Ad-Hoc Bank Campaign Market Analysis

Data Duo

Problem Description

- **Objective**: ABC Bank wants to develop a predictive model to predict whether a customer will subscribe to their term deposit product based on their past interactions with the bank or other financial institutions.

- **Data**: ABC Bank has a dataset of information related to direct marketing campaigns conducted by a Portuguese banking institution. The outcome of each campaign was recorded as either a successful subscription ('yes') or a non-subscription ('no’).

- **Task**: To build a classification model that can accurately predict whether a client will subscribe to the term deposit product ('yes') or not ('no').

Data Description

The dataset contains 41,000 rows and 21 features.

- The 'y' feature represents the target variable, indicating whether a customer subscribed to the term deposit product or not.
- The distribution of the target variable is as follows:

    > No: 36,548 customers did not subscribe to the term deposit product.

    > Yes: 4,640 customers subscribed to the term deposit product.

The data is highly imbalanced, with a ratio of approximately 1:8 for 'no' to 'yes' subscriptions.

Feature Types: The features are categorized as categorical and numerical. There are 10 categorical features and 10 numerical features apart from the response variable ‘y’.

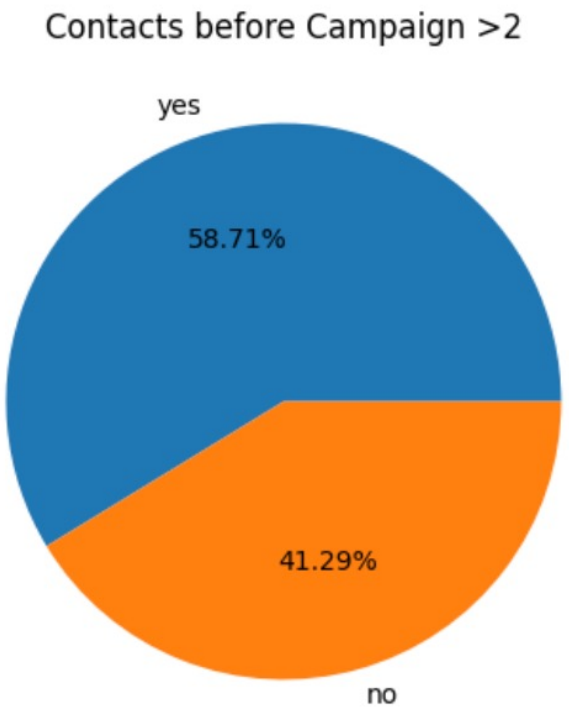
Exploratory Data Analysis

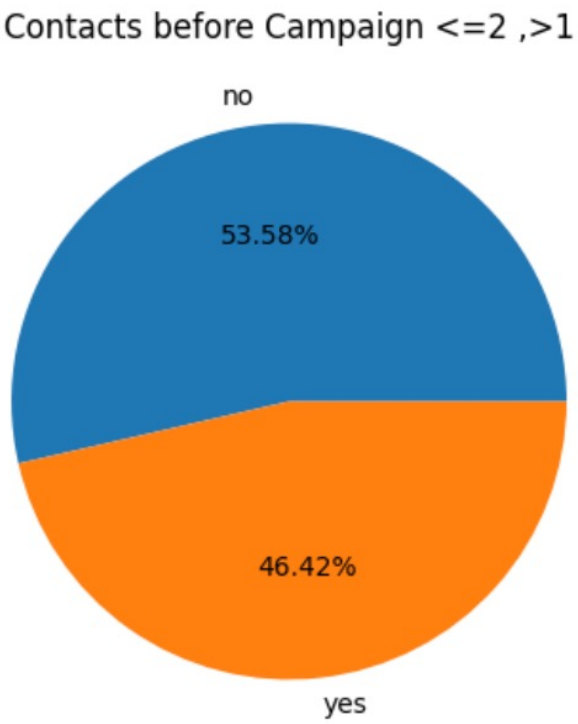
- There are 10 numerical features:
    - Three variables are discrete in nature, while the remaining seven are continuous.
    - One of the features, 'pdays', has values ranging from 1 to 27 but also includes a value of 999, which appears to be an imputed value for missing data.

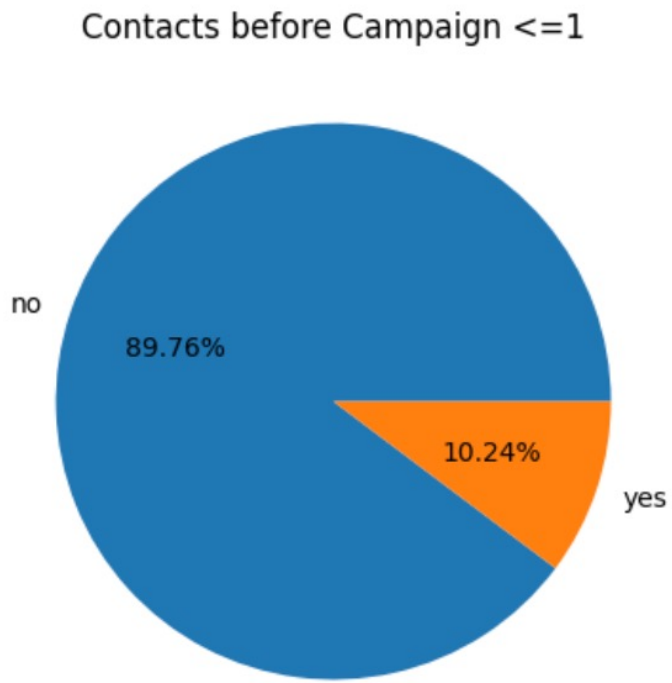
The 999 value accounts for 96% of the data for this feature. The other two discrete variables seem to have valid values.

- Among the seven continuous variables
- 'age' and 'duration' exhibit outliers

Total number of times each people were contacted.

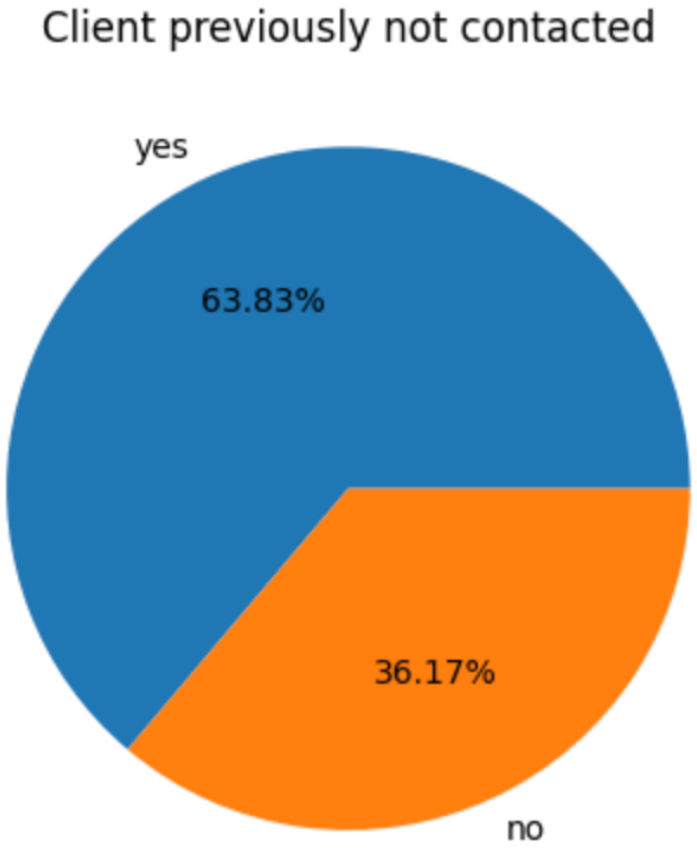


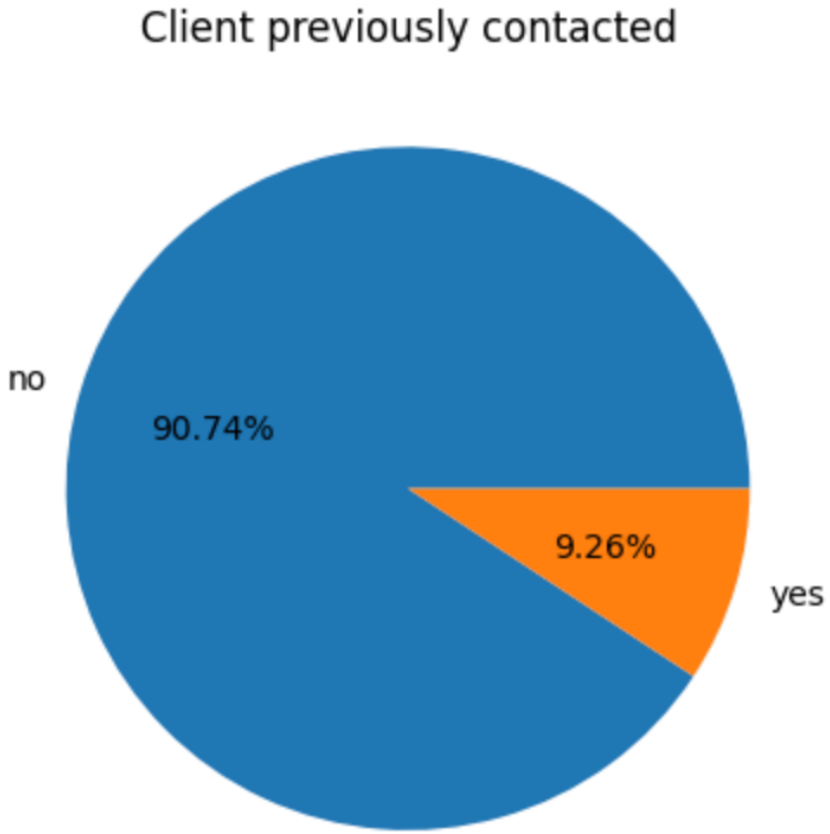




Observation: Less chance of subscribing the data, if they are contacted once or not.

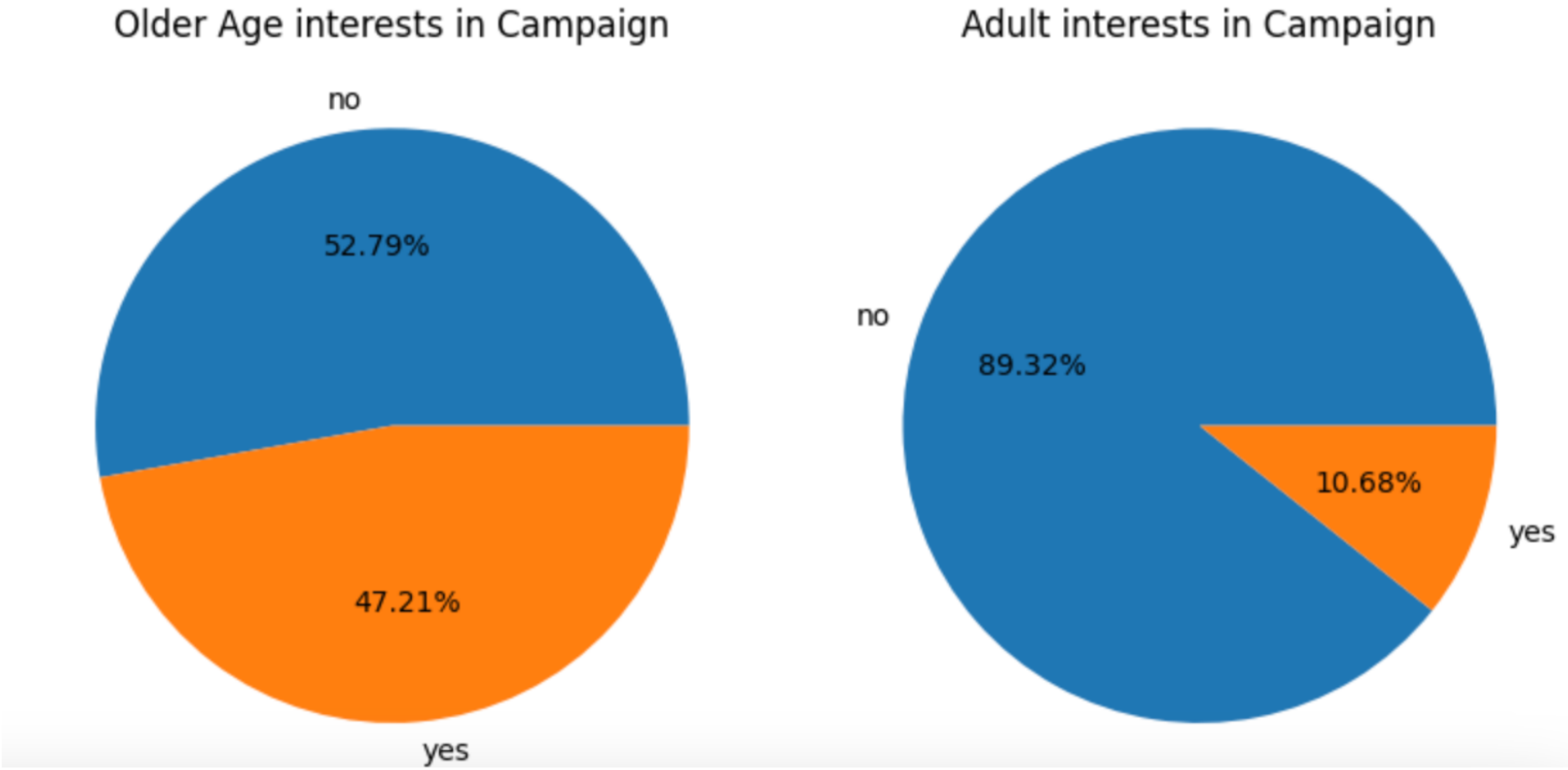
Clients contacted in the previous campaign



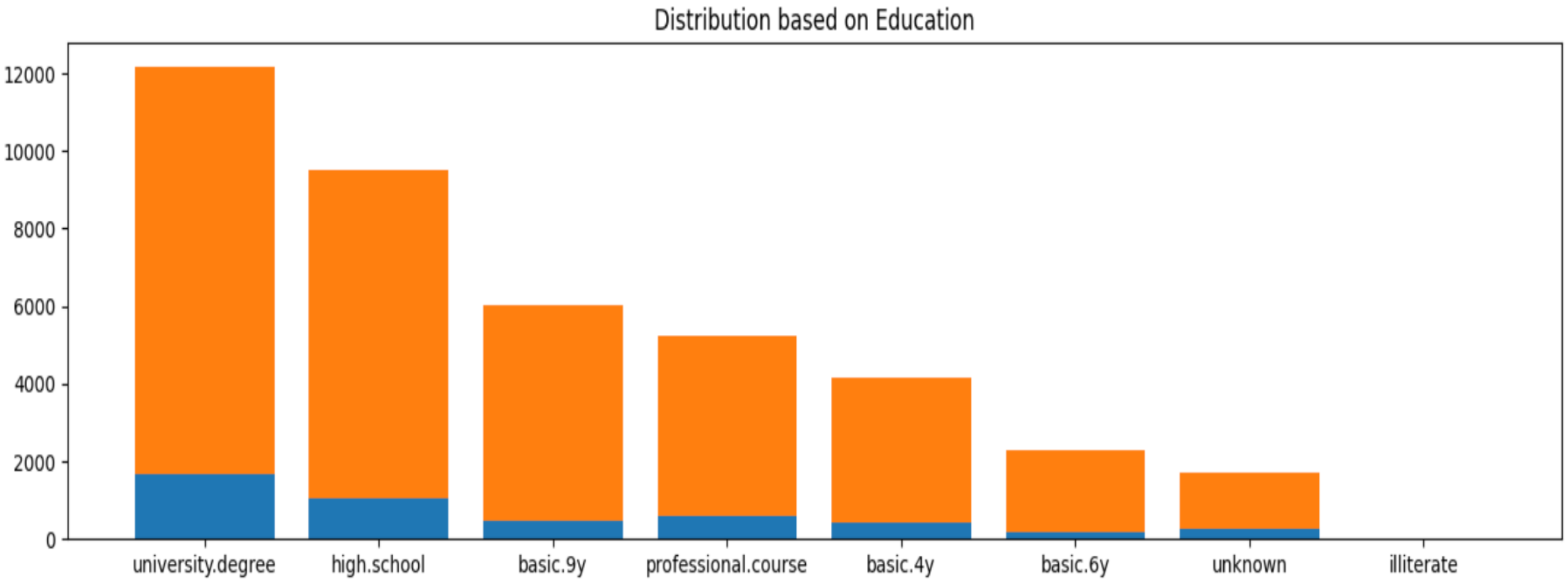


Observation : Client who were contacted in previous campaign and in the current campaign has more chance of enrolling subscription.

Impact of campaign marketing different ages

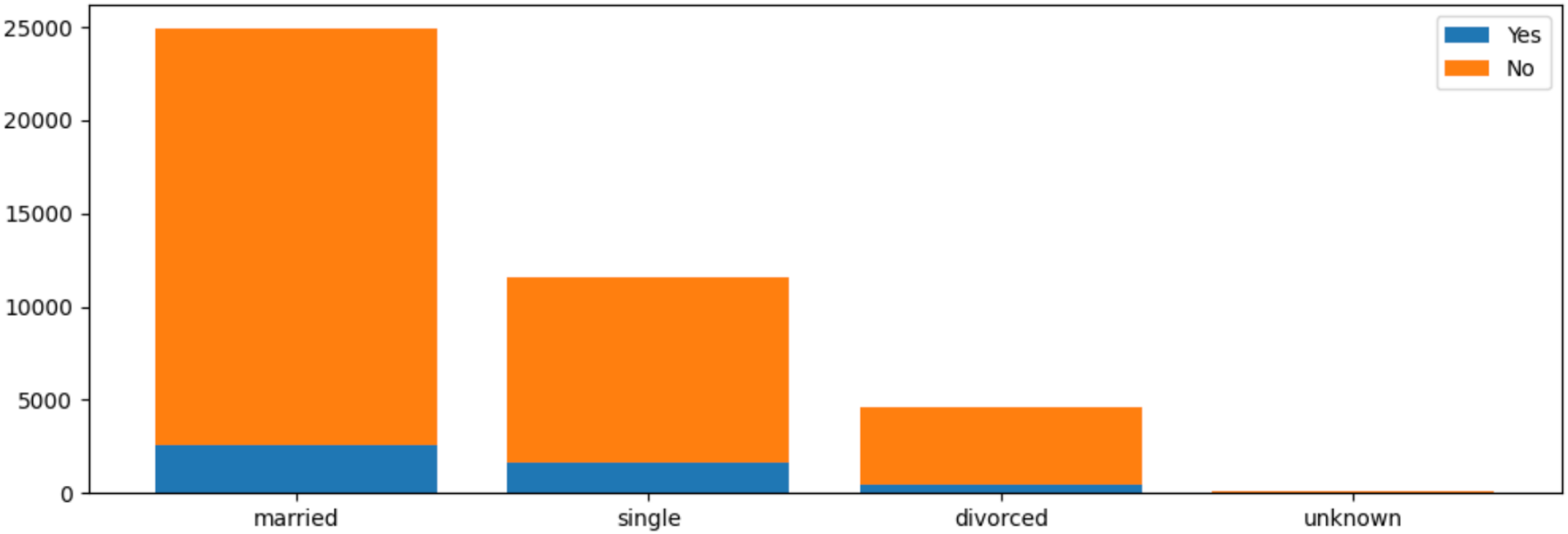


Observations: Elder people are more interested than adult in the adult



Observations: Here, ‘yes’ is blue colored bar and ‘no’ is orange colored bar.
From the plot, Educated people are more interested in the subscription compared to illiterate people.

Distribution based on Marital status

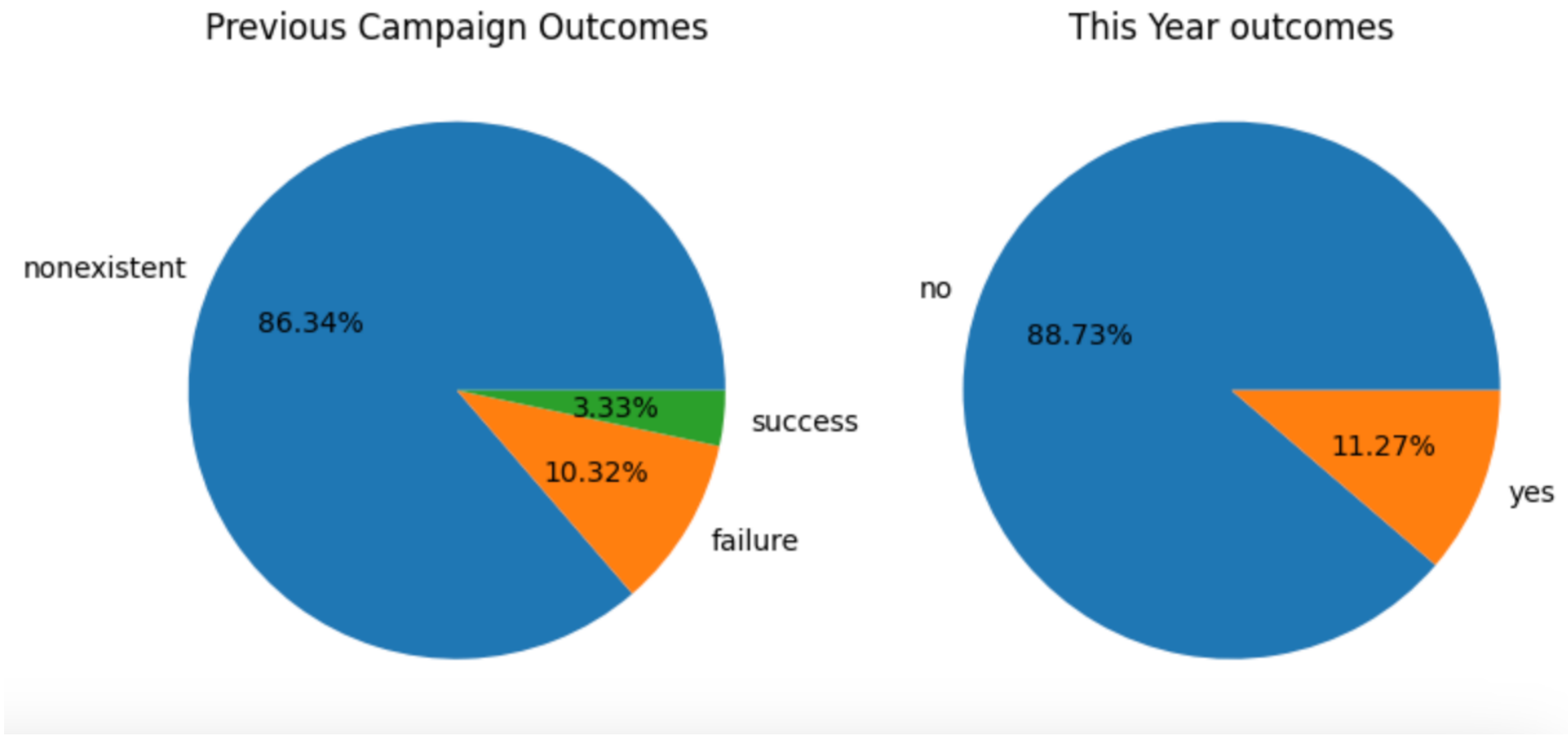


Observations: Married people are more influenced and eager to subscribe. Here, Unknown is less relevant. Hence, we can ignore it.

Distribution based on the Job status

Observations: Administrator Job status people has more interests in taking the subscriptions.

Previous outcome v/s This year outcomes



Conclusion:

- Older people are more interested in term deposit product compared to adults.
- Adults' chances of enrolling into the subscription can be increased by contacting them more than once.
- University students are likely to get involved in the campaign and show interested in the product.
- Overall, The dataset can model using

> Logistic regression: a simple linear classification model that is effective for binary

> Random forest: an ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting

> XGB boosting: a powerful boosting algorithm known for its speed and performance in classification tasks

> SVM : a versatile classification algorithm that finds a hyperplane to separate data into different classes, making it suitable for both linear and non-linear problems