

EXPLORATORY DATA ANALYSIS

INTRODUCTION

NAME: VIVIAN KERUBO MOSOMI

Email: kerubomosmi7@gmail.com

Country: Kenya

College: Jomo Kenyatta University of Agriculture and Technology (JKUAT)

Specialization: Data Science

Problem Description

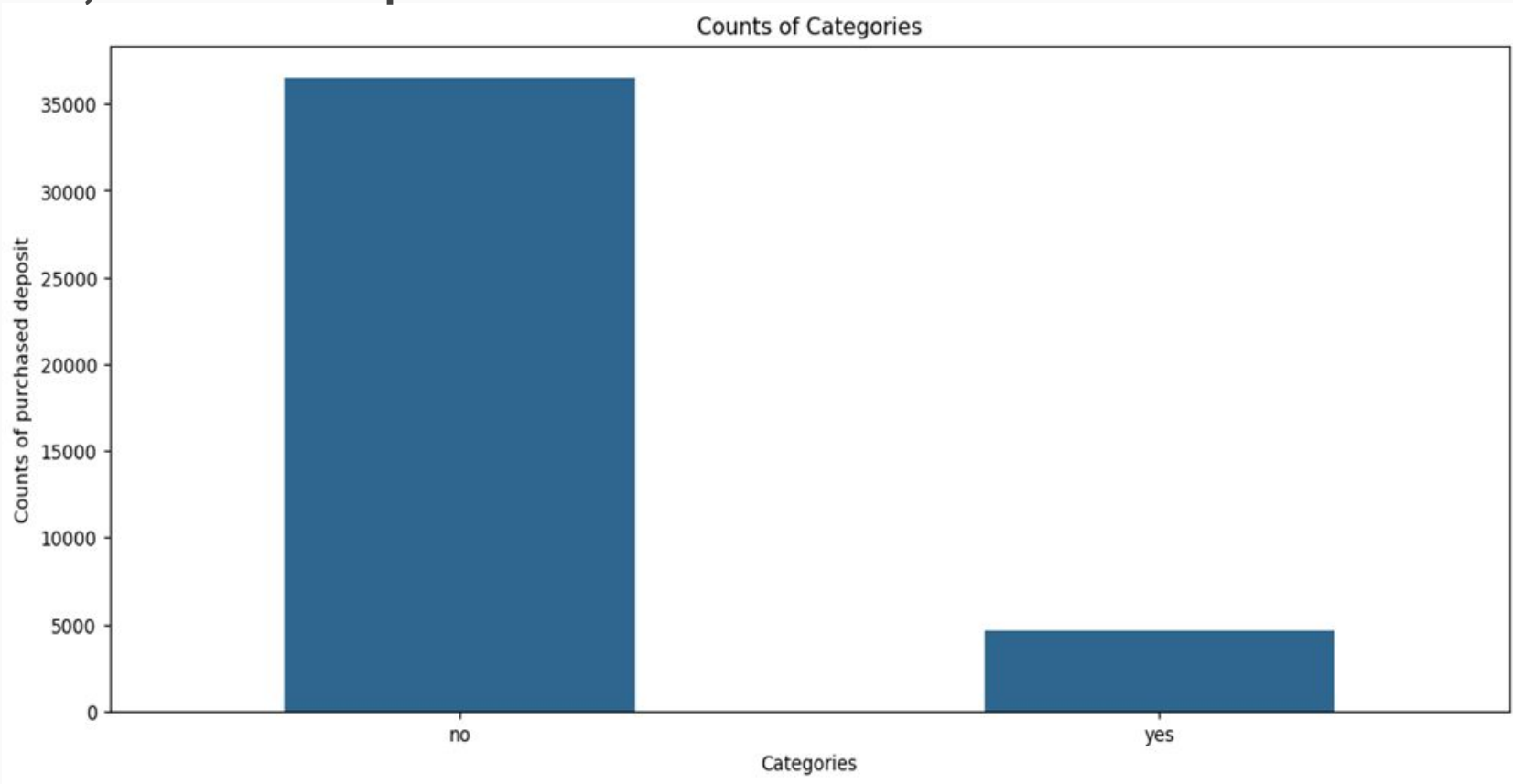
ABC Bank wants to predict whether a particular customer will buy their term deposit product based on their past interactions with the bank or other financial institutions. This will allow the bank to focus its marketing efforts on customers who are more likely to purchase the product.

GitHub Repository Link

https://github.com/Vee2002/DataGlacier_Internship/tree/vc/Data%20Glacier

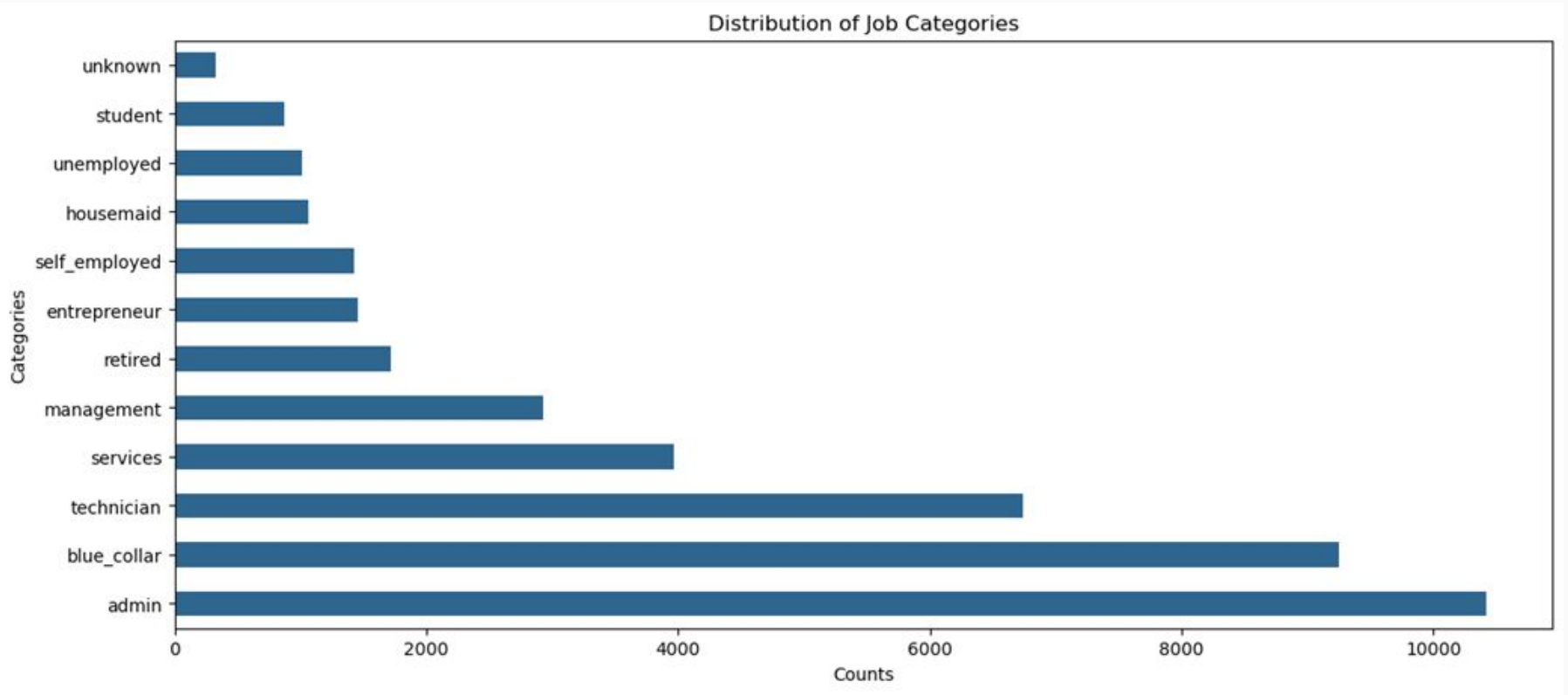
1.UNIVARIATE ANALYSIS

i) Purchased Deposit Counts



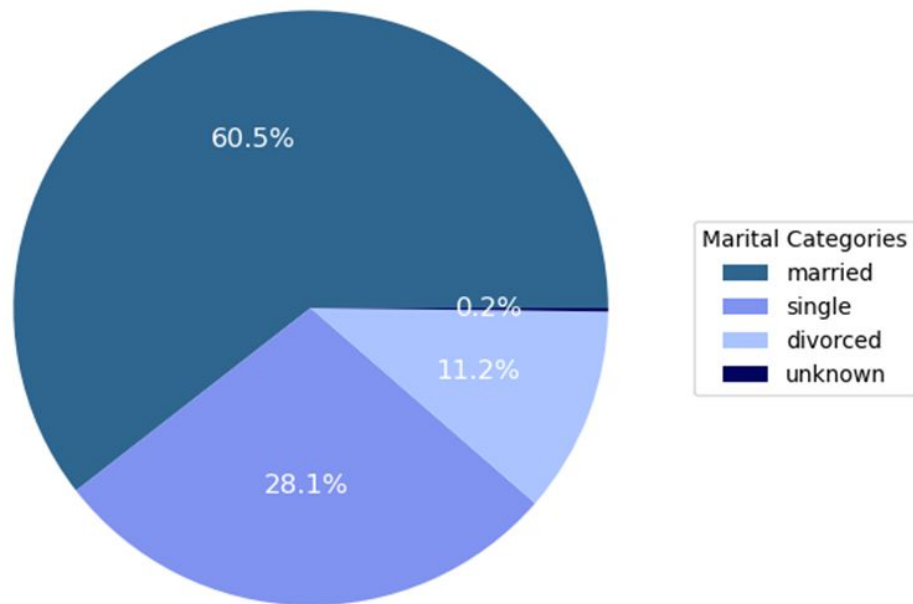
Most people did not purchase the term deposit

ii) Job Categories



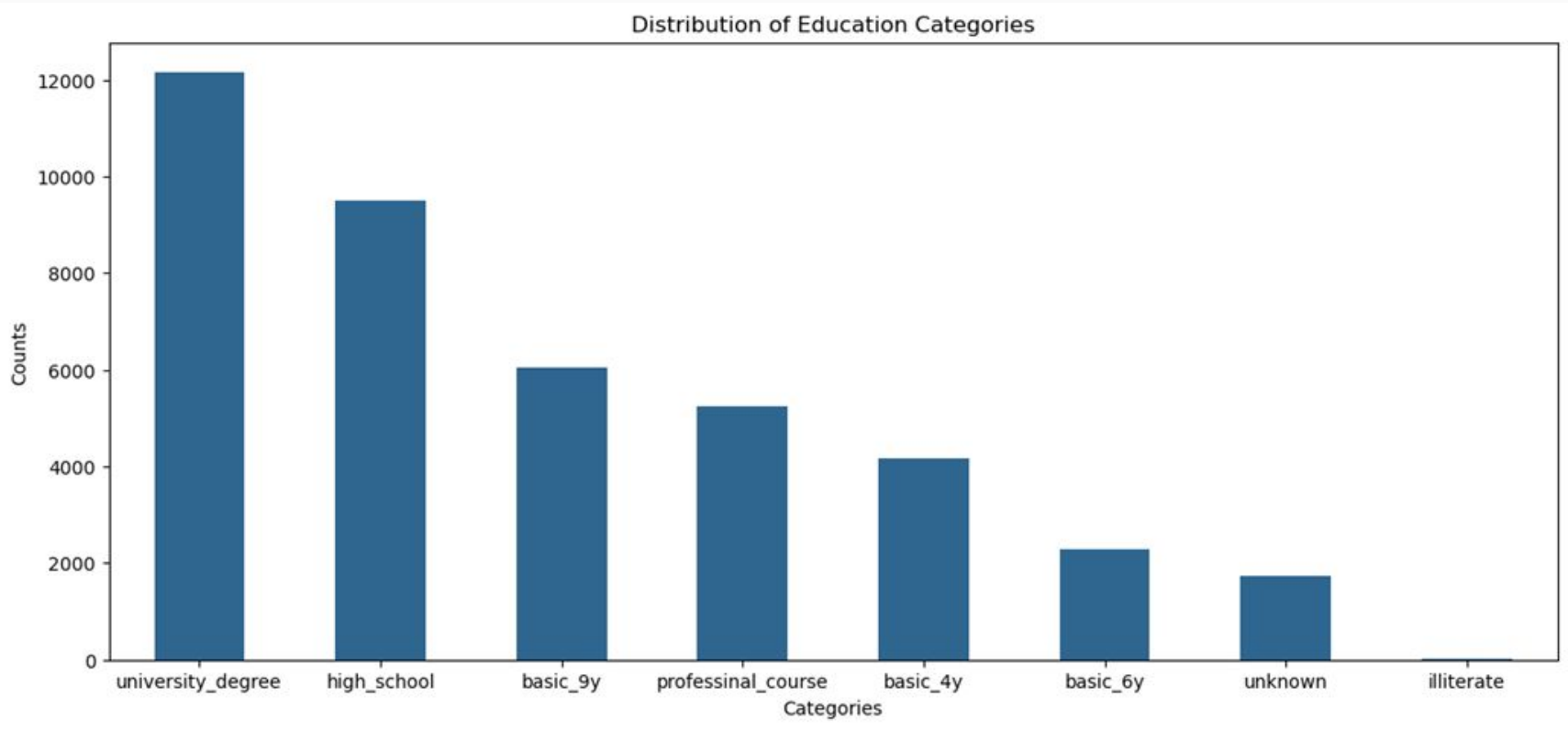
Most people work as admins, followed by blue collar jobs and technicians.

iii) Marital Categories



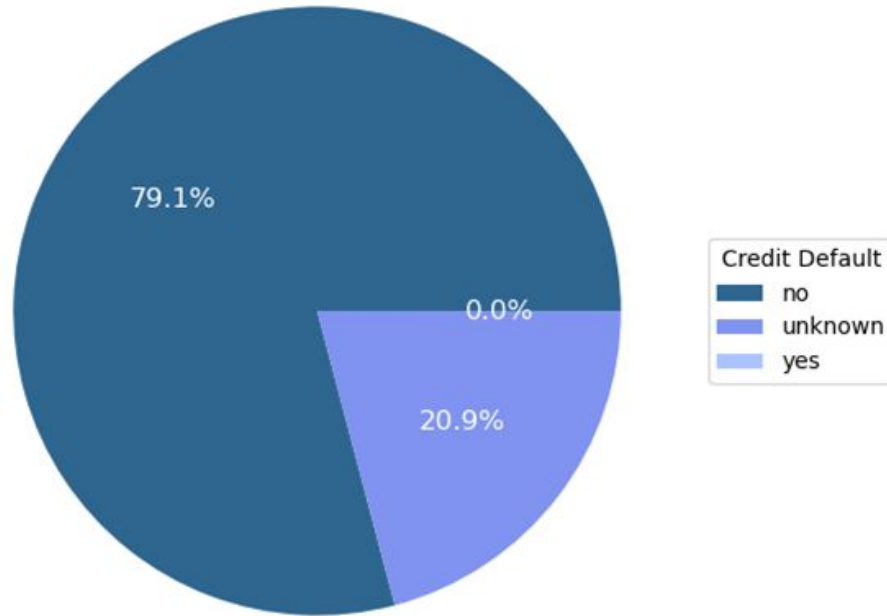
Most customers are married as they lead by 60.5%. Single people follow by 28.1%, and divorced are 11.2%.

iv) Education Categories



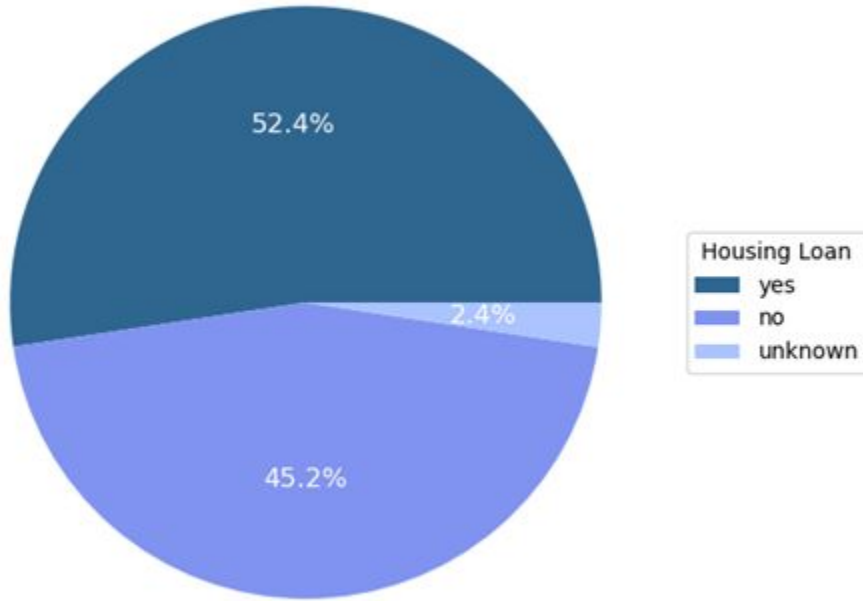
Most people are at the level of university degree followed by high school.

v) Default Loans



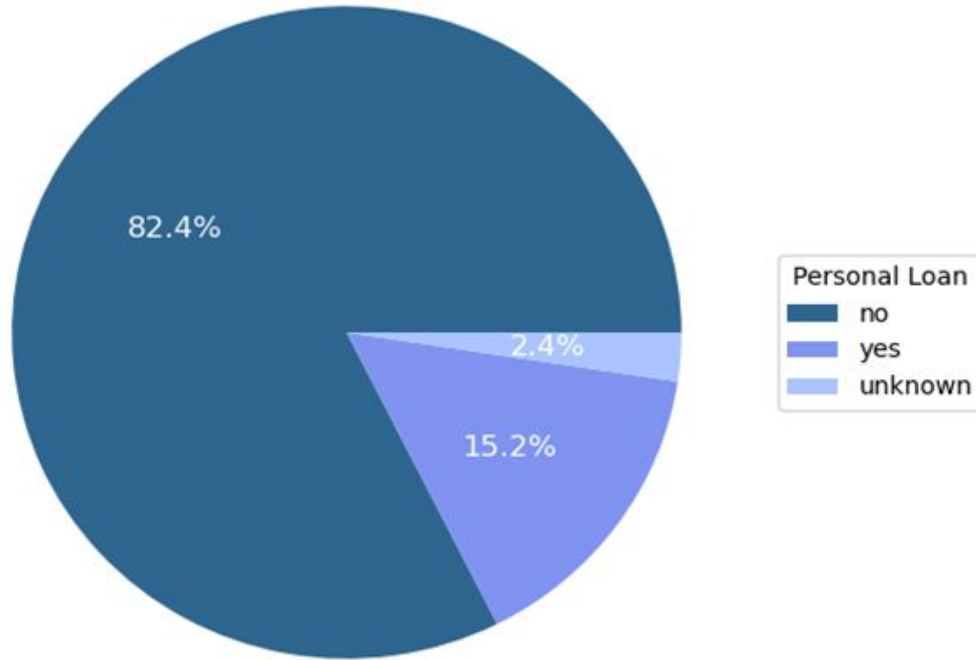
From the chart most people do not have credit defaults

vi) Housing Loans



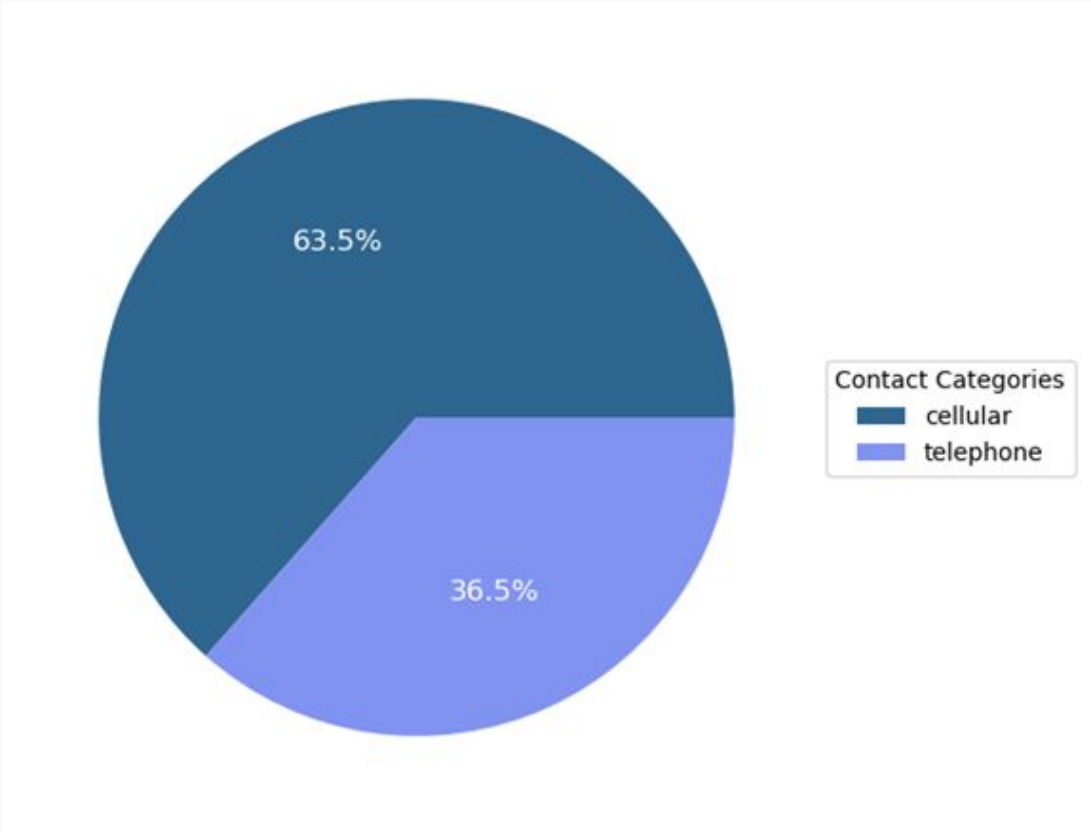
Most people have housing loans cause of the higher percentage of 52.4%

vii) Personal Loans



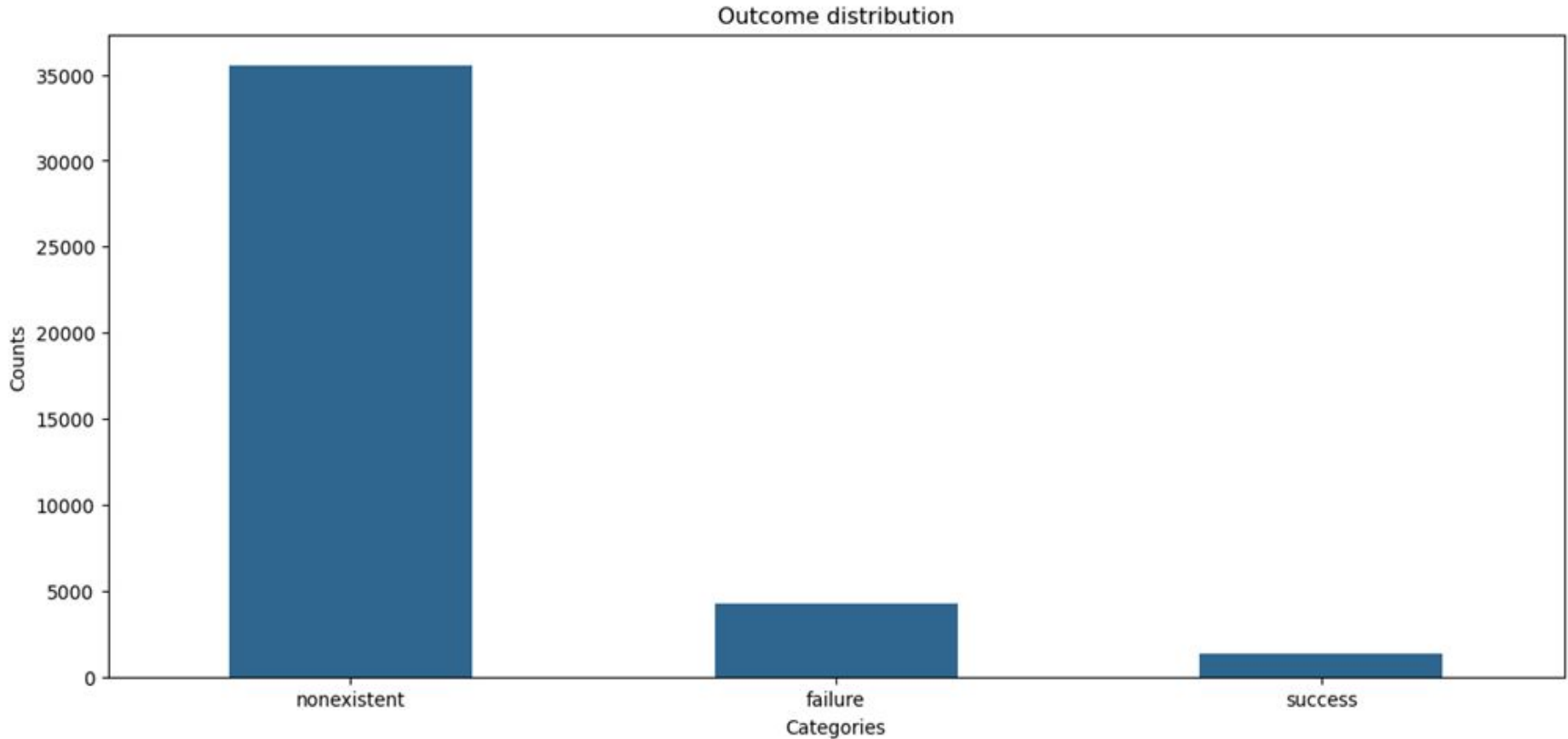
Most customers do not have personal loans

viii) Contact Categories



The contact is done to most customers via cellular not telephone.

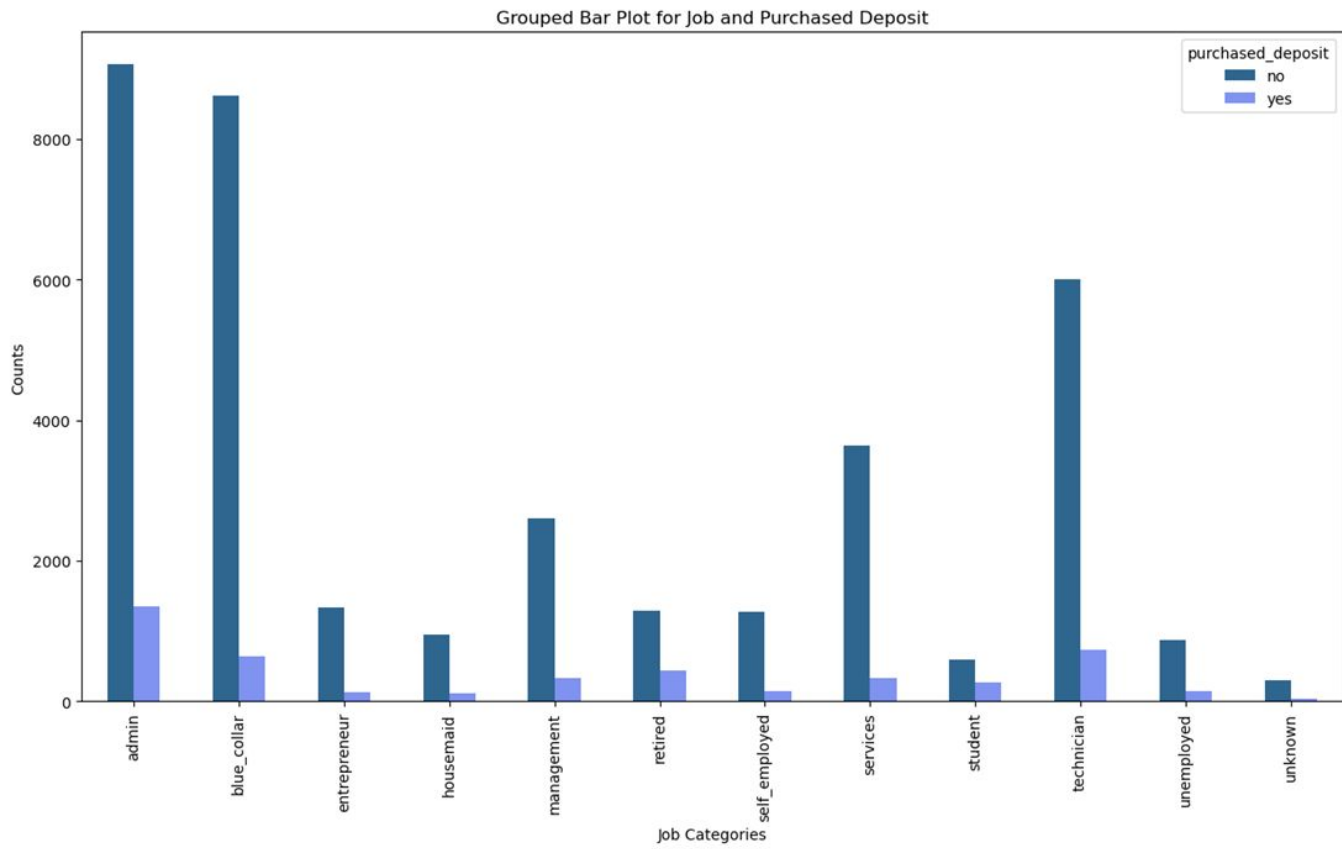
ix) Outcome of previous marketing campaigns



The above plot is communicating that most customers were not contacted during the previous marketing campaign and the failure was higher than the

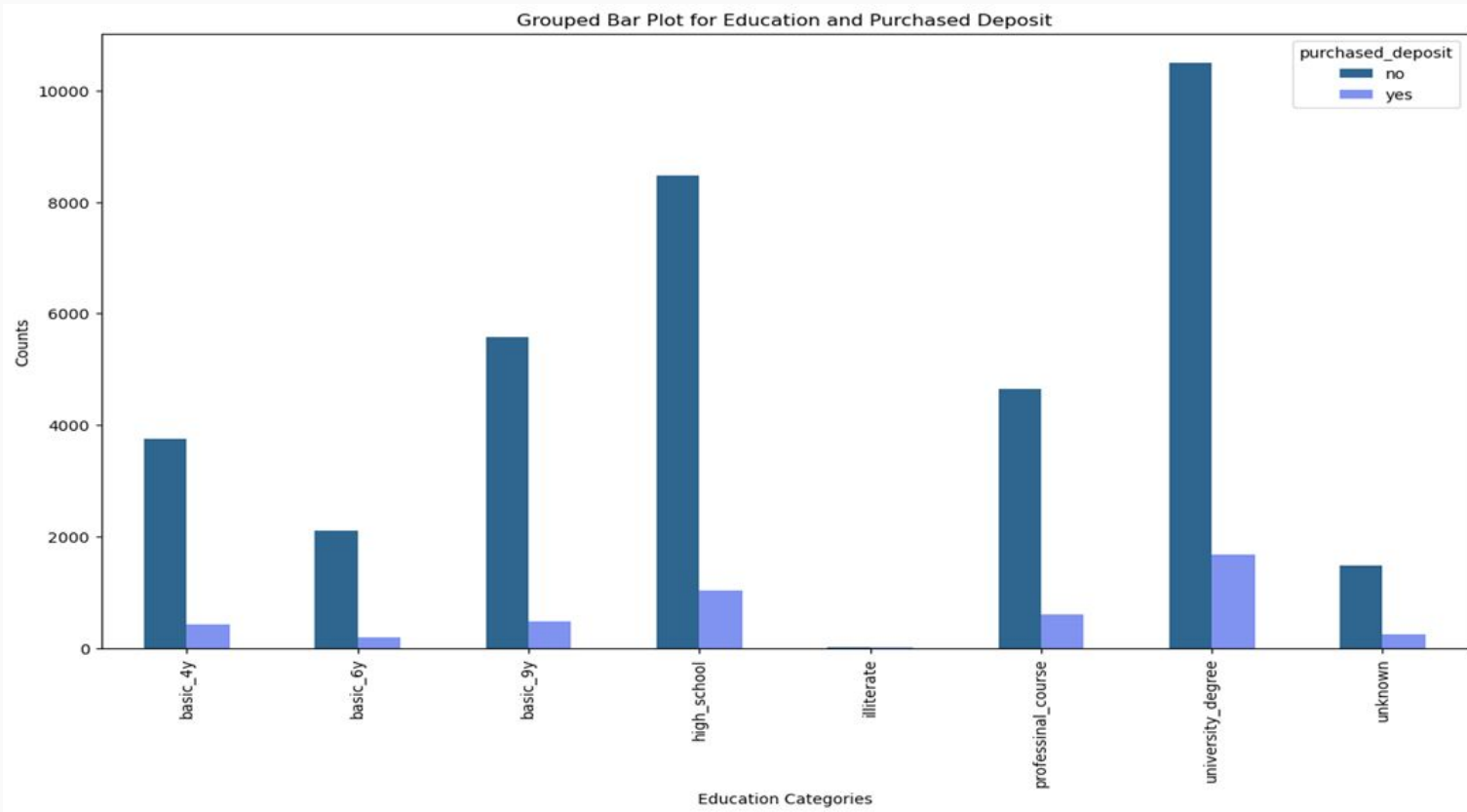
2.BIVARIATE ANALYSIS

i) Job by purchased deposit



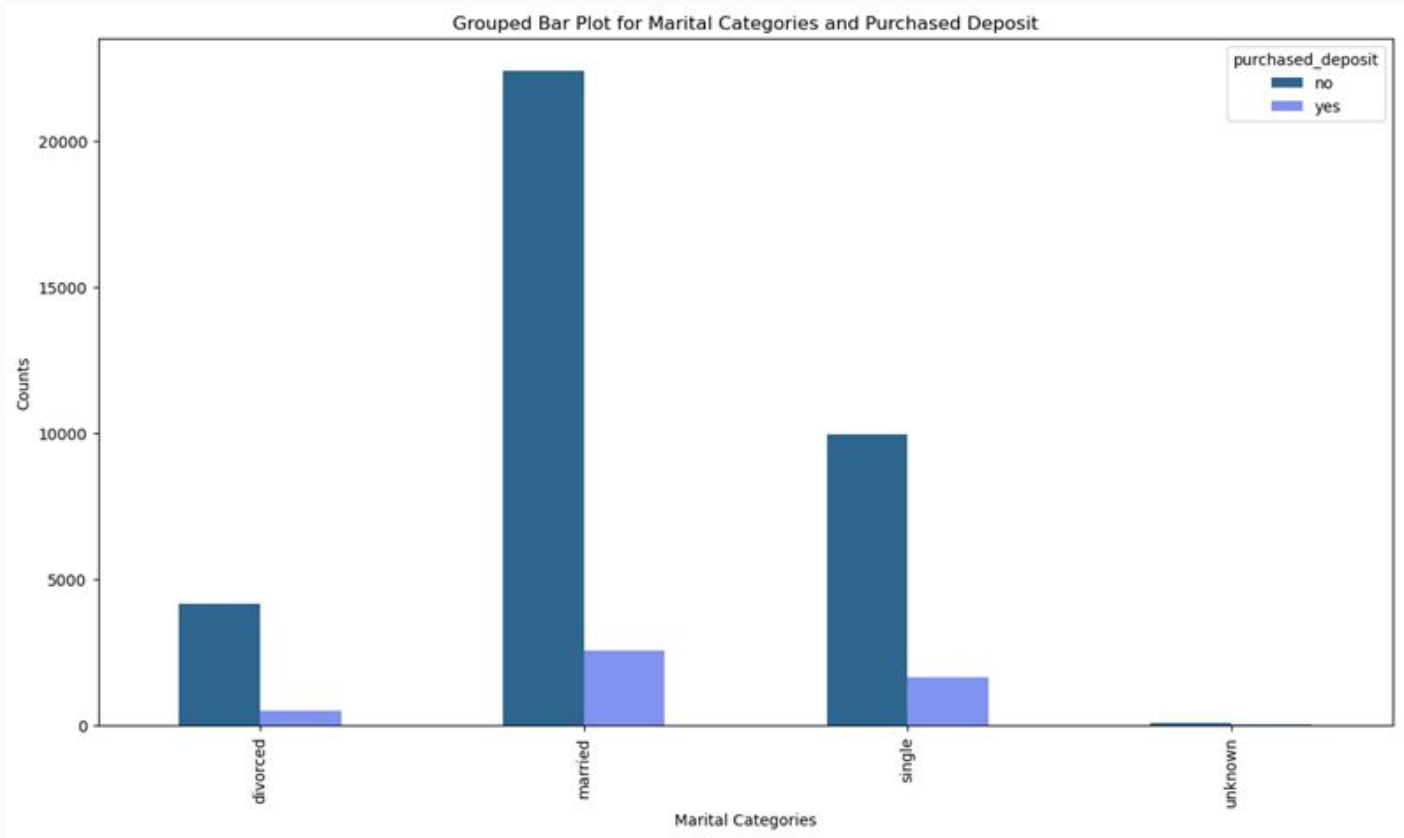
Most customers in their respective job categories did not purchase the term deposit.

ii) Education by purchased deposit



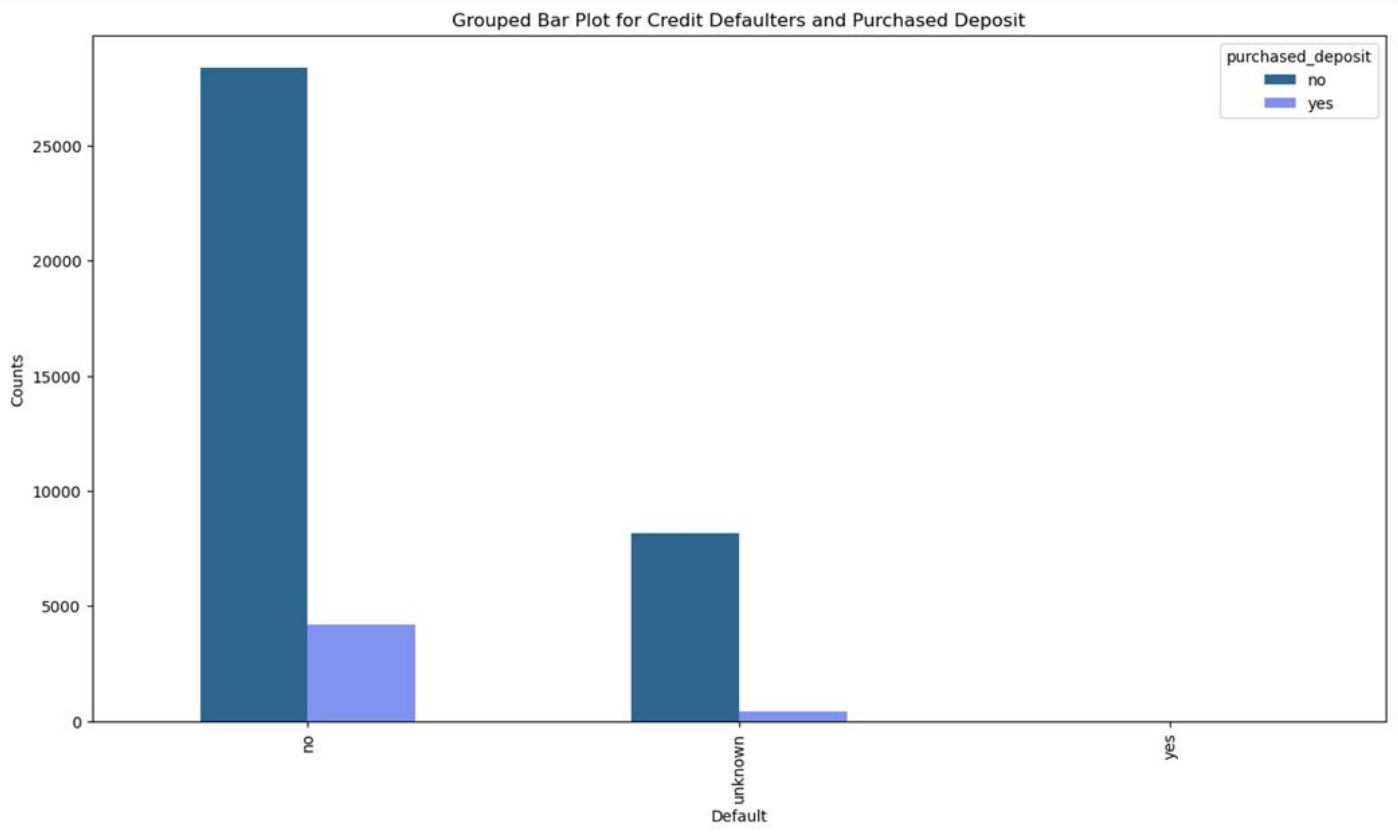
The conclusion from the above plot is also that most people did not purchase the term deposit in all

iii) Marital by purchased deposit



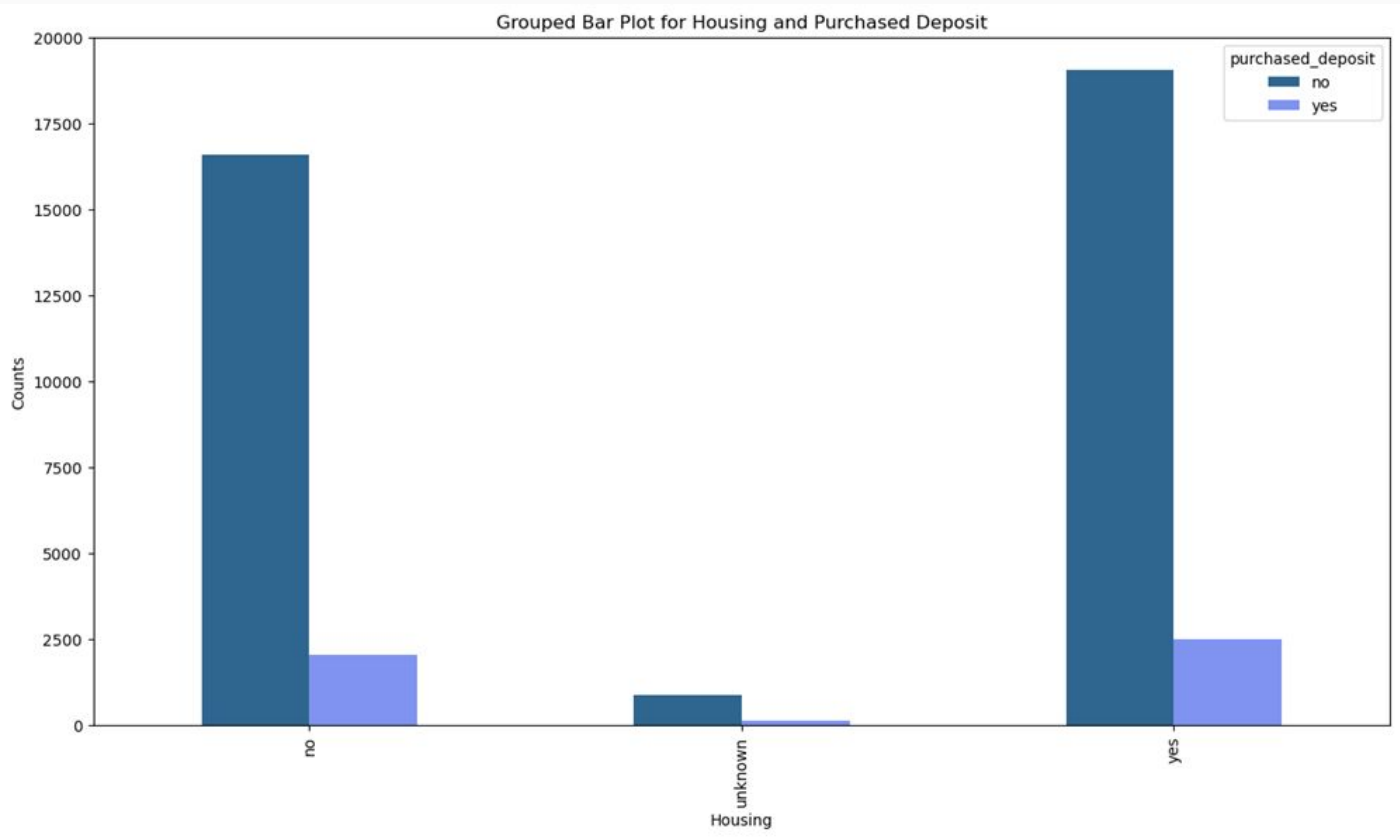
For marital categories, most customers also did not purchase the term deposit. But the leading group is the

iv) Credit Defaulters by purchased deposit



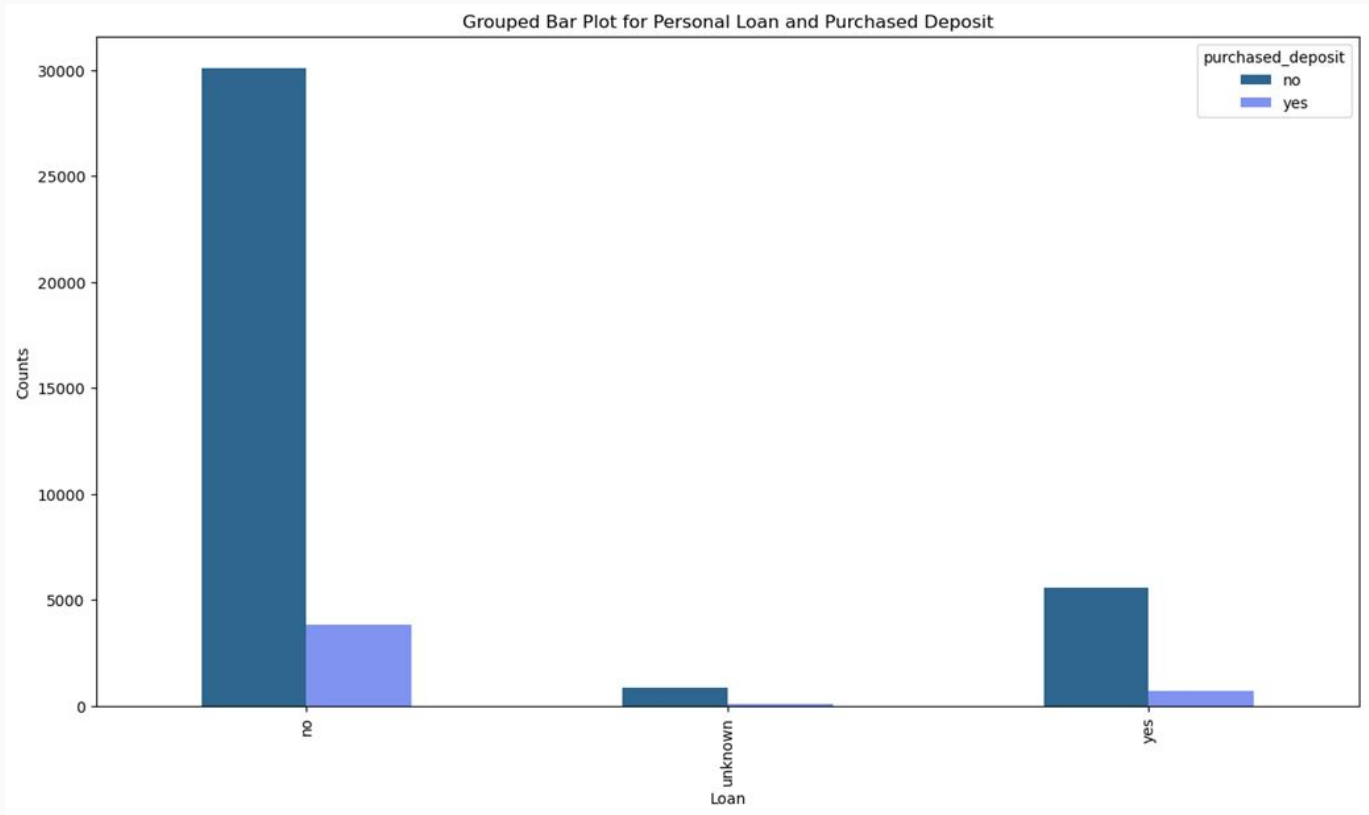
Most people who do not have credit defaults did not purchase the term deposit.

v) Housing by Purchased Deposit



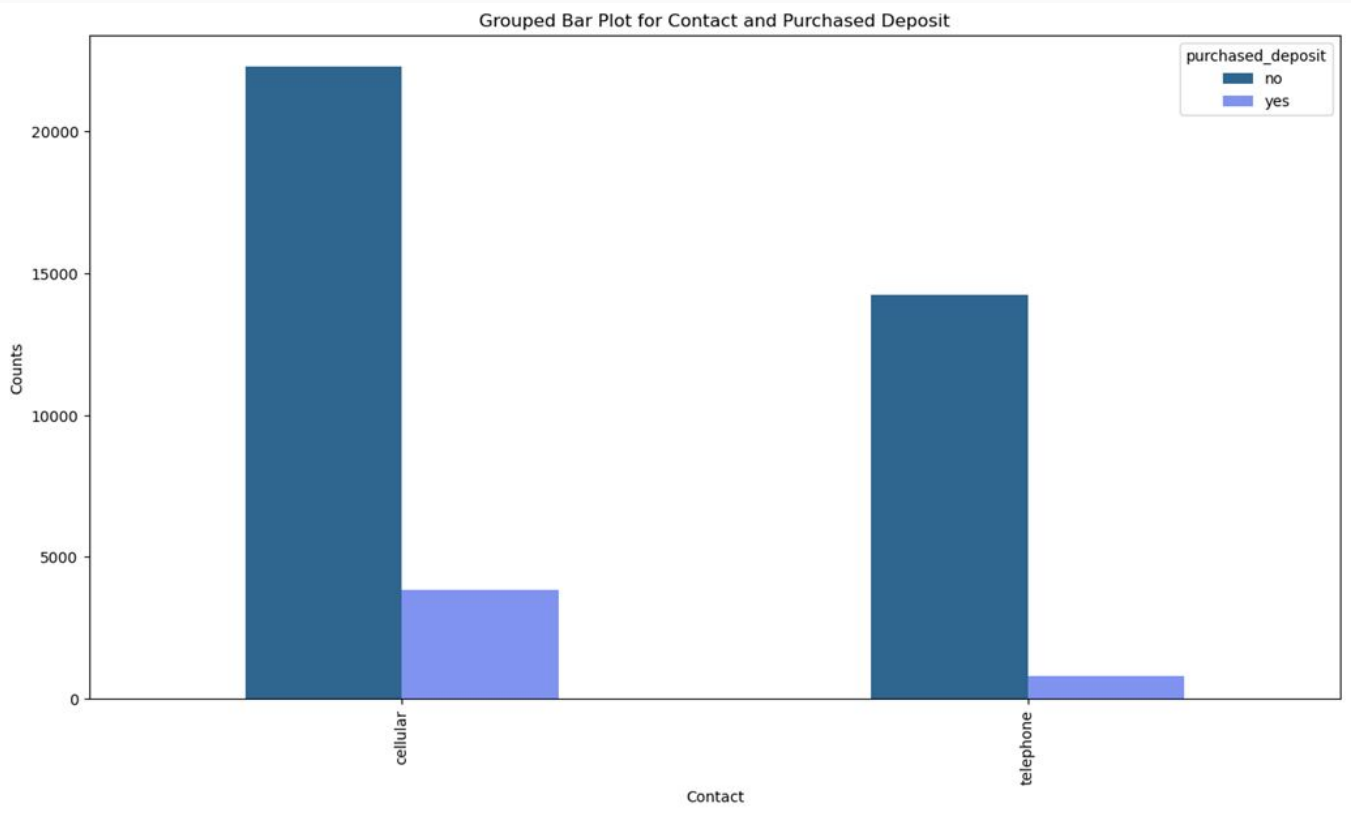
People with housing loans purchase the term deposits more.

vi) Loan Default by purchased deposit



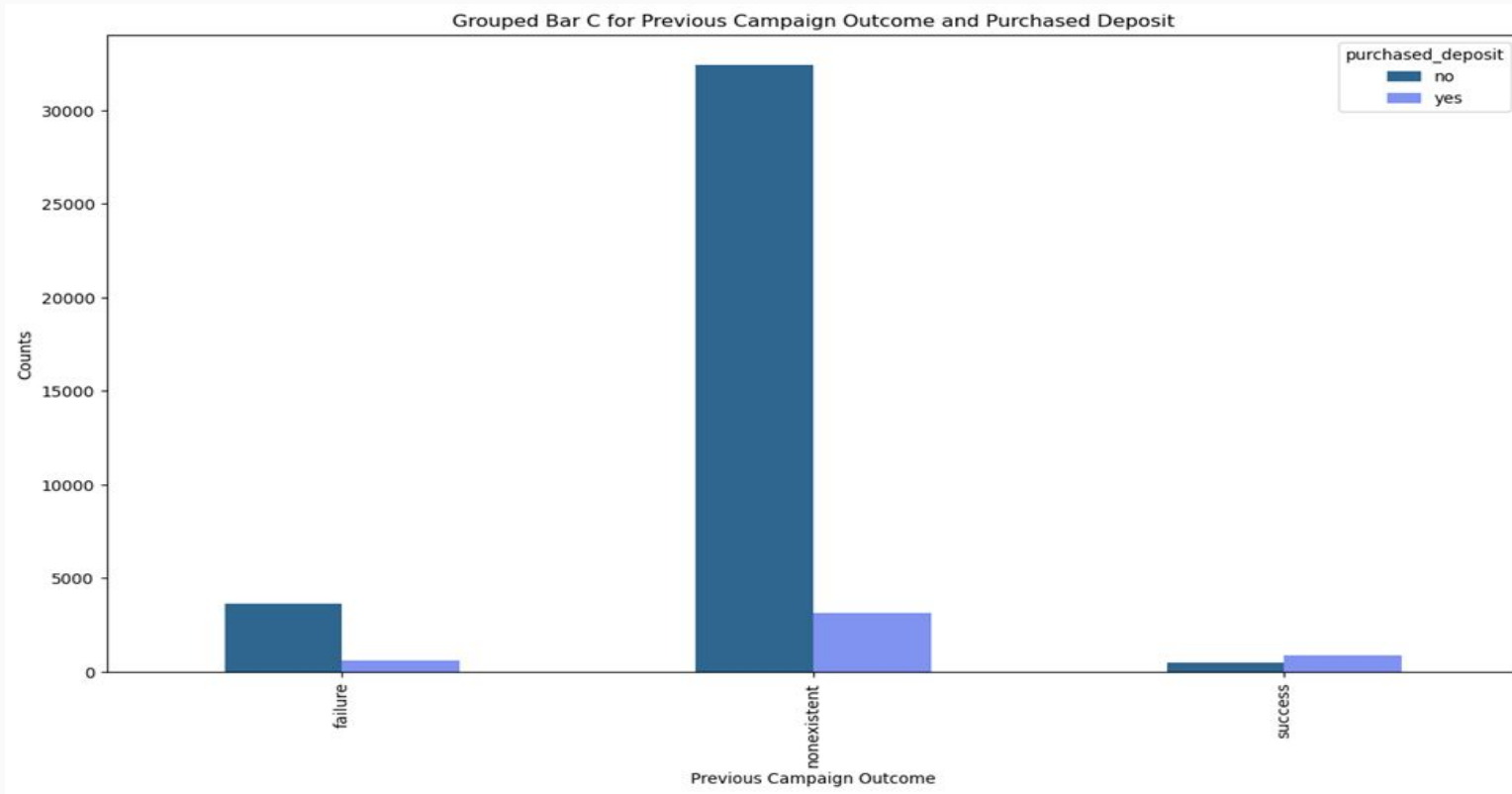
Most customers who do not have loans did not purchase the term deposit

vii) Contact by purchased deposit



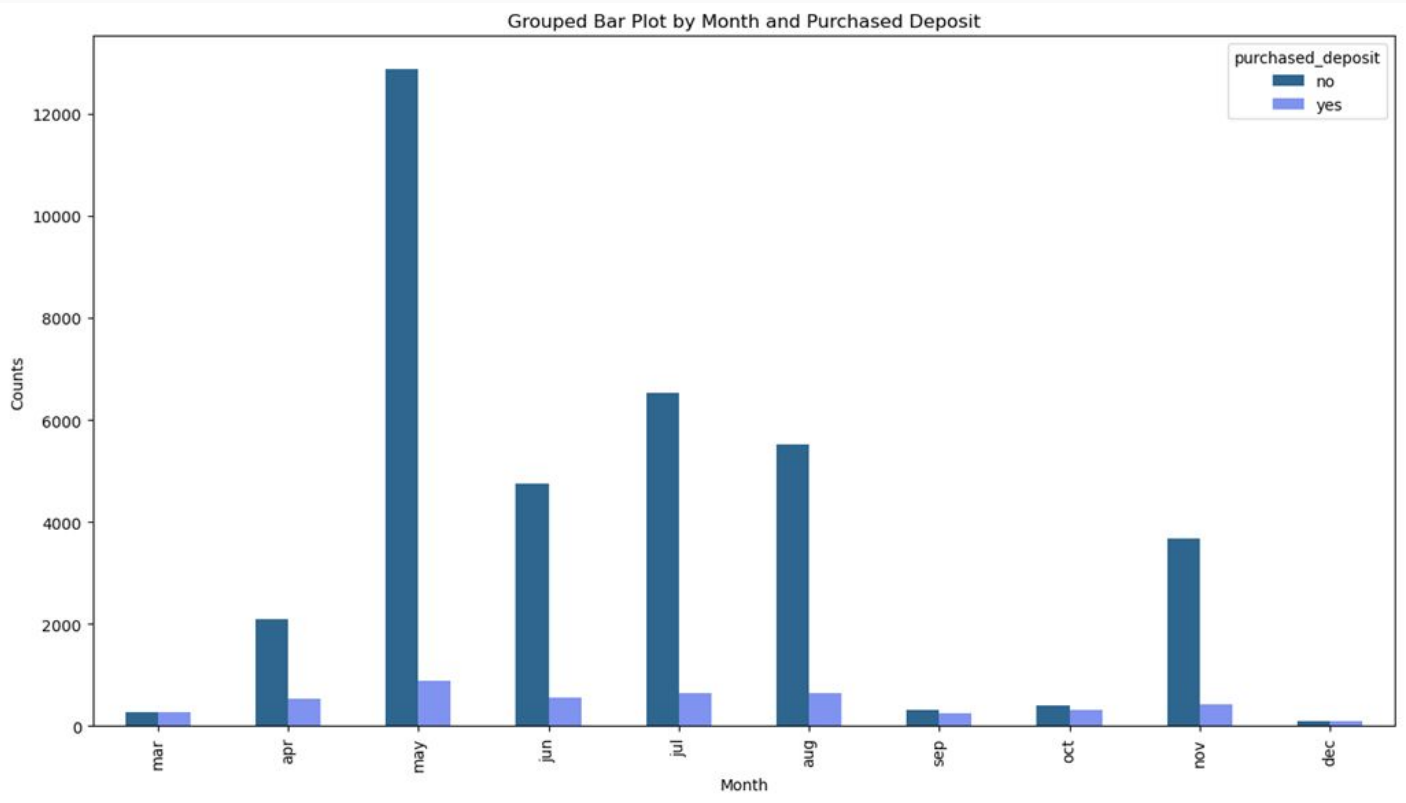
Most of the customers who were contacted by cellular and telephone did not purchase the term deposits but those who were contacted through cellular purchased the term deposit a bit more compared to those contacted through telephone

viii) Previous outcome of the campaign by purchased deposit



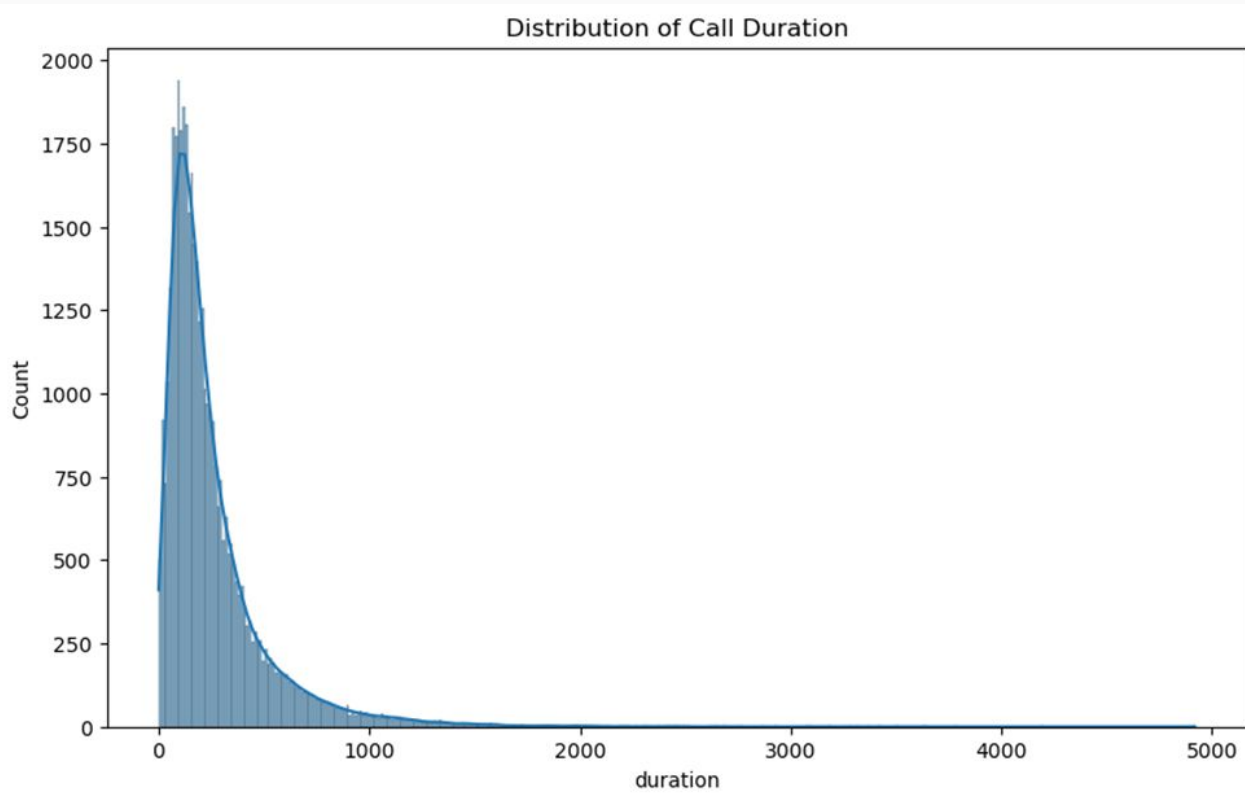
Many people were not contacted during the previous marketing campaign and hence not many people did not purchase the deposit

ix) Month by purchased deposit



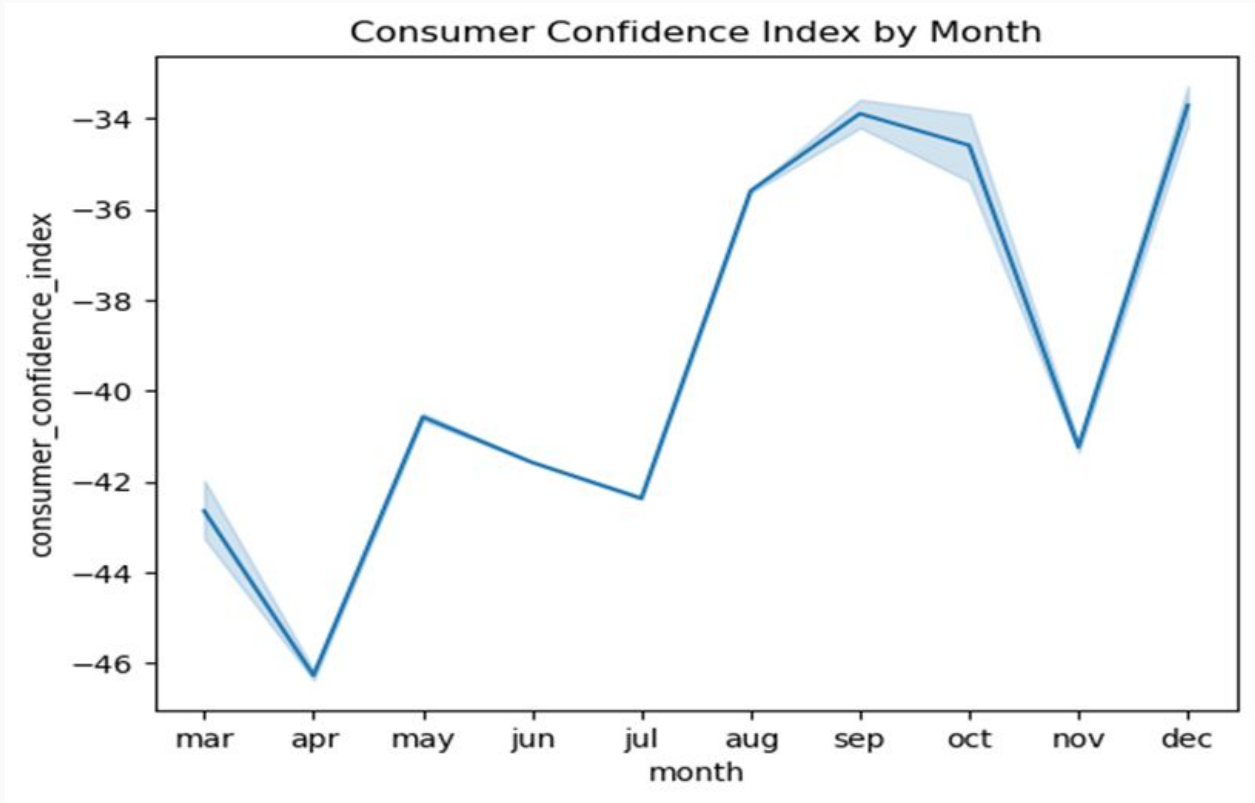
Most customers did not purchase term deposits regardless of the month but the months May, June, July, August (Mid-Year), the purchase was better

x) Distribution of call duration



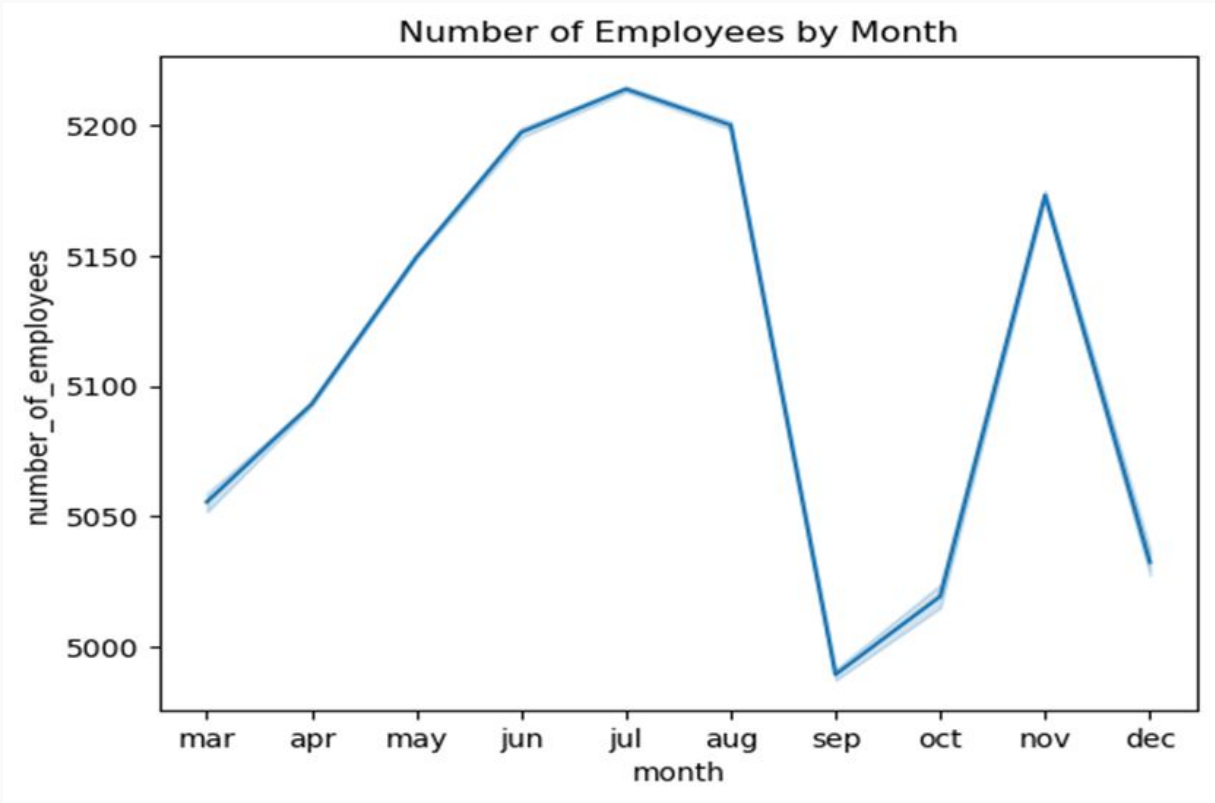
Observation is that the duration for most calls to customers approximately ranges between 0 to 500 seconds.

xiii) Consumer Confidence index by month



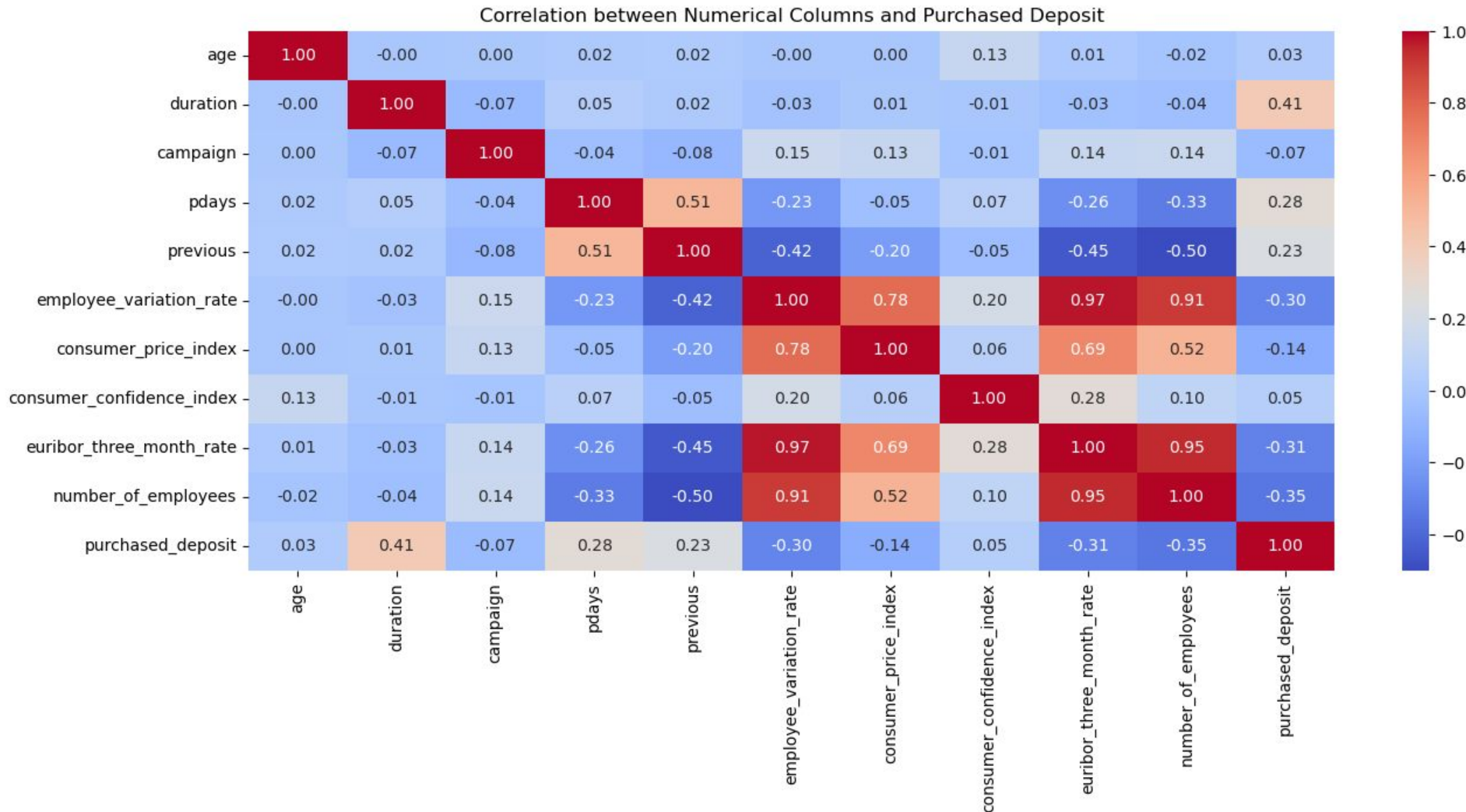
The plot shows a fluctuating trend on the consumer confidence index by months

xiv) Number of Employees by Month



According to the plot, the number of employees is on the rise from March to August

3.UNIVARIATE ANALYSIS



Observations from the correlation plot

- Duration is moderately positively correlated to purchased deposit. It has a positive correlation of 0.41
- Pdays and previous are weakly positively related to purchased deposits. With a correlation of 0.28 and 0.23 respectively.
- Age has a very weak positive correlation of 0.03 to purchased deposit.
- The following columns are negatively correlated to purchased deposit:
campaign, employee_variation_rate, consumer_price_index,
euribor_three_month_rate, number_of_employees.

4.RECOMMENDATIONS

1. Develop marketing campaigns that highlight the benefits of term deposits, such as security, returns, and suitability for long-term goals by promotions or incentives eg slightly higher interests for higher customers. The bank can consider holding webinars and workshops to educate customers on the importance of term deposits.

2. Increase the reach and frequency of marketing campaigns, particularly focusing on customers who were not contacted previously. By considering using cellular to contact customers as it is yielding more results compared to telephone.

3. Focus marketing efforts during periods of high consumer confidence or low inflation (e.g., early in the year or after mid-year). The bank can monitor economic indicators like the Consumer Confidence Index and Euribor rates to plan campaigns strategically.

4. Train agents to have meaningful, informative conversations that address customer concerns within the call duration 0-500 seconds as most calls are within that range. They can provide training to call agents on how to effectively pitch term deposits.

Recommended Models for this Dataset

The target variable for this dataset is `purchased_deposit`. That communicates whether the customer will purchase the deposit or not based on the independent variables.

Since the prediction is categorical, the following models can be used:

1. Logistic Regression
2. Decision Trees
3. Random Forest
4. Gradient Boosting Model eg XGBoost
5. Support Vector Machine