

WEEK 10 : DELIVERABLES

BANK MARKETING CAMPAIGN

‘DATA SCIENCE’

GROUP NAME: DATA SCIENCE MASTER
NAME : ABHIMANYU GANGANI
EMAIL : Agangani97@gmail.com
COUNTRY : UNITED KINGDOM
COLLEGE : ANGLIA RUSKIN UNIVERSITY
SPECIALIZATION : DATA SCIENCE

EDA :

For better understanding of the dataset we will proceed with appropriate exploratory data analysis of the dataset in depth. This will help us to get clear image of the data structure, shape and characteristics.

- **Dataset Description:**

Data columns (total 21 columns):

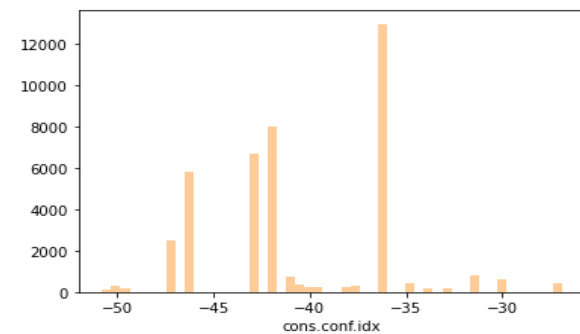
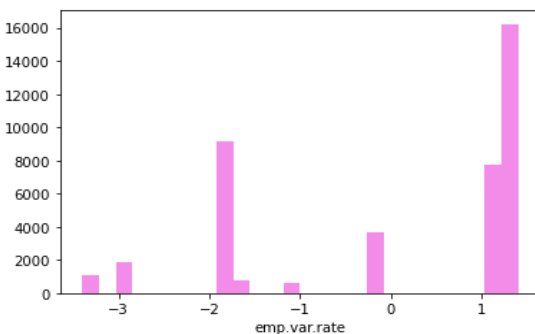
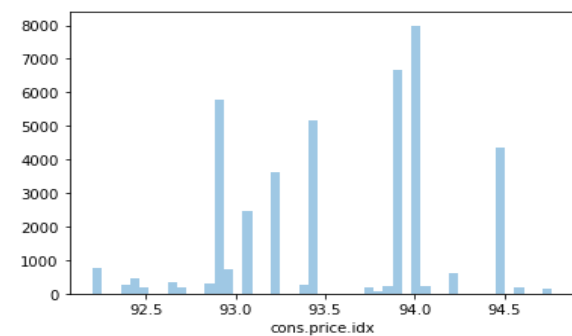
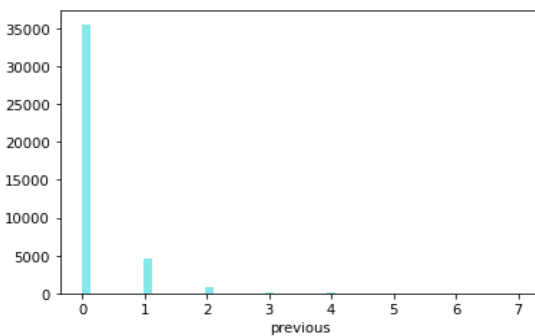
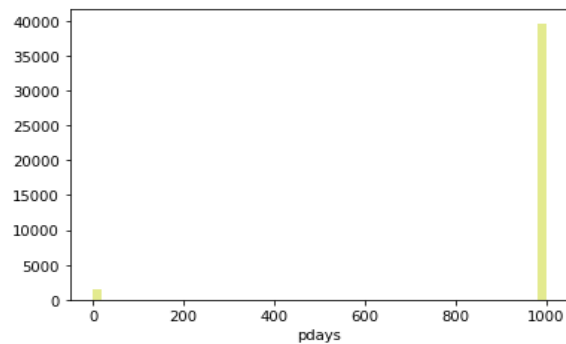
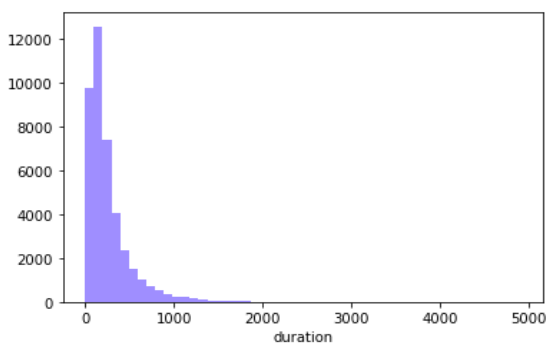
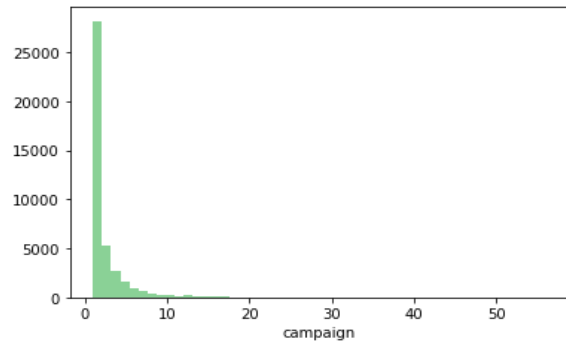
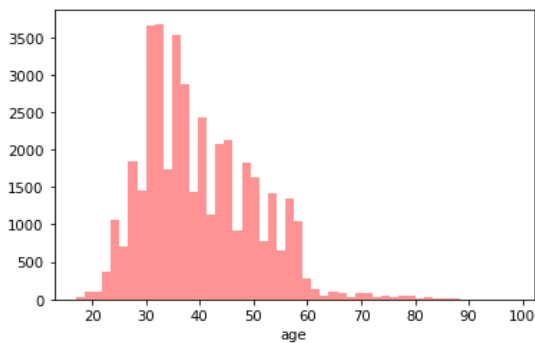
#	Column	Dtype	Description
---	-----	-----	-----
0	age	int64	Age of Client.
1	job	object	Type of Job.
2	marital	object	Marital Status.
3	education	object	Level of Education.
4	default	object	Has credit in default?
5	housing	object	Has housing loan?
6	loan	object	Has personal loan?
7	contact	object	How client has been communicated?
8	month	object	last contacted month.
9	day_of_week	object	last contacted day.
10	duration	int64	duration of communication(seconds).
11	campaign	int64	number of contacts performed in Campaign.
12	pdays	int64	number of days passed after contact.
13	previous	int64	number of total contacts performed.
14	poutcome	object	outcome of the previous campaign.
15	emp.var.rate	float64	Employment variation rate.
16	cons.price.idx	float64	Consumer price index.
17	cons.conf.idx	float64	Consumer confidence index.
18	euribor3m	float64	Euribor 3 months rate.
19	nr.employed	float64	number of employees.
20	y	object	has the client subscribed product.

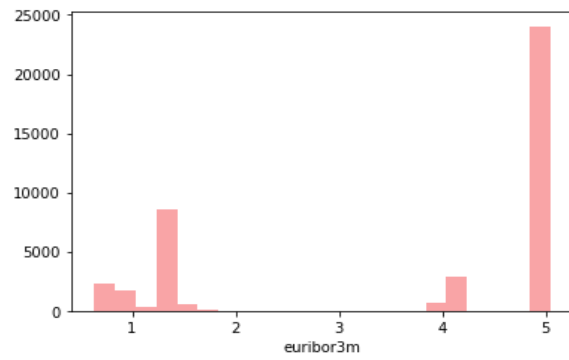
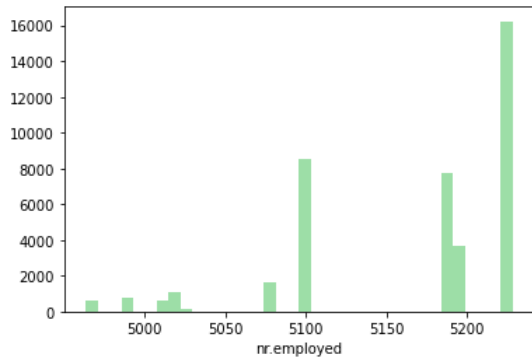
dtypes: float64(5), int64(5), object(11)

Descriptive Analysis (univariate):

- Numerical Attributes:

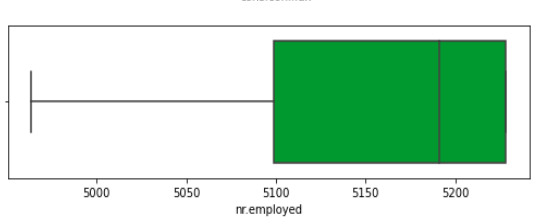
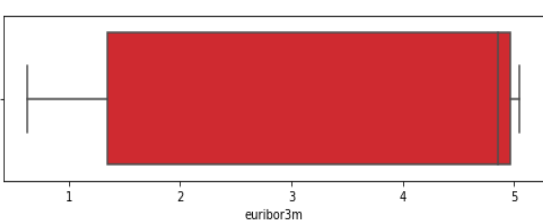
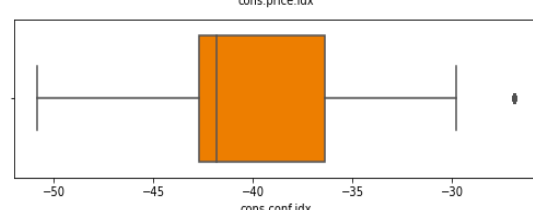
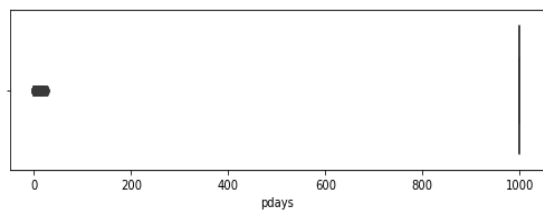
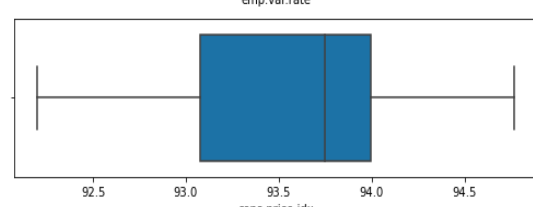
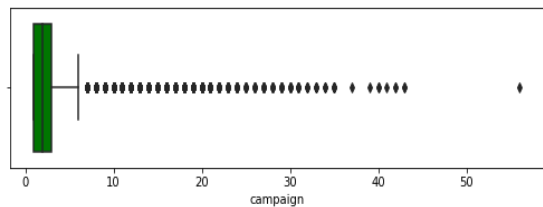
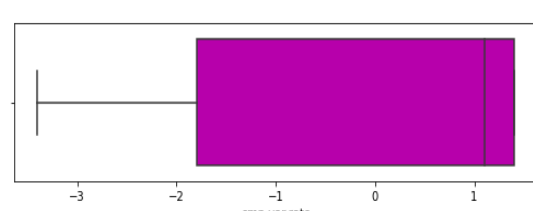
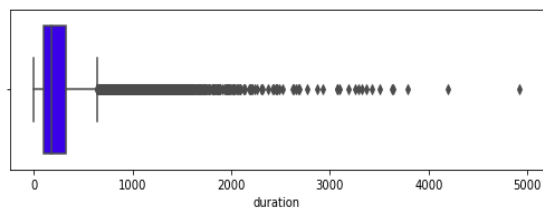
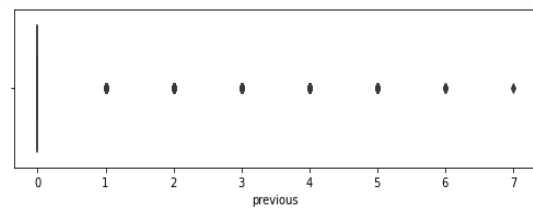
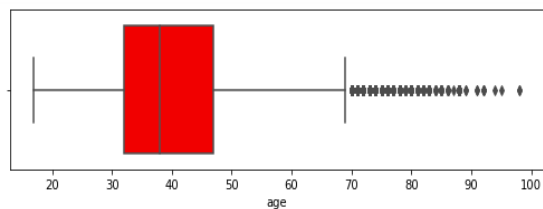
Histograms are plot for numerical variables in order to check what kind of distribution attributes follows and this will help us to detect outliers as well.





Box plots were also made to check if any outliers present in the data.

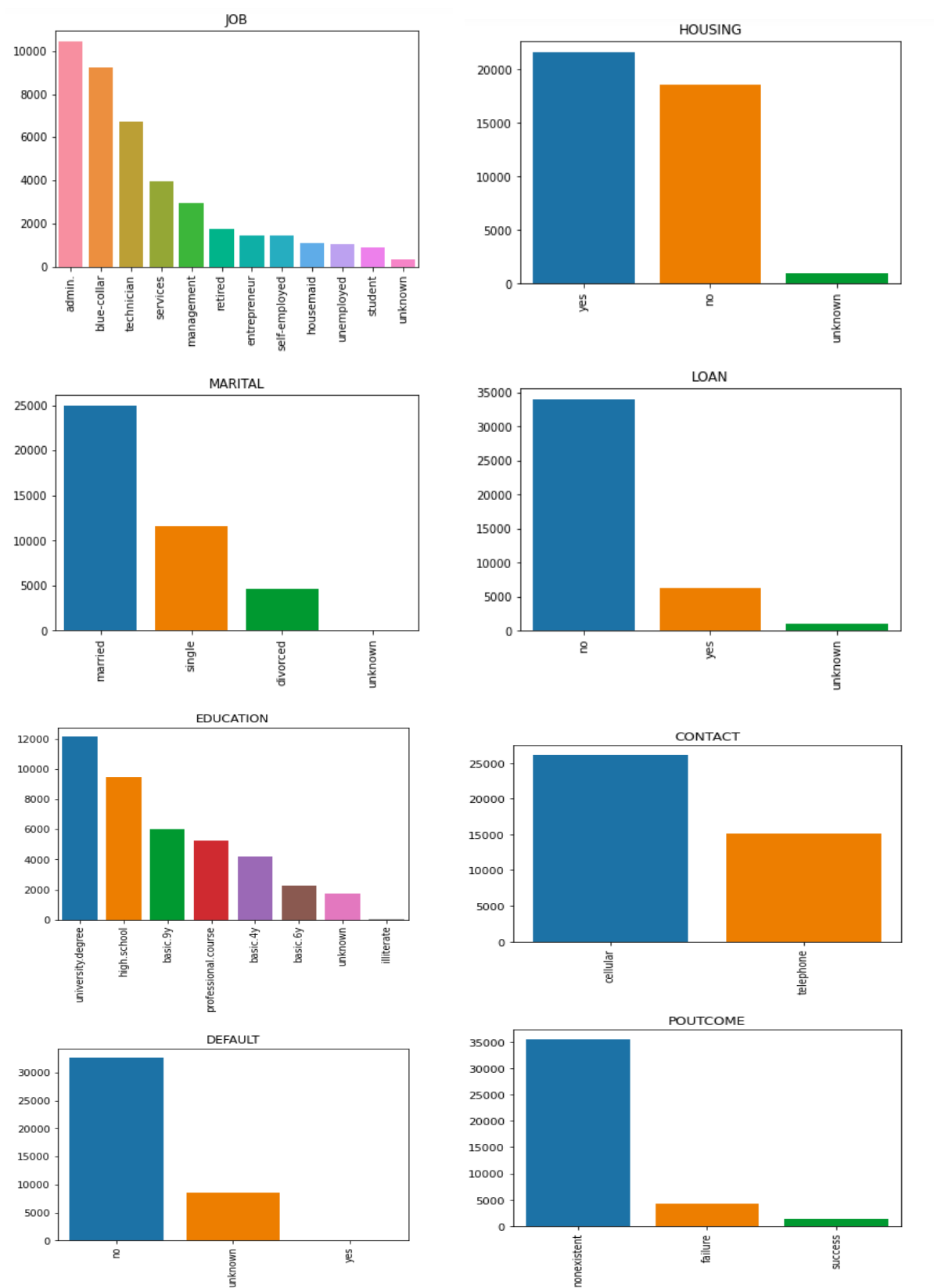
Outliers are the values which lie at above 3 standard deviation distance from the other Values in normal distribution. It might occur due to improper collection of the data. . Outliers can disturb our analysis by changing the mean on normal distribution graph. following variables consists of significant outliers.

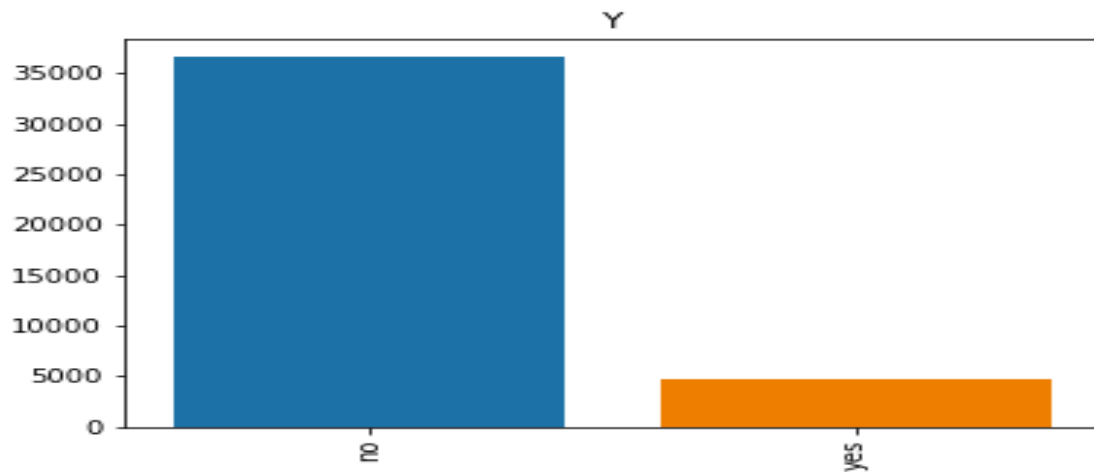


The maximum value for 'age' variable is 98 and that of 'campaign' variable is 56 and both are significant values. Since model is needed to be generalized to reflect the real world data we are not going to remove these outliers as these seems to be realistic value.

- Categorical Attributes:

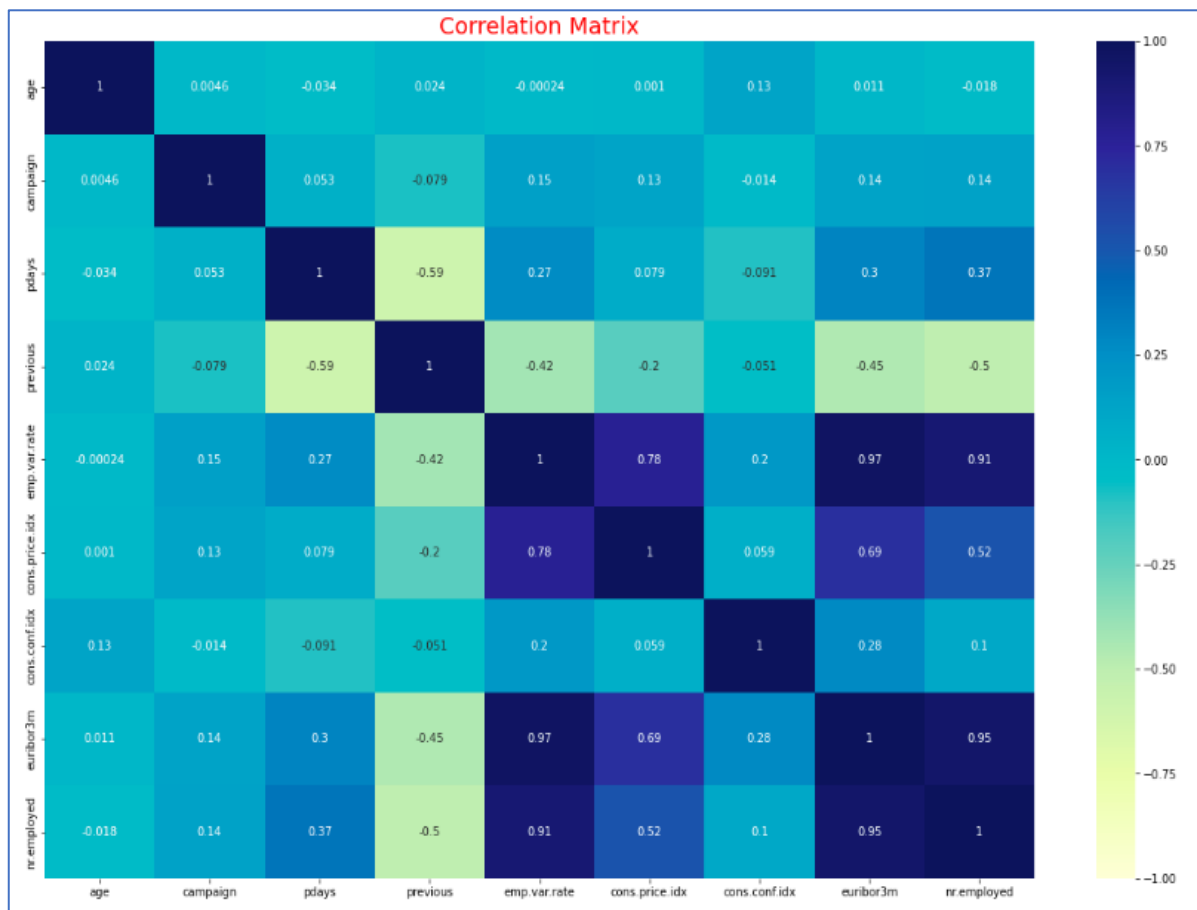
For categorical variables Bar plots were plotted to get better understanding.





Correlation Analysis (Bivariate):

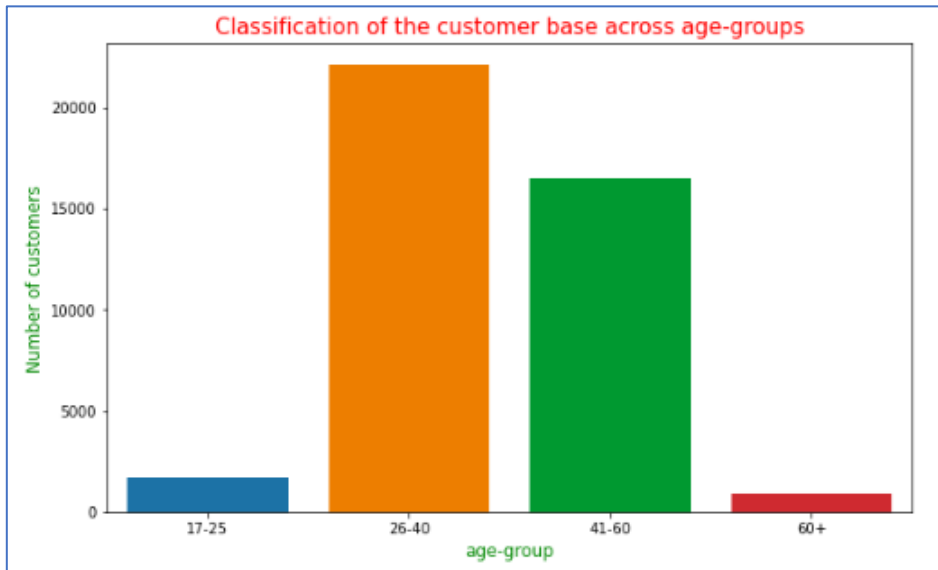
Correlation analysis (or bivariate analysis) examines the relationship between two attributes and determines whether the two are correlated. This analysis can be done from two perspectives for various possible combinations:



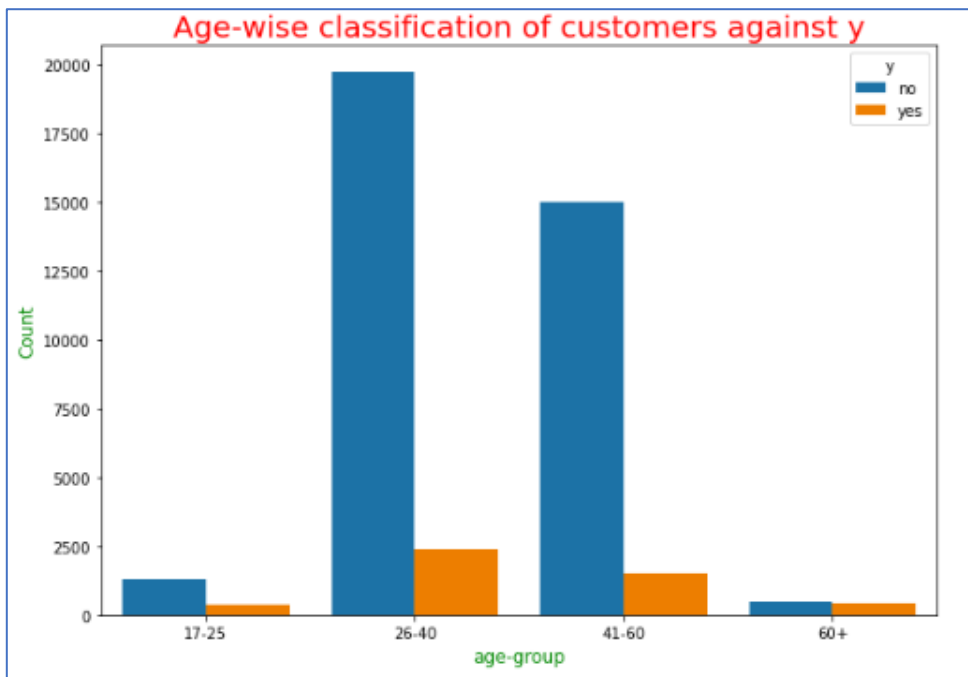
- Age-Group Feature :

Using binning we have created new feature age_group using attribute age to get better visualization.

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+.



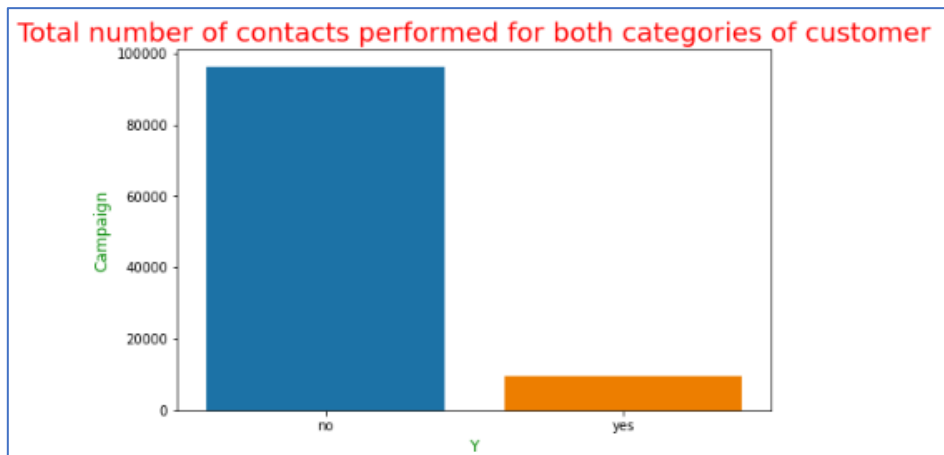
In the given customer's dataset age group 26-40 and 41-60 has highest number of customer base. This age-group variable is plotted with target variable 'Y'. Below is the bifurcation of the customer base based on the age-group and 'y'.



Around 19,000 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 nearly 15,000 people rejecting the deposit plan.

- Campaign Feature :

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



It can be concluded from the graph that 90,000 people have not opted for term deposit plan and 10,000 have subscribed the service. The conversion rate through campaign is near to 10%.

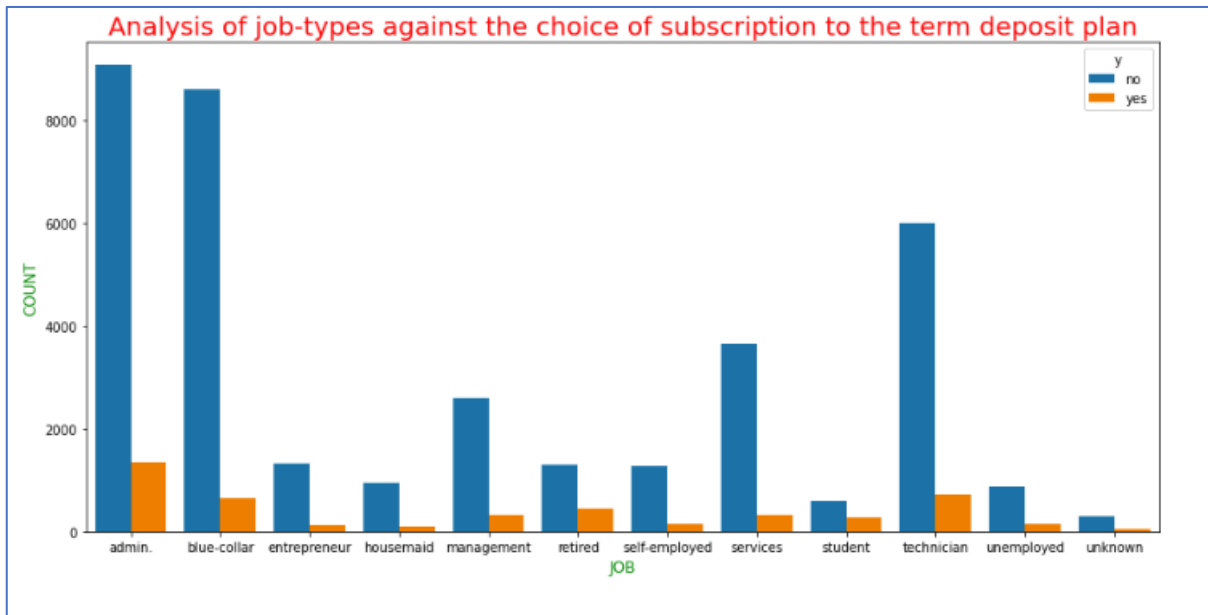
To go into more detail, the 'age_group' and 'campaign' features are considered to understand which age-group witnesses the most number of contacts.



The age-groups of 26-40 and 41-60 are targeted in this campaign with more than 90,000 contacts made. As expected, these two seem to be the target groups in the marketing campaign.

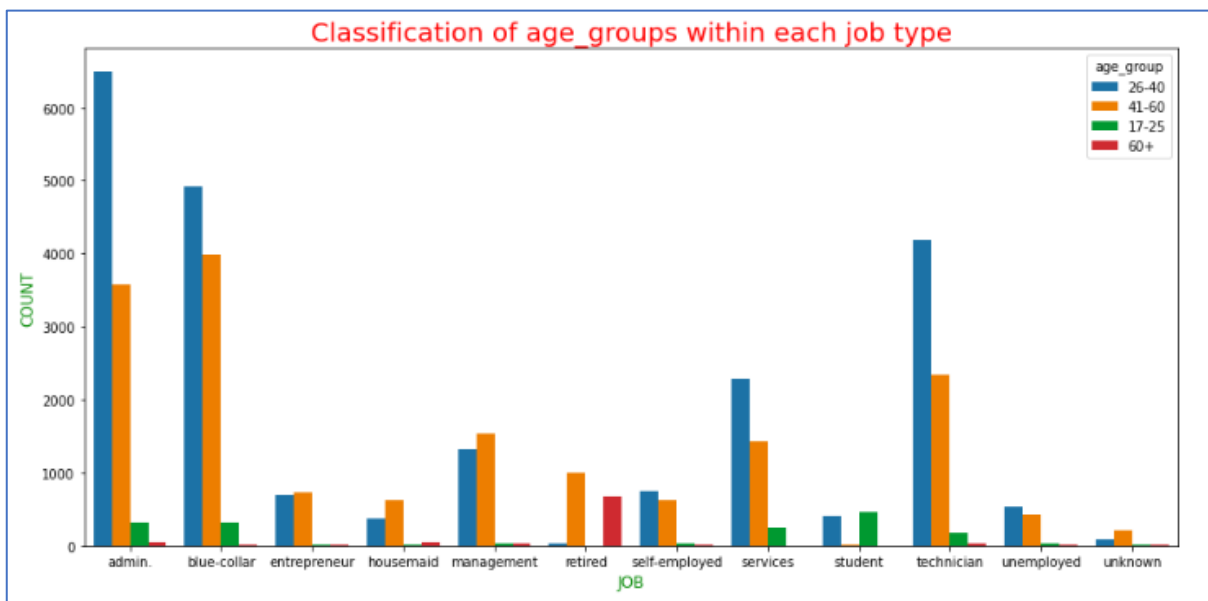
- JOB feature :

The job categories are analysed against the target variable.

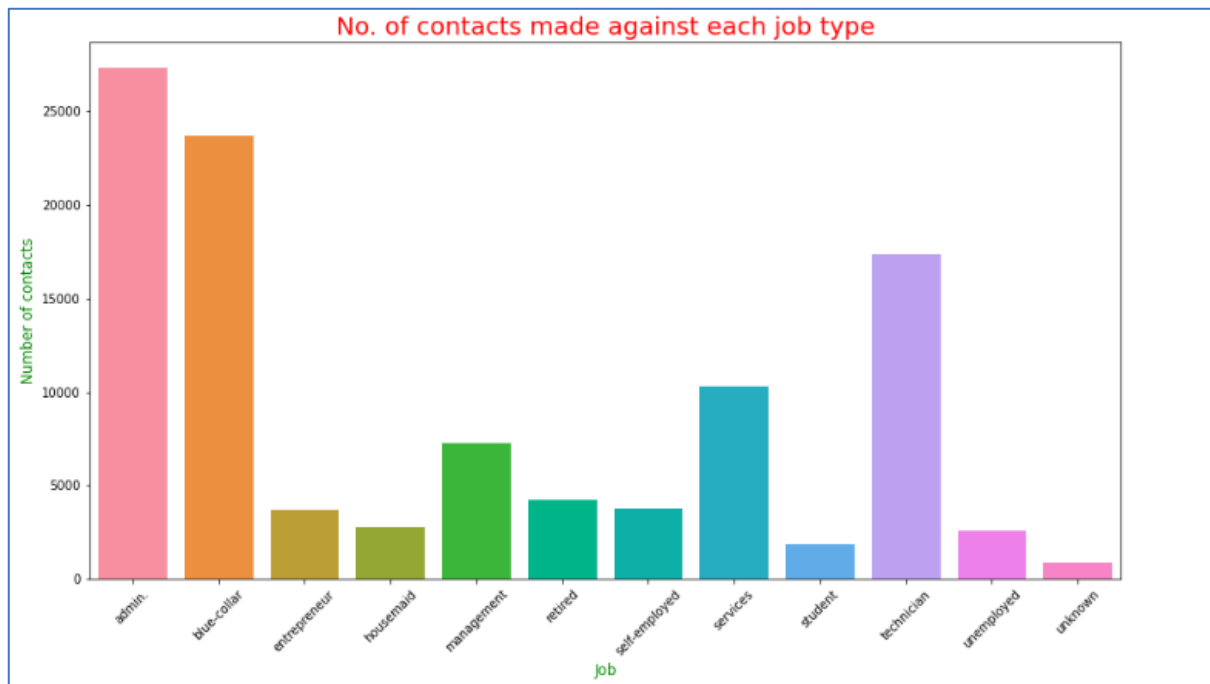


The customers in professions of 'admin.', 'blue-collar' and 'technician' have majorly rejected the term deposit plan. The former three jobs comprise approximately 64% among the customer base.

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

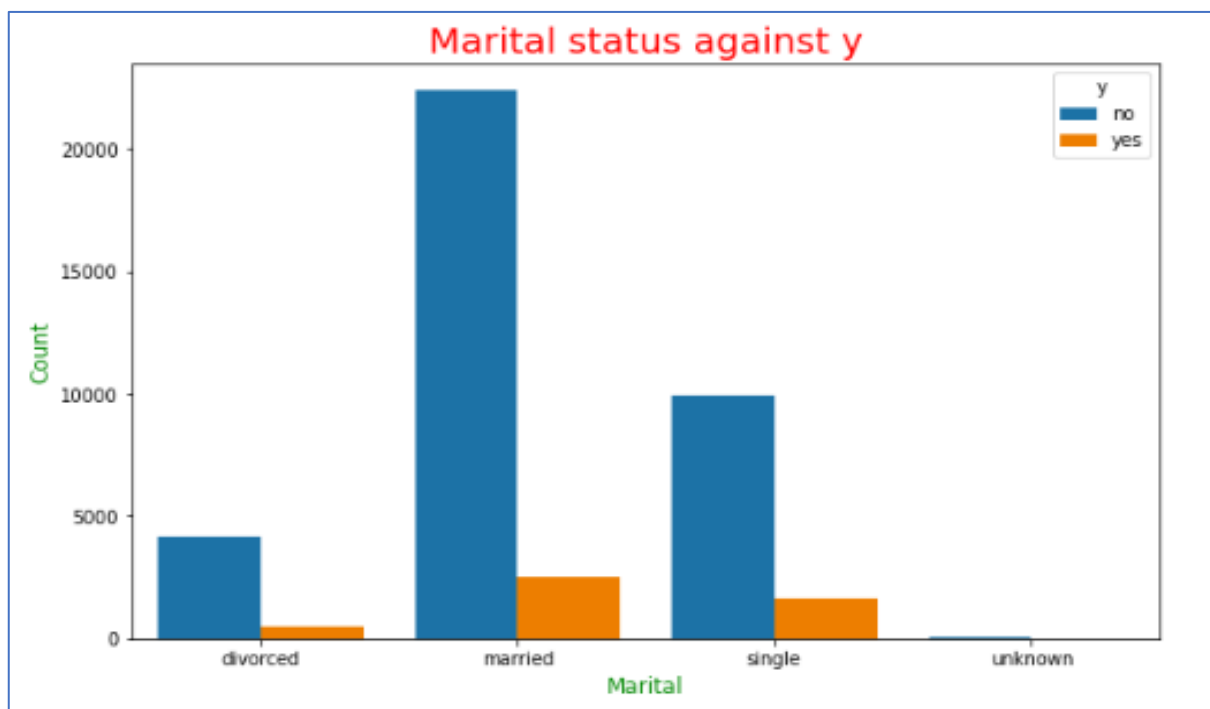


‘Admin’, ‘blue-collar’ and ‘technician’ professions have people primarily in the age-groups of 26-40 and 41-60 yrs. This is followed by looking at the ‘campaign’ feature against ‘job’. This gives an idea if a particular job profession has an effect on the number of contacts the marketing team makes.

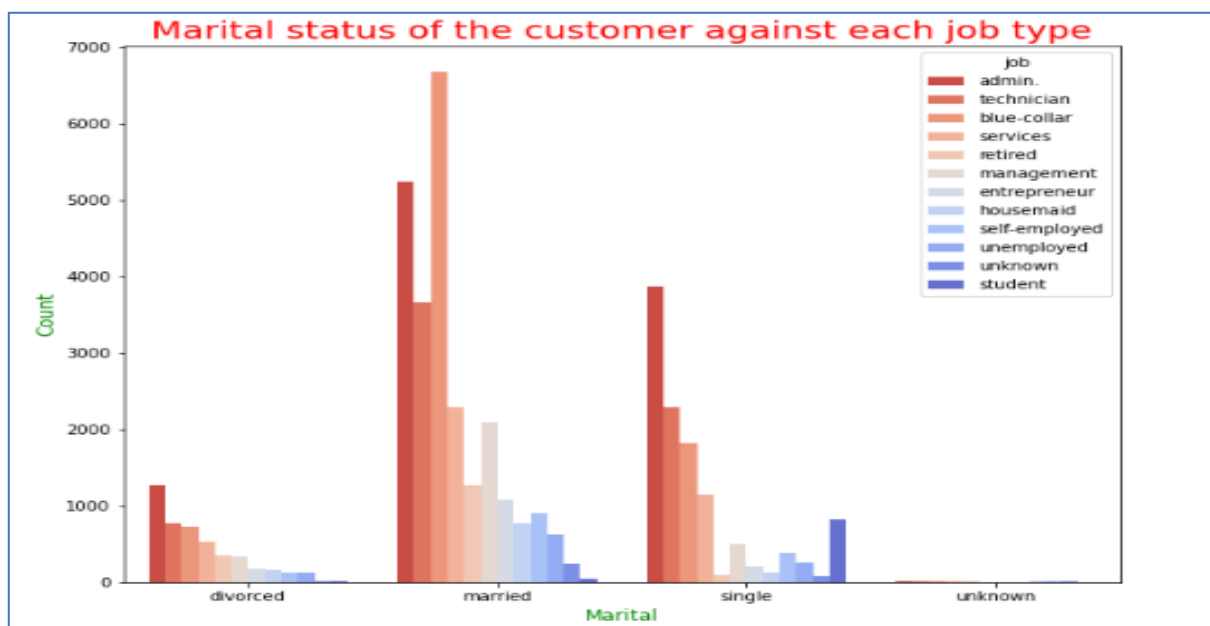
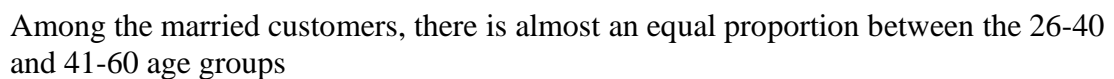


Nearly 28000 contacts are performed for those in admin jobs in this current campaign. The next two with most contacts are blue-collar and technician job types with the latter receiving approximately 20000 contacts from the marketing team.

- Marital Feature :

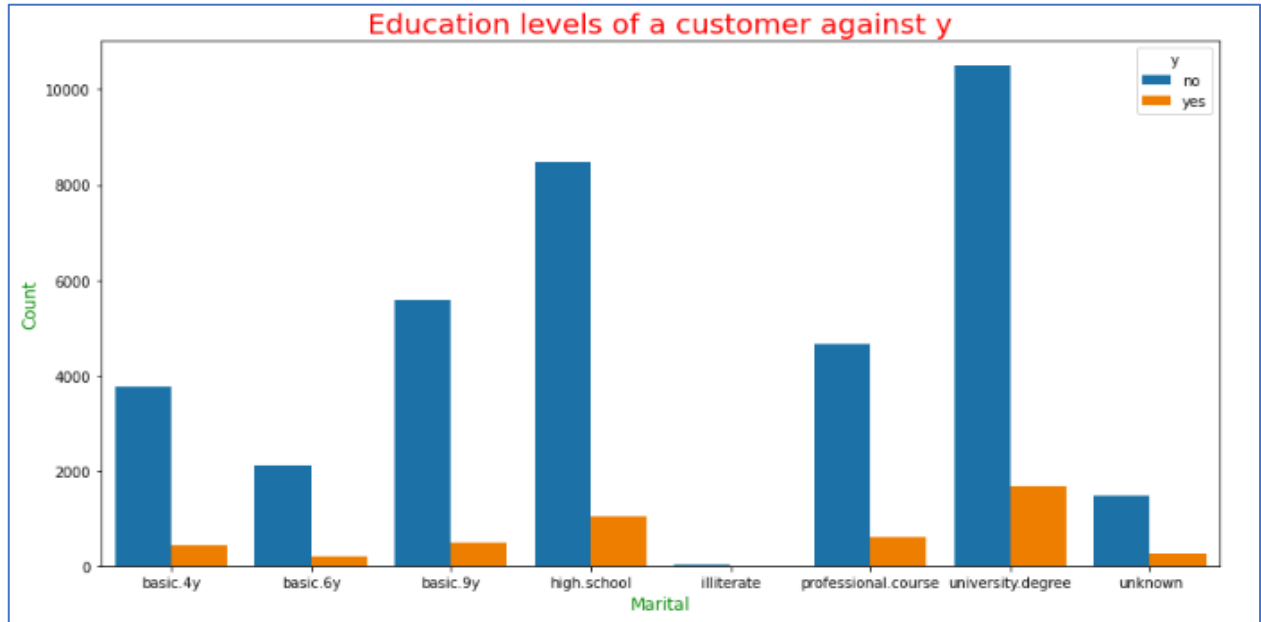


The majority of the married and single customers have shown a rejection to the term deposit plan. This is followed by looking at the 'age_group' together with the marital status.

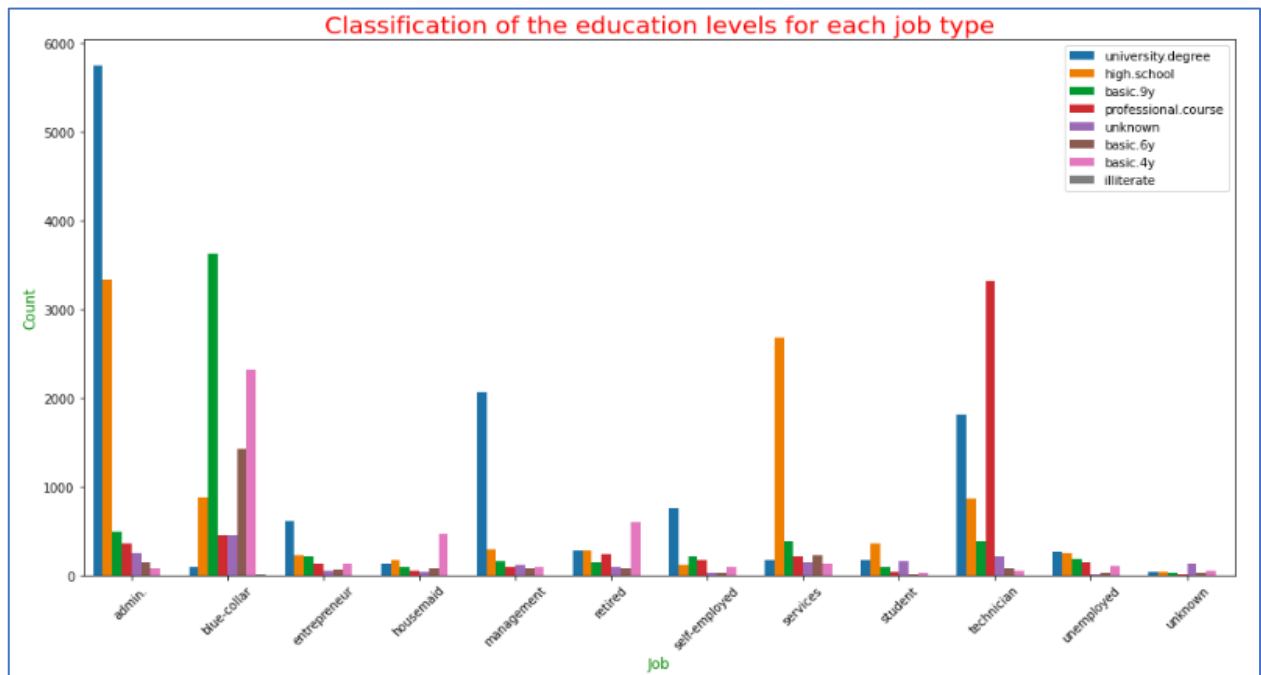


The customers in professions of admin, blue-collar and technician are around 16000 (approx.) who are married. Single customers are close to 4000 in administrative jobs.

- Education Feature :



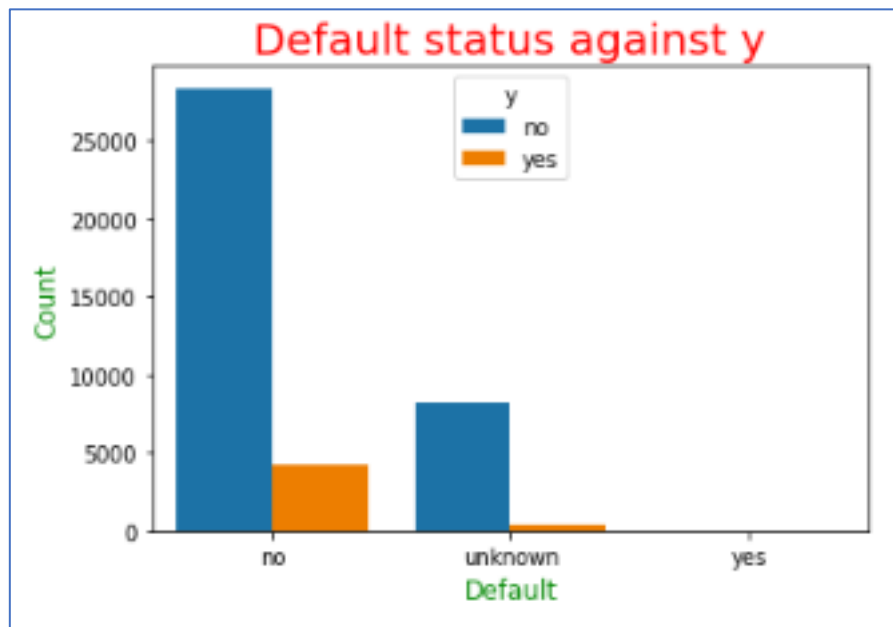
It can be observed that around 12000 clients have achieved university degree and approx. 8000 clients have completed high school. We further concluded that almost 2000 customers are having unknown education. With all the levels, there are a majority of customers rejecting the term deposit plan. To get a more deeper understanding of the variable 'education', it is explored with the 'job' feature.



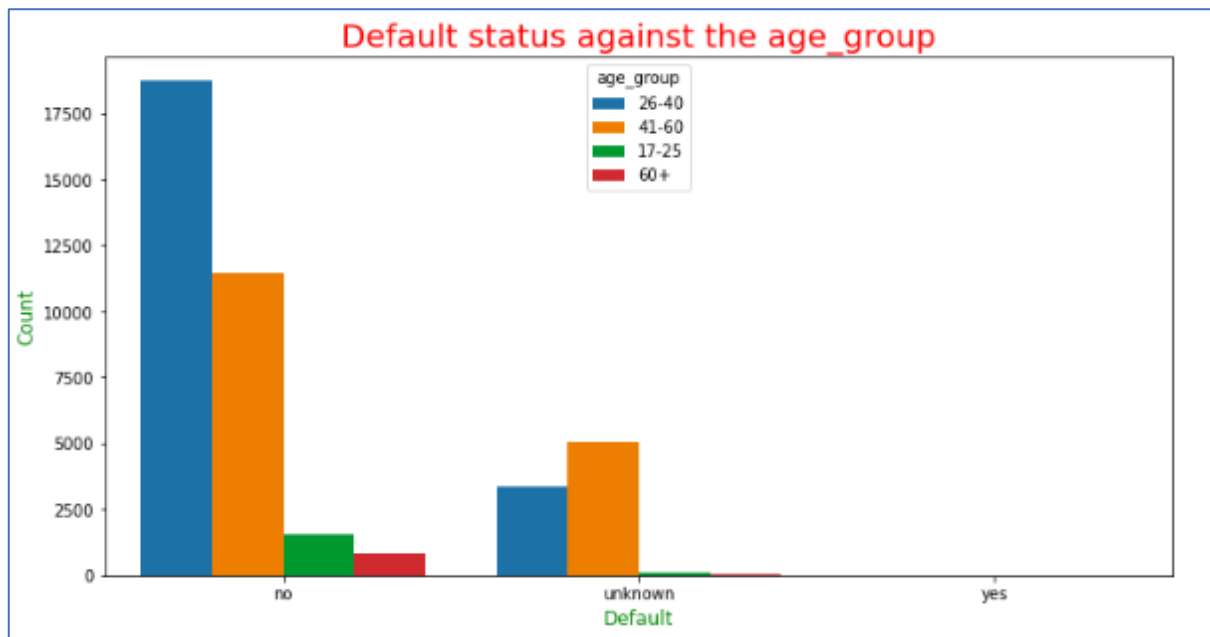
A blue-collar professional requires at least a basic.6y education as observed from the above data, with over 3000 customers achieving that level. In a management type

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 all this can be concluded from the above graph.

- Default feature :



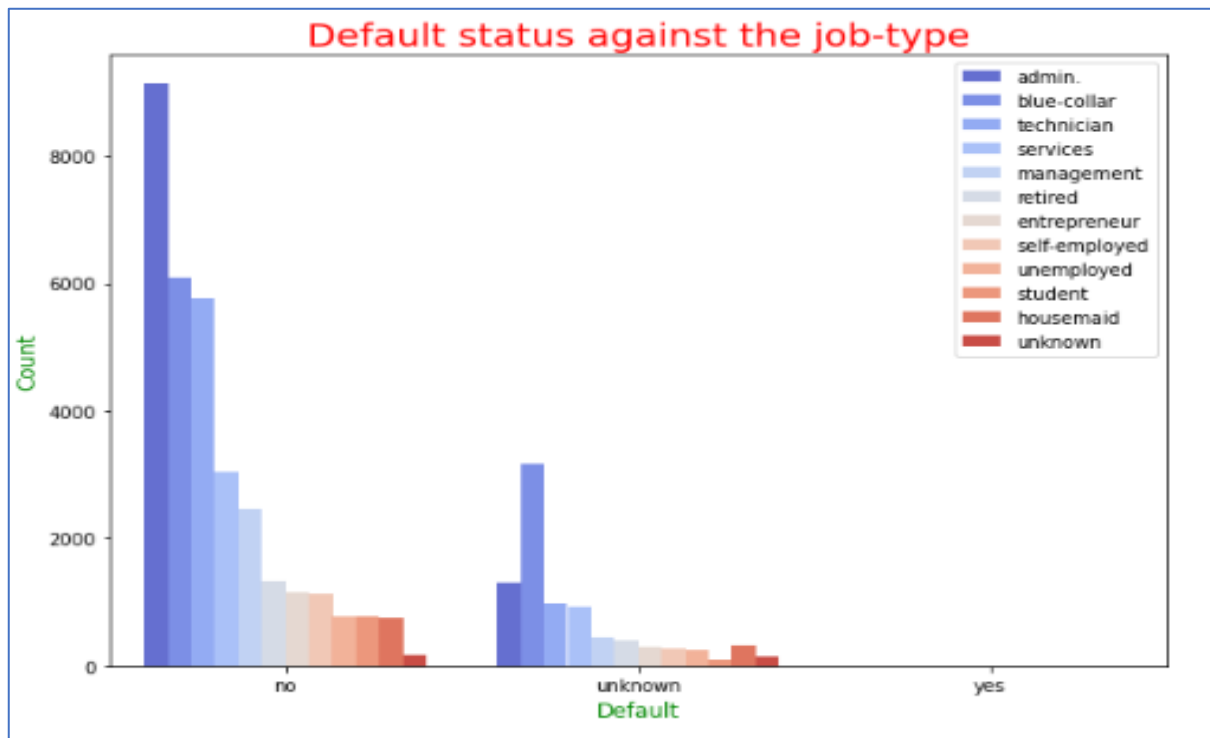
Out of the 41117 customers, 75% of them do not have a credit in default. On the other hand, there are zero customers who have defaulted. But, around 8000 people have an unknown default status on their name which sounds strange.



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

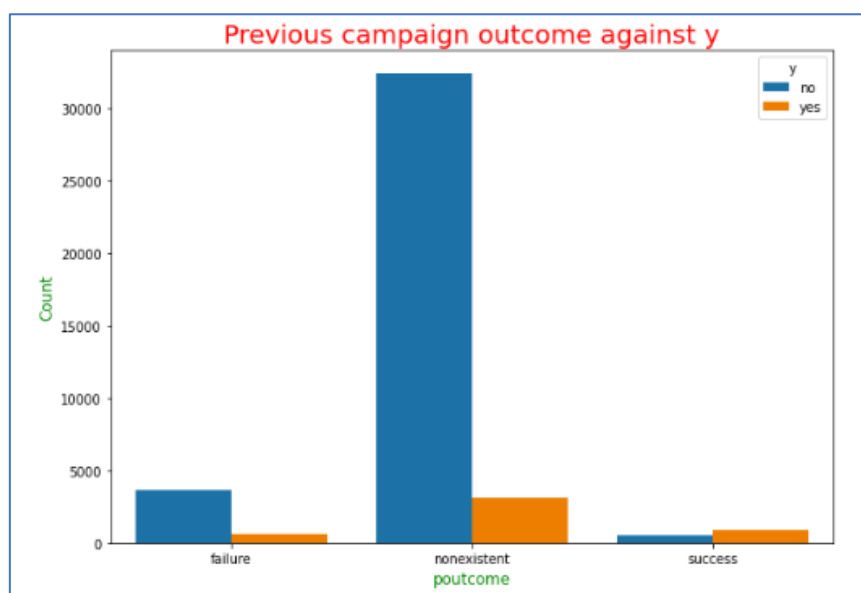
other two groups are lower as expected since they might not be financially secure enough to acquire a loan.

Across various job categories too we look at the default status.

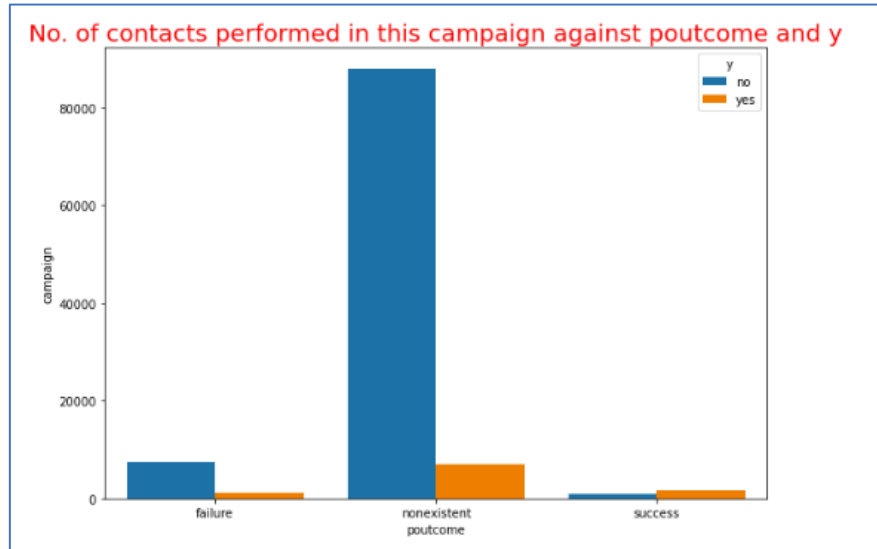


The professions of admin., blue-collar and technician lead the other roles in terms of having a no default. In addition, around 5000 customers in blue-collar jobs have an unknown status.

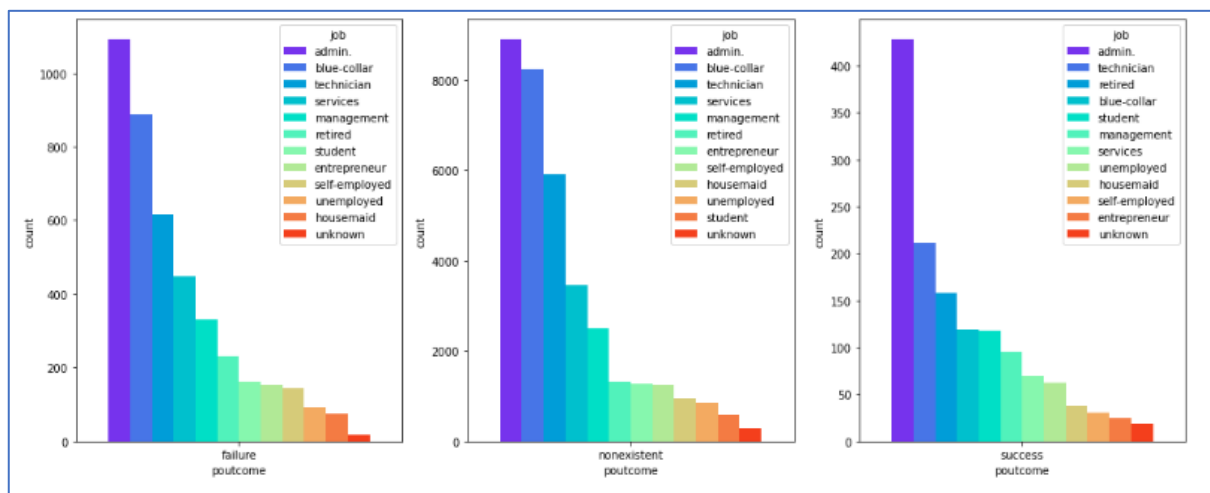
- Poutcome feature :



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.

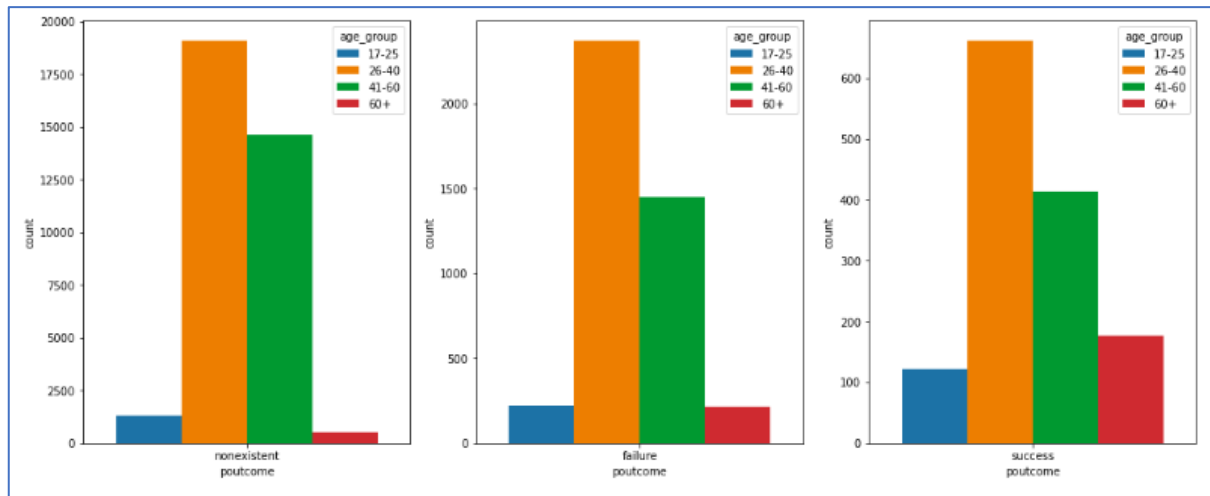


Greater than 90% of customers for whom the outcome was non-existent have rejected the term deposit plan. The effort of the marketing team can be understood by looking at the number of contacts made during this campaign. Below plot, shows that more than 80000 contacts are performed for customers who have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



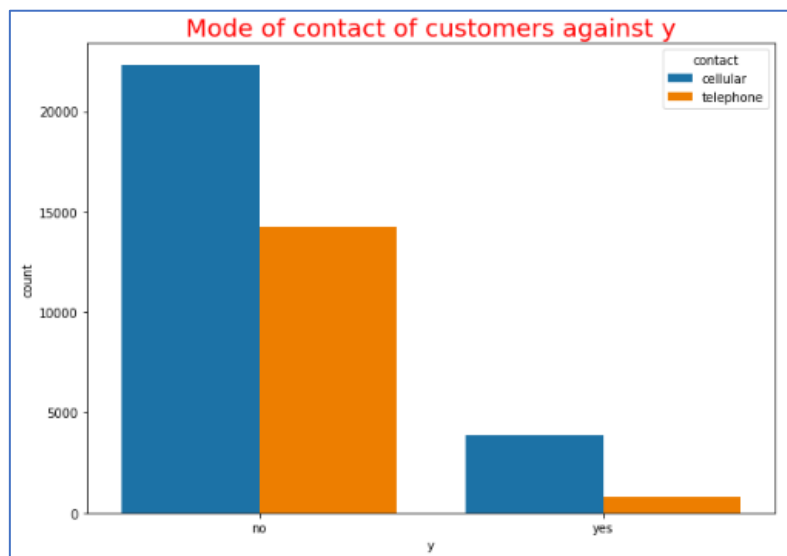
The 'job' feature is also analysed together with the 'poutcome' variable. For all the non-existent outcomes there are over 14000 customers in the roles of administrative, blue-collar and technician. For the outcome of success, customers in administrative jobs are almost double in number to the next job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

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



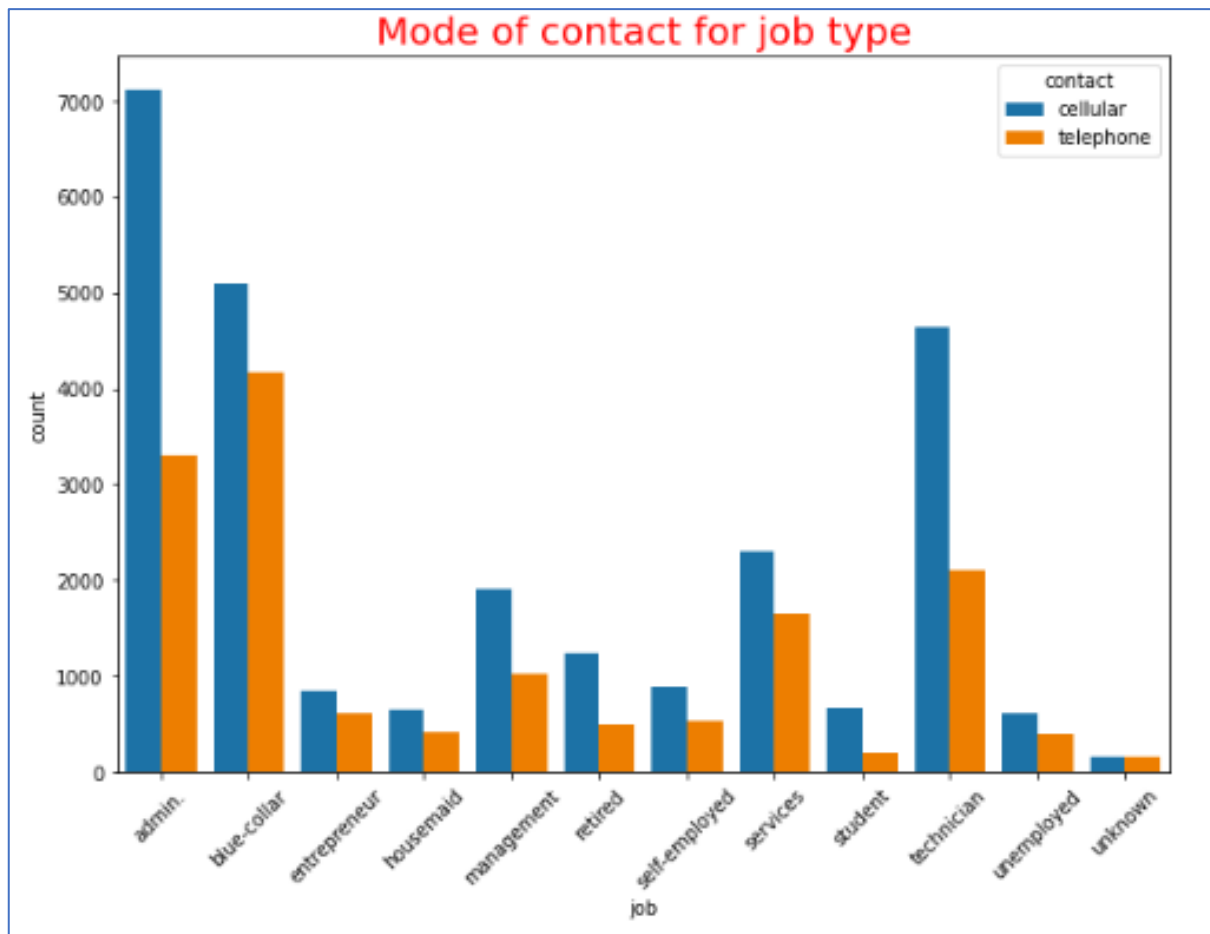
The previous campaign had been a success for over 1000 customers in 26-40 and 41-60 age-groups. For the other age-groups, the number of customers is very small in number for all the outcomes. Close to 35000 customers have a non-existent outcome associated with them for the previous campaign.

- Contact Feature :

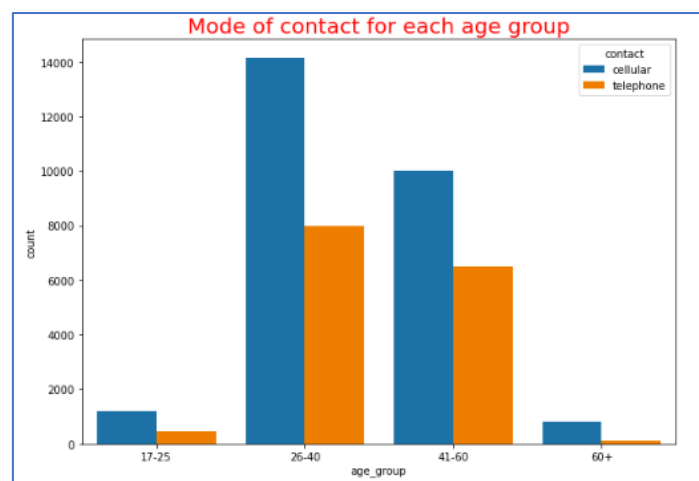


The impact of the mode of communication by the marketing team is also studied along with other features. Looking at the numbers tell that approximately 25000 customers have been contacted through a cellular mode in this campaign, with the rest through telephone. The mode of communication is analysed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



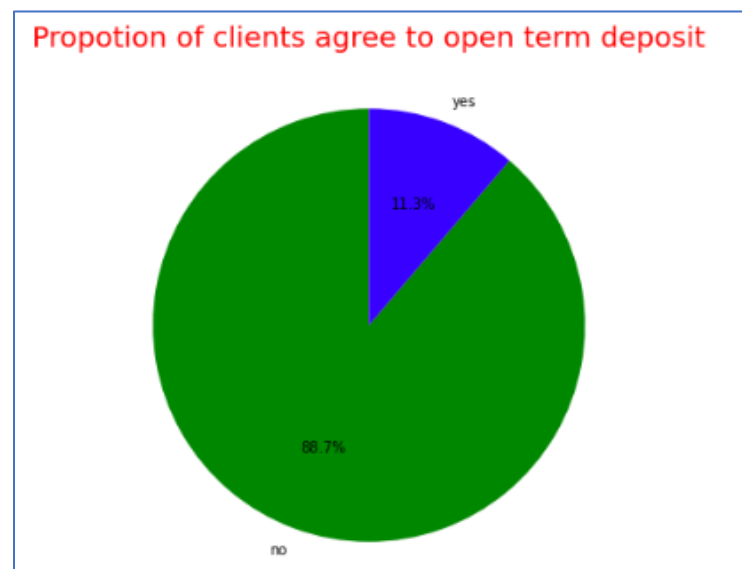
The cellular mode is done for approximately 17000 customers who are in admin., blue-collar and technician job roles. This is almost double the number for telephonic mode of communication for the same three professions. An age-groupwise analysis is also done to get more understanding of this feature.



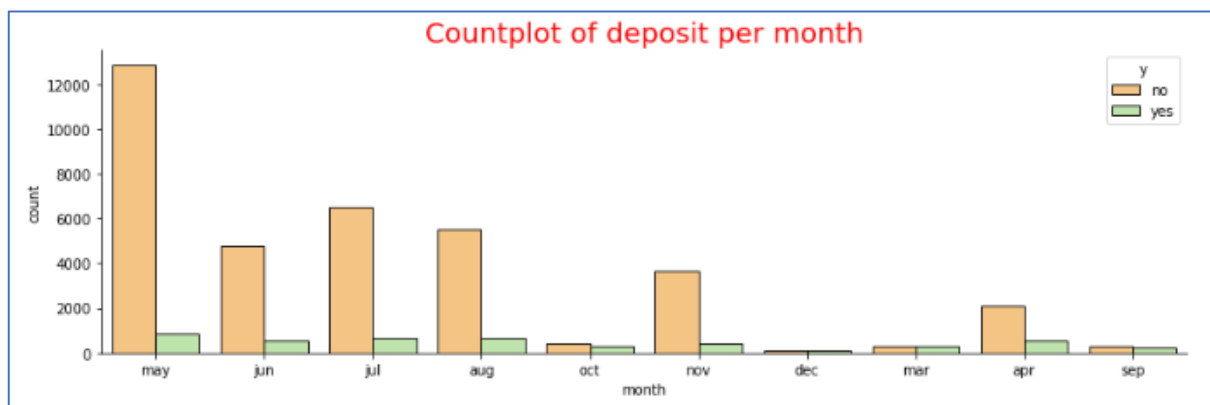
The cellular mode is used for approximately 24000 customers in the age-groups 26-40 and 41-60 yrs. For both these groups there is not a big gap in the number of customers between the two modes of communication.

- 'Y' Feature :

The propotion 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.



The most number of successful conversions were observed in the month of may as shown by below plot.



Observations and Recommendations :

- Senior citizens and students responds better to their proposal and have high conversion rate.
- Looking at the customer database, the age-groups of 26-40 and 41-60 have a higher proportion. These groups present a profitable target for the marketing team.
- Month of may is most successful as we can say most term deposits are subscribed in this month.
- It is imperative to form target groups based on socioeconomic categories. Age, income level (not always high), profession and education can accurately determine the marketing profile of a potential client.
- Increasing the time with the customer and using different means of communication tends to increase the chances of customer subscribing to the term deposit plan.
- 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/AbhimanyuGangani/Week_7_Bank_Marketing/tree/main/Week_10_Bank_Marketing