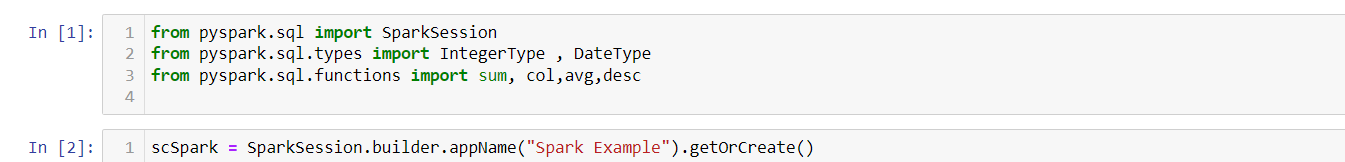
**NAME: AQSA TAUHEED (2303.KHI.DEG.011)**

**PAIRING WITH : Mohammad Hussam (2303.KHI.DEG.020)**

**& MAVIA ALAM KHAN (2303.KHI.DEG.017)**

**ASSIGNMENT NO : 5.1**

Import libraries



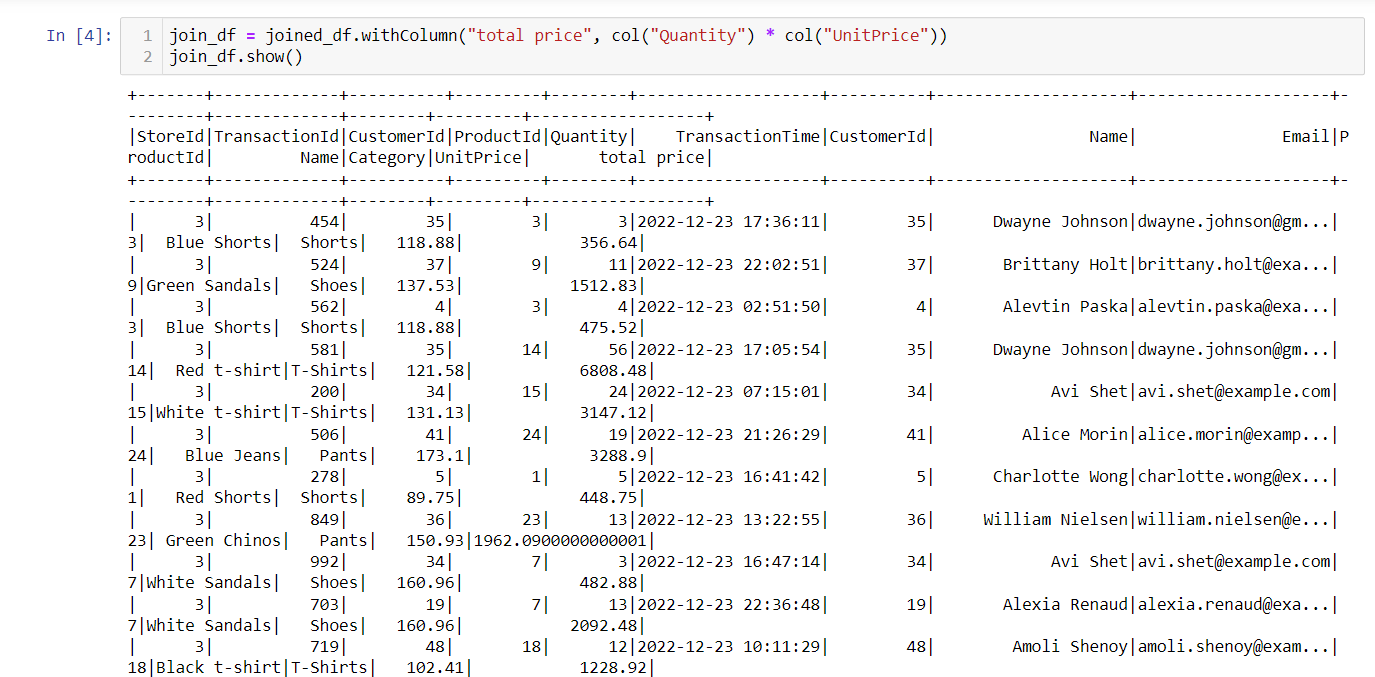
Two join conditions are defined:

join\_condition1 specifies that the "CustomerId" column in the transaction DataFrame should match the "CustomerId" column in the customer DataFrame.

join\_condition2 specifies that the "ProductId" column in the transaction DataFrame should match the "ProductId" column in the products DataFrame.

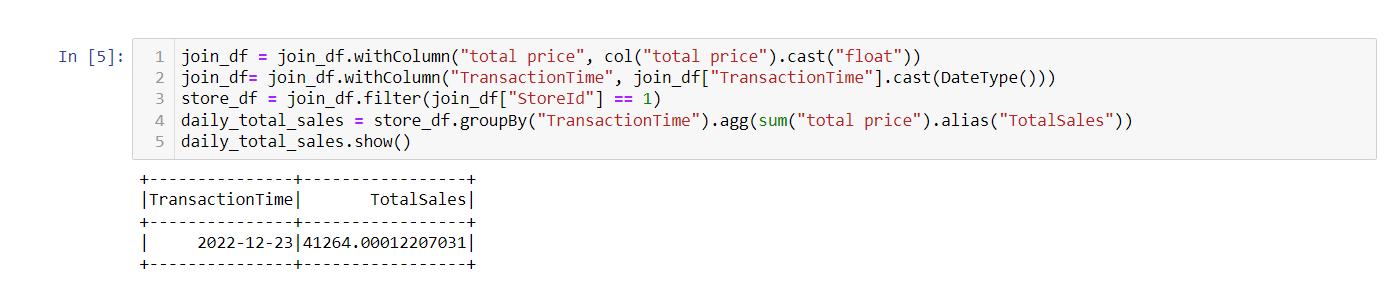


Now we add total price column in which we multiply quantity & unit price column



The daily total sales for the store with id 1

and the withColumn method is used to create a new column in the DataFrame with the casted values, and the original column is replaced by the new one.

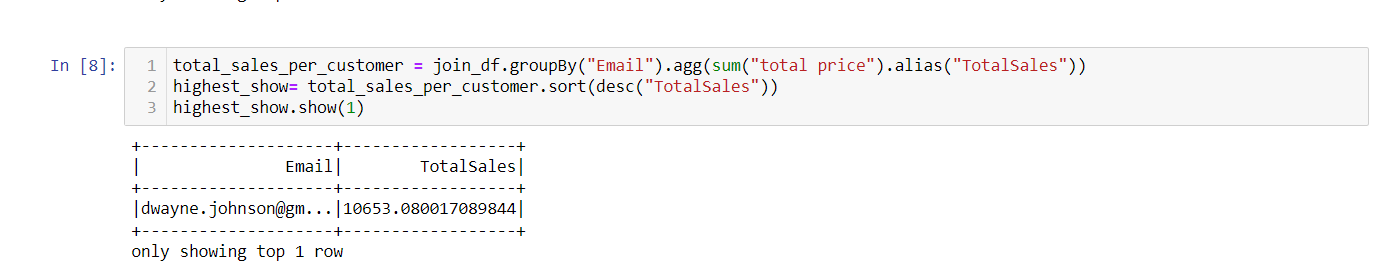


the mean sales for the store with id 2

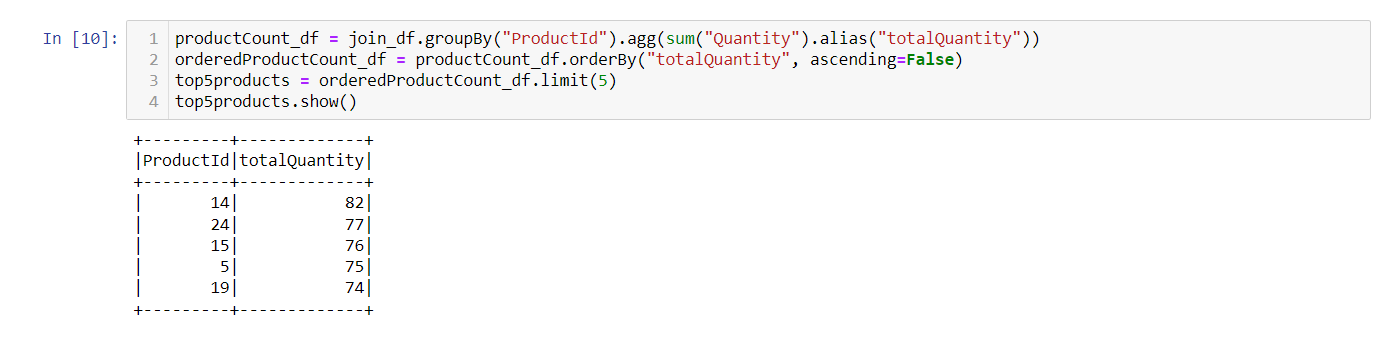


the email of the client that spent the most when summing up purchases from all of the stores

the sort method is used on total\_sales\_per\_customer to sort the DataFrame in descending order based on the "TotalSales" column. The desc function specifies the sorting order as descending.



And 5 products are most frequently bought across all stores



Here we print name of the product that purchase most so we join the table

