Group Name: Project on my own.
Name: Aditi Dadariya
Email: aditi.dadariya@gmail.com
Country: United Kingdom
College/Company: MSc (Data Science and Advanced Computing) from University of Reading
Specialization: Data Science
Project Name: Bank Marketing (Campaign)
Github Repo link: https://github.com/aditidadariya/BankMarketingCampaign

Problem description: ABC Bank wants to sell its term deposit product to customers and before launching the product they want to develop a model which help 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).

Exploratory Data Analysis: The dataset was analysed in three stages to gain insights and understand the relationships between variables.

In the first stage, a heat map was created to visualize the correlation between variables, and the Pearson Correlation coefficient was calculated. This helped identify the variables with the highest correlation. Additionally, a Pairplot was generated to provide a comprehensive view of the relationships between these highly correlated variables.

The second stage involved examining the unique values of each categorical variable and plotting them for better understanding. For numerical variables, statistical measures such as mean and standard deviation were computed to gain insights into their distribution and variability.

In the final stage, the relationship between the categorical variables and the target variable was examined using a bar plot. This visualization helped to identify any significant patterns or trends between the categorical variables and the target.

Overall, these three stages of analysis provided valuable insights into the dataset, uncovering important relationships and patterns that can inform further decision-making and analysis.

1. **Correlation between variables:** A heat map was generated to gain insights into the relationships between variables. As depicted in Fig 1, variables with a correlation coefficient greater than 0.5 are presented below along with their respective correlation values. Notably, the highest correlations are observed between EMP.VAR.RATE and EURIBOR3M, EMP.VAR.RATE and NR.EMPLOYED, and EURIBOR3M and NR.EMPLOYED. These findings suggest a potential relationship among these three variables. To further illustrate this relationship, a Pair plot is showcased in Fig 2.

   Correlation between EMP.VAR.RATE and CONS.PRICE.IDX: 0.7753

   Correlation between EMP.VAR.RATE and EURIBOR3M: 0.9722

   Correlation between EMP.VAR.RATE and NR.EMPLOYED: 0.9070

   Correlation between CONS.PRICE.IDX and EURIBOR3M: 0.6882

   Correlation between CONS.PRICE.IDX and NR.EMPLOYED: 0.5220

Correlation between EURIBOR3M and NR.EMPLOYED: 0.945

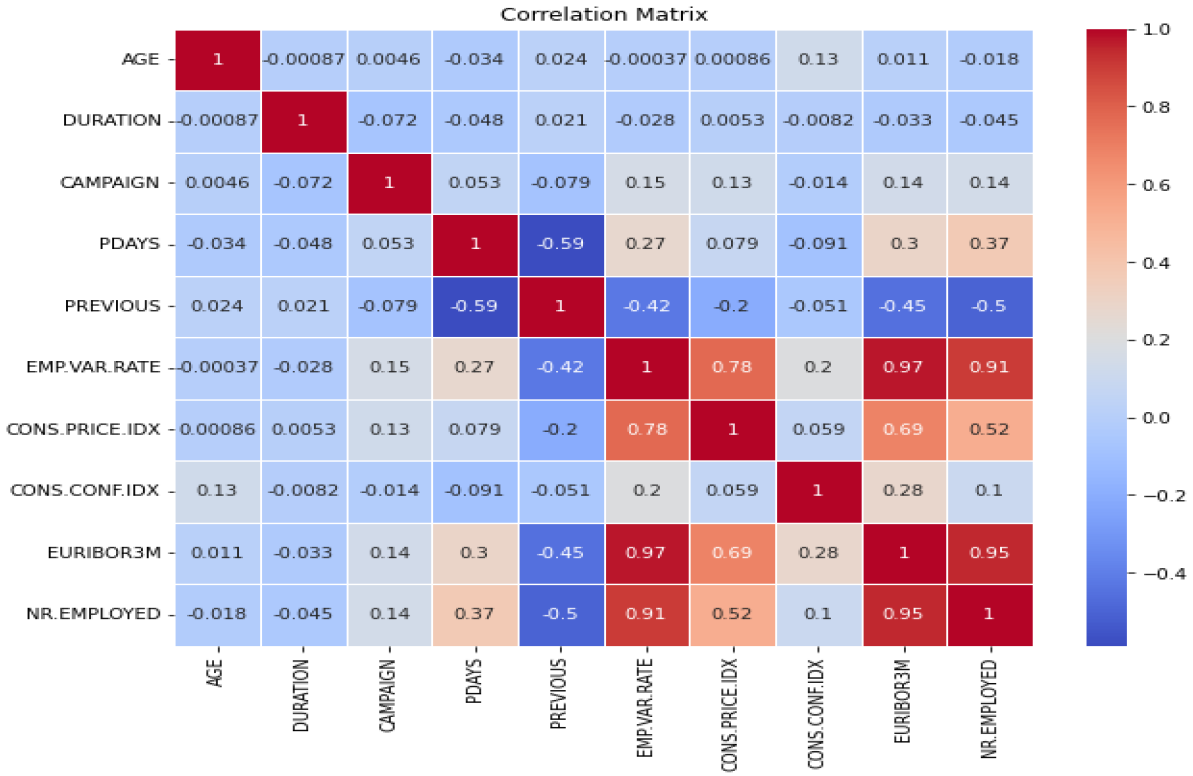

Fig 1: Heat Map

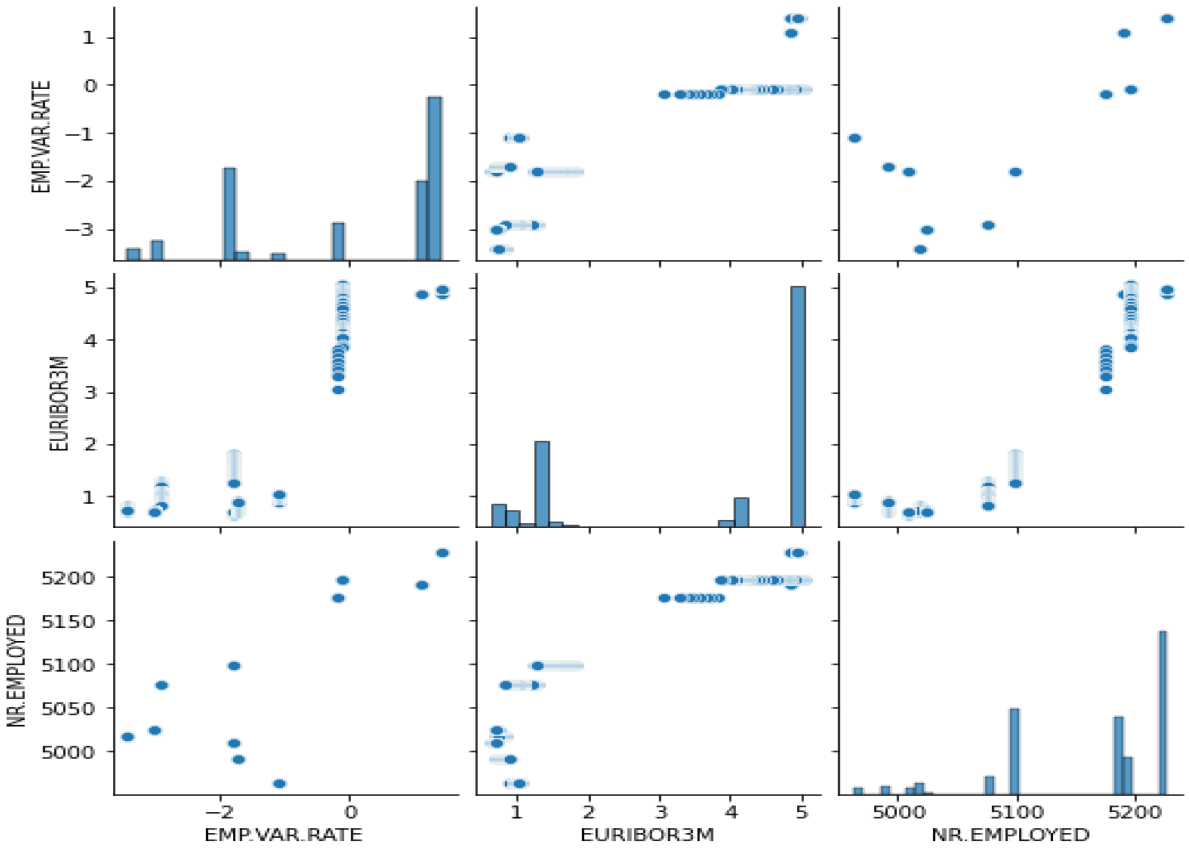

Fig 2: Pariplot

2. **Understanding each variable individually**: The unique values of categorical variables were explored and visualized using bar plots. The numerical variables were analyzed by calculating their mean and standard deviation, providing insights into their distribution. Please refer to Figures 3 to 23 for the detailed visualizations.

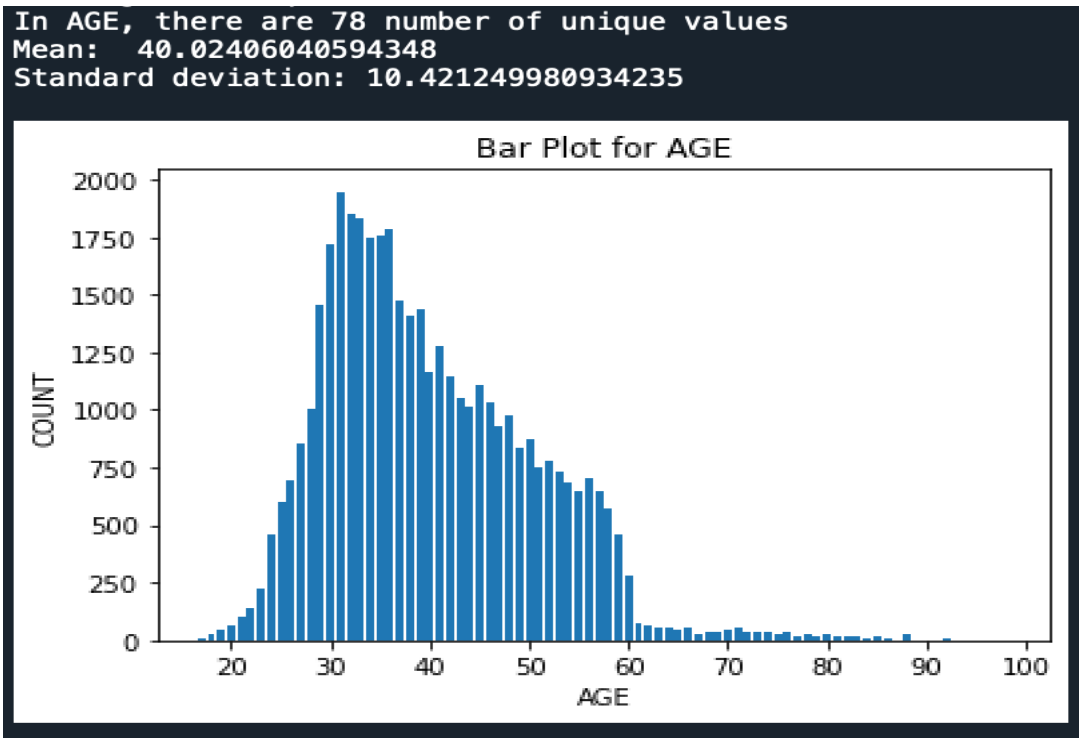

Fig 3: Age variable

Fig 3 shows that the age range of customers who are likely to be eligible for a term deposit is between 20 and 60 years.

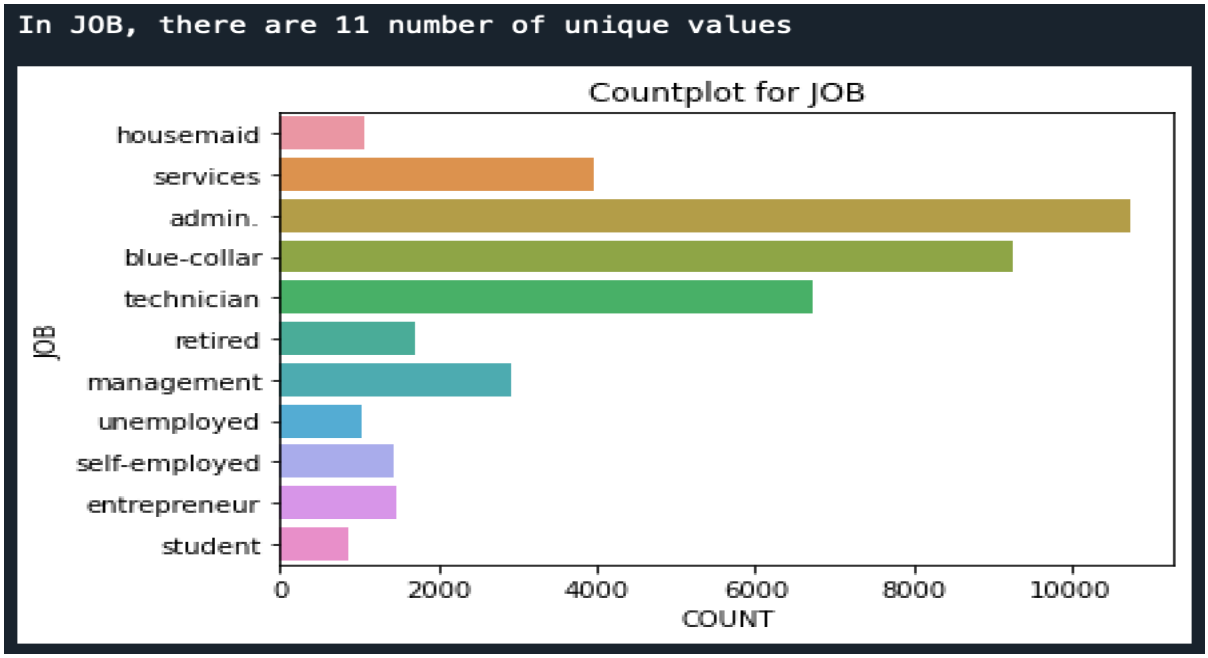

Fig 4: Job variable

Among the various job categories, administrative jobs have the highest representation, as depicted in Fig 4.

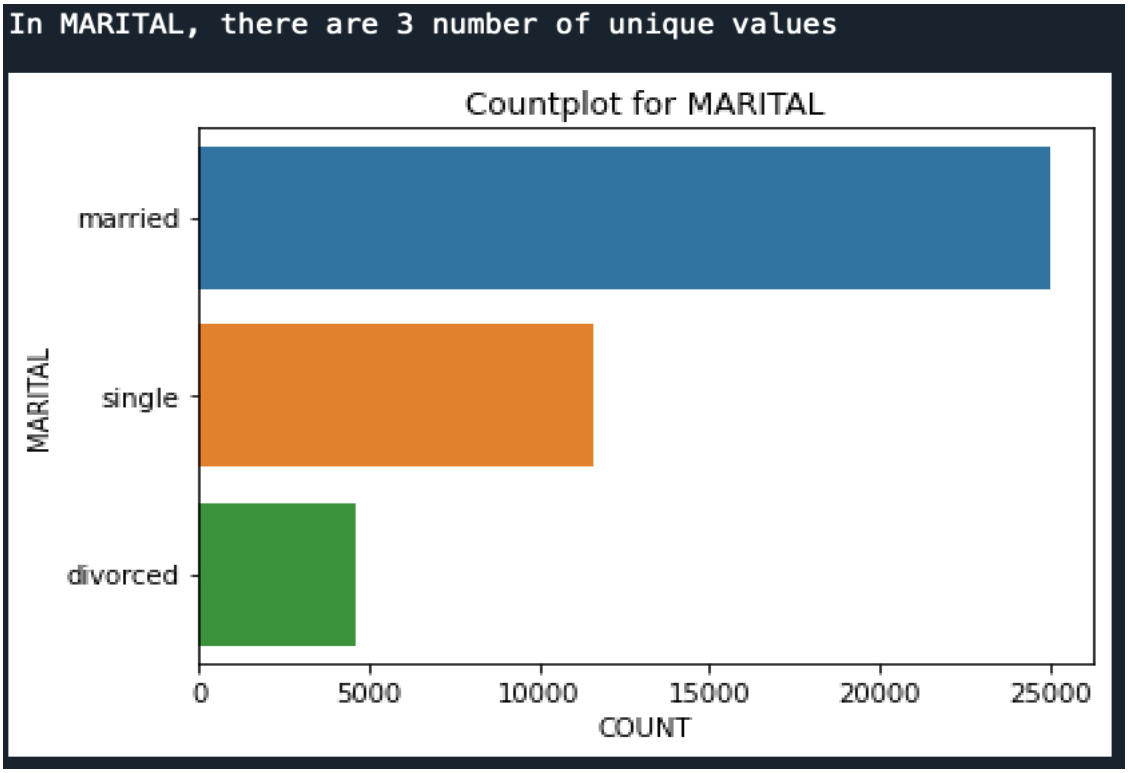

Fig 5: Marital Variable

Fig 5 indicates that the number of married customers surpasses that of single and divorced customers, suggesting their eligibility for term deposit offers.

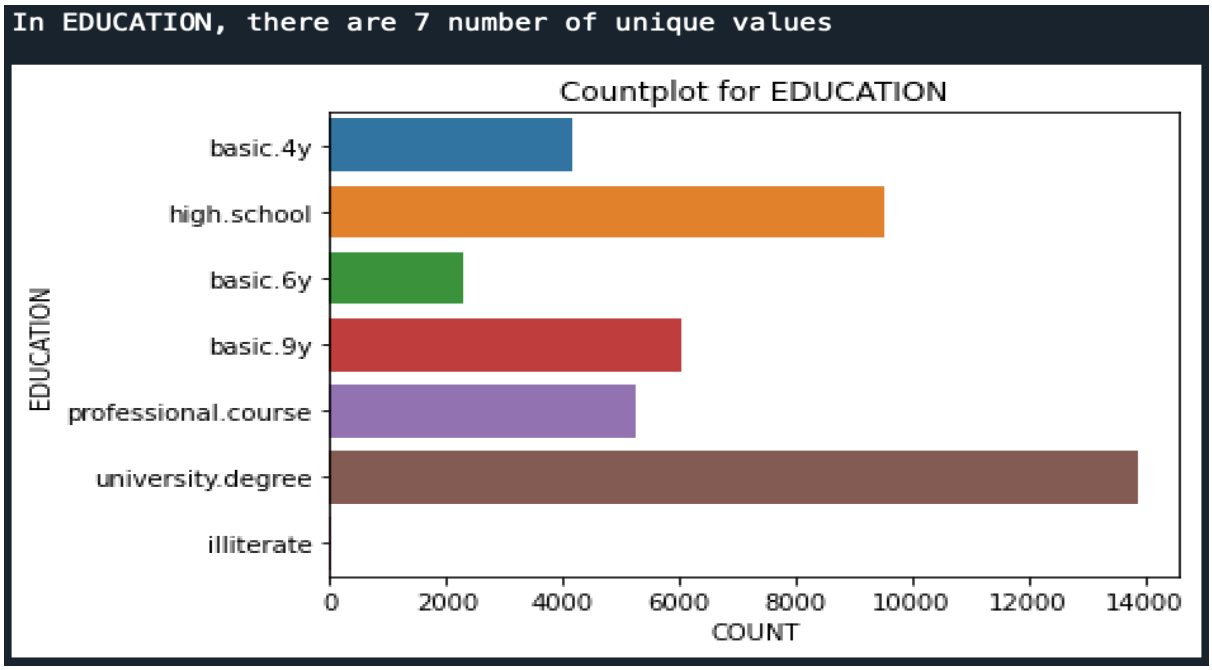

Fig 6: Education variable

Fig 6 showcases the presence of university degree holders who are potentially targeted for term deposit contact.

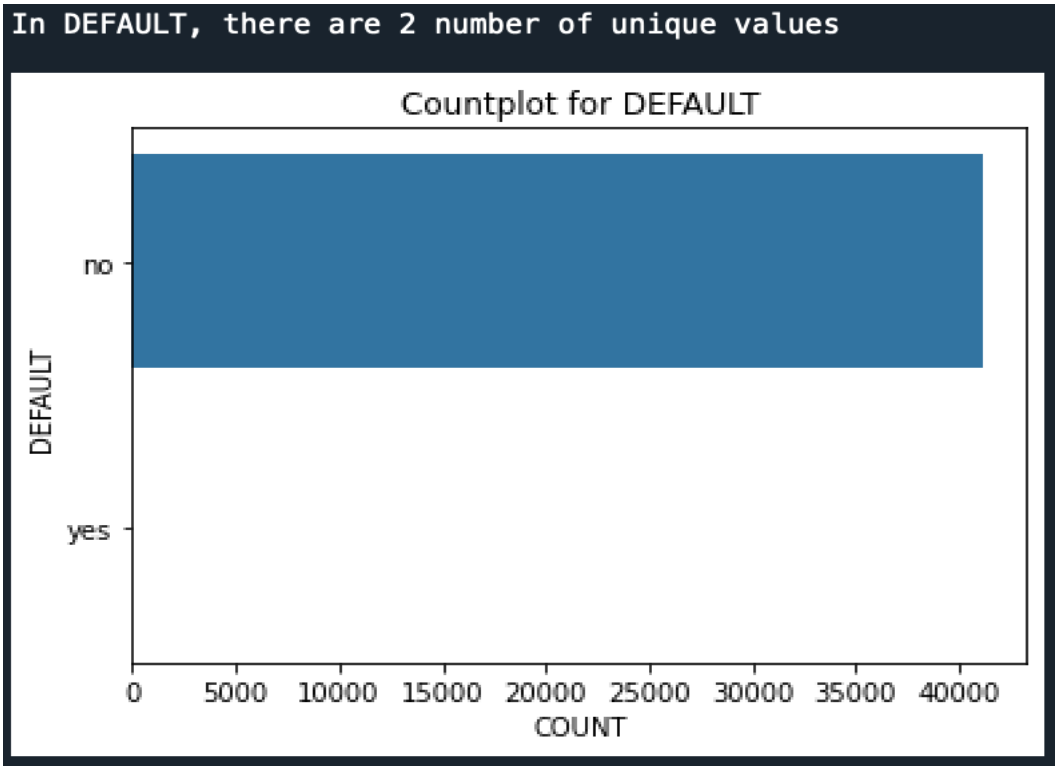

Fig 7: Default variable

Fig 7 reveals that the majority of customers do not have any credit default.

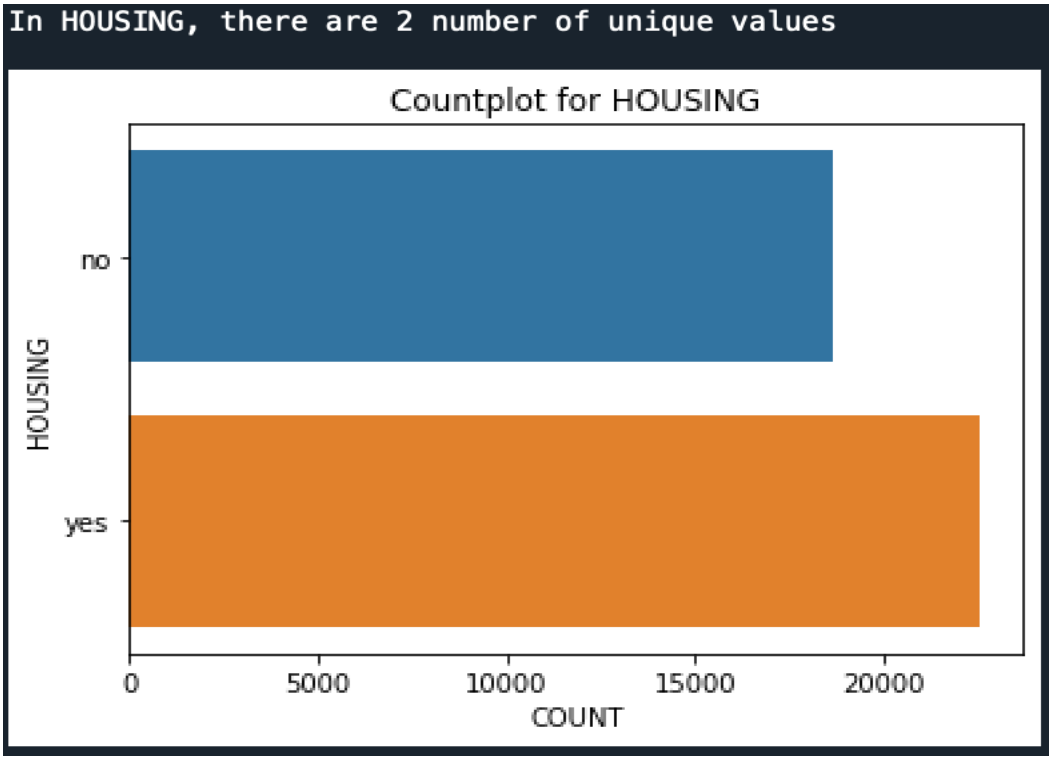

Fig 8: Housing variable

Fig 8 illustrates that the number of customers with housing loans is greater than the number of customers without housing loans.

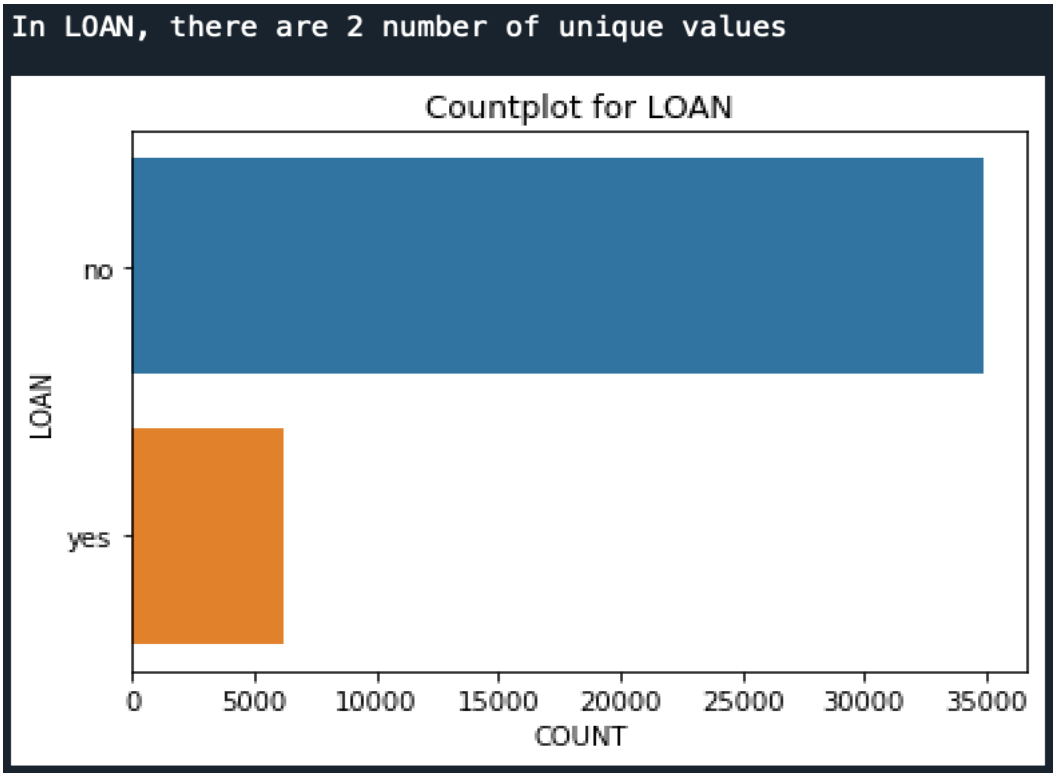

Fig 9: Loan variable

As shown in Fig 9, the count of customers with personal loans is significantly lower than the count of customers without personal loans..

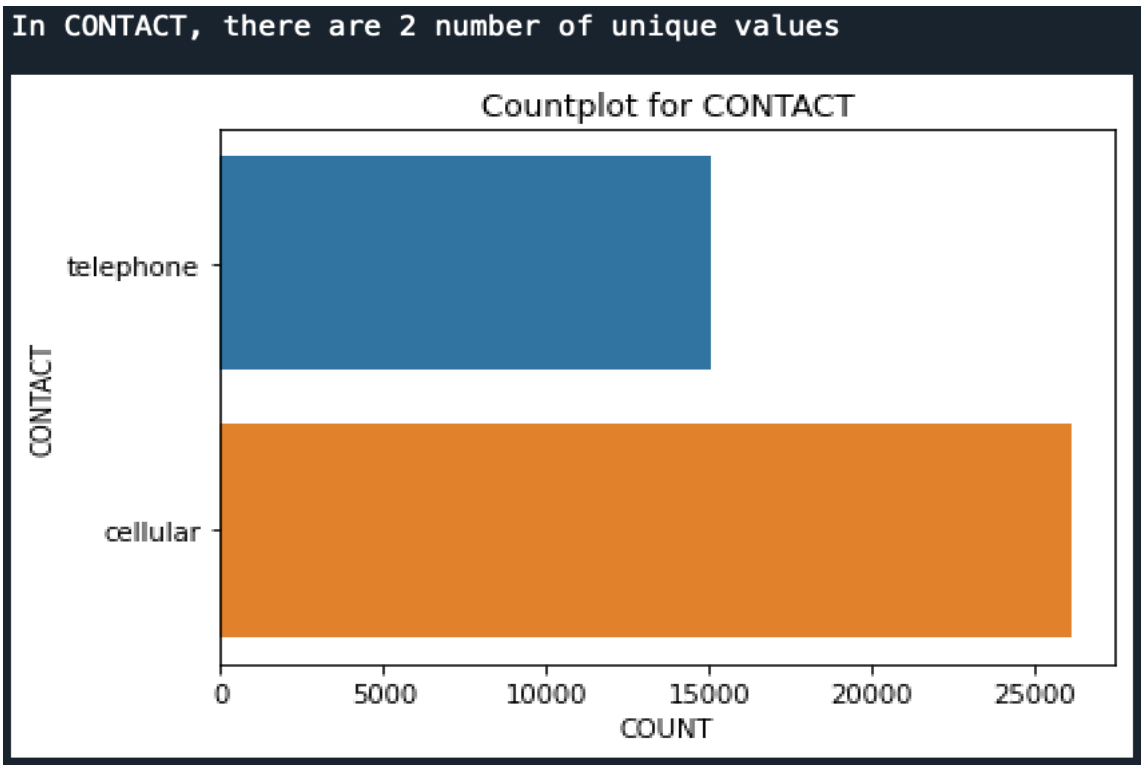

Fig 10: Contact variable

As depicted in Fig 10, the count of customers with cellular phones is higher compared to customers with telephone connections.

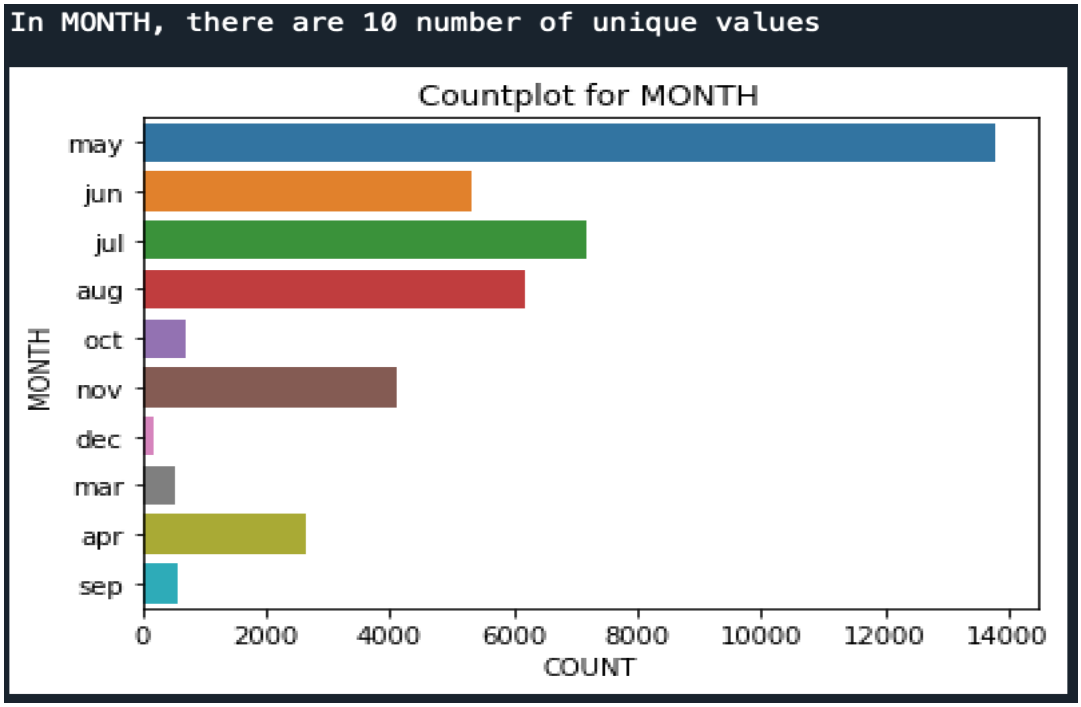

Fig 11: Month Variable

As shown in Fig 11, the majority of customers were contacted during the month of May in the past.

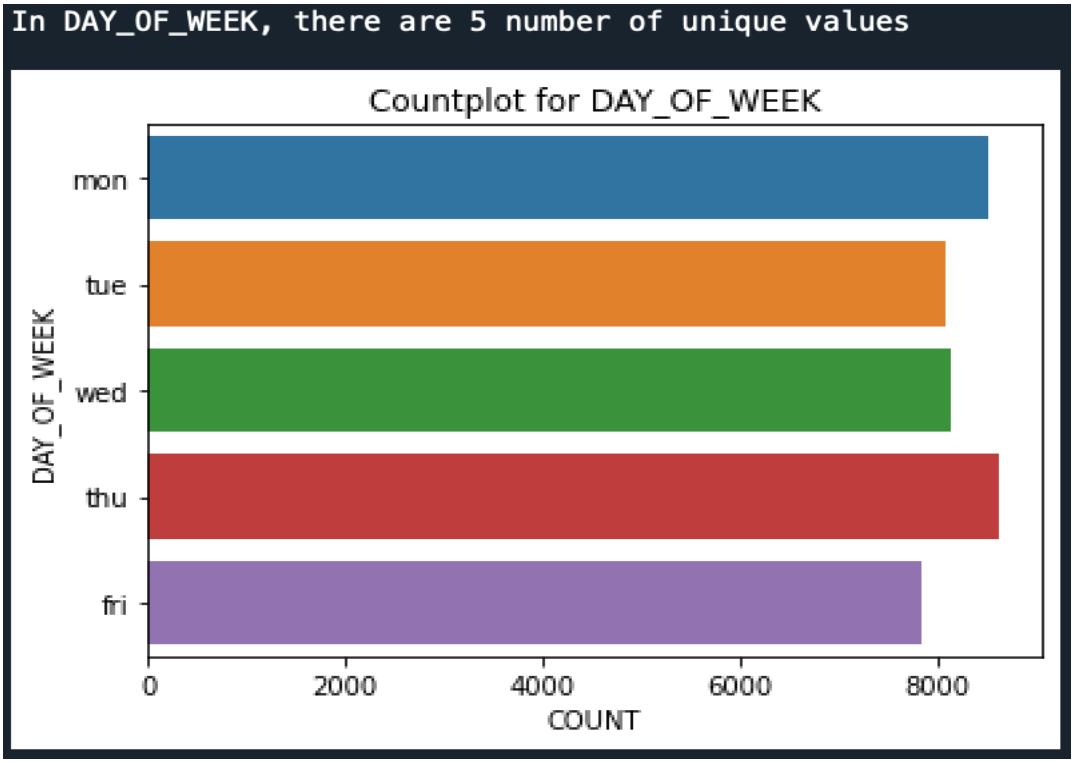

Fig 12: Day_of_Week

As depicted in Fig 12, the number of customers contacted during the weekdays shows a slight increase on Mondays and Thursdays, although the overall impact is not substantial.

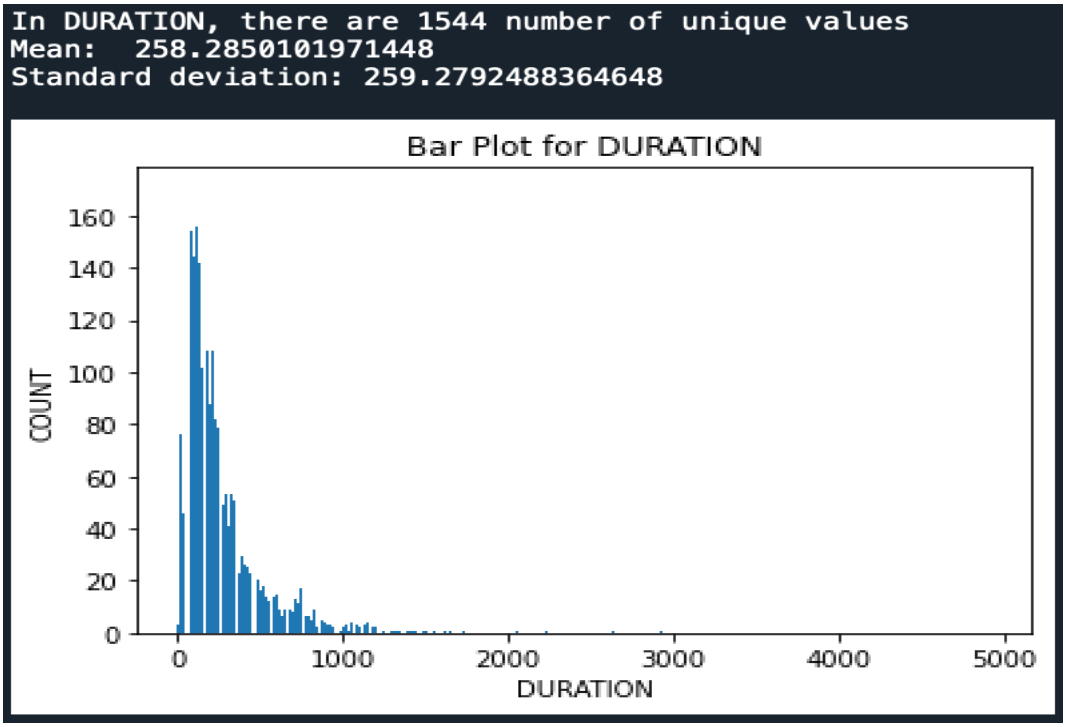

Fig 13: Duration variable

As shown in Fig 13, the duration of calls ranged from 0 to 1000 seconds. On average,the calls lasted for approximately 258.28 seconds with the customers.

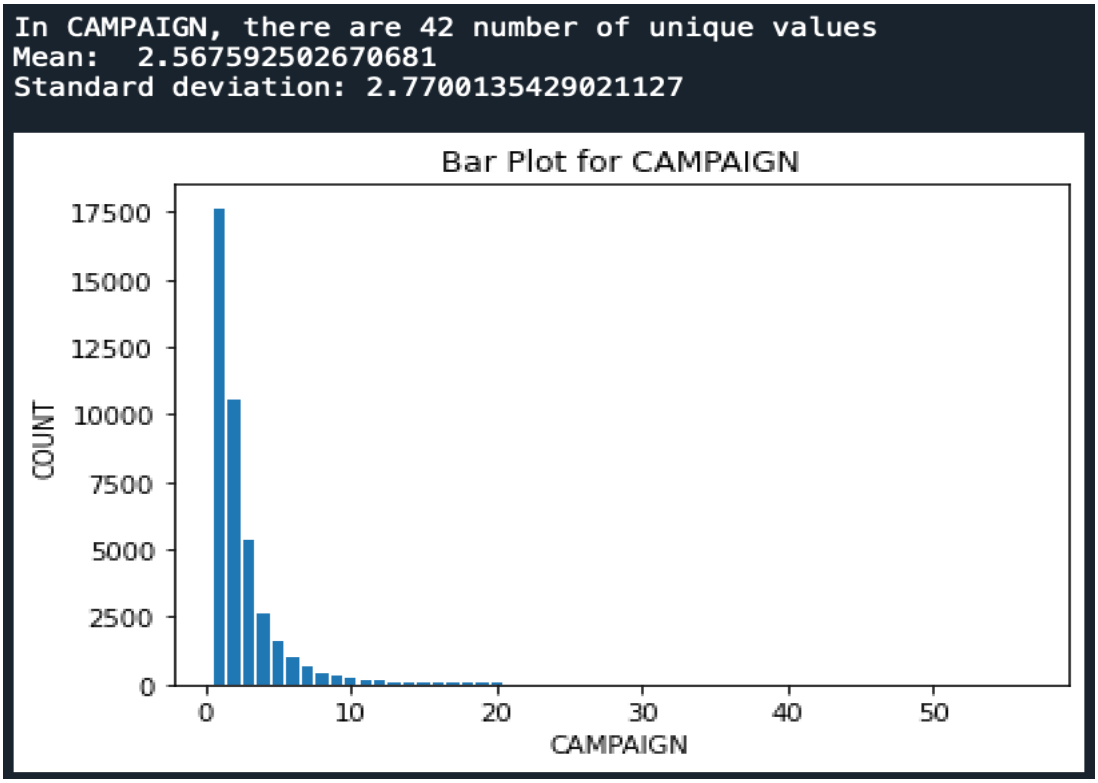

Fig 14: Campaign variable

As shown in Fig 14, the number of contacts made during the campaign ranged from 1 to 10, with a significant proportion of customers being contacted only once.

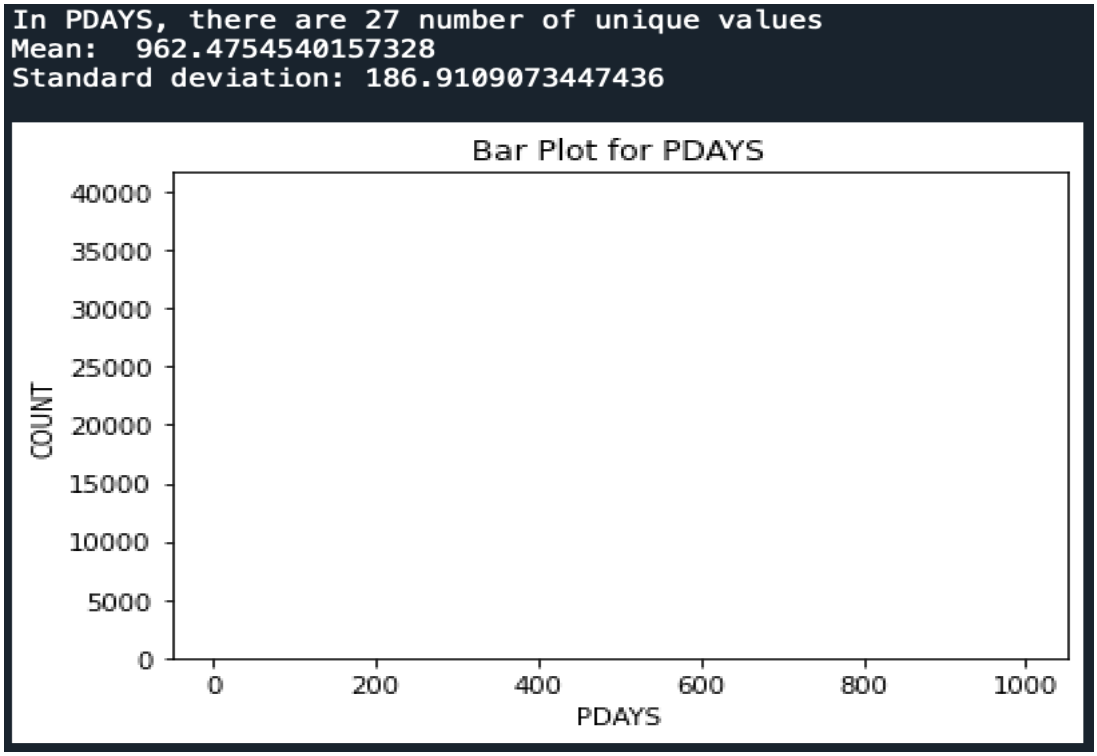

Fig 15: PDays variable

As depicted in Fig 15, the average number of times customers were contacted before the campaign was relatively low.

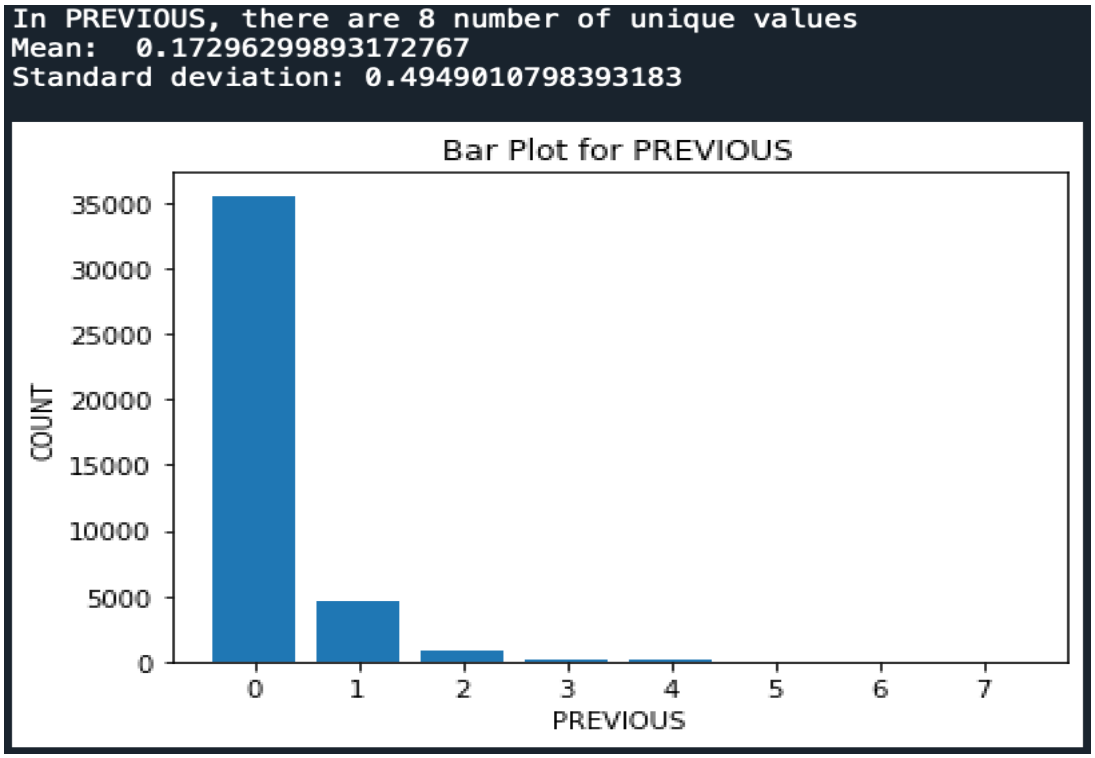

Fig 16: Previous variable

As shown in Fig 16, the number of calls made to customers prior to the campaign is relatively low. Only a small number of customers have been contacted once or twice before the campaign..

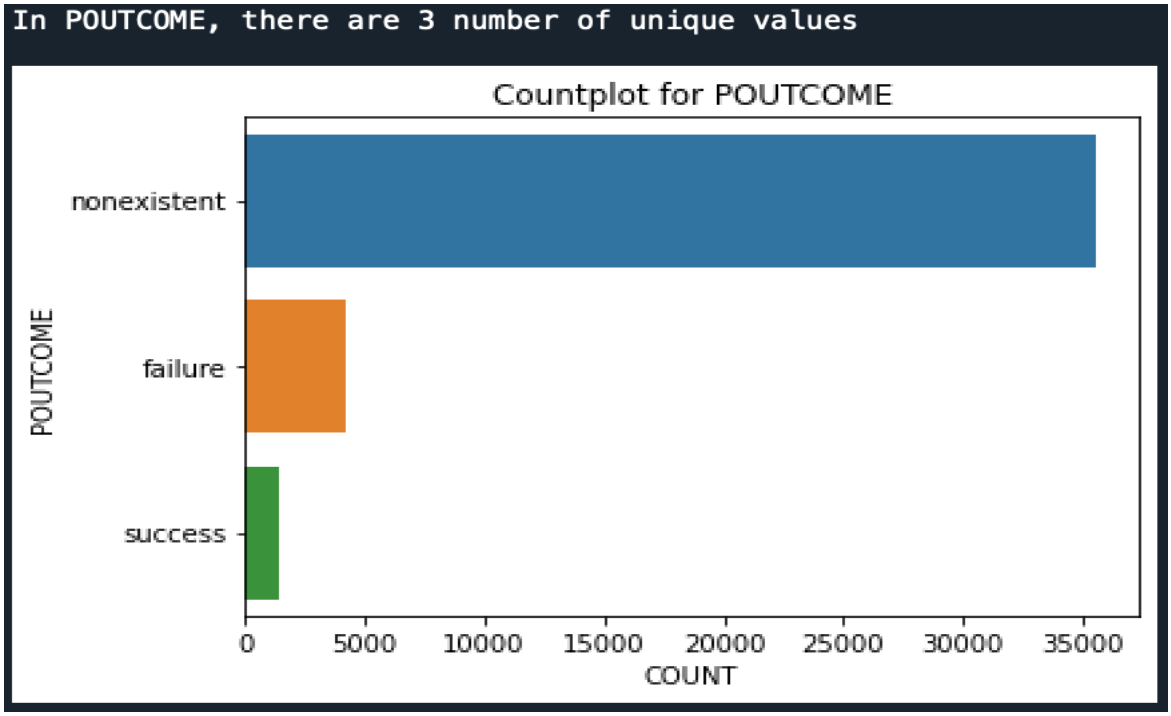

Fig 17: POutcome variable

Based on Fig 17, the majority of the previous marketing campaign outcomes are non-existent, with a smaller number of successes and failures.

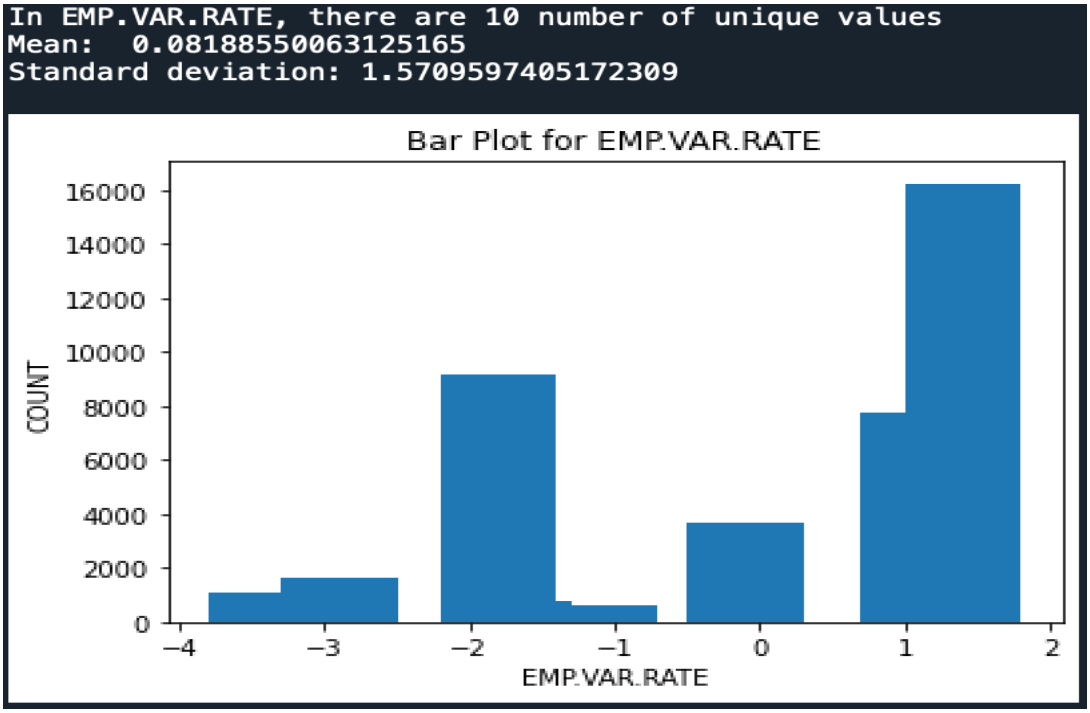

Fig 18: EMP.VAR.RATE variable

Based on Fig 18, the employment variation rate is highest between 1 and 2, with an average of 0.0818.

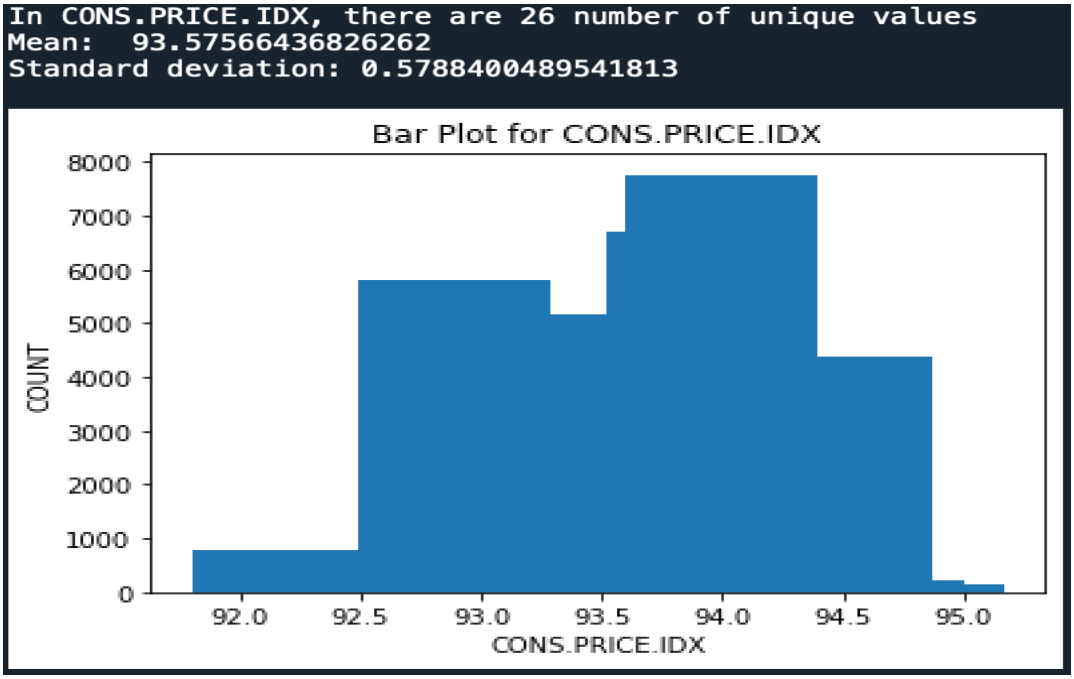

Fig 19: CONS.PRICE.IDX

According to Fig 19, the average customer price index is 93.57, with a range between 92.5 and 94.5.

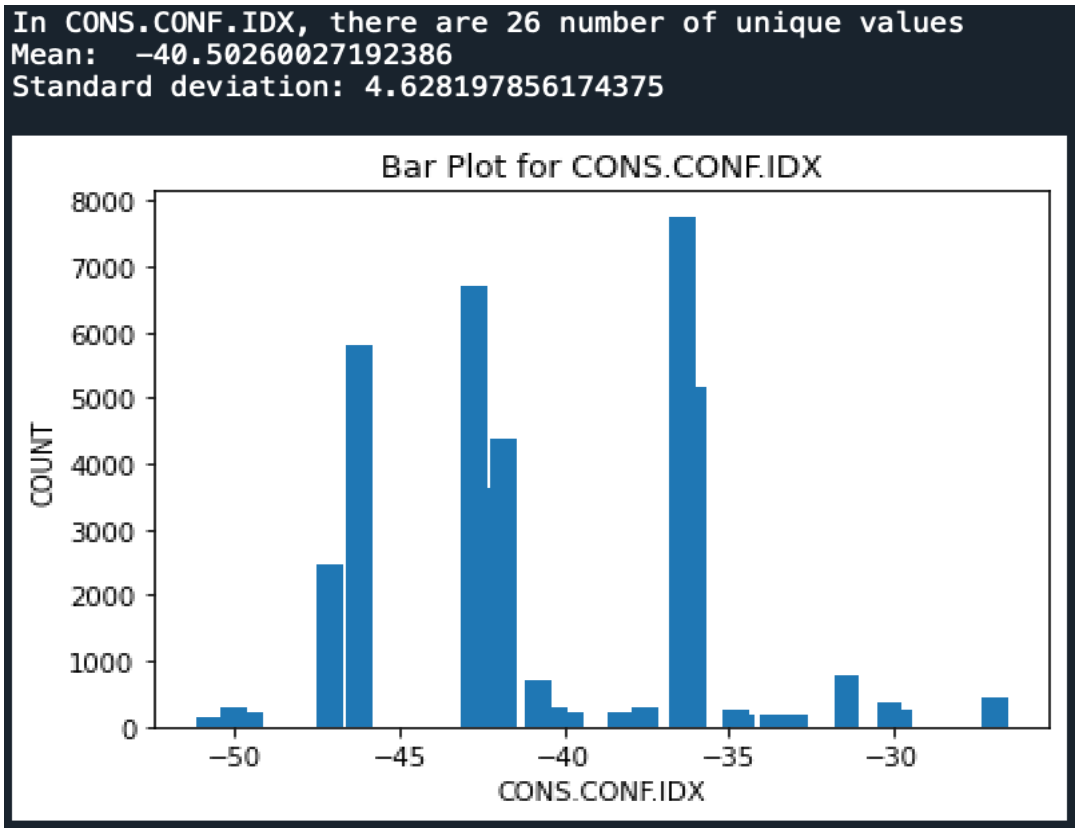

Fig 20: CONS.CONF.IDX

Fig 20 indicates that the customer confidence index ranges from -35 to -38, with an average value of -40.50.

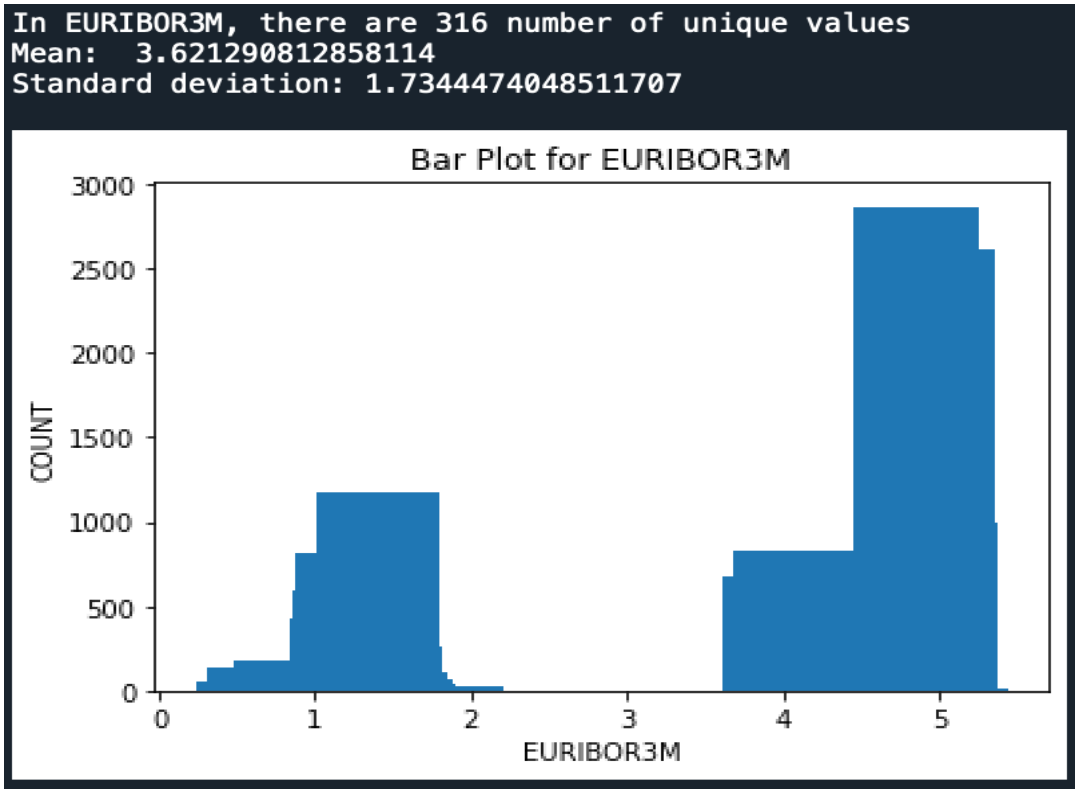

Fig 21: EURIBOR3M variable

Fig 21 shows that the maximum Euribor 3 month rate falls within the range of 4 to 5.

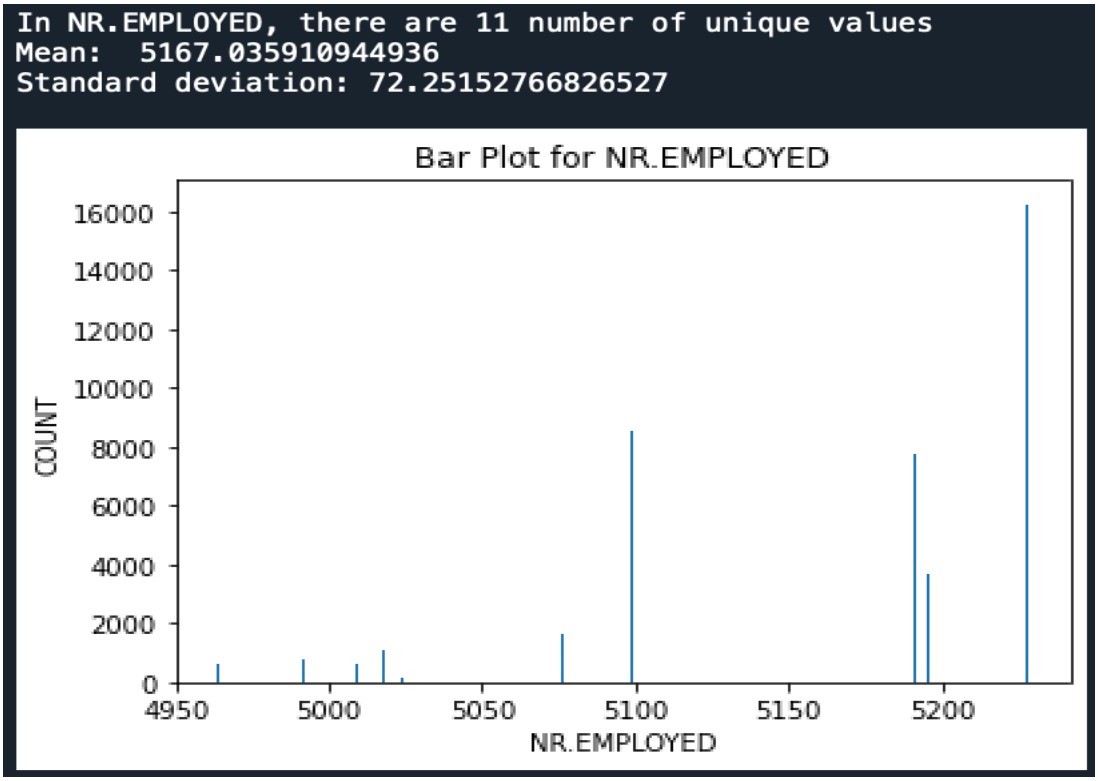

Fig 22: NR.Employed variable

Fig 22 illustrates that the average number of employees is 5167, with a maximum exceeding 5200.

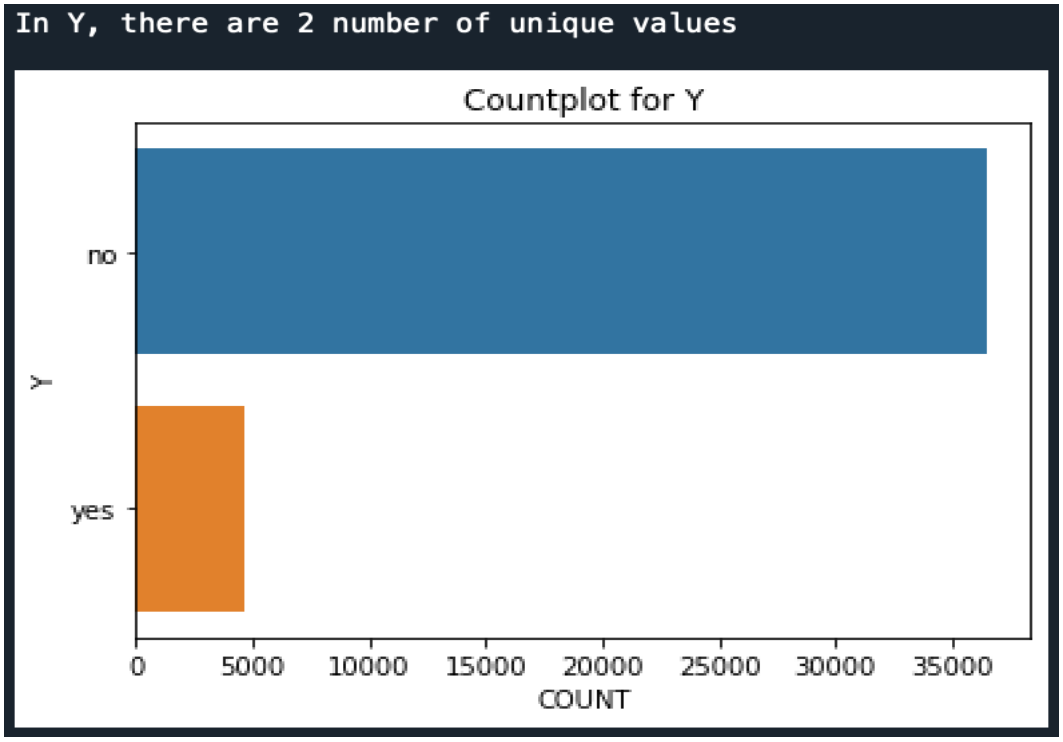

Fig 23: Y target variable

Fig 23 reveals that a small number of customers already have a term deposit.

3. **Relation of categorical variable with target variable:** Below are the plots that illustrate the relationship between categorical variable and the target variable.

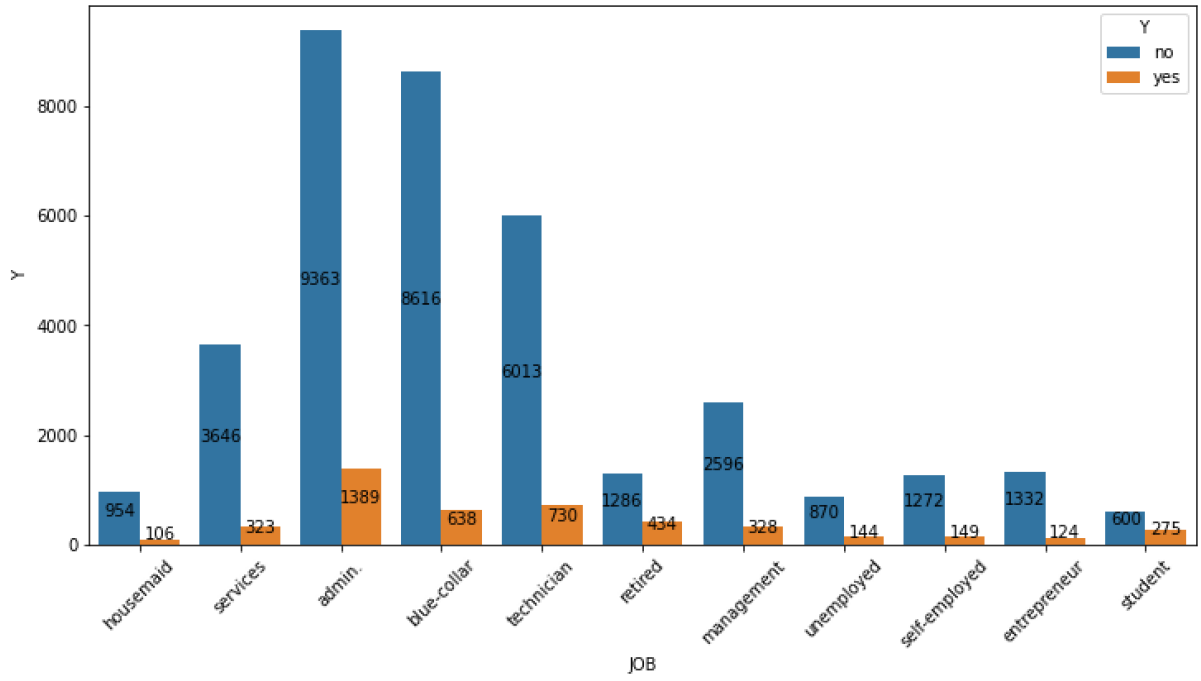

Fig 24: Job versus Y

Fig 24 demonstrates that the highest number of customers with a term deposit are employed in administrative roles.

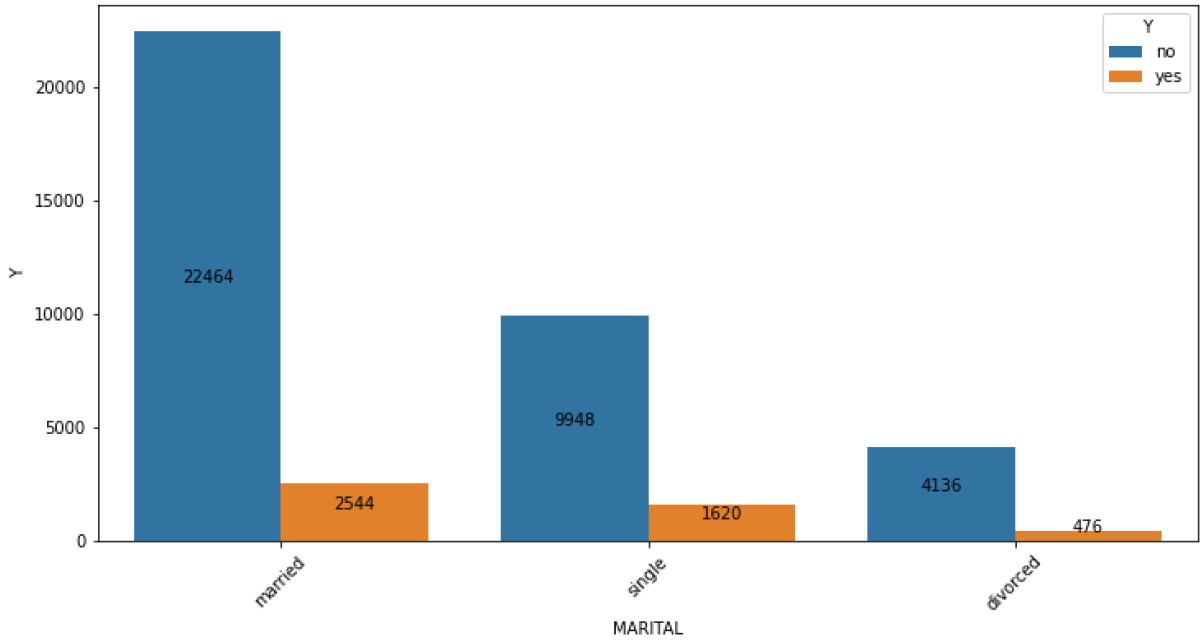

Fig 25: Marital versus Y

Fig 25 illustrates that the highest number of customers with a term deposit are married individuals.

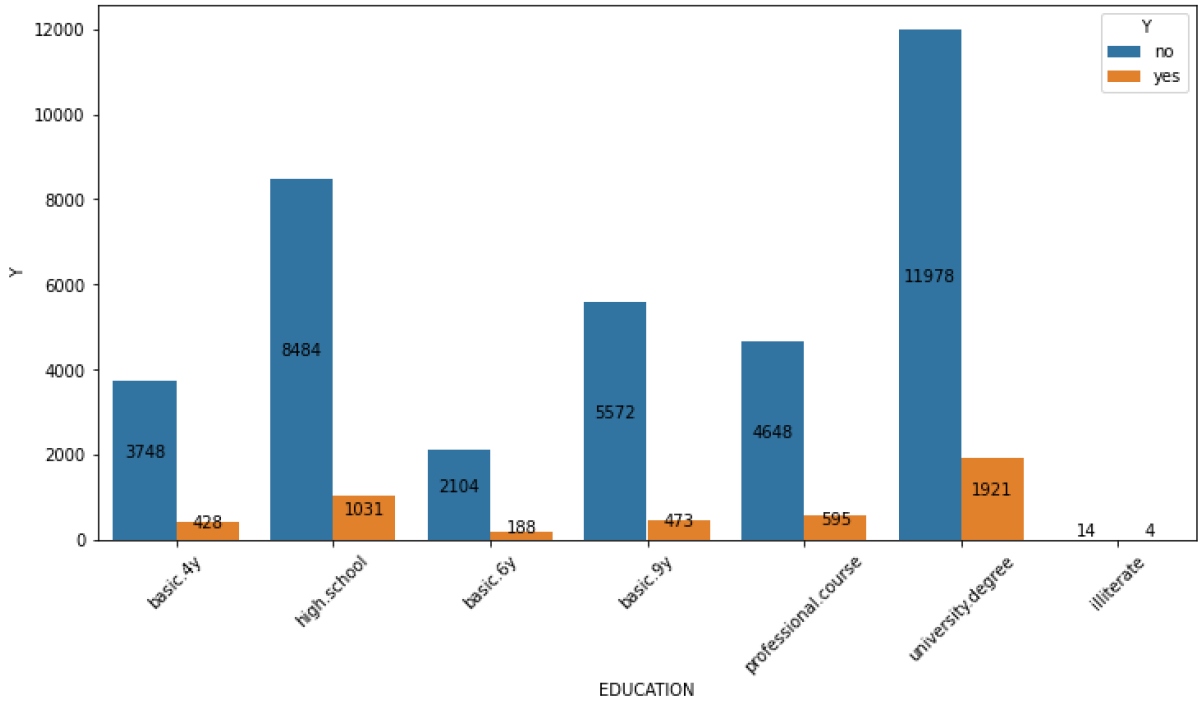

Fig 26: Education versus Y

As shown in Fig 26, the majority of customers with a term deposit are those who hold a university degree.

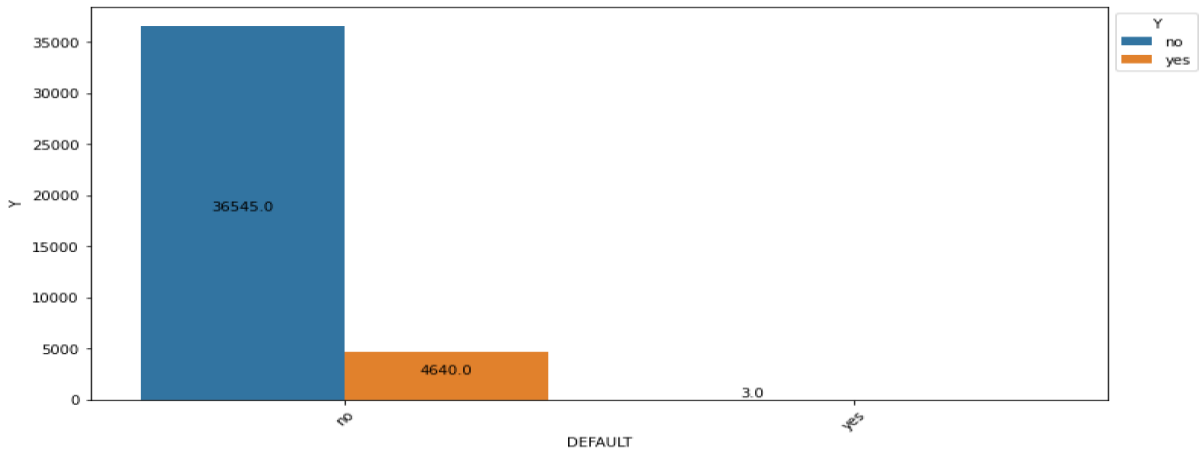

Fig 27: Default versus Y

According to Fig 27, there is a total of 4640 customers who have had a term deposit in the past.

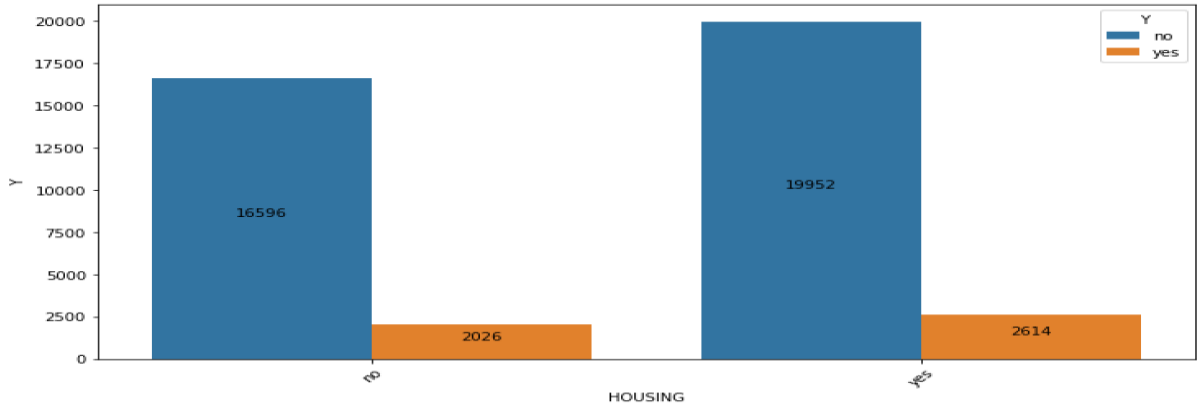

Fig 28: Housing versus Y

Based on the information shown in Fig 28, there is a relatively similar proportion between housing loan holders who have a term deposit and those who do not have a term deposit.

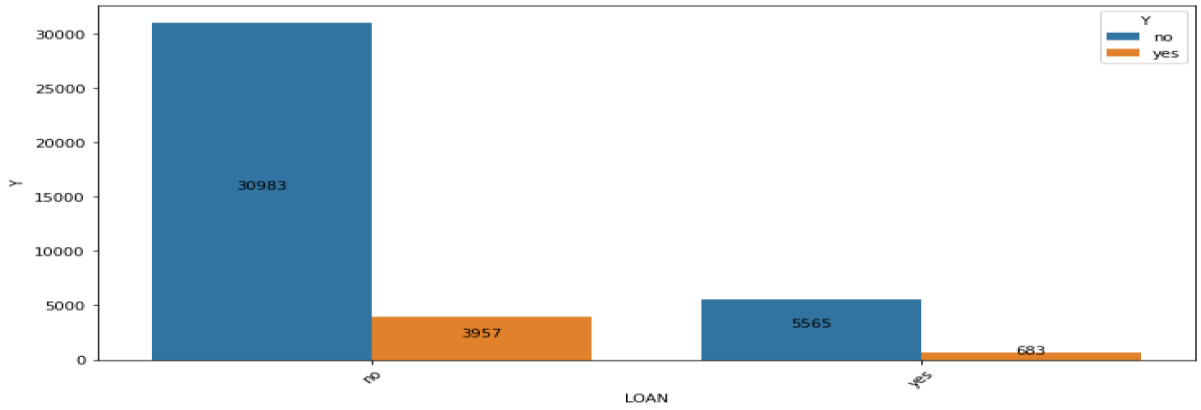

Fig 29: Loan versus Y

Based on the data presented in Fig 29, it can be observed that the number of term deposits for customers without a personal loan is significantly higher compared to those with a personal loan.

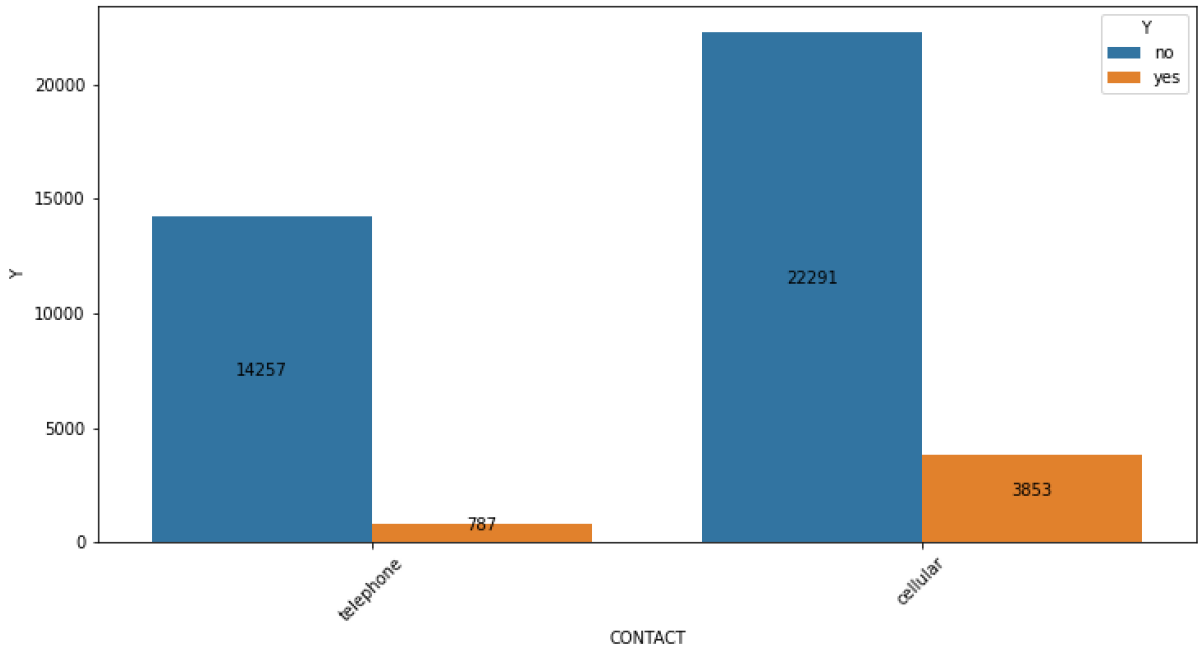

Fig 30: Contact versus Y

Based on the information depicted in Fig 30, it can be observed that the majority of customers who have opted for a term deposit have a cellular phone.

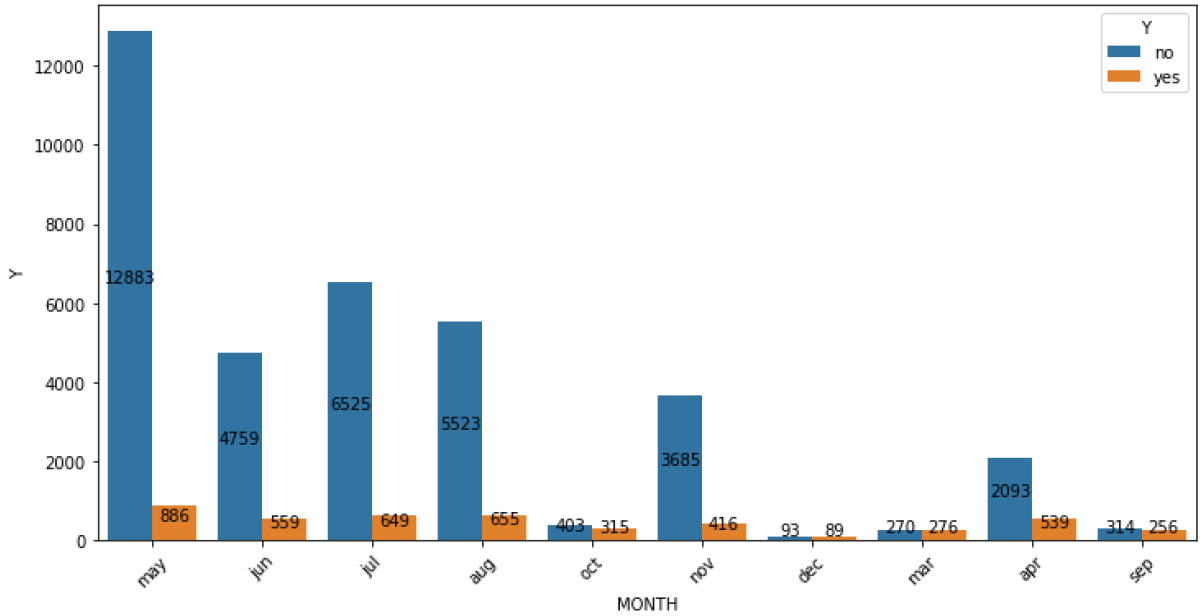

Fig 31: Contact versus Y

Based on the data visualized in Fig 31, it is evident that the majority of customers were contacted during the month of May. Furthermore, this period also witnessed the highest number of conversions to a term deposit.

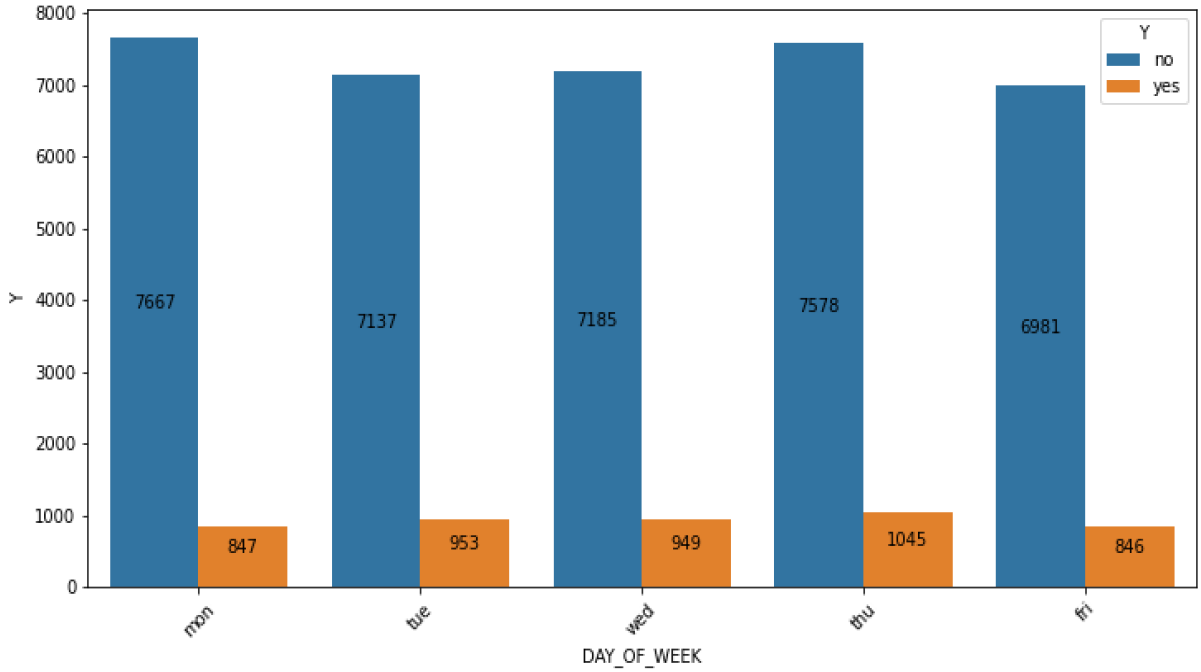

Fig 32: Day_of_Week versus Y

Based on the information presented in Fig 32, it can be observed that the majority of customers were contacted on Monday and Thursday. Additionally, there are only 1,045 customers who have a term deposit.

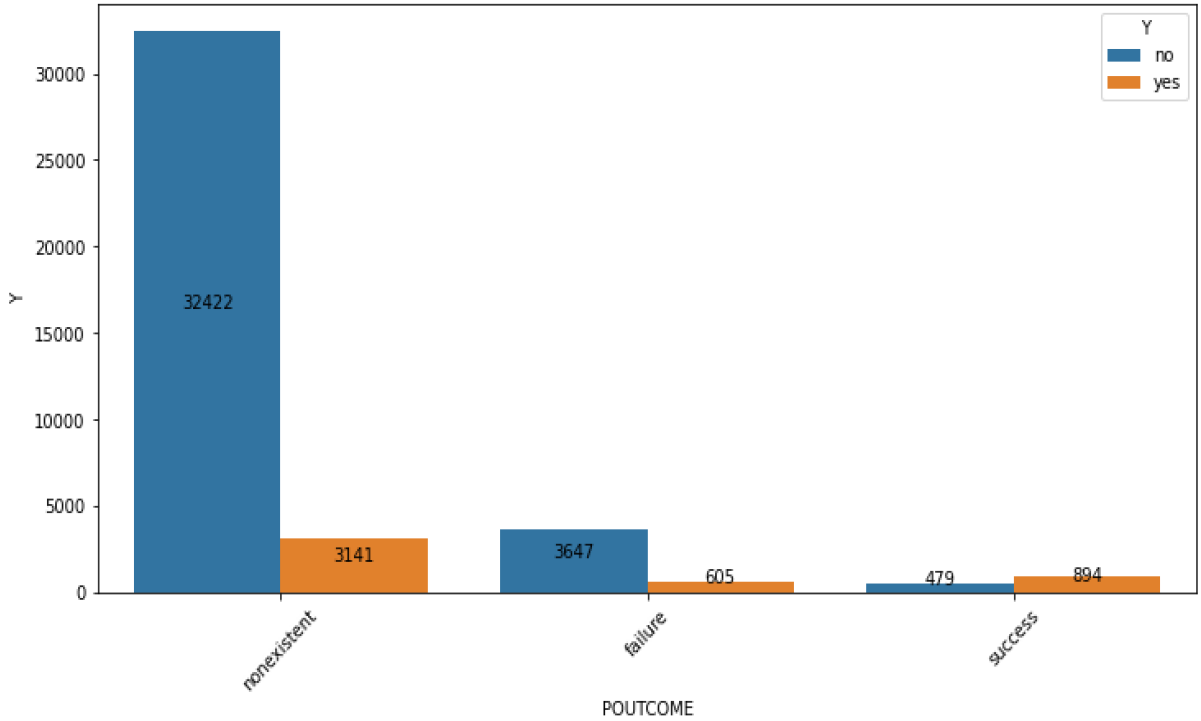

Fig 33: POutcome versus Y

According to the data presented in Fig 33, there are 3,141 non-existent customers but have a term deposit.

Final Recommendation: As analysed with the dataset, the occupation, education, contacted day and month are impacting the customers to opt for term deposit.

1. Customers employed in administrative roles are more inclined to choose a term deposit option.
2. Customers with a university degree are more likely to opt for a term deposit. This indicates that students and retirees are more receptive to being contacted and converting the marketing campaign into a term deposit.
3. The month of May has had a greater impact on converting the marketing campaign into term deposits.
4. Monday and Thursday show slightly higher optimism when it comes to contacting customers during the marketing campaign.