Data Glacier

Week 10: Final Project

Bank Marketing Campaign

- Data Science -

Table of Contents

Team member's details

	Group Name: *Data Science Enthusiasts*				
	Name	**Email**	**Country**	**College/Company**	**Specialization**
1	Amira Asta	amira.asta02@gmail.com	Tunisia	Afrikanda	Data Science
2	Vatsal Vinesh Mandalia	vatsalvm10@outlook.com	Oman	Graduated	Data Science

Problem description

ABC Bank wants to sell its term deposit product to customers and before launching the product they want to develop a model which helps them understand whether a particular customer will buy their product or not. In order to achieve this task, they approached an Analytics company to automate this process of classification. The Analytics company has given responsibility to the **Data Science Enthusiasts** Team and has asked to come up with a ML model to shortlist customers whose chances of buying the product is higher, so that ABC's marketing channel can focus only on those customers.

Business understanding

There has been a revenue decline for an ABC bank and they would like to know what actions to take. After investigation, they found out that the root cause is that their clients are not depositing as frequently as before. Knowing that term deposits allow banks to hold onto a deposit for a specific amount of time, banks can invest in higher gain financial products to make a profit.

In addition, banks also hold better chances to persuade term deposit clients into buying other products such as funds or insurance to further increase their revenues. As a result, the ABC bank would like to identify existing clients that have higher chances to subscribe for a term deposit and focus marketing efforts on such clients. The classification goal is to predict if the client will subscribe to a term deposit or not.

EDA

For the main objective of this project to be achieved, an appropriate exploratory data analysis needs to be done in order to explore the dataset provided in depth. The exploration will help to provide a clearer understanding of the data characteristics, size, and structure.

Data set features description:

Table 1: Dataset input features.

N°	Feature name	Description	Type
1	age	age	numeric
2	job	type of job	categorical
3	marital	marital status	categorical
4	education	level of education	categorical
5	default	has credit in default?	categorical
6	housing	has housing loan?	categorical
7	loan	has personal loan?	categorical
8	contact	contact communication type	categorical
9	month	last contact month of year	categorical
10	day of week	last contact day of the week	categorical
11	duration	last contact duration, in seconds	numeric
12	campaign	number of contacts performed in this campaign	numeric
13	pdays	number of days passed by after the last contact	numeric
14	previous	number of contacts performed for this client	numeric
15	poutcome	outcome of the previous marketing campaign	categorical
16	emp.var.rate	employment variation rate	numeric
17	cons.price.idx	consumer price index - monthly indicator	numeric
18	cons.conf.idx	consumer confidence index - monthly indicator	numeric
19	euribor3m	euribor 3 month rate - daily indicator	numeric
20	nr.employed	number of employees - quarterly indicator	numeric
21	y	has the client subscribed a term deposit?	binary

Descriptive analysis (univariate analysis):

- **Numerical attributes:**

Distribution histogram plots as well as density plots were done in order to quickly get a feeling for whether an attribute is Gaussian, skewed or even has an exponential distribution. It can also help you see possible outliers.

Figure 1 - Histogram plots of numerical attributes

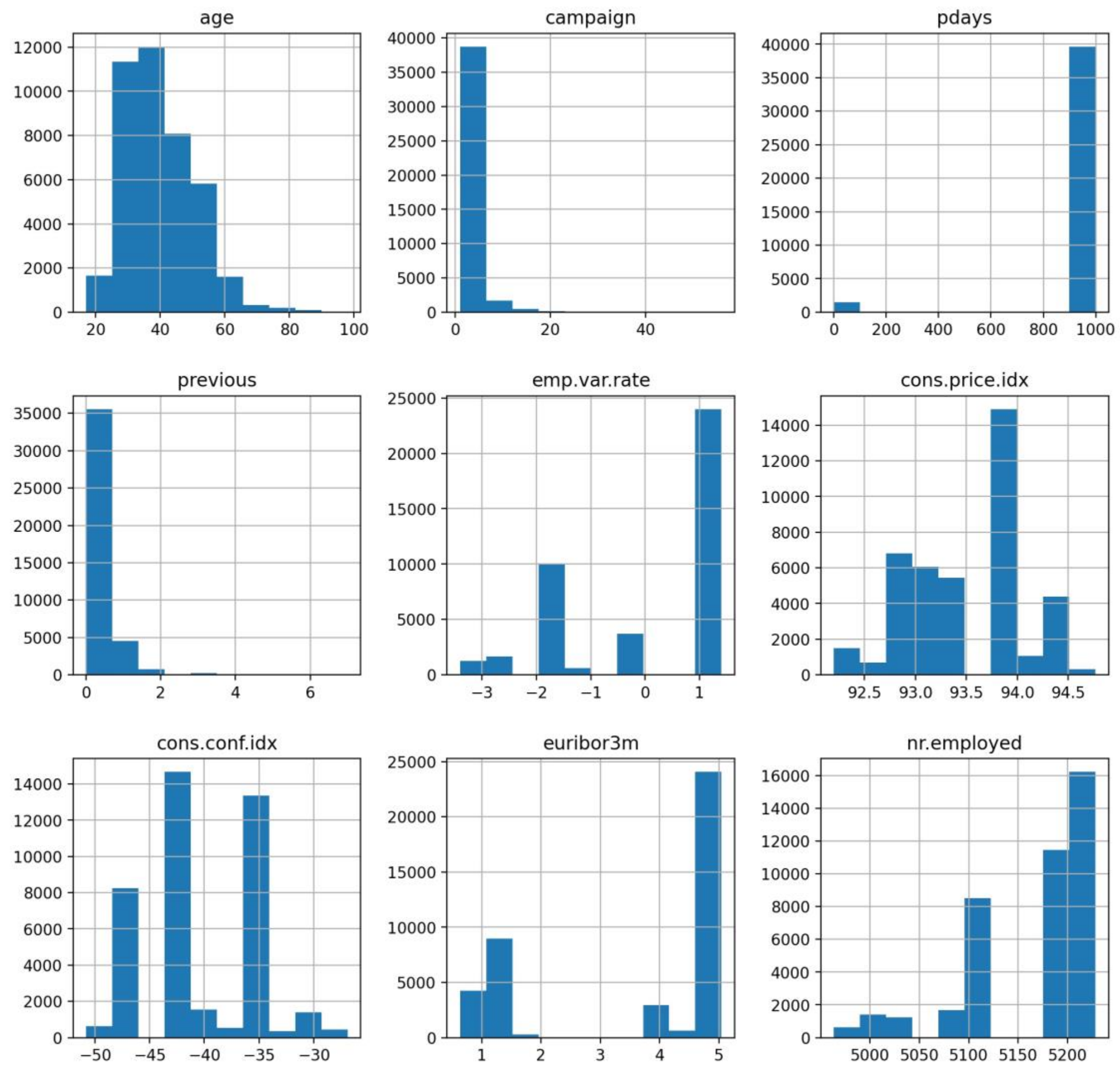


Figure 2 - Density plots of different numerical attributes

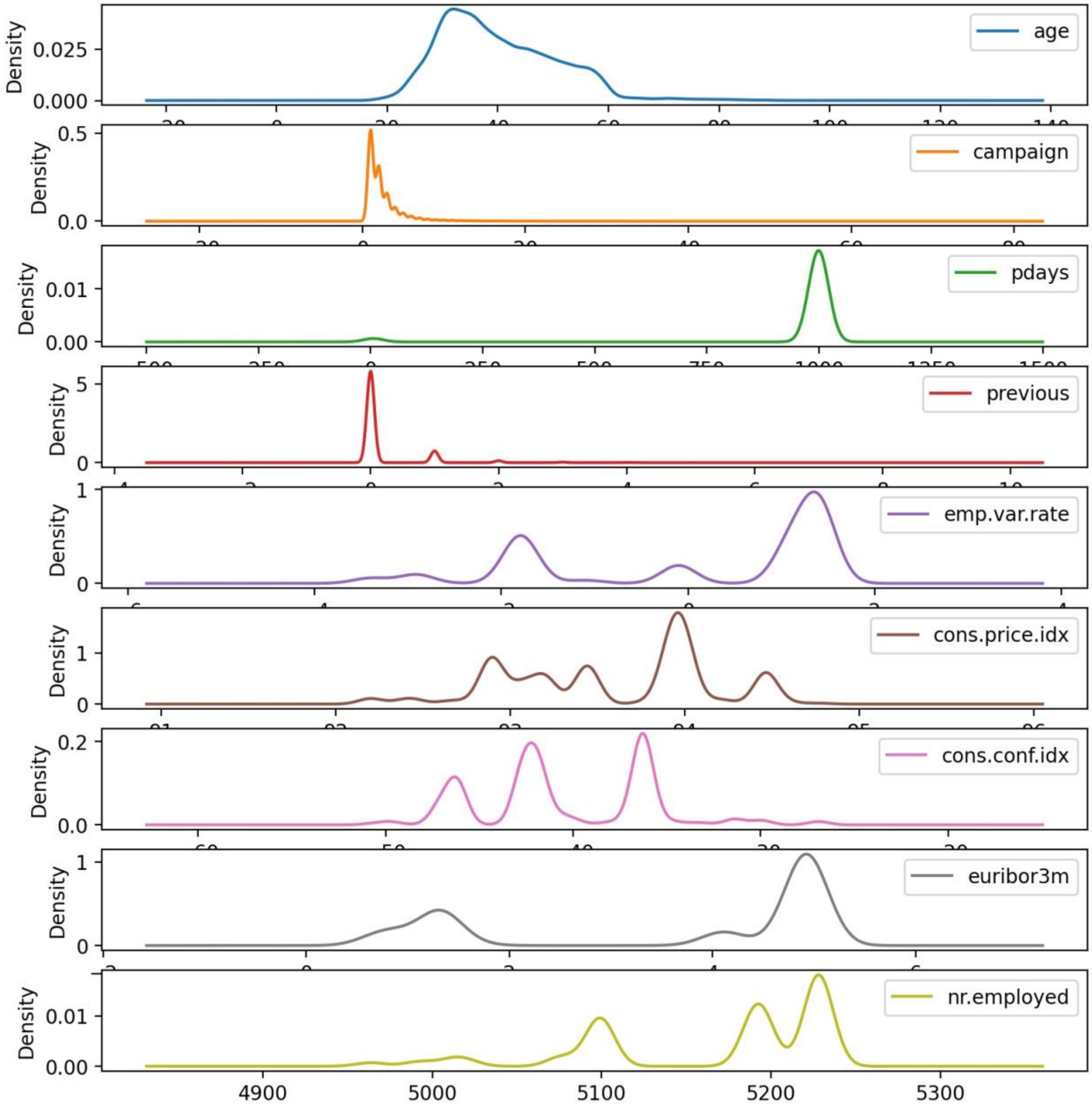
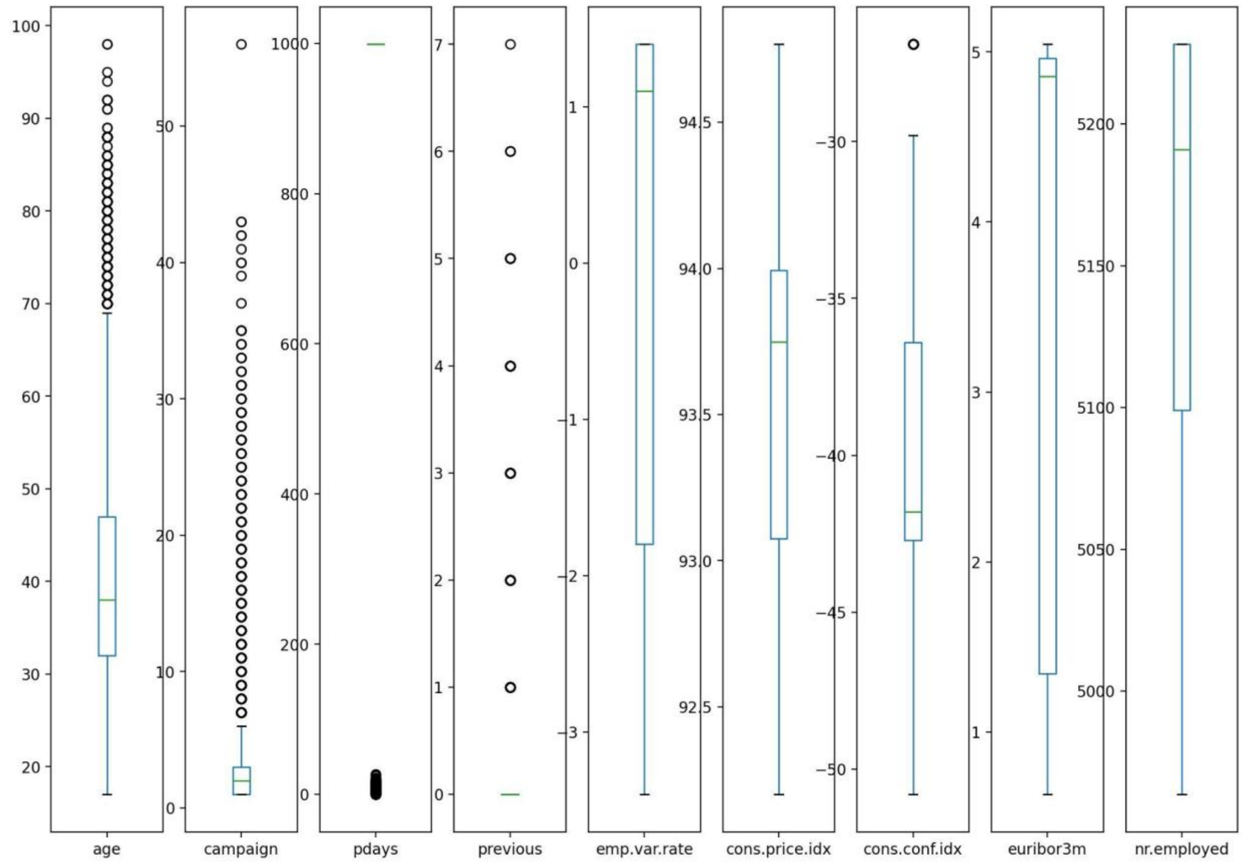
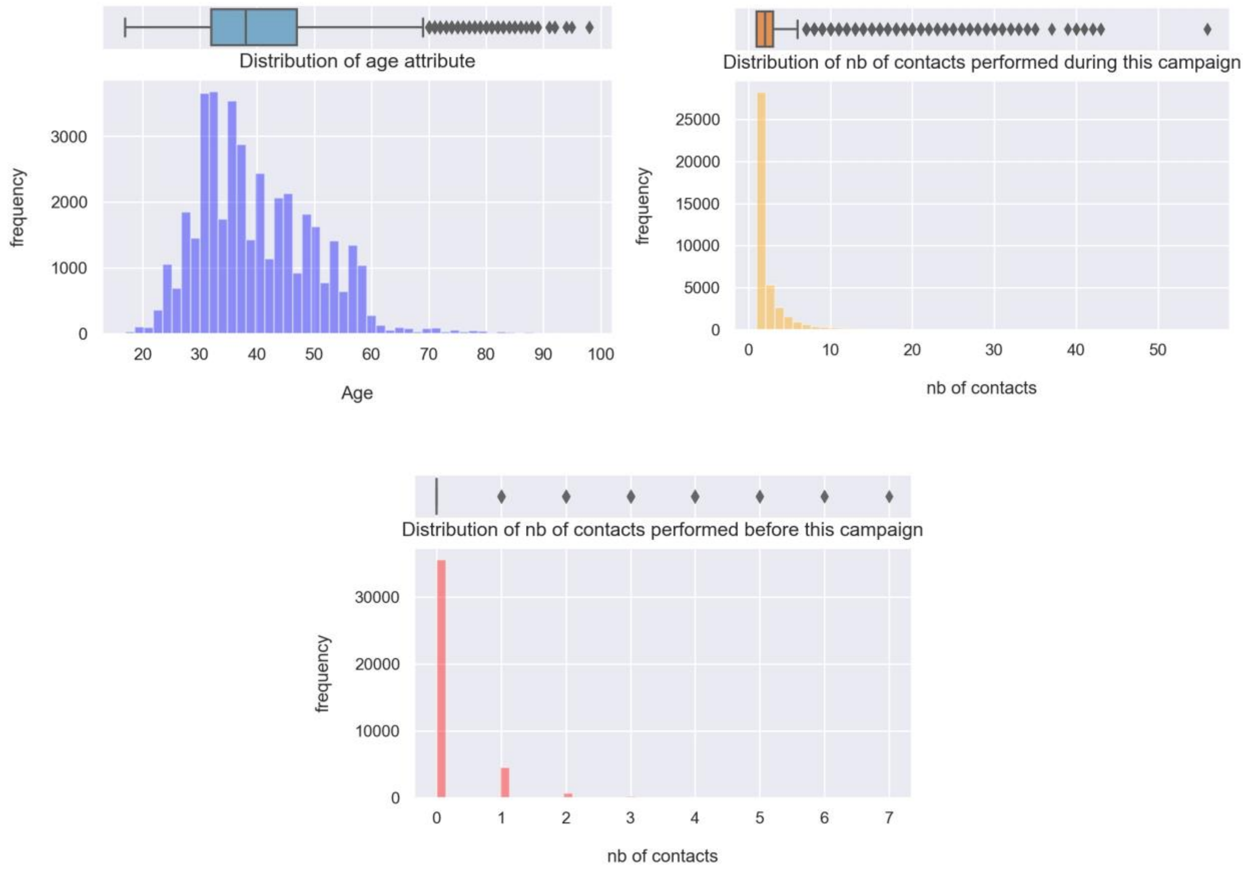


Figure 3 - Box plot of each numerical attribute



From the above plots, we noticed the existence of outlier values present in some features i.e. ‘age’, ‘campaign’ and ‘previous’. An outlier is a value/observation which lies at an abnormal distance from other values in the normal distribution. It can occur due to an error in measurement or data collection. Outliers can affect the mean of the distribution. Let’s take a closer look at these outlier values:

Figure 4 - Distribution of ‘age’, ‘campaign’ and ‘previous’ attributes

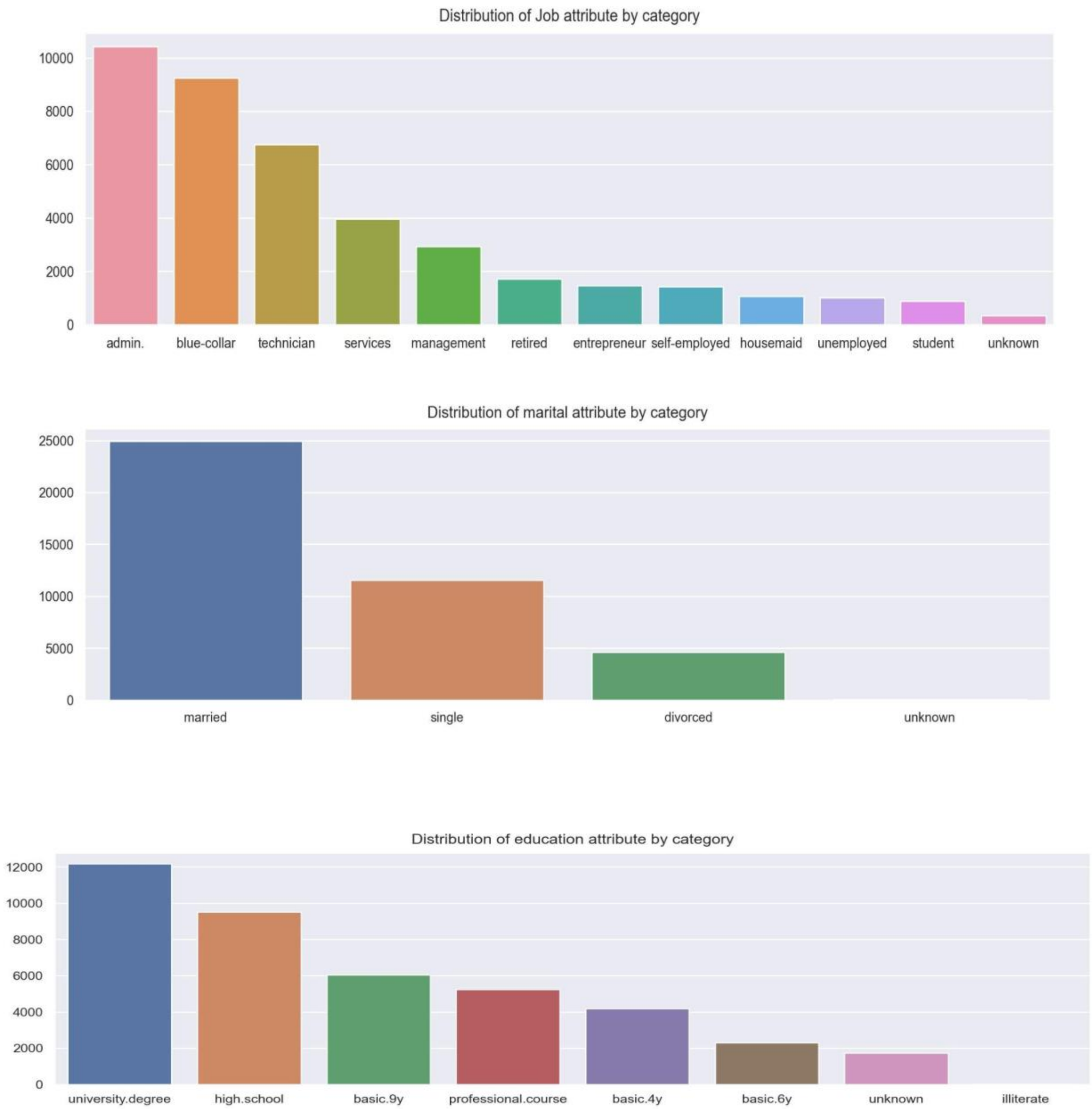


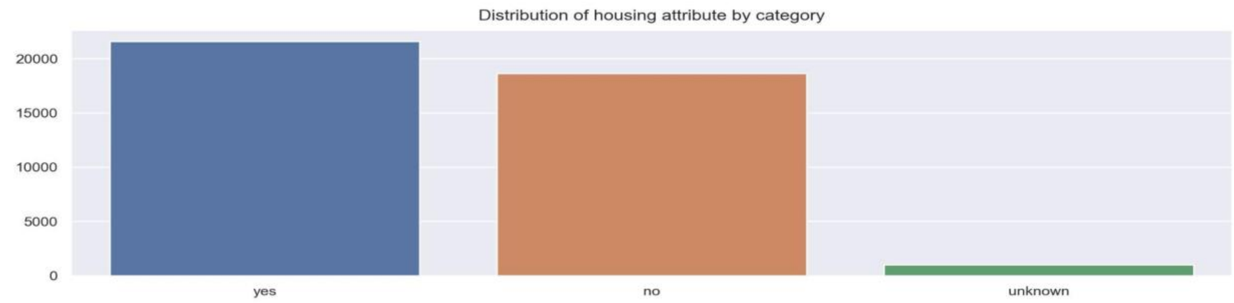
In our case, we don’t need to remove outliers from the data since the max(‘age’)=98 and max(‘campaign’)=56 are not unrealistic values. This will help with the generalization of the model later since it should reflect the real world. We notice that **97.5%** of the clients fall in the age range of **20 - 60 years old**.

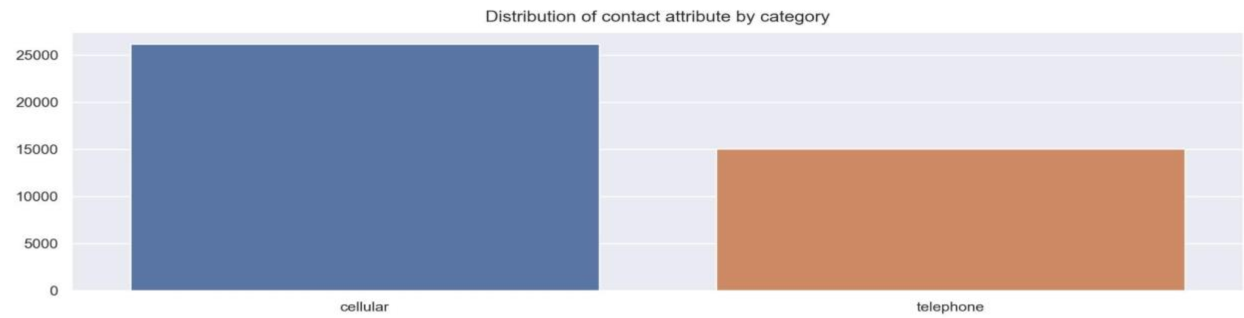
- **Categorical attributes:**

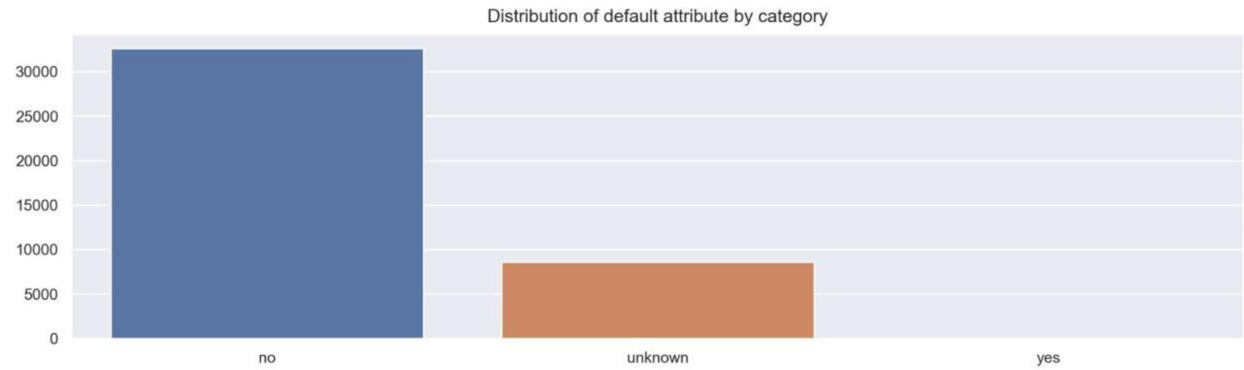
For these attributes, bar plots were created to have a better look at them.

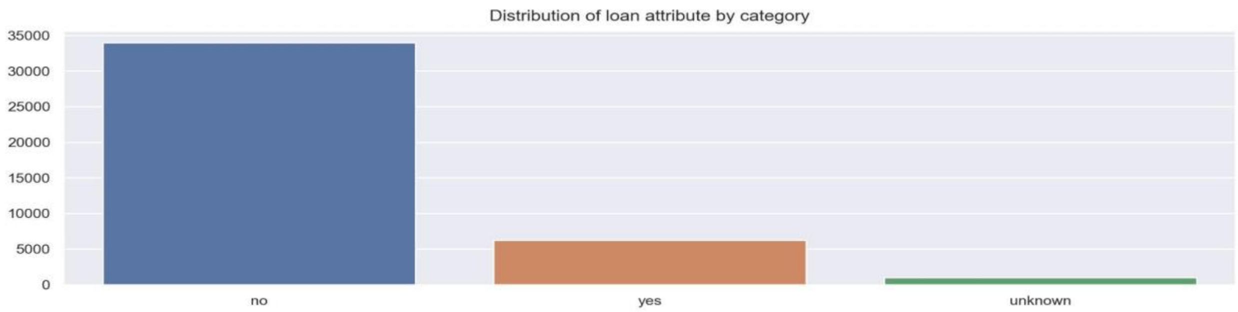
Figure 5 - Number of occurrences for each unique value of each attribute




Distribution of housing attribute by category
20000
15000
10000
5000
0
yes
no
unknown


Distribution of contact attribute by category
25000
20000
15000
10000
5000
0
cellular
telephone

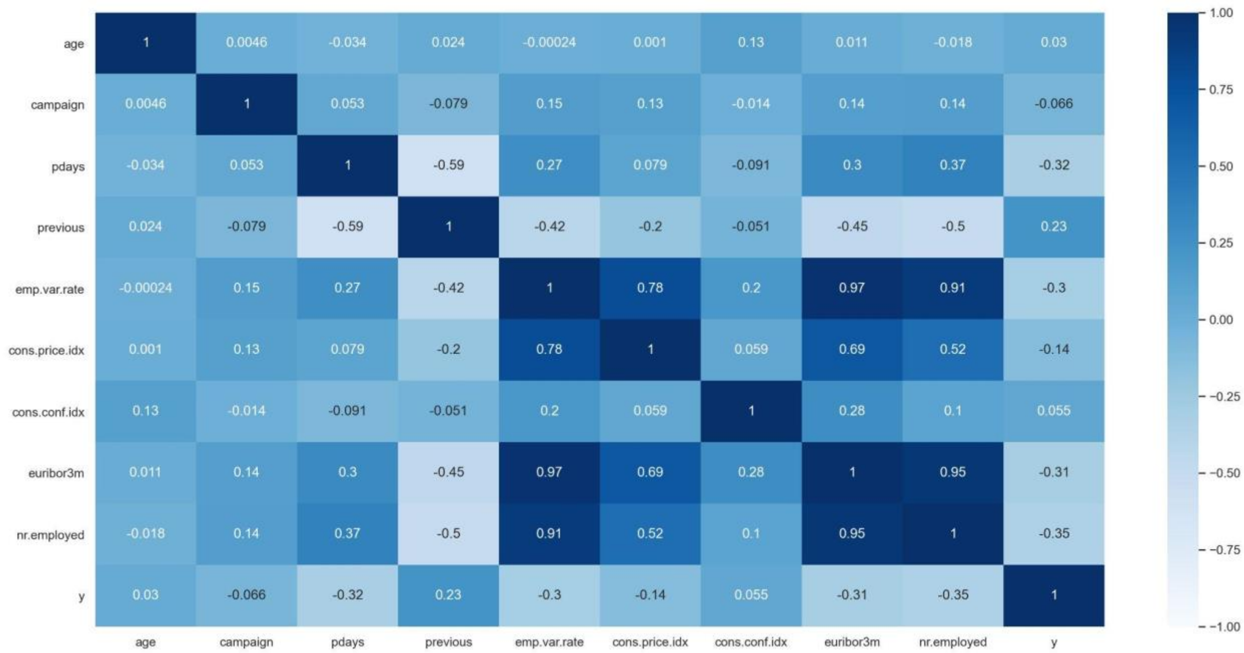

Distribution of default attribute by category
30000
25000
20000
15000
10000
5000
0
no
unknown
yes



Correlation analysis (bivariate analysis):

Correlation analysis (or bivariate analysis) examines the relationship between two attributes and determines whether the two are correlated. This analysis can be done from two perspectives for various possible combinations:

Figure 6 - Correlation matrix plot

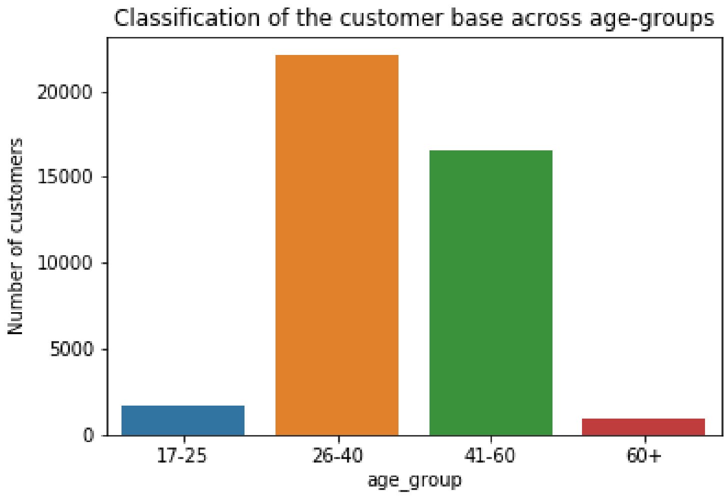


To carry on the Exploratory Data Analysis, bivariate analysis is carried out on the data.

Age-group feature:

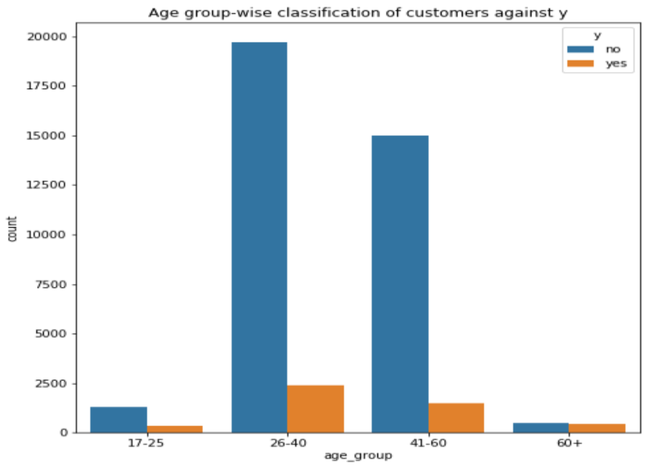
Using the 'age' column a new feature 'age_group' is created to give a more definitive analysis. The minimum and maximum values of age in this data are 17 and 98 yrs respectively. The age-groups created are as follows: 17-25, 26-40, 41-60 and 60+.

Figure 7 - Age groups



The age-groups of 26-40 and 41-60 are the dominant categories for this customer base. This new feature is analysed with the target variable 'y'. Below is the classification of the customer base based on the age-group and 'y'.

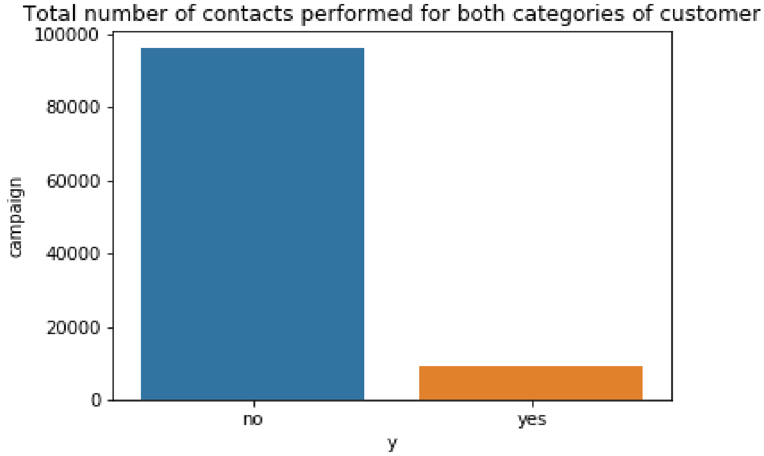
Figure 8 - Deposit per age groups



Approximately 20000 customers in the 26-40 age-group are not subscribed to the term deposit plan. It is closely followed by the 41-60 group with around 15000 people rejecting the deposit plan.

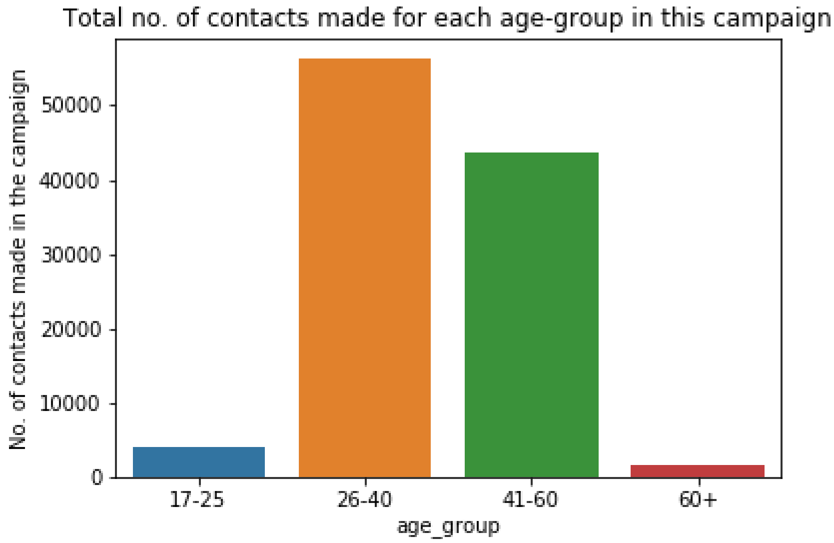
<u>Campaign feature:</u>

The ‘campaign’ feature is explored by analysing it with the target variable ‘y’.



The number of contacts made in this campaign for those not subscribing to a term deposit is almost 80000 more than those opting for this deposit scheme.

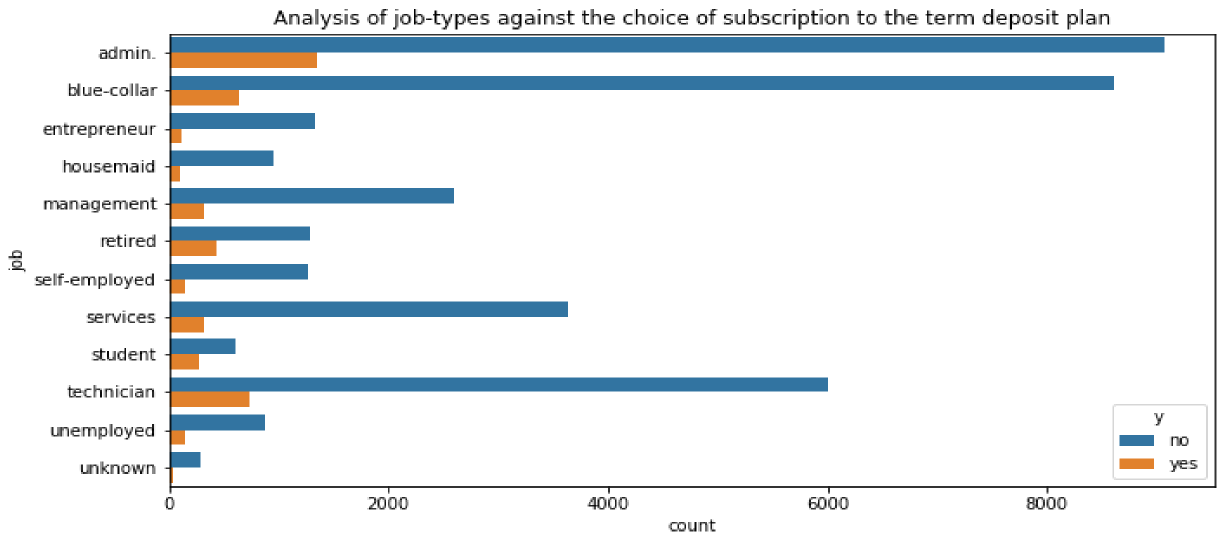
To go into more detail, the ‘age_group’ and ‘campaign’ features are considered to understand which age-group witnesses the most number of contacts.



The age-groups of 26-40 and 41-60 are targeted in this campaign with more than 90000 contacts made. As expected, these two seem to be the target groups in the marketing campaign.

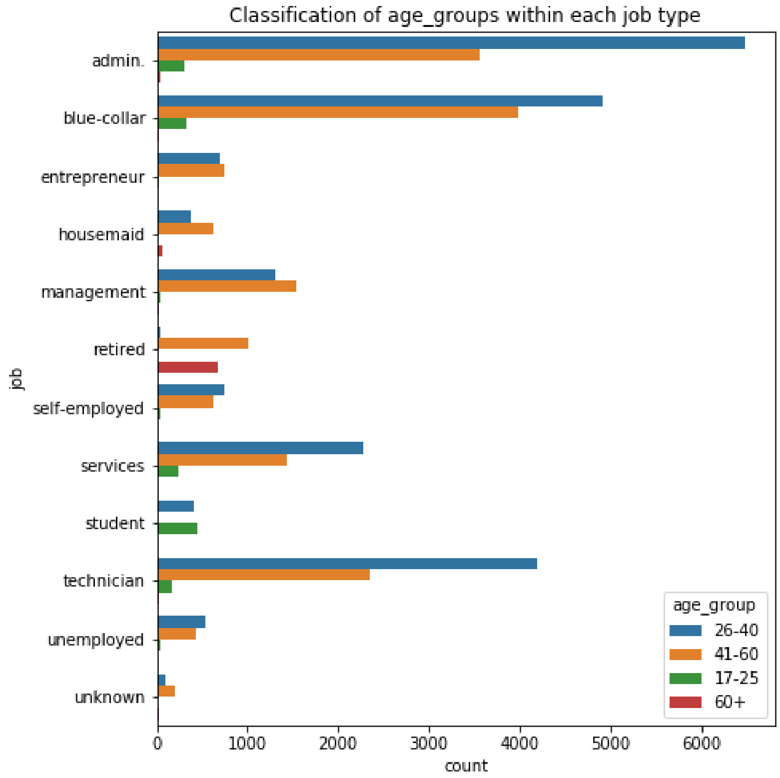
Job feature:

The job categories are analysed against the target variable.

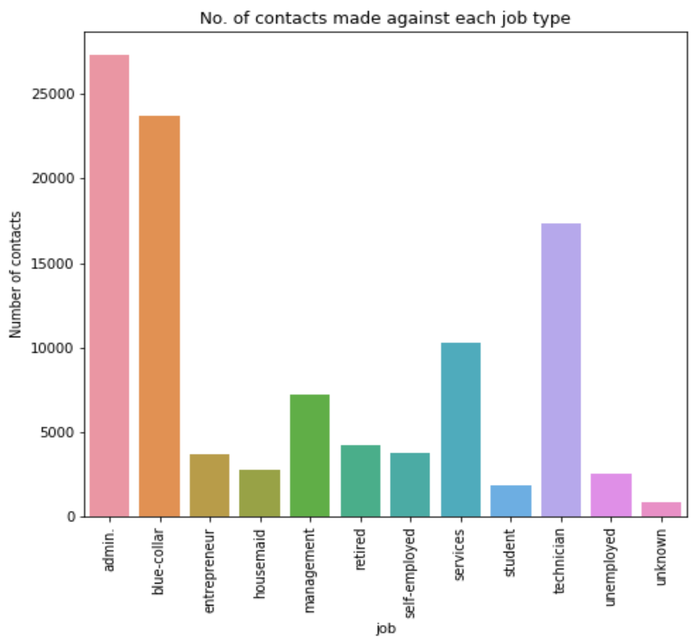


The customers in professions of ‘admin.’, ‘blue-collar’ and ‘technician’ have majorly rejected the term deposit plan. The former three jobs comprise approximately 64% among the customer base.

Another analysis is done by exploring the different job types against each age group. This will give an understanding of the customer demographic.



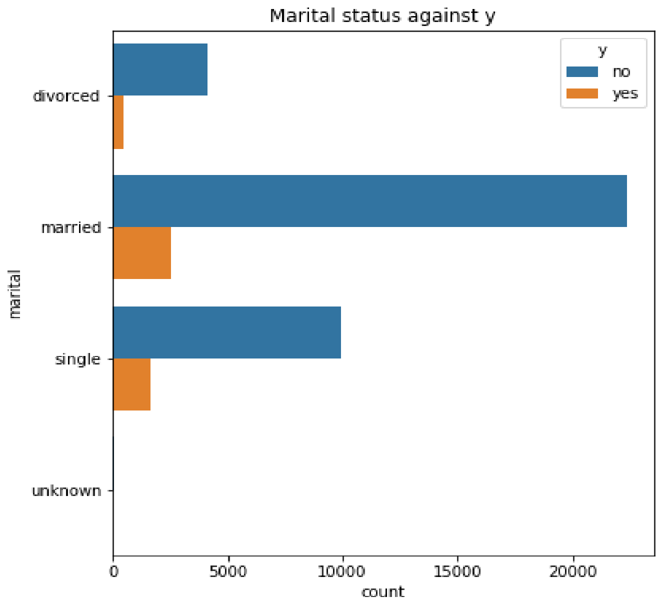
'Admin', 'blue-collar' and 'technician' professions have people primarily in the age-groups of 26-40 and 41-60 yrs. This is followed by looking at the 'campaign' feature against 'job'. This gives an idea if a particular job profession has an effect on the number of contacts the marketing team makes.



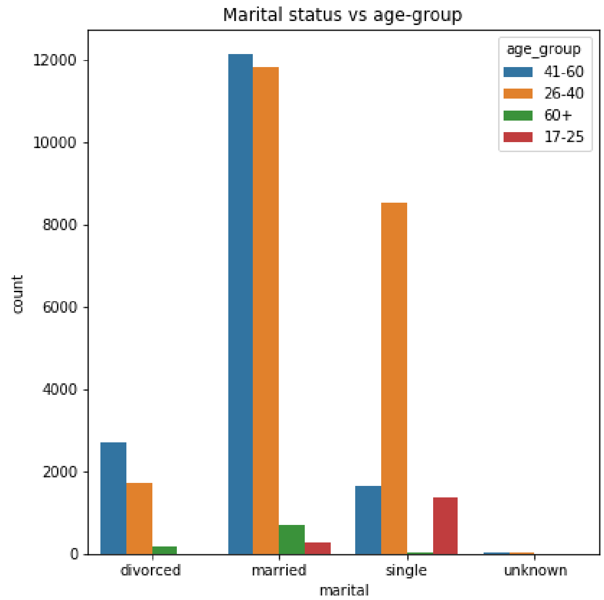
Over 25000 contacts are performed for those in admin jobs in this current campaign. The next two in line are blue-collar and technician job types with the latter receiving close to 20000 contacts from the marketing team.

Marital feature:

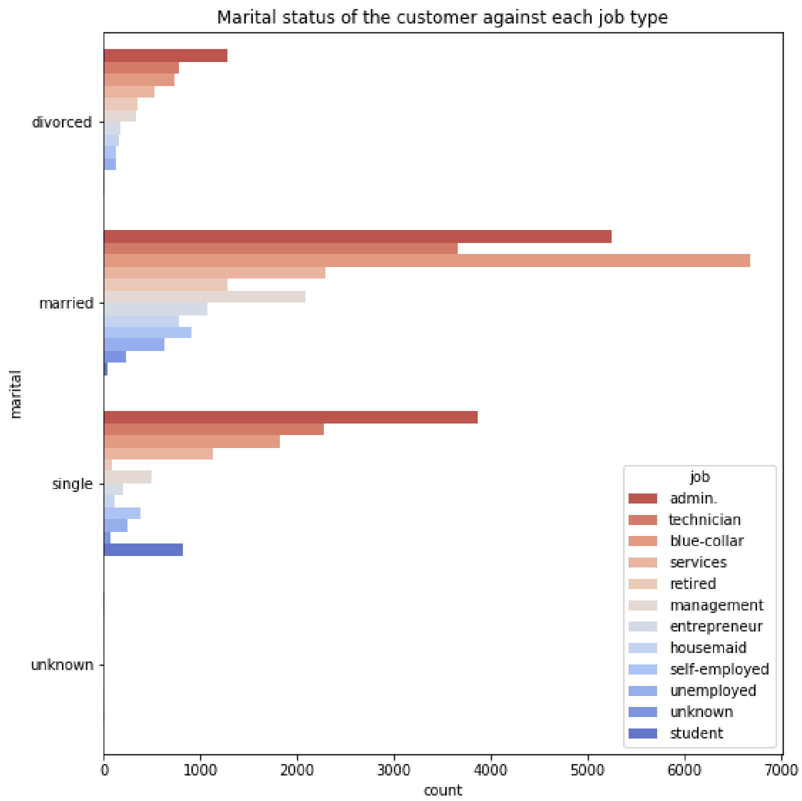
A univariate analysis of this variable shows approximately 25000 customers, which is more than 50%, having a married status. This feature is explored together with the variable defining whether the customer has subscribed or not to the term deposit.



The majority of the married and single customers have shown a rejection to the term deposit plan. This is followed by looking at the ‘age_group’ together with the marital status.

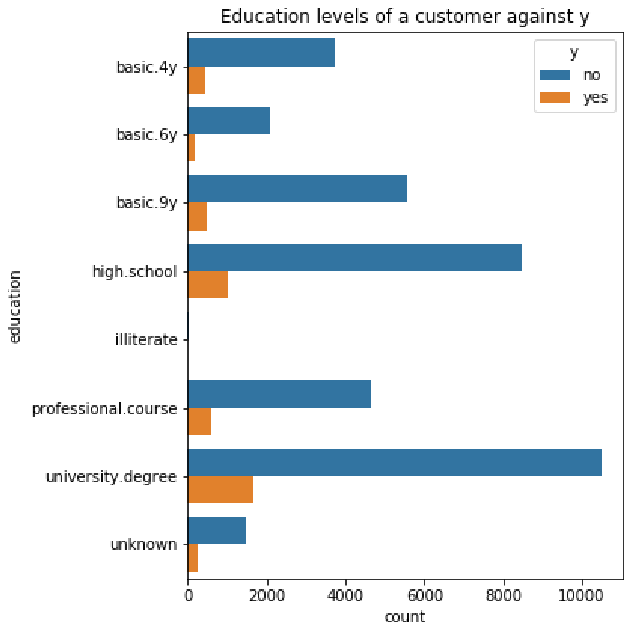


Among the married customers, there is almost an equal proportion between the 26-40 and 41-60 age groups. In addition to this, the customers in professions of admin., blue-collar and technician are around 16000 (approx.) who are married. Single customers are close to 4000 in administrative jobs.

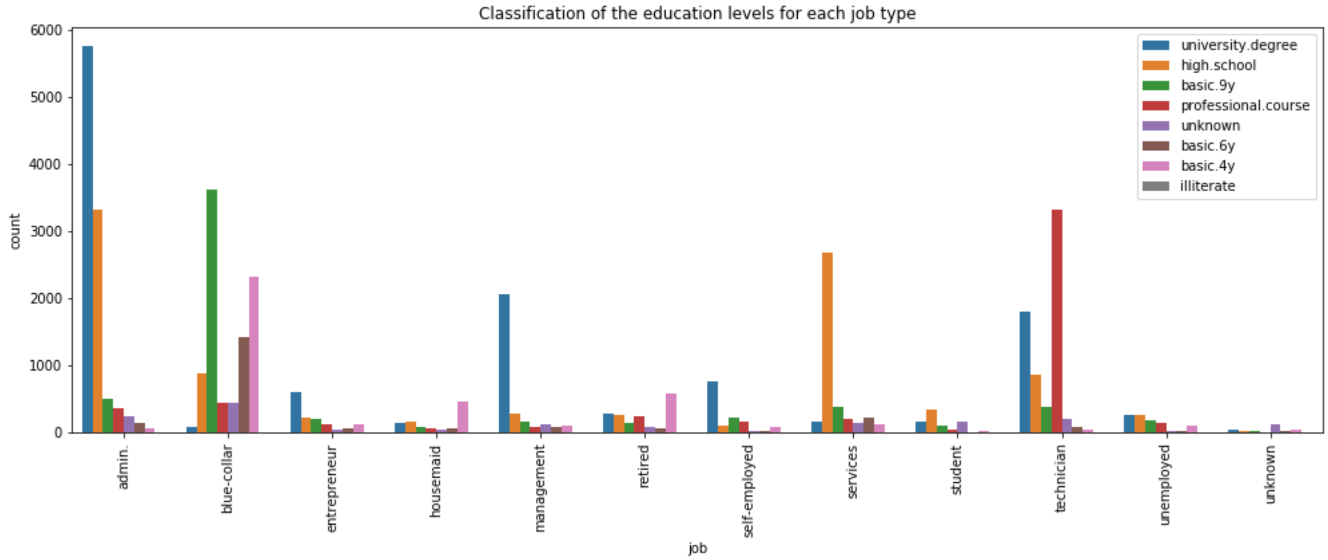


Education feature:

From the univariate analysis, there are around 12000 and 10000 customers whose education levels are a University degree or High school respectively. This feature is analysed together with ‘y’.



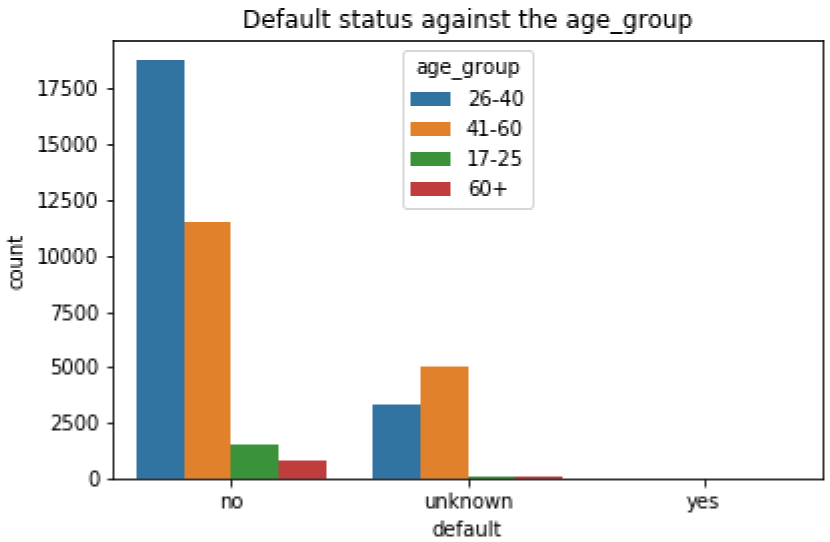
There are nearly 2000 customers with an unknown education level. With all the levels, there are a majority of customers rejecting the term deposit plan. To get a more deeper understanding of the variable 'education', it is explored with the 'job' feature.



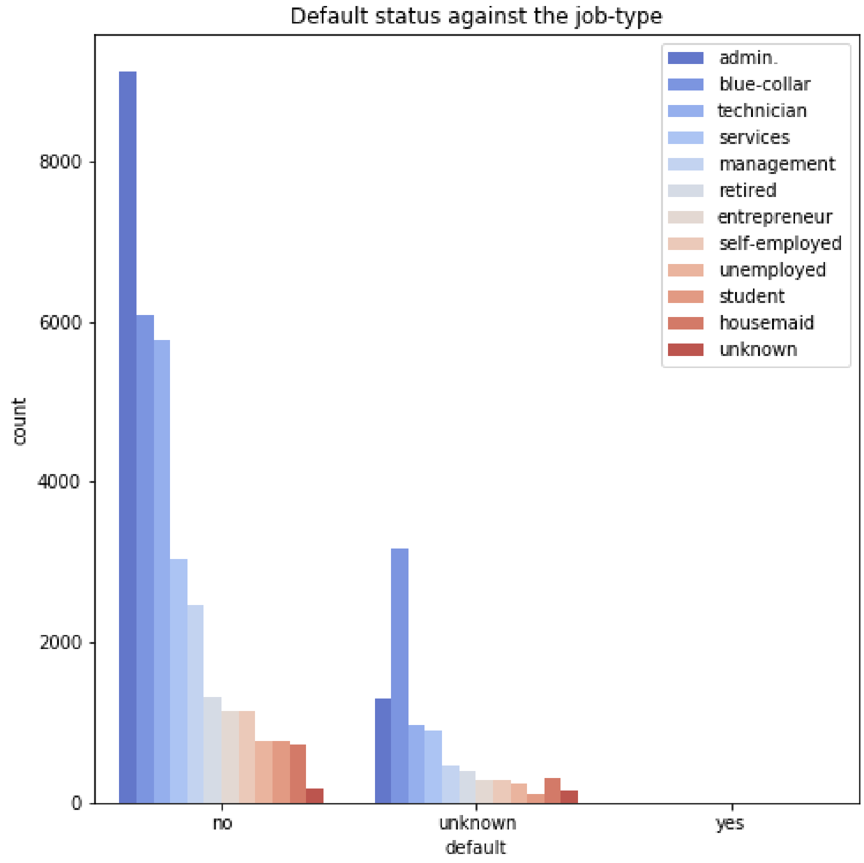
A blue-collar professional requires at least a basic.6y education as observed from the above data, with over 3000 customers achieving that level. In a management type role, a University degree appears to have more importance. In comparison to an admin. job, there are close to 8500 customers with a high school or University degree as the education attained. Finally, a role of a technician requires some sort of professional course training and the evidence of this in the plot above.

Default feature:

Out of the 41117 customers, 75% of them do not have a credit in default. On the other hand, there are zero customers who have defaulted. But, around 8000 people have an unknown default status on their name which sounds strange.



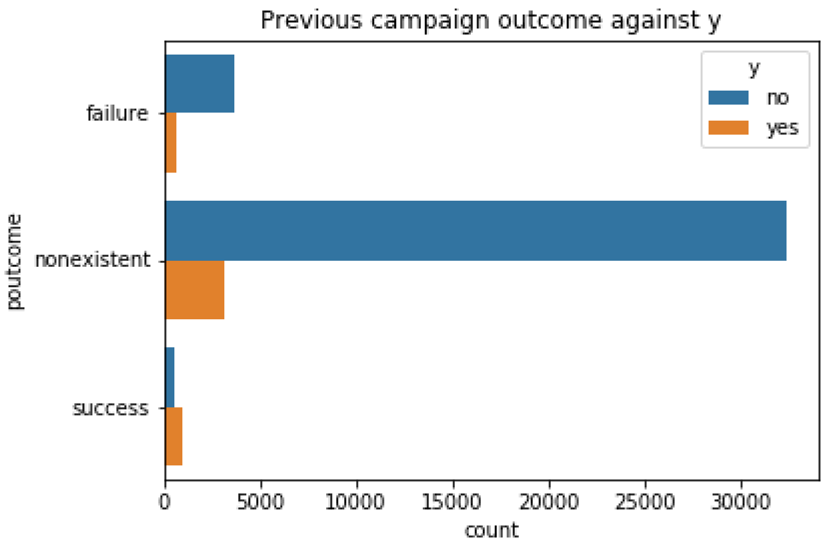
A deeper understanding of the customer base is done by looking at the default status across the different age-groups. There are approximately 28000 customers from the 26-40 and 41-60 age-groups who are clean in their default status. The numbers for the other two groups are lower as expected since they might not be financially secure enough to acquire a loan. Across various job categories too we look at the default status.



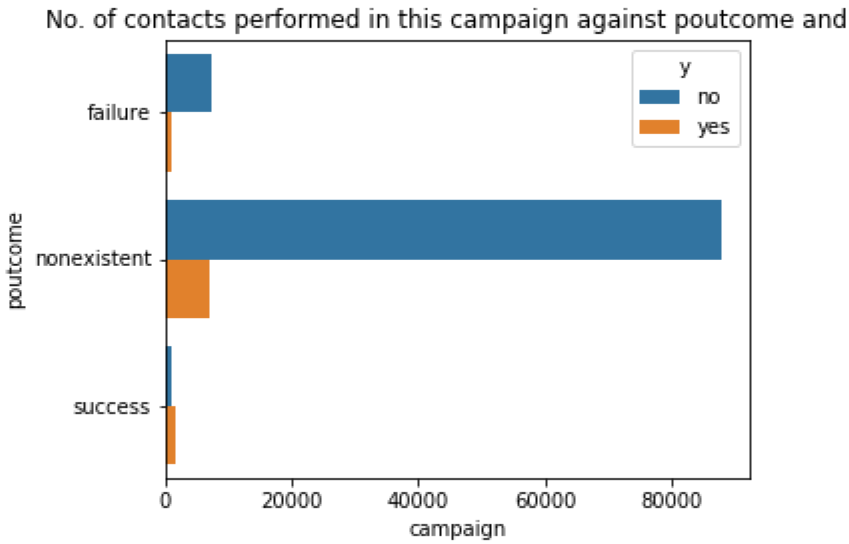
The professions of admin., blue-collar and technician lead the other roles in terms of having a no default. In addition, around 5000 customers in blue-collar jobs have an unknown status.

<u>Poutcome feature:</u>

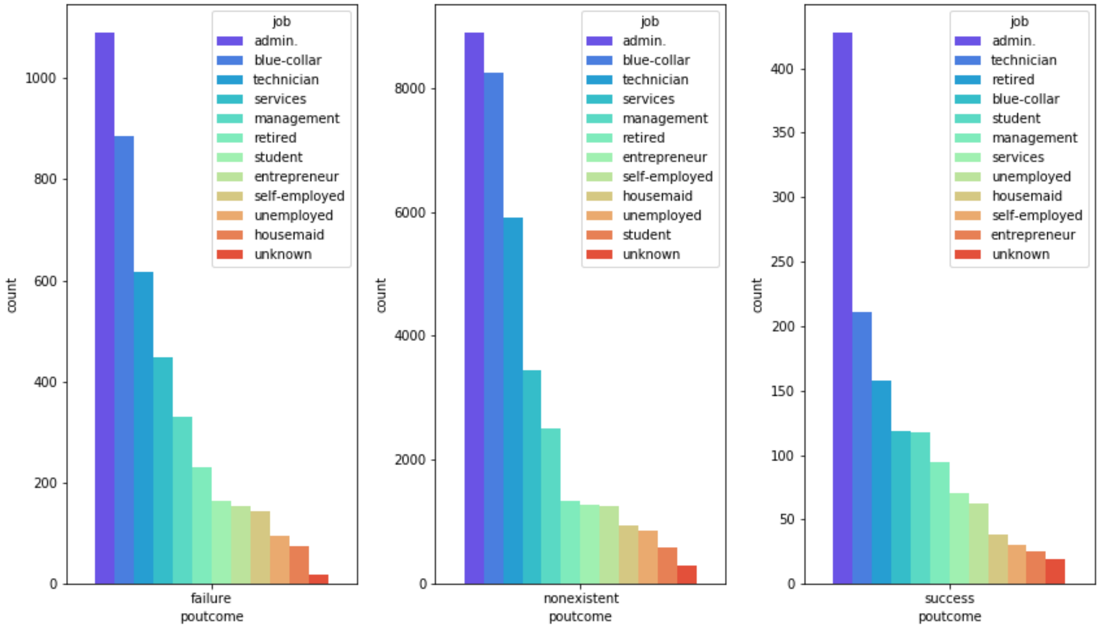
The outcome of the previous marketing campaign will surely affect the decision of the customers on whether to opt or not for the term deposit scheme offered by the bank. The results of the previous campaign show that the outcome was non-existent for 35000 customers. On the other hand, the numbers for the failure and success of the previous campaign are close to 12% of the total pool of customers.



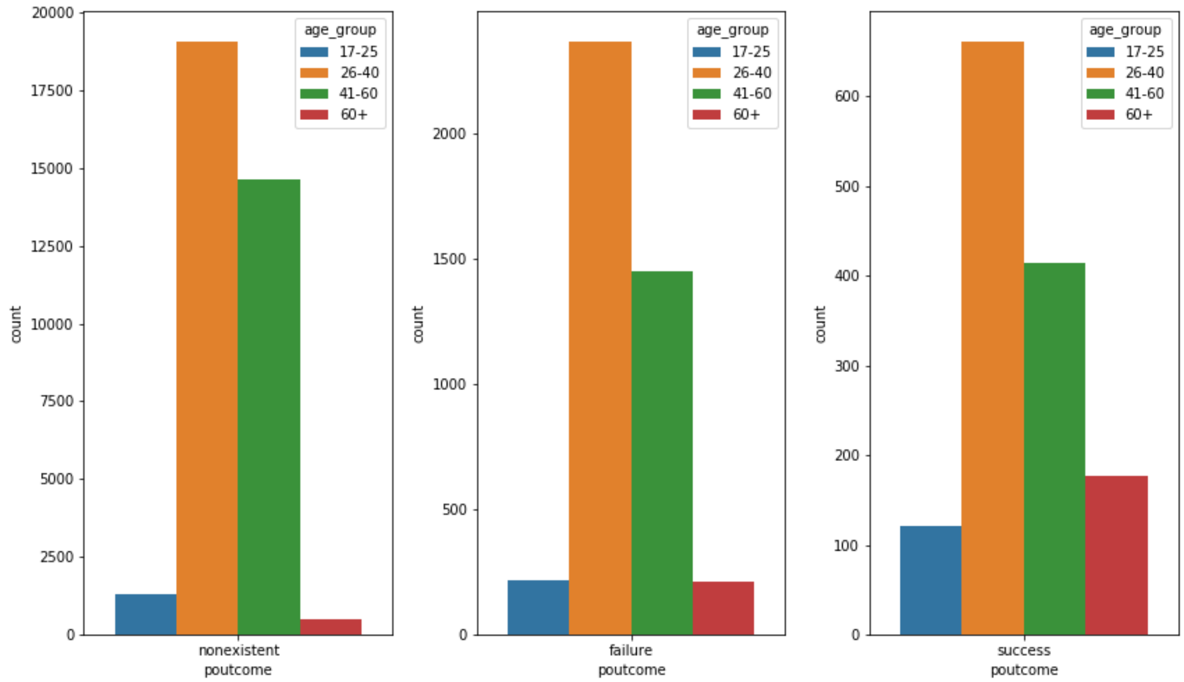
Greater than 90% of customers for whom the outcome was non-existent have rejected the term deposit plan. The effort of the marketing team can be understood by looking at the number of contacts made during this campaign. Below plot, shows that more than 80000 contacts are performed for customers who have not subscribed to the term deposit plan. This dwarfs the numbers for whom the 'poutcome' was a failure or success.



The ‘job’ feature is also analysed together with the ‘poutcome’ variable. For all the non-existent outcomes there are over 14000 customers in the roles of administrative, blue-collar and technician. For the outcome of success, customers in administrative jobs are almost double in number to the next job category.



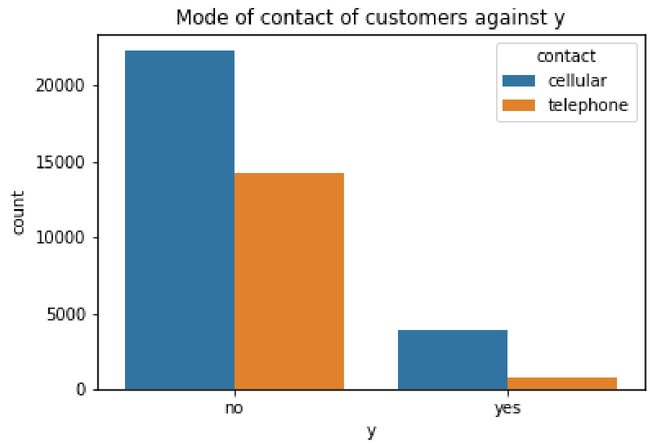
To get another perspective of the analysis of 'poutcome', the age-groups are considered here.



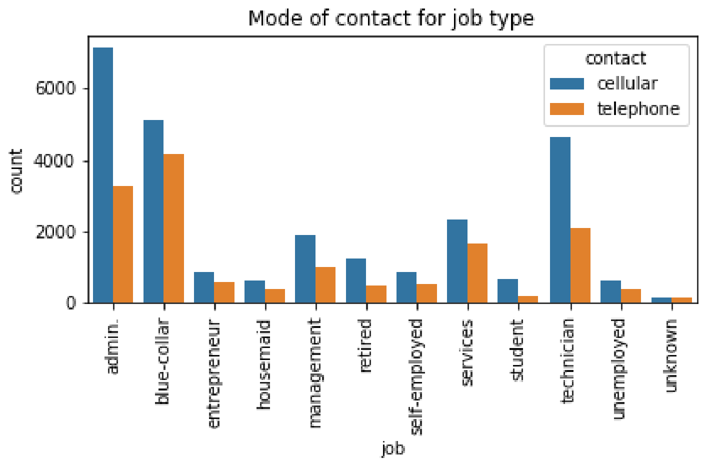
The previous campaign had been a success for over 1000 customers in 26-40 and 41-60 age-groups. For the other age-groups, the number of customers is very small in number for all the outcomes. Close to 35000 customers have a non-existent outcome associated with them for the previous campaign.

Contact feature:

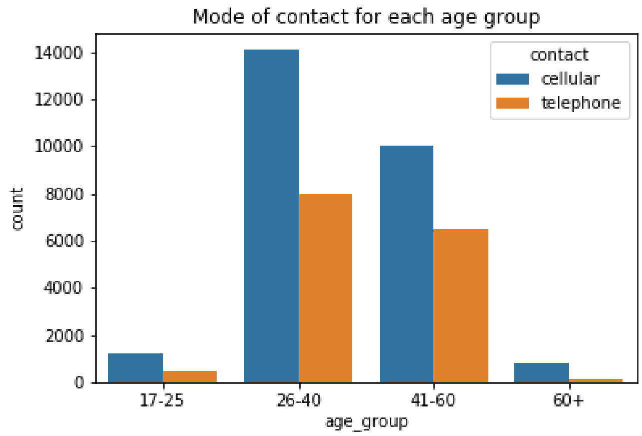
The impact of the mode of communication by the marketing team is also studied along with other features. Looking at the numbers tell that approximately 25000 customers have been contacted through a cellular mode in this campaign, with the rest through telephone. The mode of communication is analysed against the customer choice of subscription to the term deposit scheme.



Following this, the relation between the mode of communication to the customer and the various job categories is also looked at.



The cellular mode is done for approximately 17000 customers who are in admin., blue-collar and technician job roles. This is almost double the number for telephonic mode of communication for the same three professions. An age-groupwise analysis is also done to get more understanding of this feature.

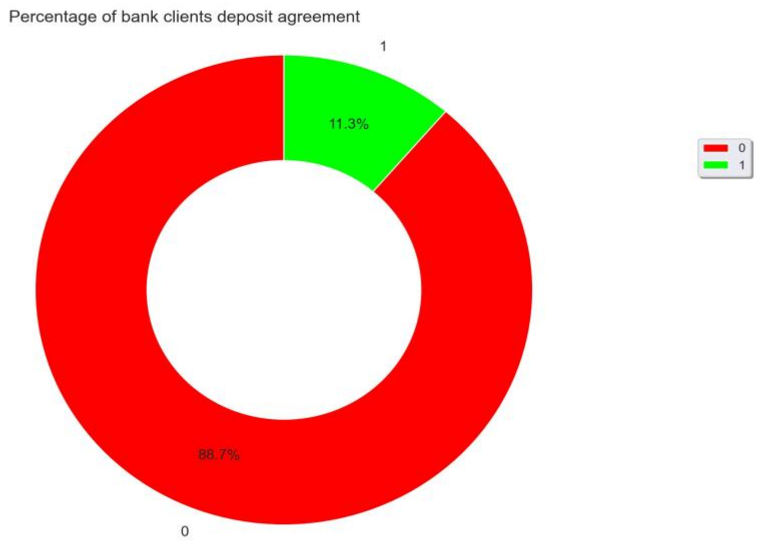


The cellular mode is used for approximately 24000 customers in the age-groups 26-40 and 41-60 yrs. For both these groups there is not a big gap in the number of customers between the two modes of communication.

Class Distribution: 'y' feature:

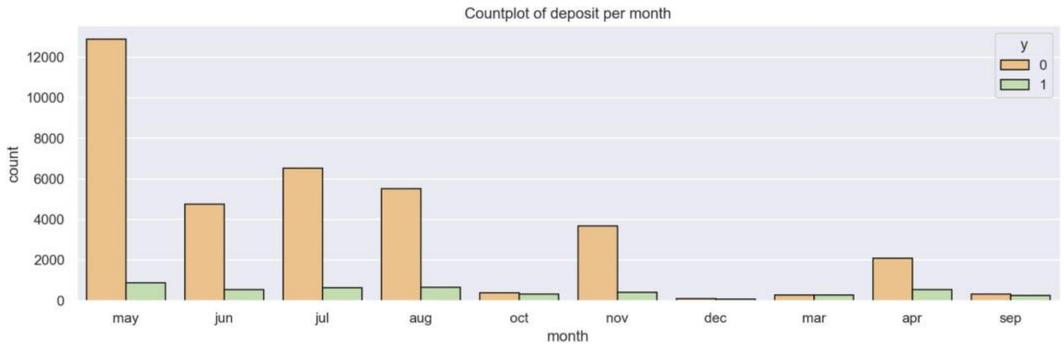
The percentage of bank clients who agree to open term deposits is low. Out of the total calls that are made, 11.3 % of them are successful.

Figure 9 - Percentage of Deposit



These successful calls were mostly done during the month of May as shown in the following figure:

Figure 10 - countplot of deposit per month



Final recommendation

1. Take into account the time of the company (May is the most effective).
2. Increase the time of contact with customers. It is possible to use other means of communication.
3. Focus on specific categories. Students and senior citizens respond better to this proposal.
4. It is imperative to form target groups based on sociological-economic categories. Age, income level (not always high), profession can accurately determine the marketing profile of a potential client.
5. Looking at the customer base, the age-groups of 26-40 and 41-60 have a higher proportion. These groups present a profitable target for the marketing team.
6. In the target groups, focus on Admin., Blue-collar, Technician, Services and Management professions.

➔ Given these factors, it is recommended to concentrate on those consumer groups that are potentially more promising.

Github Repo link:

https://github.com/AsAmira02/Bank-Marketing-Campaign-DSEnthusiasts2021

This repository includes the four datasets, model code and necessary files used in this project.