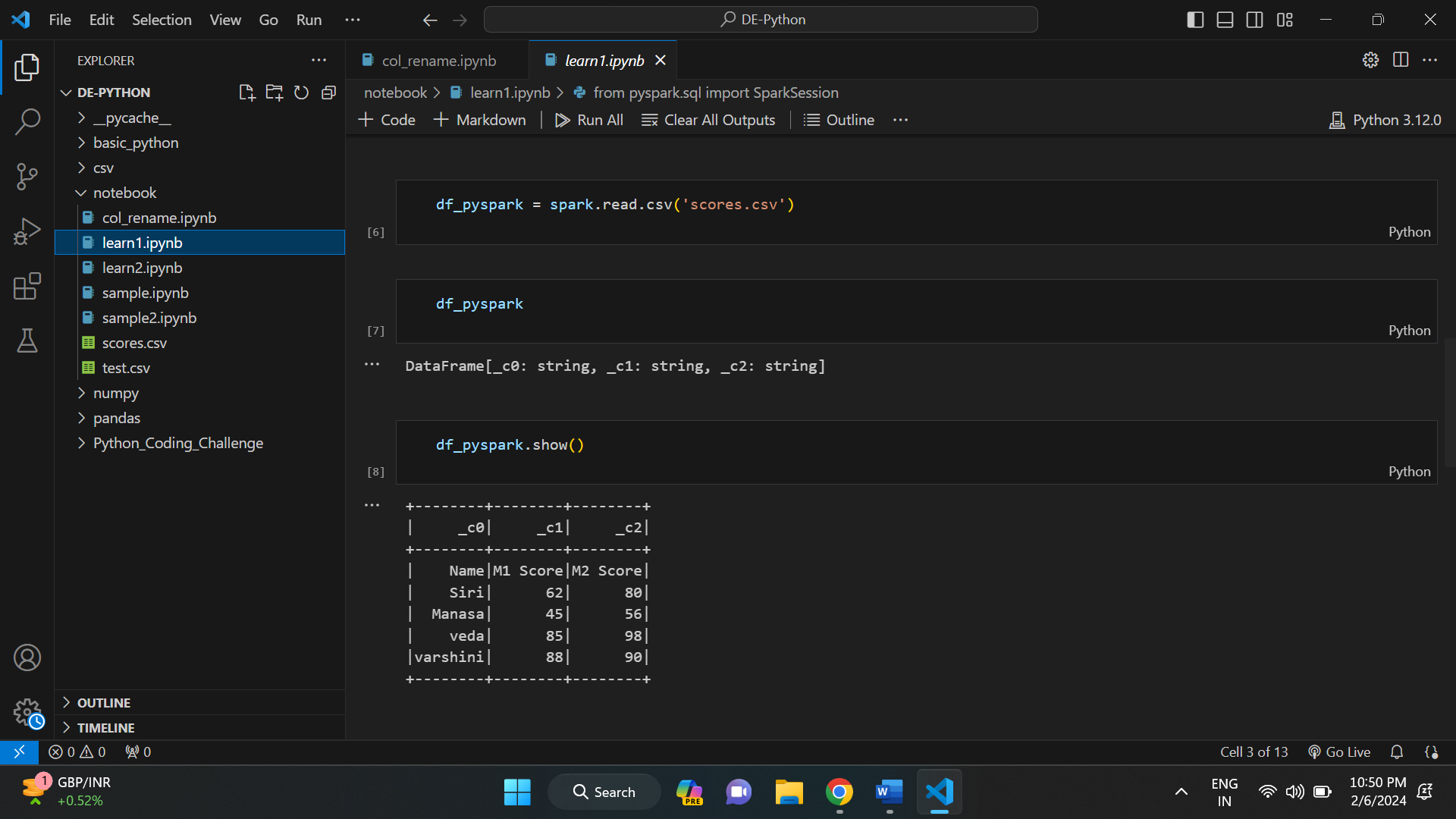
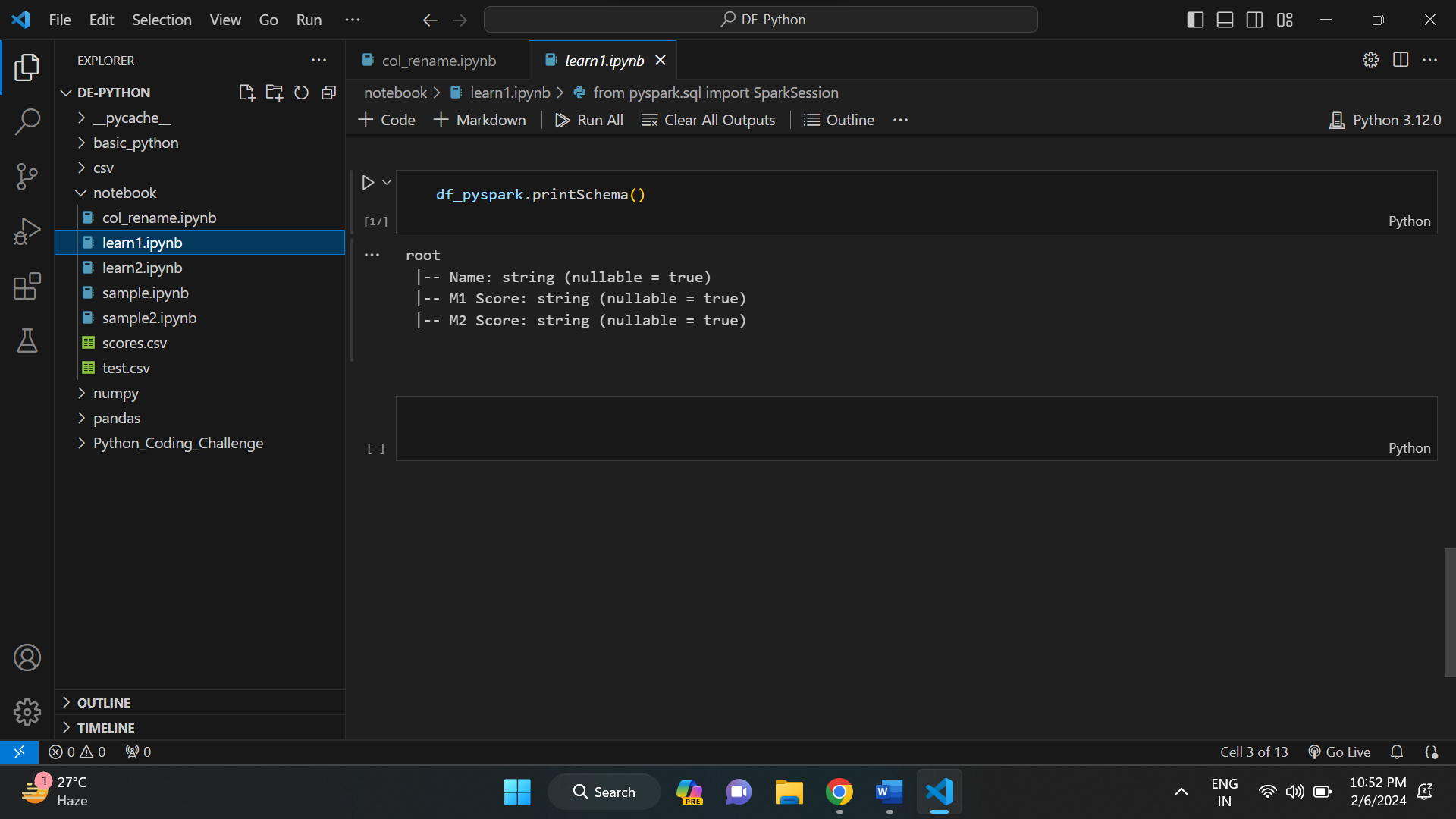
**Data Engineering Assignment – 13**

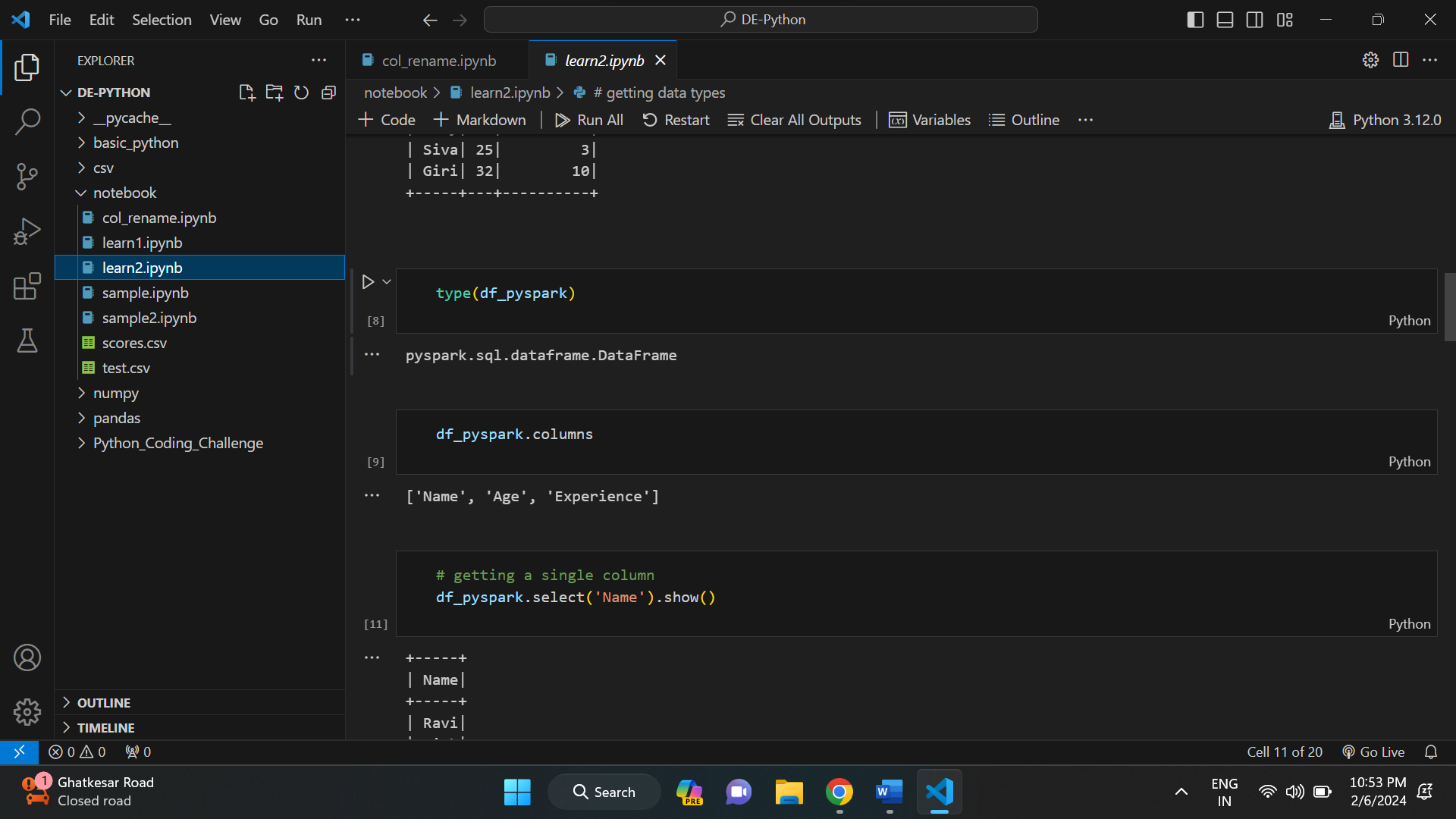
* Reading a CSV file in pyspark



* Getting dataset info using printSchema()

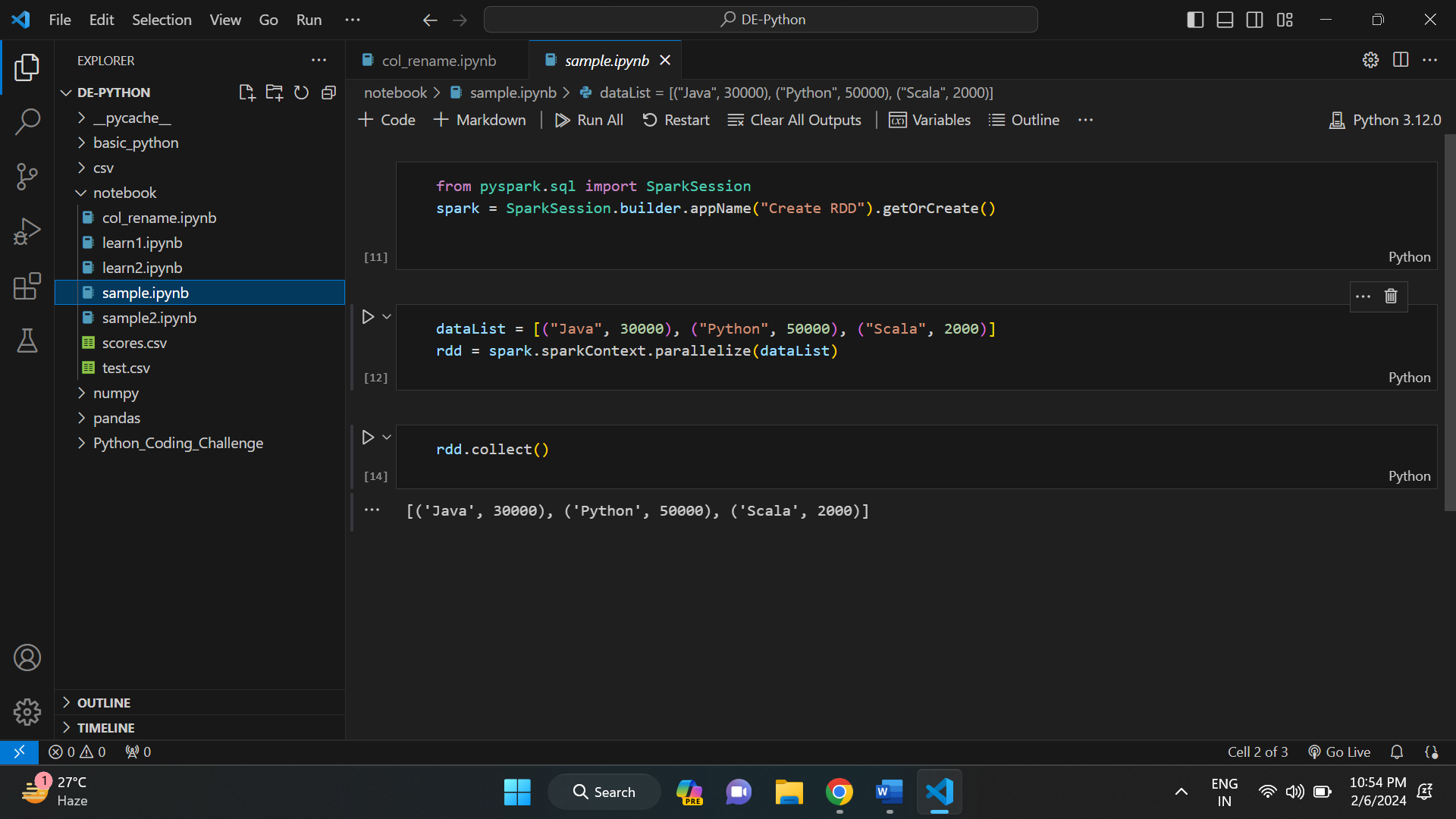


* Get data type

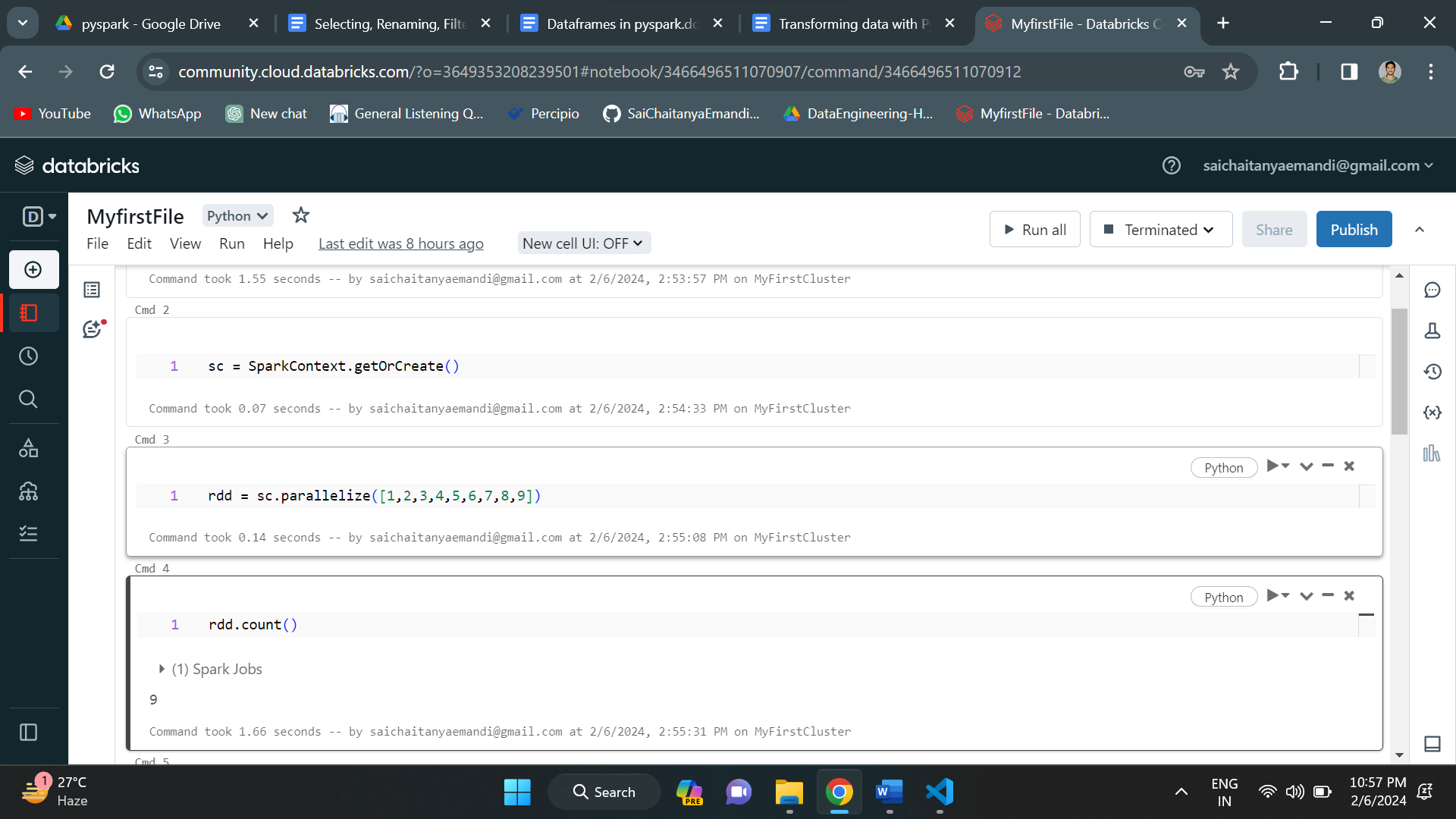


**Actions in pyspark RDD:**

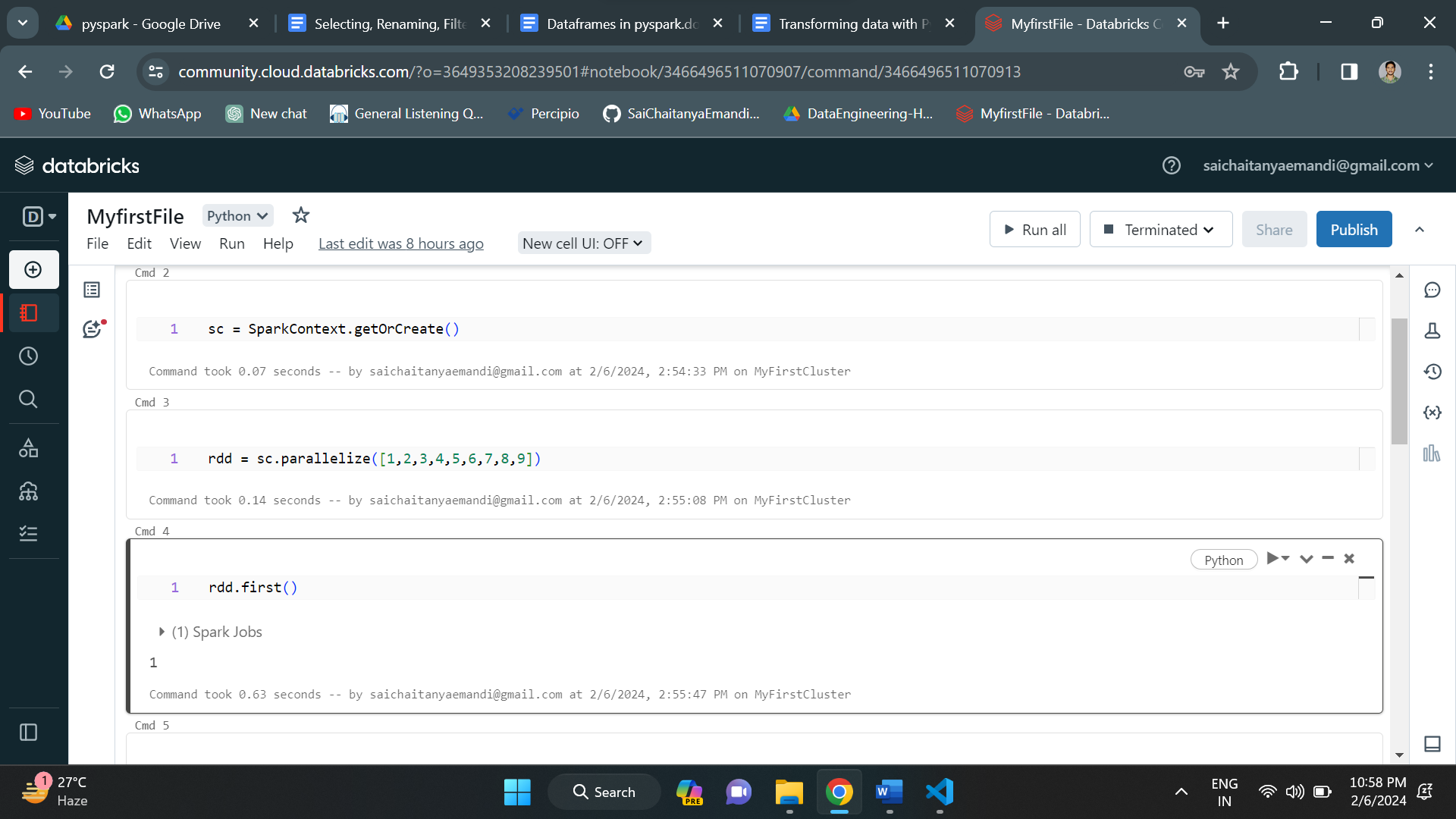
* Collect ()



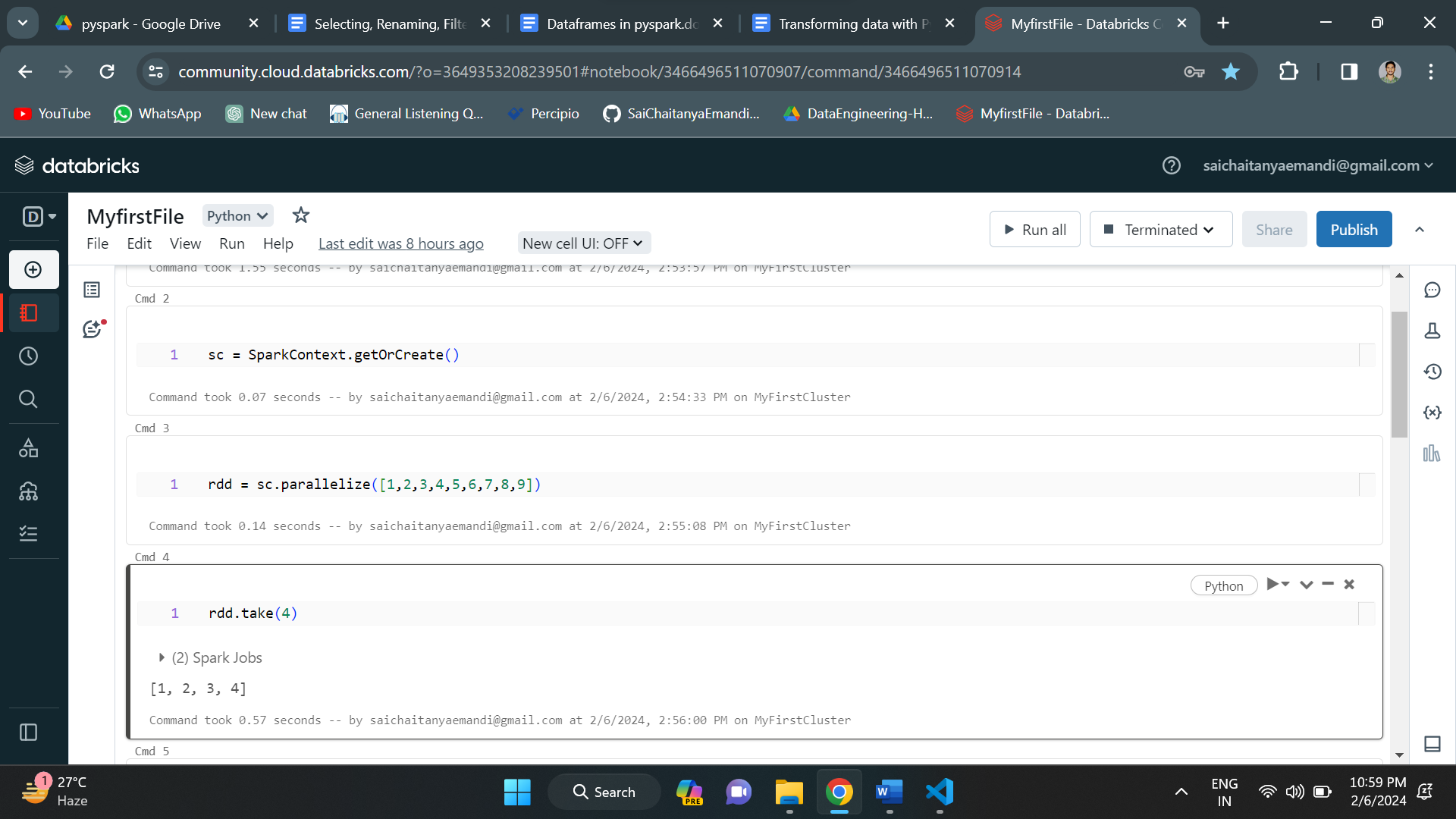
* Count ()



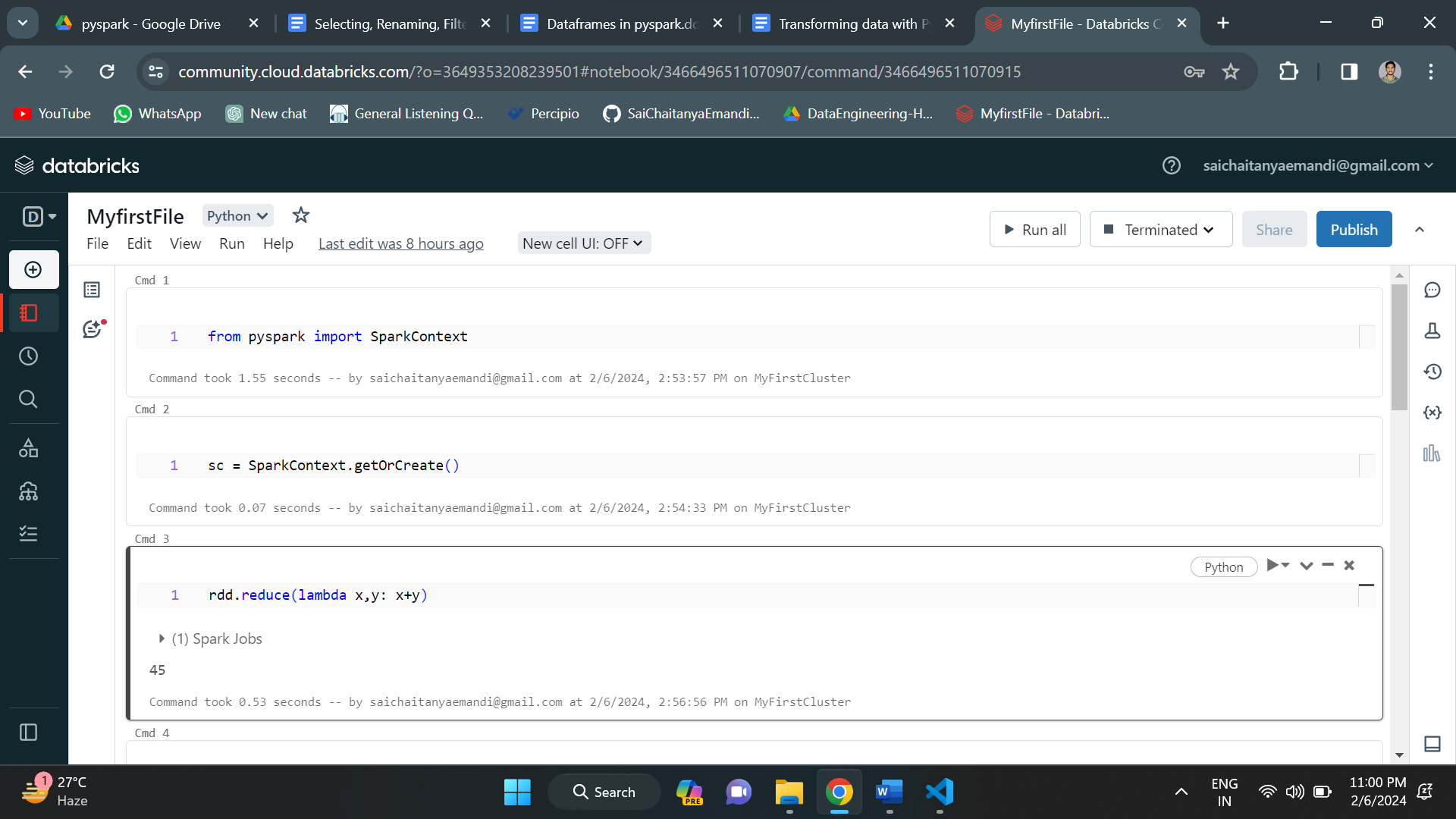
* First ()



* Take ()

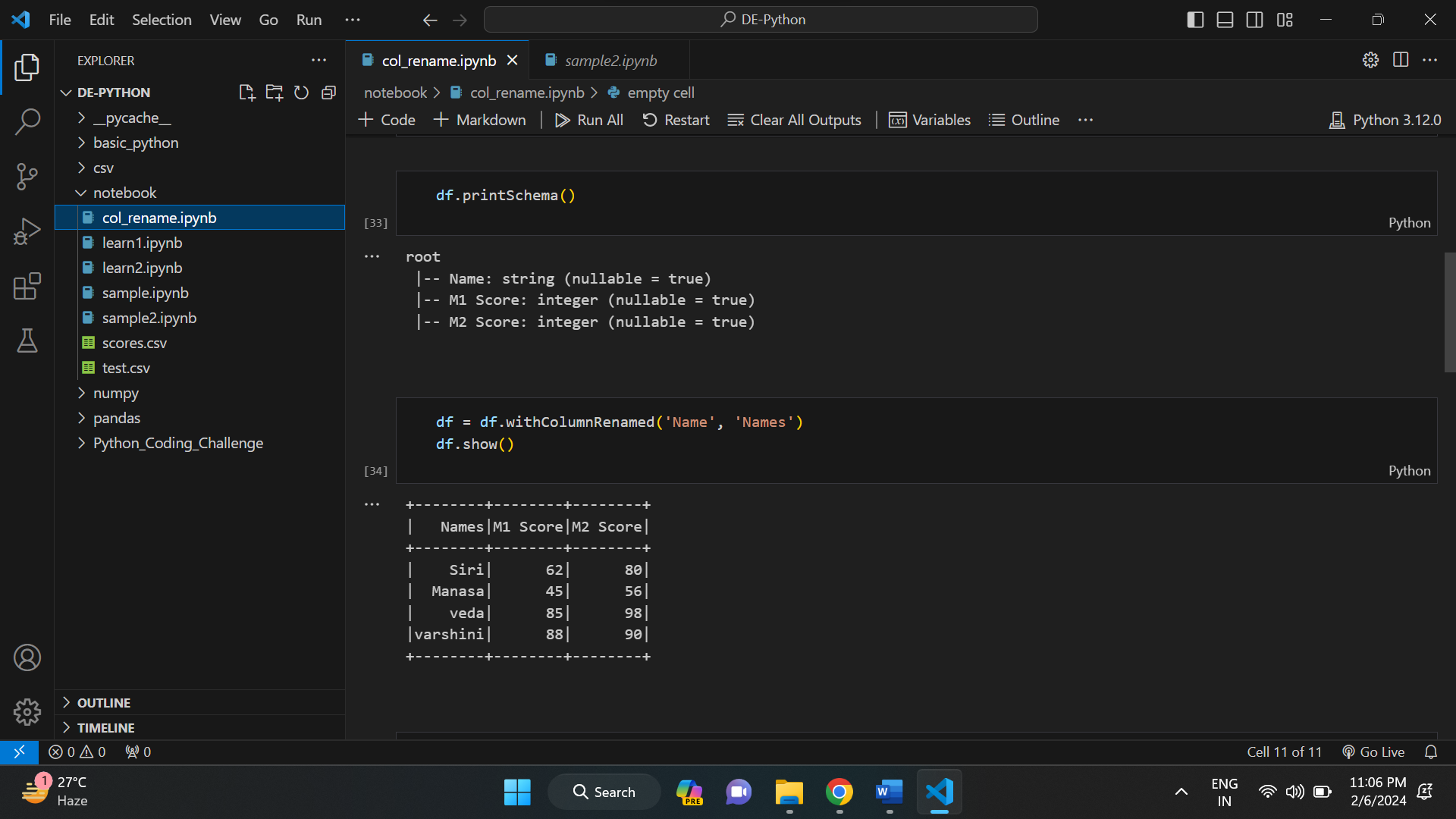


* Reduce ()

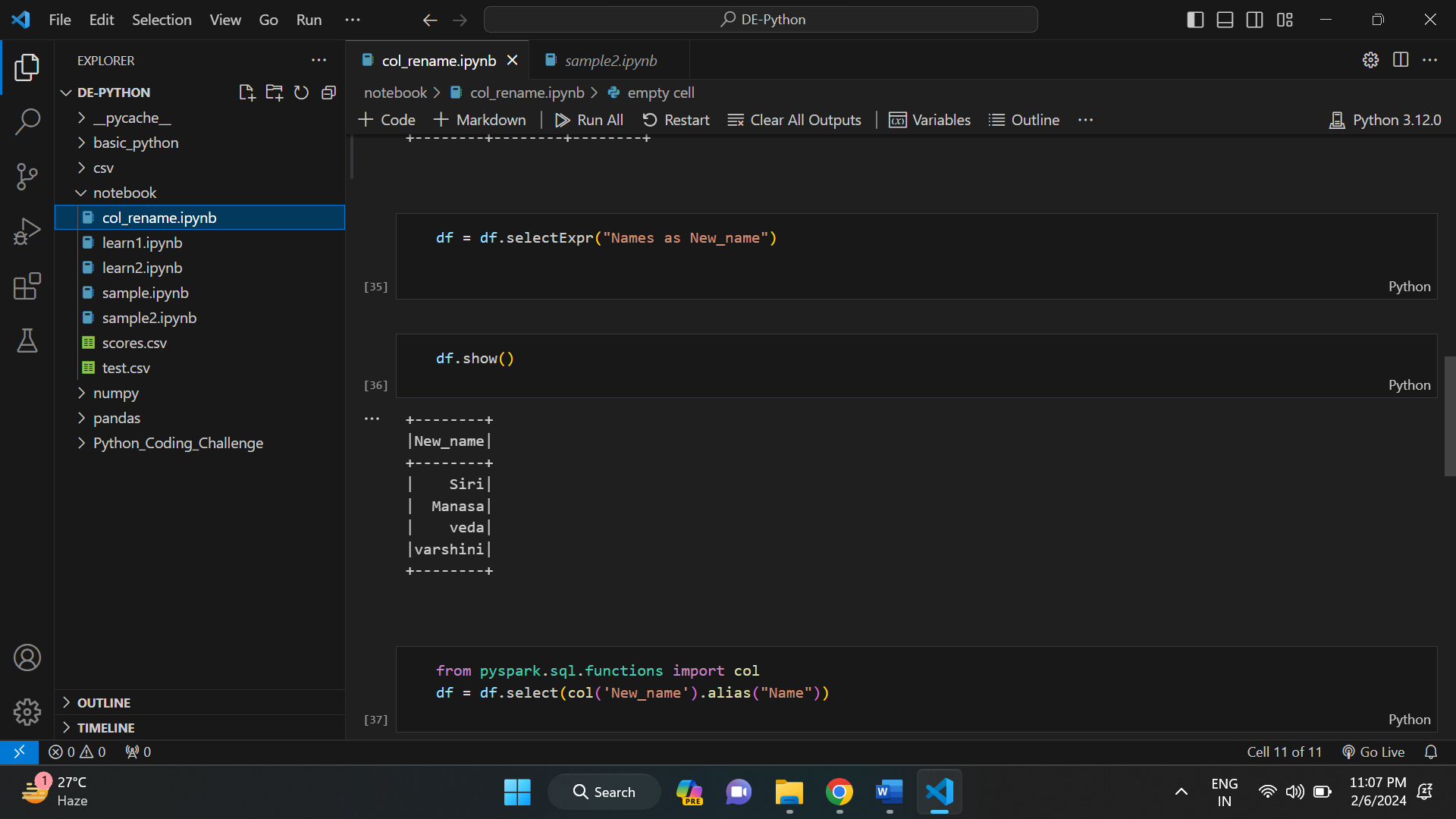


**Renaming columns using different methods:**

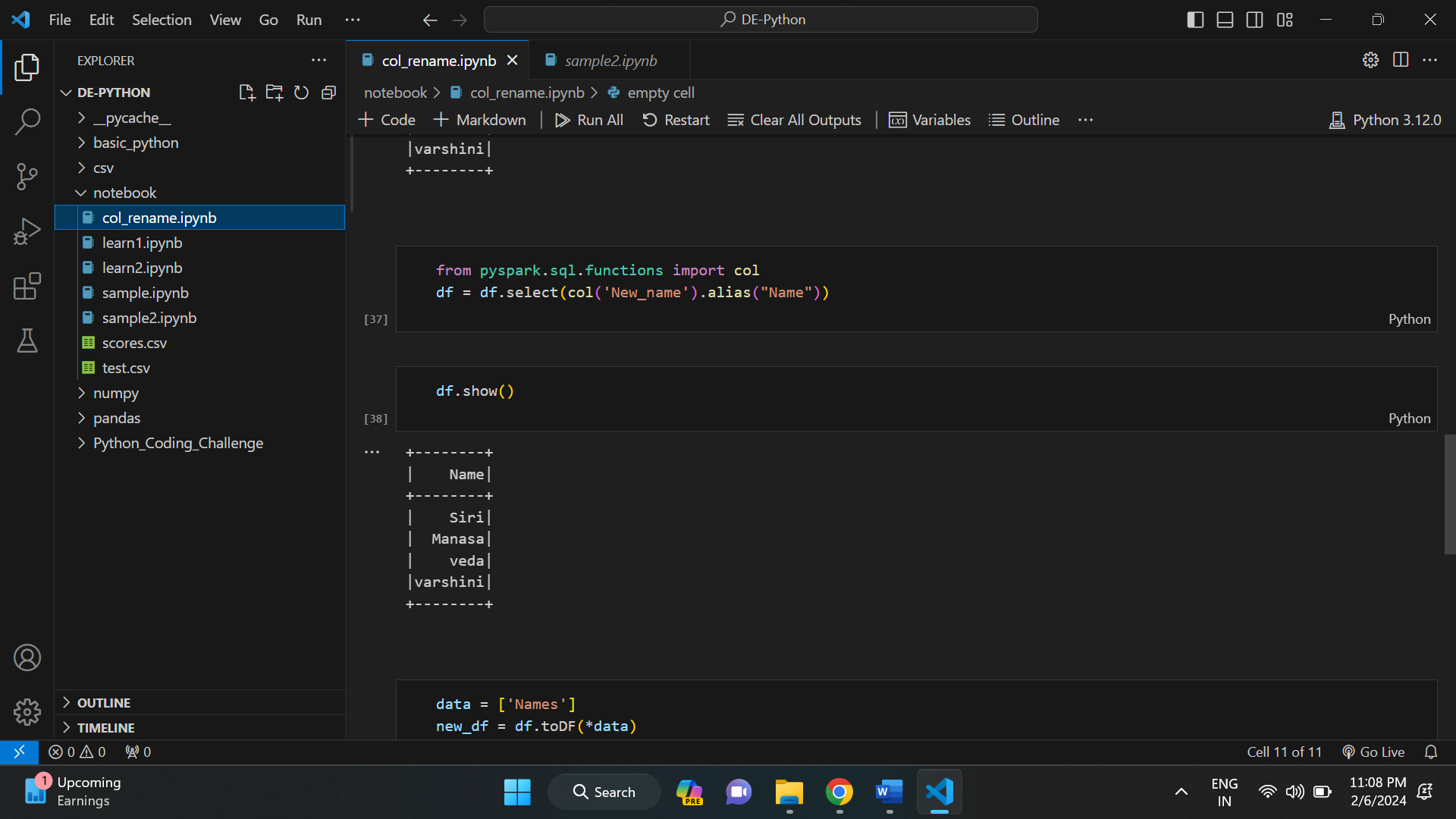
* Using withColumnRenamed () method:



* Using selectExpr ():



* Using select ():



* Using toDF ():

