Data Glacier

Your Deep Learning Partner

Bank Marketing (Campaign)

Data Girl

Date: 19 Nov 2022

LISUM 14

Fatimah Asiri

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Models

Recommendations

Executive Summary

- For this project, ABC Bank wants to sell its term deposit product to customers and before launching the product they want to develop a model which helps in understanding whether a particular customer will buy its product or not (based on the customer's past interaction with the bank or other Financial institutions).

Problem Statement

- Bank wants to use the ML model to shortlist customer whose chances of buying the product is more so that their marketing channels marketing SMS/email marketing, etc. can focus only on those customers whose chances of buying the product is more.
- -This will save resources and time (which is directly involved in the cost (of resource billing).
- Develop a model with Duration and without duration features and report the performance of the model.
- Based on PCA for the features, I choose just 27 columns from the dataset (job, contract, loan, default, age, housing, balance, marital, and education).

Approach

- Business Understanding.
- Data understanding.
- Exploratory Data Analysis.
- Data Preparation.
- Model Building (Logistic Regression, Random Forest, Decision Tree)
- Model Selection.
- Performance reporting.
- Deploy the model using a flask.
- Converting ML metrics into Business metrics and explaining results to the business.

Data Set Information :

- The data is related to direct marketing campaigns of a Portuguese banking institution.
- The marketing campaigns were based on phone calls. Often, more than one contact with the same client was required, in order to access if the product (bank term deposit) would be (‘yes’) or not (‘no’) subscribed.

- The classification goal is to predict if the client will subscribe (yes/no) to a term deposit (variable y).

Attribute Information:

Input variables:

bank client data:

1 - age (numeric)
2 – job: type of job (categorical: ‘admin.’, ‘blue-collar’, ‘entrepreneur’, ‘housemaid’, ’management’, ’retired’, ’self-employed’, 'services’, 'student’, 'technician’, 'unemployed’, 'unknown')
3 - marital: marital status (categorical: 'divorced’, 'married’, 'single’, 'unknown'; note: 'divorced' means divorced or widowed)
4 - education (categorical: 'basic.4y’, 'basic.6y’, 'basic.9y’, 'high. School’, ‘illiterate’, ‘professional. Course’, ‘university. Degree’, ‘unknown’)
5 – default: has credit in default? (categorical: ‘no’, ‘yes’, ‘unknown’)
6 – housing: has a housing loan? (categorical: ‘no’, ‘yes’, ‘unknown’)
7 – loan: has a personal loan? (categorical: 'no’, 'yes’, 'unknown')

Attribute Information: (con…)

8 - contact: contact communication type (categorical: 'cellular’, 'telephone')related to the last contact of the current campaign

9 - month: last contact month of the year (categorical: ‘Jan, ‘Feb, 'mar', ..., ‘Nov, ‘Dec')

10 - duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

11 - campaign: number of contacts performed during this campaign and for this client (numeric, includes the last contact)

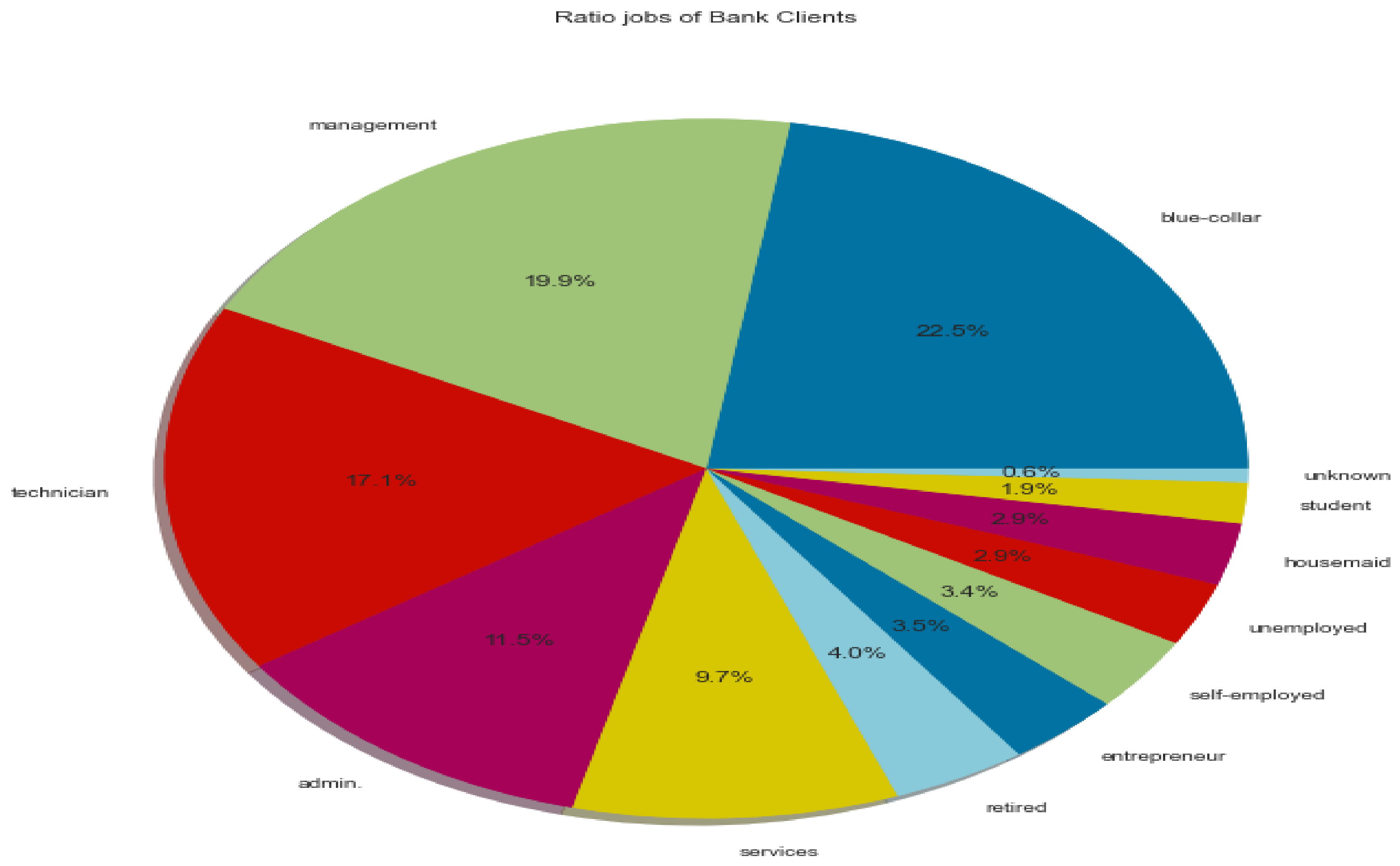
12 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means the client was not previously contacted)

13 - previous: number of contacts performed before this campaign and for this client (numeric)

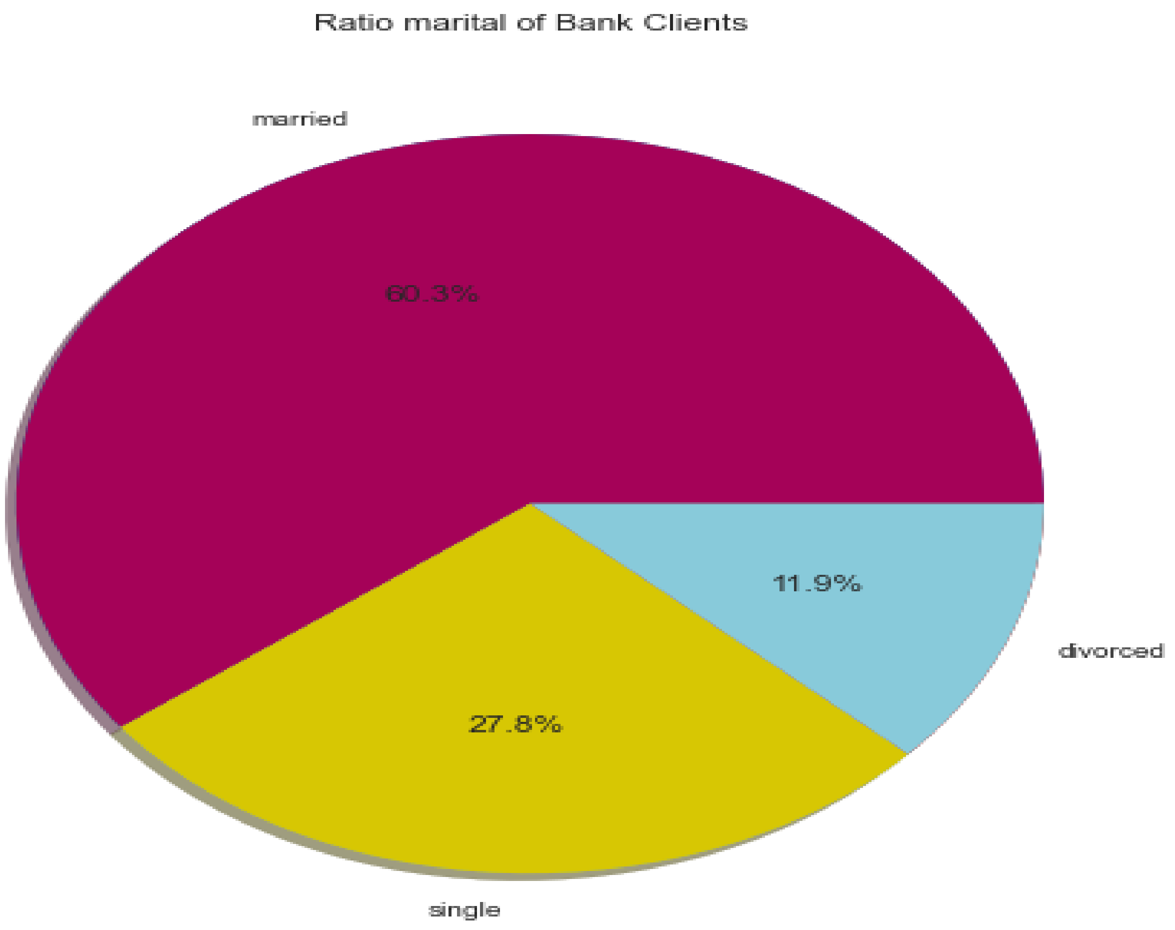
14 - poutcome: outcome of the previous marketing campaign (categorical: 'failure’, Unknown’, ‘success’)

15 – y: The classification goal is to predict if the client will subscribe (yes/no) to a term deposit (variable y).

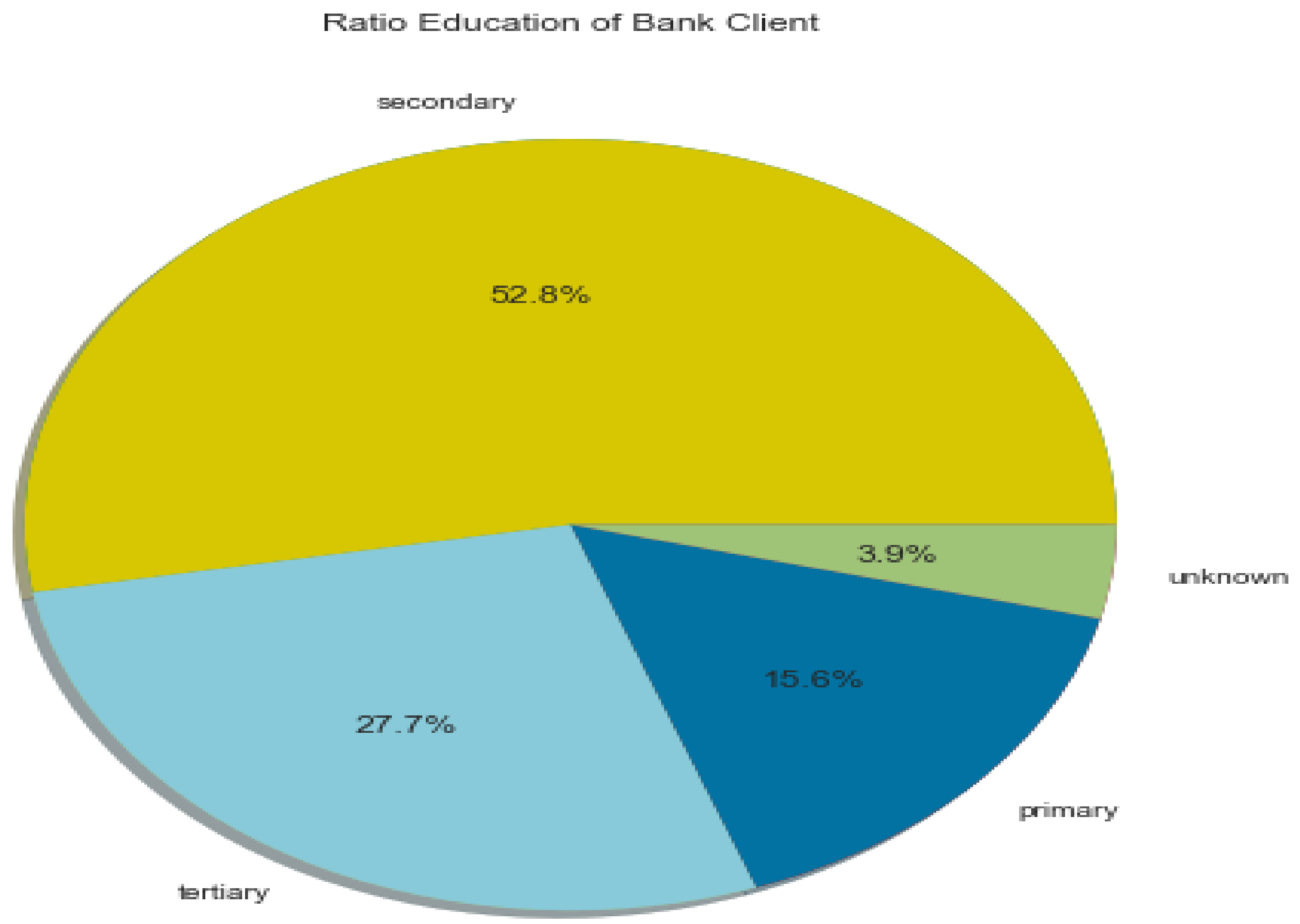
EDA - What's the jobs for bank clients?



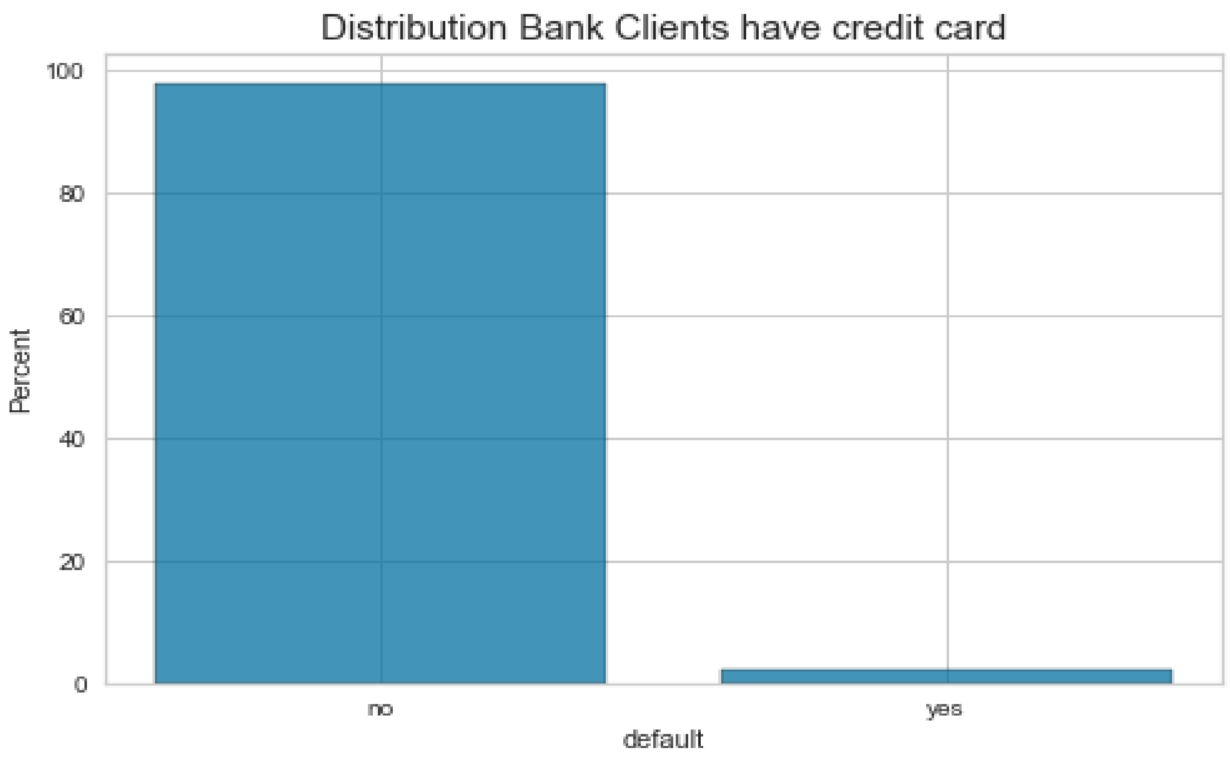
EDA - What is the marital status of bank clients?



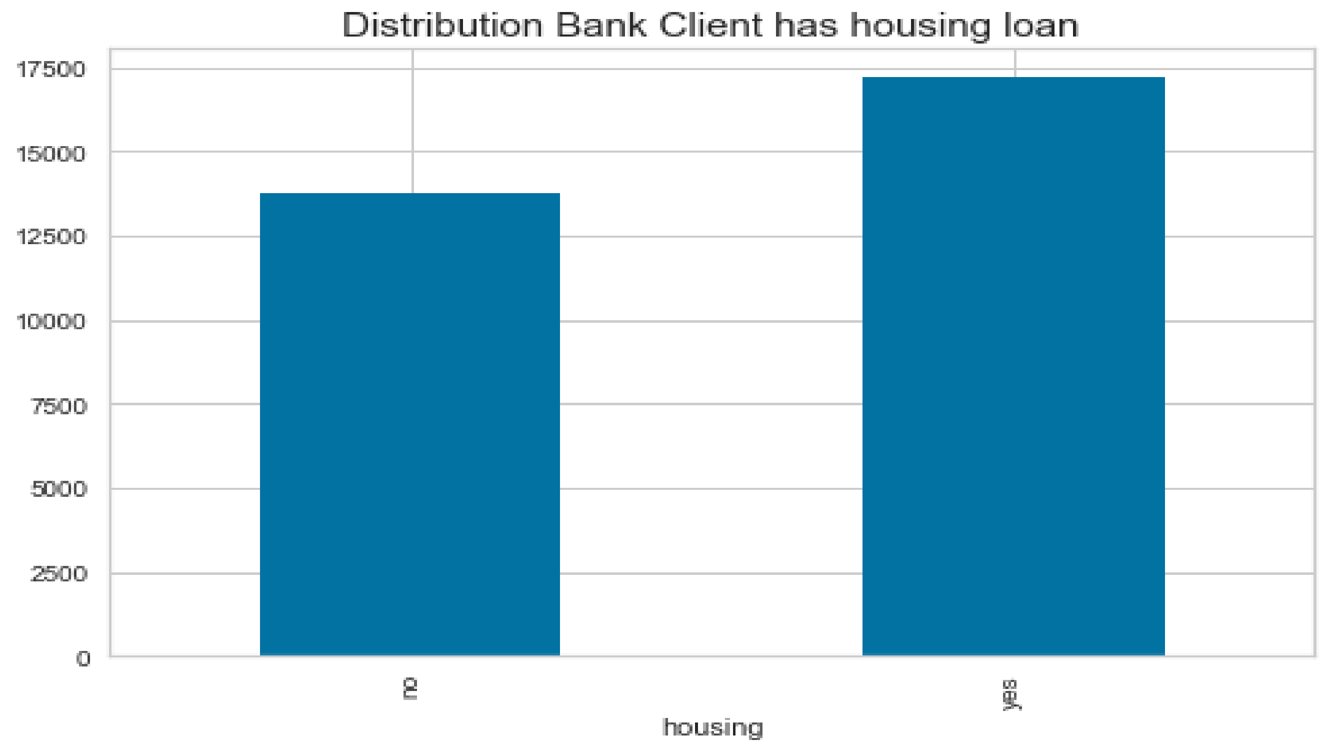
EDA - What is the education status of bank clients?



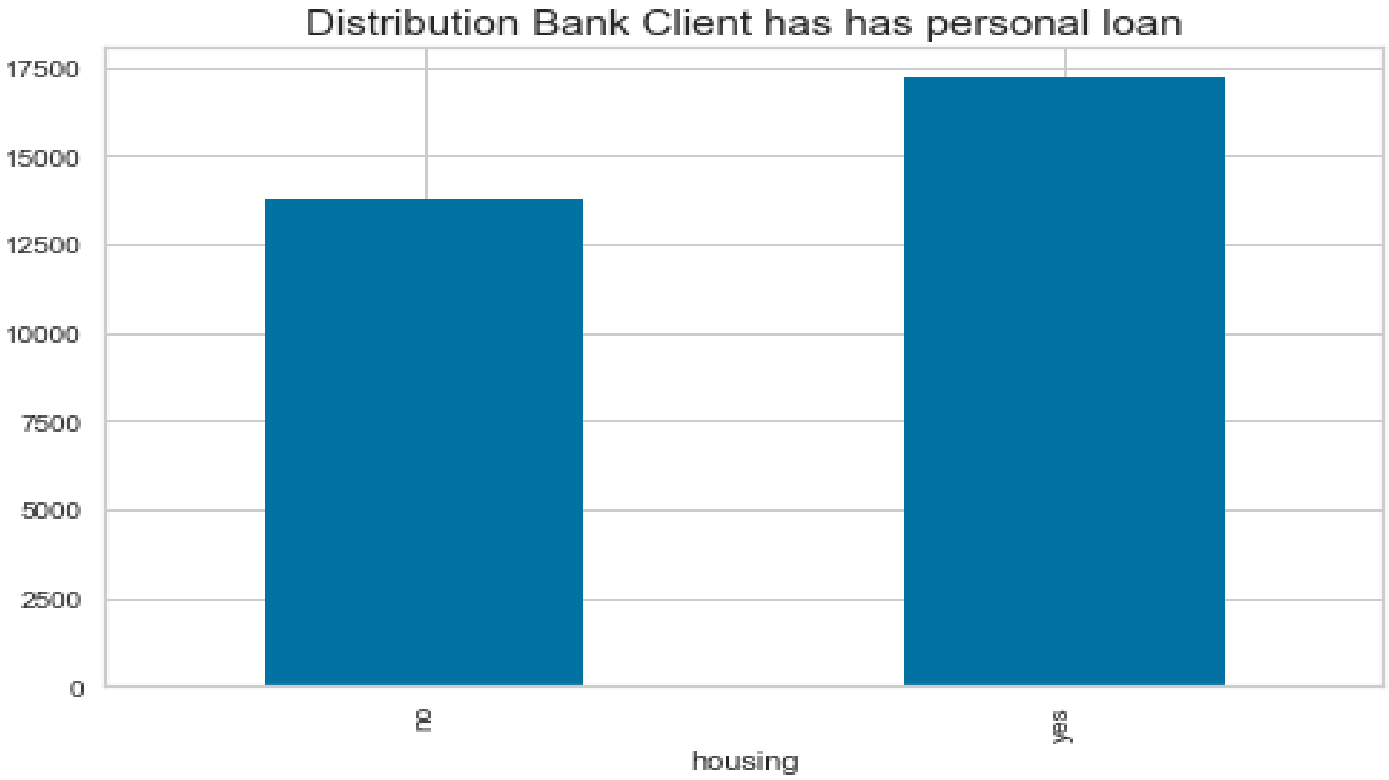
EDA - Are bank clients have credit cards?



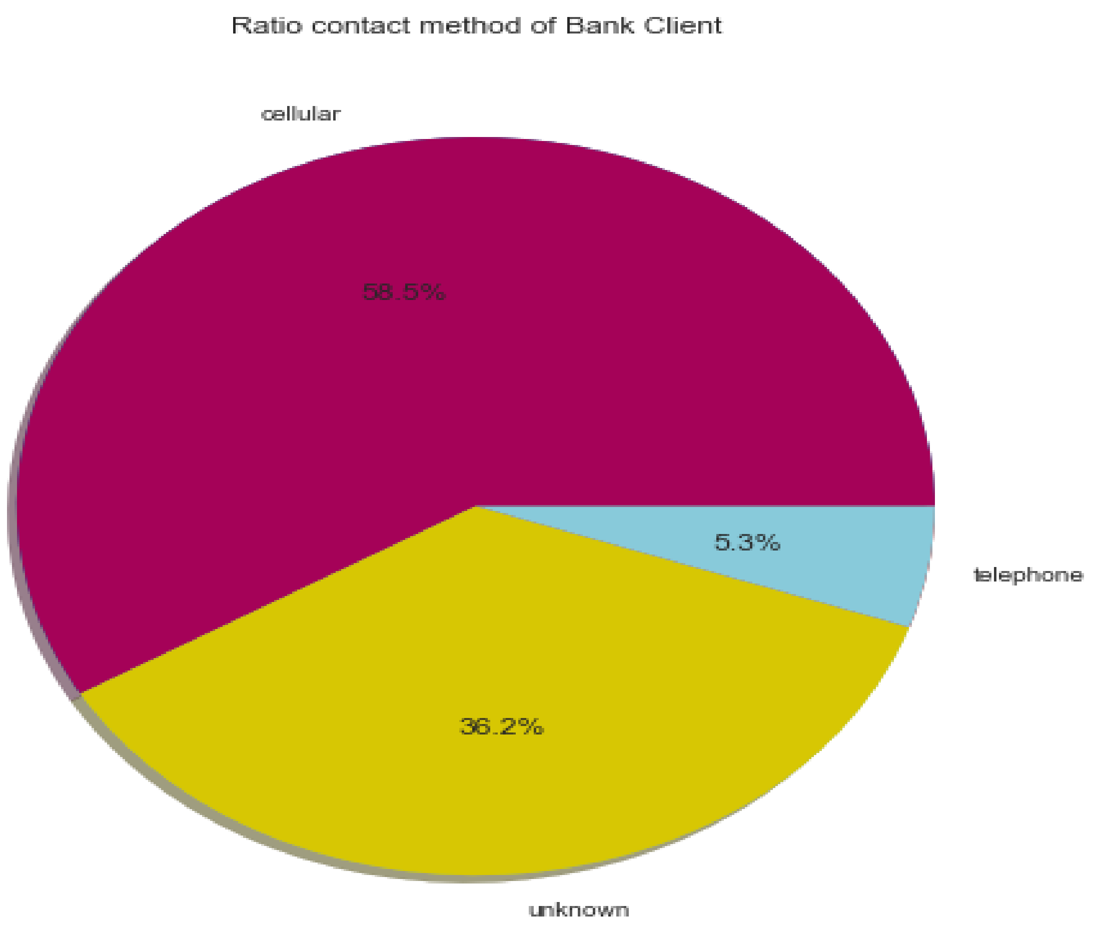
EDA - Are bank clients have housing loans?



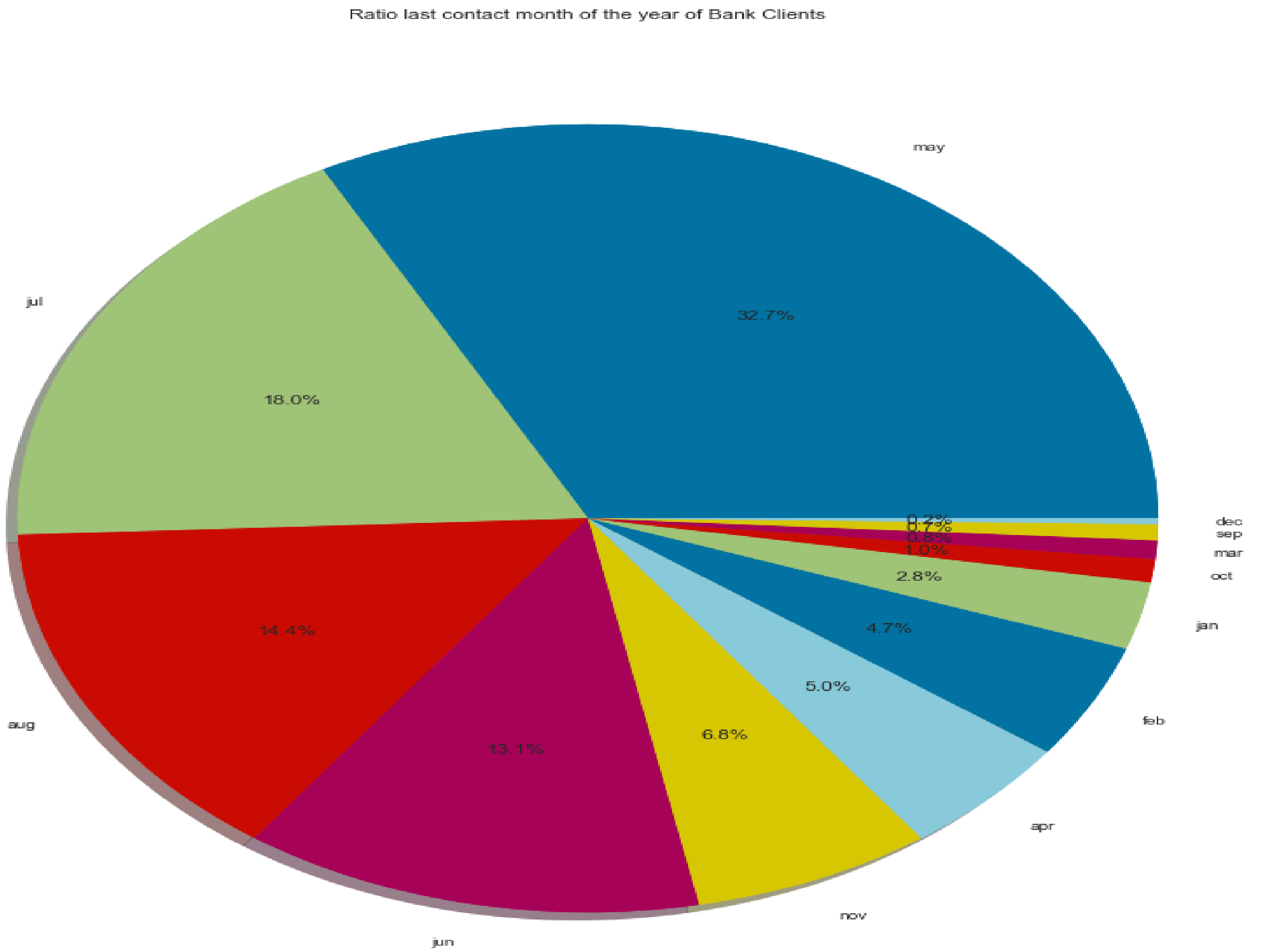
EDA - Are bank clients have personal loans?



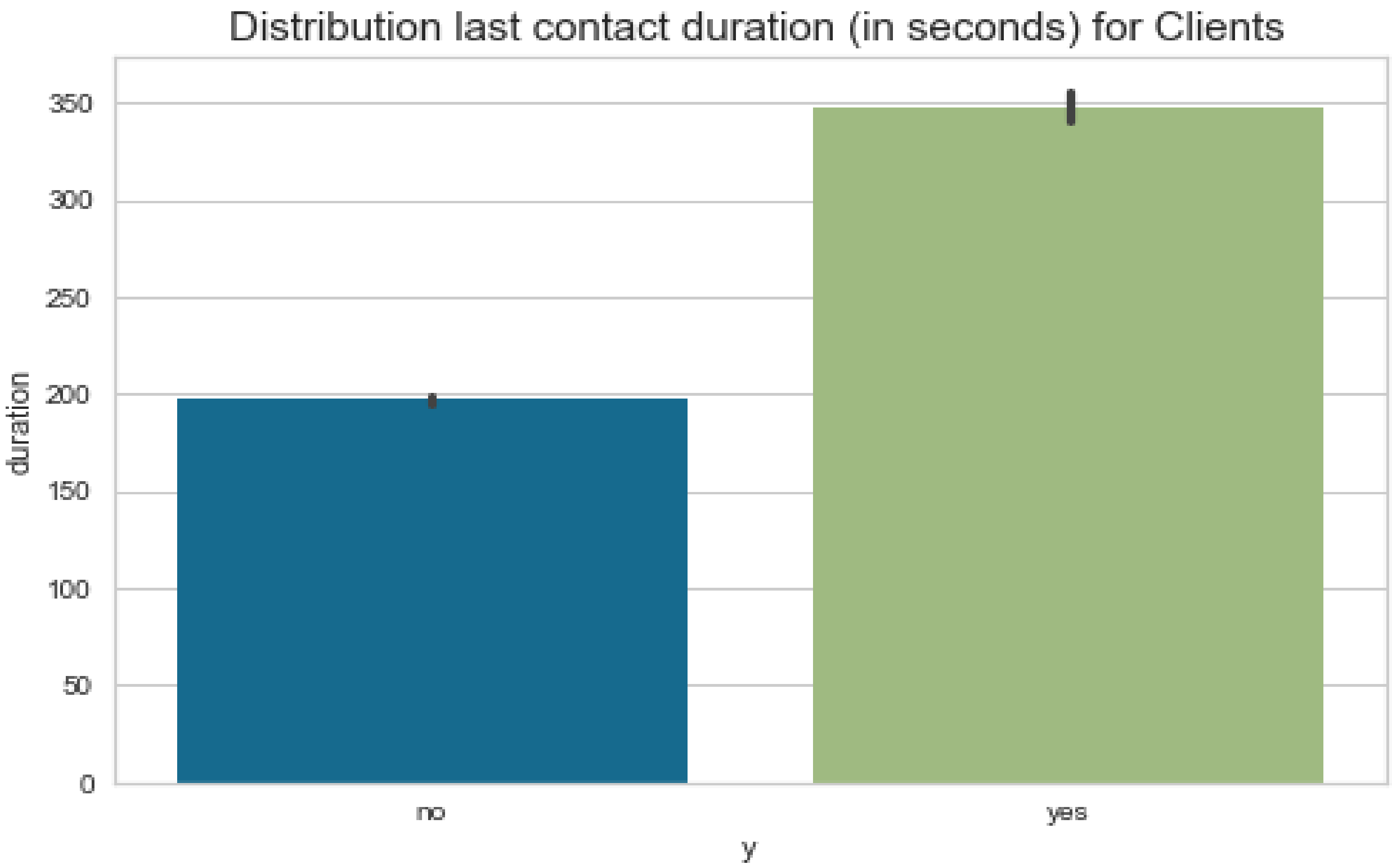
EDA - what's the method contact with bank clients?



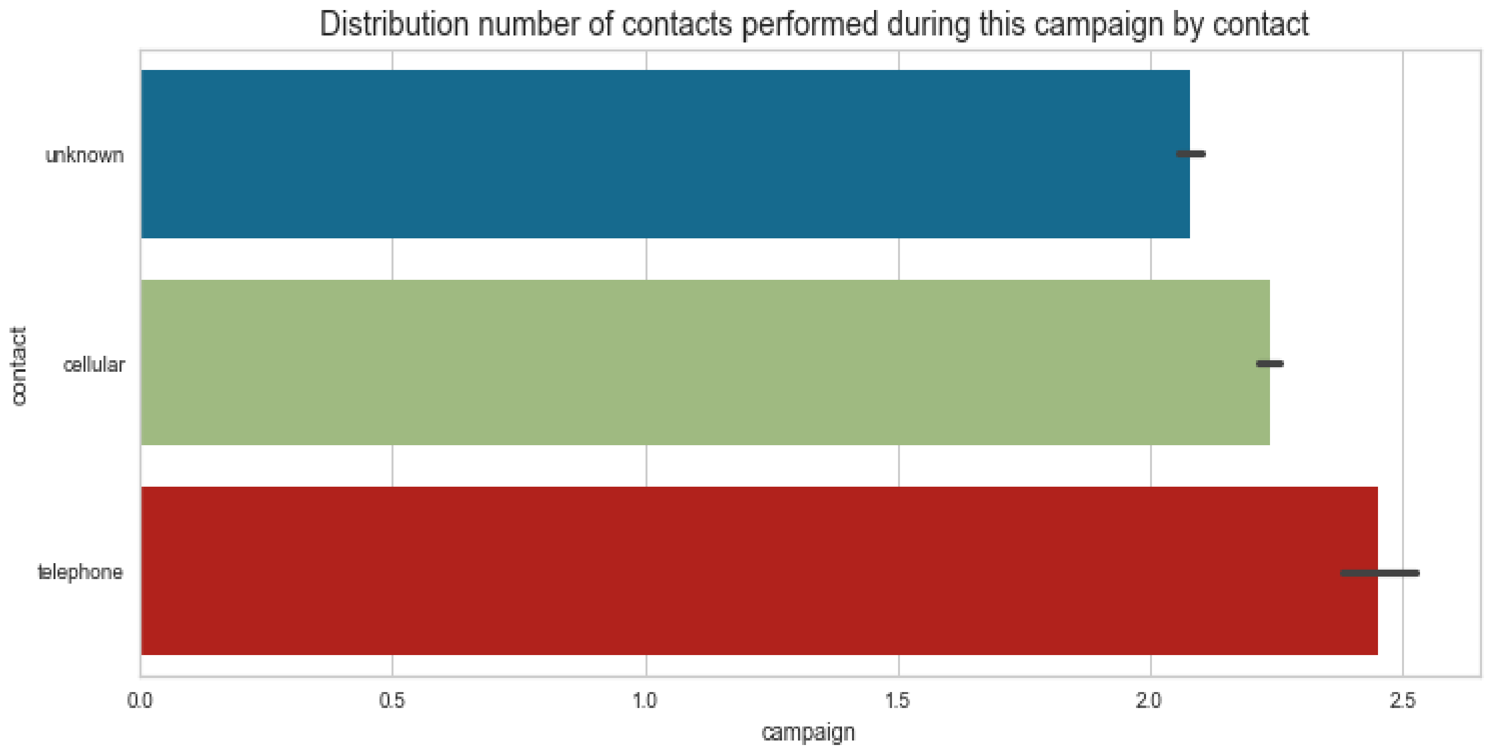
EDA - what's the last contact month of the year of bank clients?



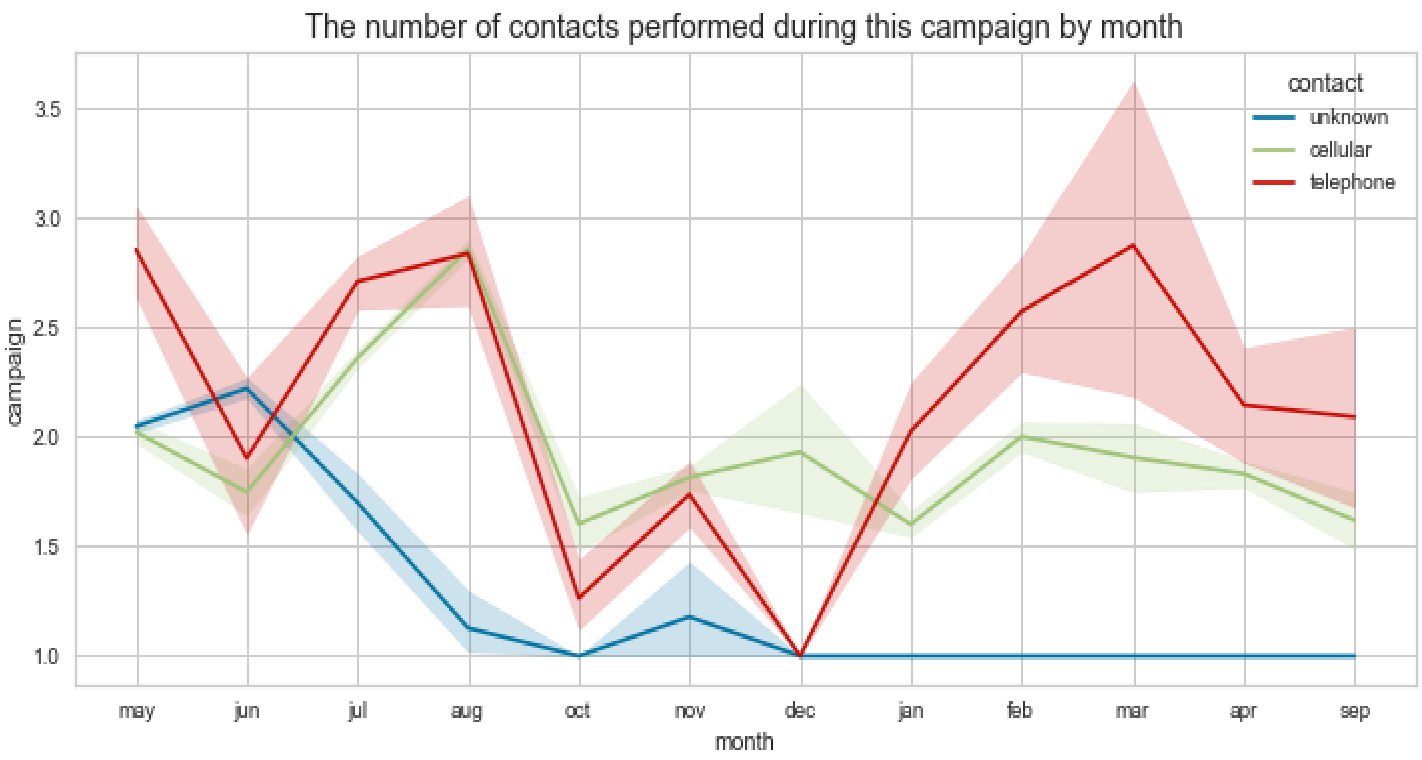
EDA - Do clients subscribe to a term deposit based on the last contact duration (in seconds)?



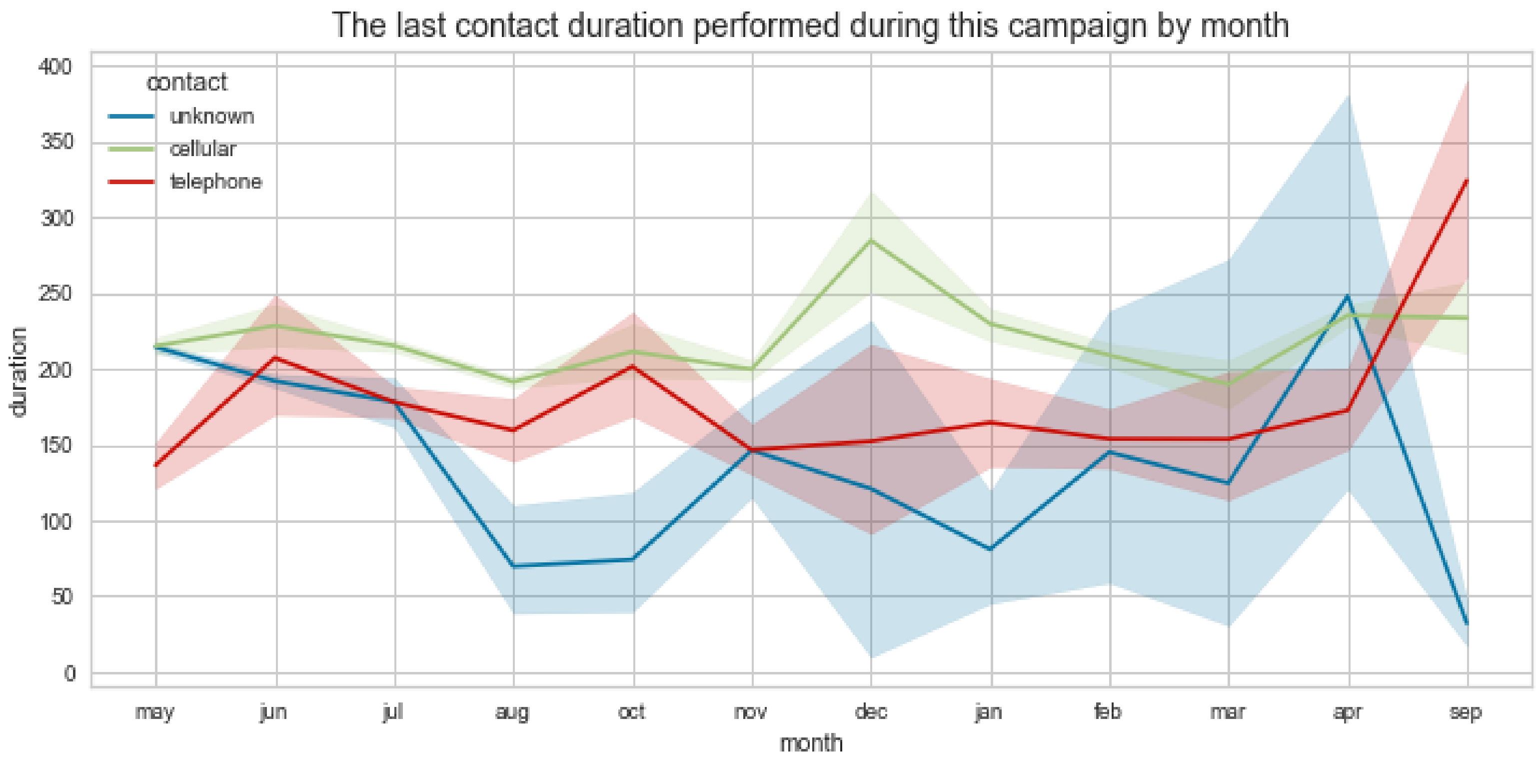
EDA - what's the method of contact performed during this campaign by contacting?



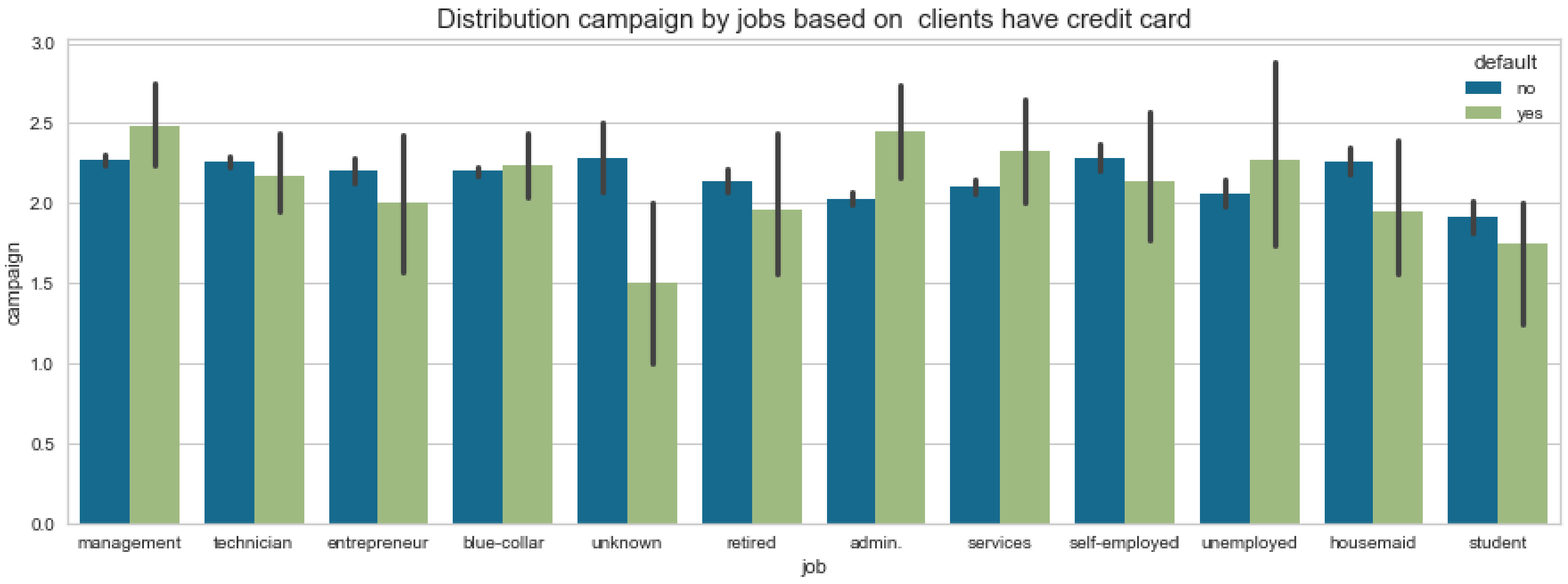
EDA - what's the highest and lowest month number of a contact in the campaign?



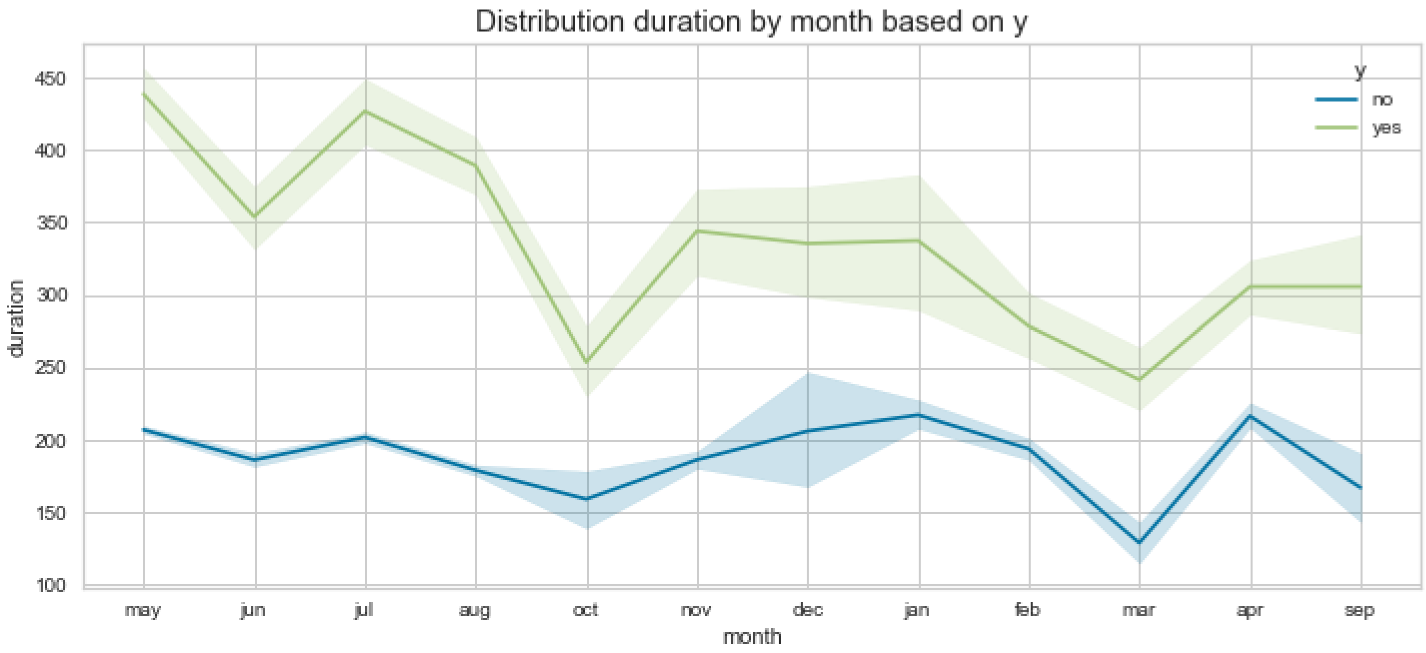
EDA - what's the last contact duration performed during this campaign by month?



EDA - Distribution campaign by jobs based on clients have credit card



EDA - Distribution duration by month based on y



EDA Summary

- The top 5 jobs for the client's bank: blue-collar (22.5%), management (19.9%), technician (17.1%), admin (11.5%), and services (9.7%)

- The lowest 5 jobs for the client’s bank: self-employed (3.4%), unemployed (2.9%), housemaid (2.9%), student (1.9%), and unknown (0.6%).

- The ratio marital of bank clients divide to 3 types married(60.3%), single(27.8%), and divorced(11.9%).

- The ratio education of bank clients divide to 4 types Secondary(52.8%), tertiary(27.7%), married(15.6%), and unknown(3.9%).

EDA Summary

- The ratio of bank clients who have a credit card (98% no) and (2% yes).

- The number of clients that have housing loans (55.51% yes) and (44.49% no).

- The number of clients that have personal loans (17.50 % yes) and (82.50% no).

- The ratio contact method of Bank Client 3 types cellular(58.5%), unknown(36.2%), and telephone(15.6%).

EDA Summary

- The ratio last contact month of the year of Bank Clients.

- The top 5 months active Ratio last contact month of the year of Bank Clients: may (32.7%), Jul (18.0%), Aug (14.4%), Jun (13.1%), and Nov (6.8%)

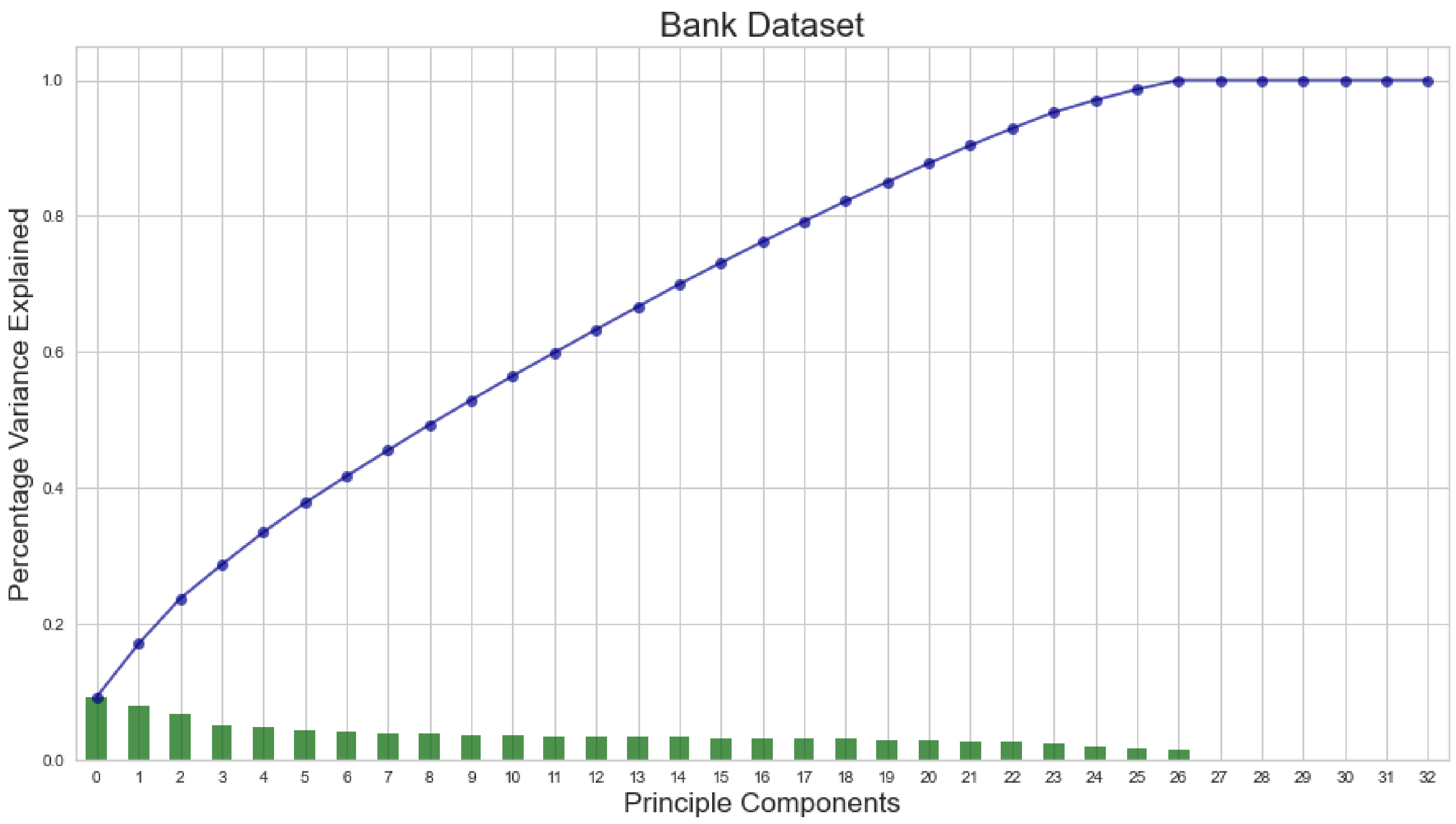
- The lowest 5 months active Ratio last contact month of the year of Bank Clients: Jan (2.8%), Oct (1.0%), Mar (0.8%), Sep (0.7%), and Dec (0.2%).

- The number of Clients who accept the campaign in the last contact duration (in seconds) is higher than the rejected.

EDA Summary

- The number of contacts performed during this campaign was sorted from top to down telephone, cellular and unknown.
- The number of contacts performed during this campaign by month sorted from top to down telephone, cellular and unknown.
- The clients that have credit cards more than they not based on jobs.
- The distribution duration by month based on y who accept is higher than rejected.

Line Plot chart of increasing variances



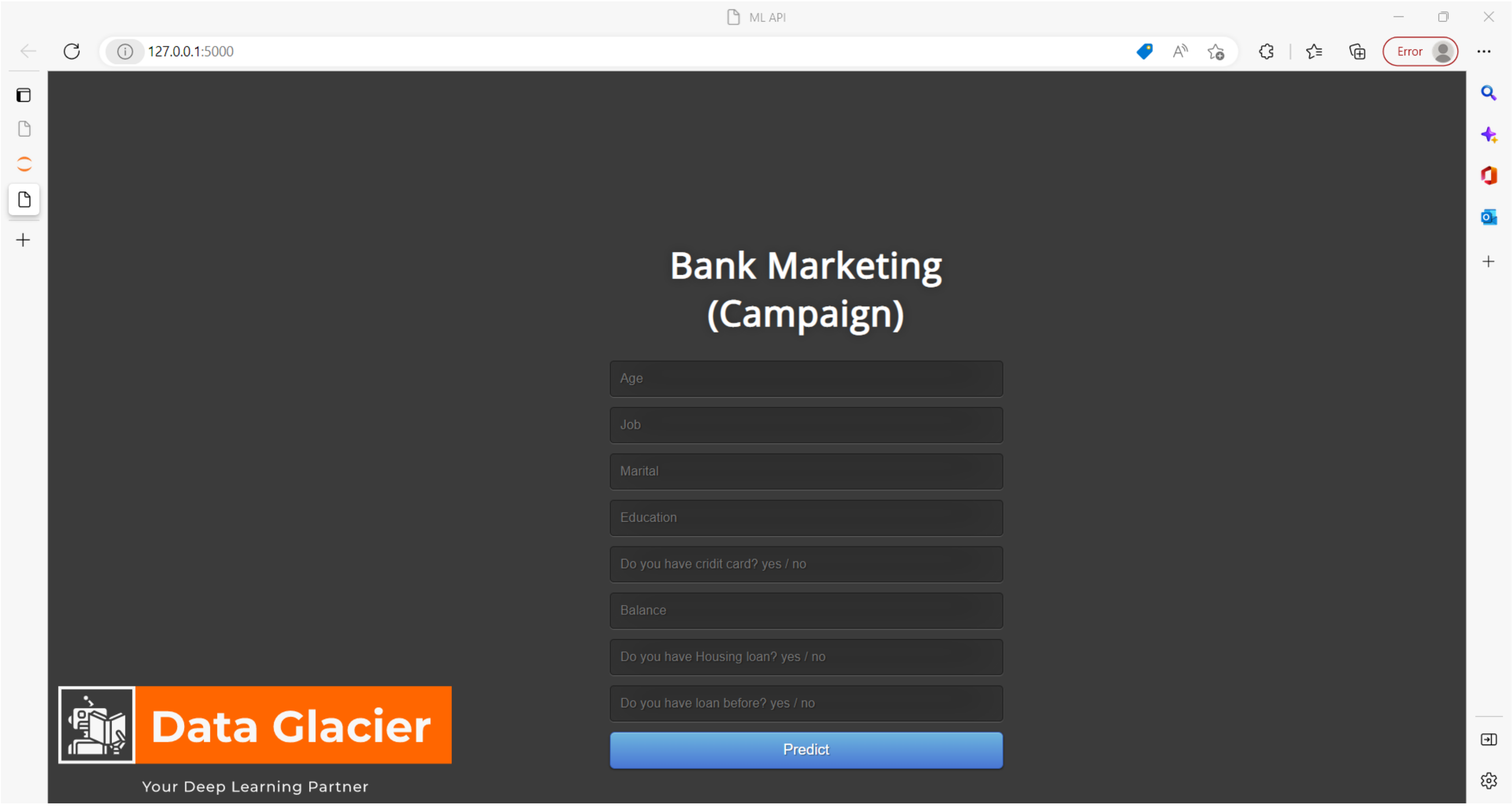
Model Building (Logistic Regression, Random Forest, Decision Tree)

Tabel of Comparing between Models

Score	Logistic Regression	Random Forest Classifier	Decision Tree
Accuracy	0.95	0.94	0.91
Precision	0.61	0.61	0.61
Recall	0.33	0.33	0.33

So, The best model is Logistic Regression based on Accuracy, Precision, Recall

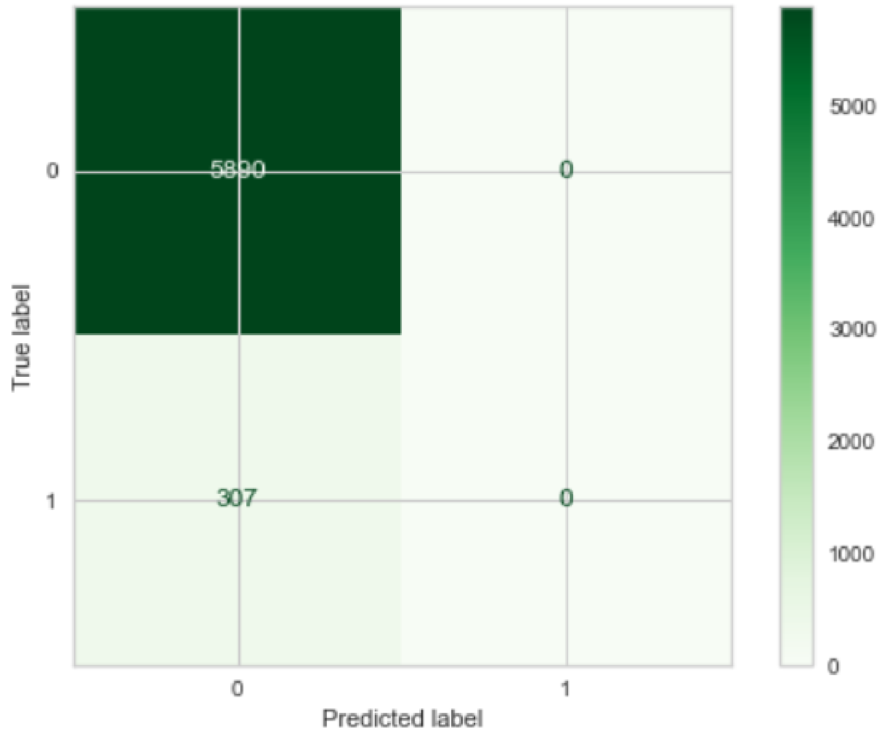
Deploy the model using a flask



Converting ML metrics into Business metrics and explaining results to the business.

1. Confusion Matrix
2. F1 Score
3. Gain and Lift charts

1. Confusion Matrix



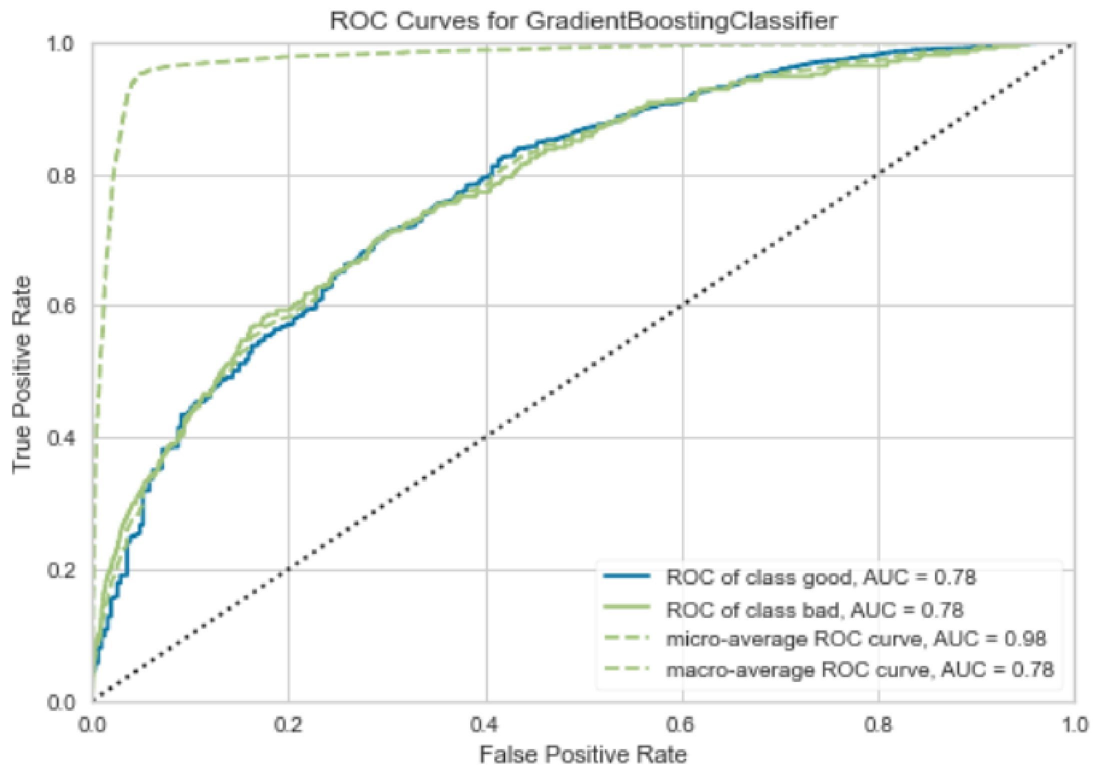
2. F1 Score

2. F1 Score

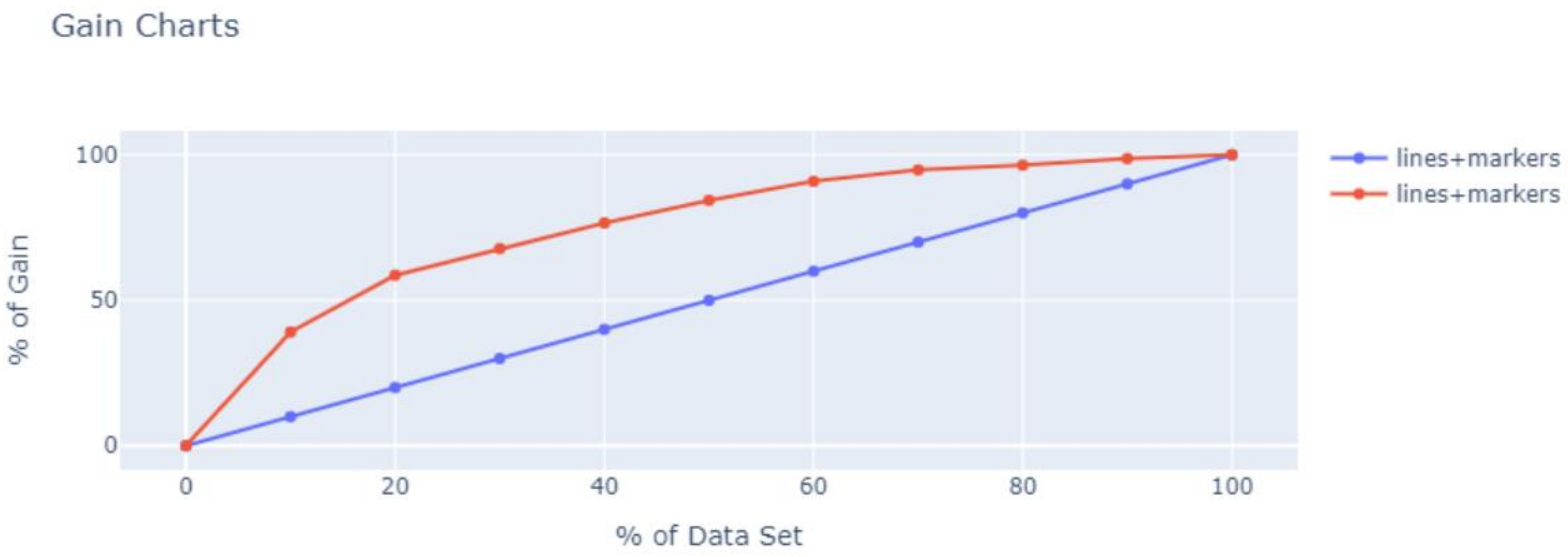
```
# F1 Score = 2 * (Precision * Recall) / (Precision + Recall)
# where:
# Precision: Correct positive predictions relative to total positive predictions
# Recall: Correct positive predictions relative to total actual positives

F1_Score = 2 * (0.613918 * 0.324924) / (0.613918 + 0.324924)
F1_Score
```

```
0.4249419864726972
```

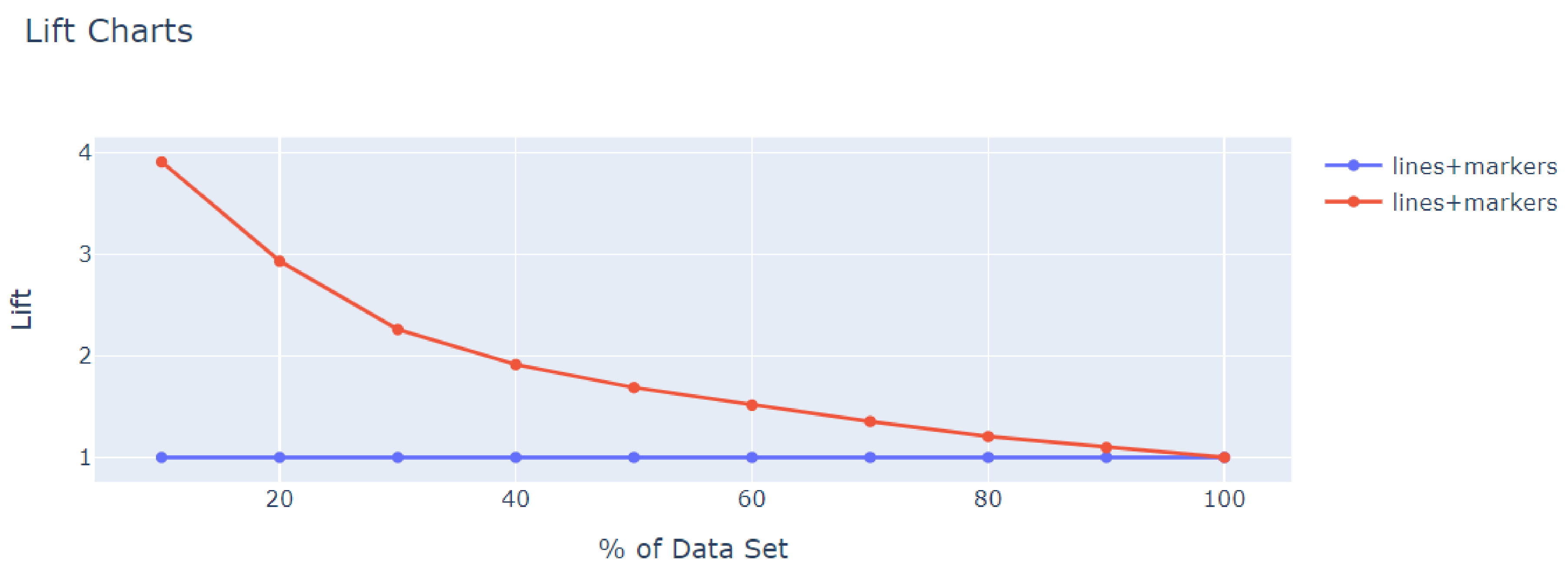
3. Gain and Lift charts



3. Gain and Lift charts



3. Gain and Lift charts



Recommendations

- The model can predict a 0.95 form result so, In the future, I want to improve the model to increase the value and the bank in another campaign.

- Employees should take the most important from clients to get the right data such as job, education, age, whether the client has a housing loan or not, whether the client has a loan before or not, whether the client has a credit card or not, balance, and what is the marital status for the client.

- After doing EDA, Feature Engineering, PCA, and Test Multiple models.

- In the Future, I will be testing more models to improve the result to help the banking sector in future campaigns.

Bank Marketing (Campaign)

Data Girl

Date: 19 Nov 2022

LISUM 14

Fatimah Asiri

Data Glacier

Your Deep Learning Partner

Thank You