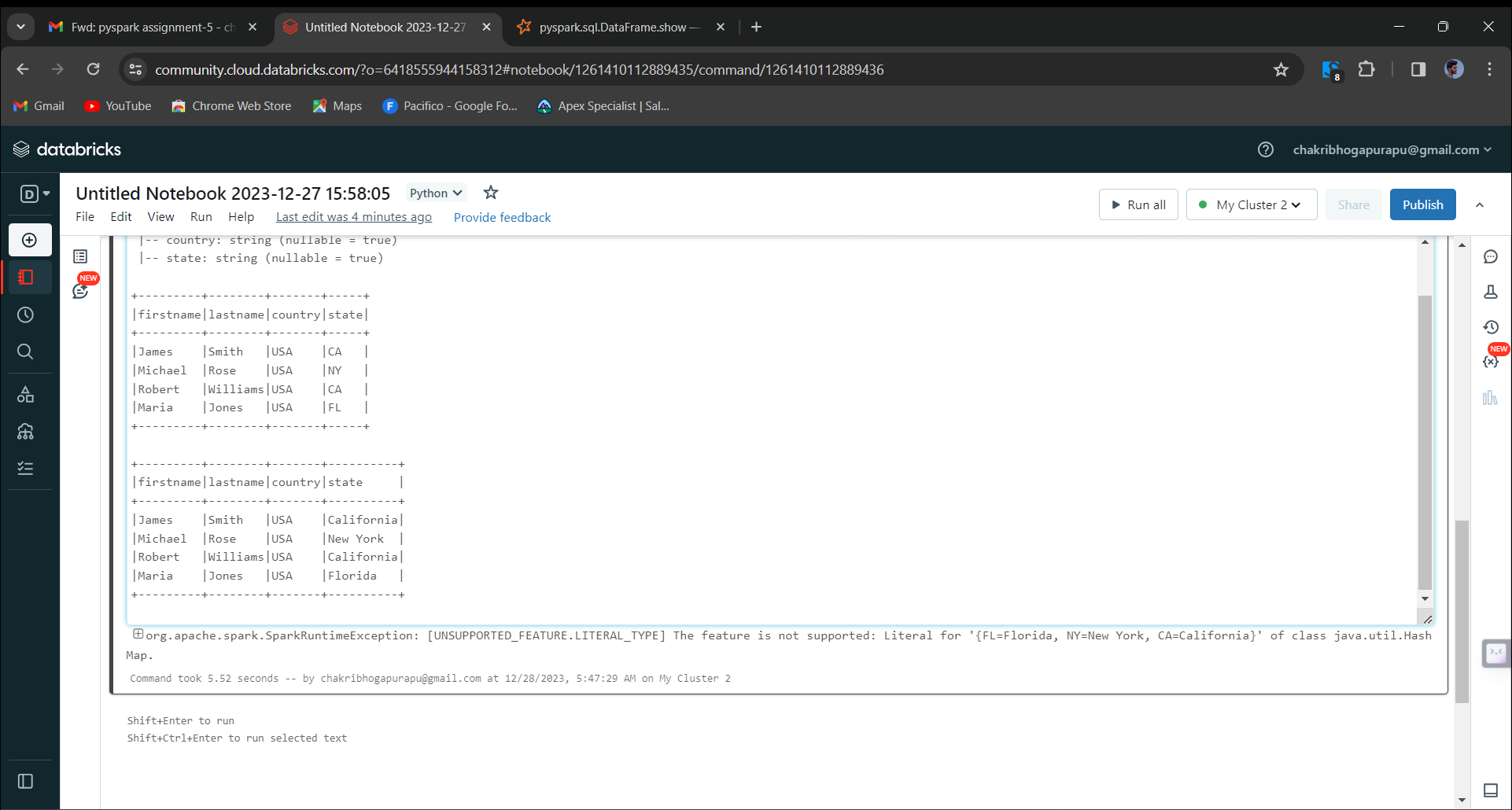
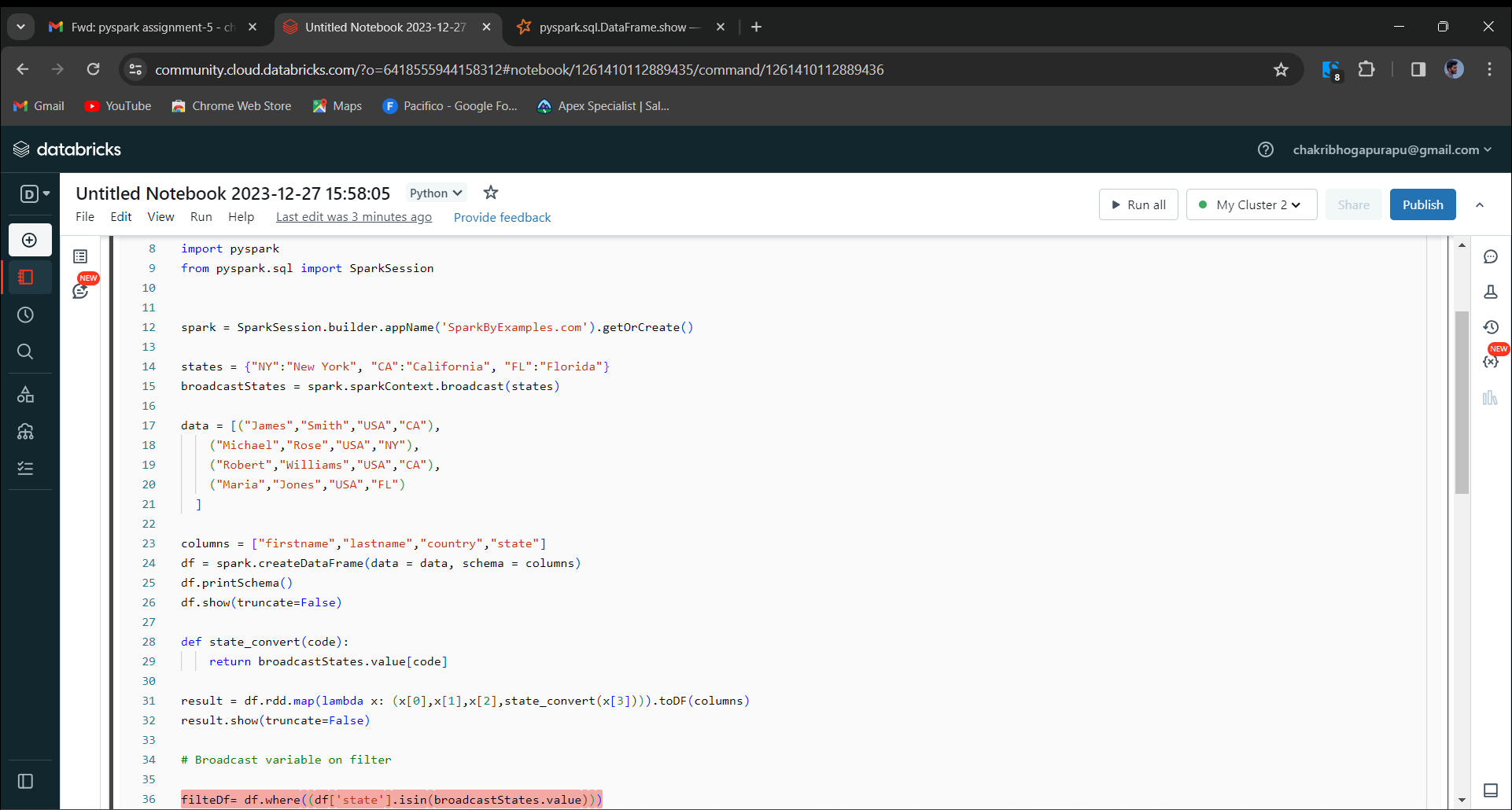




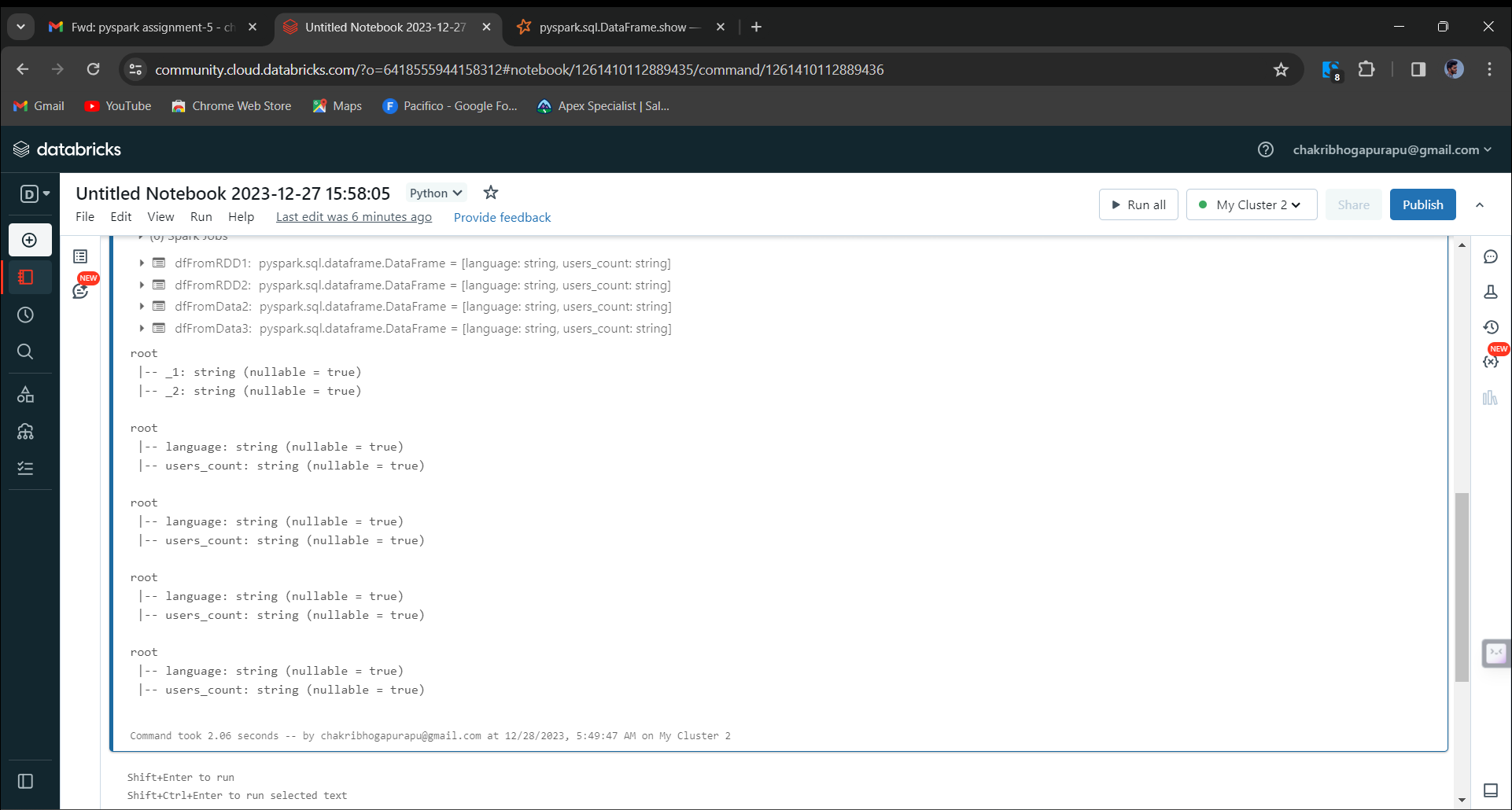
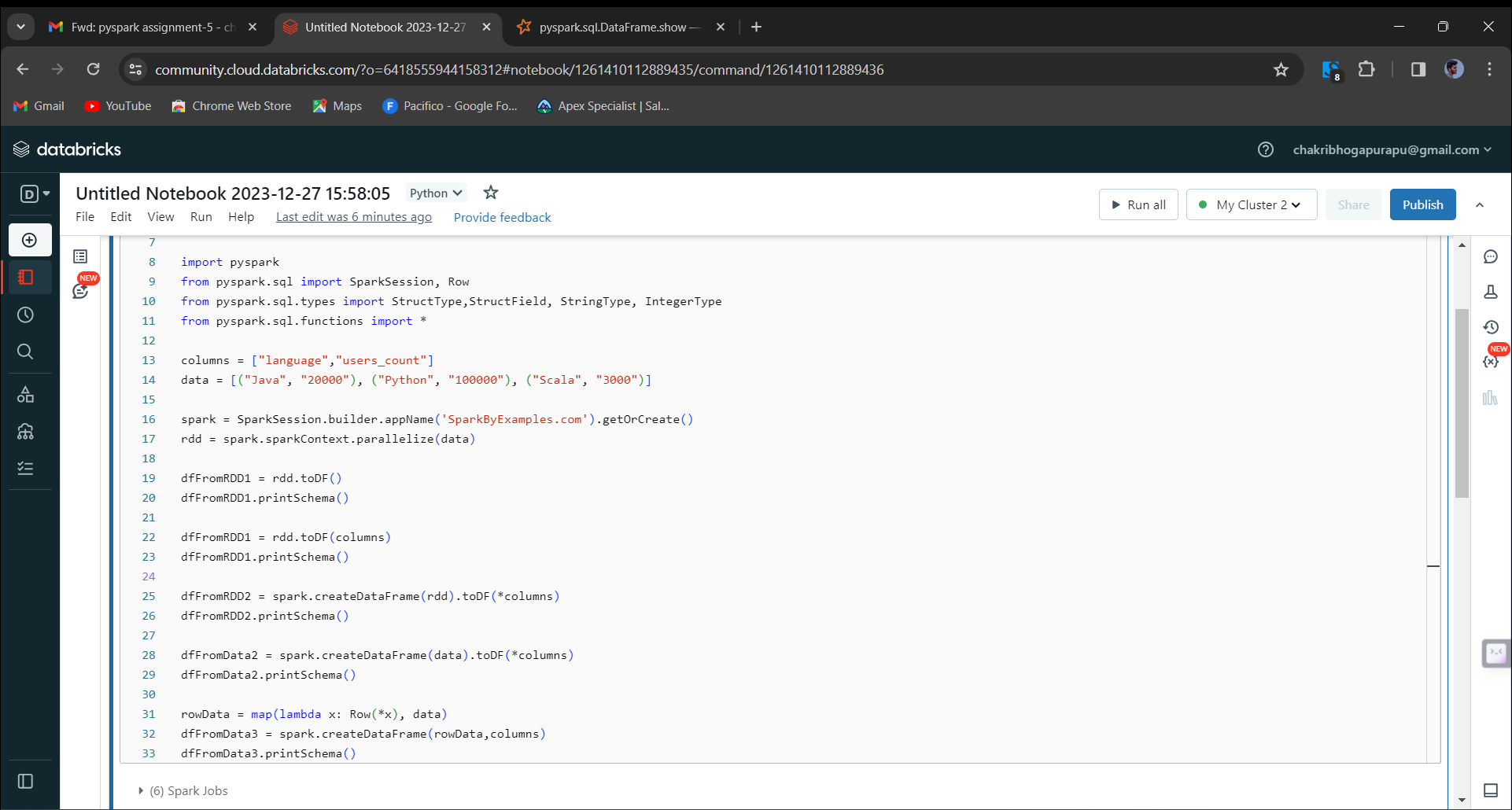
🡪Pyspark aggregate



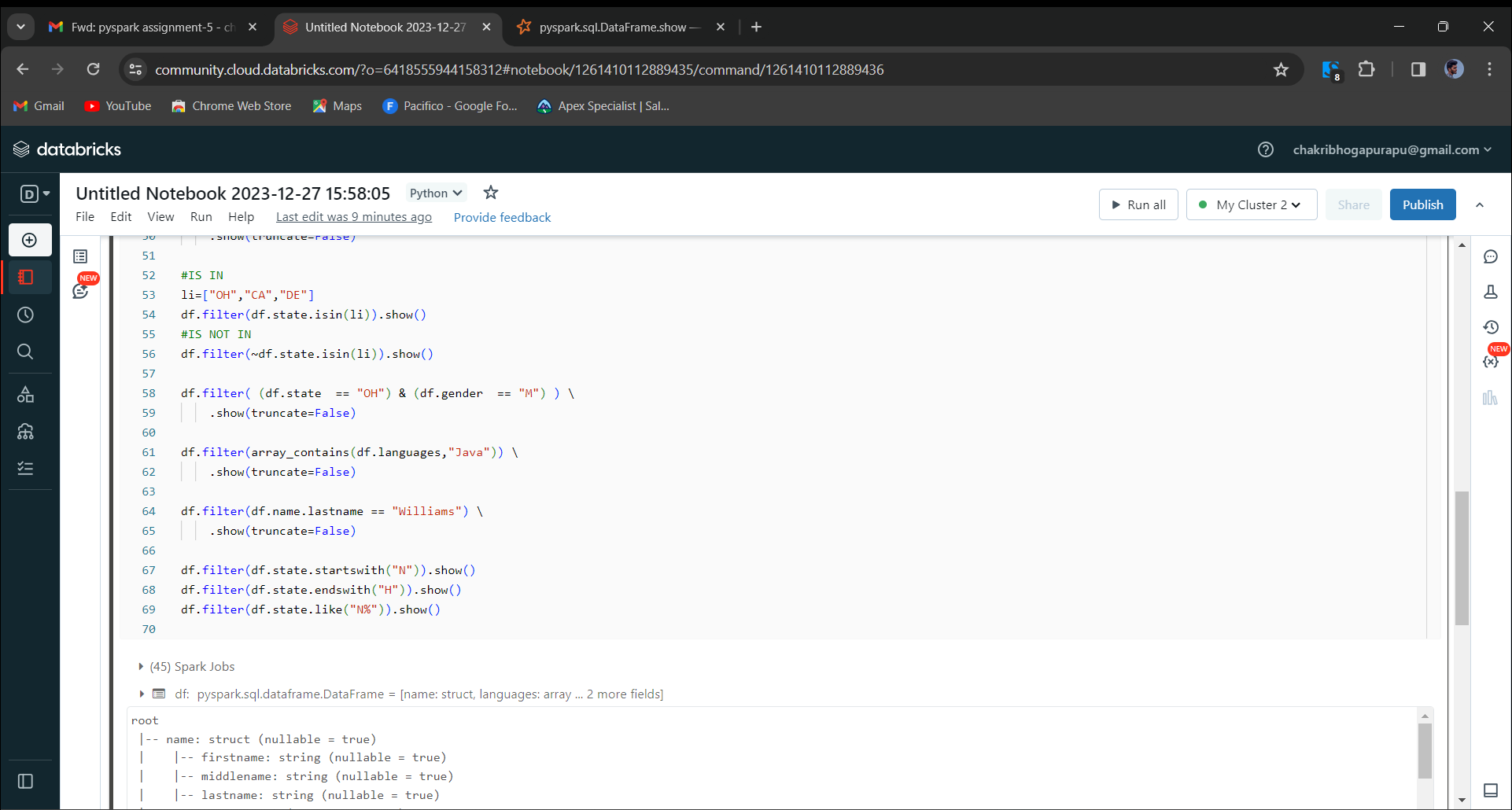
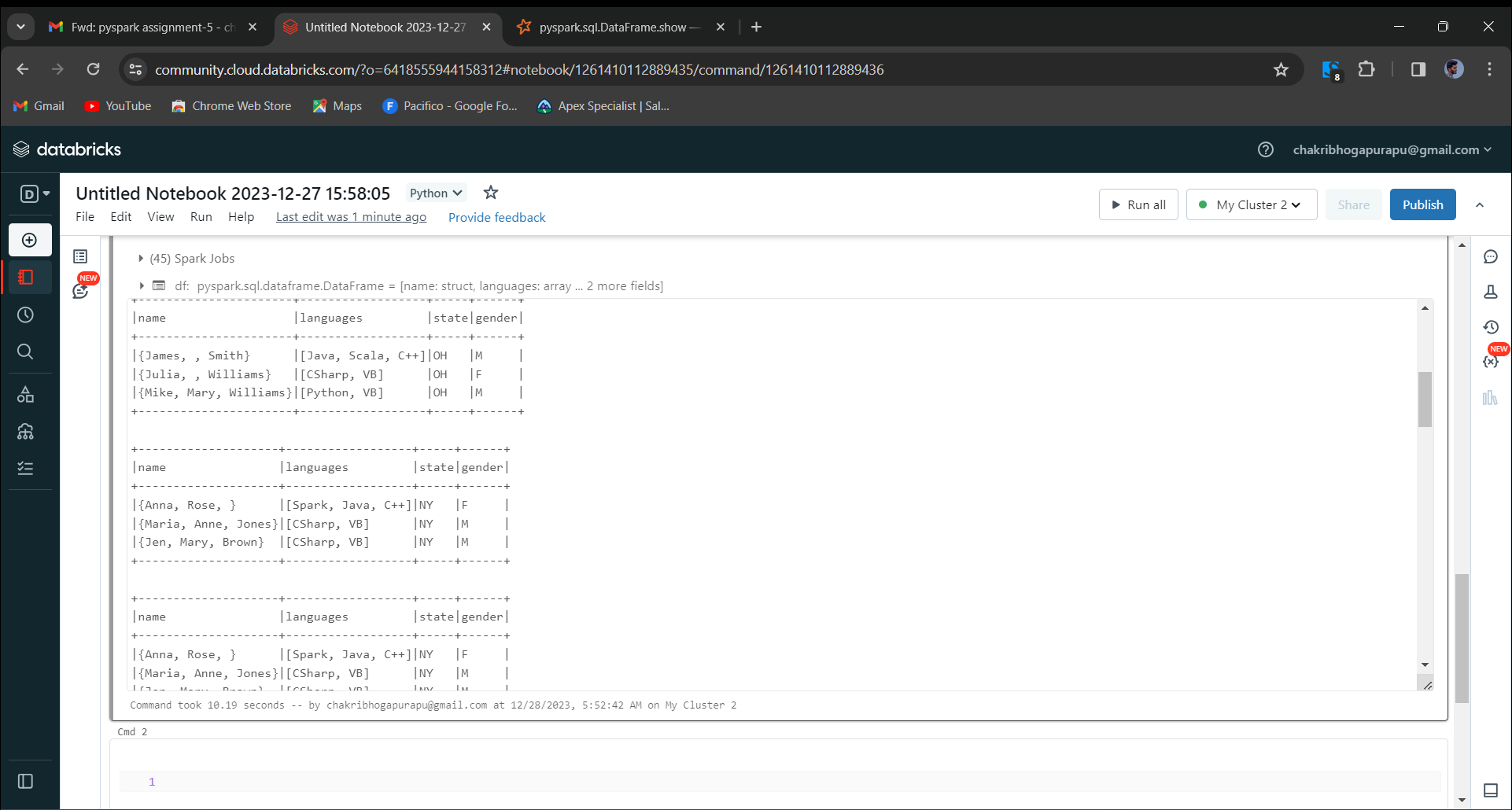
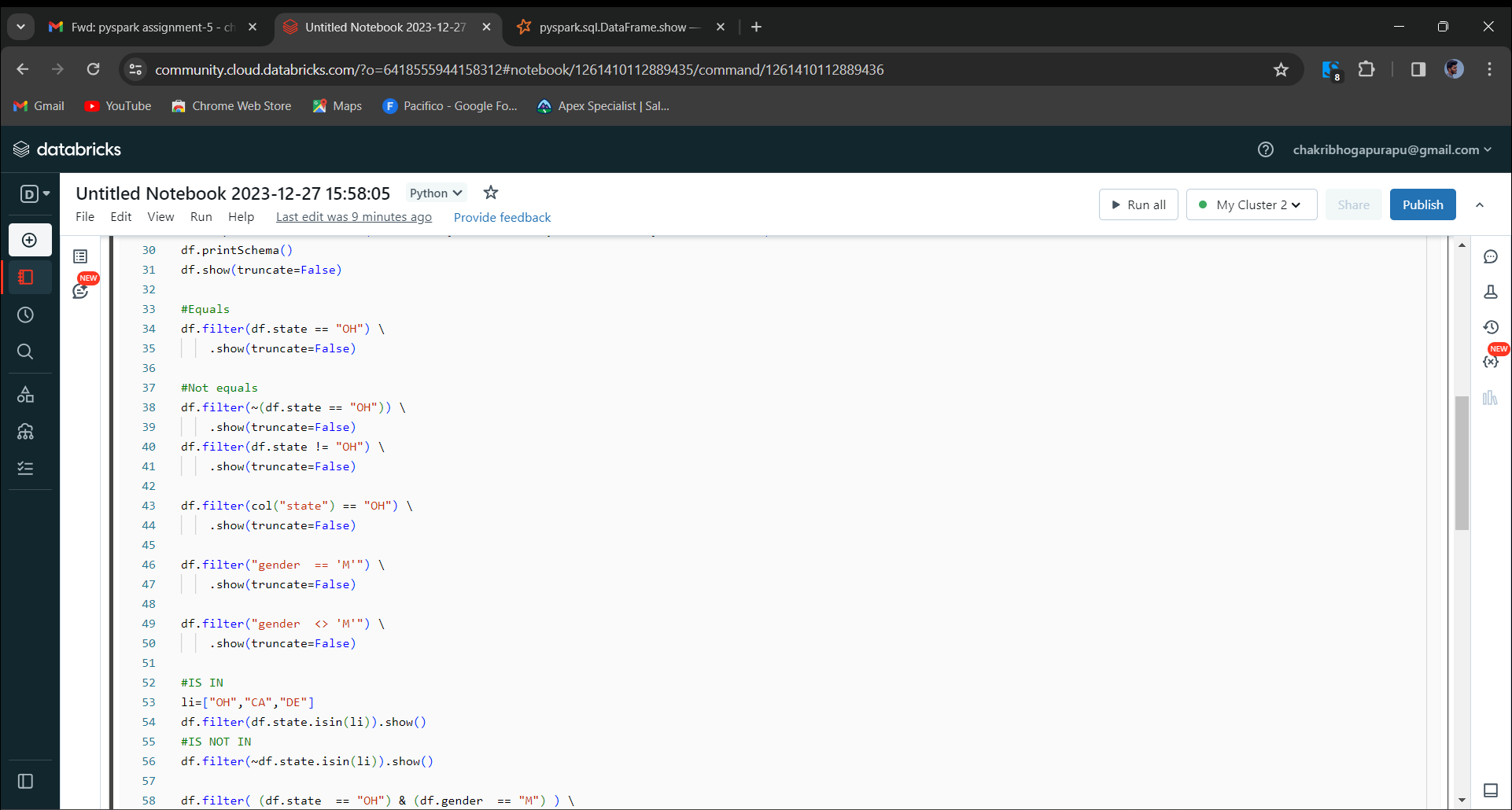
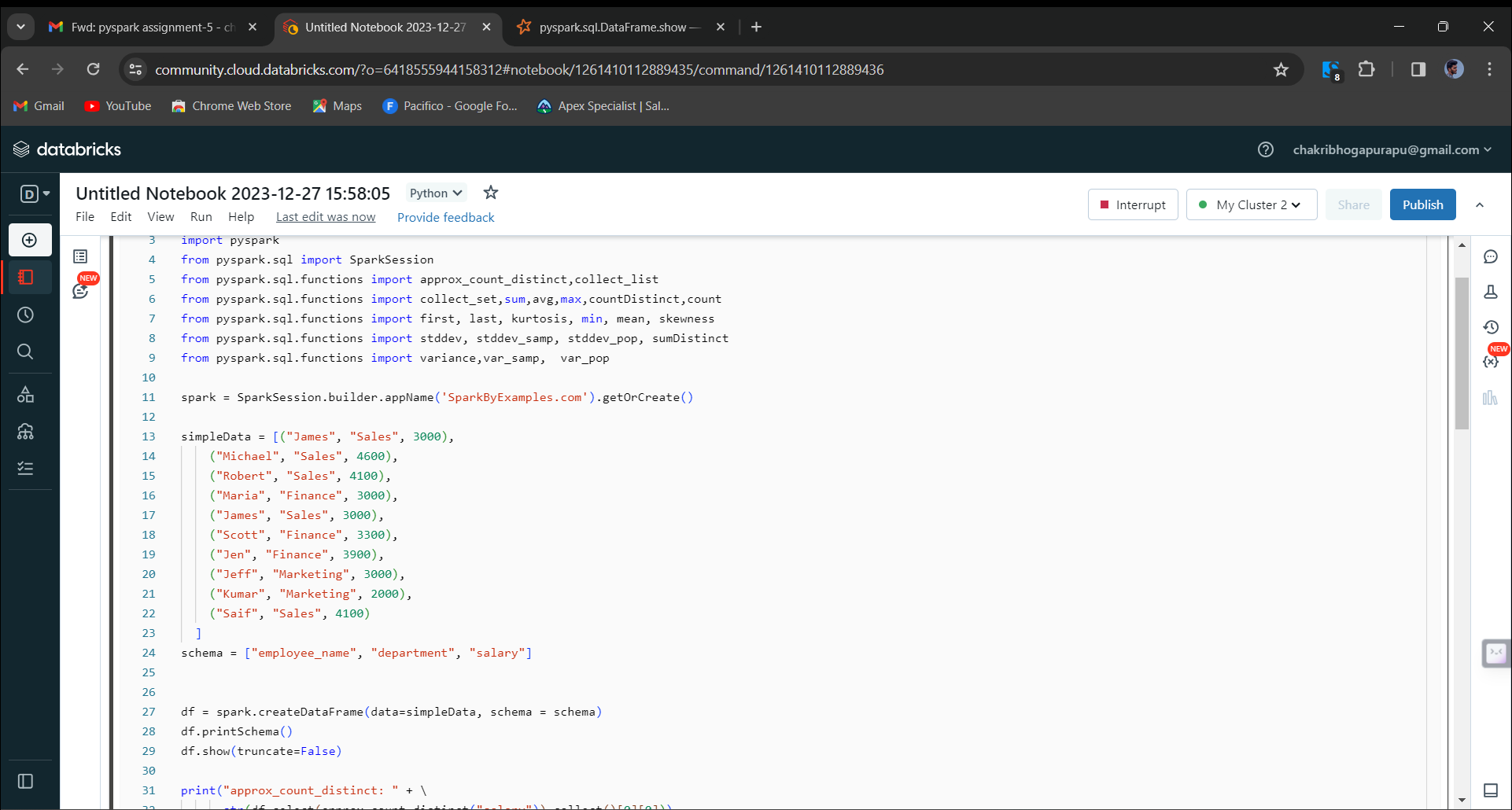
🡪Pyspark broadcast dataframe

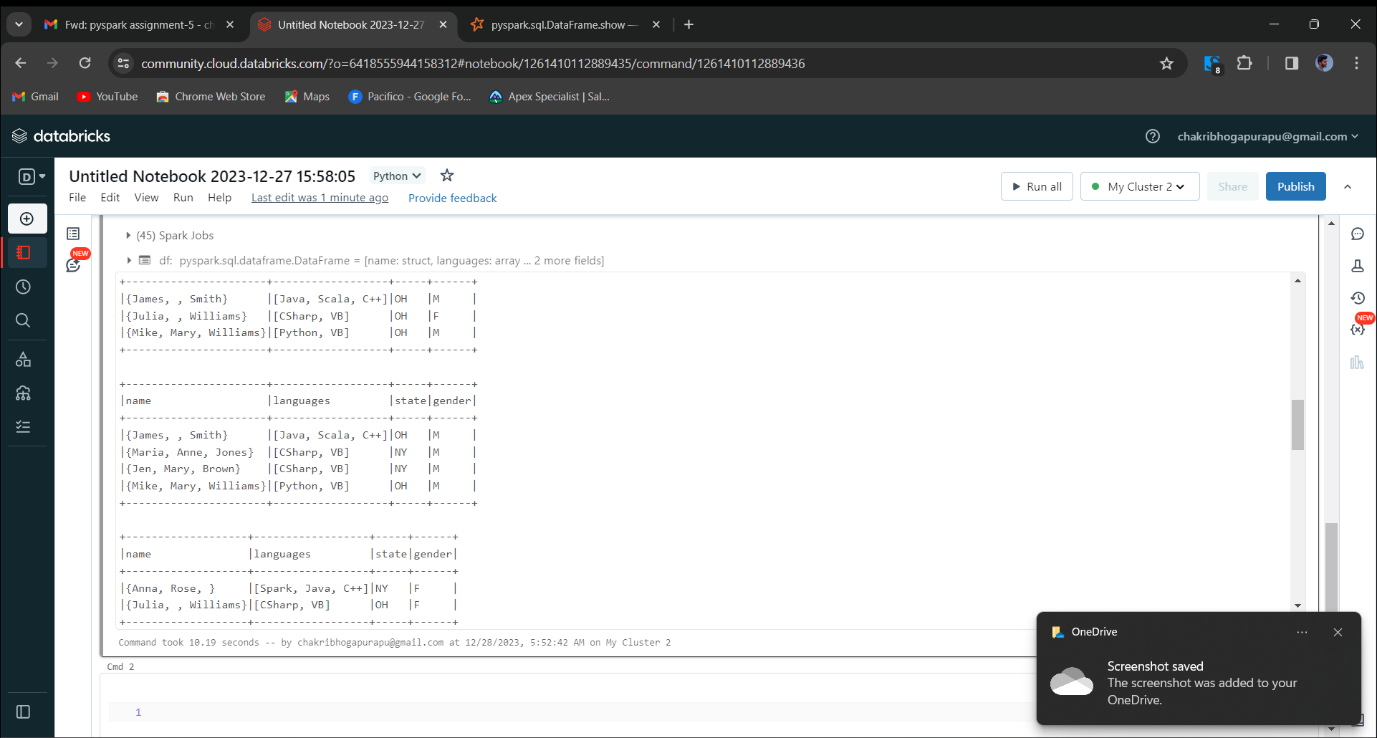
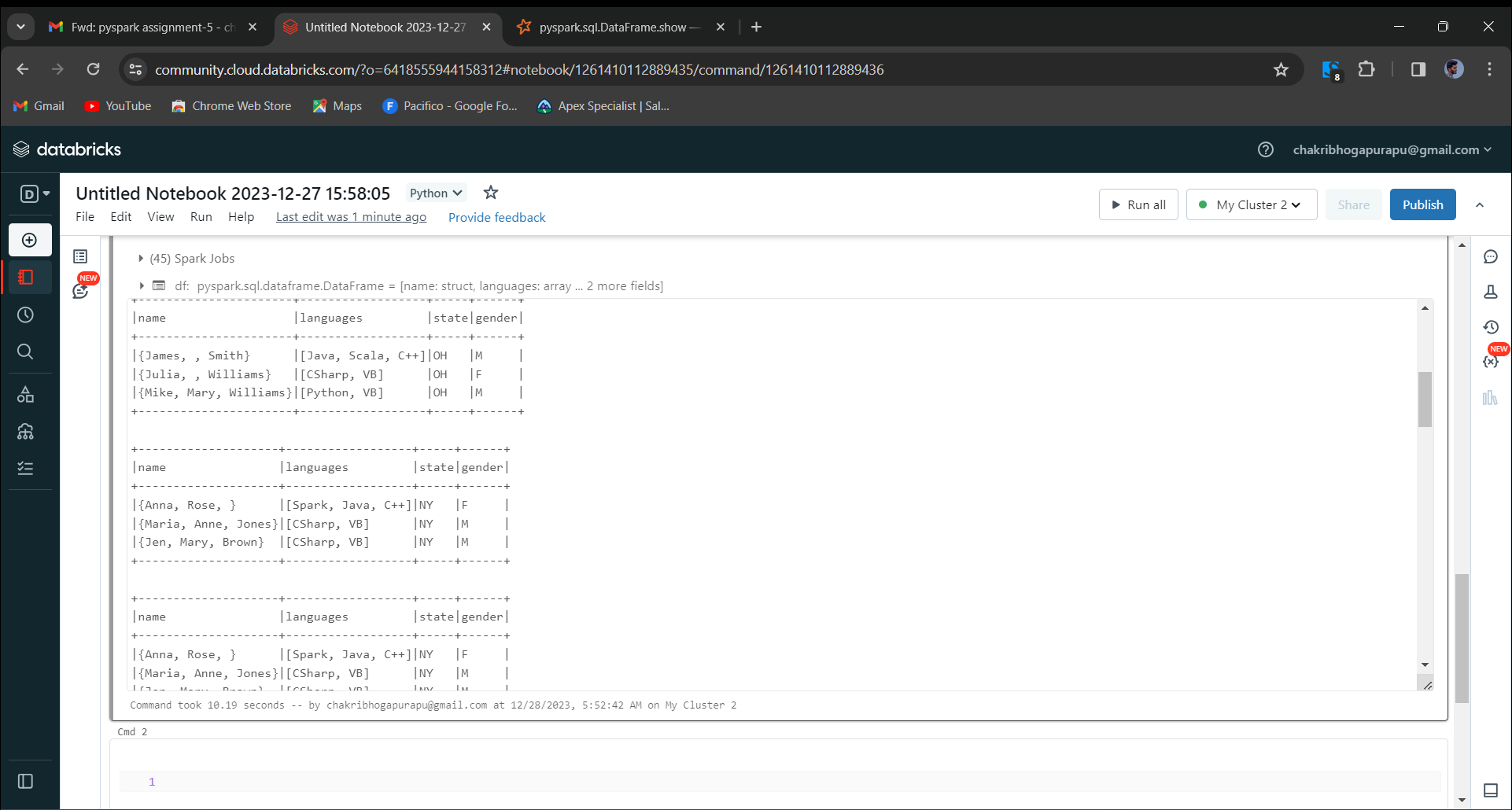
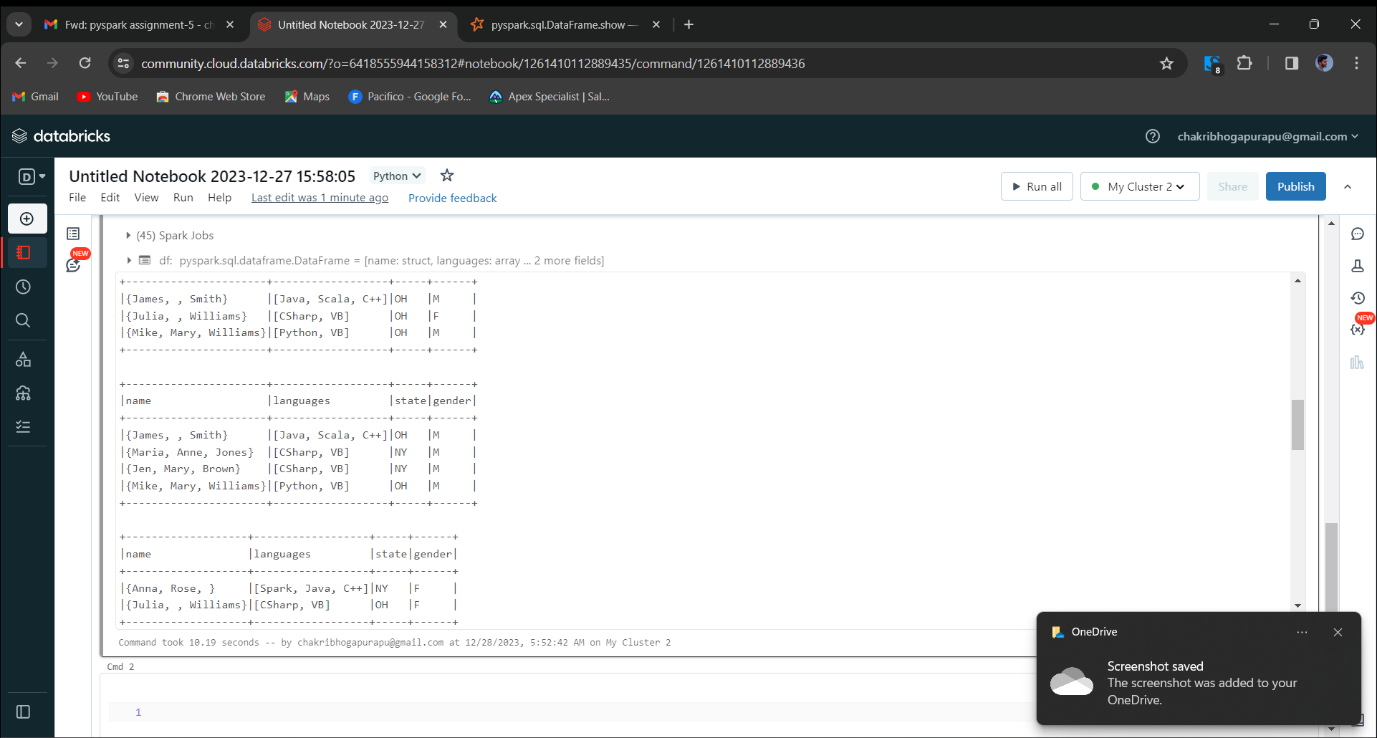


🡪Pyspark create dataframe

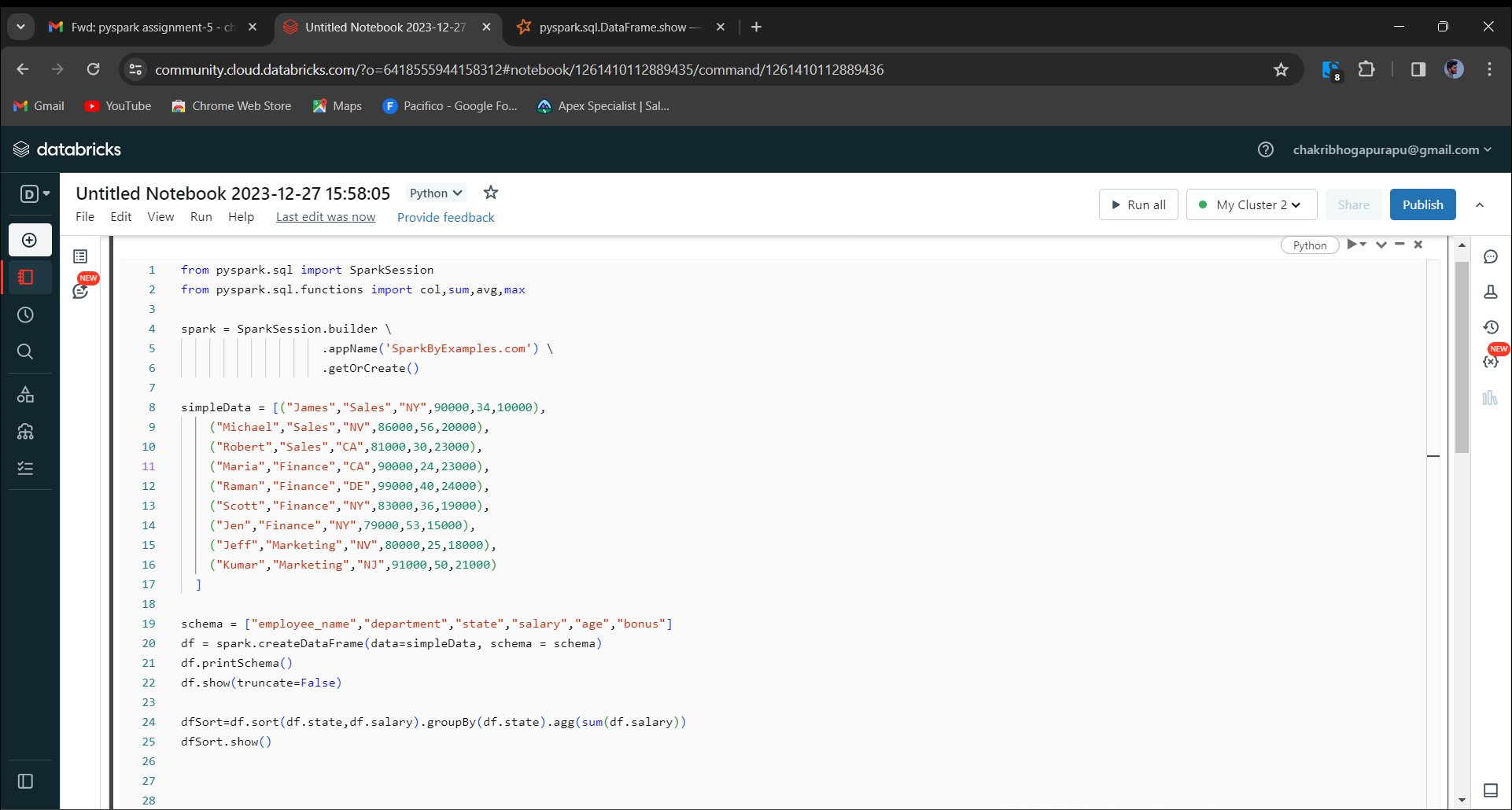


🡪Filter

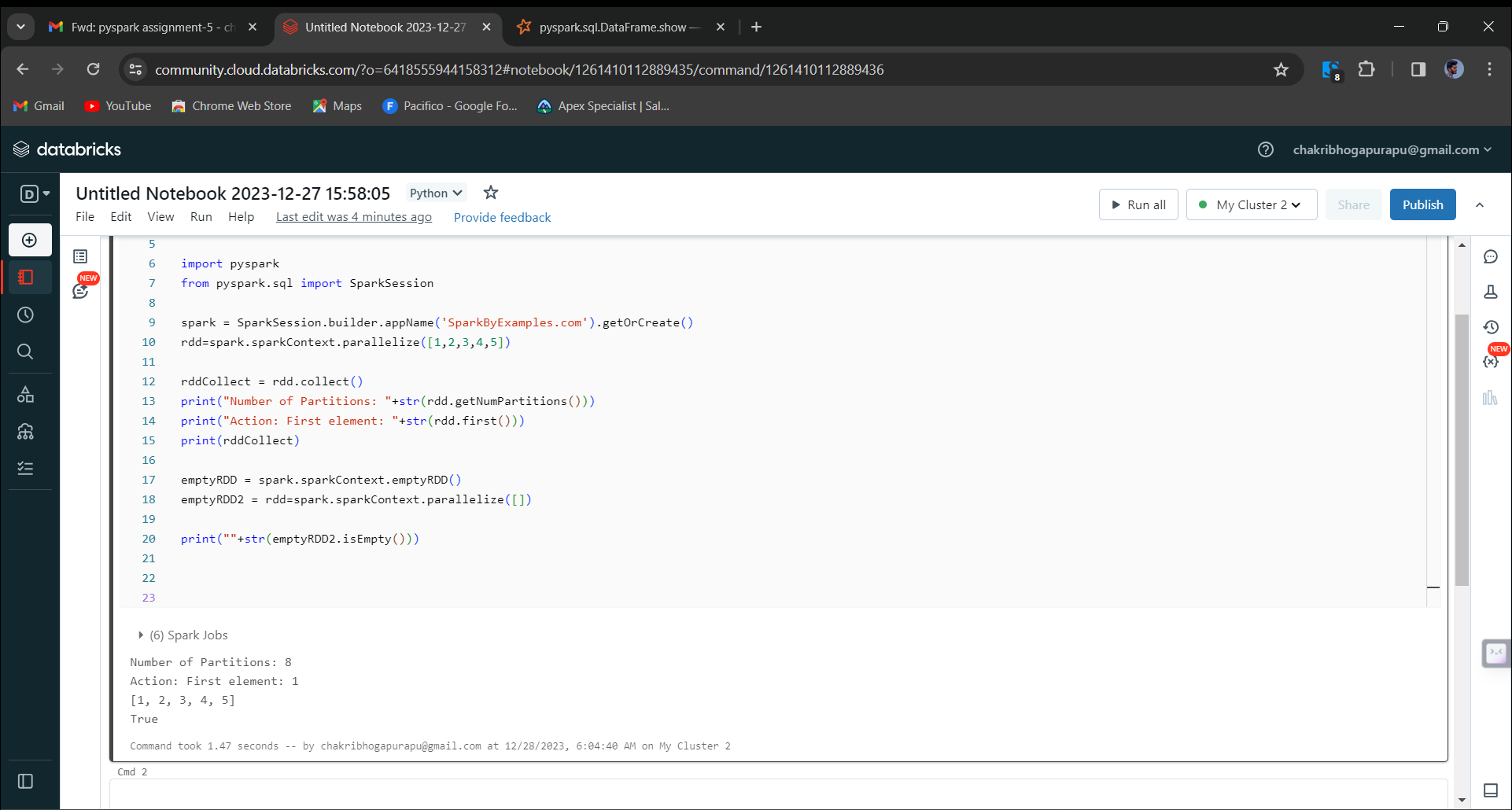




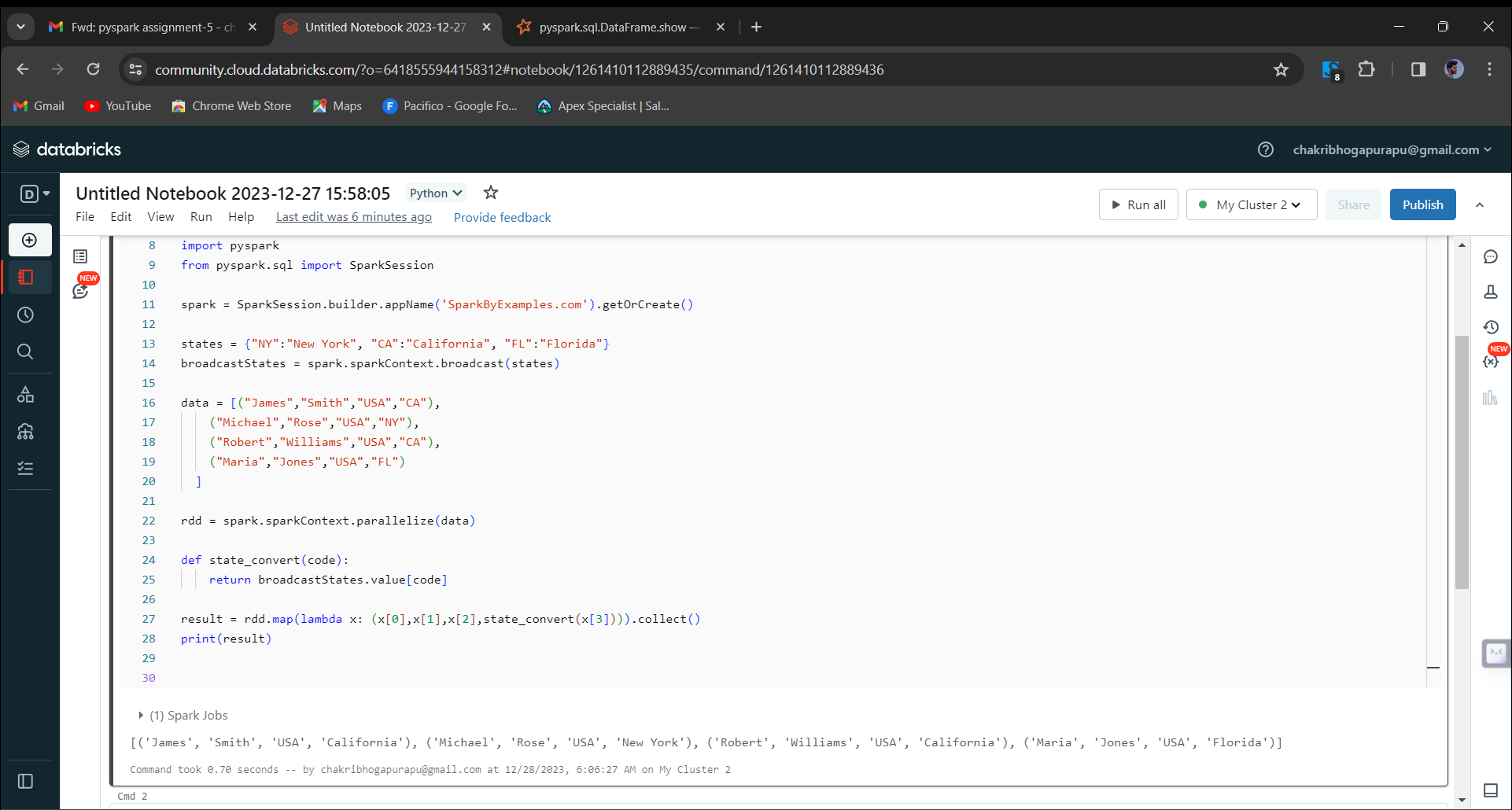
🡪GroupBy



🡪Parallelize



🡪RDD Broadcast



🡪OrderBy