

Bibek Shah root@sandbox-hdp:~

```
scala> val orderedRecords = salesRecords.orderBy("Country")
orderedRecords: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [Id: string, SalesAmt: string ... 5 more fields]

scala> orderedRecords.show(100)
+-----+-----+-----+-----+-----+-----+
|   Id|SalesAmt|PaymentType|CustomerName|        City|      Region|  Country|
+-----+-----+-----+-----+-----+-----+
|Product1| 1200|      Amex|     pamela|    Ayacucho|Buenos Aires|Argentina|
|Product1| 1200|       Visa|   Elizabeth|   The Grange|Queensland|Australia|
|Product1| 1200|       Visa|       nina|     Warwick|Queensland|Australia|
|Product1| 1200|       Visa|   Elizabeth| Wollongong|New South Wales|Australia|
|Product1| 1200|       Visa| Ellen Walton|Melbourne|Victoria|Australia|
|Product1| 1200|       Visa|       Lisa|      Perth|Western Australia|Australia|
|Product1| 1200|       Visa|      chris| Gold Coast|Queensland|Australia|
|Product1| 1200| Mastercard|       Tina|Anglesea|Victoria|Australia|
|Product1| 1200|       Visa|       Sara|Perth City|Western Australia|Australia|
|Product1| 1200|       Visa|       IMAN|Brisbane|Queensland|Australia|
|Product2| 3600|       Visa|Kees en Valesca|Scamander|Tasmania|Australia|
|Product1| 1200| Mastercard|       Dean|Rockingham|Western Australia|Australia|
|Product1| 1200| Mastercard|      Nicole|South Melbourne|Victoria|Australia|
|Product1| 1200|      Amex|     Kirsten|Melbourne|Victoria|Australia|
|Product2| 3600| Mastercard|       Piret|Bunbury|Western Australia|Australia|
|Product1| 1200| Mastercard| Elisabeth|Birkdale|Queensland|Australia|
|Product1| 1200|       Visa|       Bob|Sydney|New South Wales|Australia|
|Product1| 1200|       Visa|       Ann|Brisbane|Queensland|Australia|
|Product1| 1200|       Visa|       Amy|Parramatta|New South Wales|Australia|
|Product1| 1200|       Visa| Polyxene|Perth|Western Australia|Australia|
|Product2| 3600|   Diners|      clara|Perth|Western Australia|Australia|
|Product2| 3600|       Visa|       eva|Gisborne|Victoria|Australia|
|Product1| 1200| Mastercard|      KELI|Worongary|Queensland|Australia|
|Product2| 3600| Mastercard|      Anke|Avalon|New South Wales|Australia|
|Product1| 1200| Mastercard|      John|Griffith|New South Wales|Australia|
|Product2| 3600| Mastercard|      TRICIA|Sydney|New South Wales|Australia|
|Product1| 1200|       Visa|      Jacob|Lindfield|New South Wales|Australia|
|Product1| 1200| Mastercard|      Frank|Melbourne|Victoria|Australia|
```

*Group E - Notepad

File Edit Format View Help

Data Technology Solutions

Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

DATAFRAME OPERATION

3. orderBy

Ln 11, Col 2 150% Windows (CRLF) UTF-8



8:54 AM 30-Nov-21

Bibek
Shankhar

```
scala> salesRecords.createOrReplaceTempView("Sales")

scala> val filteredRecords = spark.sql("select * from Sales limit 10")
Hive Session ID = b438177c-982d-4255-a71b-e3cbf7e32e9
filteredRecords: org.apache.spark.sql.DataFrame = [Id: string, SalesAmt: string ... 5 more fields]

scala> filteredRecords.show
+-----+-----+-----+-----+-----+-----+
|    Id|SalesAmt|PaymentType|CustomerName|        City|Region|Country|
+-----+-----+-----+-----+-----+-----+
|Product1| 1200| Mastercard| carolina| Basildon| England|United Kingdom| |
|Product1| 1200| Visa| Betina| Parkville| ...| MO| United States|
|Product1| 1200| Mastercard| Federica e Andrea| Astoria| ...| OR| United States|
|Product1| 1200| Visa| Gouyal| Echuca| Victoria| Australia|
|Product2| 3600| Visa| Gerd W| Cahaba Heights| ...| AL| United States|
|Product1| 1200| Visa| LAURENCE| Mickleton| ...| NJ| United States|
|Product1| 1200| Mastercard| Fleur| Peoria| ...| IL| United States|
|Product1| 1200| Mastercard| adam| Martin| ...| TN| United States|
|Product1| 1200| Mastercard| Renee Elisabeth| Tel Aviv| Tel Aviv| Israel|
|Product1| 1200| Visa| Aidan| Chatou| Ile-de-France| France|
+-----+-----+-----+-----+-----+-----+
```

scala>

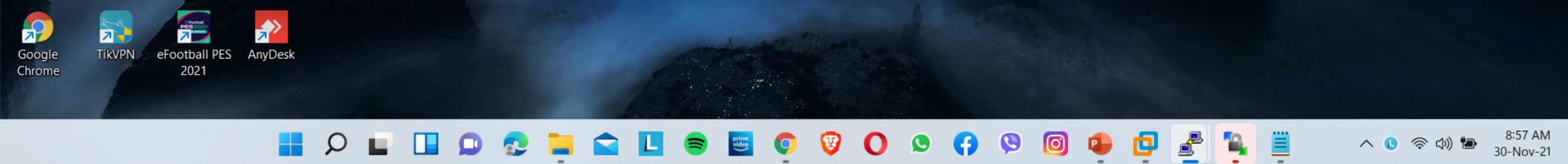
*Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

DATAFRAME OPERATION
4. createOrReplaceTempView

Ln 11, Col 2 150% Windows (CRLF) UTF-8



```
Bibek  
Shan  
scala> val groupedByDf = salesRecords.groupBy("PaymentType")  
groupedByDf: org.apache.spark.sql.RelationalGroupedDataset = RelationalGroupedDataset: [grouping expressions:  
[PaymentType: string], value: [Id: string, SalesAmt: int ... 5 more fields], type: GroupBy]  
scala> val summedDf = groupedByDf.sum("SalesAmt")  
summedDf: org.apache.spark.sql.DataFrame = [PaymentType: string, sum(SalesAmt): bigint]  
  
scala> summedDf.show  
+-----+-----+  
|PaymentType|sum(SalesAmt)|  
+-----+-----+  
| Visa | 849350 |  
| Diners | 133800 |  
| Mastercard | 458450 |  
| Amex | 188900 |  
+-----+-----+  
  
scala> salesRecords.groupby("PaymentType").sum("SalesAmt").show  
<console>:26: error: value groupby is not a member of org.apache.spark.sql.DataFrame  
    salesRecords.groupby("PaymentType").sum("SalesAmt").show  
                           ^  
  
scala> salesRecords.groupBy("PaymentType").sum("SalesAmt").show  
+-----+-----+  
|PaymentType|sum(SalesAmt)|  
+-----+-----+  
| Visa | 849350 |  
| Diners | 133800 |  
| Mastercard | 458450 |  
| Amex | 188900 |  
+-----+-----+  
scala>
```

*Group E - Notepad

Data Technology Solutions
Group E

Bibek Shah Shankhar
Davinderjit Singh
Kirandeep Kaur
Kamaldeep kaur Dhami
Anirudh Siddula

DATAFRAME OPERATION
5. groupBy

Ln 11, Col 2 150% Windows (CRLF) UTF-8