

Analyzing a Bank Marketing Campaign Dataset

And creating a machine learning model for opening a term deposit





Description of the Dataset

- The given dataset contains the results of Portugal bank marketing campaigns.
- The goal of the marketing campaigns was to find customers who are willing to open a term deposit.
- The conducted campaigns were mostly based on direct phone calls, offering bank clients to place a term deposit.



Source of the Dataset

- Direct source:
 - <https://www.kaggle.com/volodymyrgavrysh/bank-marketing-campaigns-dataset>
- Original source:
 - <https://archive.ics.uci.edu/ml/datasets/bank+marketing>
- Citation of the source
 - This dataset is public available for research. The details are described in S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

Exploratory Data Analysis (EDA)



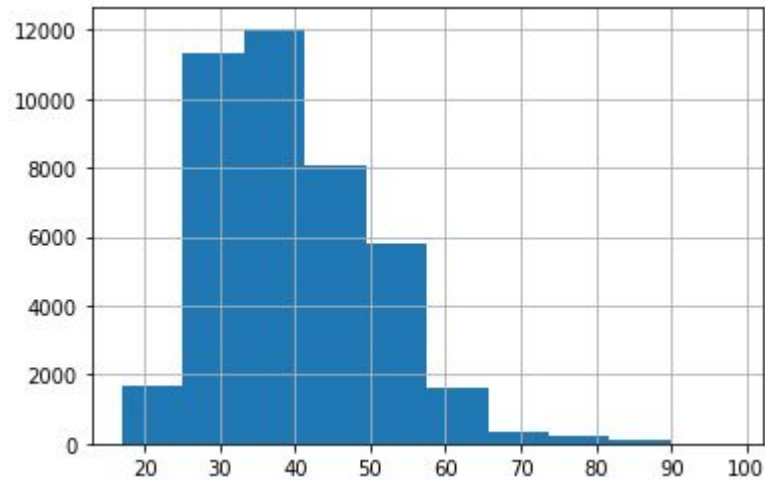
Only a small TEASER

More examples, more details and more context, see:

[ipynb-files](#)

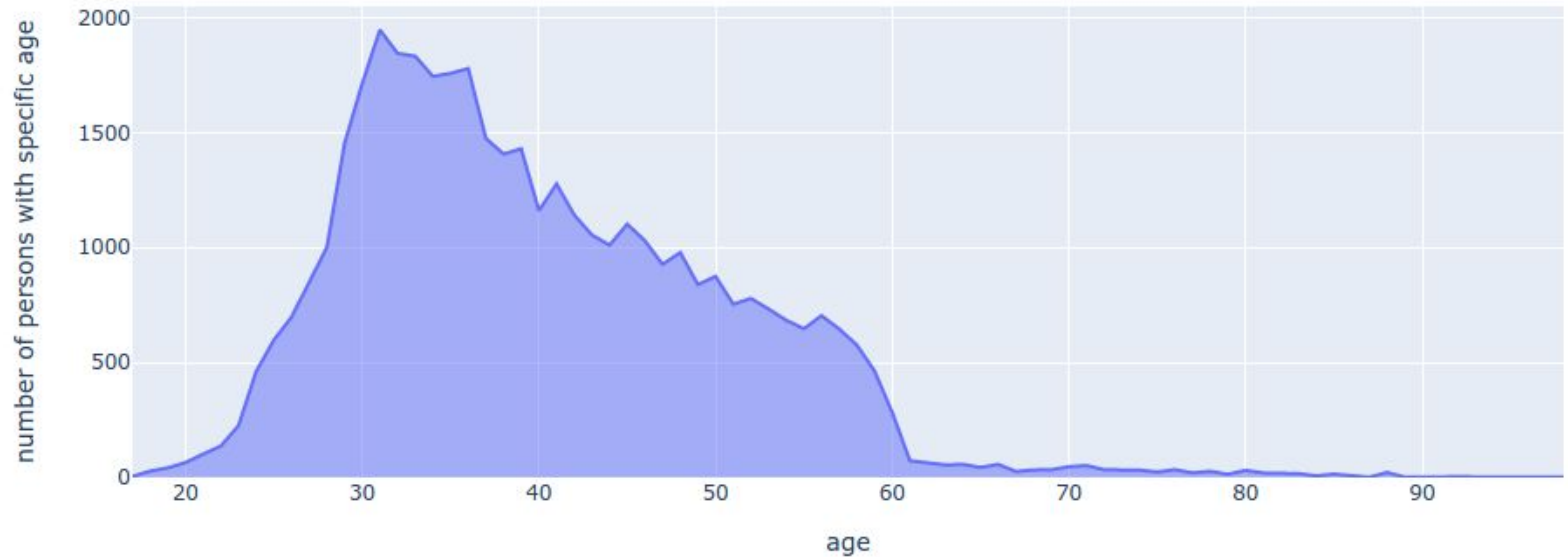


Histogram of the age of the contacted customers

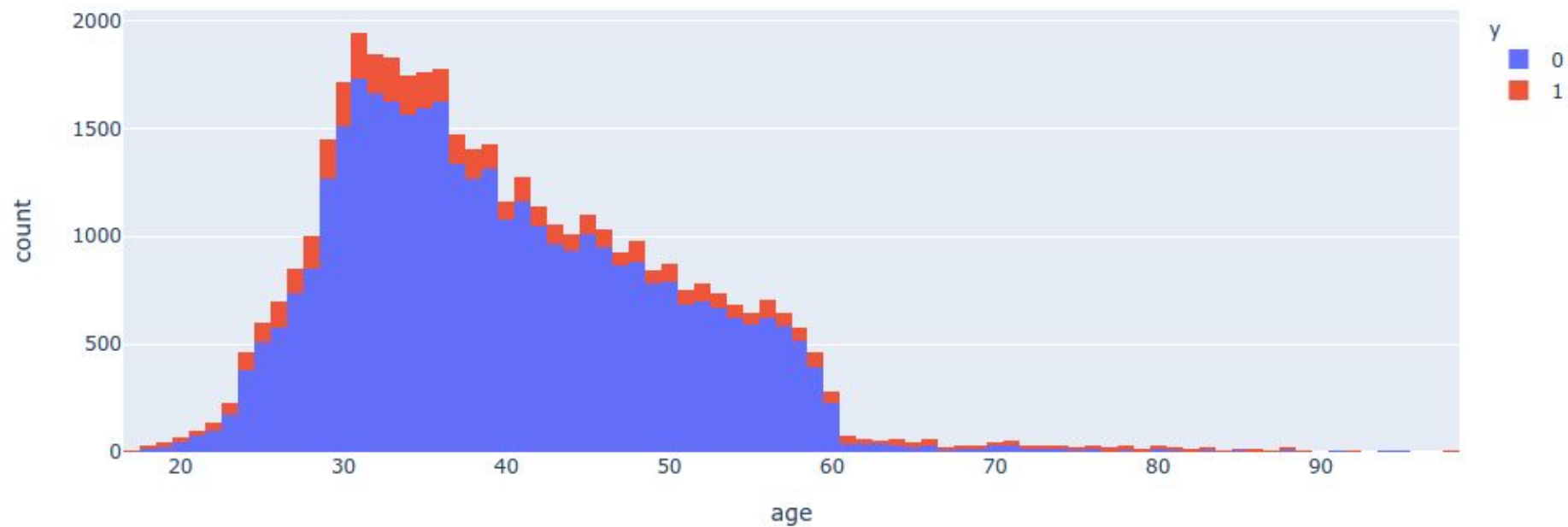


Age of customers

Exact age distribution of all contacted customers



Histogram of age grouped by subscribed/not subscribed term deposits

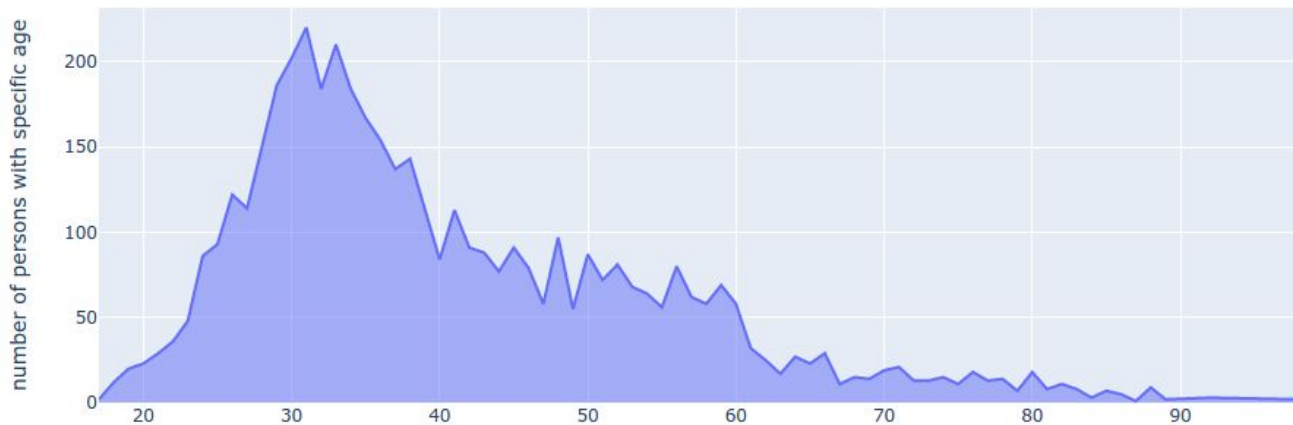




Customers, who subscribed to a term deposit

- The average customer, who actually subscribed to a term deposit is around 40.9 years old.

Exact age distribution of all contacted customers, who actually subscribed to a term deposit

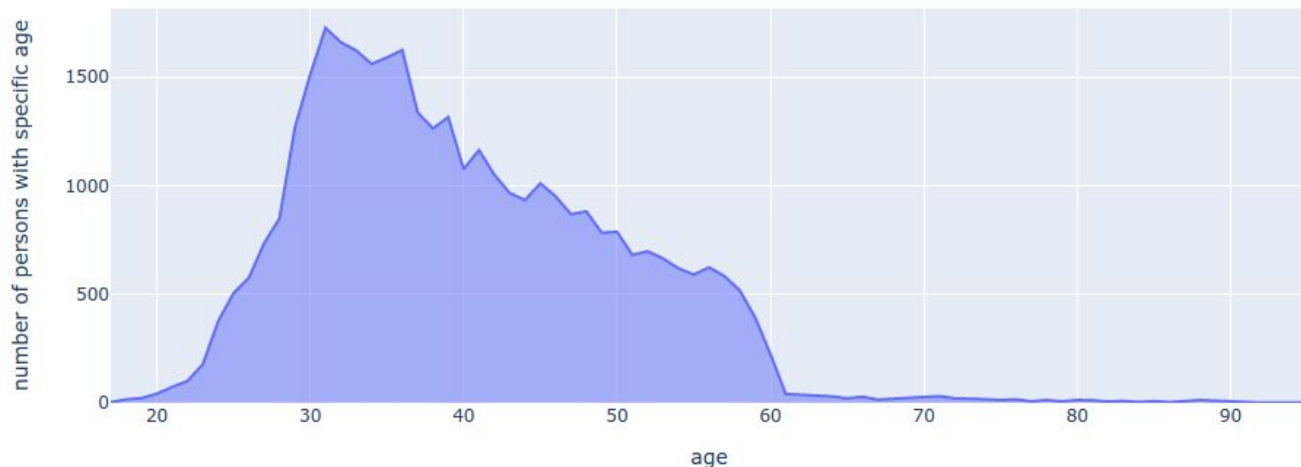




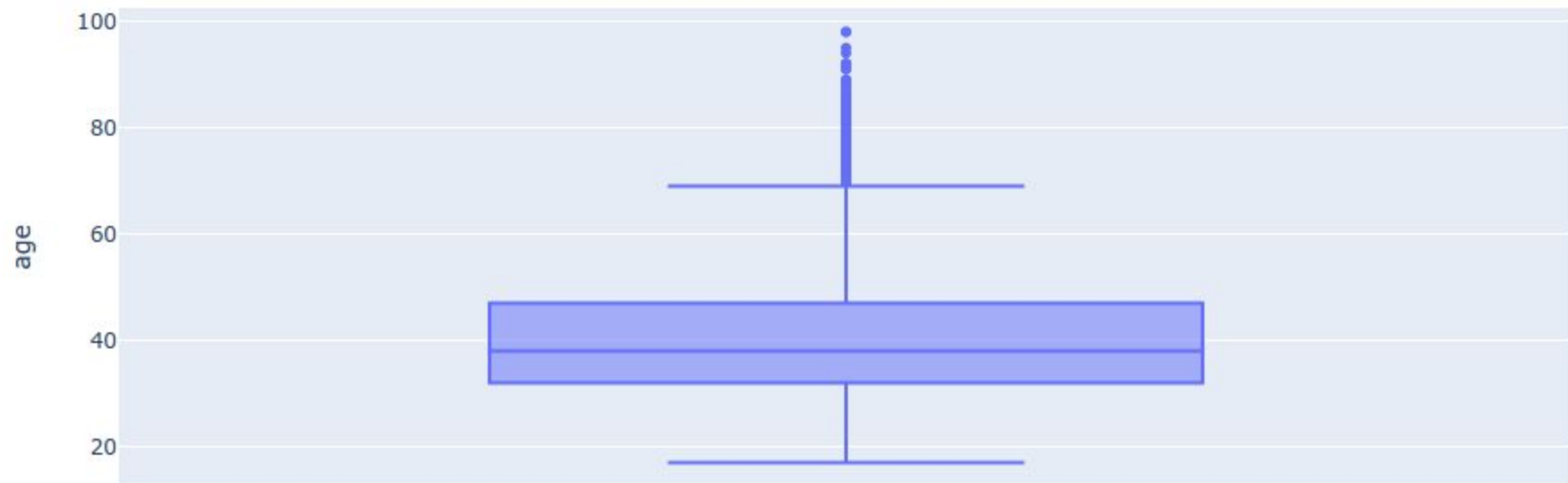
Customers, who did not subscribed to a term deposit

- The average customer, who did not subscribe to a term deposit is around 39.9 years old.

Exact age distribution of all contacted customers, who did not subscribe to a term deposit



Customers' age in general

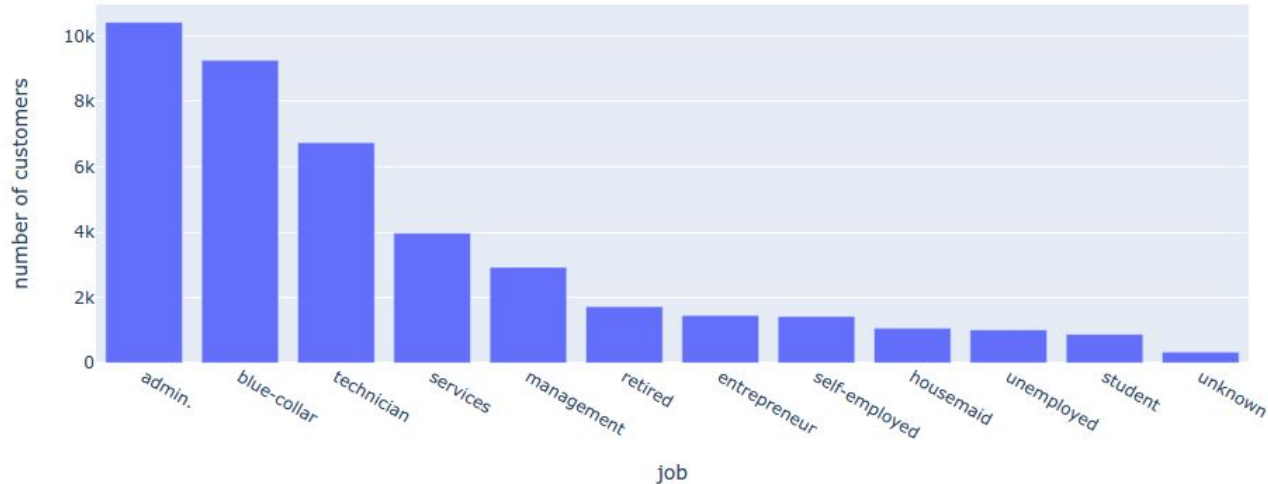


Comparison of customers' age, who subscribed to a term deposit to those who did not



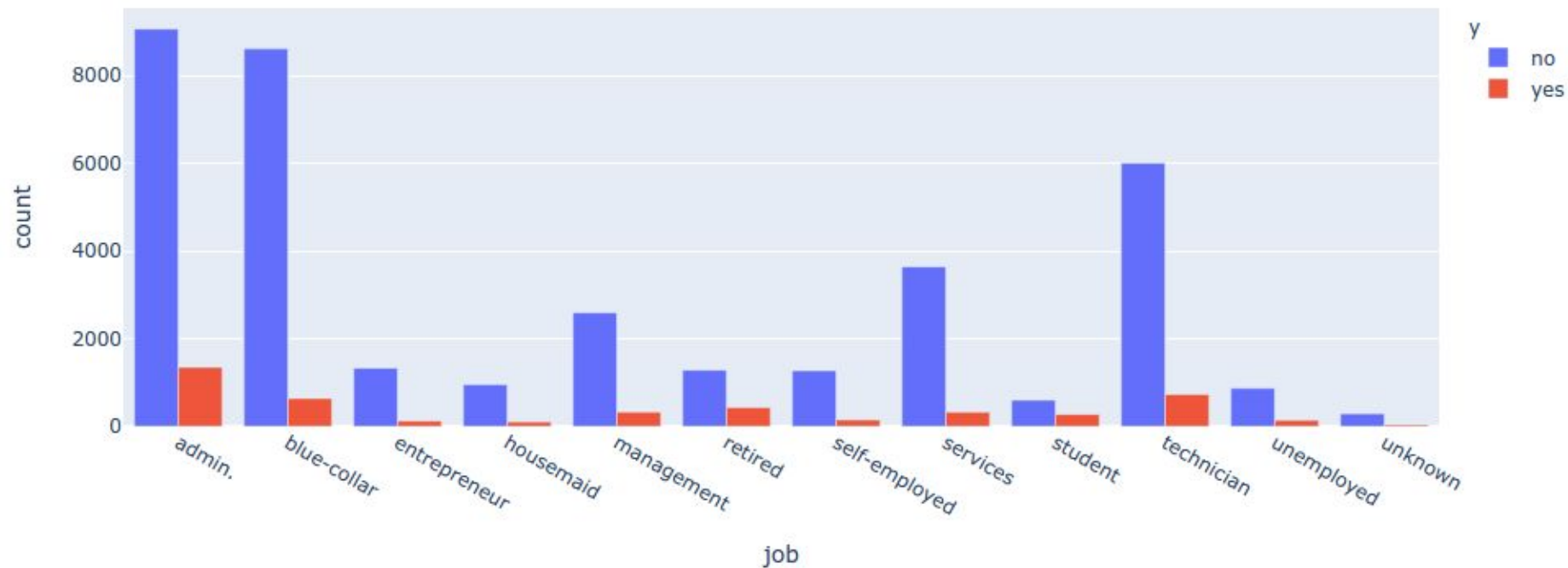
Job of customers

Size of customer groups in terms of their job

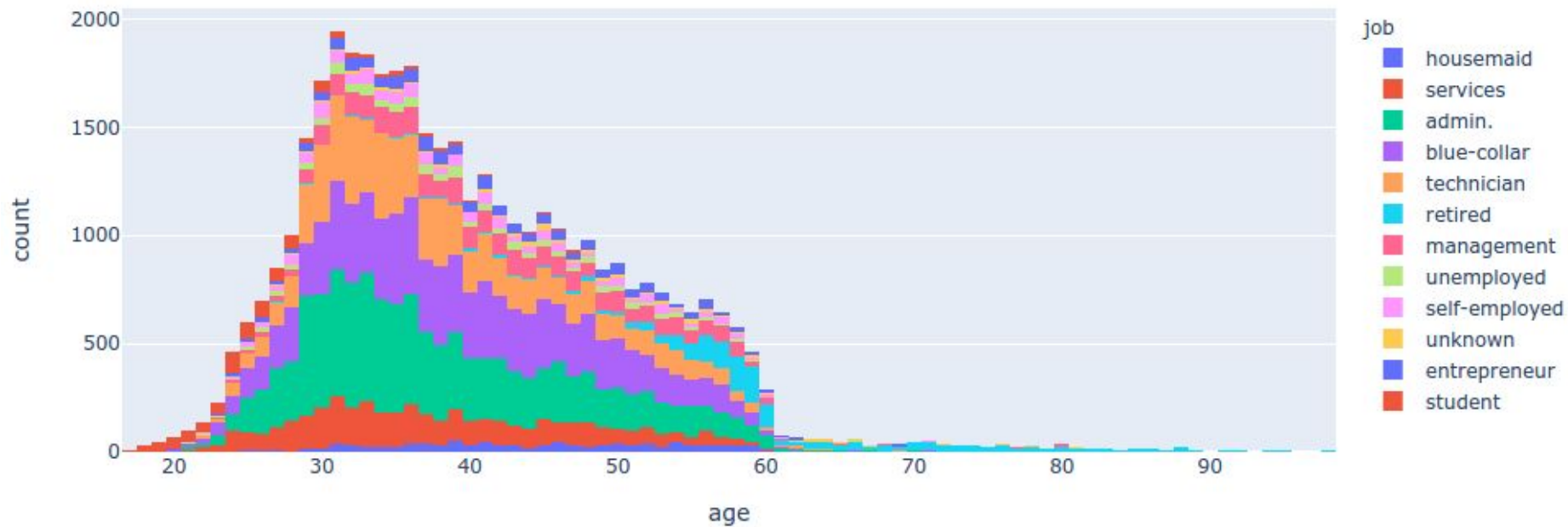


- Most customers work in the admin. sector.
- These are followed by customers working as a blue-collar.
- These are followed by technicians and customers working in services.

Comparison of success (yes) vs. no success (no) grouped by job



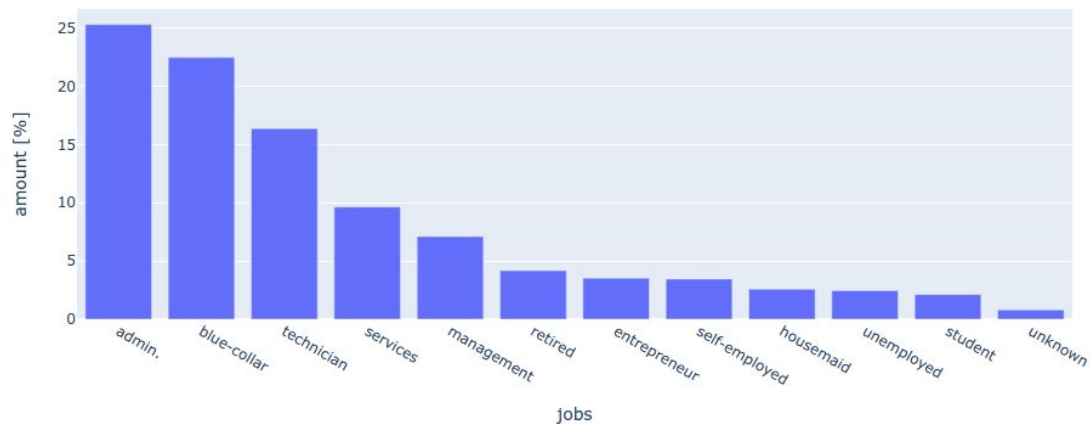
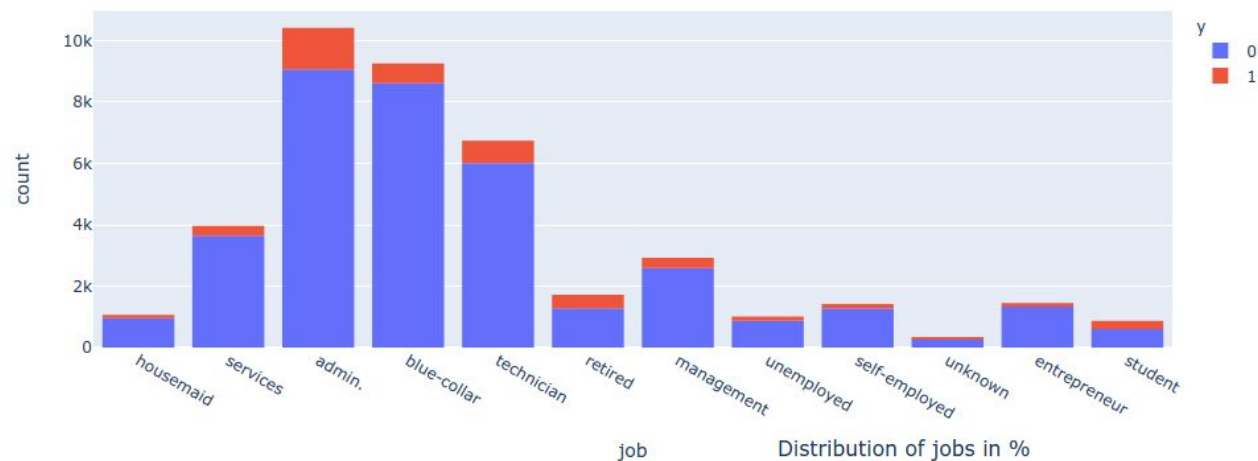
Histogram of age grouped by job





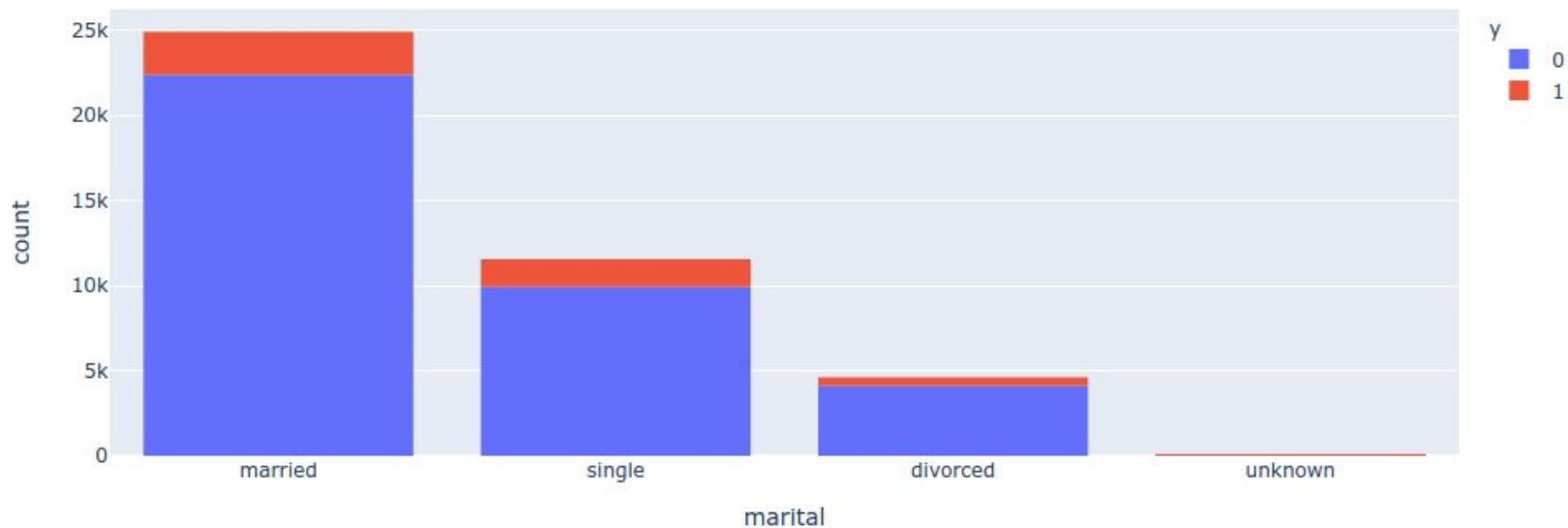
- The youngest customers are 17.
- In the age group from 17 to 19, all recorded customers are students.
- In the young age group from 17 to 22, the major customer group are students.
- Most term deposits are subscribed in the age group 29 to 39.
 - In this age group admin. jobs dominate.
- Starting with age 58, most customers are retired.
- The oldest customers are 98.
- The average customer is 40 years old.
- 50% of the customers are 38 years old.

Histogram of jobs grouped by subscribed/not subscribed term deposits

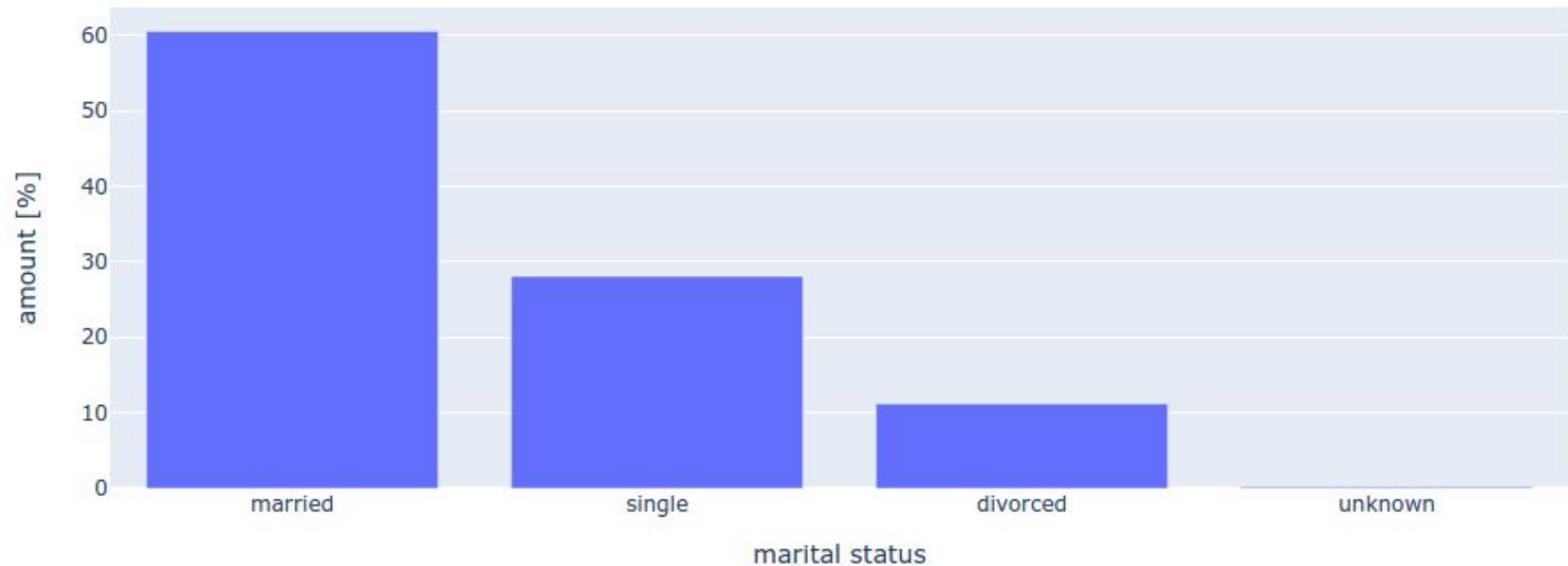


Marital status of customers

Histogram of marital status grouped by subscribed/not subscribed term deposits



Distribution of the marital status in %



- More than 60% of the customers are married.
- About 28% of the customers are single.
- Only 11% of the customers are divorced.

Phone calls

- The average phone call lasted 258.3 seconds (= about 4.3 minutes).
- The longest phone call lasted 4918 seconds (= about 82 minutes).

| | |
|-------|--------------|
| count | 41176.000000 |
| mean | 258.315815 |
| std | 259.305321 |
| min | 0.000000 |
| 25% | 102.000000 |
| 50% | 180.000000 |
| 75% | 319.000000 |
| max | 4918.000000 |

- The number of contacts performed during this campaign and for a client is on average 2.568.
- The median shows, that there are two contacts.

Histogram of contact methods grouped by subscribed/not subscribed term deposits





Success rate of customers which were contacted at most twice

- 28202 customers have been contacted at most twice.
- The success rate of customers, which have been contacted at most twice is about 12.4%.

Contrarily: Customers, which were contacted more than twice:

- 12974 customers were contacted more than twice.
- The success rate for customers who were contacted more than twice is about 8.7%.



- Customers who actually subscribed to a term deposit in the last campaign, had 2.05 contacts on average.
 - This may be one indicator out of others to find out whether a customer will subscribe or not, or whether it is worth to again contact the same customer.
- Astonishingly, there was one customer, where 56 contacts were performed during the last campaign!
 - **Let's further inspect this case:**
 - The customer (32 years old, married, with university degree), who had 56 contacts finally did not subscribe to a term deposit, so it was not worth the effort.



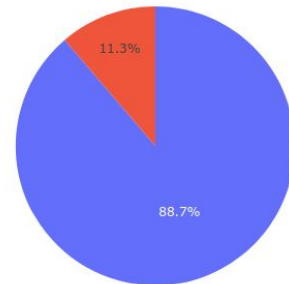
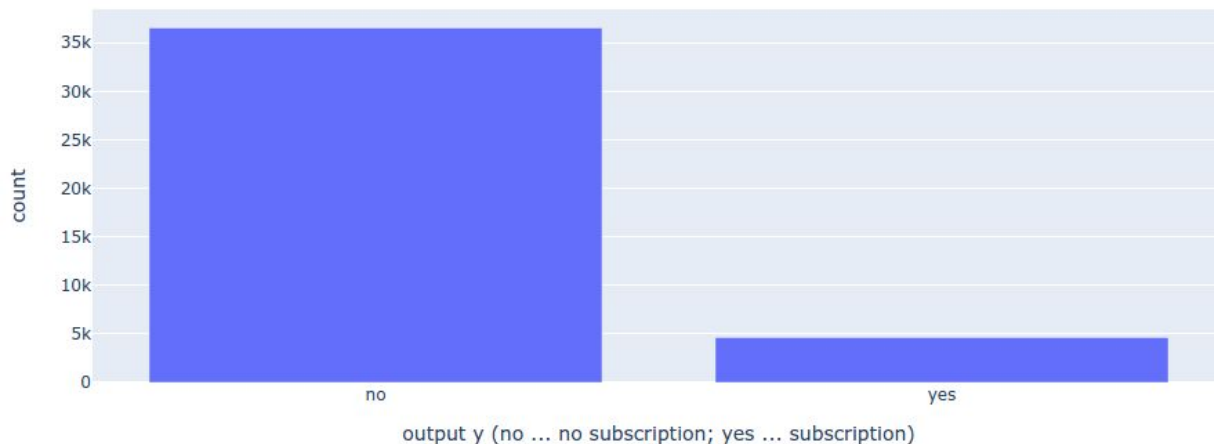
Number of days of the last contact from a previous campaign

- The average number of days that passed by after the client was last contacted from a previous campaign is 962.46 days.
- Most customers have been contacted 999 days ago before his campaign.



Subscribers vs. non subscribers

Illustration of customers with subscriptions and no subscriptions



- The output variable to predict is very imbalanced.

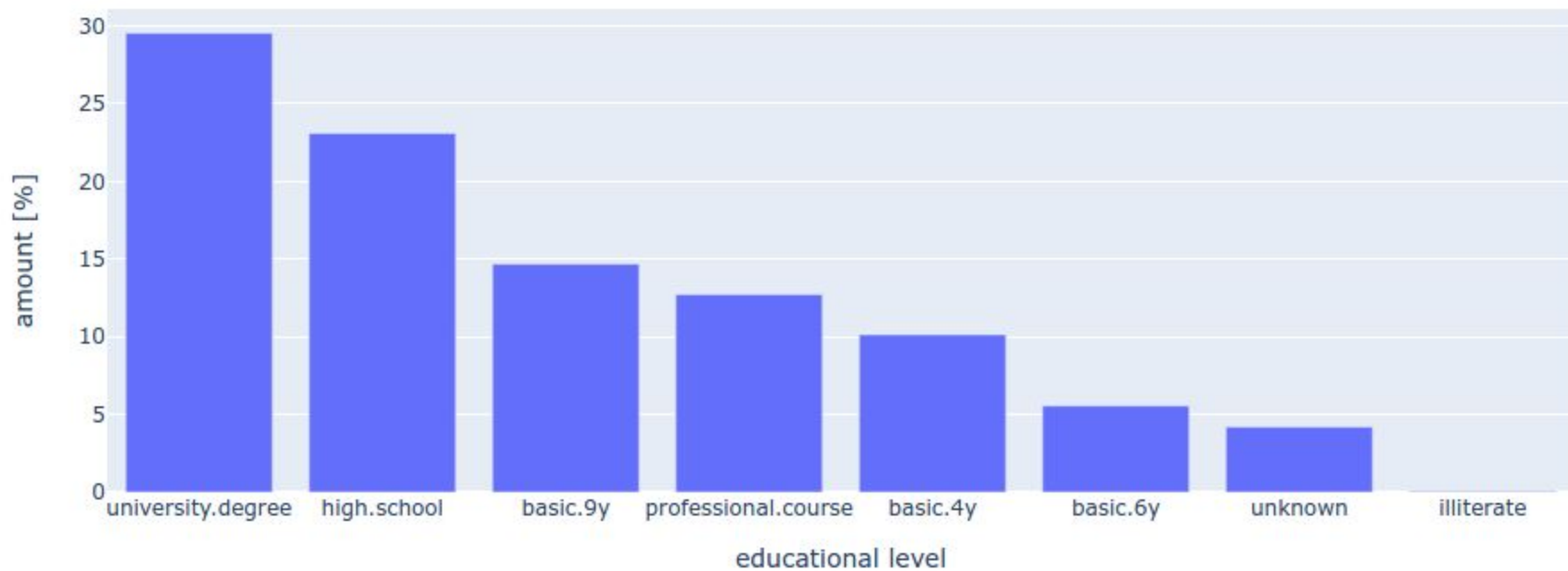


Education level of customers

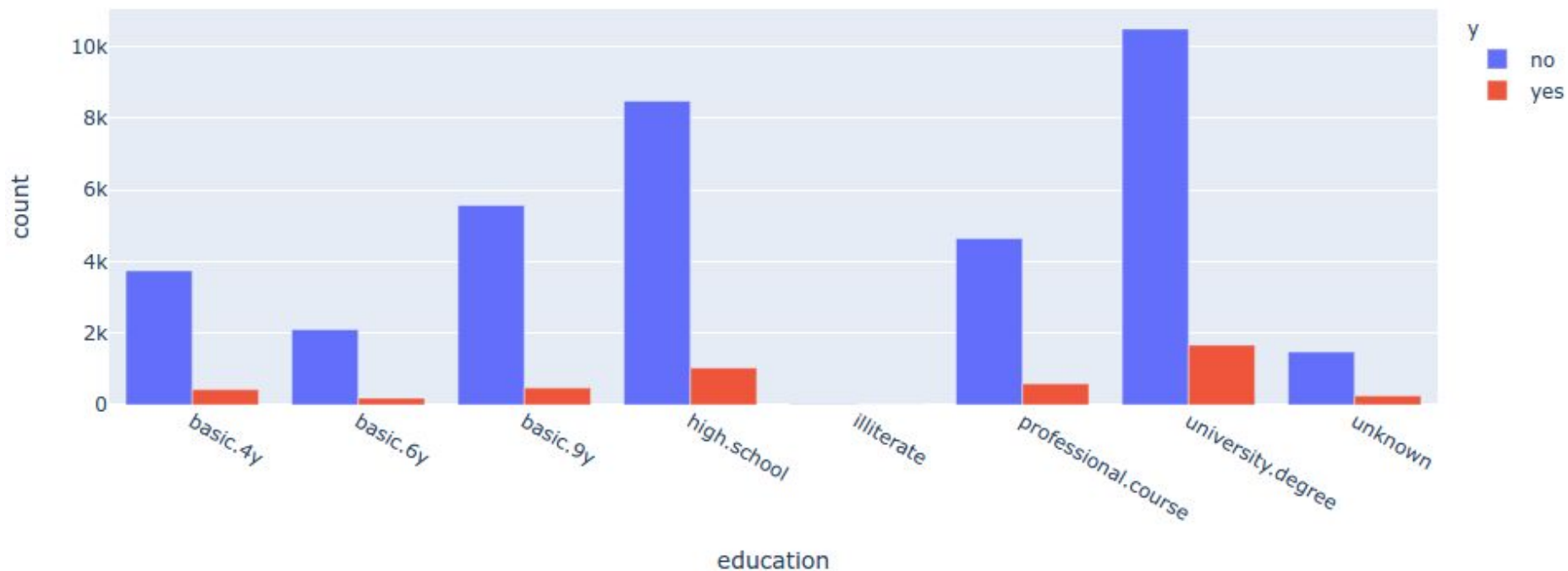
Histogram of education grouped by subscribed/not subscribed term deposits



Distribution of educational background of customers in %



Comparison of success (yes) vs. no success (no) grouped by education





- The highest success rate was achieved with illiterates (22.2%) and with people having an unknown education.
- Since illiterates are a very small target group and people with an unknown educational background are not an easy target group, they will be skipped, thus:
 - The success rate of the most promising target groups are (in descending order):
 1. Persons having a university degree (13.7%)
 2. Persons having made professional courses (11.4%)
 3. Persons having a high school degree (10.8%)



For which jobs were the campaign most successful?

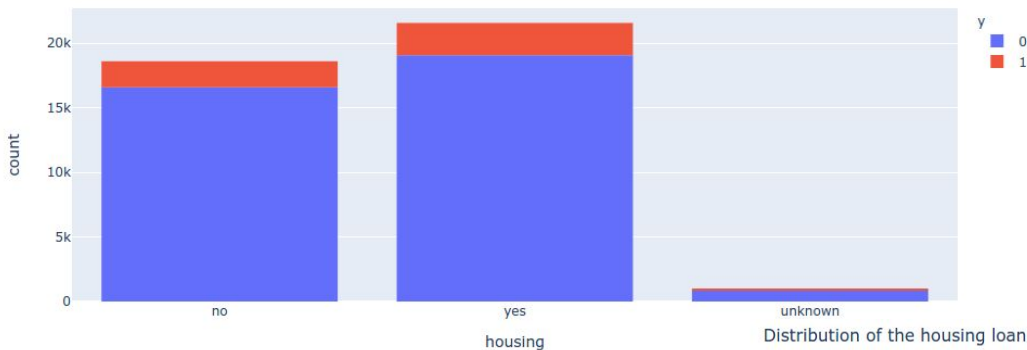
The highest success rates were achieved with the following jobs (descending):

1. students (31.4%)
2. retired people (25.3%)
3. unemployed (14.2%)
4. admin. personnel (13.0%)

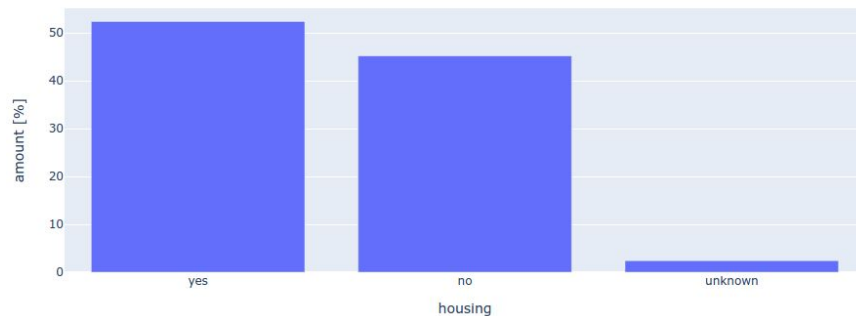


Customers' housing loan status

Histogram of housing loan status grouped by subscribed/not subscribed term deposits



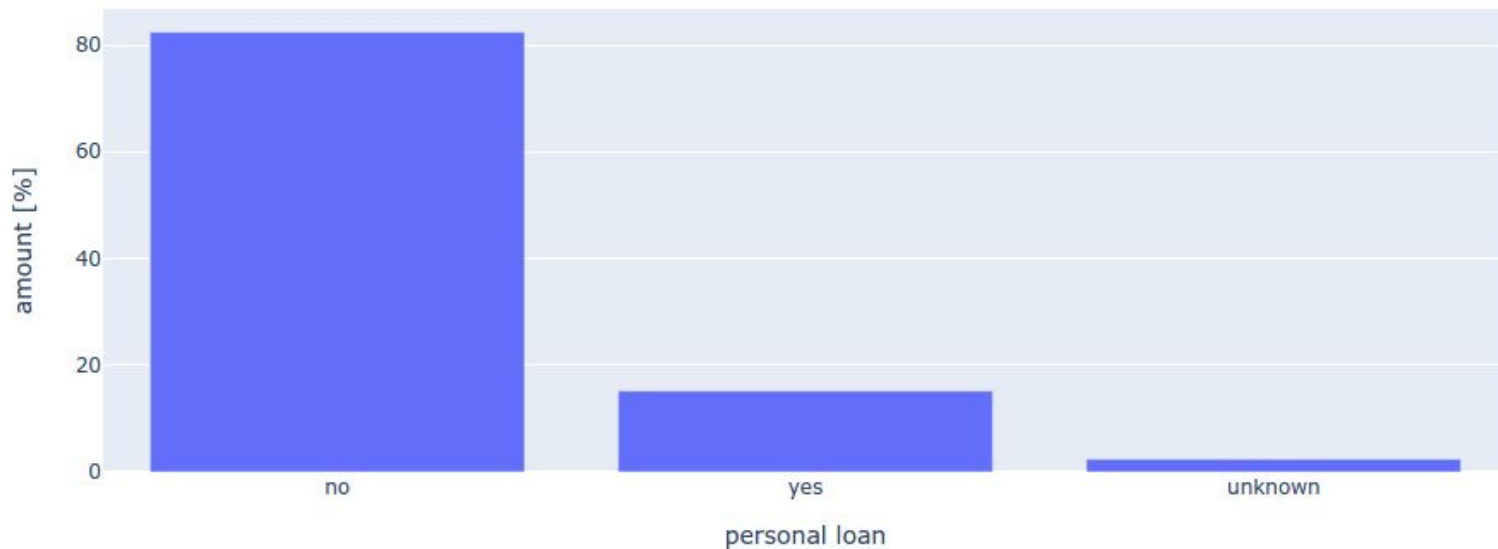
Distribution of the housing loan status of the customers in %





Customers' personal loan status

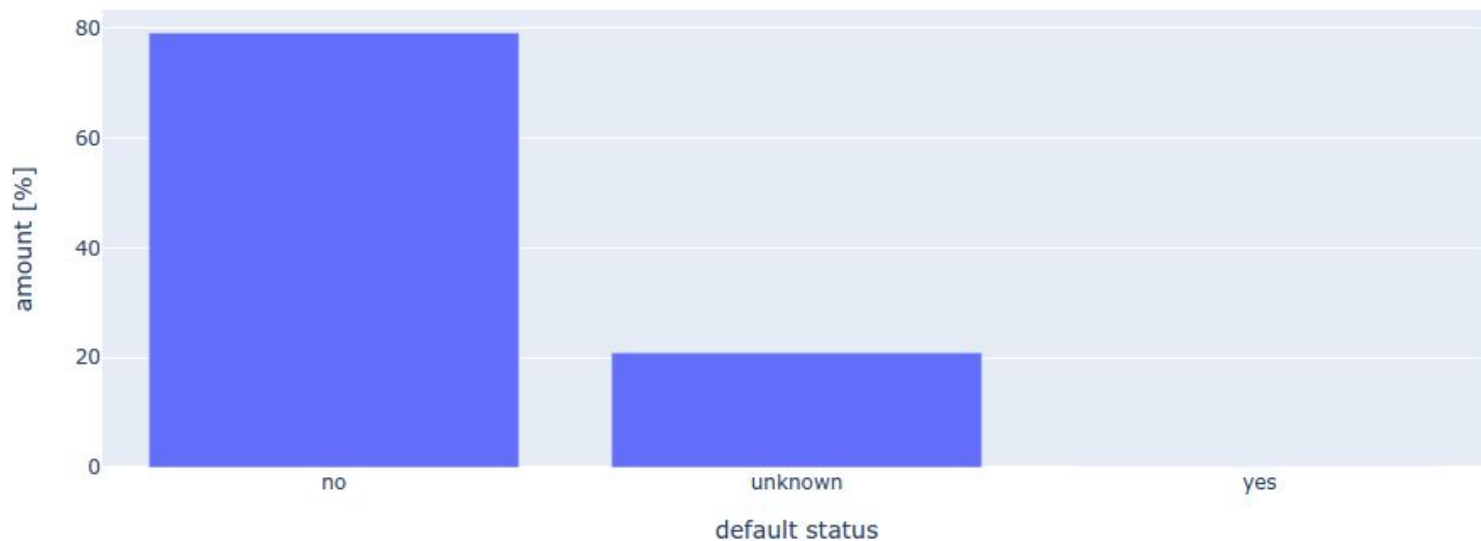
Distribution of the personal loan status of the customers in %





Customers' default loan status

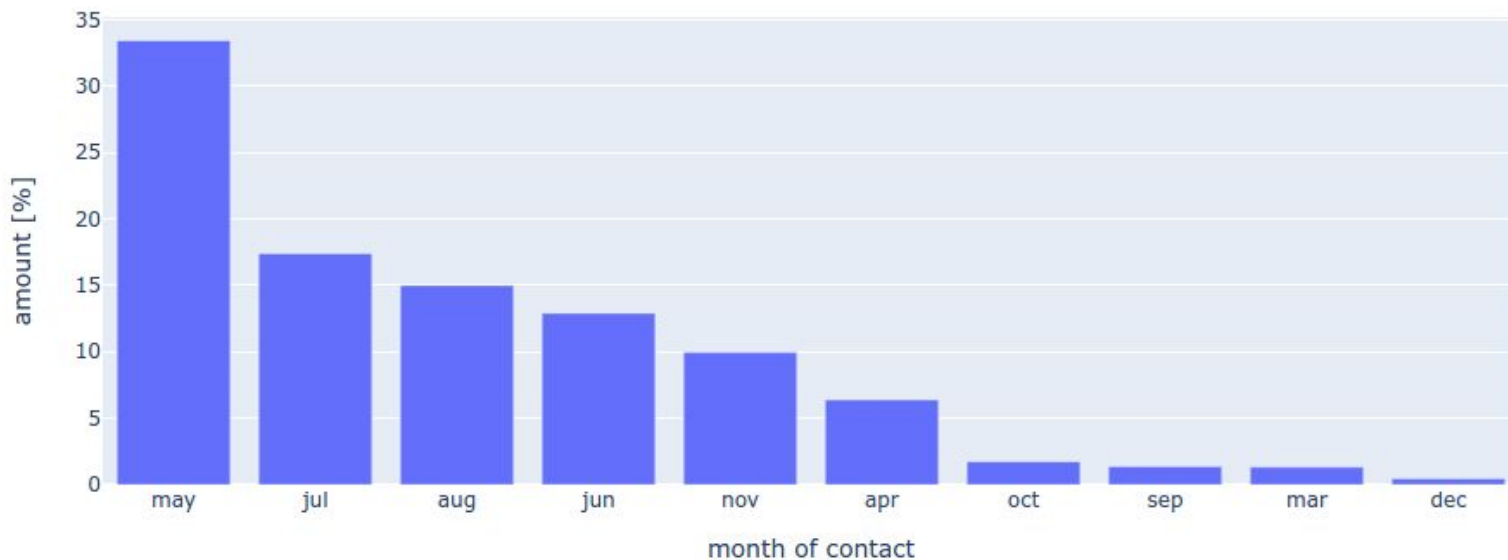
Distribution of the default status of the customers in %





Months on which customers were contacted

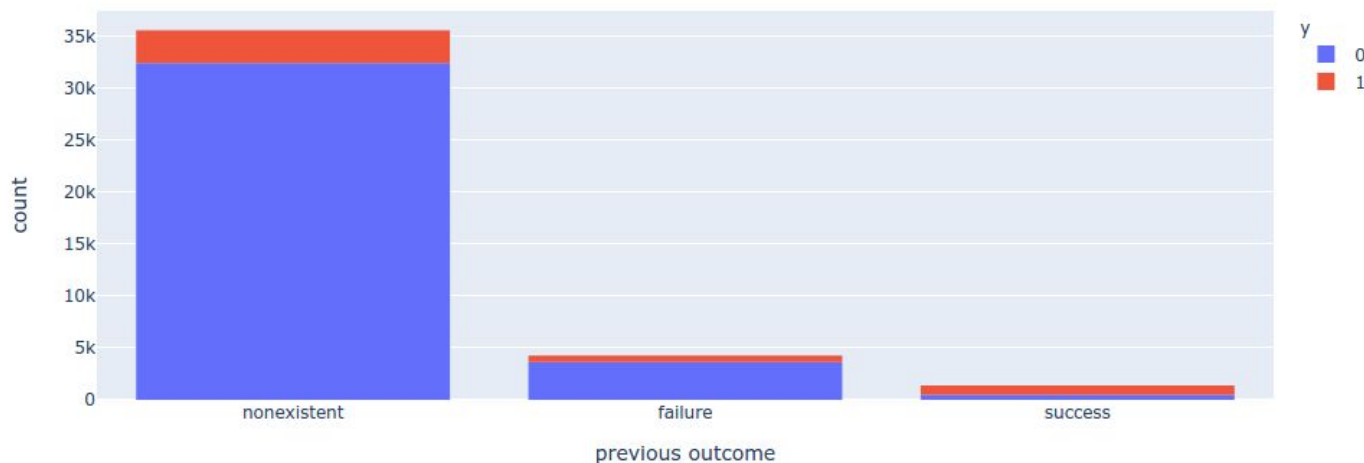
Distribution of the month where the customers were contacted in %





Outcome of previous campaigns

Histogram of previous outcomes grouped by subscribed/not subscribed term deposits



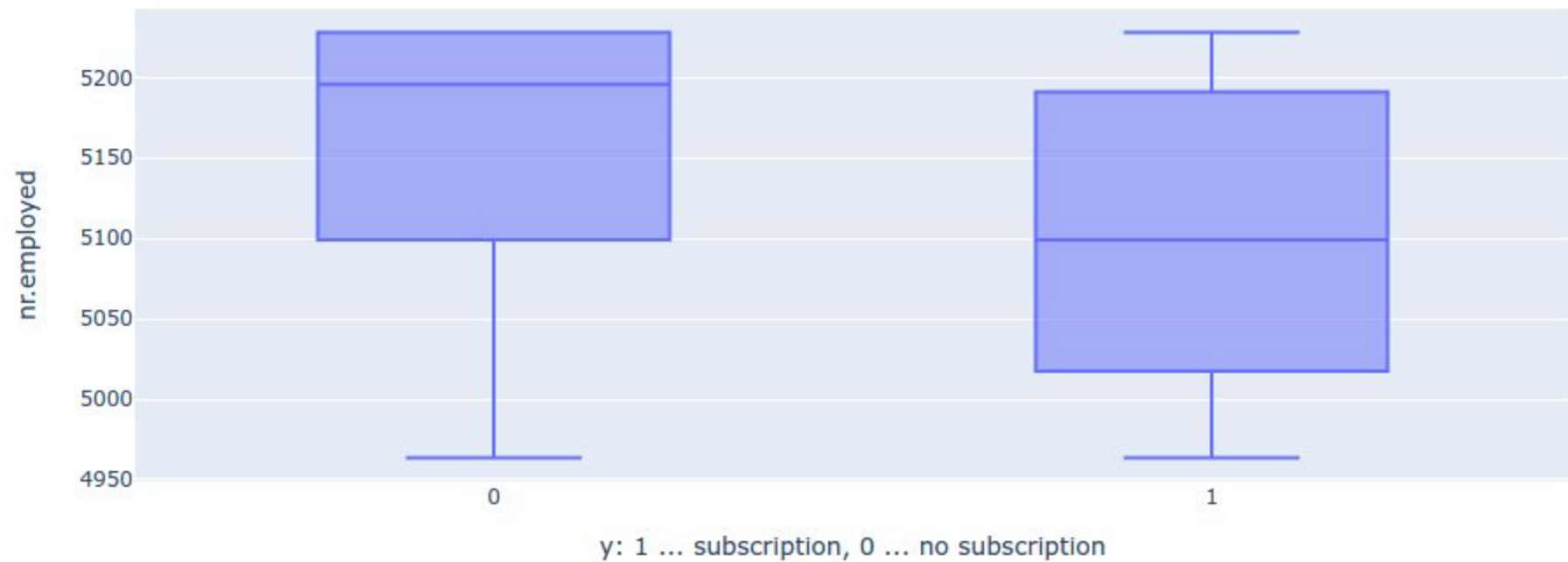
- Most customers (86.3%) have not subscribed to a term deposit previously.
- For 10.3% of the customers a failure occurred.
- Only 3.3% of the customers have actually subscribed to a term deposit previously.

- Most of the customers did not subscribe to a term deposit before, as they are new customers.
 - But there are also several customers, who subscribed to one or more term deposits (see above).
 - One customer even subscribed to seven term deposits in the past!
-
- Most customers, who subscribed to a term deposit this time, had no deposit previously.
 - Six subscribers, who have subscribed six times before, now subscribed for another term deposit in the latest campaign.
-
- Most customers, who did not subscribe to a term deposit this time, had no deposit previously.
 - One subscriber, who has subscribed seven times before, did not subscribe for a term deposit in the latest campaign.

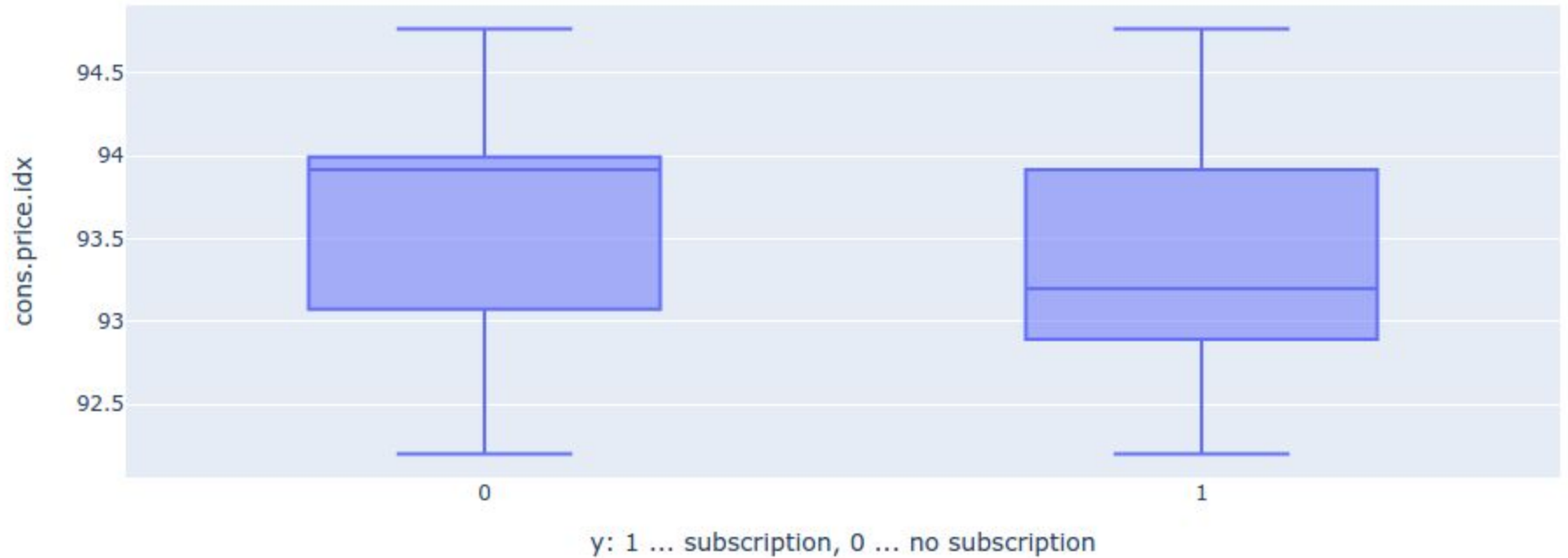
Box plots for euribor3m for subscriptions and non subscriptions



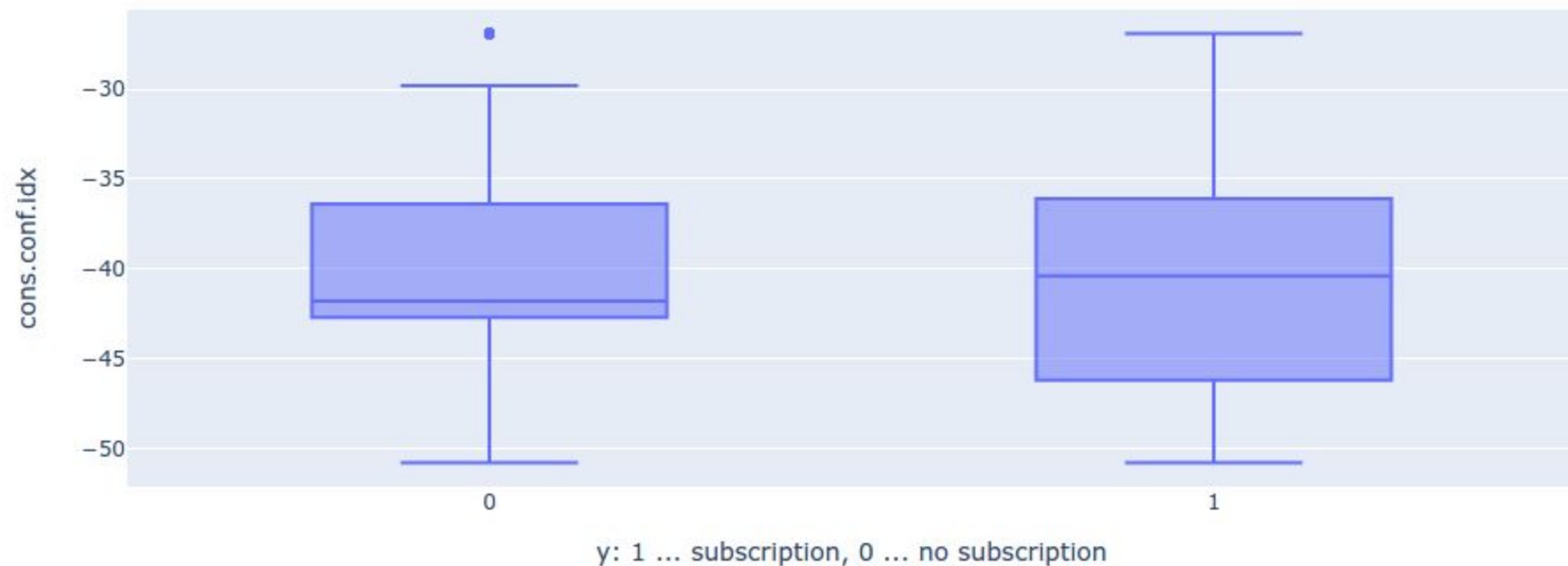
Box plots for nr.employed for subscriptions and non subscriptions



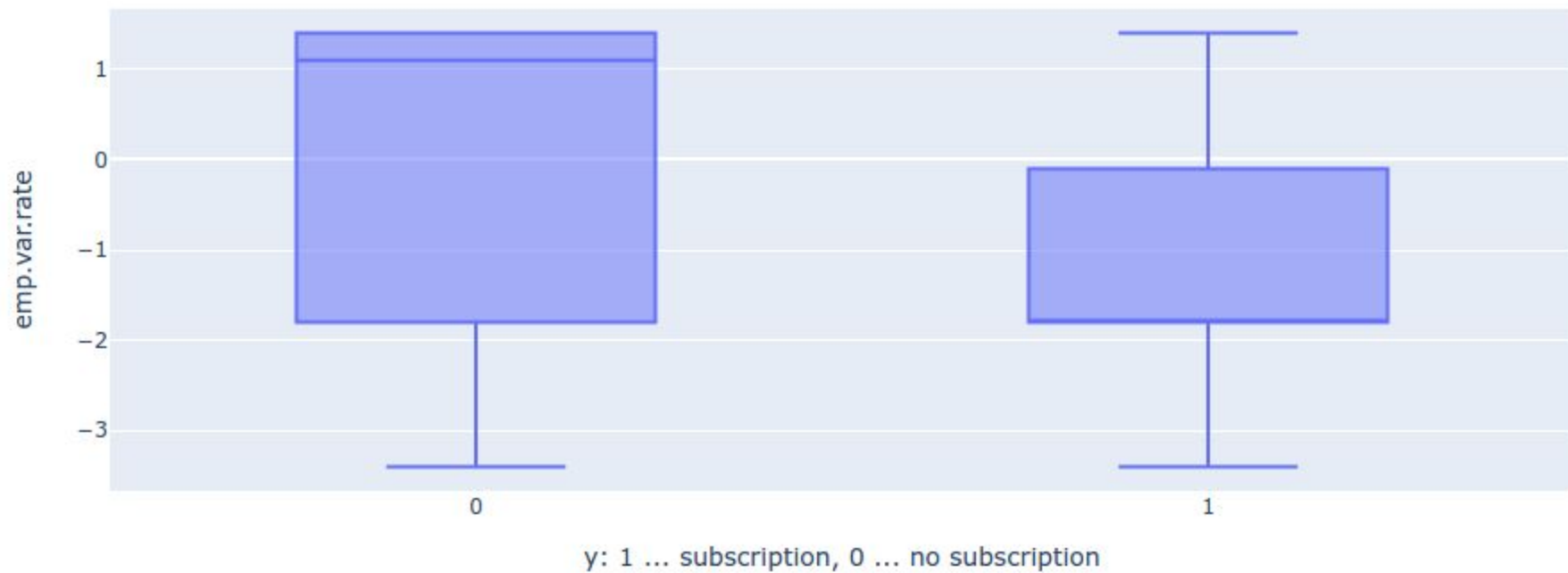
Box plots for cons.price.idx for subscriptions and non subscriptions



Box plots for cons.conf.idx for subscriptions and non subscriptions



Box plots for emp.var.rate for subscriptions and non subscriptions





Thank you for your attention!

More examples, more details and more context, see:

[ipy nb-files](#)

