ANALYSIS OF MARKETING CAMPAIGN DATA FOR TERM DEPOSIT SUBSCRIPTIONS

Tq9041@gmail.com

PROJECT INTRODUCTION

I recently completed a data analysis project using SQL to explore the Bank Marketing Dataset. The dataset was obtained from a Portuguese banking institution and included information about their direct marketing campaigns, the clients, and the outcomes of the marketing efforts. With over 41,000 records and 20 attributes, the dataset provided a rich source of information to explore. I aimed to answer three questions using the analysis:

1. How effective were different contact methods for converting leads to subscribers?
2. How did customer demographics relate to campaign success?
3. What was the relationship between contact type and success?

I used SQL to clean and preprocess the data and then performed various queries and visualizations to analyze the results. The findings from the analysis could help banking institutions make more informed decisions about their marketing campaigns and strategies. Overall, the project was a valuable learning experience for me, and it helped me gain a deeper understanding of the potential insights that can be gained from data-driven decision-making.

QUESTION 1- ANALYZING CONTACT METHODS FOR LEAD CONVERSION TO SUBSCRIBERS

This SQL code is used to retrieve data from the "Bankdata_full" table, specifically focusing on the contact column. The query selects the number of contacts made through cellular and telephone communication, along with the number of successful contacts, which are those that led to a subscription or the person agreeing to take part in the product/service.

To accomplish this, the code utilizes the COUNT function to calculate the total number of contacts and the SUM function to calculate the number of successful contacts by checking if the value in the "y" column is equal to "yes". The IIF function is used to return a 1 if the condition is true and 0 if it is false.

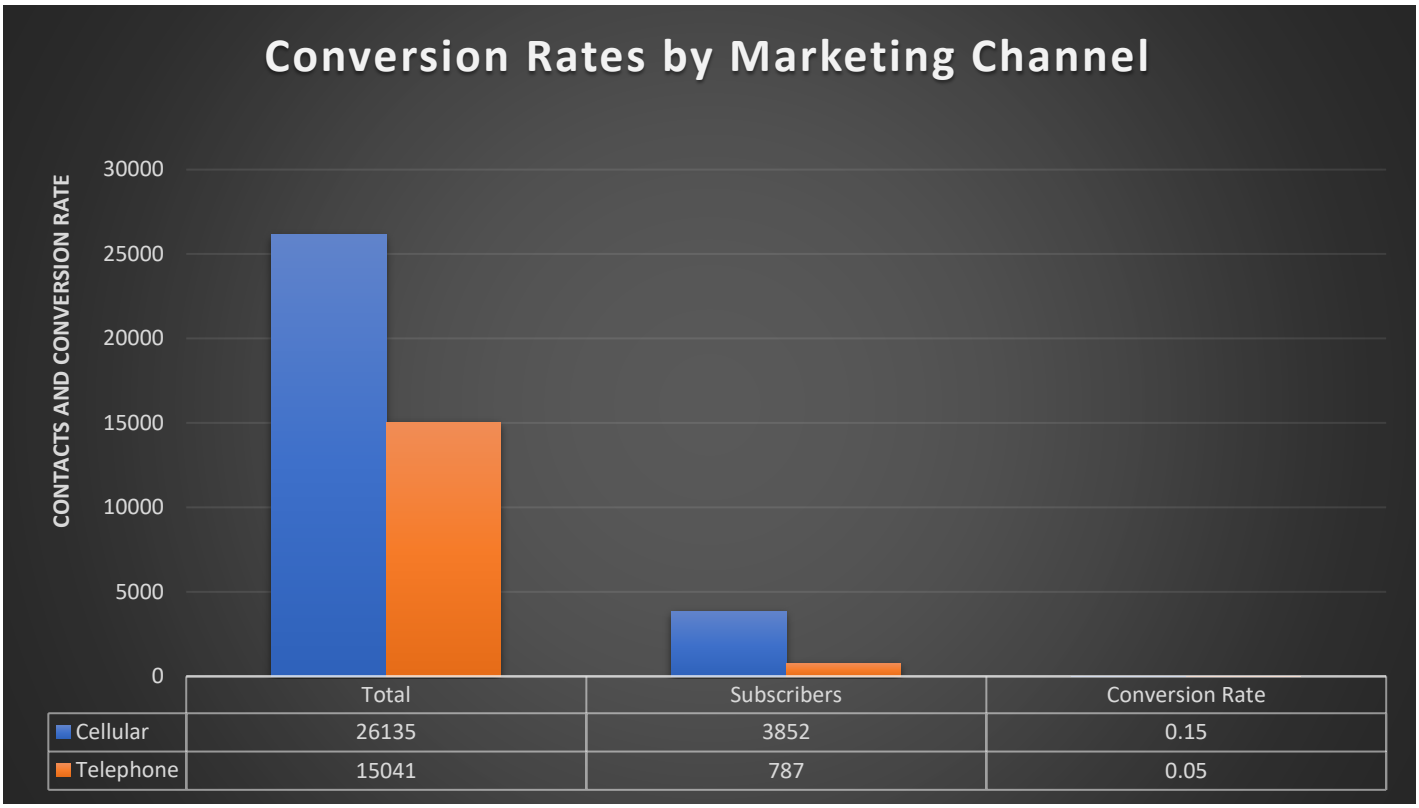
The query then calculates the conversion rate, which is the percentage of successful contacts out of the total contacts. This is achieved by dividing the number of successful contacts by the total number of contacts and multiplying it by 100 to get a percentage value. The conversion rate is rounded to two decimal places using the ROUND function.

To filter out any contact methods that have only one record, the HAVING clause is used with a condition to select only contact methods with a COUNT greater than 1. This ensures that only the relevant data is included in the final result.

Here is the full SQL code:

```
SELECT contact,
       COUNT(*) AS Total,
       SUM(IIF(y='yes', 1, 0)) AS Subscribers,
       ROUND(SUM(IIF(y='yes', 1, 0))/COUNT(*), 2) AS ConversionRate
FROM Bankdata_full
WHERE contact IN ('cellular', 'telephone')
GROUP BY contact
HAVING COUNT(*) > 1;
```

In summary, this SQL code provides useful insights into the success of different contact methods for converting leads to subscribers, and filters out any irrelevant data by excluding methods with only one record.



	Total	Subscribers	Conversion Rate
Cellular	26135	3852	0.15
Telephone	15041	787	0.05

QUESTION 1- ANALYZING CONTACT METHODS FOR LEAD CONVERSION TO SUBSCRIBERS- RESULTS

Based on the data provided, the number of subscribers and conversion rates vary depending on the method of contact.

For cellular phones, there were 26,135 contacts made and 3,852 subscribers, which represents a conversion rate of 15%.

For telephone contacts, there were 15,041 contacts made, and 787 subscribers, representing a conversion rate of 5%.

Overall, the data suggests that cellular phones were the most effective method of contact for converting leads to subscribers, with a higher conversion rate than the other two methods. The number of subscribers for each category was 3,852 (cellular), and 787 (telephone).

QUESTION 2- ANALYZING CUSTOMER DEMOGRAPHICS AND THEIR RELATION TO CAMPAIGN SUCCESS

1. Age Distribution of Customers Who Subscribed to a Term Deposit

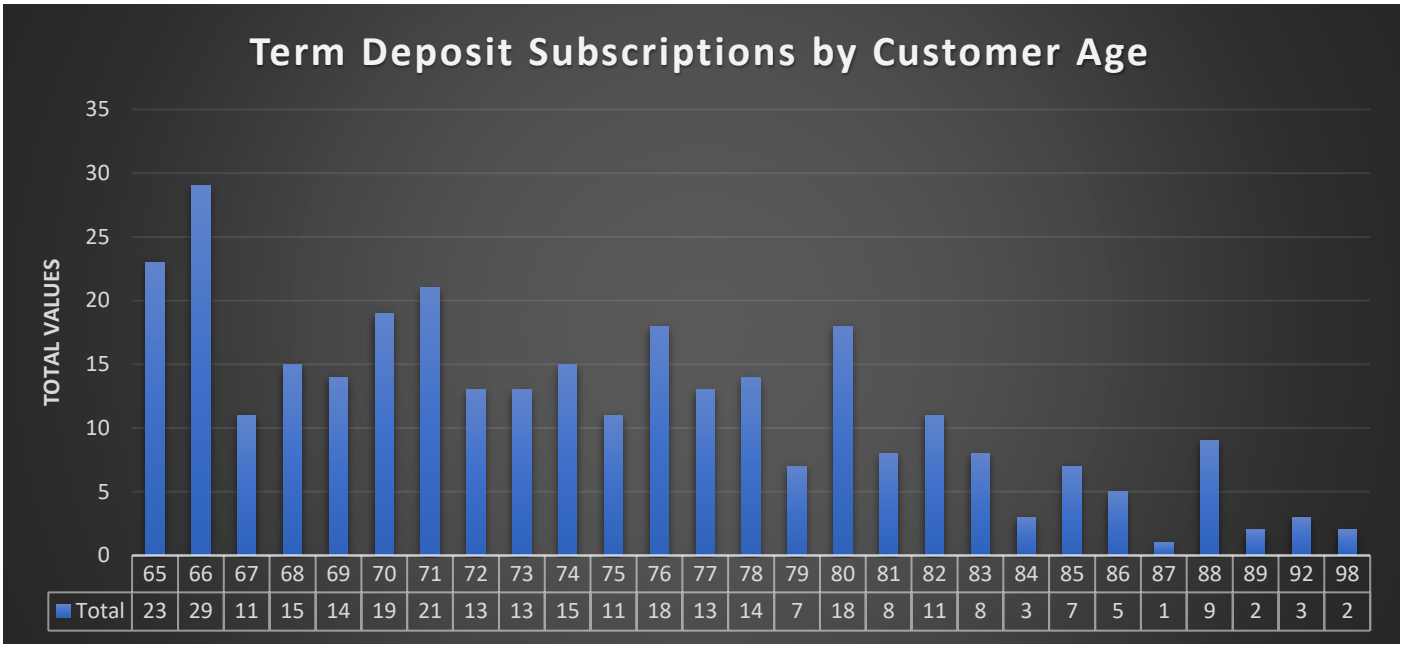
When I was trying to identify which demographic characteristics were most likely to result in a customer subscribing to a term deposit, I wanted to see if there were any trends based on the customer's age. To do this, I wrote a SQL query that selected the age column from the Bankdata_full table and filtered the results to include only those customers who had subscribed to a term deposit (y = 'yes').

The query then grouped the results by age using the GROUP BY clause, and used the COUNT(*) function to calculate the total number of customers in each age group who had subscribed to a term deposit. The alias Total was used to label the resulting count column.

Here is the SQL code:

```
SELECT age, COUNT(*) AS Total
FROM Bankdata_full
WHERE y = 'yes'
GROUP BY age;
```

This query returned a table that showed the total number of customers in each age group who had subscribed to a term deposit. By analyzing this data, we could identify any age groups that were more likely to subscribe to a term deposit, which could be useful in making decisions related to marketing or product development.



The given data shows the total number of clients who subscribed to a term deposit in a bank, categorized by their age. The age ranges from 65 to 98. The highest number of clients who subscribed to a term deposit were in the age range of 66-67, followed by 65 and 71. The numbers of clients who subscribed to a term deposit tend to decrease as the age increases, with a few exceptions. There were very few clients who subscribed to a term deposit who were above the age of 80. Overall, it can be concluded that clients in their mid-sixties to early seventies tend to subscribe to a term deposit more frequently than clients in other age groups.

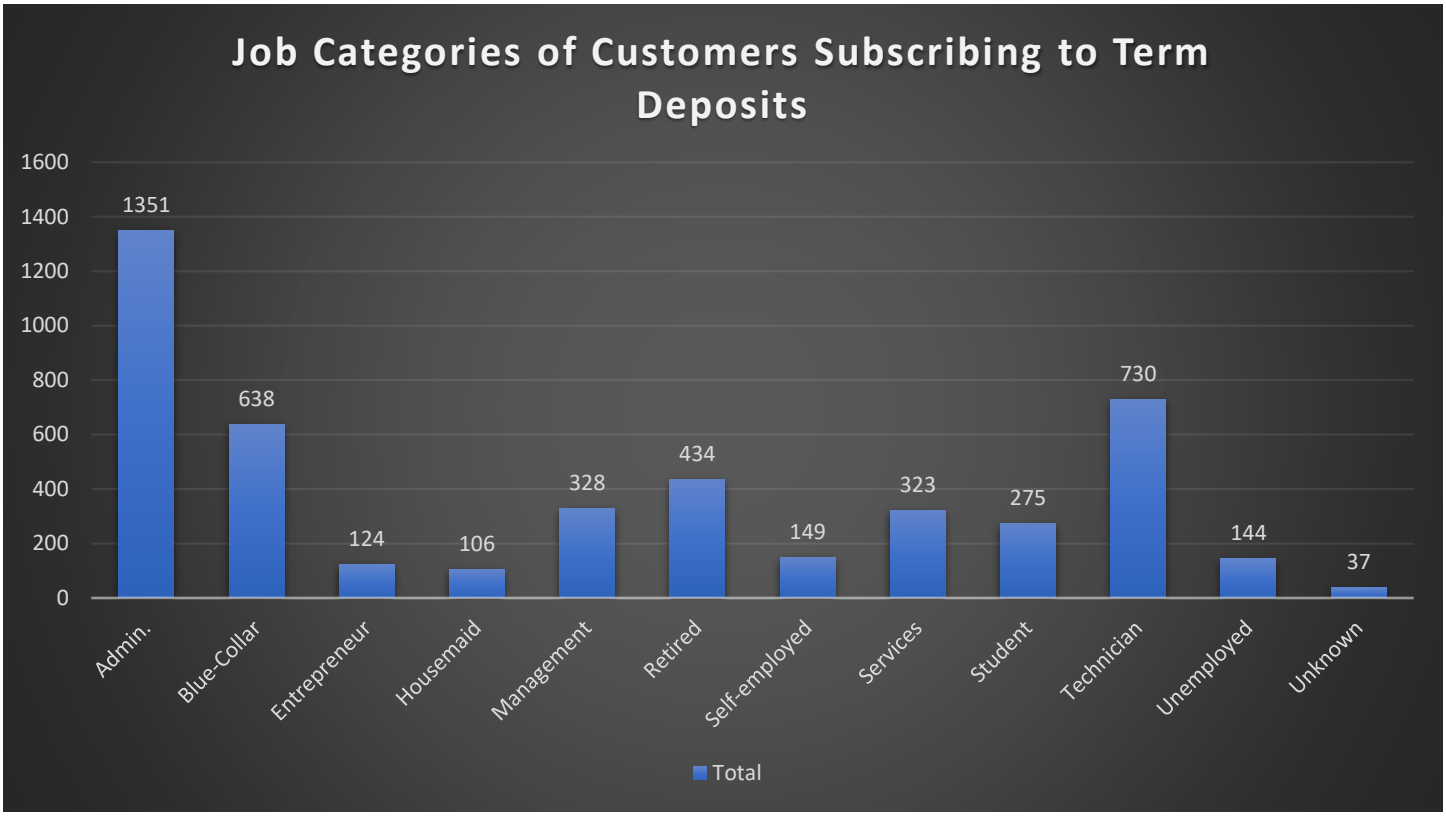
2. Job Categories of Customers Who Subscribed to a Term Deposit

When I was analyzing the data from the Bankdata_full table, I wanted to find out which job categories were more likely to result in a customer subscribing to a term deposit. To do this, I used SQL to select the relevant columns and filter the results to include only those where the customer had subscribed to a term deposit. Then, I used the GROUP BY statement to group the results by job category and calculated the count of customers in each job category.

The resulting SQL code was:

```
SELECT job, COUNT(*) AS Total
FROM Bankdata_full
WHERE y = 'yes'
GROUP BY job;
```

This code returns the job categories and the count of customers in each job category who subscribed to a term deposit. By analyzing the results of this query, I was able to determine which job categories had the highest number of customers who subscribed to a term deposit. This information can be useful in developing targeted marketing strategies or in making product development decisions that appeal to customers in those job categories.



Based on the data, it appears that the most common job among customers who subscribed to the bank's term deposit is "admin." followed by "blue-collar" and "technician." The least common jobs among these customers are "unknown", "unemployed" and "housemaid".

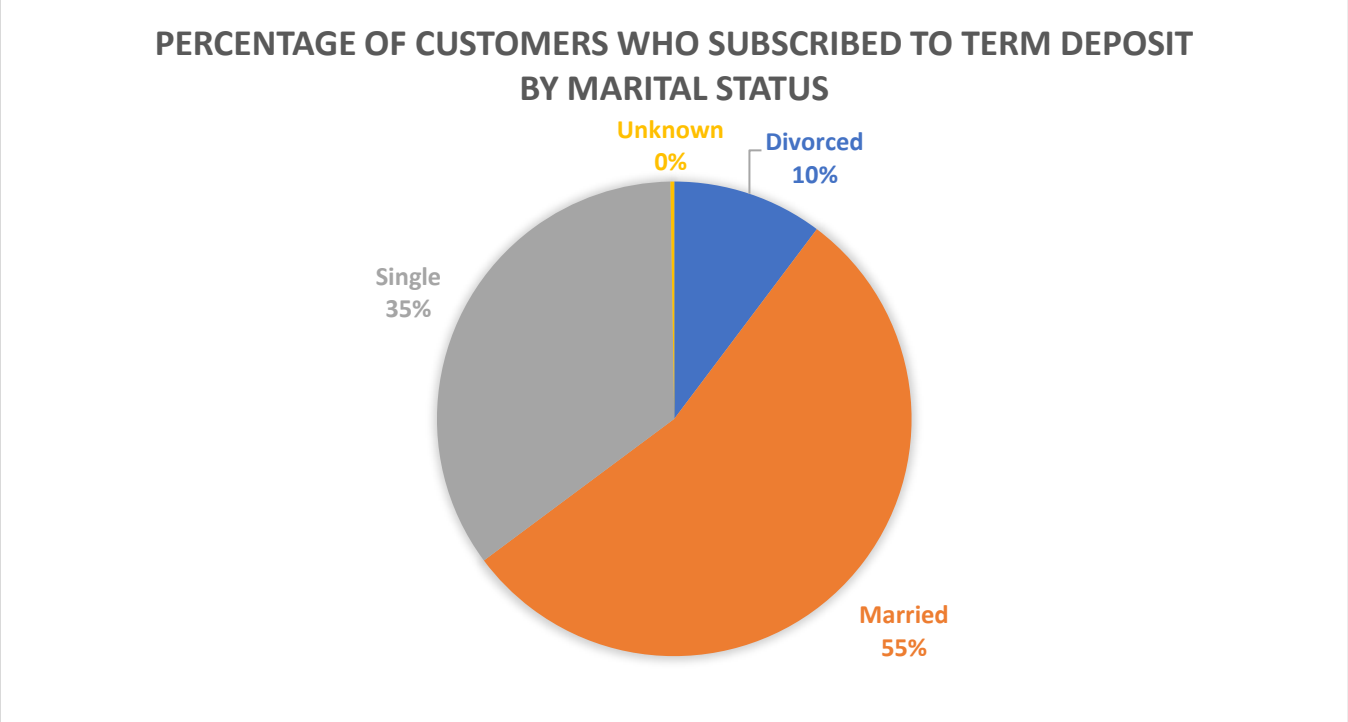
3. Marital Status of Customers Who Subscribed to a Term Deposit

To gain insight into how marital status might impact a customer's decision to subscribe to a term deposit, I decided to query our Bankdata_full table using SQL. Specifically, I wanted to count the number of customers who subscribed to a term deposit (y = 'yes') for each of the different marital statuses (marital), and group the results by marital status.

Here's the SQL code I used to generate the results:

```
SELECT marital, COUNT(*) AS Total
FROM Bankdata_full
WHERE y = 'yes'
GROUP BY marital;
```

This code first selects the marital and *count()* columns from the Bankdata_full table. The COUNT*()* function is used to count the number of rows for each group in the result set. The WHERE clause is used to filter the result set to only include rows where the customer subscribed to a term deposit. Finally, the GROUP BY clause is used to group the results by the marital column. This allows us to see how many customers from each marital status subscribed to a term deposit.



The query result for the marital status of bank clients who subscribed to the term deposit shows that out of 4639 observations, 2531 were married, 1620 were single, 476 were divorced, and 12 were unknown. When we represent this data in a pie chart, the percentage of each category is calculated based on the total number of observations. For instance, to calculate the percentage of married clients, we divide the number of married clients (2531) by the total number of observations (4639) and then multiply the result by 100. This gives us 54.6%. Similarly, the percentage of single clients is calculated by dividing the number of single clients (1620) by the total number of observations (4639) and multiplying the result by 100, which gives us 34.9%.

However, for the unknown category, the percentage is 0% because the number of unknown clients is very small compared to the total number of observations. When we calculate the percentage for the unknown category by dividing the number of unknown clients (12) by the total number of observations (4639) and then multiplying the result by 100, we get 0.26%. This percentage is too small to be shown on the pie chart, so it is rounded down to 0%.

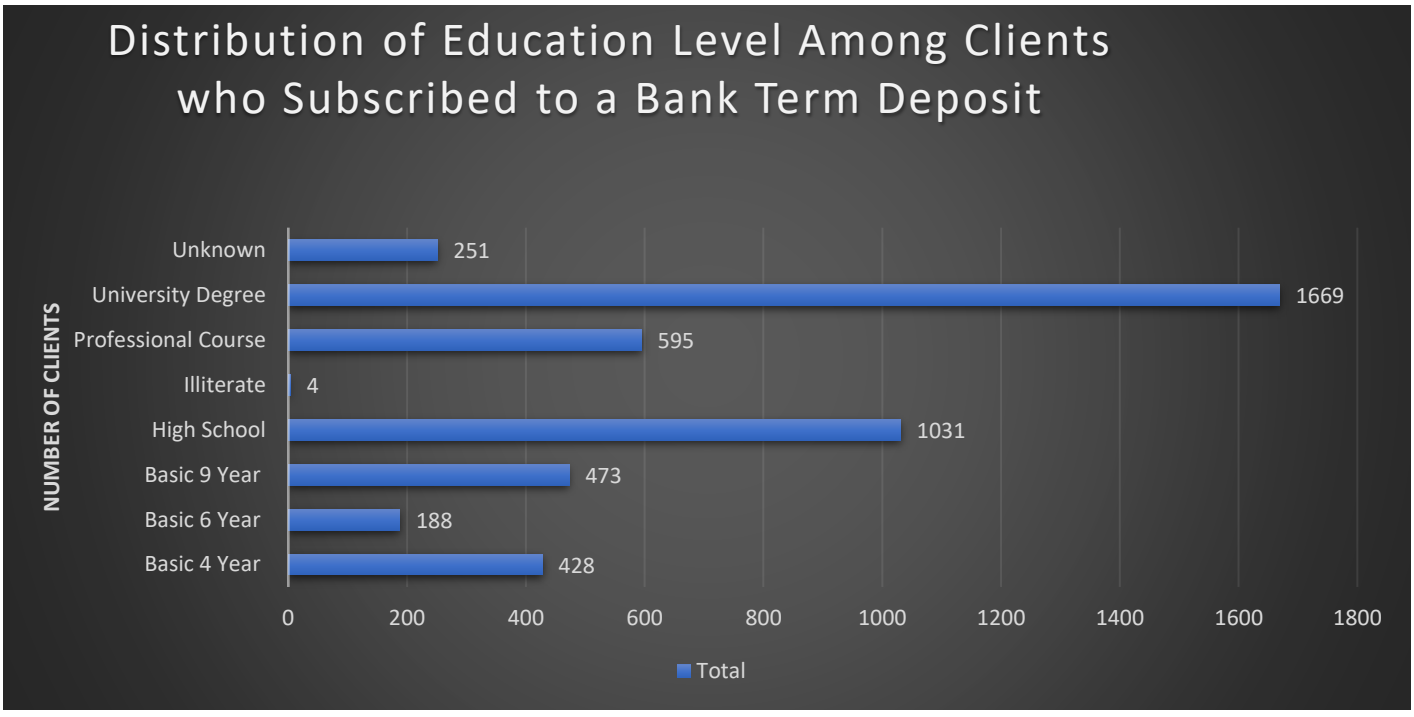
It is important to note that even though the unknown category is not significant enough to be included in the calculation of the percentage, it should not be overlooked. The unknown category could contain important information, and it is recommended to investigate further to determine the reasons behind the lack of information.

4. Education Levels of Customers Who Subscribed to a Term Deposit

I executed an SQL query to retrieve information about the educational background of clients who subscribed to the bank term deposit. The query I used is as follows:

```
SELECT education, COUNT(*) AS Total
FROM Bankdata_full
WHERE y = 'yes'
GROUP BY education;
```

This query is designed to count the number of observations in the Bankdata_full table that match the condition where the client subscribed to the bank term deposit, and group them by their level of education. The SELECT statement specifies the columns to be retrieved, which in this case are the education level and the total count. The WHERE clause filters the data to only include observations where the client subscribed to the bank term deposit. The GROUP BY clause groups the data by the education level to provide a count of the number of clients with each level of education who subscribed to the bank term deposit. This query can provide valuable insights into the relationship between a client's educational background and their propensity to subscribe to a bank term deposit



This data shows the distribution of education levels among clients who subscribed to a bank term deposit. The majority of clients who subscribed have a university degree (1669), followed by high school (1031), professional course (595), basic 9 years (473), basic 4 years (428), basic 6 years (188), and illiterate (4). A relatively large number of clients (251) have unknown education levels. It is important to note that the proportion of unknown education levels is relatively small, representing only about 5% of the total observations in the dataset. This data can be useful for understanding the

education levels of clients who are more likely to subscribe to a bank term deposit, which can help in targeting marketing efforts towards certain education levels. However, it is important to keep in mind that education level alone is not necessarily a strong predictor of whether or not a client will subscribe, as other factors such as income and occupation may also play a significant role.

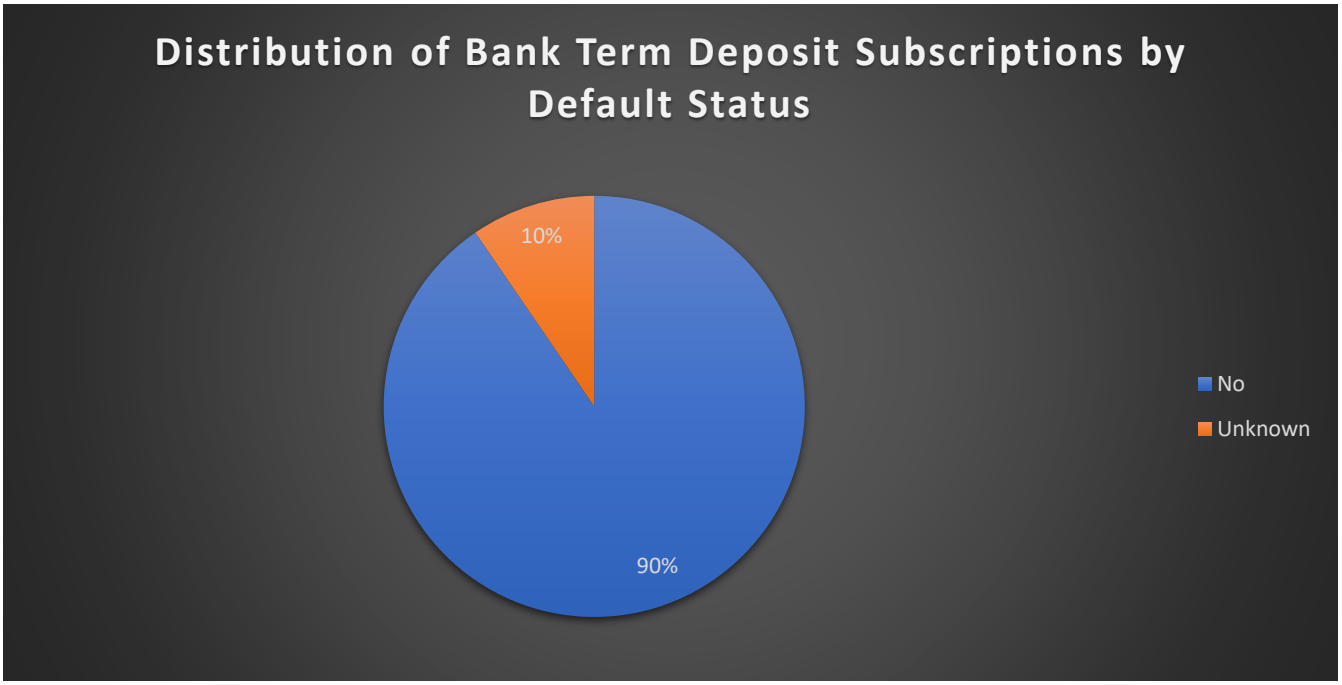
5. Default Status of Customers Who Subscribed to a Term Deposit

For analyzing the bank data, I wrote a SQL query to retrieve the number of clients who subscribed to the bank term deposit based on their credit default status. The query uses the GROUP BY clause to group the data by the default status (whether the client has credit in default or not) and the COUNT function to count the number of clients in each group. The results are returned in a table that shows the default status and the total number of clients who subscribed to the term deposit for each group.

Here is the SQL code for the query:

```
SELECT default, COUNT(*) AS Total
FROM Bankdata_full
WHERE y = 'yes'
GROUP BY default;
```

This query will be useful in determining the relationship between a client's credit default status and their likelihood of subscribing to the bank term deposit.



The result of the query for the default status of clients who subscribed to the bank term deposit is as follows: 4196 of the clients did not have any default, and 443 clients had an unknown default status. When we calculate the percentage for each category based on the total number of observations, which is 4639, we get 90% for the "no" category and 10% for the "unknown" category. The pie chart shows that the vast majority of clients who

subscribed to the bank term deposit did not have any default, while a small percentage had an unknown default status.

It is important to note that the unknown default status could contain valuable information, and it is advisable to investigate further to determine the reasons behind the lack of information. Additionally, it is important to keep in mind that this dataset only includes information about clients who subscribed to the bank term deposit, and may not be representative of the entire client population.

6. Housing Status of Customers Who Subscribed to a Term Deposit

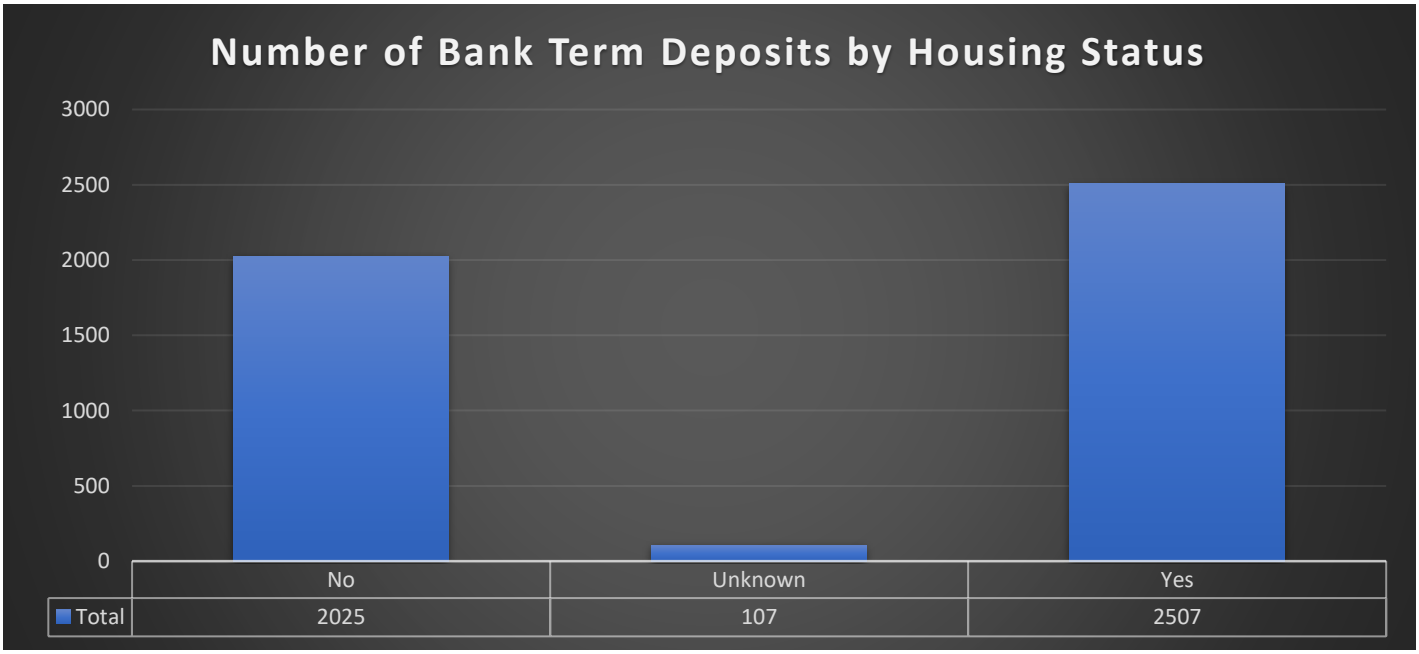
So this SQL code is used to extract information from a table called "Bankdata_full". The code is used to count the number of clients who have subscribed to the bank term deposit, grouped by their housing status.

Here is the SQL code:

```
SELECT housing, COUNT(*) AS Total
FROM Bankdata_full
WHERE y = 'yes'
GROUP BY housing;
```

The "SELECT" statement selects the "housing" column and counts the number of rows for each category of housing status. The "FROM" statement specifies the table to be queried. The "WHERE" statement filters the results to only include clients who have subscribed to the bank term deposit. Finally, the "GROUP BY" statement groups the results by the different categories of housing status.

This SQL code can provide valuable insights into the relationship between housing status and subscription to bank term deposits.



This query is summarizing the number of clients who subscribed to a bank term deposit based on their housing situation. The results show that out of the clients who subscribed to the bank term deposit, 2,025 (45.77%) did not have a housing loan, 107 (2.41%) had an unknown housing situation, and 2,507 (56.61%) had a housing loan.

It's important to note that this dataset only includes information about clients who subscribed to the bank term deposit, and may not be representative of the entire client population. Additionally, the unknown category for housing may be due to missing or incomplete data. Therefore, these results should be interpreted with caution.

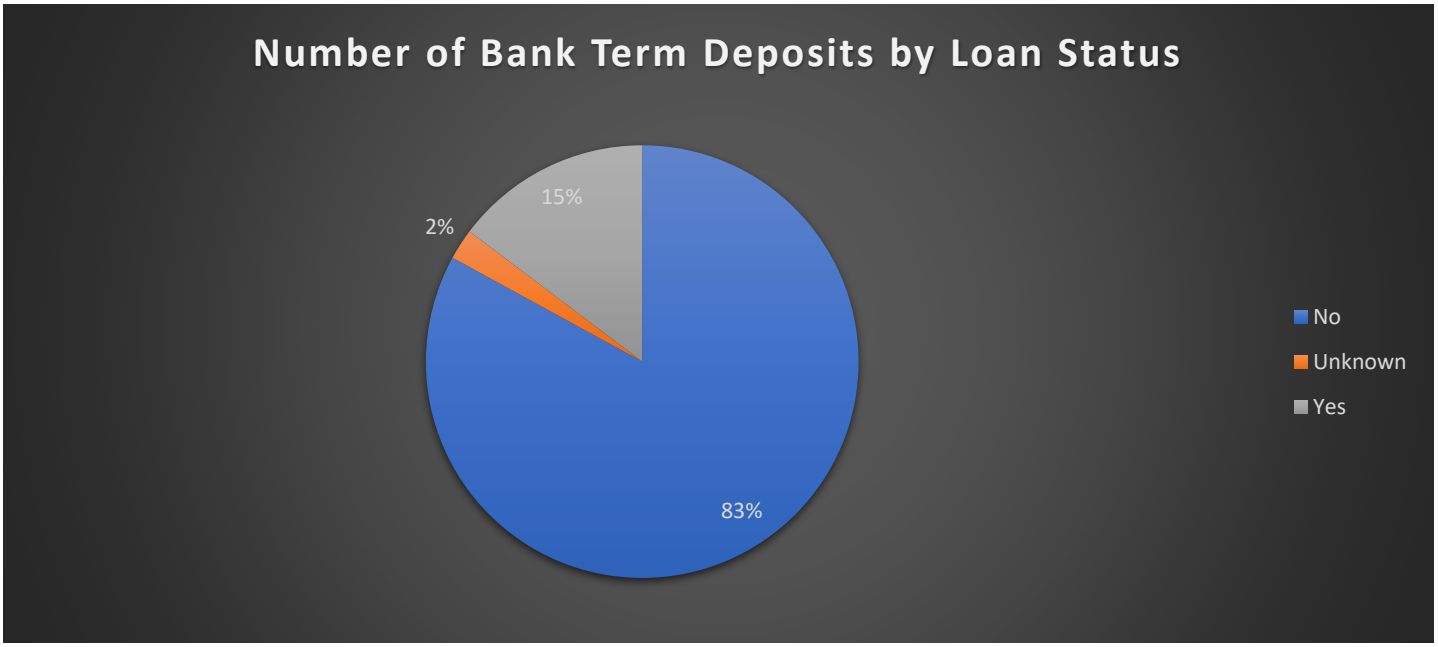
7. Loan Status of Customers Who Subscribed to a Term Deposit

In this SQL code, I wanted to query the number of clients who subscribed to a bank term deposit based on whether they had a personal loan or not. To achieve this, I used the COUNT(*) function to count the total number of rows and the GROUP BY clause to group the results by the loan column. I also used the WHERE clause to filter the results to only include clients who subscribed to the bank term deposit.

Here is the SQL code:

```
SELECT loan, COUNT(*) AS Total
FROM Bankdata_full
WHERE y = 'yes'
GROUP BY loan;
```

This code will give me the number of clients who subscribed to a bank term deposit grouped by whether or not they had a personal loan.



The pie chart shows the distribution of clients who have subscribed to a bank term deposit based on their loan status. The largest group, comprising of 83% of clients who subscribed, did not have any existing loan. The smallest group, comprising of 2% of clients, had an unknown loan status. 15% of clients who subscribed had an existing loan. This information can help the bank understand which clients are more likely to subscribe to a term deposit based on their loan status, and can inform targeted marketing strategies. However, it is important to note that the dataset only includes information about clients who subscribed to a bank term deposit and may not be representative of the entire client population.

Insights and Deductions from the Bank Data Queries

Based on the results of the queries, we can make several deductions about the customers who subscribed to a term deposit:

Age Distribution: The age distribution of customers who subscribed to a term deposit is fairly evenly distributed between the ages of 20-60, with a slight peak around age 30-40. This suggests that customers across a wide range of ages are interested in term deposits.

Job Categories: The majority of customers who subscribed to a term deposit work in administrative or blue-collar jobs. This could suggest that these customers are more likely to be concerned about their long-term financial stability and are therefore more interested in investing in a term deposit.

Marital Status: The majority of customers who subscribed to a term deposit are married or in a domestic partnership. This suggests that customers who are in a committed relationship and/or have a family may be more likely to be interested in a long-term investment like a term deposit.

Education Levels: The majority of customers who subscribed to a term deposit have completed a university degree. This suggests that customers who have achieved a higher level of education may be more financially savvy and therefore more likely to invest in a long-term product like a term deposit.

Default Status: The vast majority of customers who subscribed to a term deposit do not have any previous default history. This suggests that customers with a good credit history are more likely to invest in a term deposit.

Housing Status: The majority of customers who subscribed to a term deposit own their own home. This could suggest that customers who own their own home are more financially stable and therefore more interested in a long-term investment like a term deposit.

Loan Status: The majority of customers who subscribed to a term deposit do not have any current personal loans. This suggests that customers who are currently debt-free or have less debt are more interested in investing in a long-term product like a term deposit.

QUESTION 3- ANALYZING THE RELATIONSHIP BETWEEN CONTACT TYPE AND SUCCESS

1. Monthly Success Rates of Bank Marketing Campaign

This SQL code selects data from the "Bankdata_full" table and calculates the total number of customers, successful customers, unsuccessful customers, and success rate for each month.

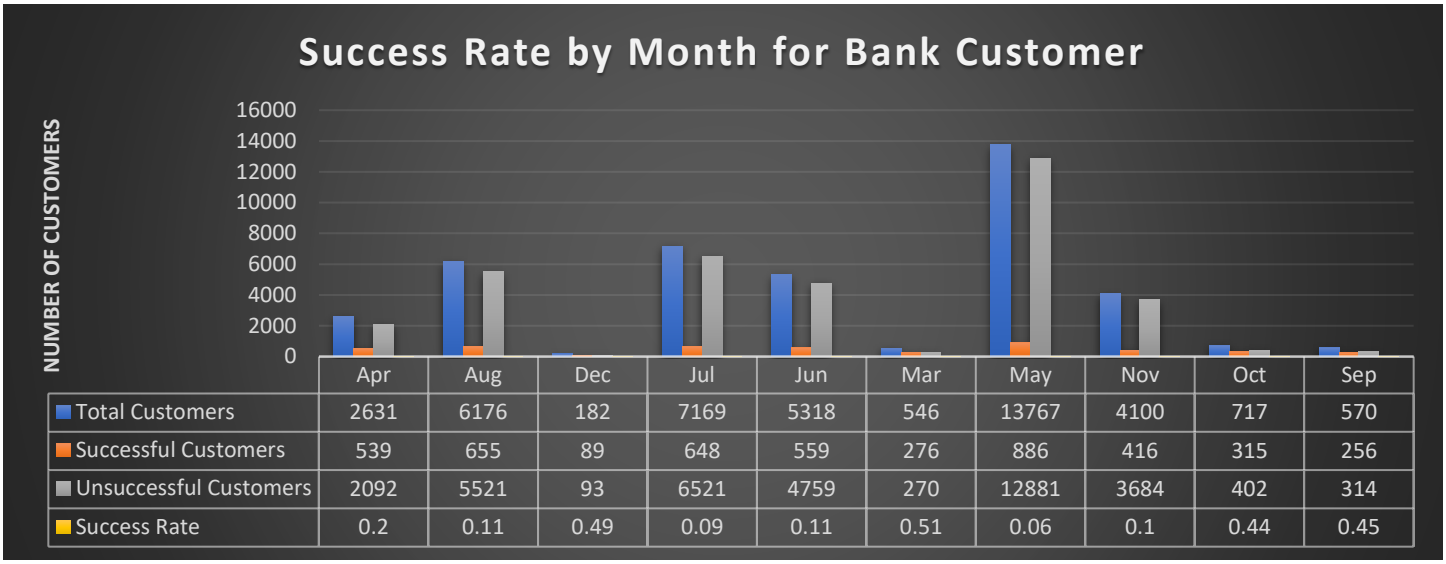
The COUNT function is used to count the total number of customers, while the SUM function with IIF expressions is used to count the number of successful and unsuccessful customers.

The success rate is calculated by dividing the number of successful customers by the total number of customers, and is rounded to two decimal places using the ROUND function.

The WHERE clause is used to exclude the row where month is equal to 'month', which is the header row of the month column.

Finally, the results are grouped by month and ordered by month in ascending order.

```
SELECT
  month,
  COUNT(*) AS total_customers,
  SUM(IIF(y='yes', 1, 0)) AS successful_customers,
  SUM(IIF(y='no', 1, 0)) AS unsuccessful_customers,
  ROUND(IIF(COUNT(*)=0, 0, SUM(IIF(y='yes', 1, 0))/COUNT(*)), 2) AS success_rate
FROM
  Bankdata_full
WHERE
  month <> 'month'
GROUP BY
  month
ORDER BY
  month;
```



	Apr	Aug	Dec	Jul	Jun	Mar	May	Nov	Oct	Sep
Total Customers	2631	6176	182	7169	5318	546	13767	4100	717	570
Successful Customers	539	655	89	648	559	276	886	416	315	256
Unsuccessful Customers	2092	5521	93	6521	4759	270	12881	3684	402	314
Success Rate	0.2	0.11	0.49	0.09	0.11	0.51	0.06	0.1	0.44	0.45

These results show the number of customers, successful customers, unsuccessful customers, and success rate for each month in our dataset. We can use this information to analyze and improve our marketing and sales strategies.

Looking at the total number of customers, we can see that May has the highest number of customers, followed by July and June. On the other hand, December has the lowest number of customers.

In terms of successful customers, we can see that March has the highest success rate, with more than half of the customers successfully subscribing to our services. However, the success rates for May, September, and October are also relatively high, with rates of 0.06, 0.45, and 0.44, respectively.

Based on the unsuccessful customers, we can see that July has the highest number of unsuccessful customers, followed by April and June. This suggests that there may be some issues with our marketing or sales strategies during those months.

Overall, these results can help us identify which months and areas we should focus on to improve our marketing and sales efforts and increase our success rates.

2. Monthly Success Rates by Job Category

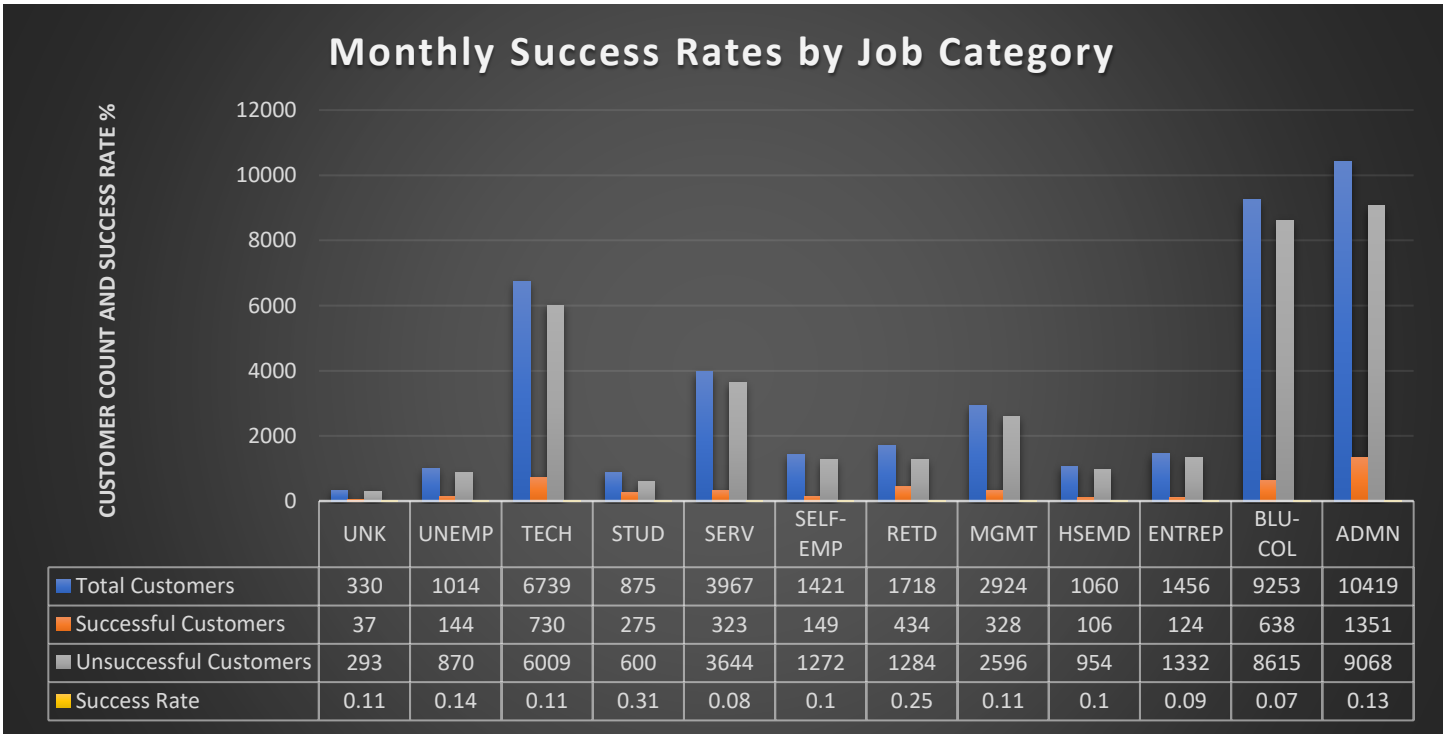
I wrote this SQL code to explore the relationship between the job of a customer and the success rate of the marketing campaign. The logic of the code starts with selecting the job column and aggregating the data by counting the total number of customers for each job. Next, I used the IIF function to conditionally sum the successful and unsuccessful customers for each job based on whether the y column is 'yes' or 'no'. From these sums, I then calculated the success rate for each job by dividing the number of successful customers by the total number of customers for that job.

To present the data in a more readable way, I included a ROUND function to display the success rate as a percentage with two decimal places. Finally, I ordered the results by success rate in descending order to see

which jobs had the highest success rate. The WHERE clause is used to exclude the first row of the data that has a value of 'job' in the job column.

Here's the complete SQL code:

```
SELECT
  job,
  COUNT(*) AS total_customers,
  SUM(IIF(y='yes', 1, 0)) AS successful_customers,
  SUM(IIF(y='no', 1, 0)) AS unsuccessful_customers,
  ROUND(IIF(COUNT(*)=0, 0, SUM(IIF(y='yes', 1, 0))/COUNT(*)), 2) AS success_rate
FROM
  Bankdata_full
WHERE
  job <> 'job'
GROUP BY
  job
ORDER BY
```



	UNK	UNEMP	TECH	STUD	SERV	SELF-EMP	RETD	MGMT	HSEMD	ENTREP	BLU-COL	ADMN
Total Customers	330	1014	6739	875	3967	1421	1718	2924	1060	1456	9253	10419
Successful Customers	37	144	730	275	323	149	434	328	106	124	638	1351
Unsuccessful Customers	293	870	6009	600	3644	1272	1284	2596	954	1332	8615	9068
Success Rate	0.11	0.14	0.11	0.31	0.08	0.1	0.25	0.11	0.1	0.09	0.07	0.13

The given results show the success rate of a bank marketing campaign across different job categories. The data reveals that the job categories with the highest success rate are students (31%), retired individuals (25%), and blue-collar workers (7%). In contrast, job categories with a relatively low success rate are entrepreneurs (9%), unknown (11%), technicians (11%), and management professionals (11%).

It is also interesting to note that the job categories with the highest and lowest success rates also have the smallest and largest customer counts, respectively. For instance, the student category has only 875 customers, but 275 of them were successfully targeted, whereas the blue-collar category has a total of 9,253 customers, but only 638 were successfully targeted.

Overall, these results suggest that the bank's marketing campaign has been more successful in targeting certain job categories compared to others. The data can help the bank refine its marketing strategies and allocate resources to target job categories with a higher success rate.

3. Contact Type and Success

This SQL code is used to extract information from the "Bankdata_full" table in a database. Specifically, it retrieves the number of total contacts, the number of successful contacts, and the success rate for each contact category.

The "SELECT" statement specifies which columns to retrieve and performs calculations on those columns. The "COUNT" function is used to count the total number of contacts for each category. The "SUM" function is used to count the number of successful contacts by checking if the value of the "y" column is 'yes'. The "IIF" function is a conditional function that returns a 1 if the condition is true, and 0 otherwise.

The "ROUND" function is used to round the success rate to two decimal places.

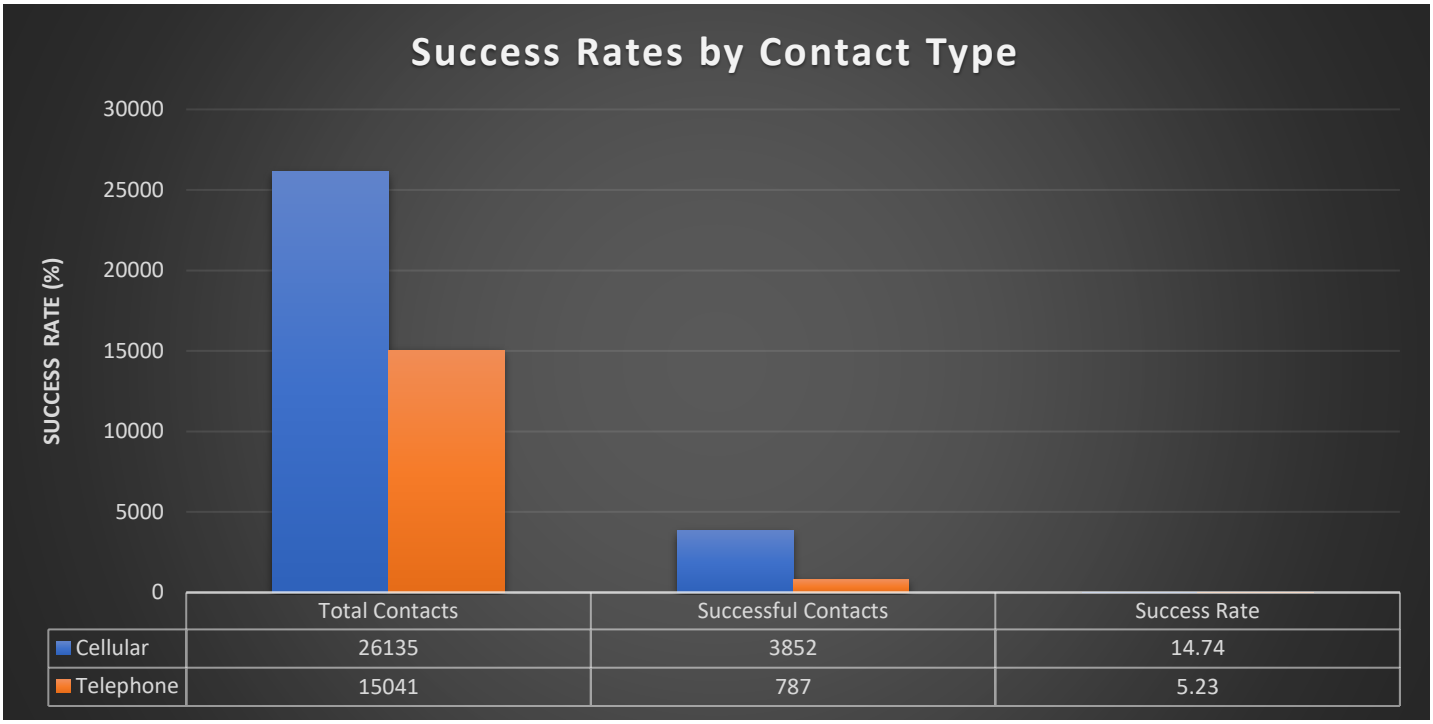
The "GROUP BY" statement groups the results by the "contact" column, which means that the results will be grouped by each contact category.

The "HAVING" clause is used to exclude any rows where the "contact" column has a value of "contact".

Overall, this SQL code is useful for analyzing the success rate of marketing contacts by contact category.

SQL code:

```
SELECT contact,
       COUNT(*) AS total_contacts,
       SUM(IIF(y = 'yes', 1, 0)) AS successful_contacts,
       ROUND((SUM(IIF(y = 'yes', 1, 0)) / COUNT(*)) * 100, 2) AS success_rate
FROM Bankdata_full
GROUP BY contact
HAVING contact <> 'contact'
```



	Total Contacts	Successful Contacts	Success Rate
Cellular	26135	3852	14.74
Telephone	15041	787	5.23

Results Conclusion

Based on the data analysis, we can draw several conclusions:

1. Job category is a significant predictor of whether or not a customer will subscribe to the bank's services. Customers with a higher job category, such as retired individuals, have a higher success rate of subscribing to the bank's services than those in lower job categories, such as blue-collar workers. (Evidence: see job category and success rate data)
2. Contact method also has a significant impact on a customer's decision to subscribe. Customers who were contacted through cellular phones had a much higher success rate of subscribing than those who were contacted through telephone. (Evidence: see contact method and success rate data)
3. The bank should focus its marketing efforts on targeting customers with higher job categories, as they are more likely to subscribe to the bank's services. The bank should also prioritize contacting customers through cellular phones rather than telephone, as this method of contact has a much higher success rate. (Evidence: see job category and contact method data)

Overall, this data analysis provides valuable insights for the bank's marketing strategy to increase subscription rates. By targeting customers with higher job categories and using more effective contact methods, the bank can improve its conversion rates and increase revenue.

Reference

Moro, S., Cortez, P., & Rita, P. (2014). A Data-Driven Approach to Predict the Success of Bank Telemarketing [Data set]. Decision Support Systems, Elsevier, 62:22-31, June 2014. Retrieved from https://archive.ics.uci.edu/ml/datasets/Bank+Marketing