



**Data Glacier**

Your Deep Learning Partner

# Exploratory Data Analysis

**Project: Bank Marketing Campaign**

**Name: Ishwar Balasaheb Bankar**

**Batch Code: LISUM28**

**Specialization: Data Science**

# Agenda

Problem Statement

Dataset

EDA

Recommendations

# Problem Statement

- 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 in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution). This is an application of the organization's marketing data.
- **Objective:** 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.

# Dataset

<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

**bank-additional-full.csv:** 20 inputs (+ 1 target variable) and 41118 observations

## Assumptions:

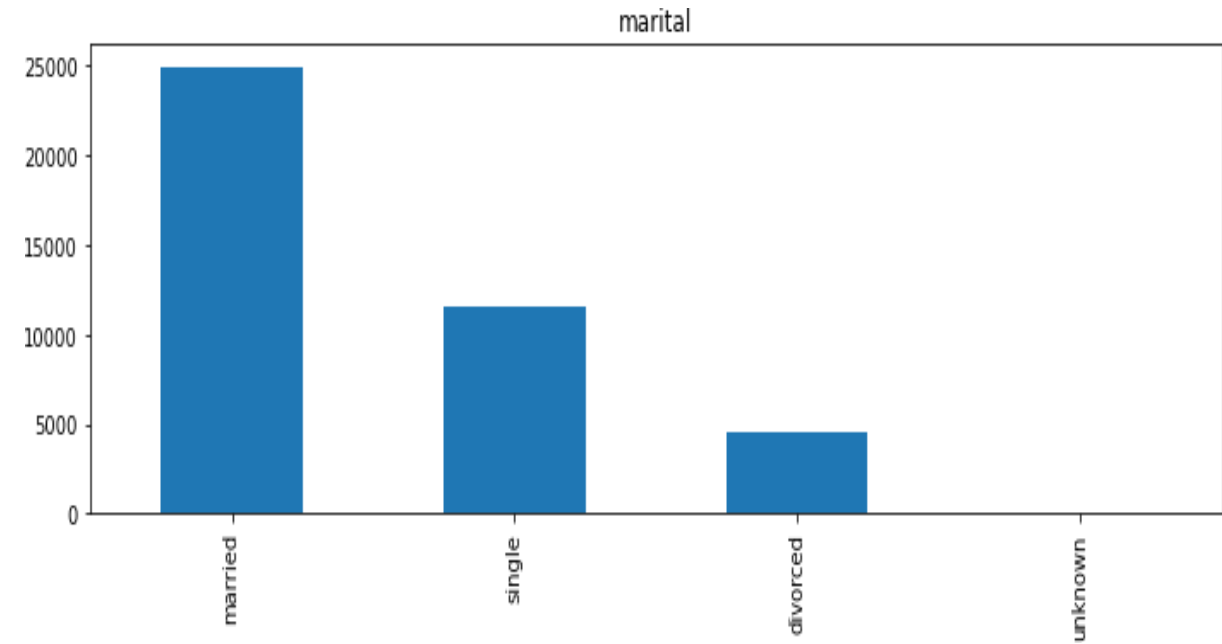
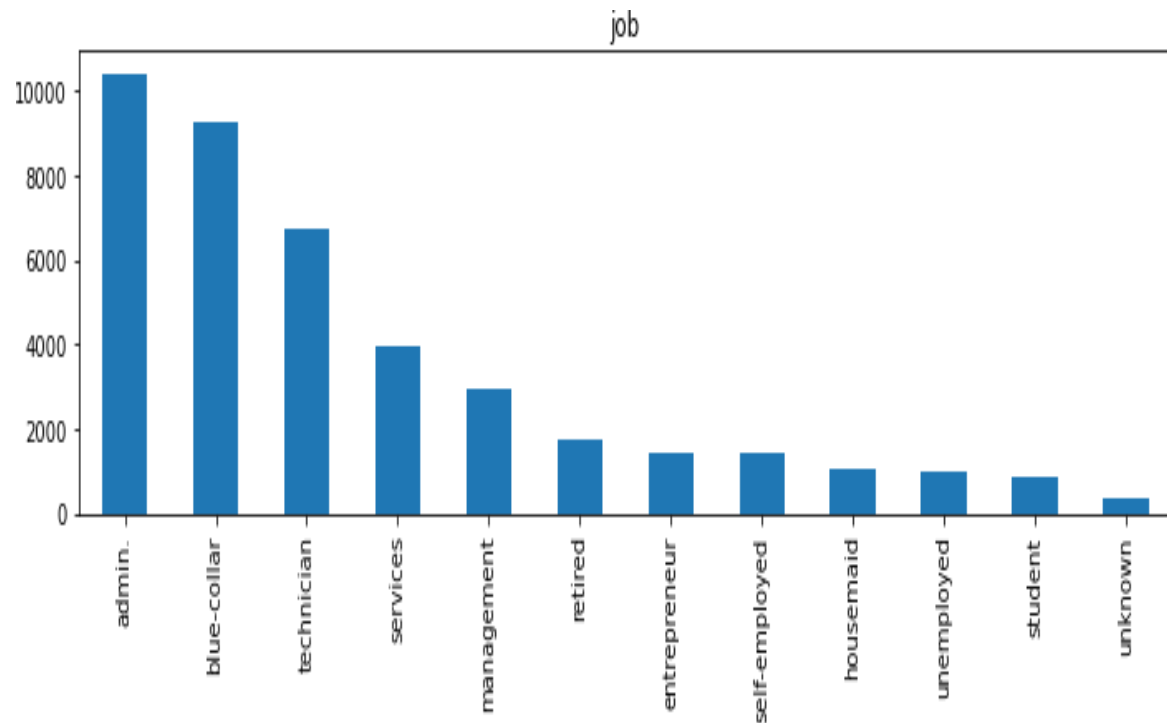
- 'Duration' feature is dropped as suggested in the dataset description
- A frequently occurring missing value 'unknown' is considered as another category for the categorical features.
- Duplicate rows were deleted from the dataset.

# EDA

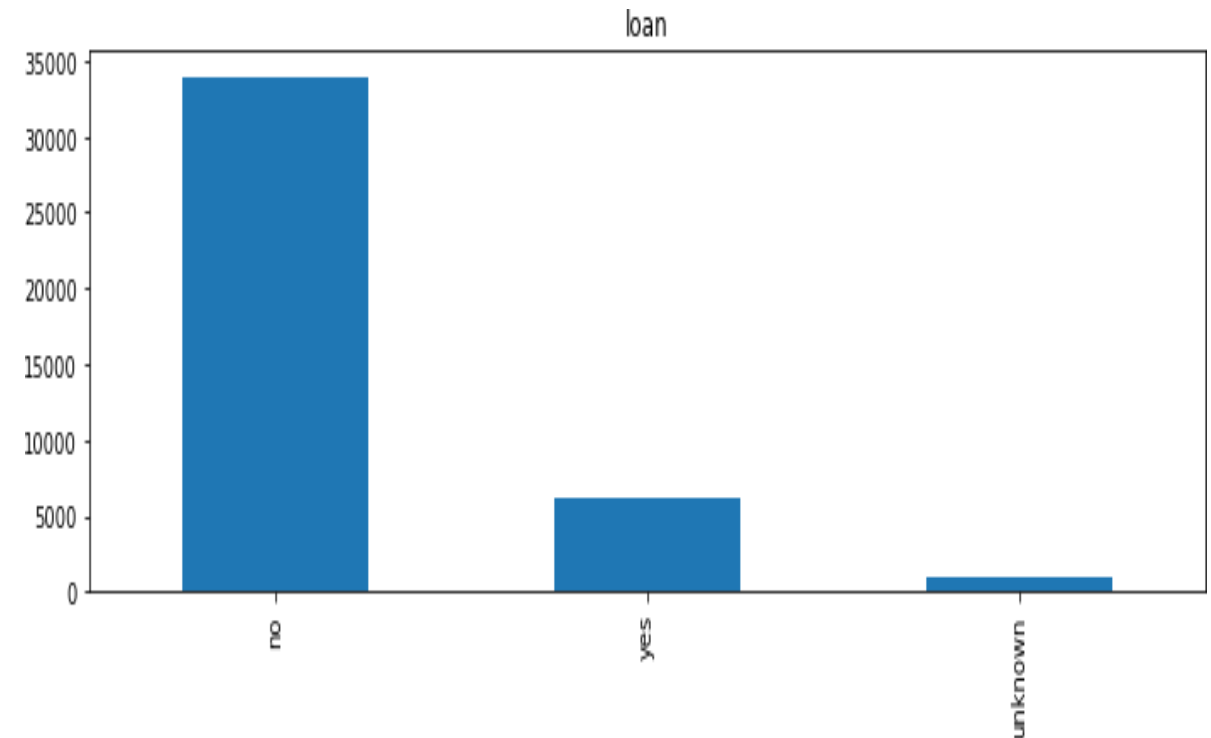
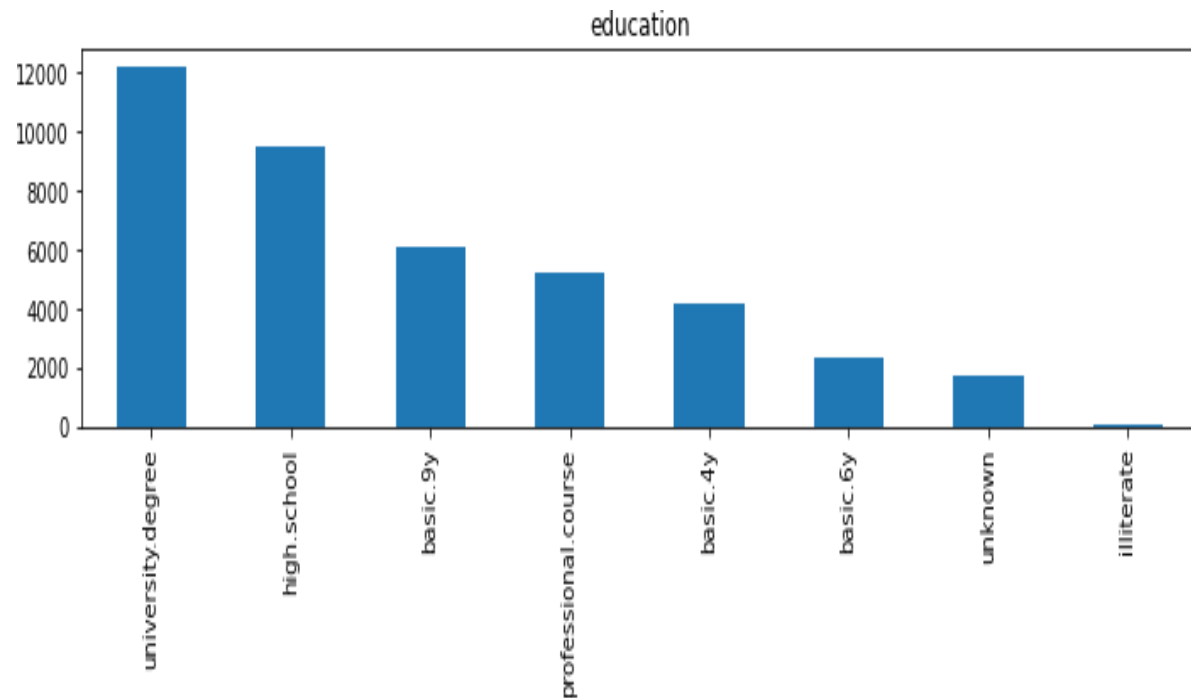
# EDA

## 1. Univariate Analysis

- In Jobs, most of the customers are administrative staffs and technicians.
- Married customers have been communicated more for subscription to the deposit more compared to single

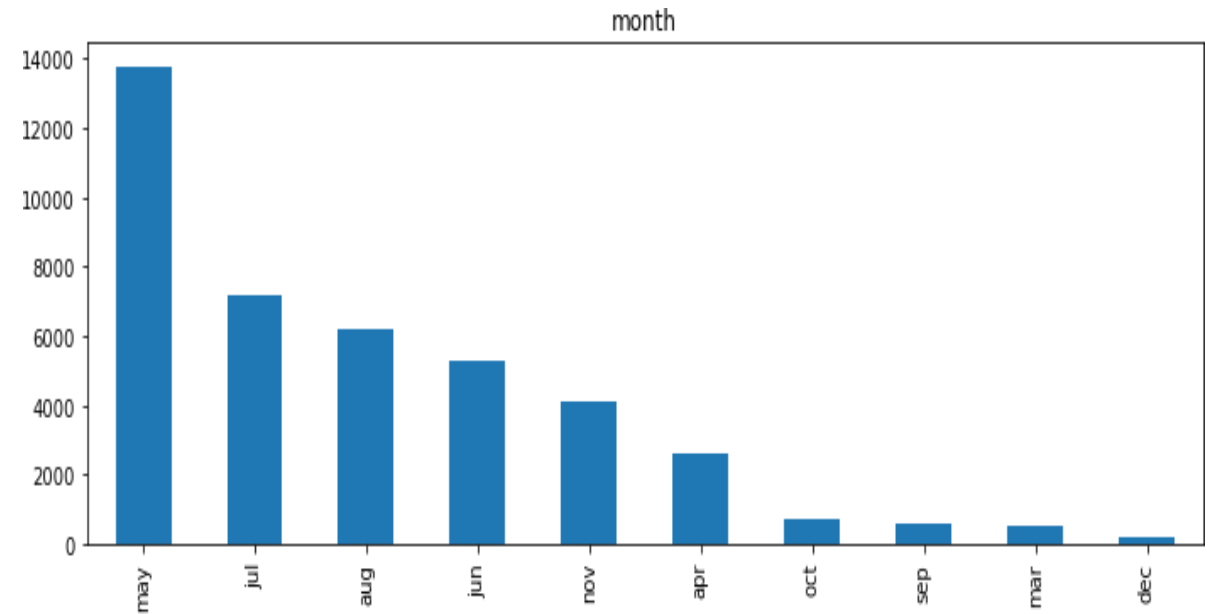
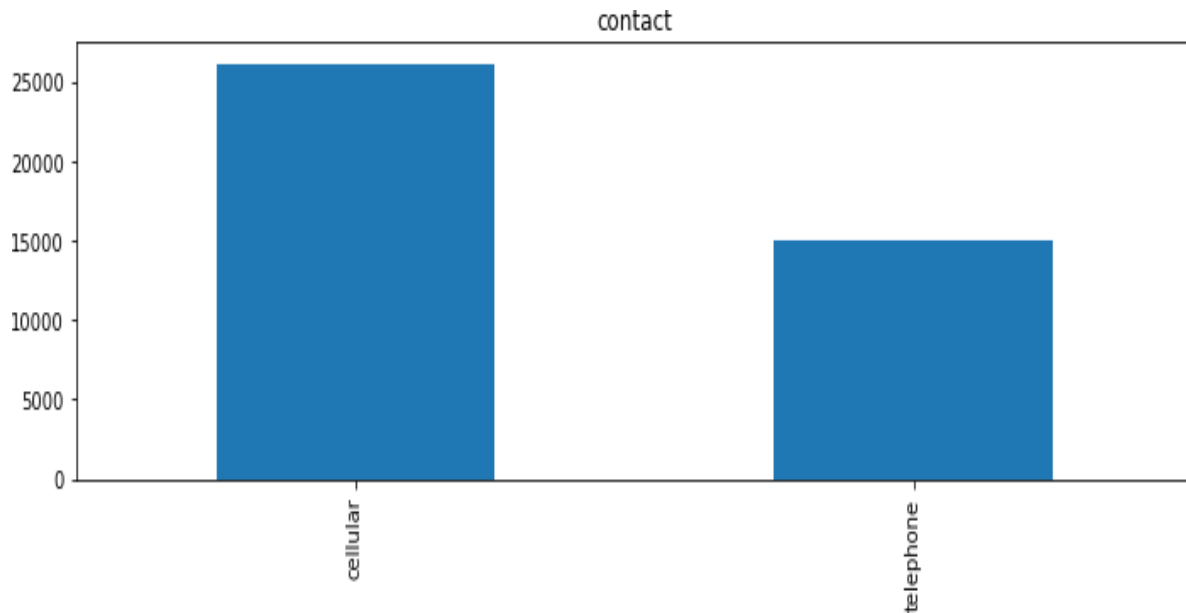


- Most of the customers are university graduates followed by high school degree
- Majority of the customers do not have a personal loan

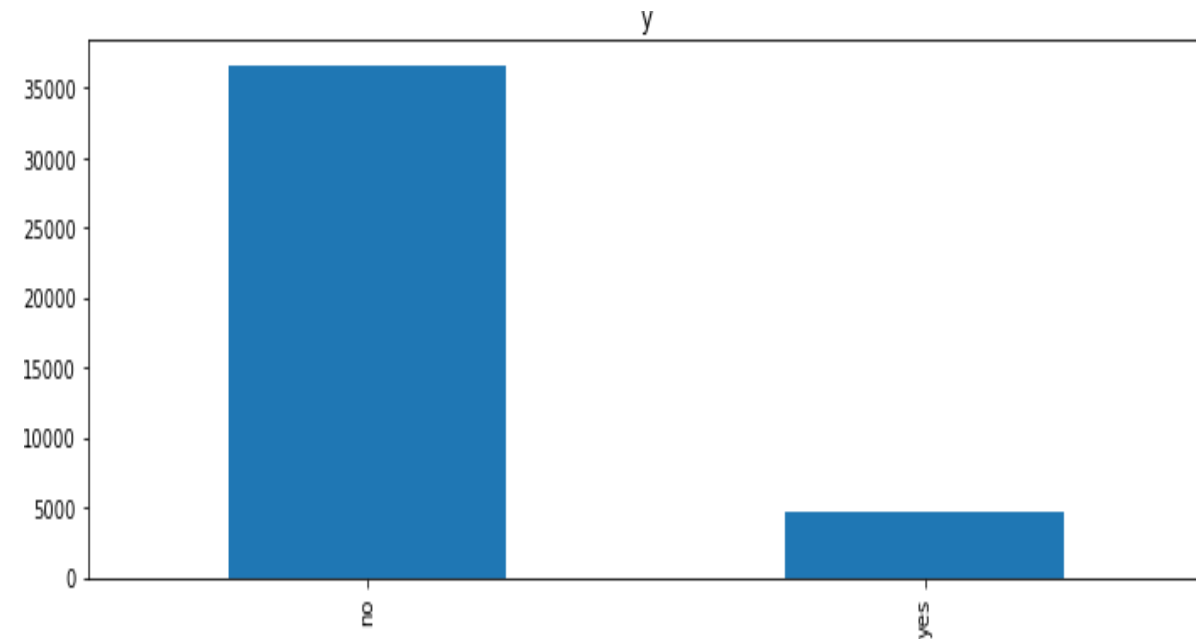
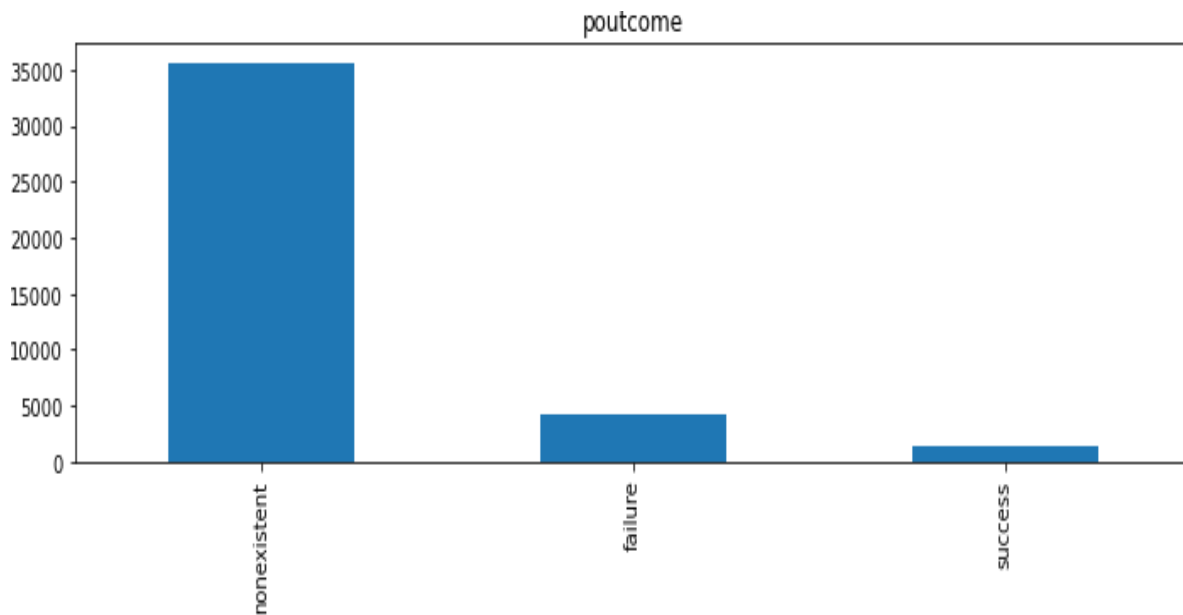




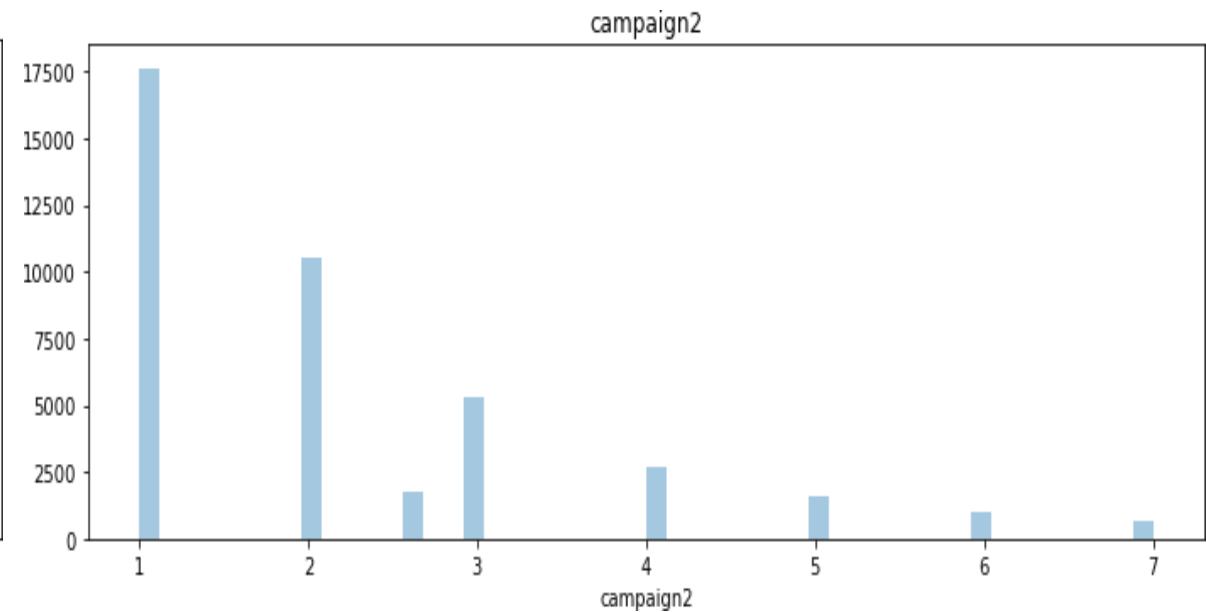
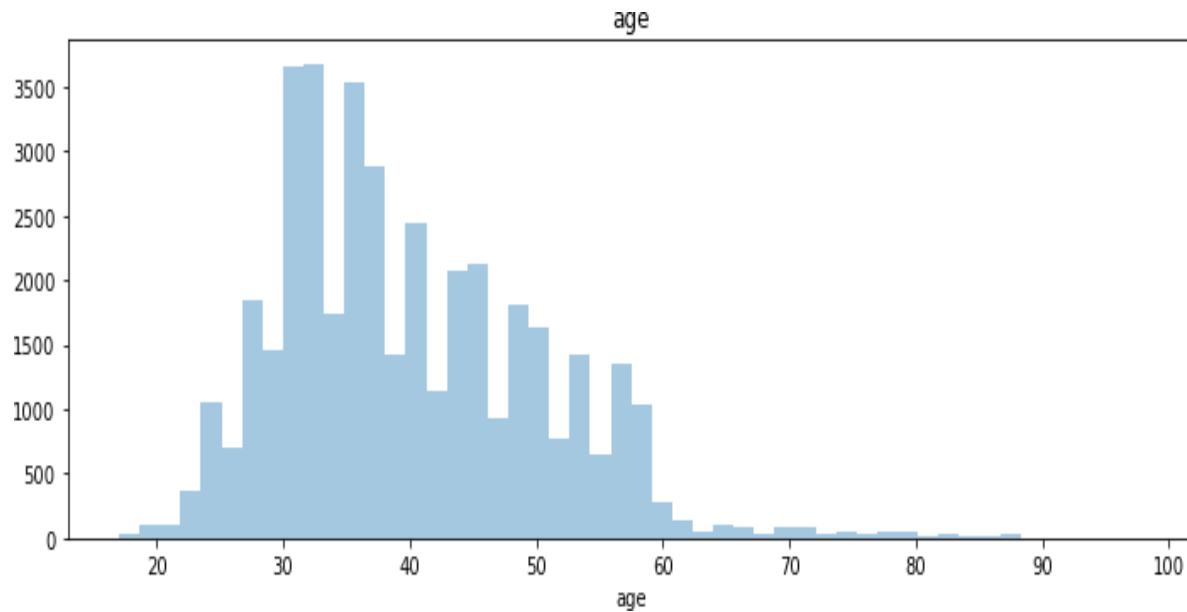
- The type of communication used to contact customers is mostly cellular compared to telephone
- May seems to be the month with most contacts made



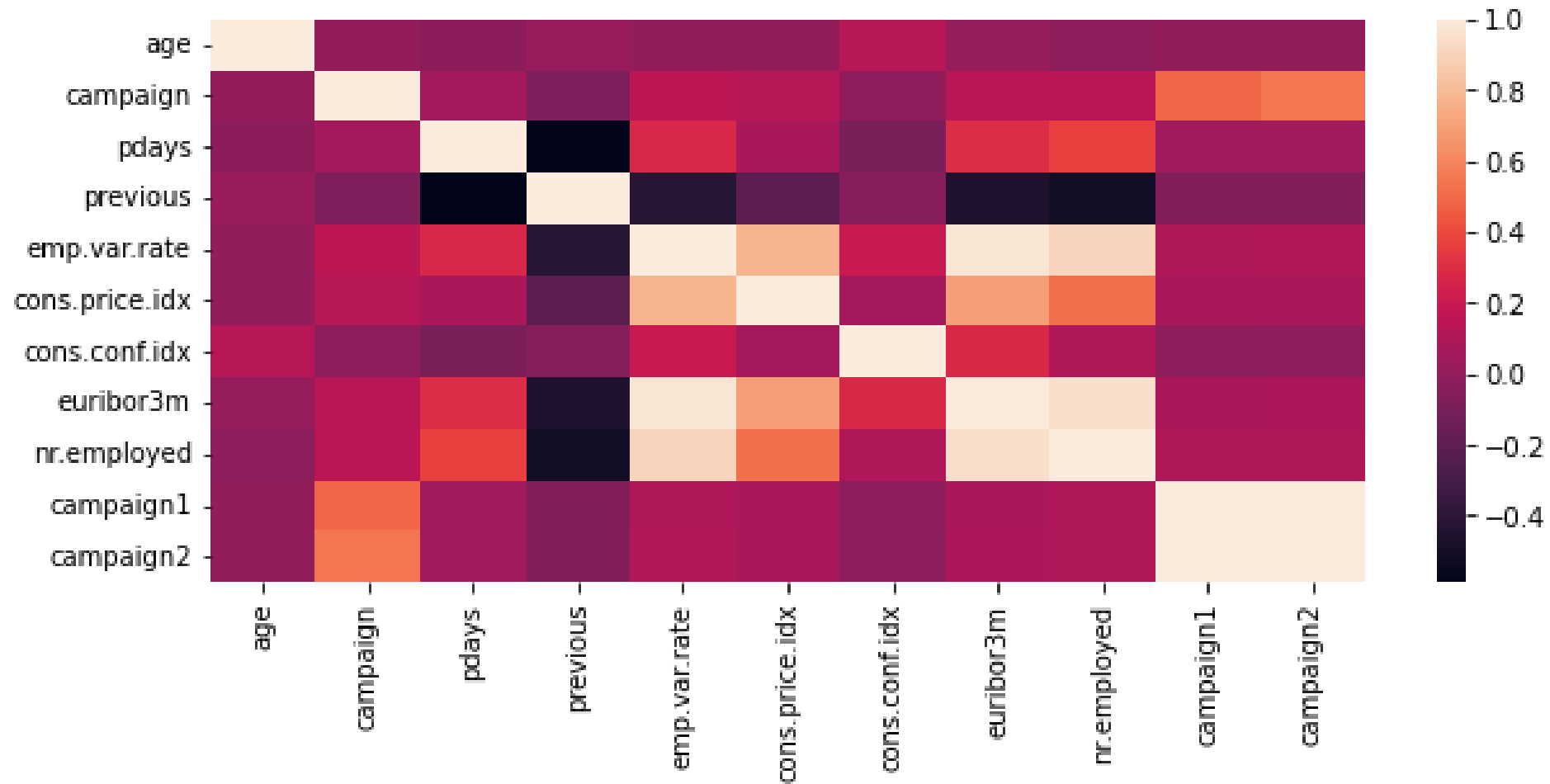
- Most of the previous campaign results whether succeeded or failed is non existent
- In the previous campaign, the percentage of people who subscribed to the deposit is less to those who did not.



- The age group of customers contacted mostly fall between 20 to 60
- Number of follow ups made to a customer is less in the previous campaign



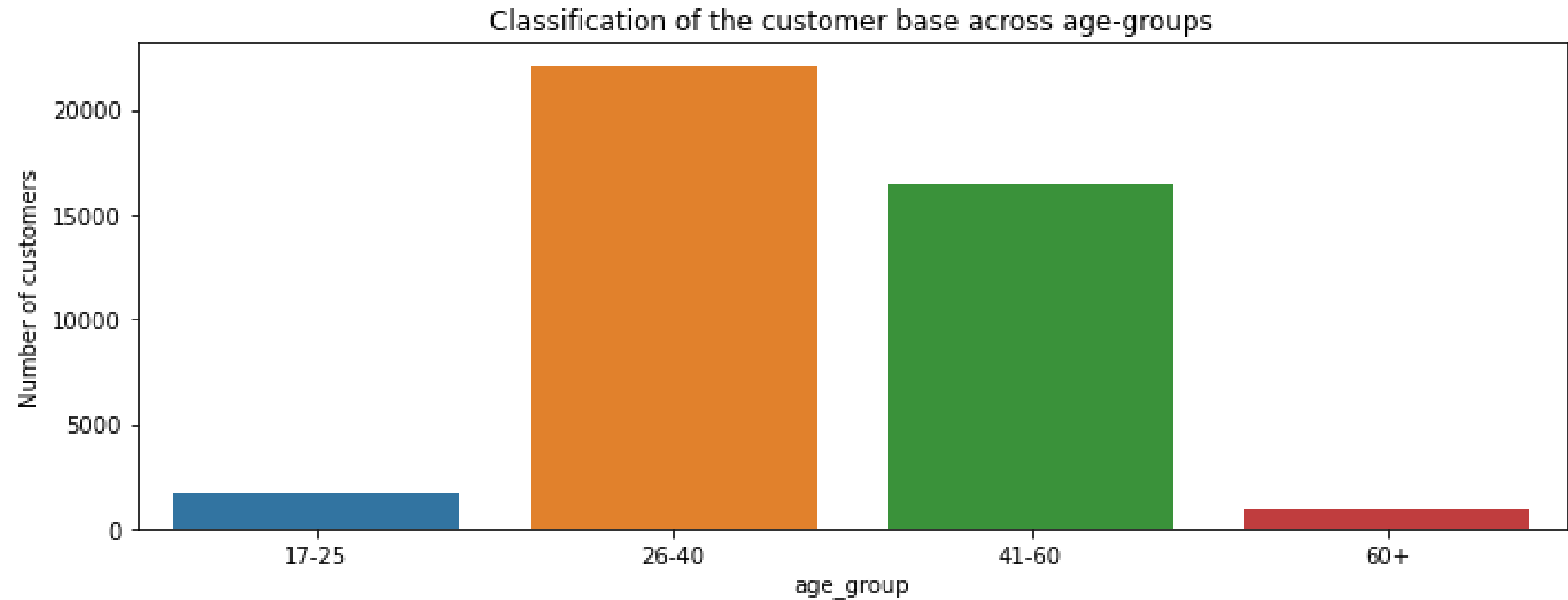
# Correlation map



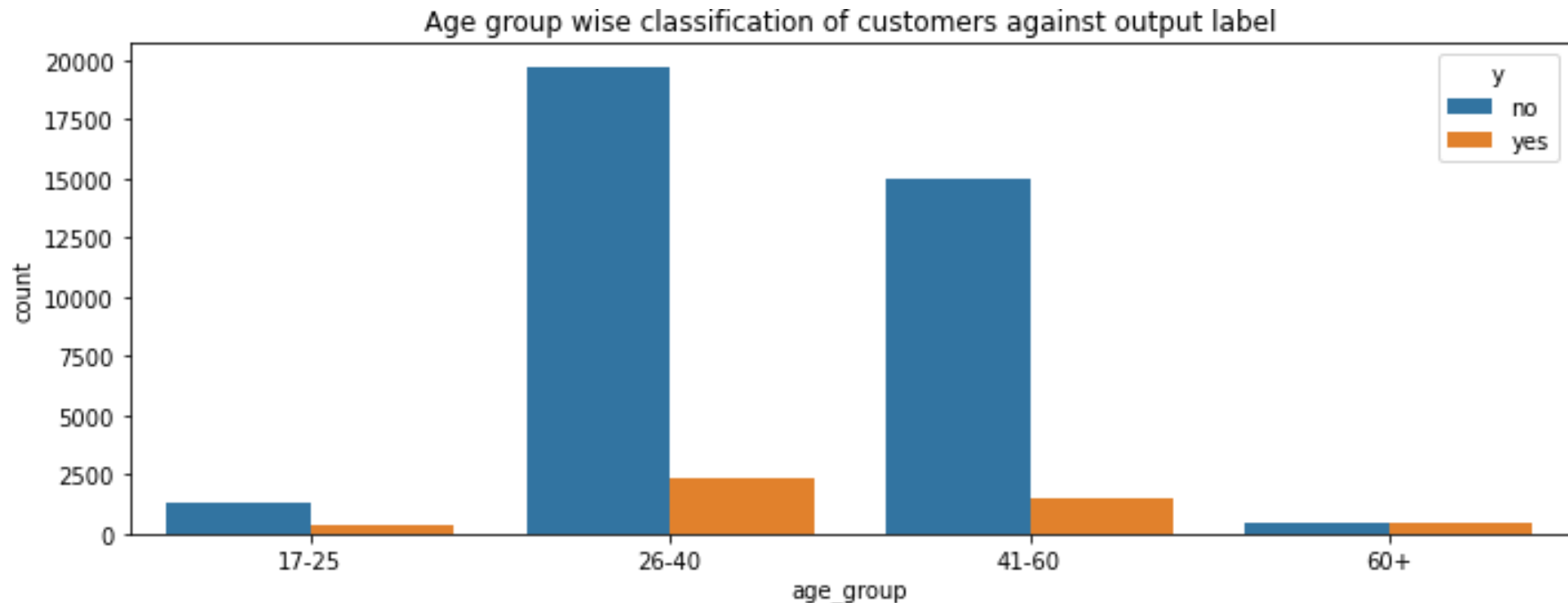
# EDA

## 2. Bivariate Analysis

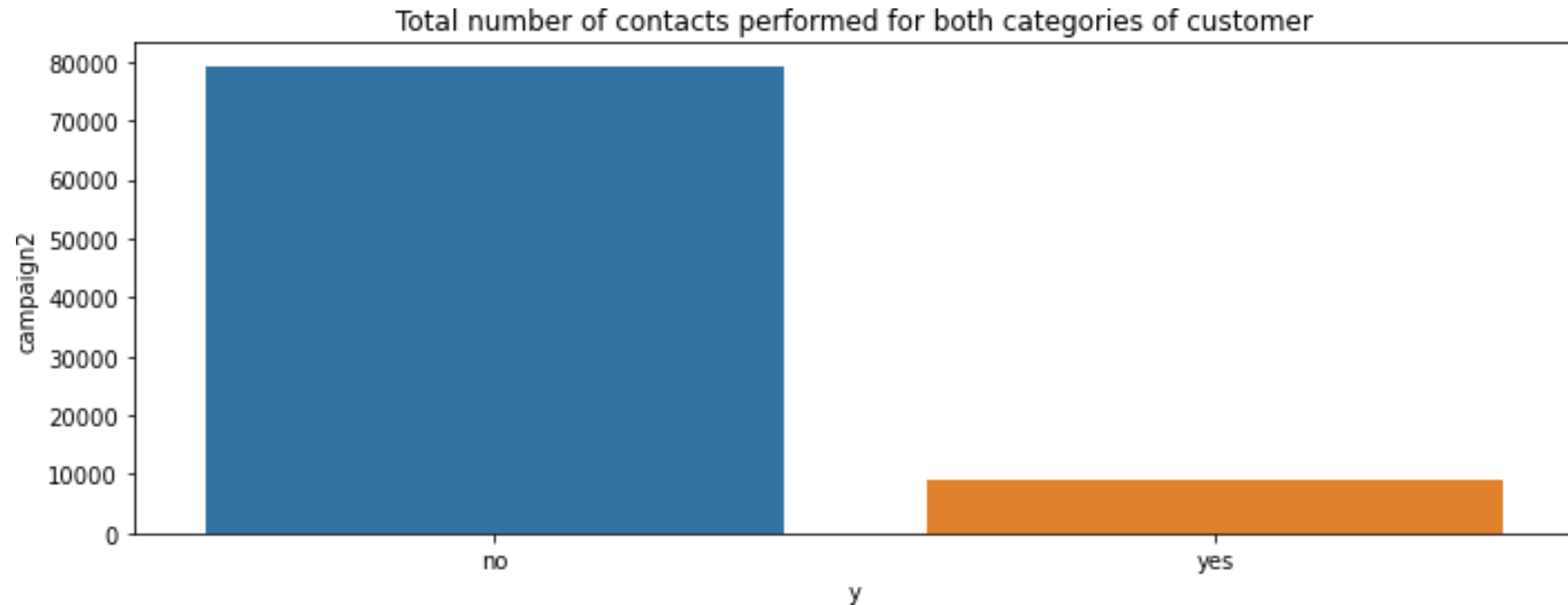
- Classification of the customer base across age-groups



- Looking at relation between different age groups and the output label y
- In the age-groups of 26-40 and 41-60 yrs, majority of the people are not subscribed to the term deposit plan

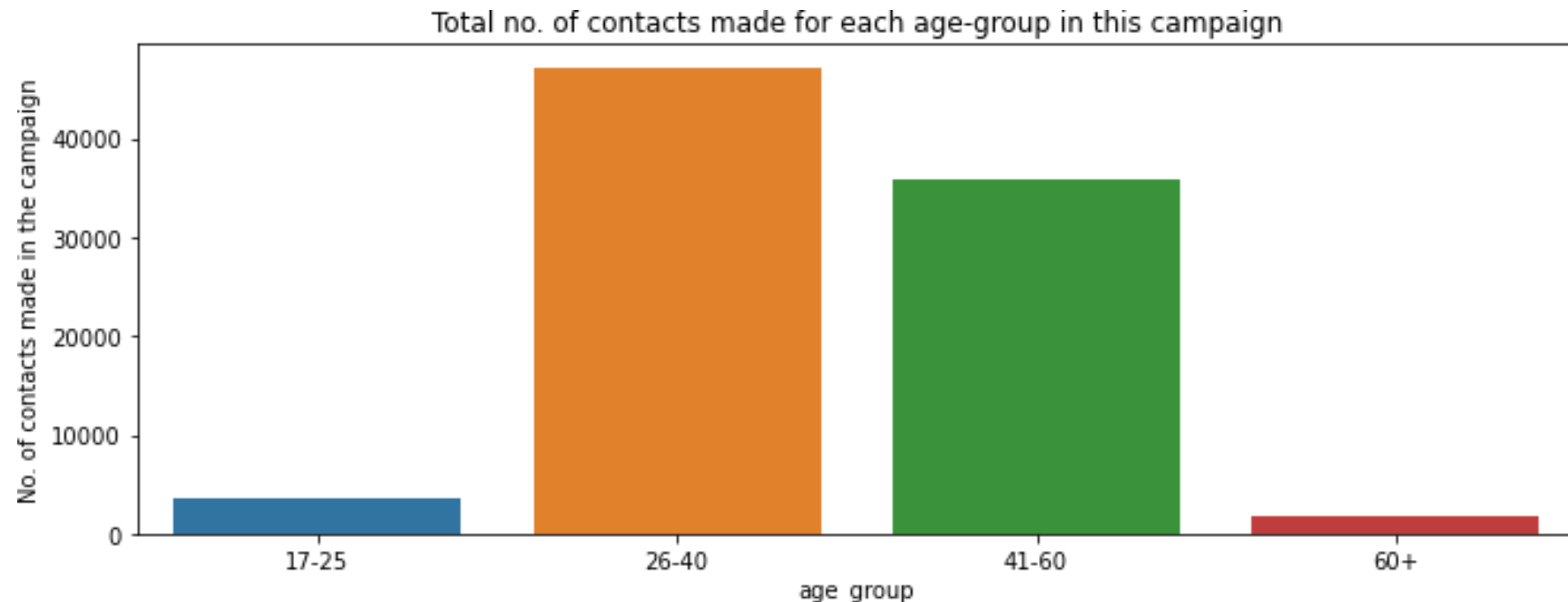


- Looking at relation between Number of contacts made to the customer (campaign) and the output label  $y$
- When a greater number of contacts is made to the customer, they haven't subscribed to the term deposit plan

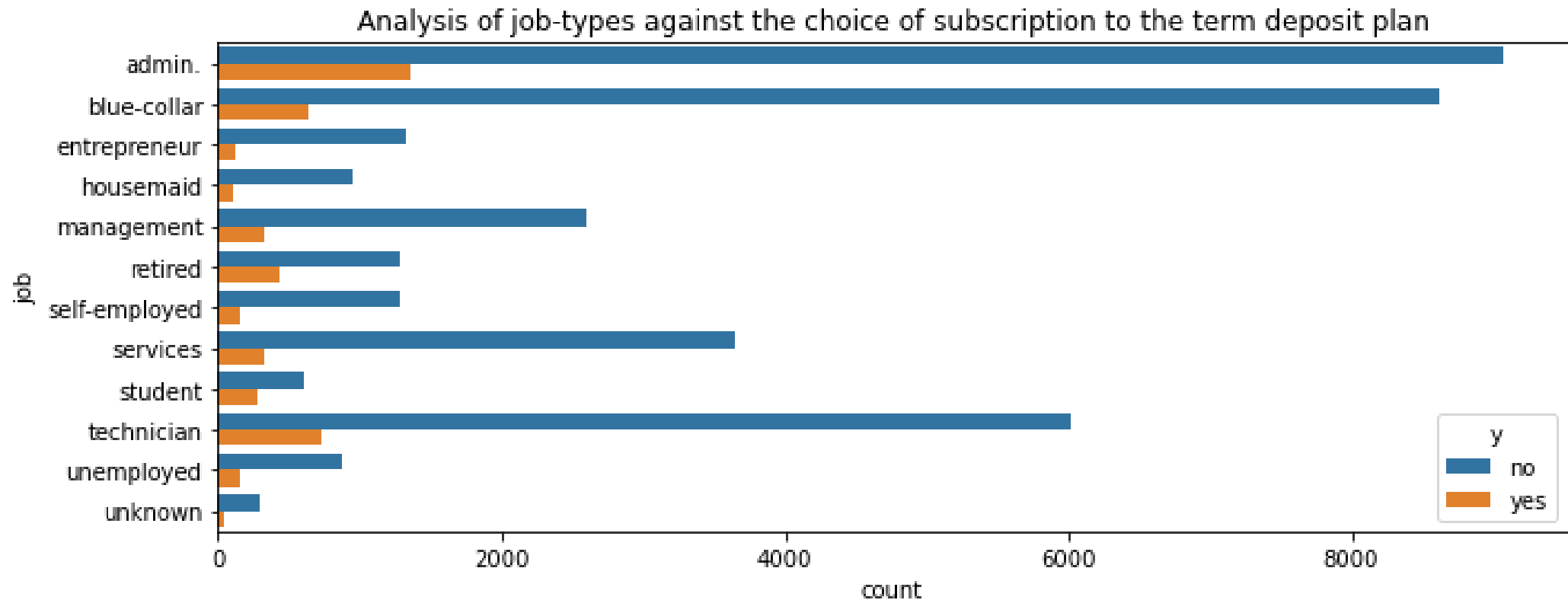




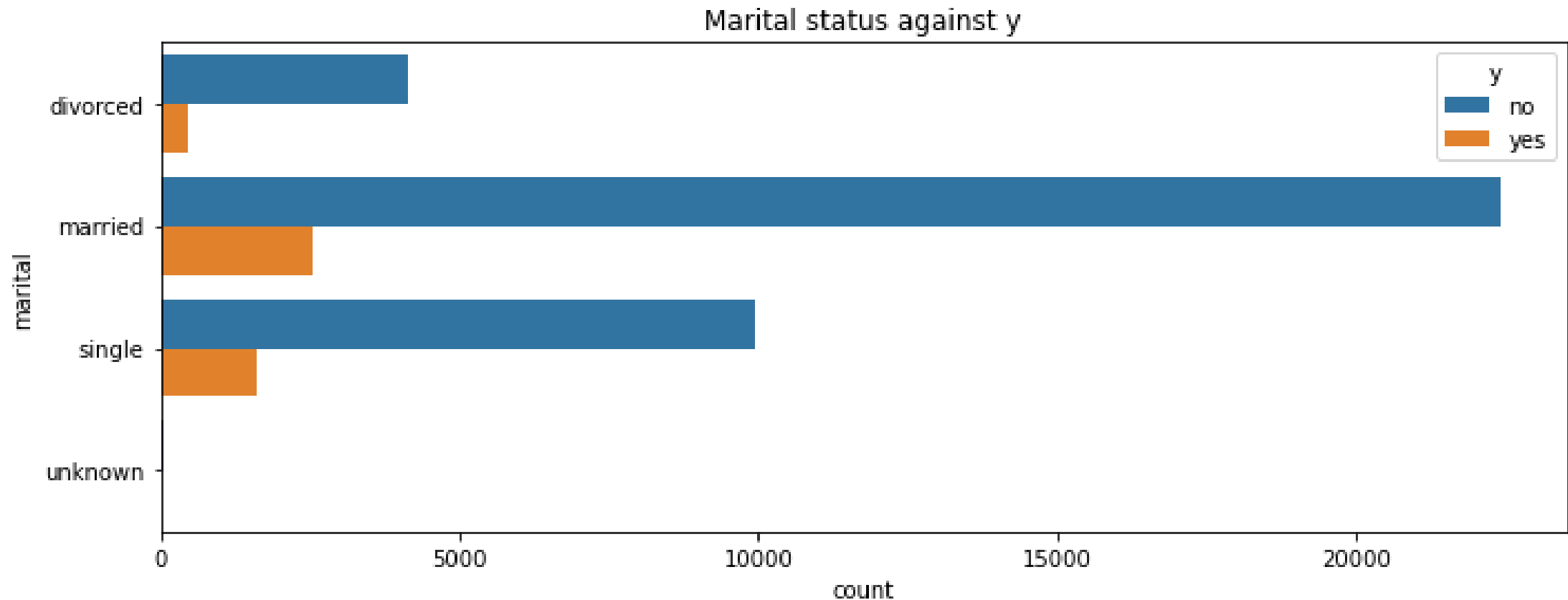
- Looking at relation between 'age\_group' and 'campaign' that is number of contacts performed for each age group
- The 26-40 and 41-60 age-groups witness majority of the contacts made in this campaign. These two age-groups seem to be the target groups for the bank.



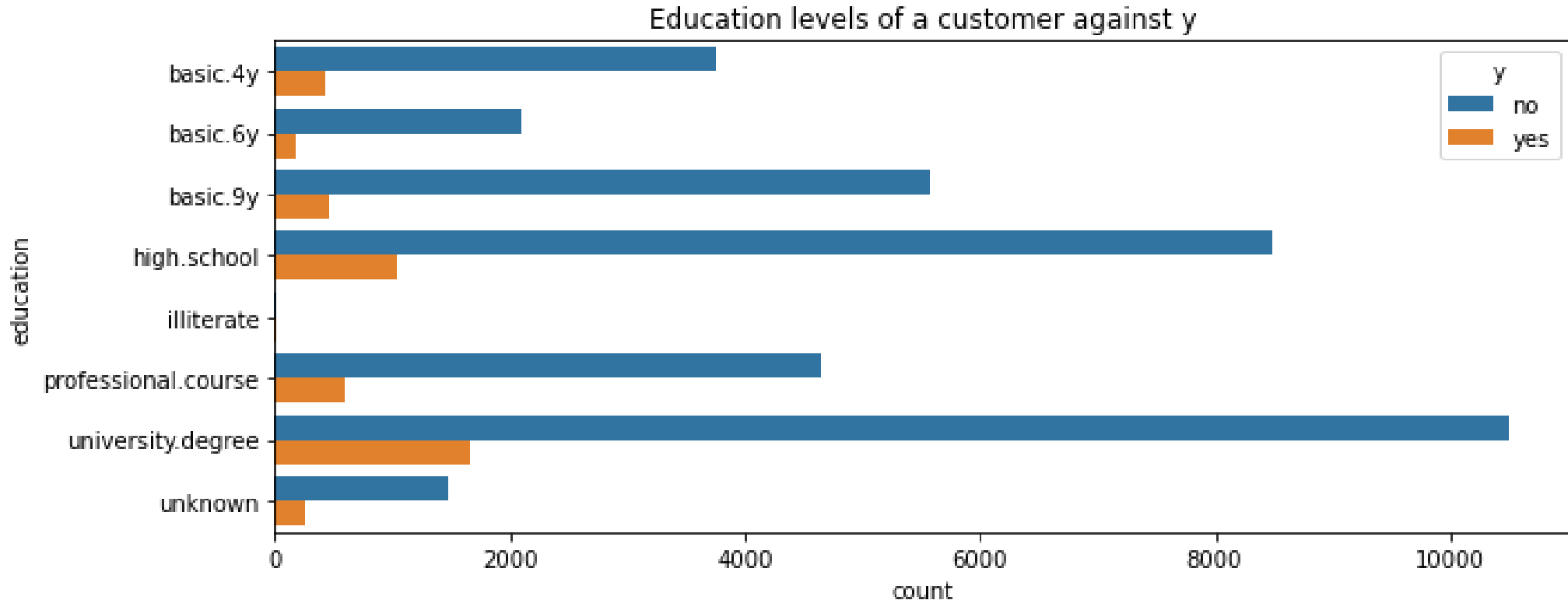
- Looking at relation between job and the output label  $y$
- Looking at the jobs, 'admin', 'blue-collar' and 'technician' are the prominent jobs and most of the customers in these jobs have rejected the term deposit plan.



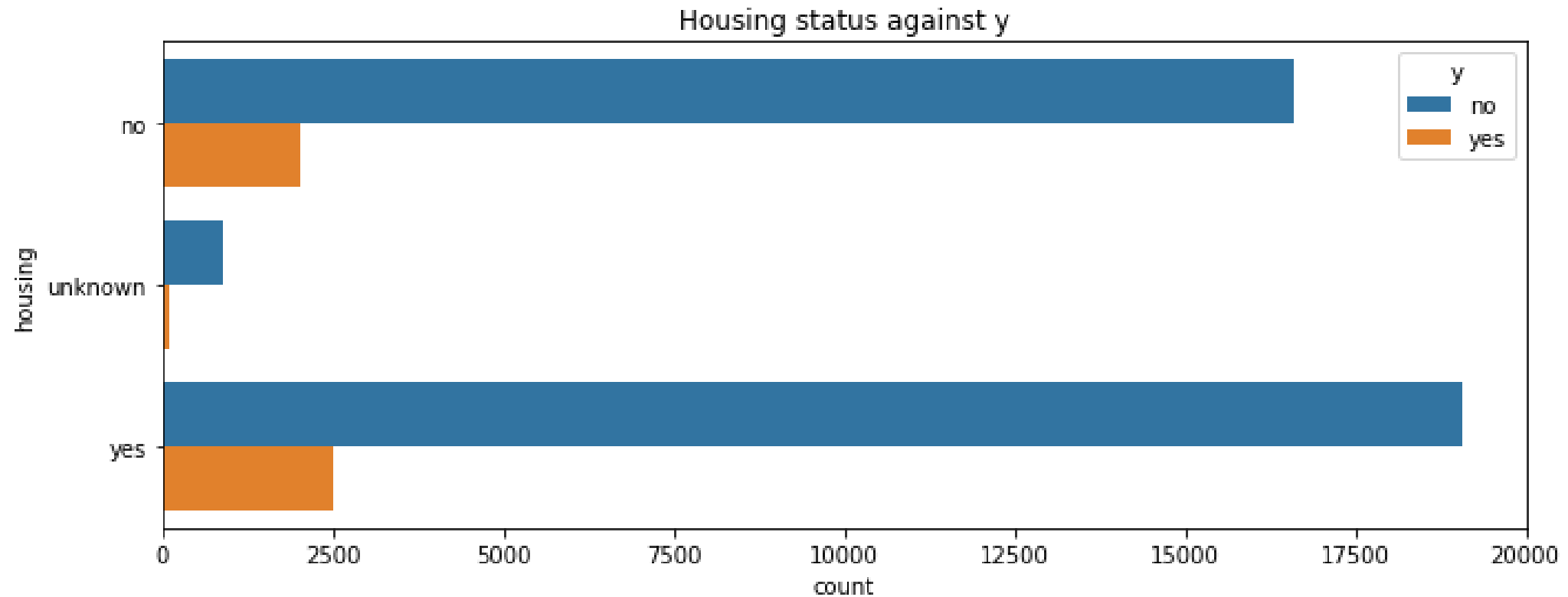
- Analysing marital status and the output label
- married and single customers are the majority of the customer base and comparatively married customers have taken the term deposit



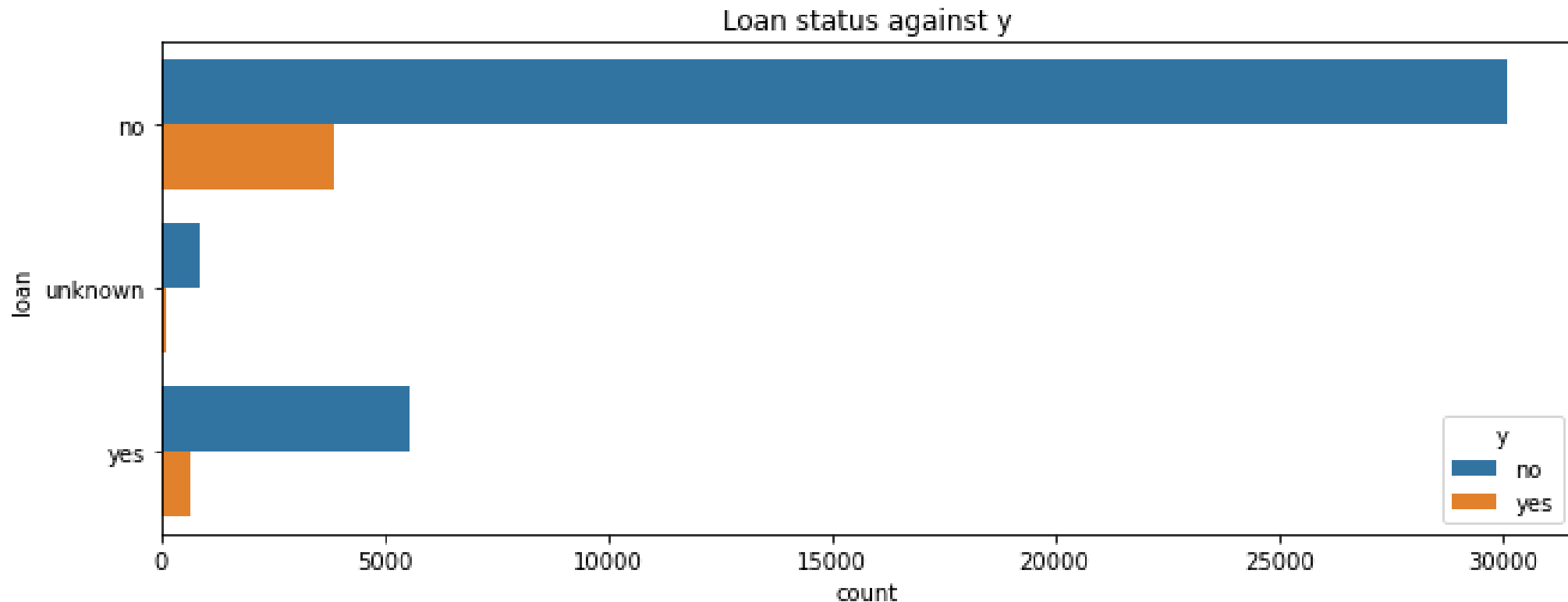
- Analysing the different education levels of a customer against the choice of subscription
- Customers with university degree have subscribed to the term deposit more



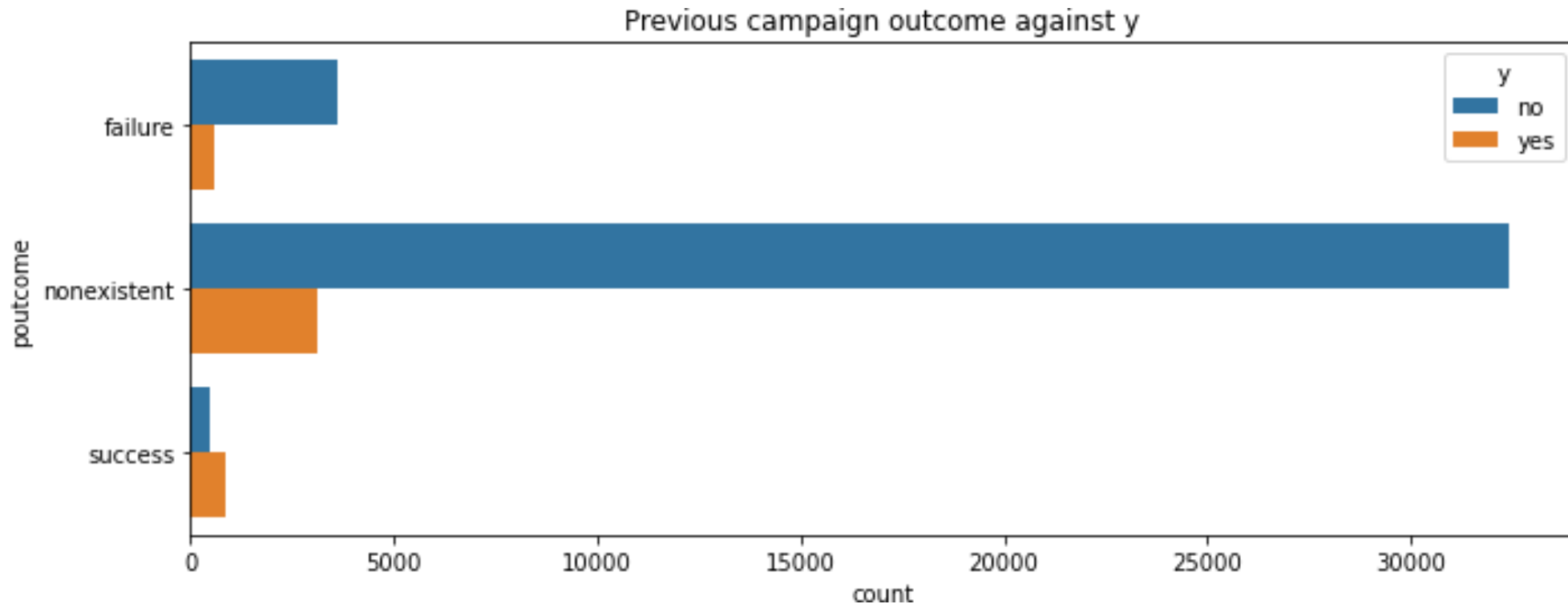
- Analysing housing status and y
- Number of customers who have subscribed to the term deposit is comparatively more for those with housing loan



- Analysing loan status and y
- Number of customers who have subscribed to the term deposit is comparatively less for those with personal loan



- Analysing poutcome and y
- The success rate of previous marketing campaign has resulted in more number of people subscribing to the term deposit



# Recommendation to improve campaign

1. **May is the most effective month to contact customers**
2. **Increase the time of contacts made per customer**
3. **Give more focus on university graduate students and high school degree students**
4. **age-groups of 26-40 and 41-60 have a higher proportion among customers, therefore these groups present a profitable target for the marketing team.**
5. **Target the admins and technicians for more subscriptions**



# Recommended models for this dataset

1. **Logistic Regression**
2. **Naïve Bayes**
3. **Decision Tree**
4. **Random Forest**
5. **Ada Boost**

**Hyper parameter tuning and model evaluation will be performed in order to determine the best model and the important features**

# Thank You