



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

BANK MARKETING CAMPAIGN

TEAM MEMBER

NAME: FREDRICKA DAVID

EMAIL: frederickadavid@gmail.com

COUNTRY: NAMIBIA

COLLEGE/COMPANY: OMDENA

SPECIALIZATION: DATA SCIENCE

Agenda

Problem Statement

Data Analysis

Data Cleaning

EDA

Final Recommendations

GitHub Repo Link

Problem Statement

Overview

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).

Solution

By analysing the dataset we will be able to convert this problem into a machine learning classification and build a model to predict whether a client will subscribe a term deposit or not.

Data Analysis

- **21 Features**
- **41188 rows**

Assumptions:

- **The data seems to be cleaned and a little bit skewed, however you can see that the variables have outliers that need to be cleaned with data cleaning process.**
- **There are no null values, but there are some "unknown" values.**
- **We've got 41188 rows, some of the columns has 85% of the values repeated.**
- **The value types seem to be correct but as said there are some "unknown" values.**

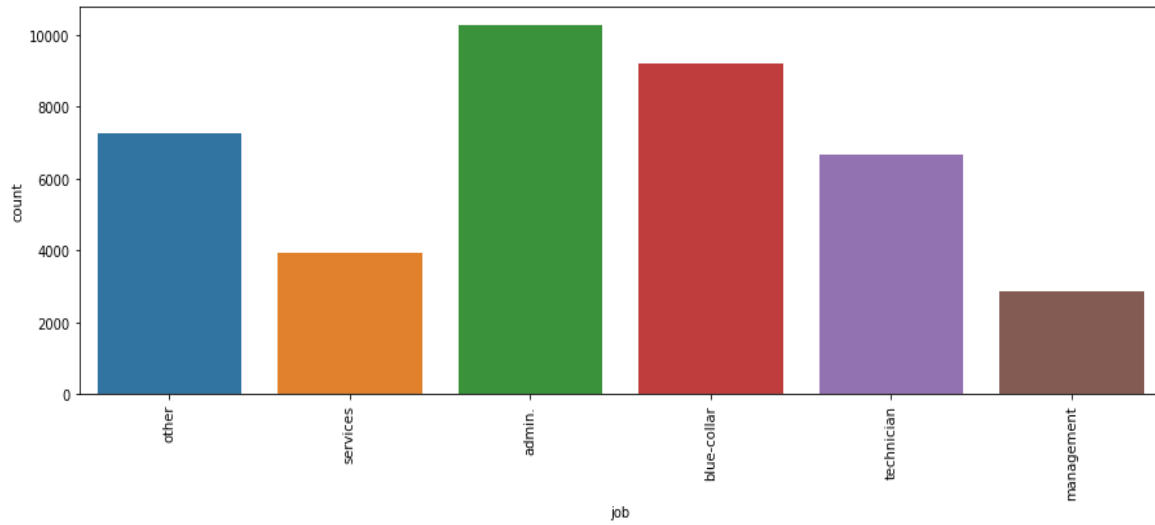
Data Cleaning

The most used techniques to treat the data frame were:

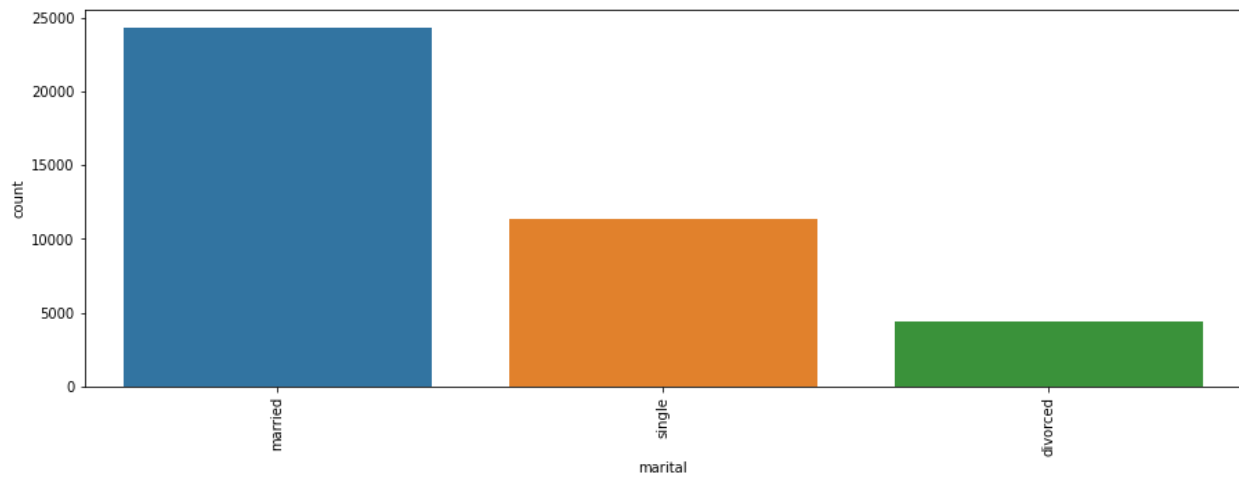
- Dropping columns and rows**
- Replacing missing values with the mode**
- Filling outliers with ffill/bfill based on interquartile range**
- Grouping text features that doesn't appear many time**

Cleaned dataset reduced from (41188, 21) to (40165, 18)

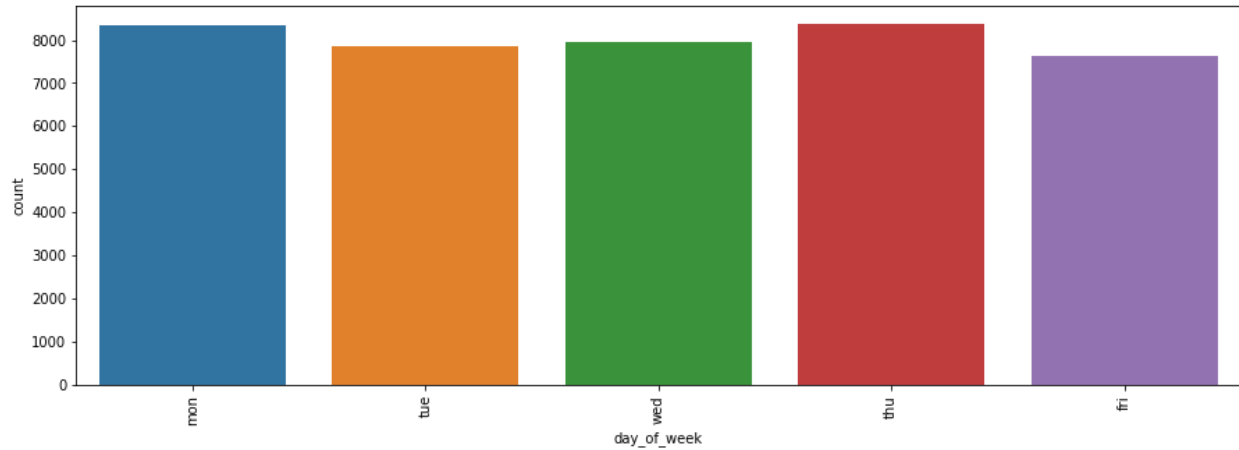
Data Cleaning



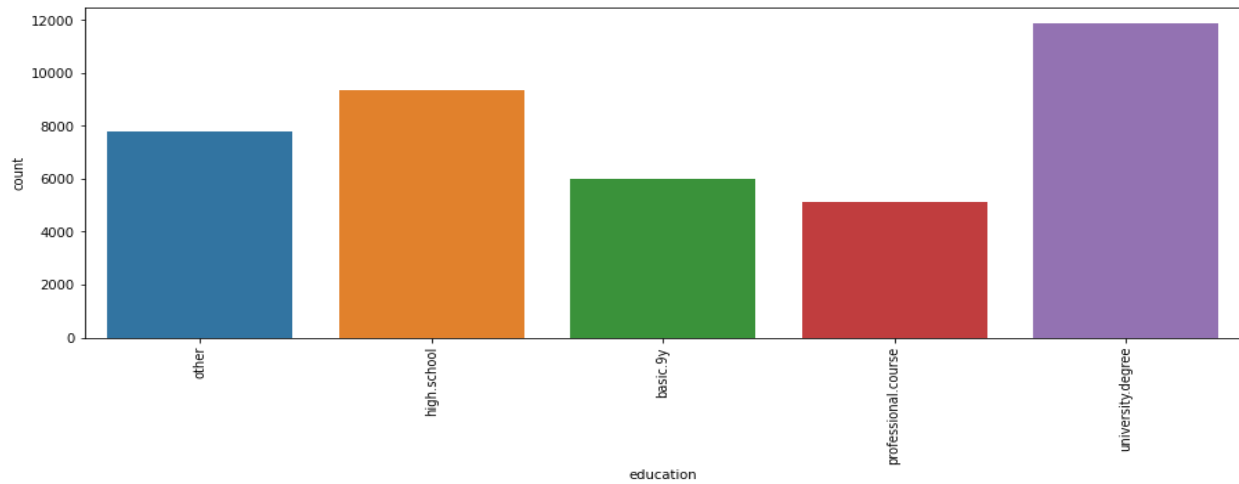
Job: Has the highest number of subscriptions to a term deposit work on admin.



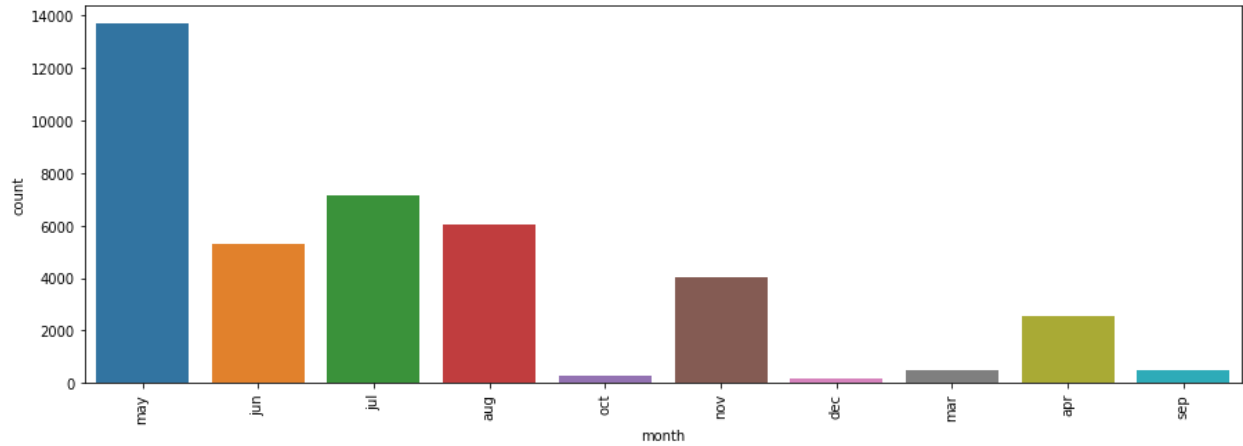
Marital: Most of the clients that were approached were married.



Day of Week: There's no significant difference between the numbers of clients that were approached and the people that subscribed.

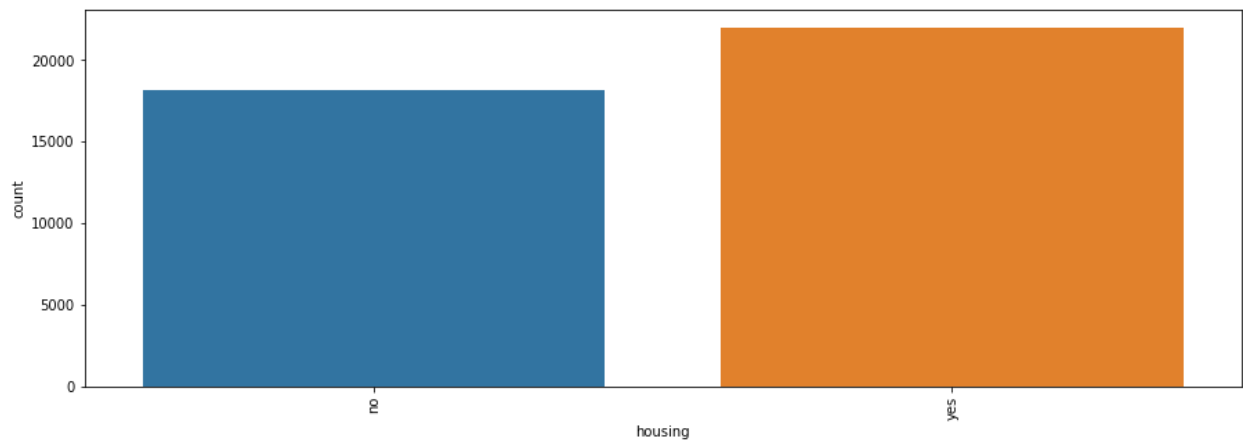


Education: In the education column we can see that most people who has subscribed have a university degree.

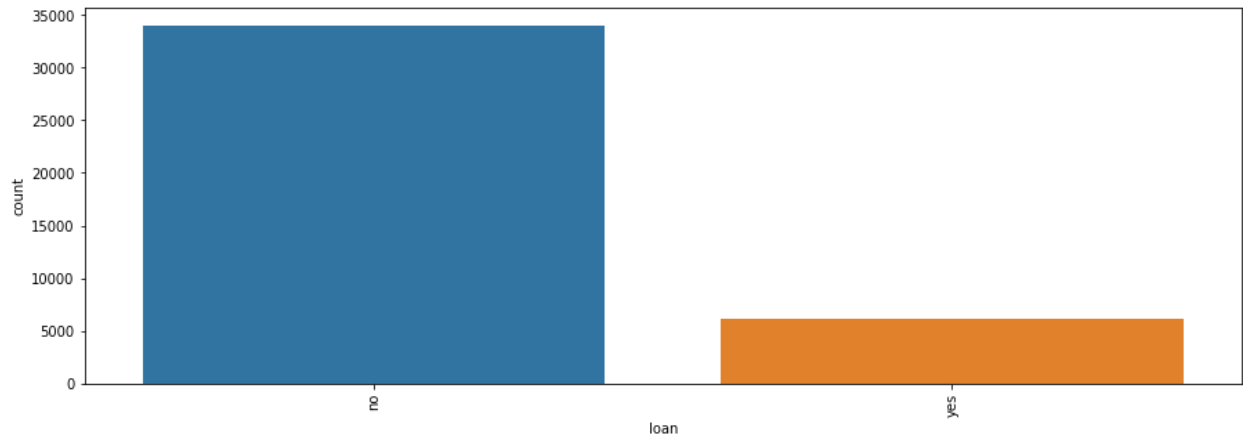


Month: The last contact month of year was way bigger in May.

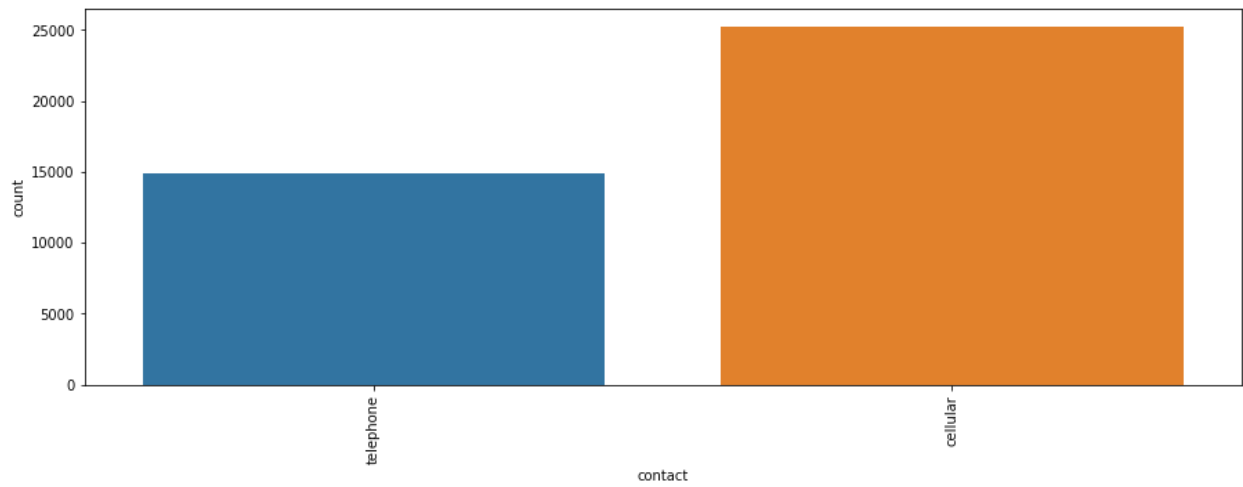
EDA – Numerical and Attributes



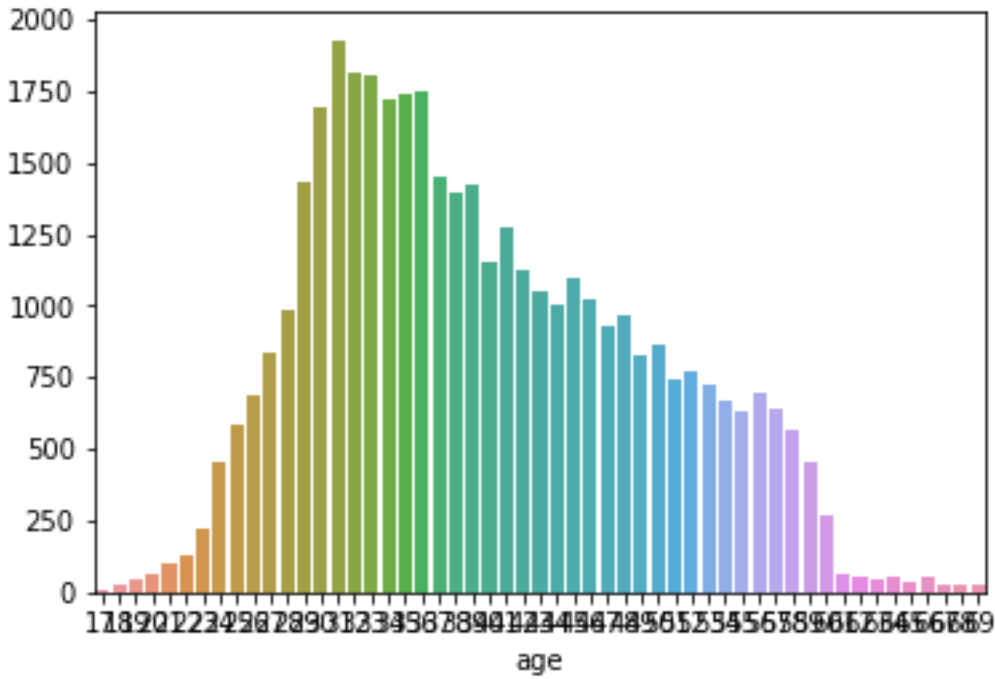
Housing: A housing loan does not have much of effect on the number of term deposits purchased.



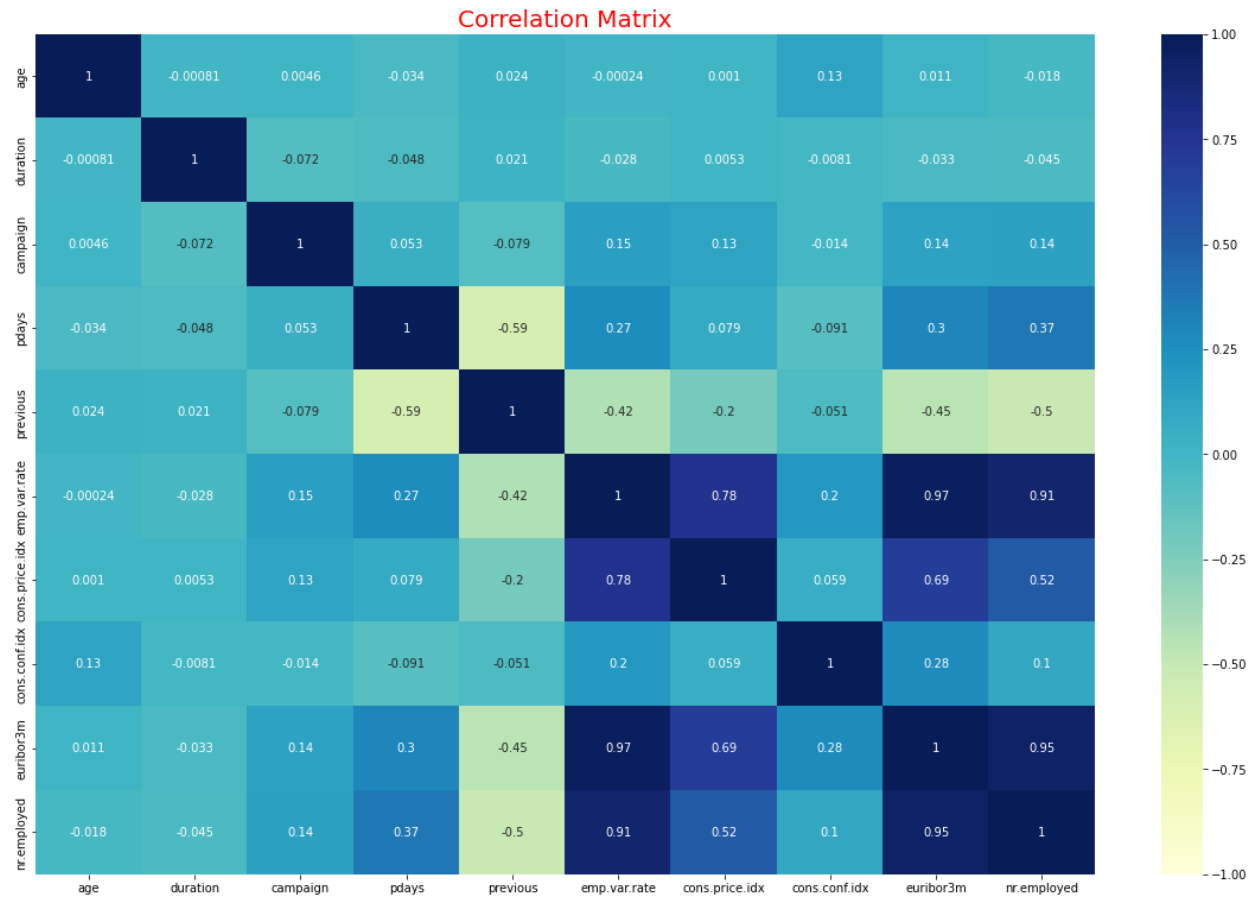
Loan: Most of the clients approached did not have a personal loan.



Contact: The contact by phone seems to be more effective.



The relation between the total number of applications the bank received and the age of the applicant. It shows that most applicant's age varies from 25-60 years old.



There is high correlation between columns 'euribor3m', 'nr.employed' and 'emp.var.rate'.

Final Recommendations

There is a direct relation between the loan and the acceptance of the application which is an important feature for classification.

Some jobs have higher priority for acceptance which is an important feature for classification.

Some features do not have effect on the output as day and month.

Model Recommendations

- Naives Bayes
- Linear Regression
- Random Forest Classifier
- Decision Tree Classifier

GitHub Repo Link

[sharuhinda/bank_marketing_campaign \(github.com\)](#)

[Fredricka23/Bank-Marketing-Campagain---Week-7: Specialization Project \(github.com\)](#)