Bank Marketing Campaign

Aly Medhat

Data Glacier

Your Deep Learning Partner

TABLE OF CONTENTS

01 Introduction

02 EDA & Recommendations

03 Model Recommendations

04 Evaluation

05 GIT Repo.

Data Glacier

Introduction

BACKGROUND

ABC Bank wants to sell its term deposit product to customers and needs a model that can predict weather a particular customer would buy a product or no.

OBJECTIVE

Build a ML model for the bank that shortlist customer whose chance of buying the product is more.

Data Glacier

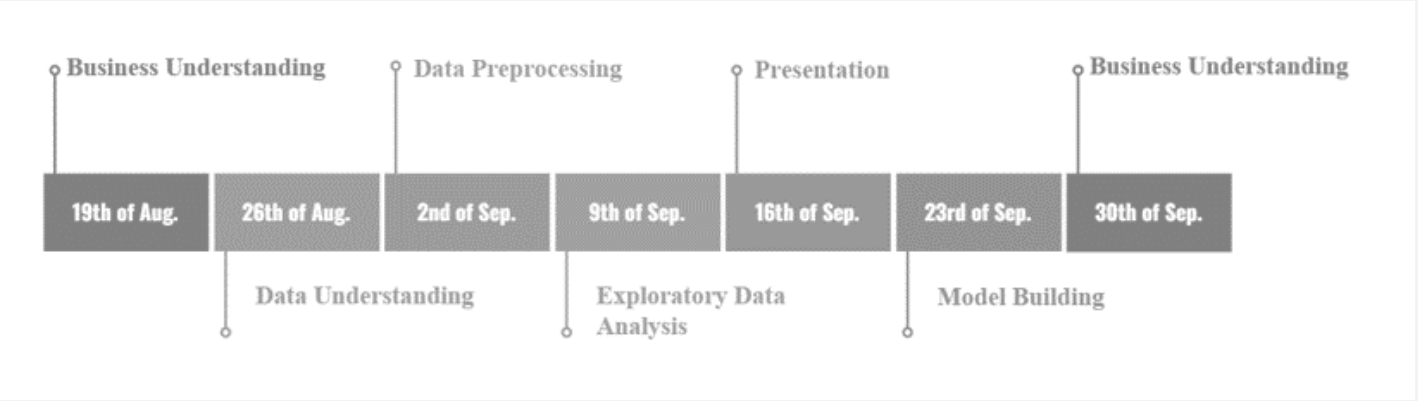
Introduction

BUSINESS UNDERSTANDING

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls.

Data Glacier

Project Lifecycle



Data Glacier

TABLE OF CONTENTS

01 Introduction

02 EDA & Recommendations

03 Model Recommendations

04 Evaluation

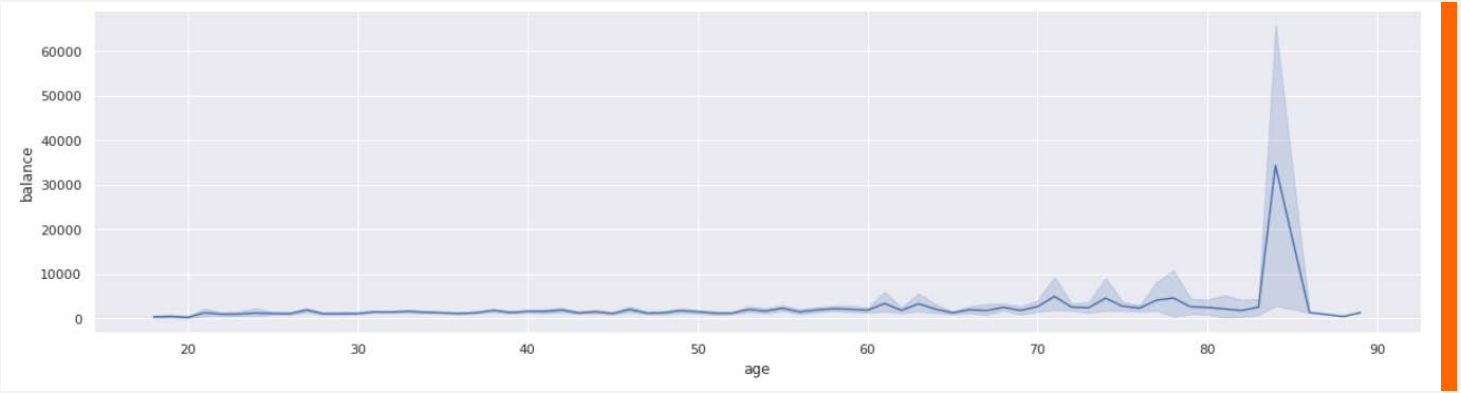
05 GIT Repo.

Data Glacier

ANALYZING THE DATA

"Exploratory Data Analysis"

The relation between the bank balance and the client age

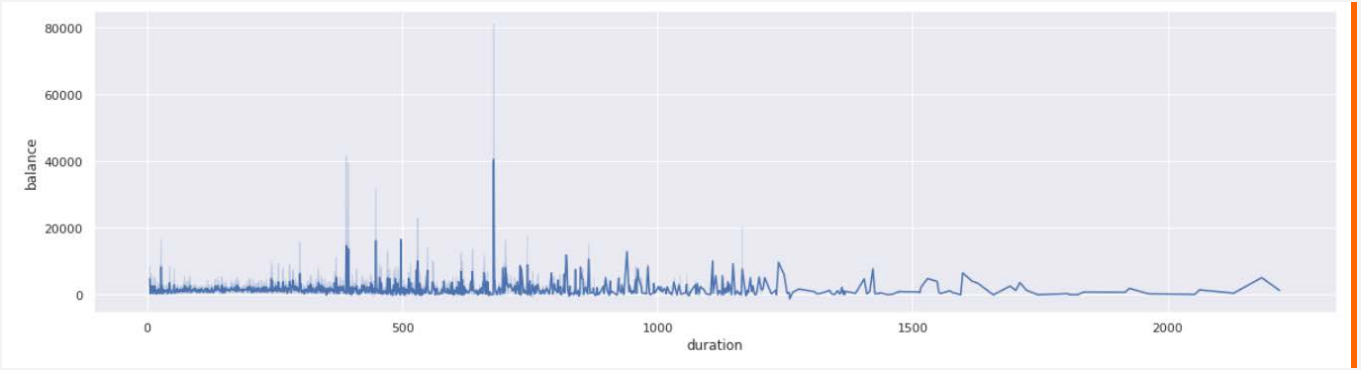


- The graph shows the relation between the account balance and the age of the client. This mayn't help in the final prediction but gives an insight of the data

ANALYXING THE DATA

"Exploratory Data Analysis"

The relation between the bank balance and the client age

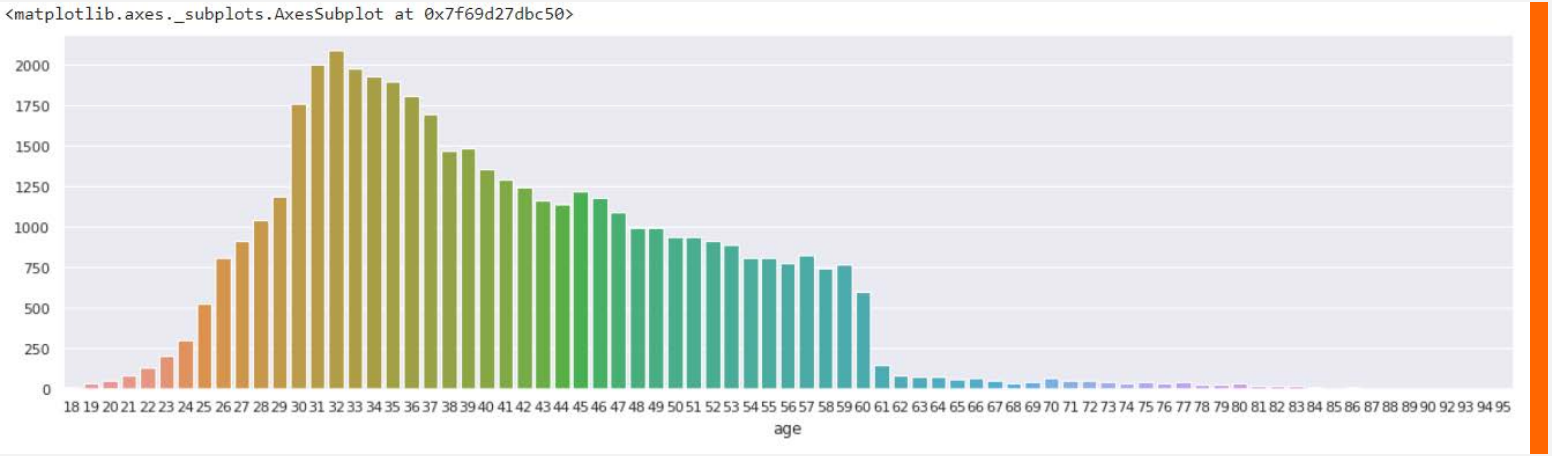


- The graph shows the relation between the account balance and the duration. This mayn't help in the final prediction but gives an insight of the data

ANALYXING THE DATA

"Exploratory Data Analysis"

The relation between the total number of applications and the client's age

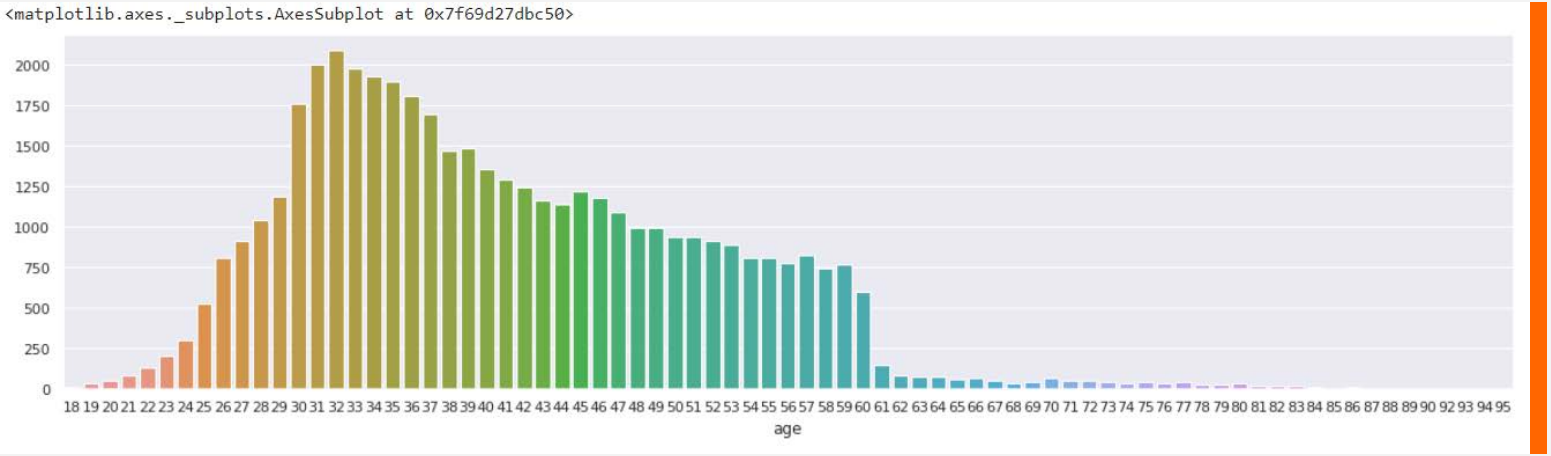


- The graph shows the relation between the total number of applications the bank received and the age of the applicant. It can be seen that most applicants age varies from 25-60 years old.

ANALYZING THE DATA

"Exploratory Data Analysis"

The relation between the accepted applications and the client's age

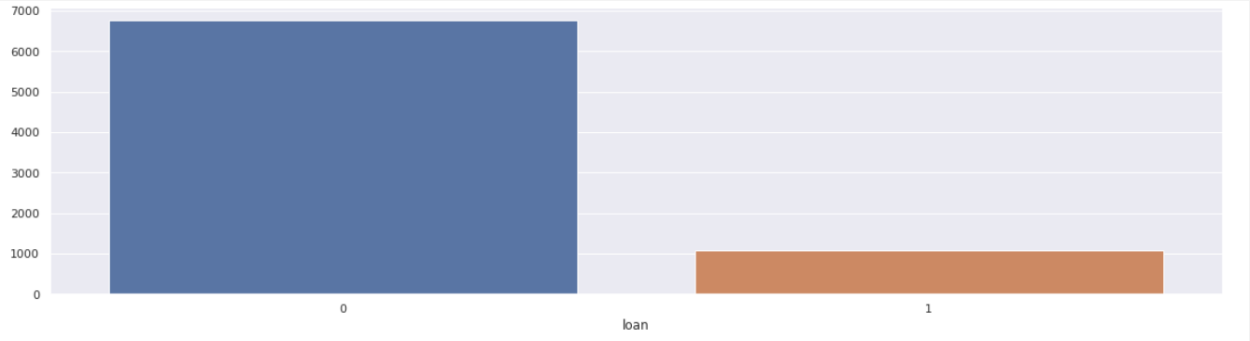


- The graph shows the relation between the number of accepted applications and the age of the applicant. The graph is similar to the previous graph meaning that there is no dependency of accepting the application and the applicant age.

ANALYZING THE DATA

"Exploratory Data Analysis"

The relation between the accepted applications and if they had a load

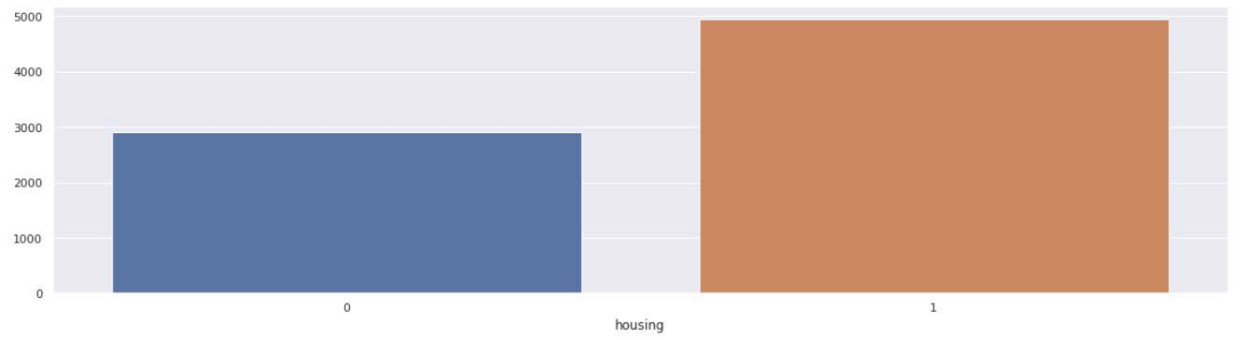


- It can be seen that most of the accepted applications didn't have a prevois loan. Meaning that it can be taken into consideration in decision making

ANALYZING THE DATA

"Exploratory Data Analysis"

The relation between the accepted applications and if they have a house

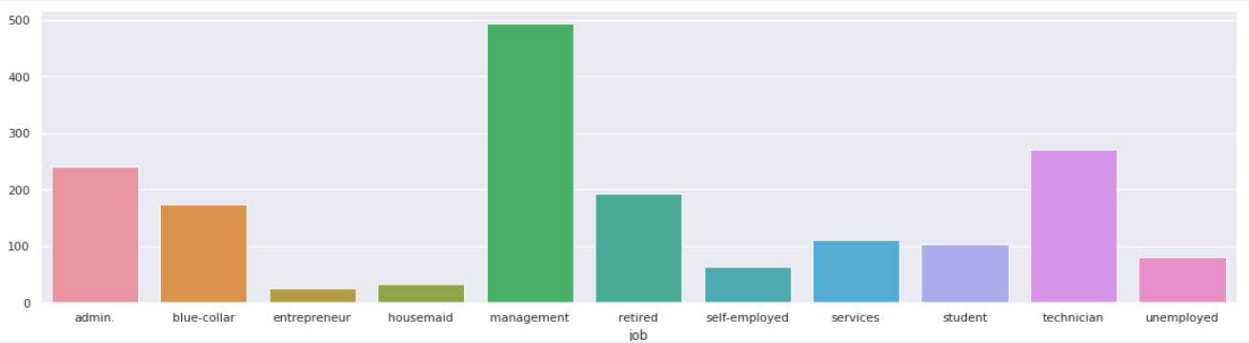


- It can be seen that most of the accepted applications have house. The relation is not that high still it should bet taking in consideration in decision making

ANALYZING THE DATA

"Exploratory Data Analysis"

The relation between the accepted applications and the applicant's job

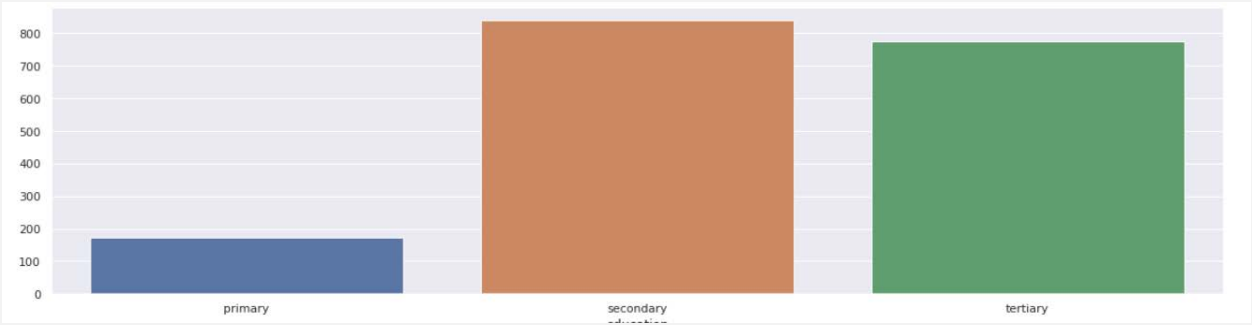


- It can be seen some jobs have higher probability of getting accepted. This means it should be taken into consideration.

ANALYZING THE DATA

"Exploratory Data Analysis"

The relation between the accepted applications and the applicant's education

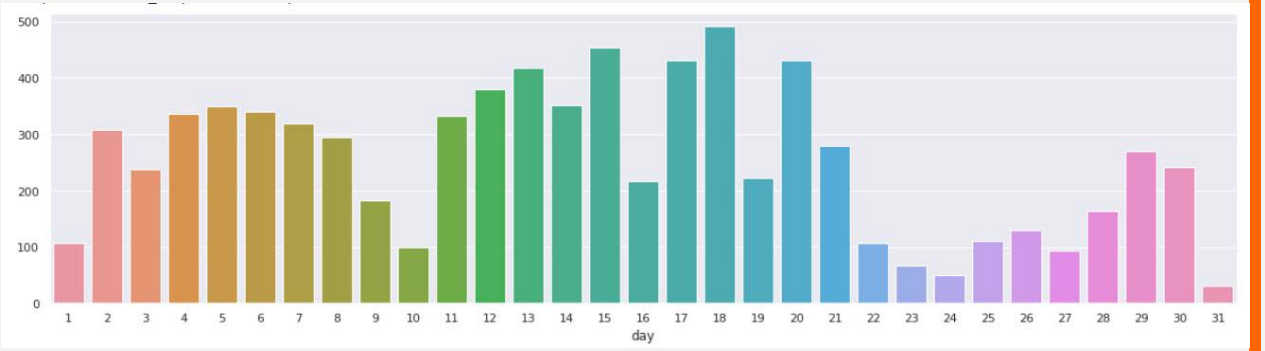


- It can be seen education influences the probability of getting accepted. This means it should be taken into consideration.

ANALYZING THE DATA

"Exploratory Data Analysis"

The relation between the accepted applications and the application day

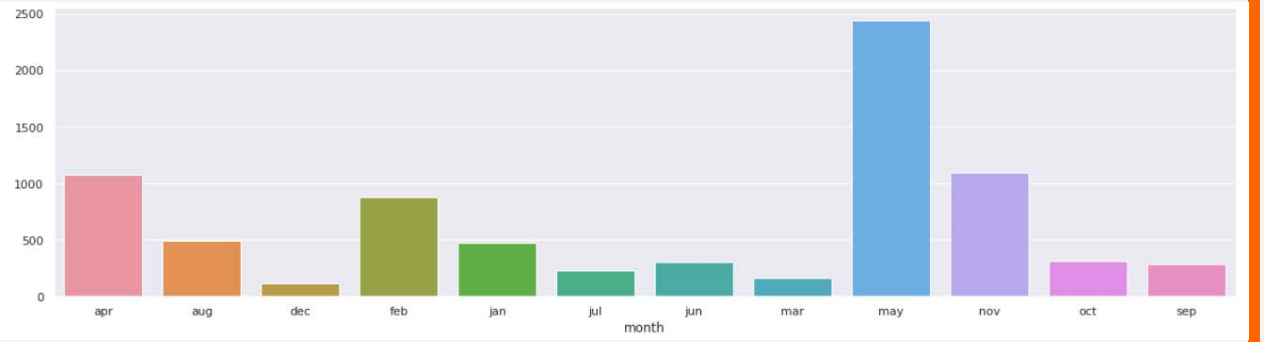


- It can be seen that the day of application does not influences the probability of getting accepted. This means it should not be taken into consideration.

ANALYZING THE DATA

"Exploratory Data Analysis"

The relation between the accepted applications and the application month



- It can be seen that the month of application does not influences the probability of getting accepted. This means it should not be taken into consideration.

Final Recommendations

1. Outliers are not to be removed.
2. There is a direct relation between the loan and the acceptance of the application (Important feature for classification).
3. Some jobs have higher priority for acceptance (important feature for classification).
4. Some features do not have effect on the output as day and month (Not to consider).

Data Glacier
Your Deep Learning Partner

TABLE OF CONTENTS

01 Introduction

02 EDA & Recommendations

03 Model Recommendations

04 Evaluation

05 GIT Repo.

Data Glacier

Model Recommendation

- The first model will be Naïve Base because:
  - It is very fast at classifying unknown records.
  - It is simple and easy to implement.
  - It doesn't require as much training data.
  - It handles both continuous and discrete data.
  - It is highly scalable with the number of predictors and data points.

- The second model will be Linear regression because:
  - It is easier to implement, interpret, and very efficient to train.
  - It can interpret model coefficients as indicators of feature importance.
  - It can interpret model coefficients as indicators of feature importance.

Model Recommendation

- The Third model will be Ensemble Models (logistic regression, xgboost, and random forest) using max voting because:
    - Reduce the spread in the average skill of a predictive model.
    - Improve the average prediction performance over any contributing member in the ensemble.
    - Has better performance than singles model.

- The Fourth model will be Ensemble Models (Bagging) because:
    - Allows many weak learners to combine efforts to outdo a single strong learner
    - Higher predictive accuracy, compared to the individual models

Model Recommendation

- The Third model will be Ensemble Models (logistic regression, xgboost, and random forest) using max voting because:
    - Reduce the spread in the average skill of a predictive model.
    - Improve the average prediction performance over any contributing member in the ensemble.
    - Has better performance than singles model.

- The Fourth model will be Ensemble Models (Bagging) because:
    - Allows many weak learners to combine efforts to outdo a single strong learner
    - Higher predictive accuracy, compared to the individual models

Model Recommendation

- The Fifth model will be Stacking classifier (RF, KNN - DT, RF) because:
    - Stacking can yield improvements in model performance.
    - Stacking reduces variance and creates a more robust model by combining the predictions of multiple models.

Data Glacier

Your Deep Learning Partner

TABLE OF CONTENTS

01 Introduction

02 EDA & Recommendations

03 Model Recommendations

04 Evaluation Recommendations

05 GIT Repo.

Data Glacier

MODEL EVALUATION

The models can be evaluated using different methods, here three evaluation methods shall be used.

1. **Accuracy:** The number of classifications a model correctly predicts divided by the total number of predictions made.

2. **Confusion Matrix:** is a summary of prediction results on a classification problem.

3. **F1 Score:** The F-score is a way of combining the precision and recall of the model.

MODELS EVALUATION

1. Naïve Bayes

Accuracy	Confusion Matrix		F1 score
	7539	**139**	
0.8930.853	778	183	0.28526

MODELS EVALUATION

2. Linear Regression

Accuracy	Confusion Matrix		F1 score
	7190	**488**	
0.87486	593	368	0.405063

MODELS EVALUATION

3. Ensemble Model Max Voting

Accuracy

0.9007

Confusion Matrix

	7490	188
669		
292		

F1 score

0.405274

MODELS EVALUATION

4. Ensemble Model Bagging

Accuracy	Confusion Matrix		F1 score
	7494	**209**	
0.9018	639	297	0.4119

MODELS EVALUATION

5. Stacking

Accuracy	Confusion Matrix		F1 score
	7446	**232**	
0.8996	692	269	0.404123

TABLE OF CONTENTS

01 Introduction

02 EDA & Recommendations

03 Model Recommendations

04 Evaluation

05 GIT Repo.

Data Glacier

GitHub Repository link

https://github.com/alymedhat10/Bank-Marketing-Campaign-.git

Data Glacier

Your Deep Learning Partner

THANK YOU

Data Glacier