Bank Marketing Campaign

1st May 2023

Preeti Verma	vermapreeti.dataanalyst@gmail.com	Data Science
Thanuja	thanujayadav953@gmail.com	Data Science
Abishek james	abishekjames1998@gmail.com	Data Science

Table of Content

Executive Summary

- A. ABC Bank wants to sell its deposit term product to the customers
- B. By evaluating and implying the EDA and afterwards analyzing it , we tried to understand the data and its pattern on the opt policy.
- C. Different attributes were considered for analysis
- D. The bank data has information of 41189 customers with 21 columns
- E. First we focused on the customers who bought the Policy

Problem Description

- Business Understanding, Data understanding, Exploratory data Analysis, Data Preparation

- Identifying age group, marital status, educational level, and job description of the clients who purchased the policy.

- Is contacting clients before of after the campaign beneficial ?

- What is the time taken for the communication of those clients who purchased the policy?

Data Sets

- The following datasets were used:
    1. bank_additional_full.csv – banking data containing all outcomess
    2. bank_additional.csv – Customer's demographic details

- Time Period of Data: 01/05/2008 to 31/11/2010

Data Cleaning & Combining

Approach

- The data cover a period from May 2008 to November 2010.
- There are 2 datasets, the second dataset is a sample of the first dataset.
- There are 10 integers and 11 categorical variables.
- The missing values in both datasets are presented by an "unknown" string. We changed it to
  NaN.
- There are missing values in six variables namely, job, marital status, education, default, housing, and loan. This will be imputed using various methods.
- There are 12 duplicates in the first dataset and no duplicates in the sample dataset, this will be dropped since they are minimal and will not affect our analysis
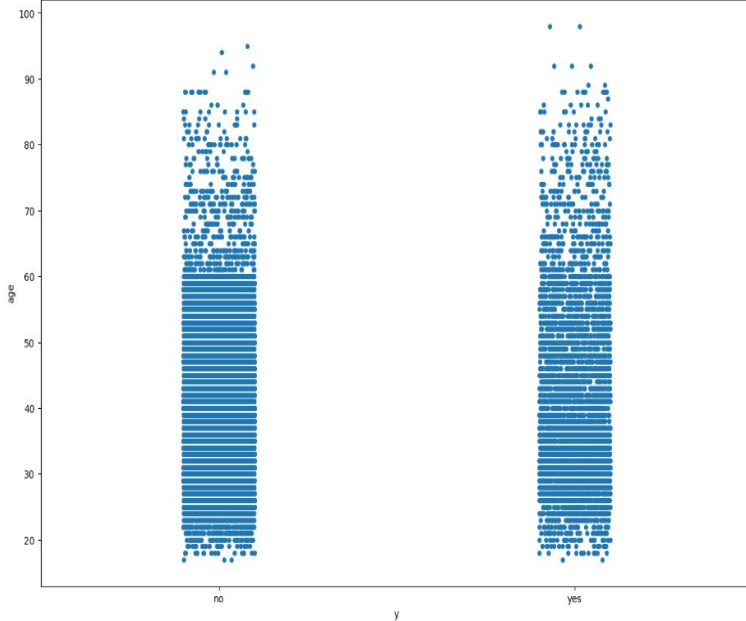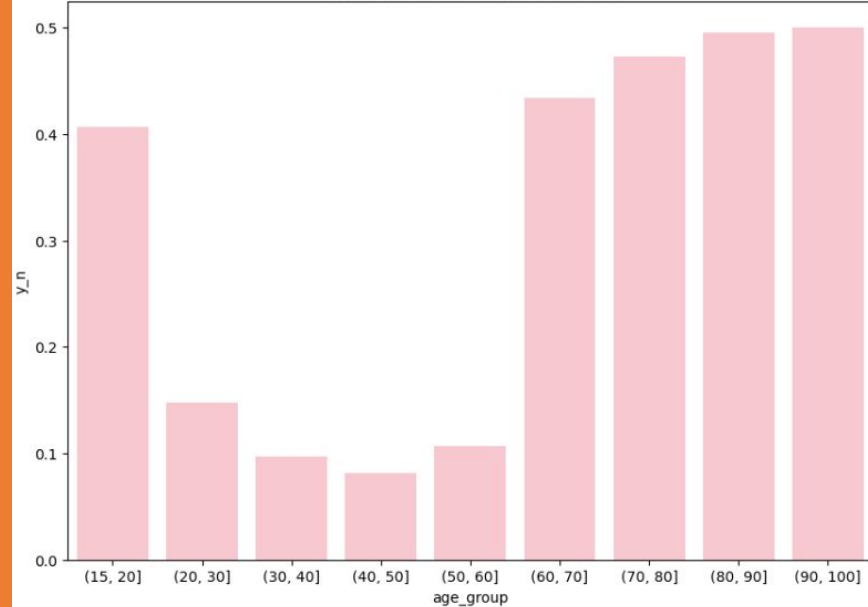
Exploratory Data Analysis

Customer vs Age

- Those who bought the policy are more likely to be between 20-60 years
- Those who are above 70 years have already bought the policy


Age Vs Subscription


Success proportion per Age Group

Distribution with Target Variable

- The highest number of interested customers comes from University Degree holders. Surprisingly, we see that the next highest number comes from customers with just a high school degree

- The highest number of interestes customers comes from those working in admin. The next highest categories are blue-collar and technician.


Distribution of Education with Y
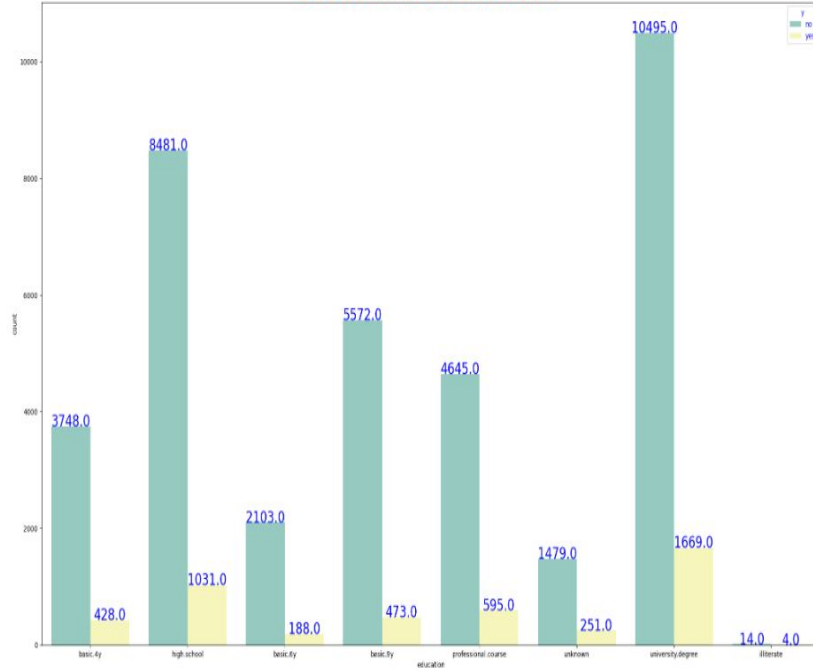
Distribution of Job with Y
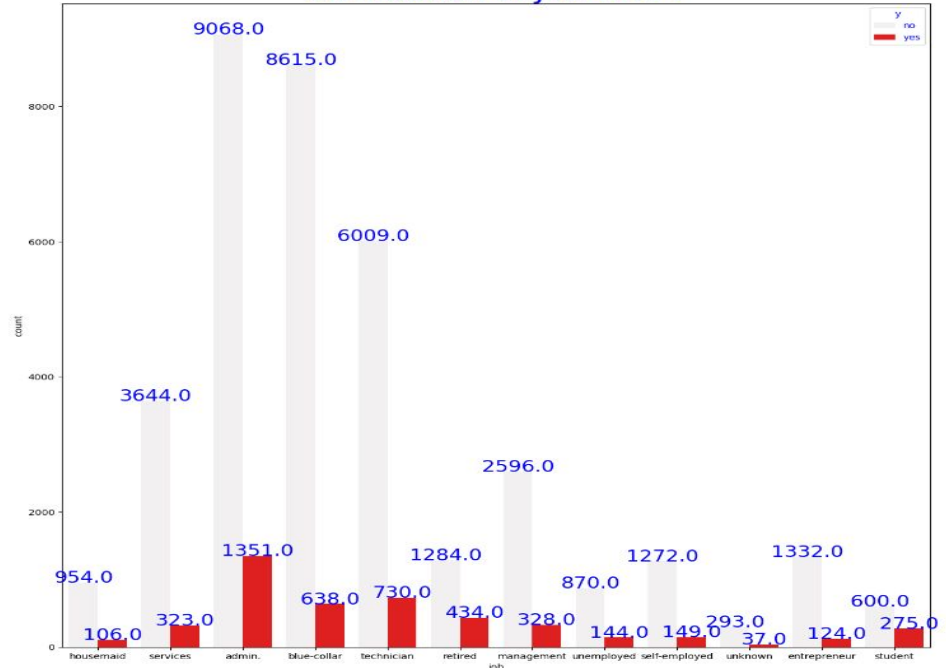
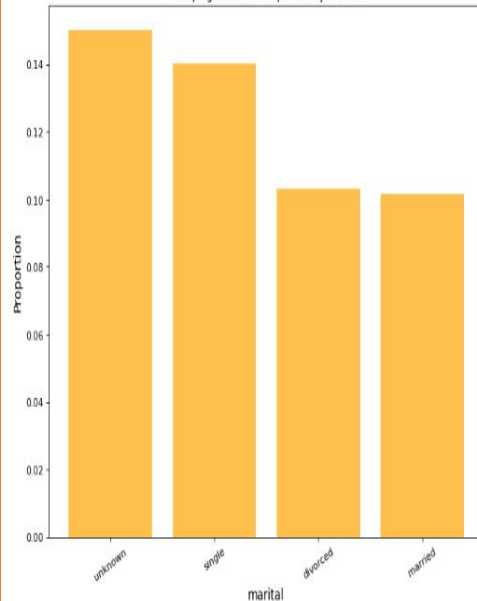# Success Proportion

- Approximately 32% of students and 25% of retirees say 'yes' to open deposit account with the bank
- Singles open more accounts when compared to divorced or married clients
- 65% of the people who agreed for previous campaign agreed for this campaign as well
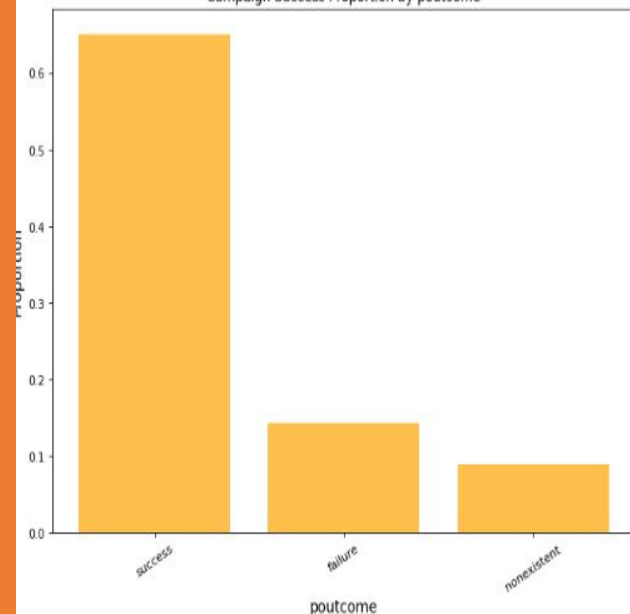

Campaign Success Proportion by job
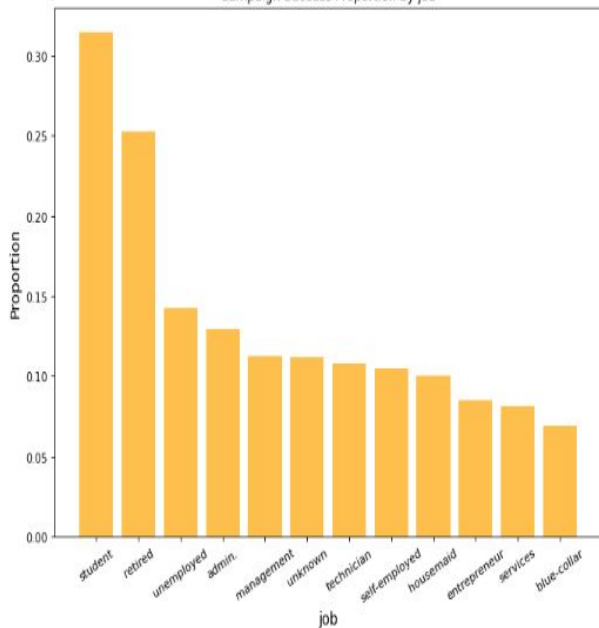
Campaign Success Proportion by marital

Campaign Success Proportion by poutcome

Is there any relationship between variables?
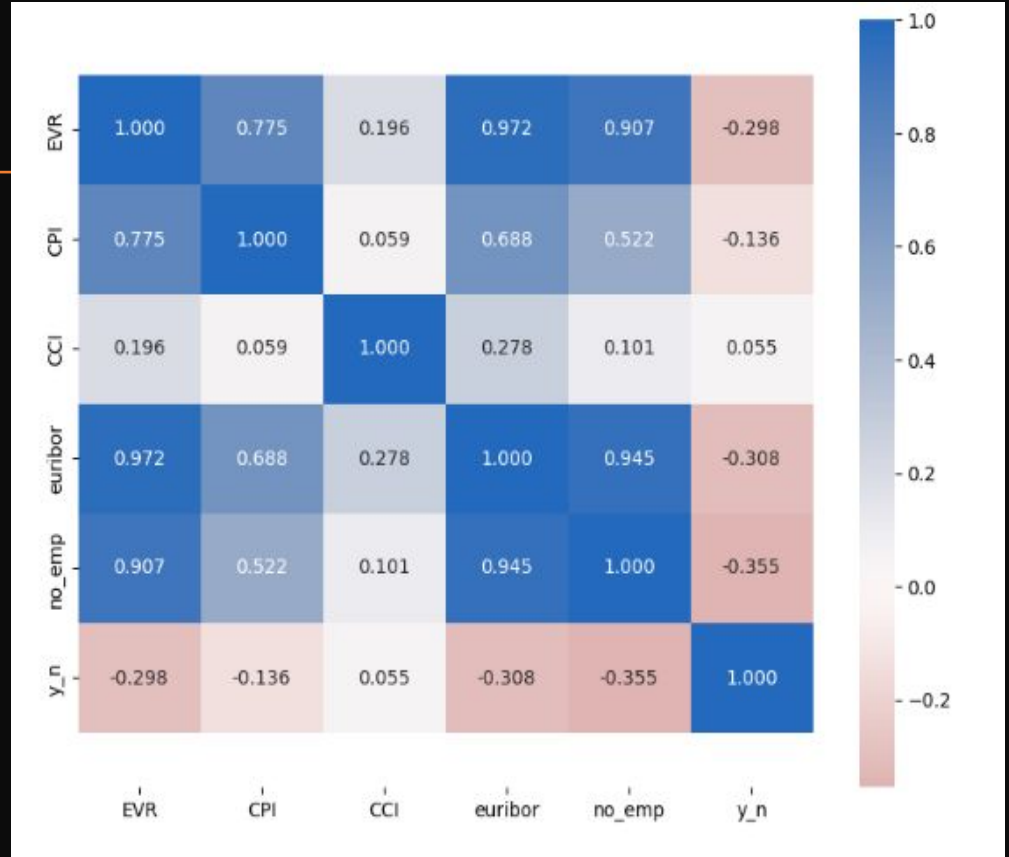
Yes, there is a strong relation between EVR, euribor, no_emp.

Our variable of interest i.e. 'y_n' appears to have some correlation with EVR, euribor and no_emp.. There also seems to be very strong correlation in EVR, euribor and no_emp with each other.

when euribor is less than 2, EVR is always less than -1. Similarly, when euribor is more than 3, EVR is always more than 0.5.
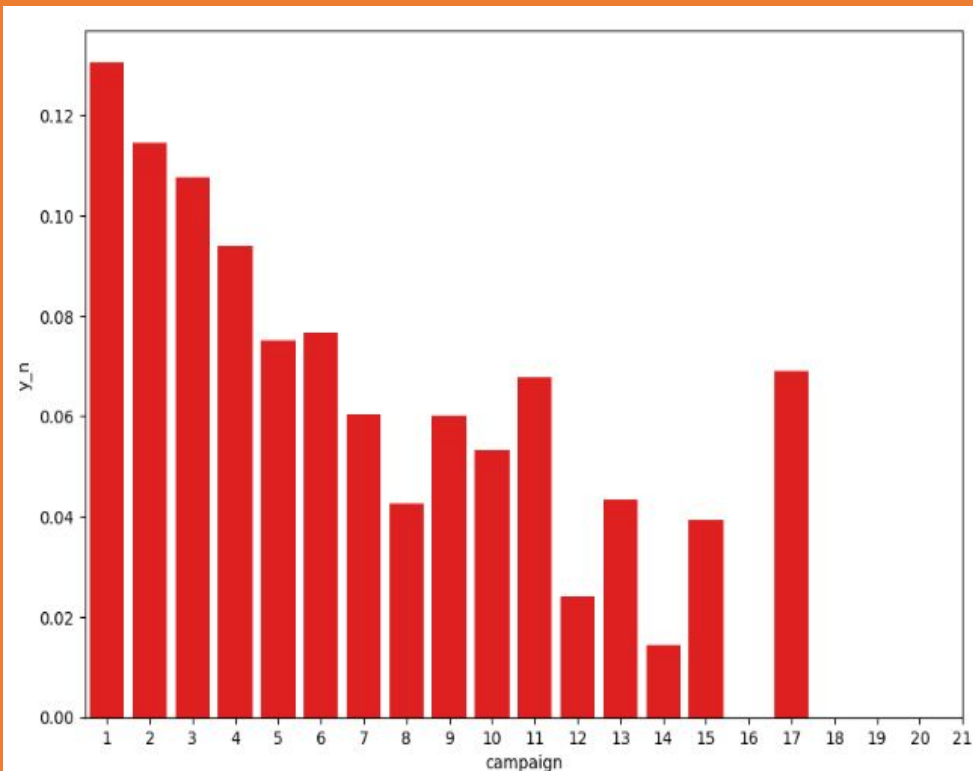
Proportion by campaign calls

Distribution of call duration is heavily concentrated until 400 seconds

Most number of customers say yes to opening the account in very first call. As the number of calls increase, the probability decreases for the customer agreeing to opening the account.
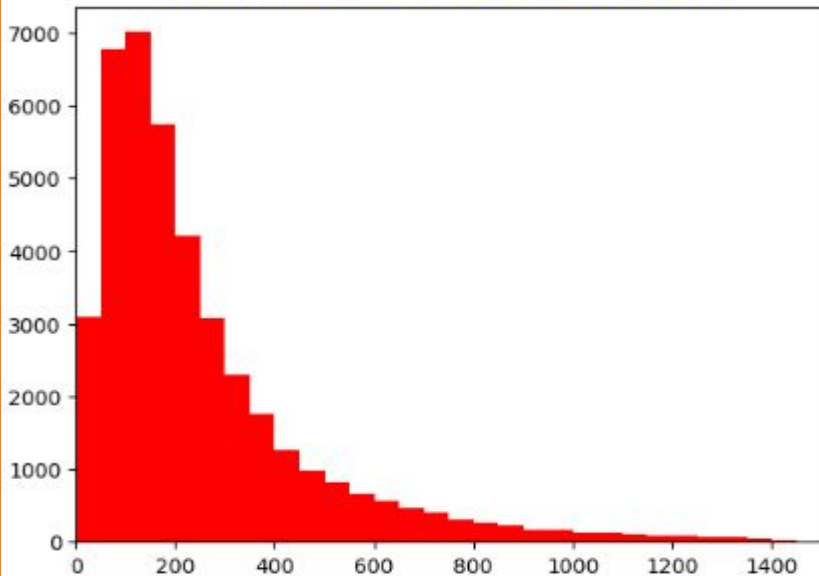
Age with euribor<2

In Pie Chart we see that When EVR < -1 and euribor < 2, 24.46% of people agree to opening term deposit. That's a very large ratio and can certainly help to approach customers

When we compare this bar chart with overall data trend, there are some considerable changes. We see that clients upto 60 years of age, open considerably higher percentage of term deposits when euribor 3 month rate is less than 2. However, clients with more than 60 years of age, follow almost the same trend as in overall data
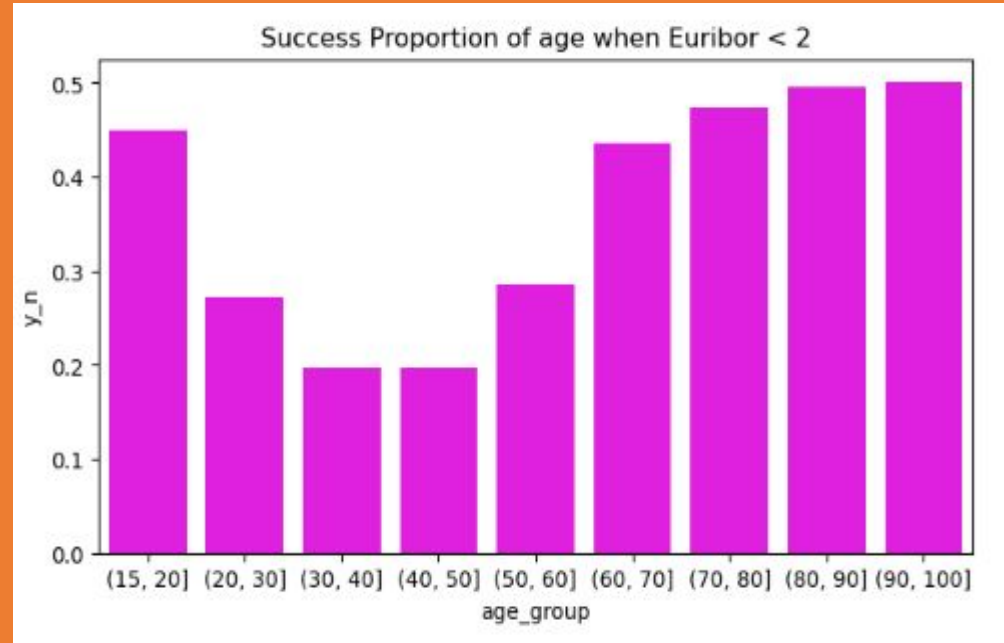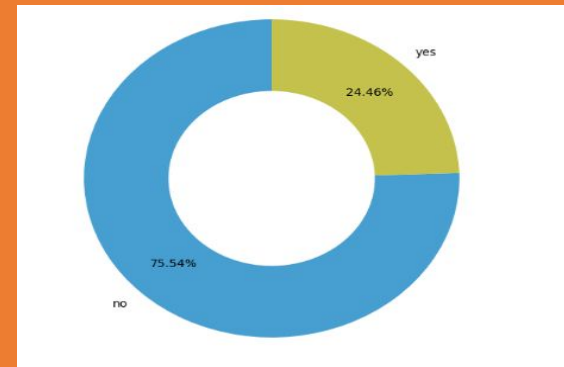
Success ratio with education & job

Generally looking at the heatmap, we see that students, technician, retired and unemployed clients are more likely to open term deposit. Also clients who have studied at university, prodessional course and basic 4y. Clients with unknown educational background are most likely to open term deposit
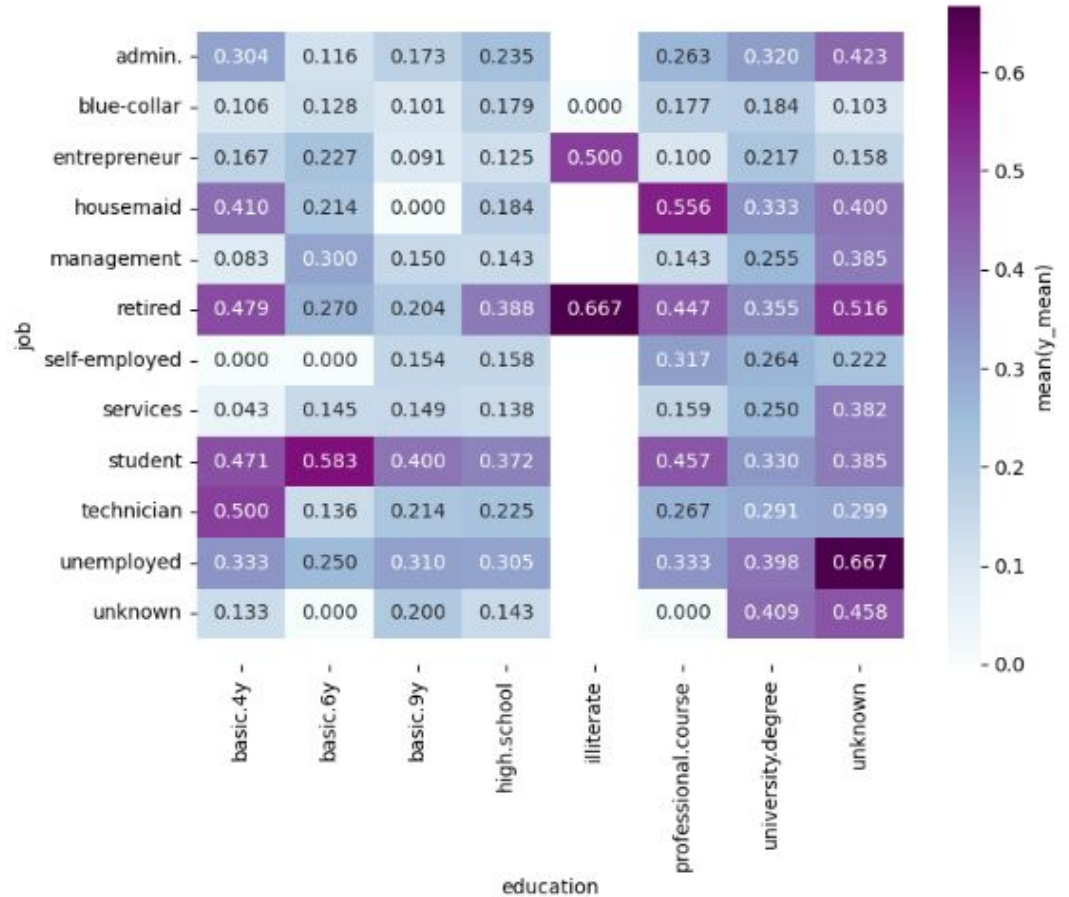
Exploratory Data Analysis (EDA) Summary

- Our variable of interest i.e. 'y_n' appears to have some correlation with EVR, euribor and no_emp.
- Approximately 32% of students and 25% of retirees say 'yes' to open deposit account with the bank.
- Singles open more accounts when compared to divorced or married clients.
- 65% of the people who agreed with the previous campaign agreed with this campaign as well. That is a very significant number and certainly could be used by the sales team to select target customers.
- Although most clients fall in the age range of 25-60, they are the ones who are least likely to open an account. * Success rate for the calls is more for clients above 60 years of age.
- Most customers say yes to open the account on the very first call. As the number of calls increases, the probability decreases for the customer agreeing to open the account. In fact, the probability is almost zero after 26 calls. The probability after 11 calls is less than 3%.

Recommendations

- Those Retired are more likely to buy a policy than others
- Those who are between the ages of 20-50 are more likely those who bought the policy and also there are those above 70 also have opted for the policy.
- Those Married are more likely to buy between the age of 60 – 80.
- When the Euribor 3-month rate is less than 2, clients under the age of 60 are twice more likely to open a term deposit.
- We should notice that bank called only half the number of customers when Euribor was less than 2 as compared to otherwise. So, we should target to call more customers during this period for a better success rate.
- In overall data, students and retirees are much more likely to open term deposits. 45% of clients above 60 years of age agreed to open term deposits. But we have only 2.5% of customers in this range.
- It is recommended to include more students, retirees, and clients older than 60 years of age to get a better success rate in the overall dataset.
- Clients for whom previous campaign outcome was successful, agreed to 65% of the calls. This is a significant figure and can be used to target customers in further campaigns

Recommendation Models

The data is related with direct marketing campaigns of a Portuguese banking institution The marketing campaigns were based on phone calls Often, more than one contact to the same client was required, in order to access the product

The following are the recommended models that may be applied to the data

- **Decision tree classifier**
- **Random Forest classifier**
- **Support Vector classifier**
- **Logistic Regression**

Machine Learning Models

Logistic Regression Models

```
Train Result:

Accuracy score: 0.5118

Classification Report:
              precision    recall  f1-score   support

           0       0.51      0.61      0.56     31860
           1       0.51      0.41      0.46     31860

    accuracy                           0.51     63720
   macro avg       0.51      0.51      0.51     63720
weighted avg       0.51      0.51      0.51     63720


Confusion Matrix:
 [[19408 12452]
 [18657 13203]]

Average Accuracy:        0.5118
```

Random Forest Models

```
print_score(classification_report_random_forest,x_train_right,y_train_right, X_test,y_test, train= False)
```

```
Test Result:

Accuracy score: 0.7325

Classification Report:
              precision    recall  f1-score   support

           0       0.97      0.74      0.84      7955
           1       0.06      0.42      0.10       283

    accuracy                           0.73      8238
   macro avg       0.51      0.58      0.47      8238
weighted avg       0.94      0.73      0.82      8238


Confusion Matrix:
 [[5914 2041]
 [ 163  120]]
```

Decision Tree Classifier

```
Accuracy score: 0.9656

Classification Report:
              precision    recall  f1-score   support

           0       0.97      1.00      0.98      7955
           1       0.50      0.00      0.01       283

    accuracy                           0.97      8238
   macro avg       0.73      0.50      0.49      8238
weighted avg       0.95      0.97      0.95      8238


Confusion Matrix:
 [[7954      1]
```

Recommended Model

Decision Tree Classifier is the best model for the bank campaign marketing

Thank You