**Business Requirements Specification (BRS)**

1. **Background of the project**

* The bank direct marketing data analysis is important work that helps the banks predict whether customers will sign long term deposits with the banks. The method that can predict such customers' needs can be profitable to the banks for improving their marketing campaign strategies.
* Term deposits are a major source of income for a bank. A term deposit is a cash investment held at a financial institution. Your money is invested for an agreed rate of interest over a fixed amount of time, or term.
* In our project, we analyzed data from the UCI Machine Learning Repository called Banking - Analysis of Bank Marketing dataset.The Data is related to direct marketing campaigns. The bank has various outreach plans to sell term deposits to their customers such as email marketing, advertisements, telephonic marketing, and digital marketing.Often, more than one contact with the same client was required,in order to access if the product (bank term deposit) would be (or not) subscribed.

**b) Approach:**

The Capabilities Needed for Banking - Analysis of Bank Marketing

To gain meaningful insight from your data, you need the following key capabilities:

* Centralized location to view data (PYSPARK)
* Real-time data updates
* Data visualization tools (GRAFANA)
* Anytime, anywhere access to data
* Time-based data analysis

Without these capabilities, you won’t be able to take action from your data.

**b) Goals of project**

1. To Increase subscription of term deposit.
2. New customer acquisition.
3. Identifying the most effective way of marketing.

**c) Problem Statement**

* Analyze the business perspectives and provide the hidden inference from the data.Also deploy the machine learning model for predicting the subscription of the client.

**d) Solution / Proposed System:**

* To verify whether the client’s have subscribed the term deposit or not.
* We analyzed the number of clients contacted through cellular and telephone.
* Analyzed the data of more clients in the age group and found the count of clients have subscribed to the term deposit.
* We have analyzed the count of the client's term deposit with respect to their loan status.
* We identified which age group has the highest and lowest number of clients subscriptions.
* Based on the clients education status we found the subscription count.

**e) Users of project**

* Customers are bank depositors, who entrust their money, banking facilities, for storage and, for which they receive interest and applicants or those who need temporary, some additional money for paying interest or commission.

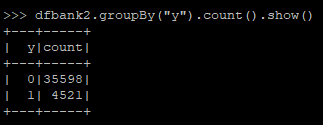
**f) Requirements**

**Business Queries:**

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

We have analyzed whether the client’s have subscribed the term deposit or not and the mean value.

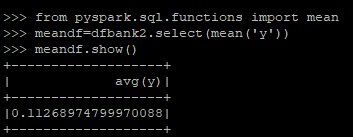
**dfbank2.groupBy(“y”).count().show()**



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

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

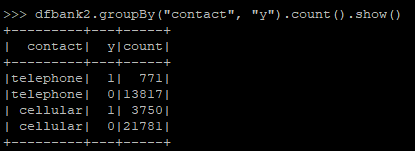
**meandf.show()**



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

We have selected the ‘contact’ column and analyzed the number of clients contacted through cellular and telephone.

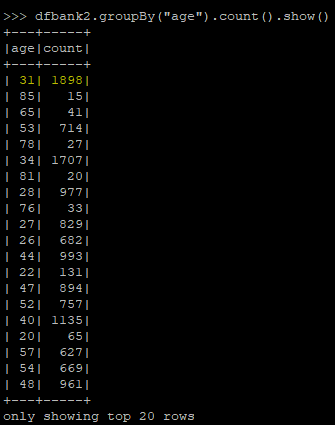
**dfbank2.groupBy(“contact”,”y”).count().show()**



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

We Analyzed the data of more clients in the age group and found the count of clients that have subscribed to the term deposit.

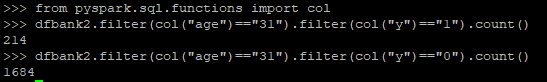
**dfbank2.groupBy(“age”).count().show()**

****

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

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

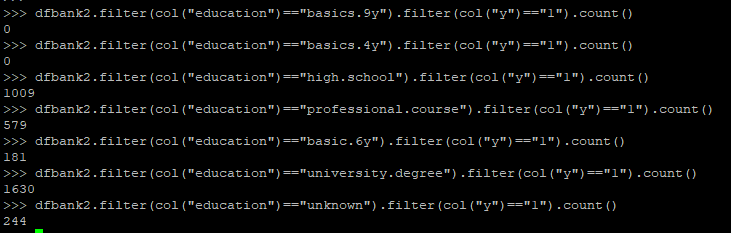
**dfbank2.filter(col("age")=="31").filter(col("y")=="0").count()**

****

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

We have analyzed based on the clients education status we found the subscription count.

**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.**

We have analyzed the count of the client's term deposit with respect to their loan status.

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

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

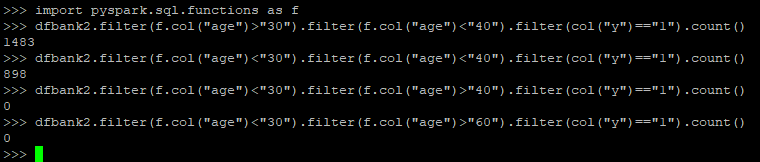


**6.Which age group has the highest and lowest number of clients subscriptions.**

We identified which age group has the highest and lowest number of clients subscriptions.

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

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

****

**Solution:**

1. **WHAT IS THE TOTAL COUNT OF CLIENT’S SUBSCRIBED THE TERM DEPOSIT AND THEIR AVERAGE ??**

Analyzed the final subscriber’s count,out of 41,188 client’s **4,521** clients have subscribed the term deposit.and the remaining **35598** clients are non-subscribers, and the average of the total number of subscribers and non-subscribers are **0.11268974799970088.**

**2. WHICH MARKETING APPROACH IS EFFECTIVE?**

Comparing cellular and telephone marketing approach , more number of clients subscribed the term deposit when contacted through cellular i.e out of **25,531** clients who are contacted through cellular **3,750** clients have subscribed the term deposit whereas in other out of **14,558** clients contacted through clients only **771** clients have subscribed.

**3. WHICH AGE GROUP HAS HIGHER NUMBER OF SUBSCRIPTIONS??**

Analyzing the entire data of the clients , we have came to know that there are more clients in the age group of **31** , i.e there are **1898** count and analyzed that under this age there are less number of count who have subscribed compared to not subscribed count.

**4. WHEN CONSIDERING THE CLIENT'S EDUCATION STATUS ,WHICH**

**CATEGORY HAS HIGHER NO SUBSCRIPTION?**

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

**5.. WHAT IS RATIO OF CLIENT’S WHO HAVE SUBSCRIBED THE TERM DEPOSIT WHO HAVE LENDED A LOAN PREVIOUS AND CLIENTS WHO HAVE NOT LENDED ANY LOAN?**

In accordance with the analysis , clients who have not taken any loan previously are higher in subscription of the Term Deposit.

**6. WHICH AGE GROUP HAS HIGHER NUMBER OF SUBSCRIPTIONS AND WHICH HAS THE LOWEST??**

Analyzing the clients details,the clients between the age **30 - 40** have subscribed the term deposit more, and clients at the age of **60** and above are null i.e no subscription have been made in this age group.

**Exploratory Data Analysis:**

1. Analysis of the dataset.
2. The number of missing/null values.
3. The number of unique values in each variable.

**Non-Functional Requirements**

**Performance Requirements**

* Increase Customer Satisfaction Performance Requirements Customers must be able to use banking services 24 hours a day, 365 days a year, with the shortest possible downtime for backup and maintenance.
* Increase Product Offerings: The new services enable banks to collect a larger share of their clients' assets.
* Banks will be able to add new services and products to their homepage using the internet banking system.
* Decrease Overall Costs: It will help a bank save money in two ways: it will reduce the cost of processing transactions and the number of branches needed to serve an equivalent number of customers.

**Conclusion:**

Based on the term deposit count the bank can focus on the following :

1.Clients in the age group of 30 to 40.

2.Clients who are University graduates.

3.Comparing cellular and telephone marketing approaches,more clients subscribed to the term deposit when contacted through cellular.

4. Based on the loan status more clients who have not lended any loan have subscribed to term deposit.

**TEAM B**

Vaishali S

Sathya Narayanan S

Manju R

Monisha K

Pooja Bedi V