



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis Presentation

Project name: **Bank Marketing Campaign**

Team: **Data Science Enthusiasts**

Date: **August 18th, 2021**

Agenda

Executive Summary

Data Understanding

EDA

EDA Summary

Recommendations

Team member's details

Group Name: <i>Data Science Enthusiasts</i>					
	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

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.

Executive Summary

- **The Client:**

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.

- **Problem statement:**

Build a Classification ML model to shortlist customers who are most likely to buy the term deposit product. This would allow the marketing team to target those customers through various channels.

- **Analysis:**

The Analysis of this data is divided into the following parts:

- Data Understanding
- Univariate analysis
- Bivariate analysis
- Model recommendations

Data Understanding

- **Datasets description:**

Four datasets provided:

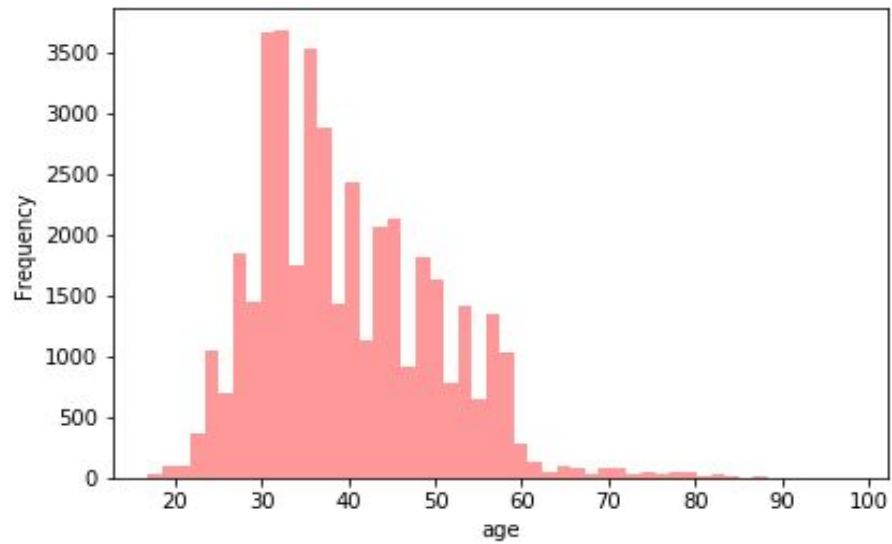
- bank-additional-full: 20 inputs (+1 target variable) and 41119 observations
- bank-additional: 20 inputs (+1 target variable) and 4119 observations
- bank-full: 17 inputs (+1 target variable) and 45211 observations
- bank: 17 inputs (+1 target variable) and 4521 observations

- **Assumptions:**

- Timeline of observations - May 2008 to November 2010.
- 'Duration' feature is dropped to give realistic predictions from the classification model.
- A frequently occurring missing value 'unknown' is considered as another category for the categorical features.
- Duplicated rows were deleted from the dataset.

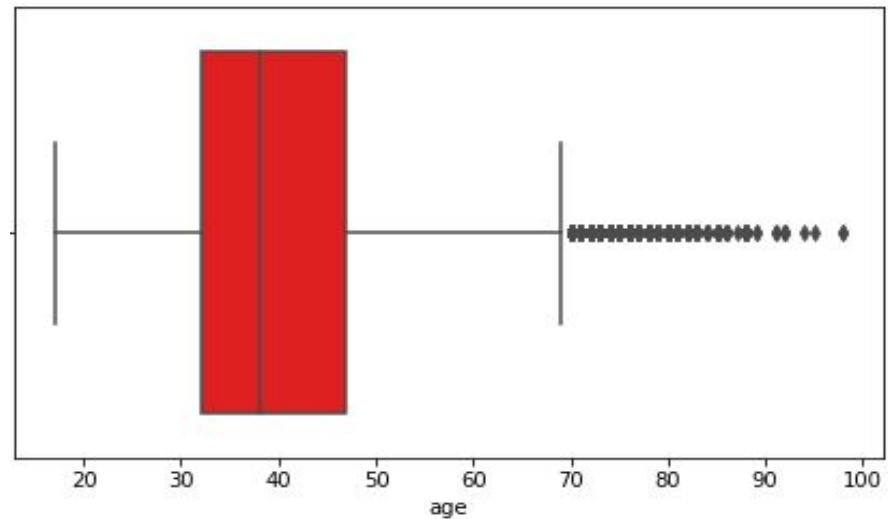
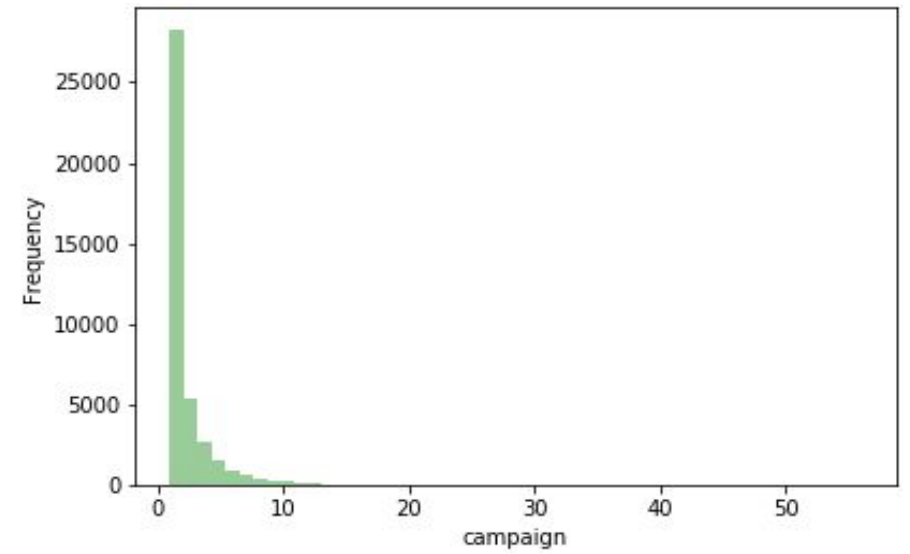
EDA

- Univariate Analysis -



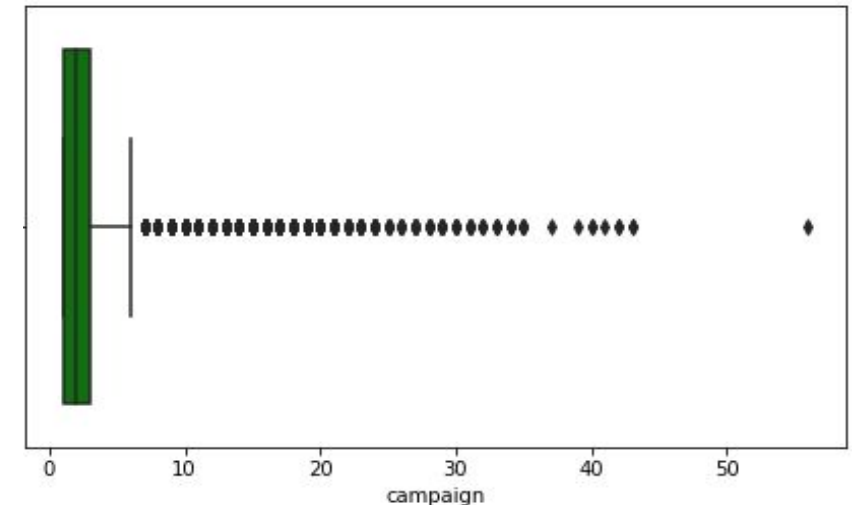
→ Positive skewness observed in these two distributions.

→ 97.5% of the clients fall in age range of 20 - 60 years old.

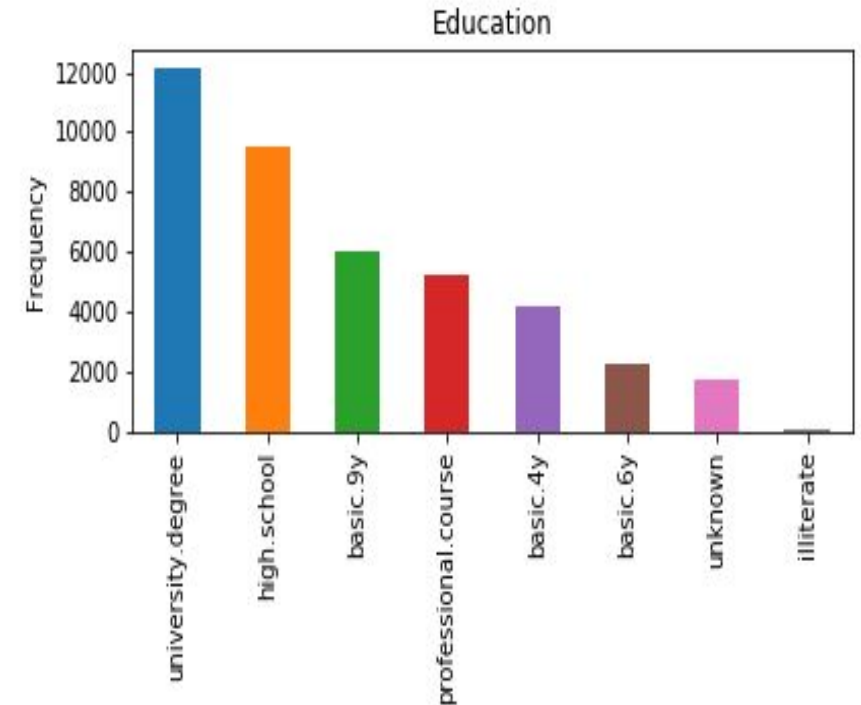
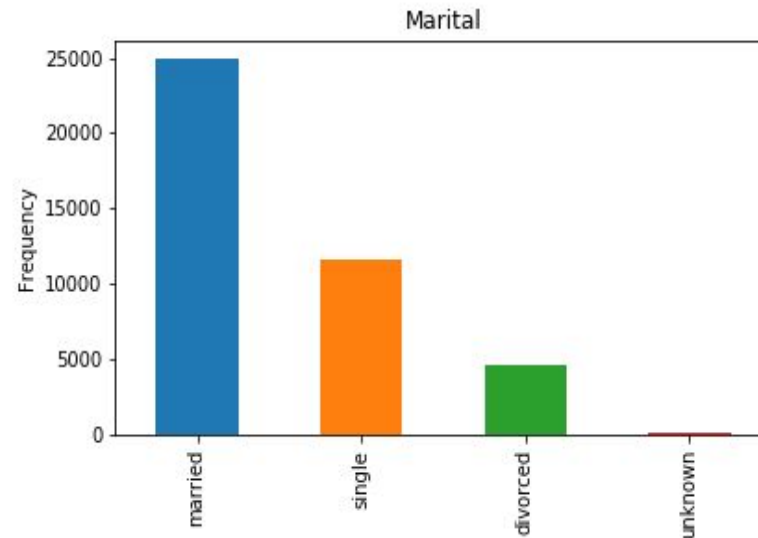
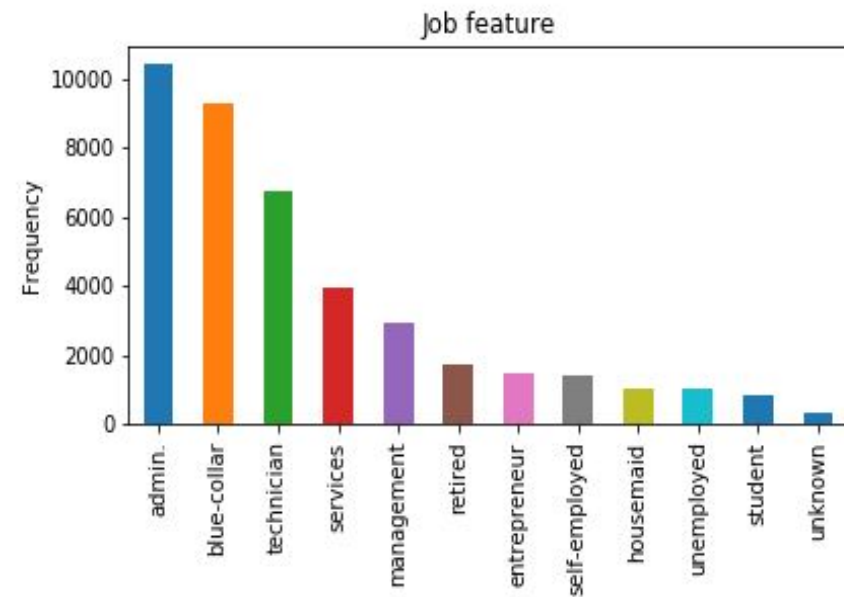


→ Box plots of 'age' and 'campaign' features show that data points outside the whiskers are outliers.

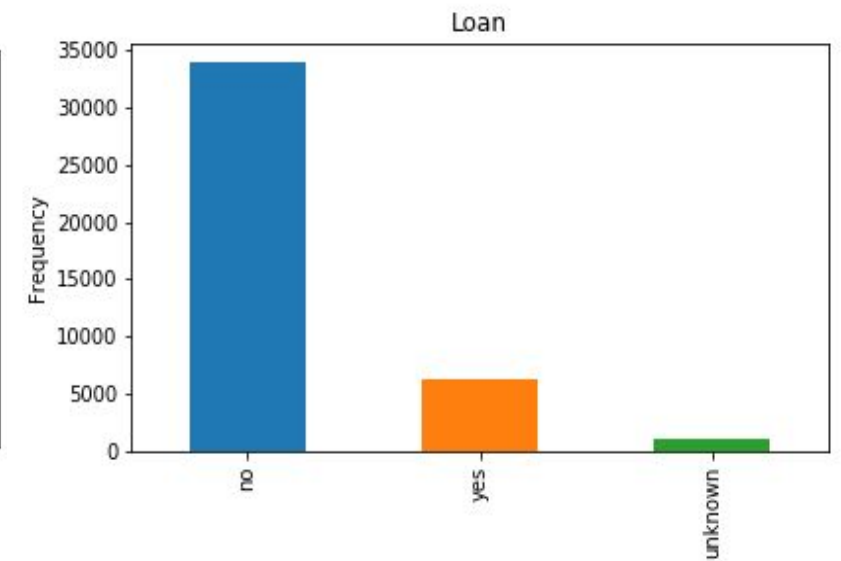
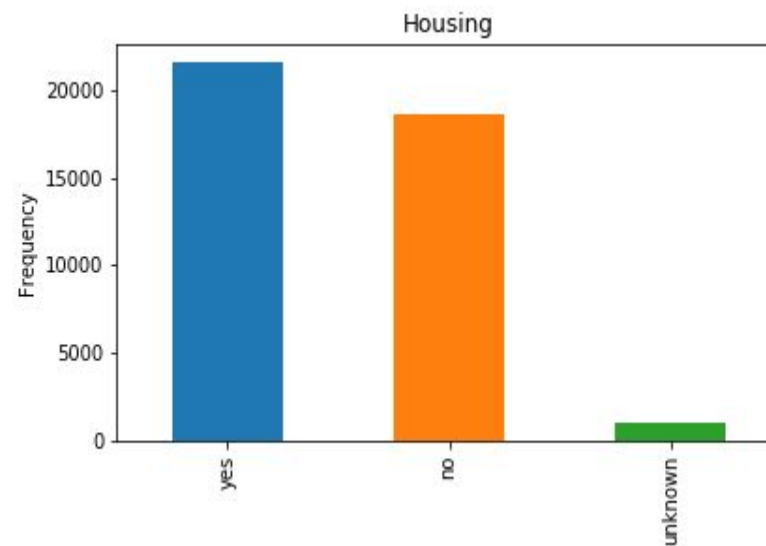
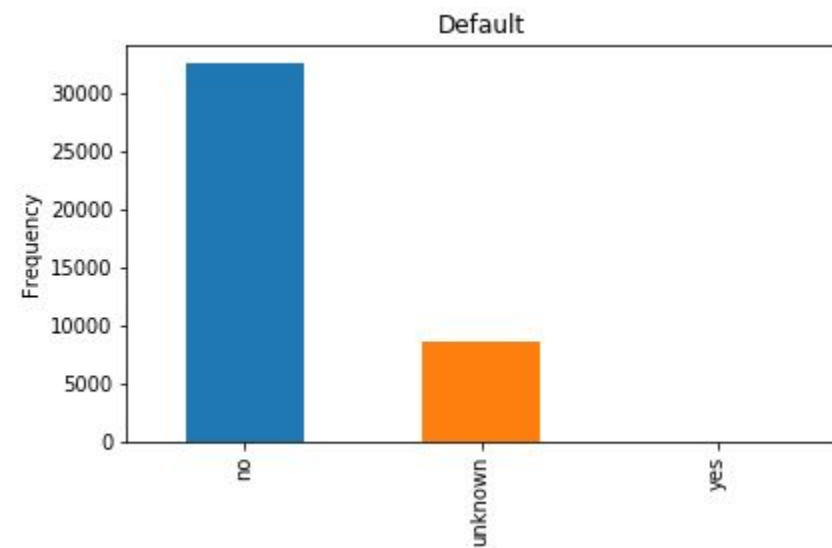
→ No outlier treatment is required since they are realistic data points.

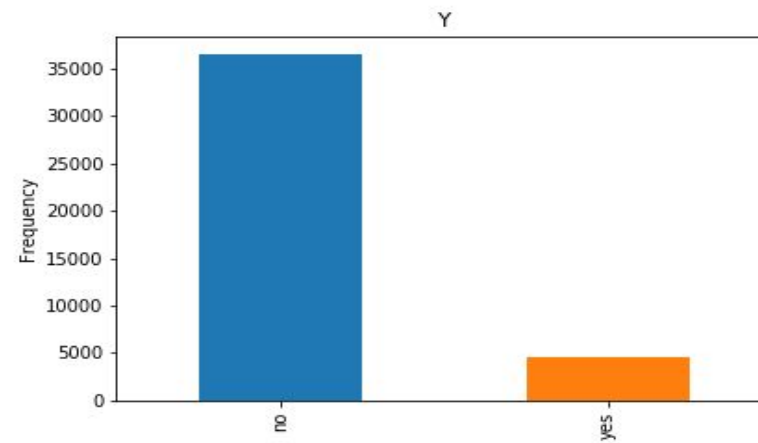
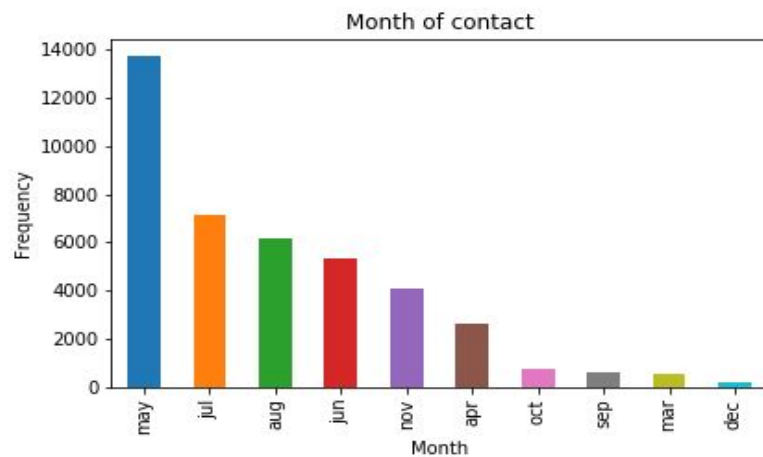
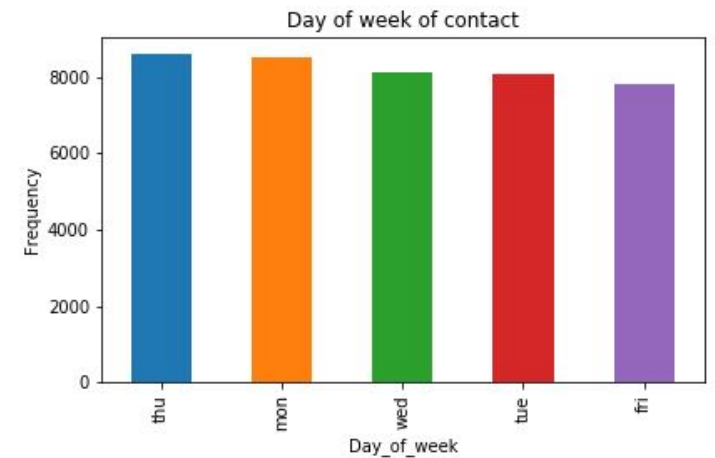
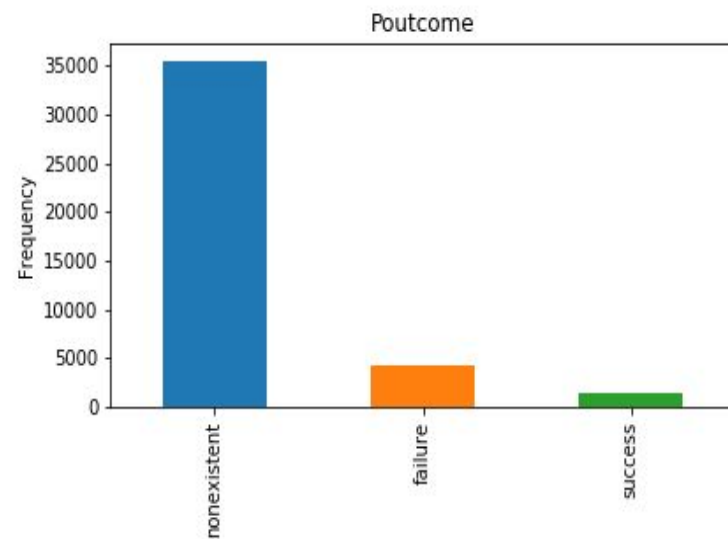
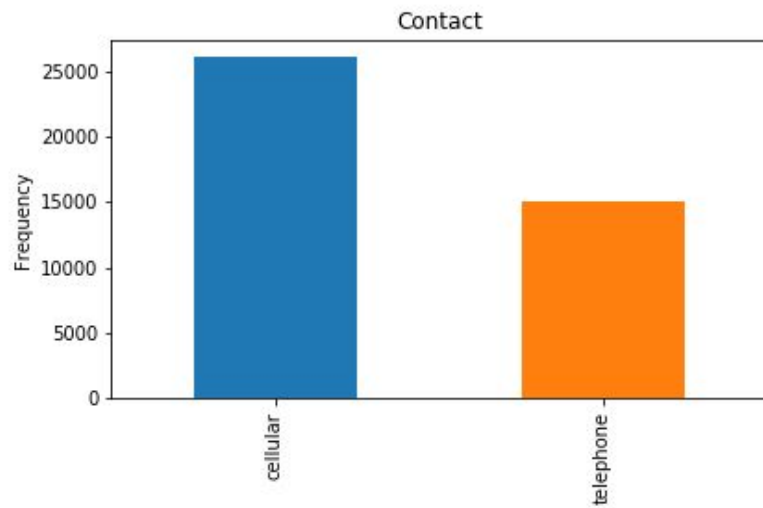


- Administrative staff and technical specialists opened the deposit most of all. In relative terms, a high proportion of pensioners and students might be mentioned as well.
- Majority of the clients are employed in admin, blue-collar, technician and services job types.
- Although in absolute terms married consumers more often agreed to the service, in relative terms the single was responded better.



- Home ownership does not greatly affect marketing company performance.
- The difference is evident between consumers who already use the services of banks and received a loan.

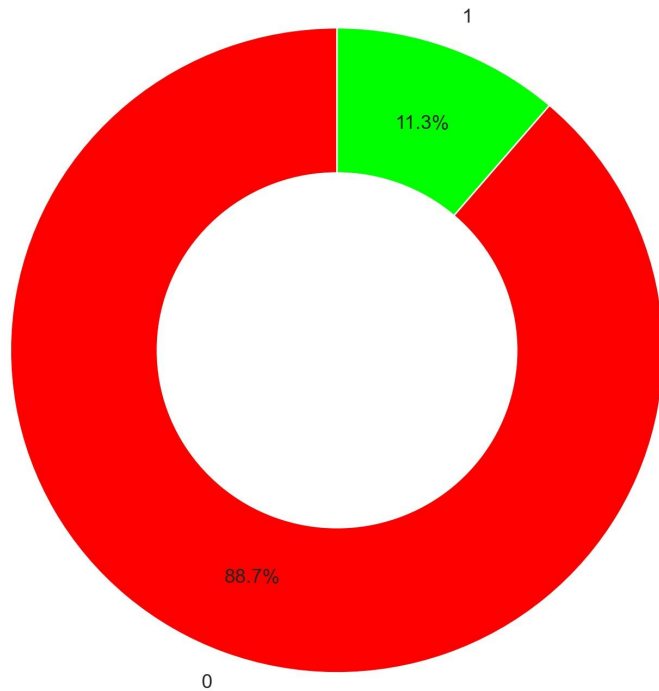




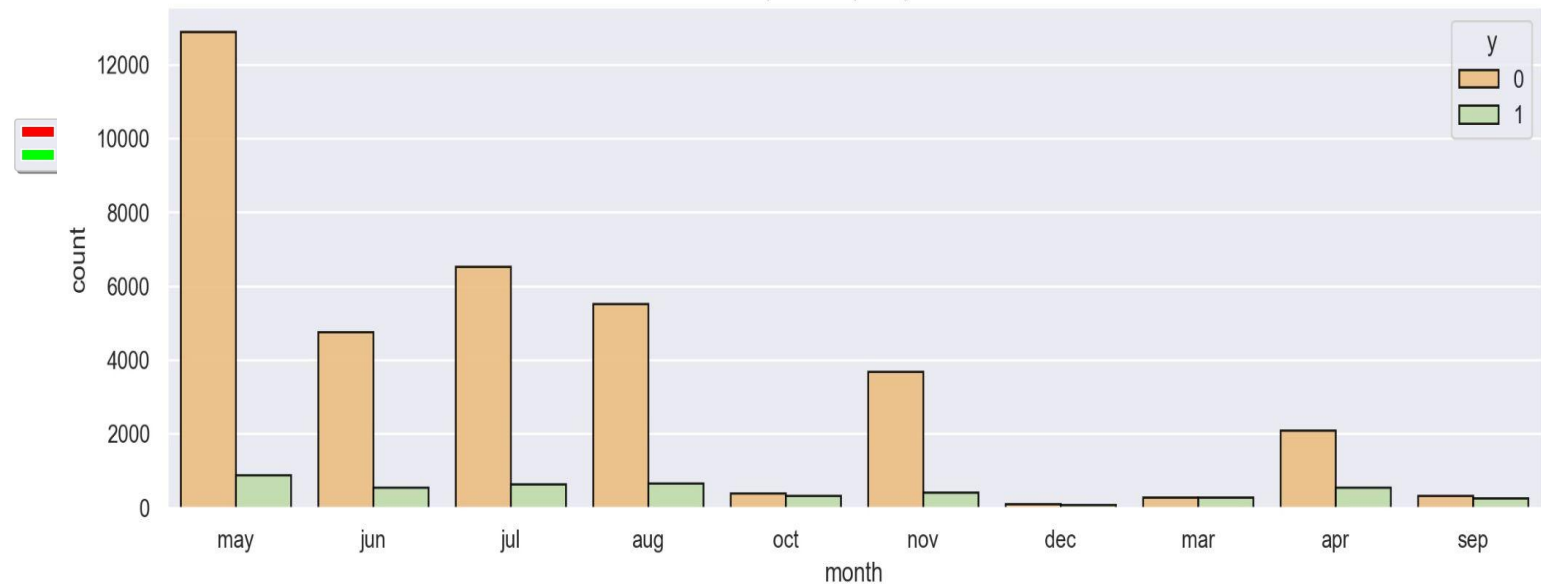
- Univariate plots of categorical features of this data.
- Among the customer base, more than 35000 of them have rejected the term deposit plan.

- Out of the total calls that are made, 11.3 % of them are successful.
- It looks like the month of May is the most promising one for accepting offers.

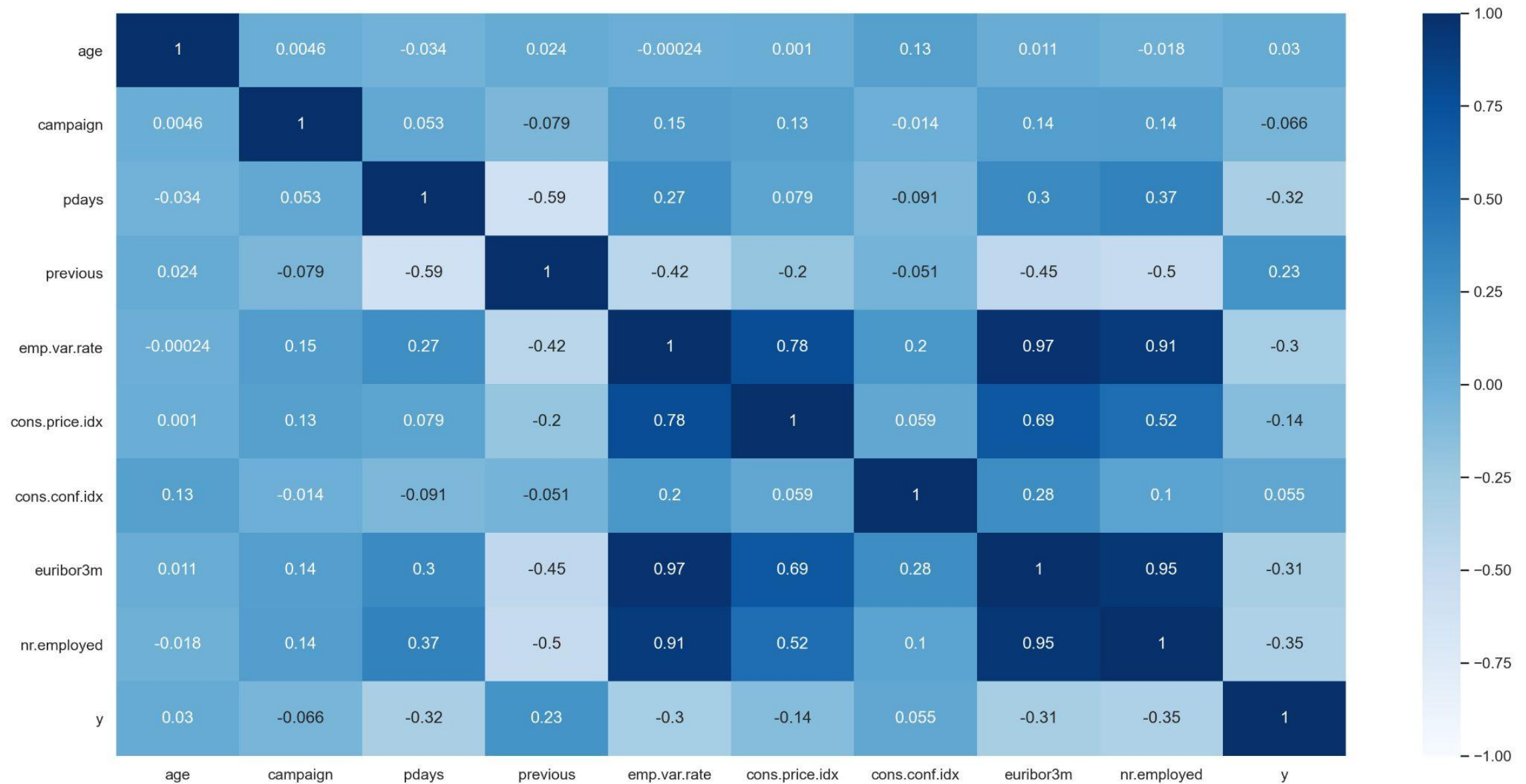
Percentage of bank clients deposit agreement



Countplot of deposit per month

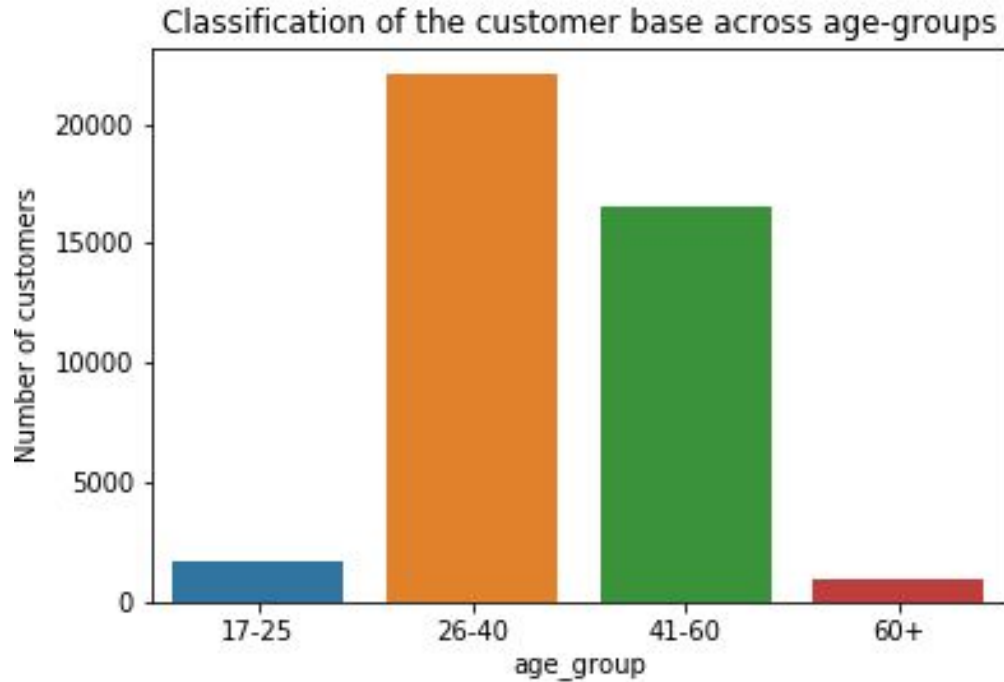


→ Highly correlated features (employment rate, consumer confidence index, consumer price index) may describe clients state from different social-economic angles. Their variance might support model capacity for generalization.



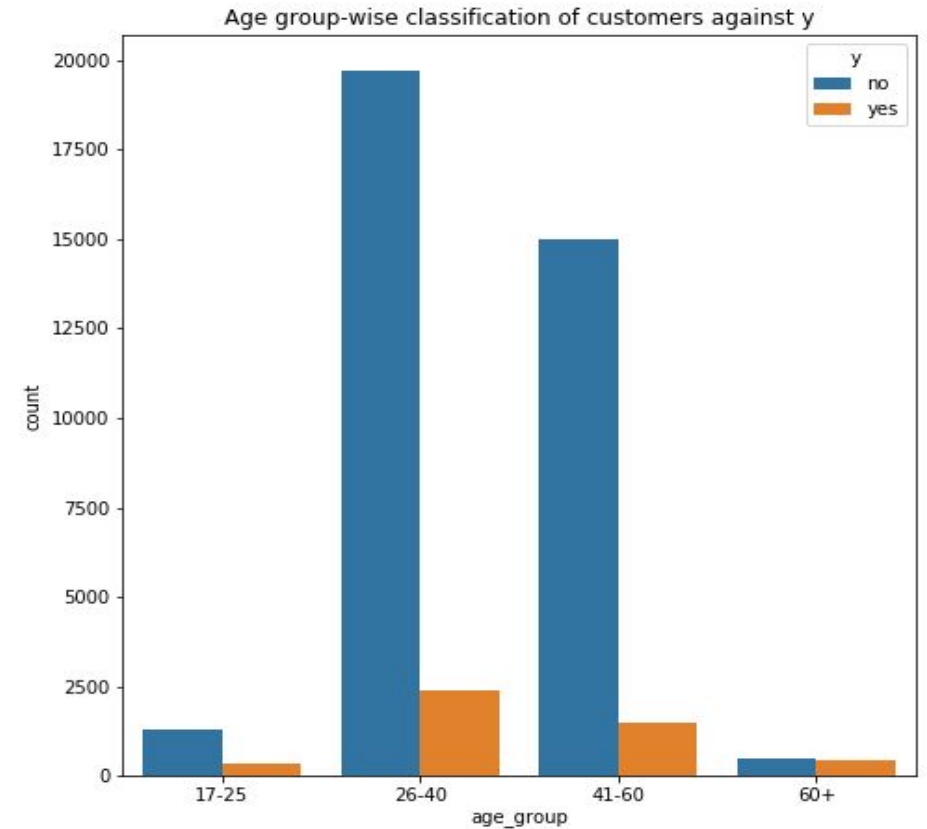
- Bivariate Analysis -

Age-group: New column engineered from the 'age' feature to allow more interactive analysis.



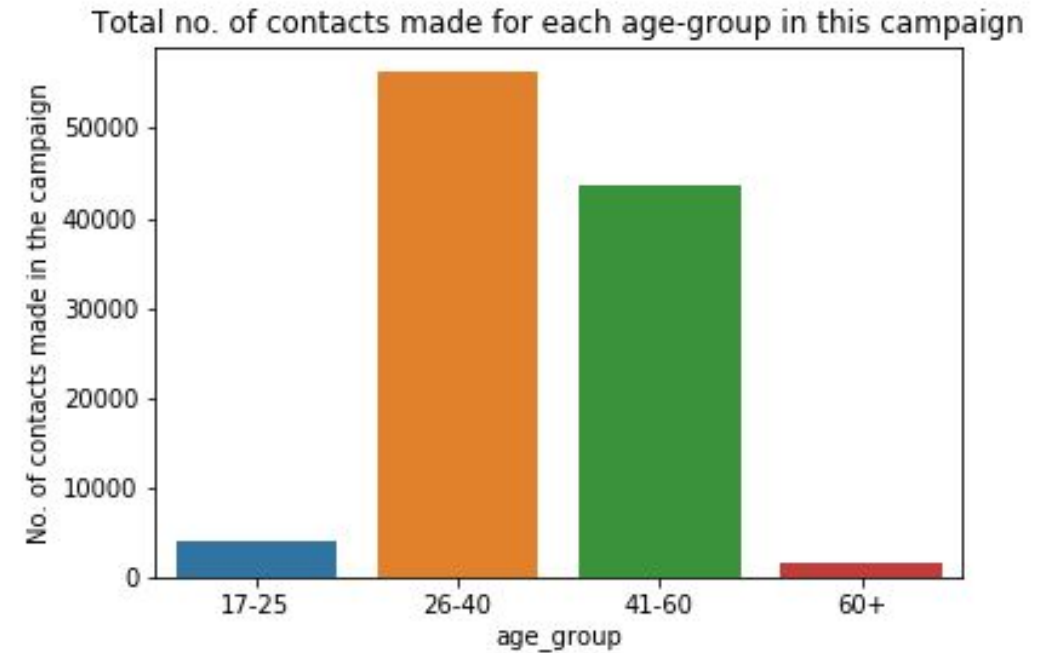
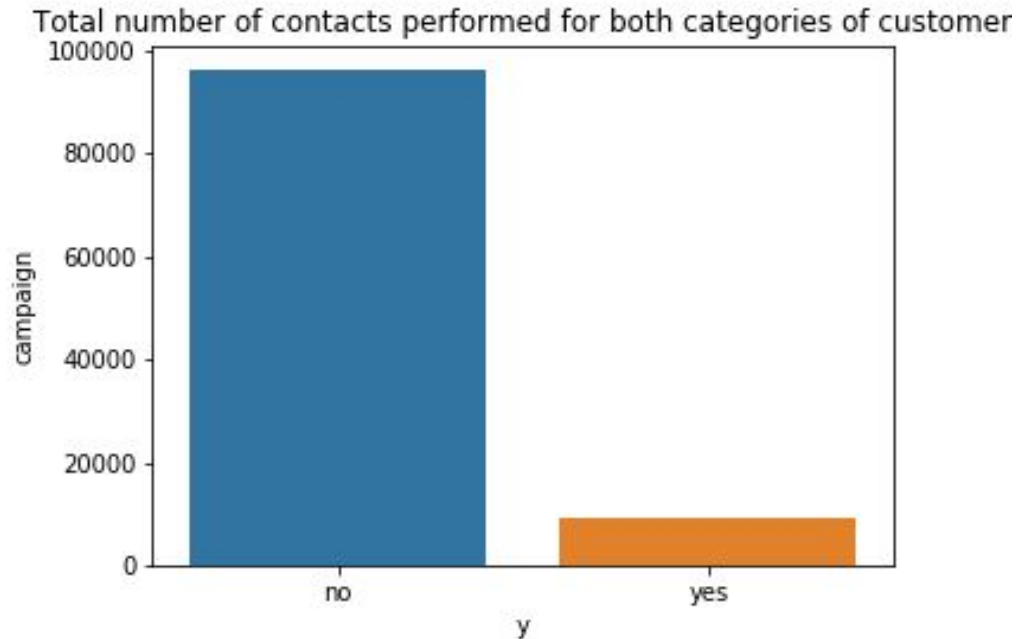
→ The 26-40 and 41-60 age-groups occupy a proportion of 97% of this customer base.

→ Approximately 83% of customers in these age-groups have rejected the term deposit plan.



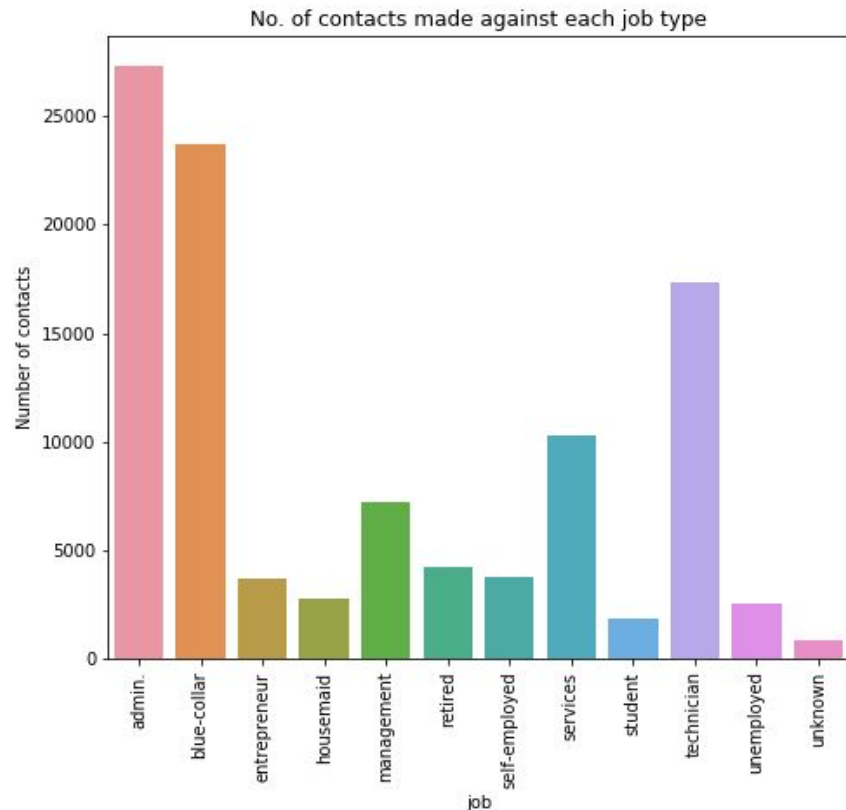
Campaign: Exploring the number of contacts performed in this marketing campaign

- The 26-40 yrs group witnesses over 50000 contacts in the current campaign.
- There is close to a 39% gap in the number contacts made between 41-60 and 17-25 and 60+ combined.



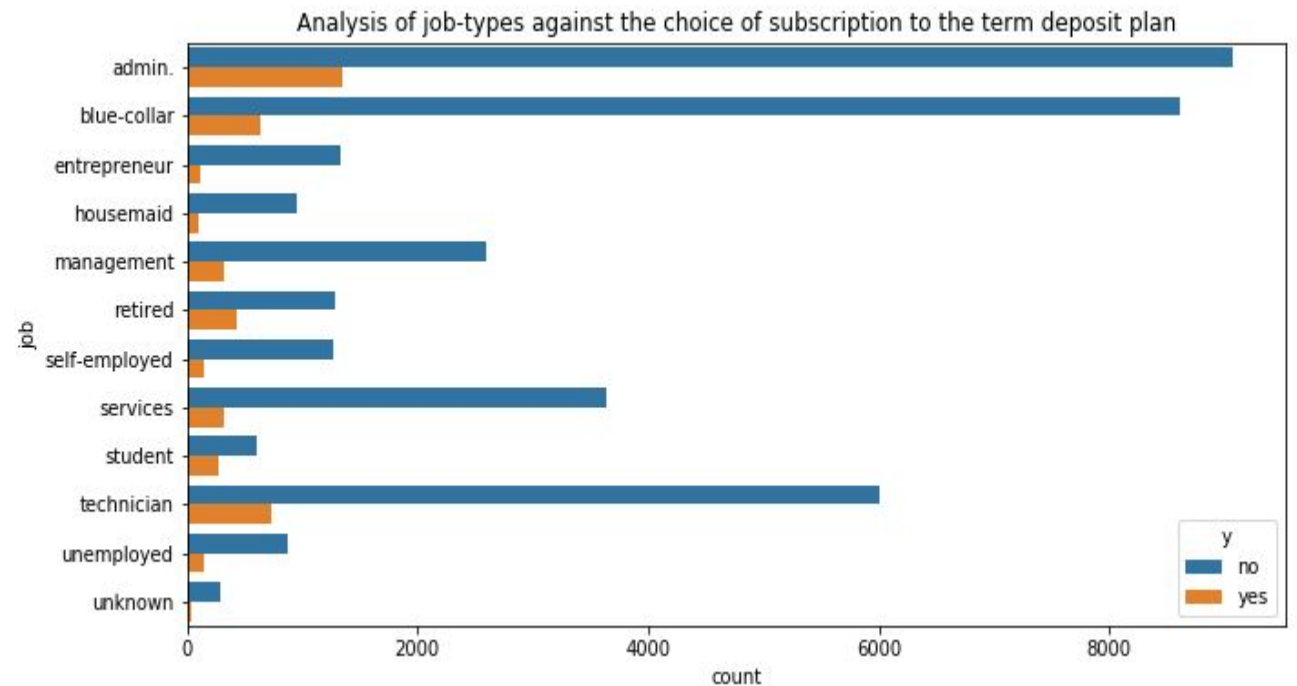
- The marketing team makes over 95000 contacts to customers who have not opted for the term deposit scheme.

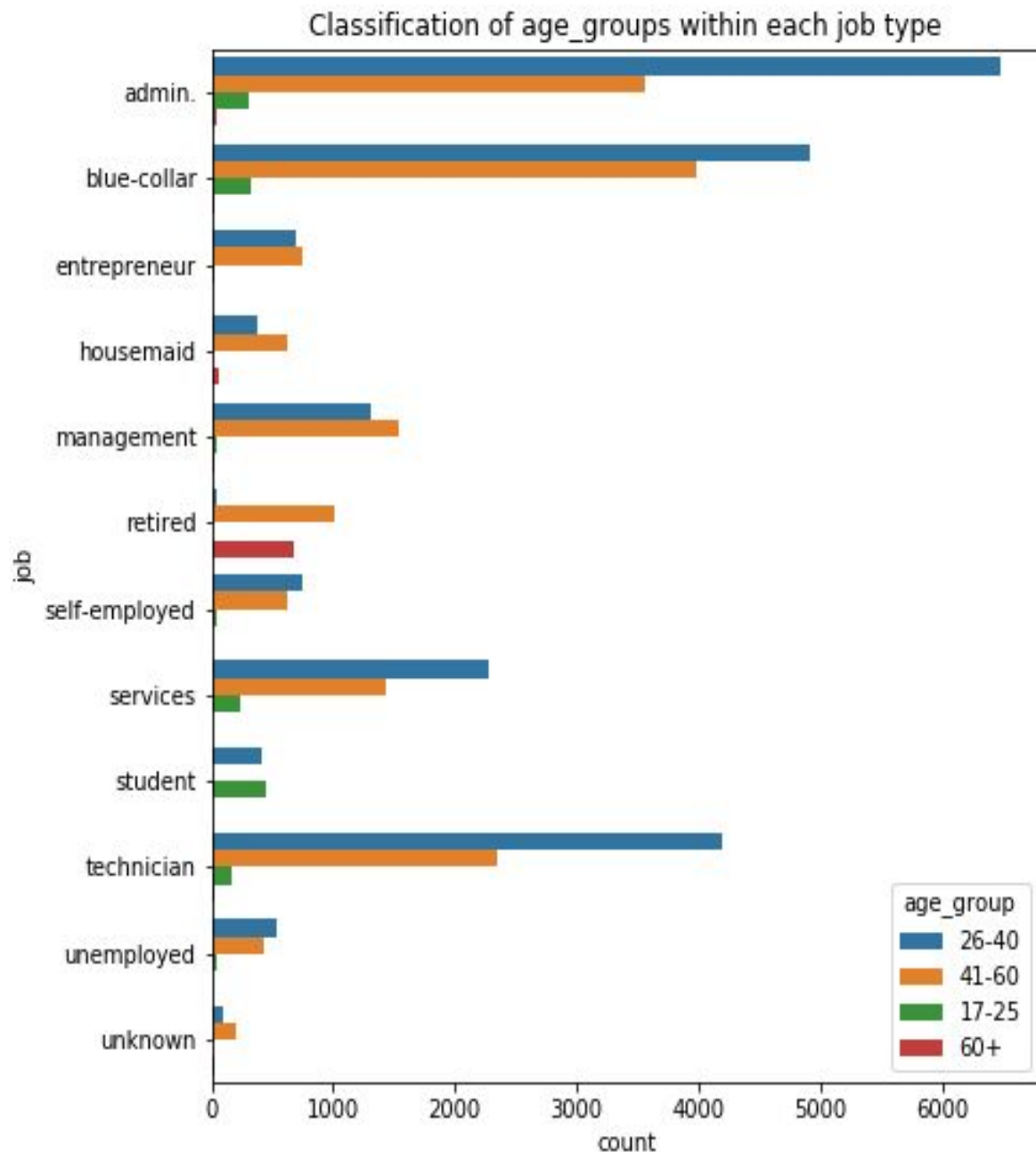
Job: Look at different job categories across the customer base



→ In the four jobs mentioned above, only 13.5% of the customers have accepted the term deposit plan.

- In the customer base, approximately 63% of them are in admin, blue-collar and technician professions.
- The next common jobs are management and services occupying around 41% of all customers.

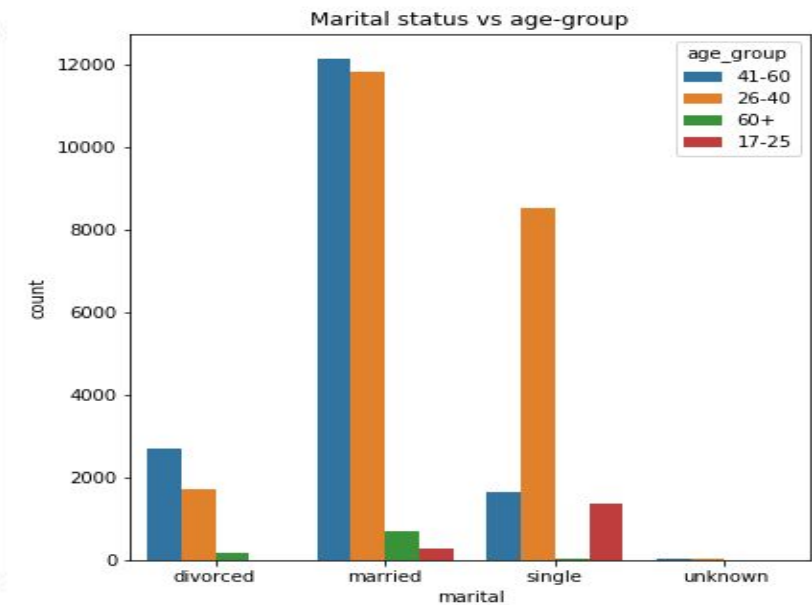
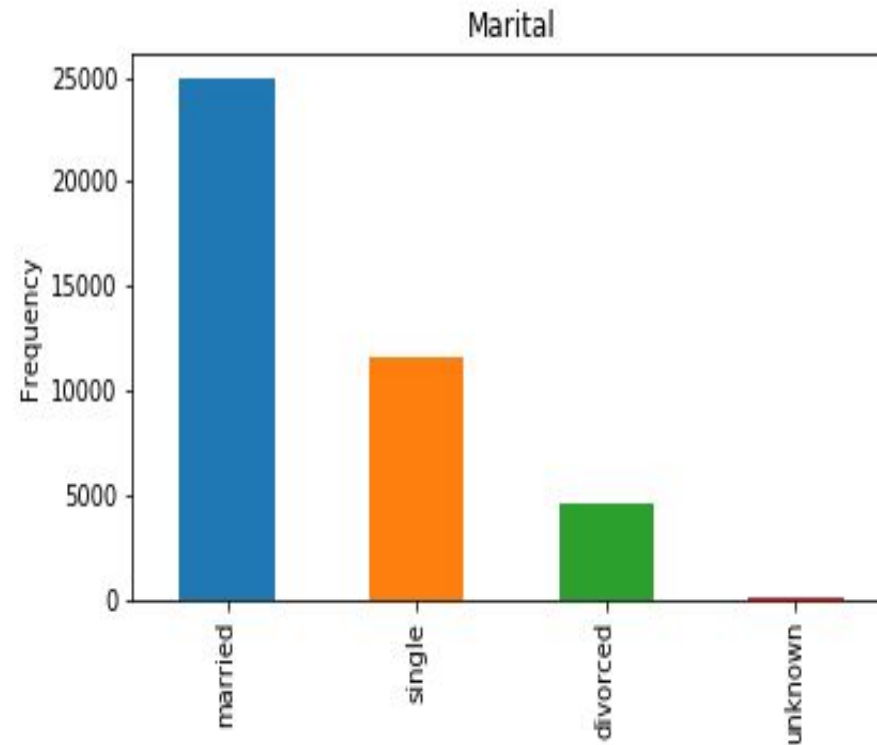
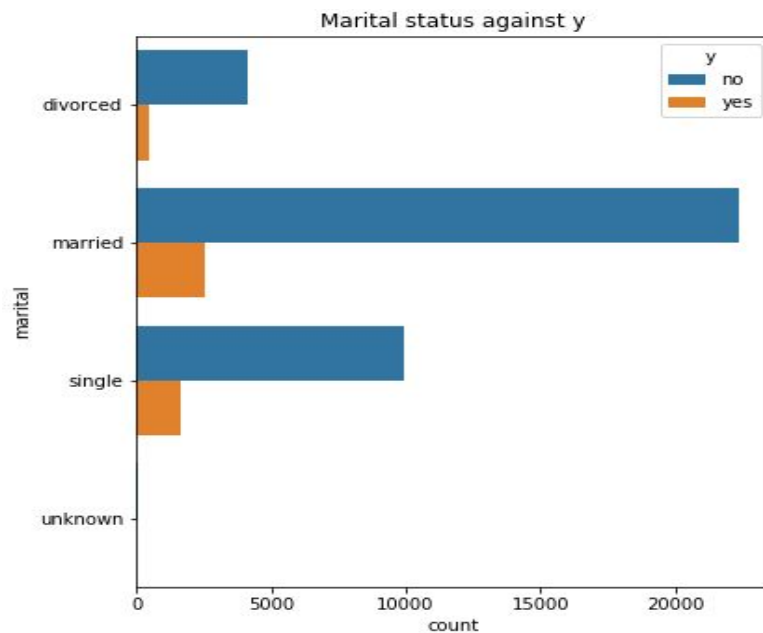




- The 26-40 and 41-60 age group is prominent in these particular four professions namely: admin, blue-collar, technician and services.
- In roles of entrepreneur, housemaid and management the 41-60 yrs group leads the other category.
- An unknown job category comprises only a few 100 or 200 in numbers.

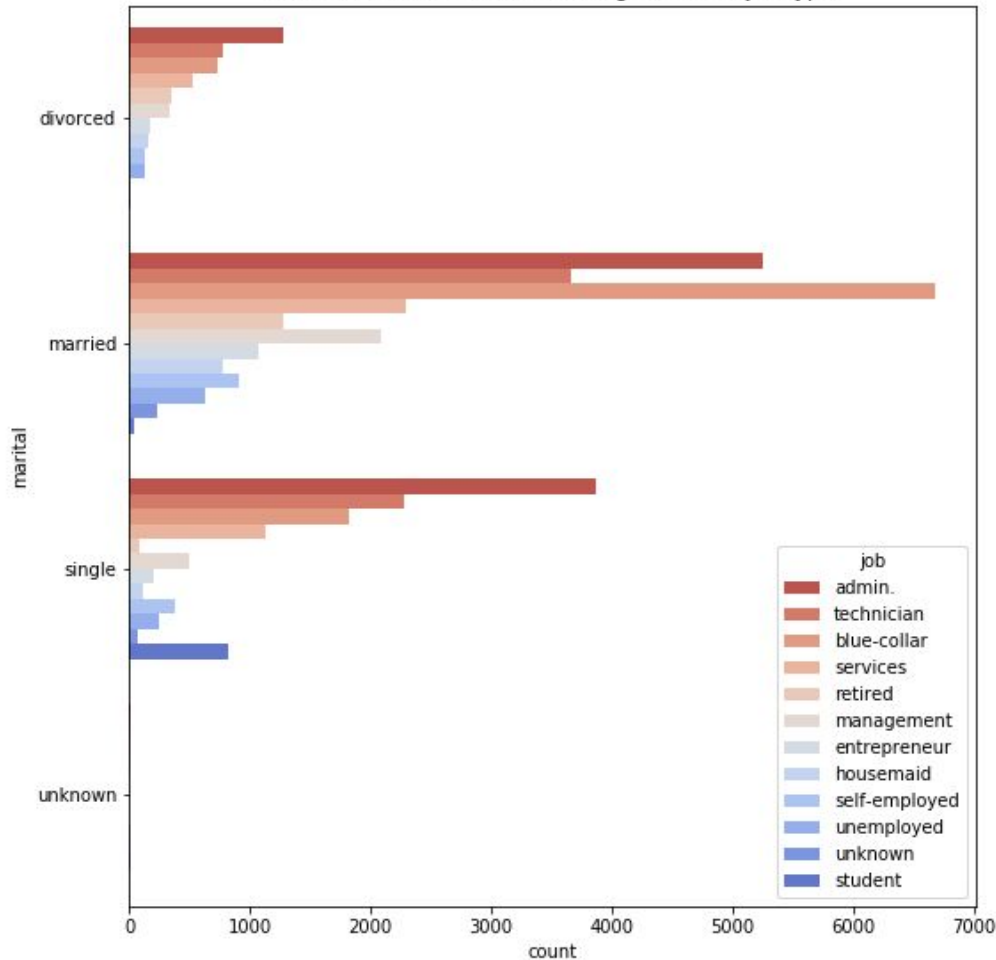
Marital: Analysing the marital status of customers

- Out of the 41117 customers, there are more than 50% of them married, and close to 10000 singles.
- Although married consumers more often agreed to the service, in relative terms the singles responded better.
- A fourth category of unknown has little to nothing customers in it.

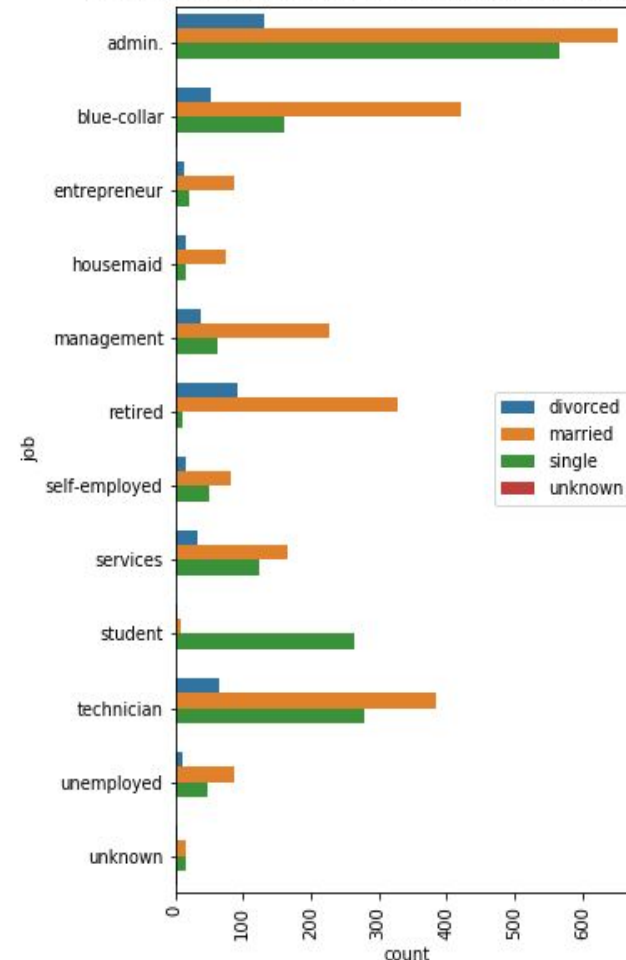


- Married people in admin, blue-collar and technician jobs are around 15000 in number.
- Among single customers, there are almost double the number in the profession of admin than in technician.

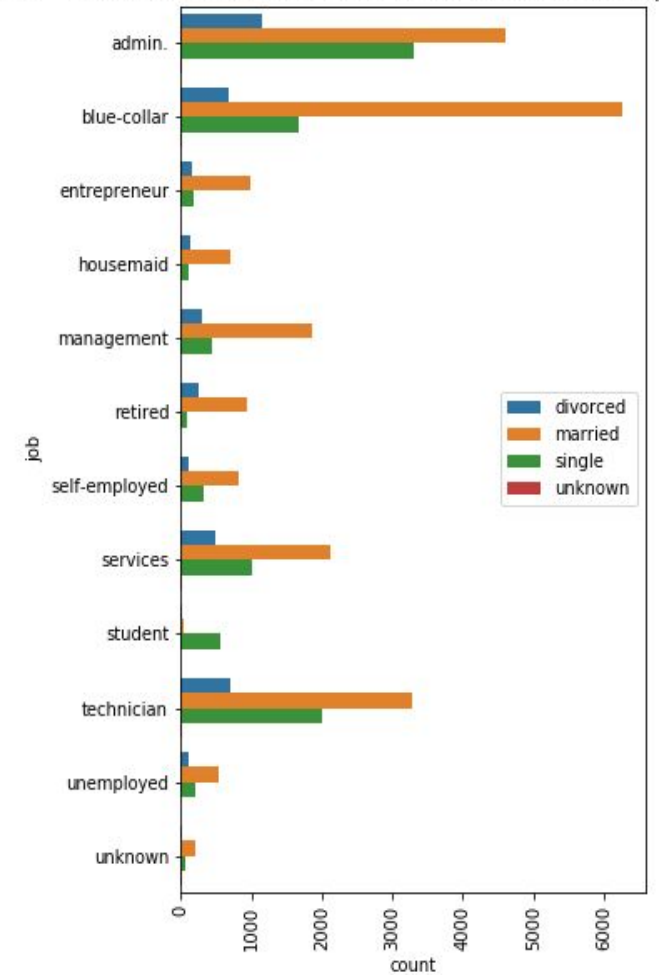
Marital status of the customer against each job type



No. of customers who have subscribed to the term deposit plan

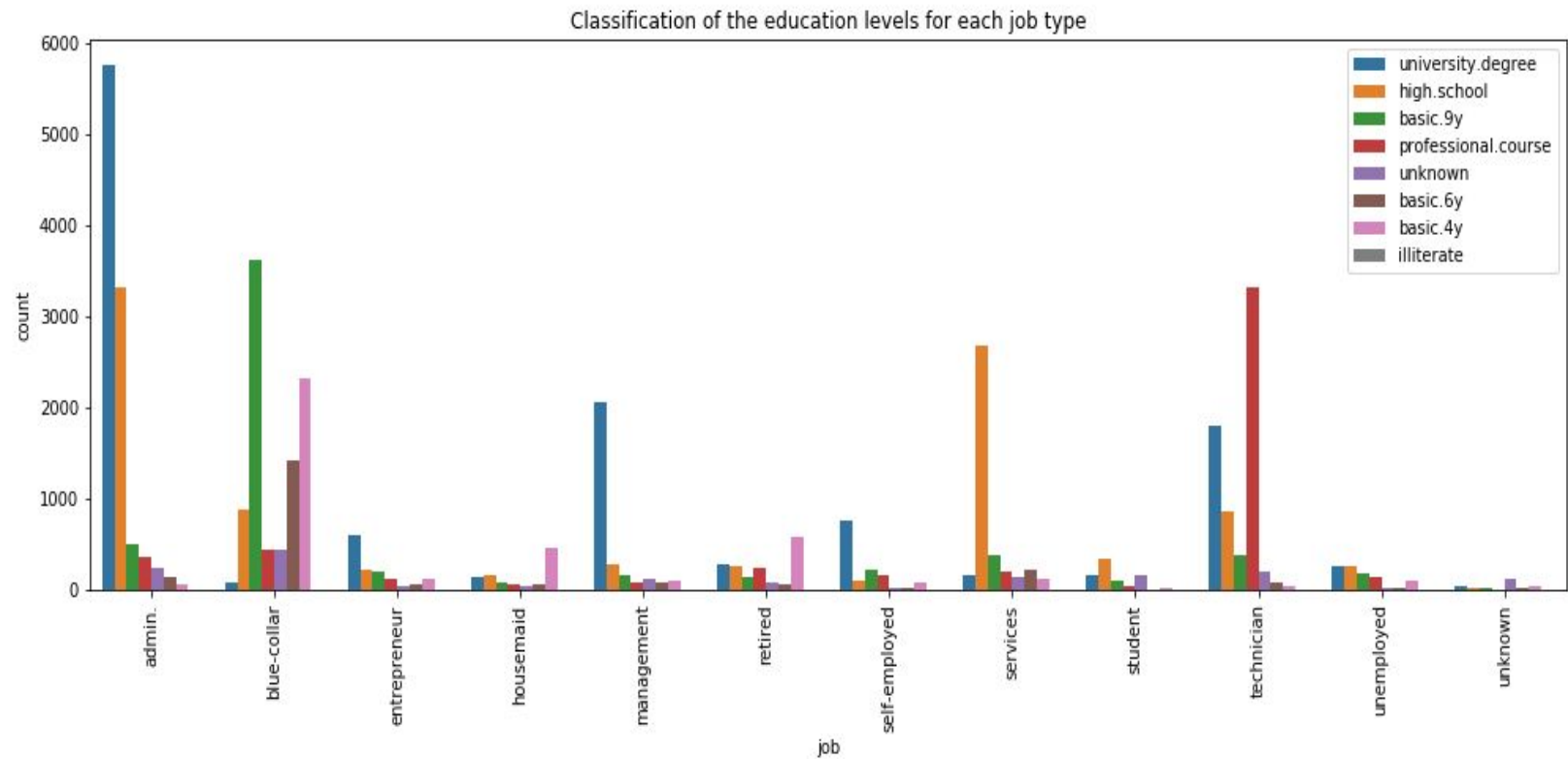
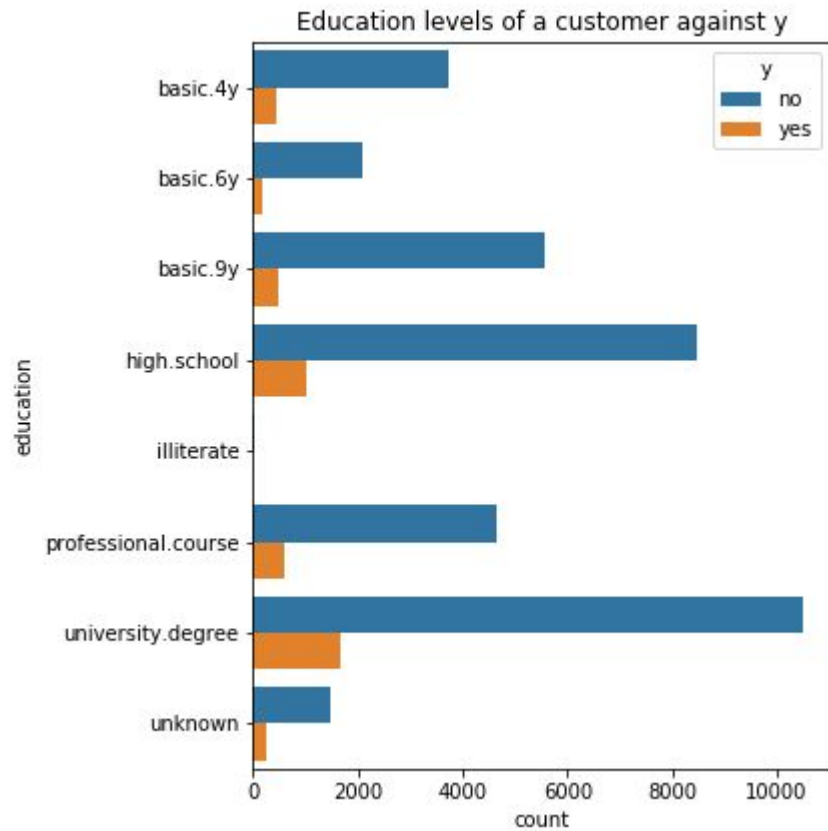


No. of customers who have not subscribed to the term deposit plan



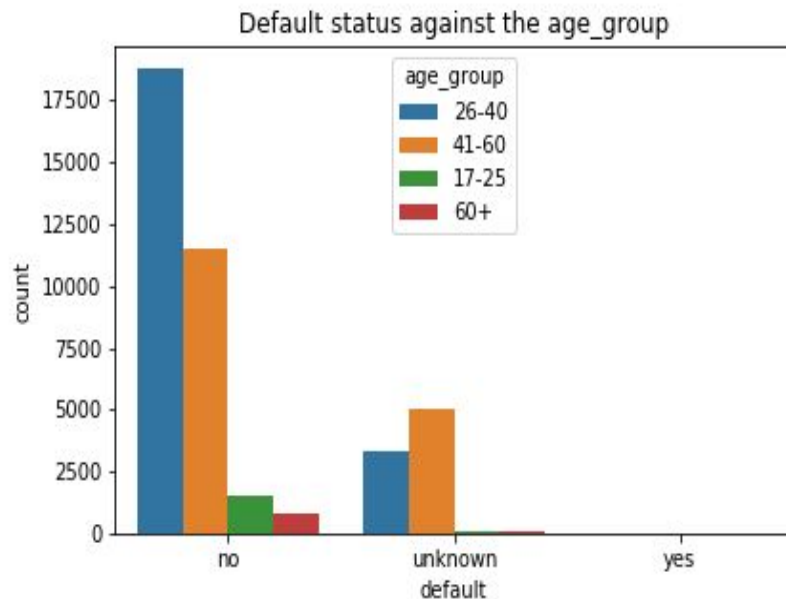
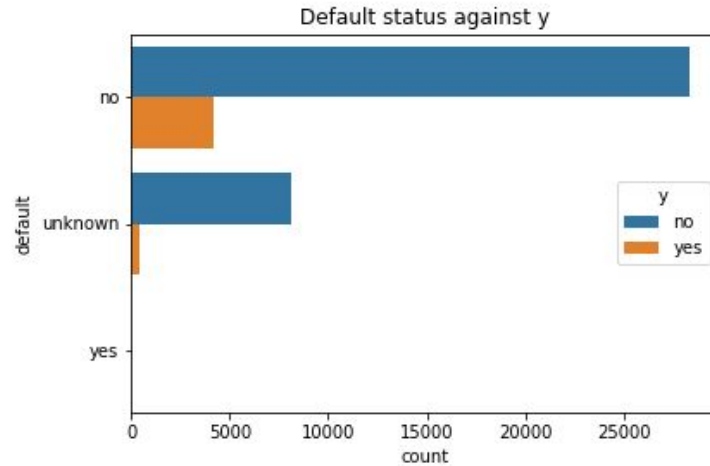
Education: Education levels of a customer

- Over 10000 customers with an education level of 'university degree' show rejection to the term deposit plan.
- There are close to 1000 customers with an unknown education level.



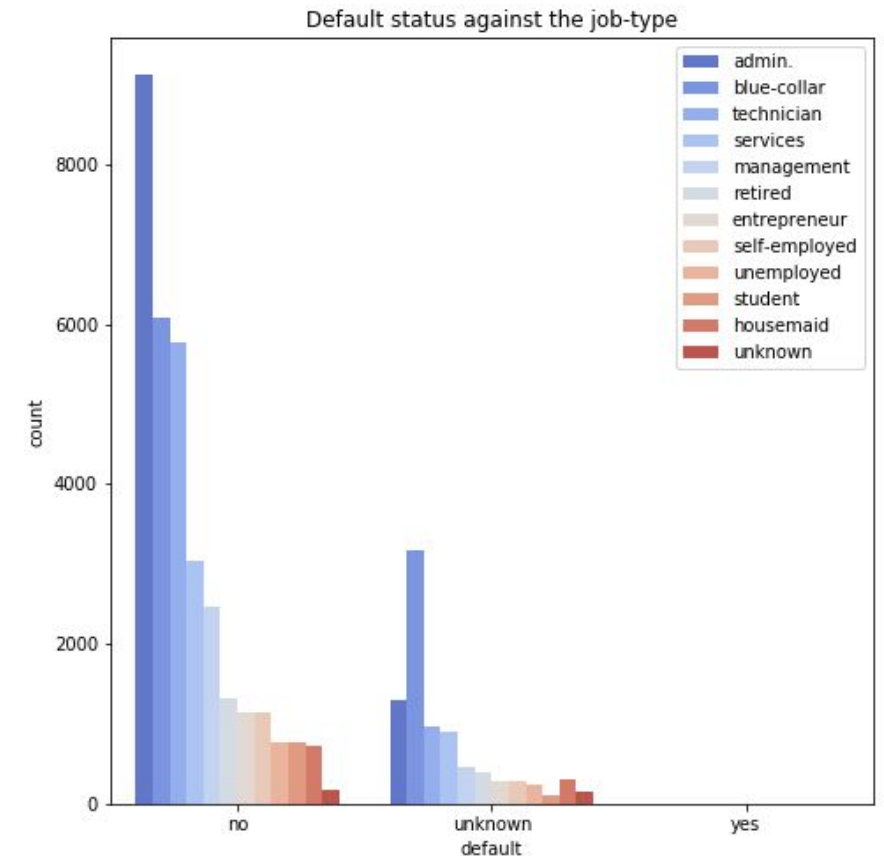
Default: The default status of customer

- Over 50% of the customer base have no credit in default. Among them, 25000 have not accepted the term deposit plan.
- Close to 7500 customers have not declared their default status.

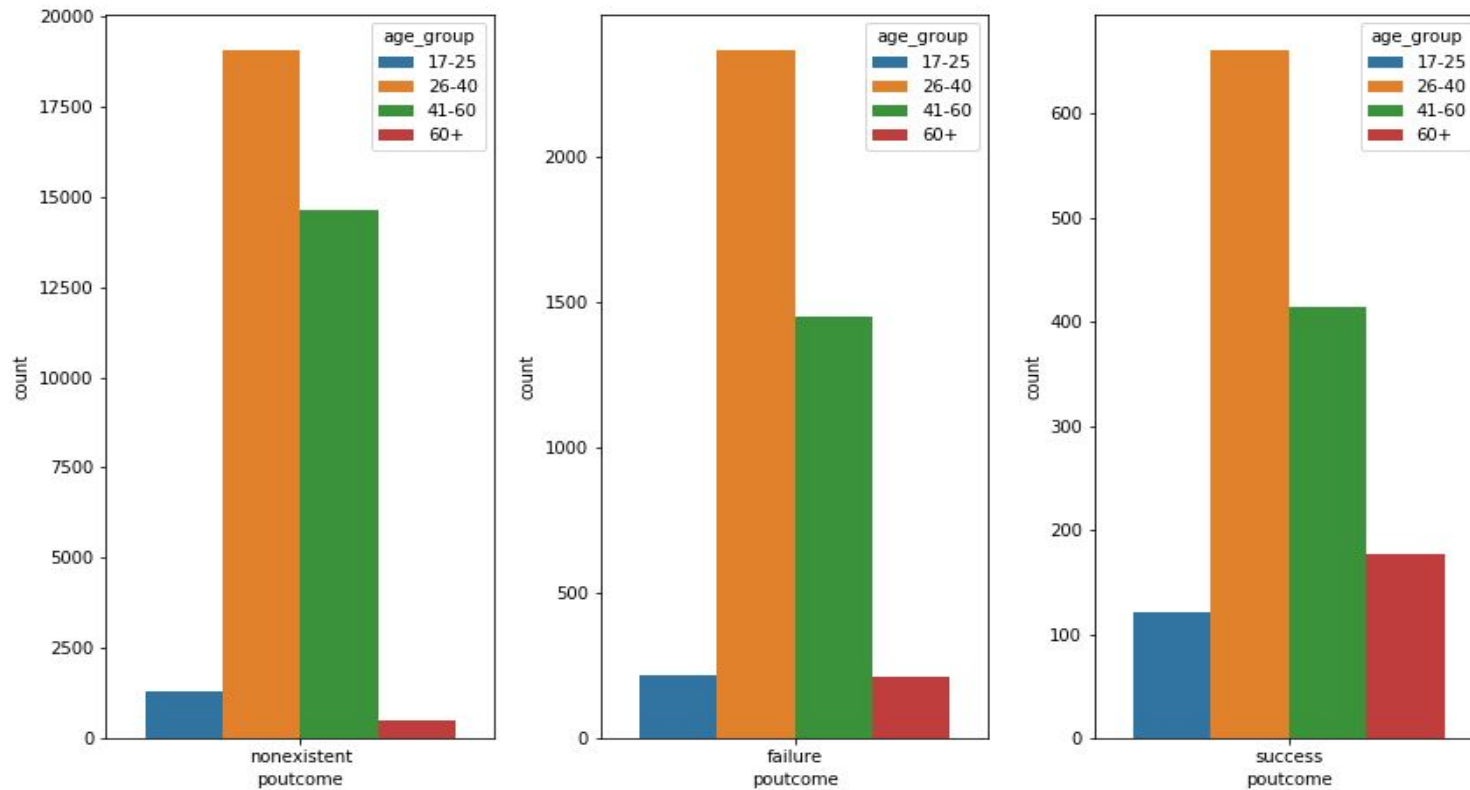


- The 26-40 and 41-60 age groups have an unknown default status.
- Similarly, these two groups are in a majority among customers who have no defaults.

- The professions of admin, blue-collar and technician are greater than 20000 in number who are no defaulters.
- In addition, 50% of blue-collar professionals have an unknown default status.

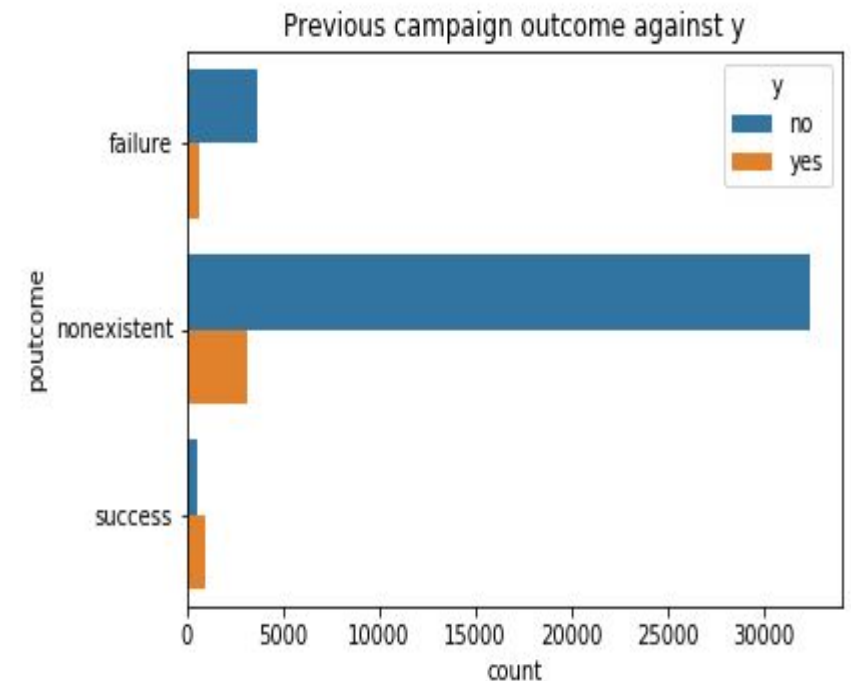


Poutcome: Outcome of the previous marketing campaign for the customer

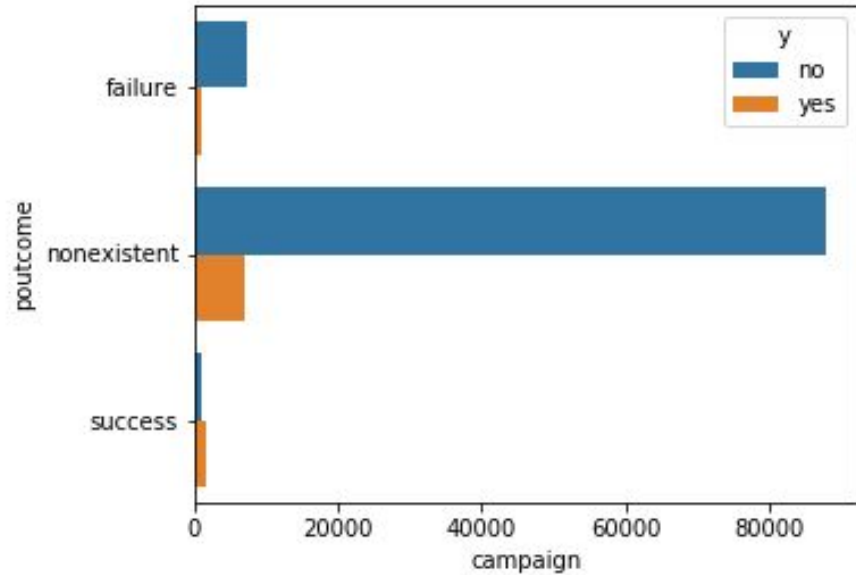


→ Out of 35000 customers for whom the previous campaign was non-existent, 31000 have rejected the term deposit plan.

- Greater than 90% of customers for whom the outcome of the previous campaign was non-existent are in 26-40 and 41-60 age groups.
- Both age-groups follow a similar trend for the rest of the outcomes.
- On the other hand, the previous campaign was a success for close to 1500 customers.

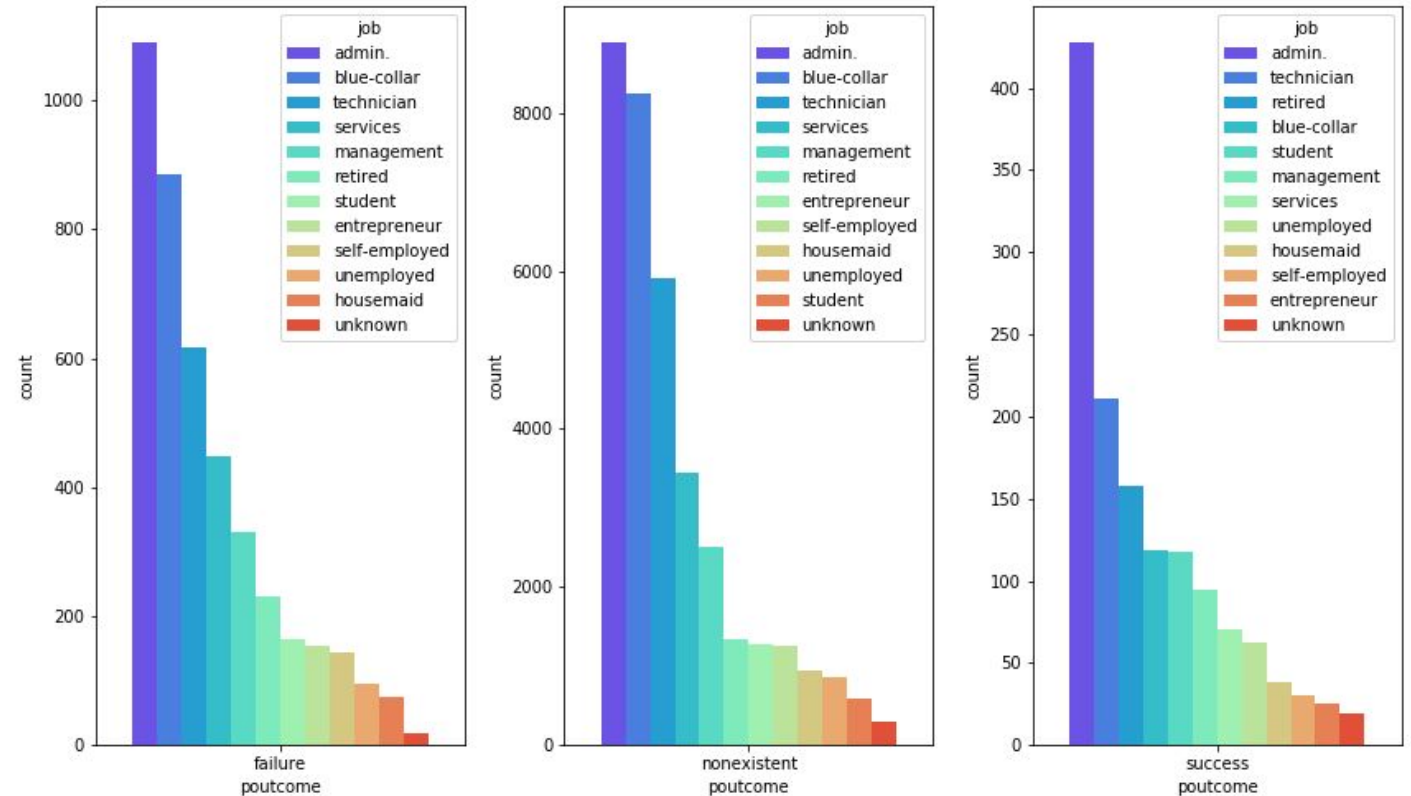


No. of contacts performed in this campaign against poutcome and



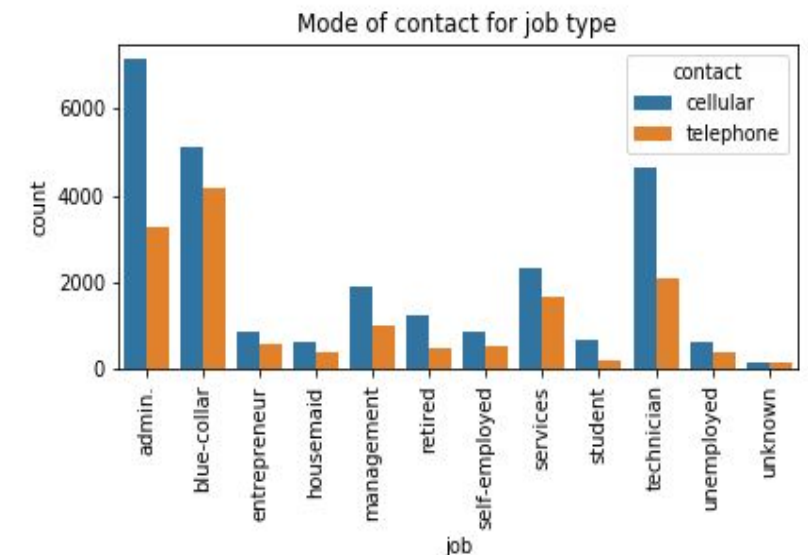
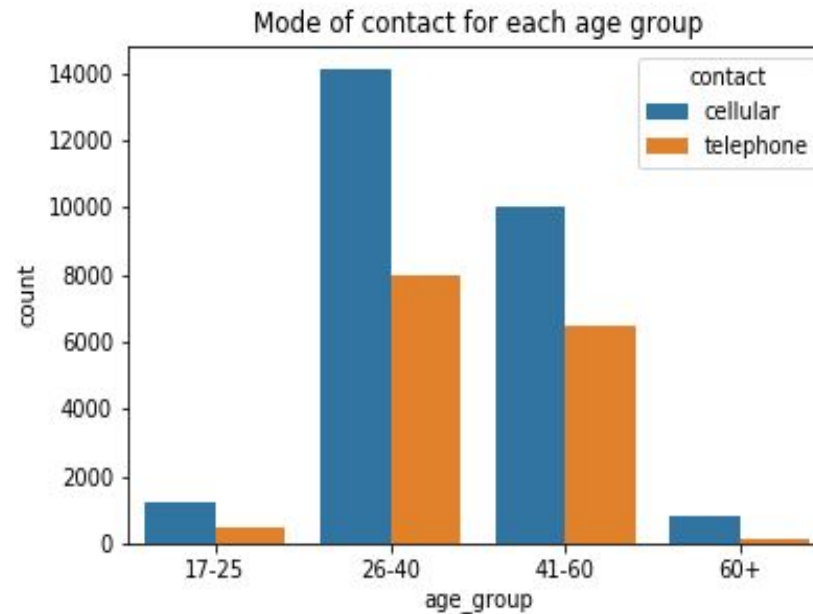
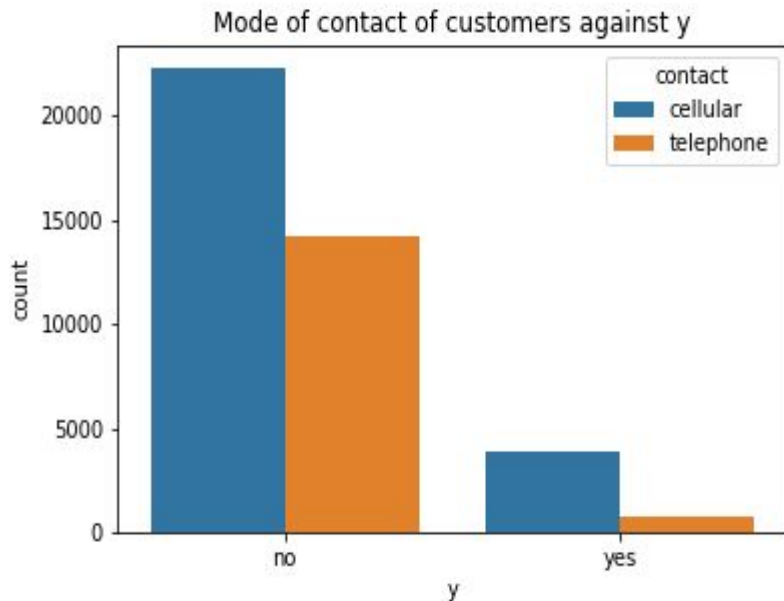
- Customers in jobs of admin, blue-collar and technician for whom the previous campaign outcome was non-existent are more than 20000 in number.
- However, for the outcome of success, the number of admin professionals are almost double the number of the next job category.

- For all non-existent outcomes, close to 85000 contacts have been made by the marketing team.
- This dwarfs the small number of contacts made to customers belonging to the other two outcomes of the previous campaign.



Contact: Mode of communication for the customer

- Out of the 37000 customers not opting for the term deposit plan, 22000 of them had been contacted through cellular communication.
- For the 26-40 age-group, almost double the number of customers were contacted through cellular means than telephone. In addition, for 41-60 yrs, there is a gap of 4000 in the number of customers contacted through the two modes of communication.
- For all job types, cellular mode seems to be the preferred mode of communication. For example, close to 7000 customers in admin roles are communicated through cellular means. This number is more than double the number contacted through telephone.



EDA Summary

What general recommendation can be offered to improve the campaign and make it more successful ?

Micro targeting can help increase the overall effectiveness of this marketing campaign.

1. Take into account the best time to contact potential clients (May is the most effective month).
2. Increase the time of contact with customers. It is possible to use other means of communication besides cellular.
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, 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.

Model Recommendation

What type of ML model to use ?

In order to predict the client subscription for a deposit term, we will use a predictive ML model to help us identify potential customers. As a start and after the performed EDA on the provided dataset, we choose to test out the following set of models since we don't know yet what algorithms will do well on this dataset.

The following algorithms selected include:

- **Linear Algorithms:**

Logistic Regression (LR)

Linear Discriminant Analysis (LDA).

- **Nonlinear Algorithms:**

Classification and Regression Trees (CART),

Support Vector Machines (SVM),

Gaussian Naive Bayes (NB)

k-Nearest Neighbors (KNN).

- **Ensemble Methods:**

Boosting Methods: AdaBoost (AB) and Gradient Boosting (GBM).

Bagging Methods: Random Forests (RF) and Extra Trees (ET).

Thank You