**Banking - Analysis of Bank Marketing**

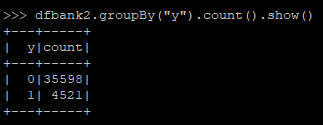
**Requirements:**

**Business Queries:**

**1. Subscribed a Term Deposited or not:**

* To verify whether the client’s have subscribed the term deposit or not ,we have used the groupby() function in the final variable ‘y’(total no of clients subscribed the term deposit or not).

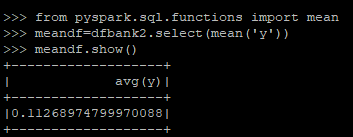
**dfbank2.groupBy(“y”).count().show()**



**from pyspark.sql.functions import mean**

**meandf=df\_spark.select(mean('y'))**

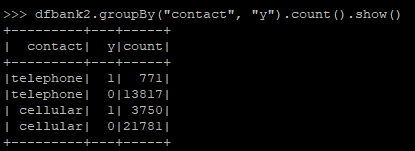
**meandf.show()**



**2. Effective way of contacting the clients:**

* We have considered the factors cellular and telephone marketing approaches ,and analysed the number of clients contacted through cellular and telephone ,and we came to know that ,the number of subscribers are higher while contacted through cellular.

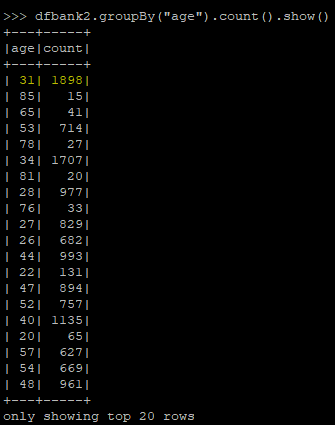
**dfbank2.groupBy(“contact”,”y”).count().show()**



**3. Filtered the ‘age’ and ‘y’ column:**

* We have Filtered the age column and analysed that there are many people are under the age 31,and further analysing that prediction we came to know that the subscription is less while comparing the total number of clients under that age.

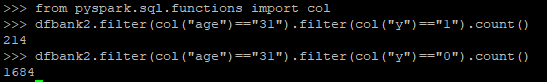
**dfbank2.groupBy(“age”).count().show()**

****

**from pyspark.sql.functions import col**

**dfbank2.filter(col("age")=="31").filter(col("y")=="1").count()**

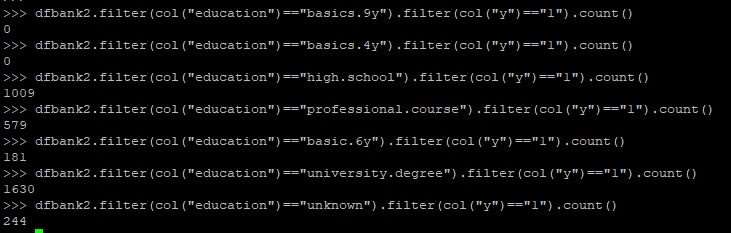
**dfbank2.filter(col("age")=="31").filter(col("y")=="0").count()**

****

**4. Based on the clients education status:**

We have analysed the client’s with respect to their loan status,and by verifying it,we came to a conclusion that ,the client’s who have not lended any personal loan,have subscribed the term deposit.

**dfbank2.filter(col("education")=="university.degree").filter(col("y")=="1").count()**



**5. Based on the count of the client's term deposit with respect to their loan status.**

Using education as a parameter and analysing the dataset,the clients with a degree from university have a higher rate of subscriptions compared to the other education categories.

**dfbank2.filter(col("loan")=="1").filter(col("y")=="1").count()**

**dfbank2.filter(col("loan")=="0").filter(col("y")=="0").count()**

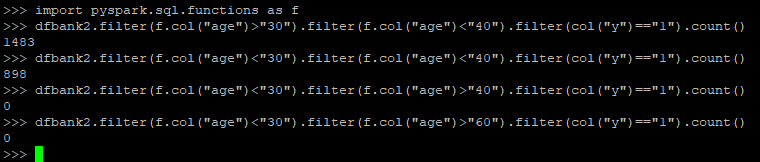


**6. Which age group has the highest and lowest number of clients subscriptions.**

While analysing the age group,the clients between the age 30 to 40 are higher in count,the subscriptions are higher in this range,and there are no clients above the age 60.

**import pyspark.sql.functions as f**

**dfbank2.filter(f.col("age")>"30").filter(f.col("age")<"40").filter(col("y")=="1").count()**

****

**TEAM B**

Vaishali S

Sathya Narayanan S

Manju R

Monisha K

Pooja Bedi V