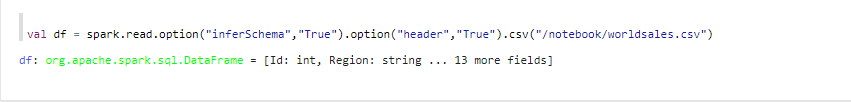
Zeppelin Assignment 1 – Mohammed Ali

1. ) Load data into a Spark dataframe



2.) Create Schema

A black text on a white background

Description automatically generated

3.) Print the Schema

A screenshot of a computer program

Description automatically generated

4.) Aggregate the dataframe via group by “Region” and count

A close up of a computer screen

Description automatically generated

5.) Create a separate dataframe with the above group by results

A screenshot of a computer code

Description automatically generated

6.) Create two views using the “createOrReplaceTempView” command

A screenshot of a computer

Description automatically generated

7.) View on “Salesview” from the first dataframe

**“Refer to 7”**

8.) View on “Regionview” from the second dataframe

**“Refer to 7”**

9.) Using SQL select all from “Regionview” view and show in a line graph.

A screenshot of a computer

Description automatically generated

10.) Using SQL, from the “Salesview” view, Select the region and sum of units sold, and group by region

A screenshot of a computer

Description automatically generated

11.) Select from the “Salesview” view – the region and sum of total\_profit and group by region and display in a Bar chart

A blue squares with different colored squares

Description automatically generated with medium confidence

12.) Using SQL select from the “Salesview” view – show the total profit as profit, the total revenue as revenue and the total cost as cost from “Salesview”, group by region.

A white rectangular object with text

Description automatically generated with medium confidence

13.) The client is in the process of opening up a new store and they are looking at the best location to do so - They need to see the avg profit in each region as a percentage (pie chart) compared to other regions.

A white background with black border

Description automatically generated

Sidenote: I used Ambari and Docker since the server for Ambari was not accessible. I tried different ways to access Zeppelin and was unsuccessful. That was until I started to use Docker and installed the container for Apache Zeppelin. From there it was easy to host the server and complete the required queries.