

AI Project

ML Marathon Dataset by
Azure Developer
Community

A decorative graphic consisting of three parallel diagonal stripes in teal, light gray, and black, extending from the bottom left towards the top right of the slide.

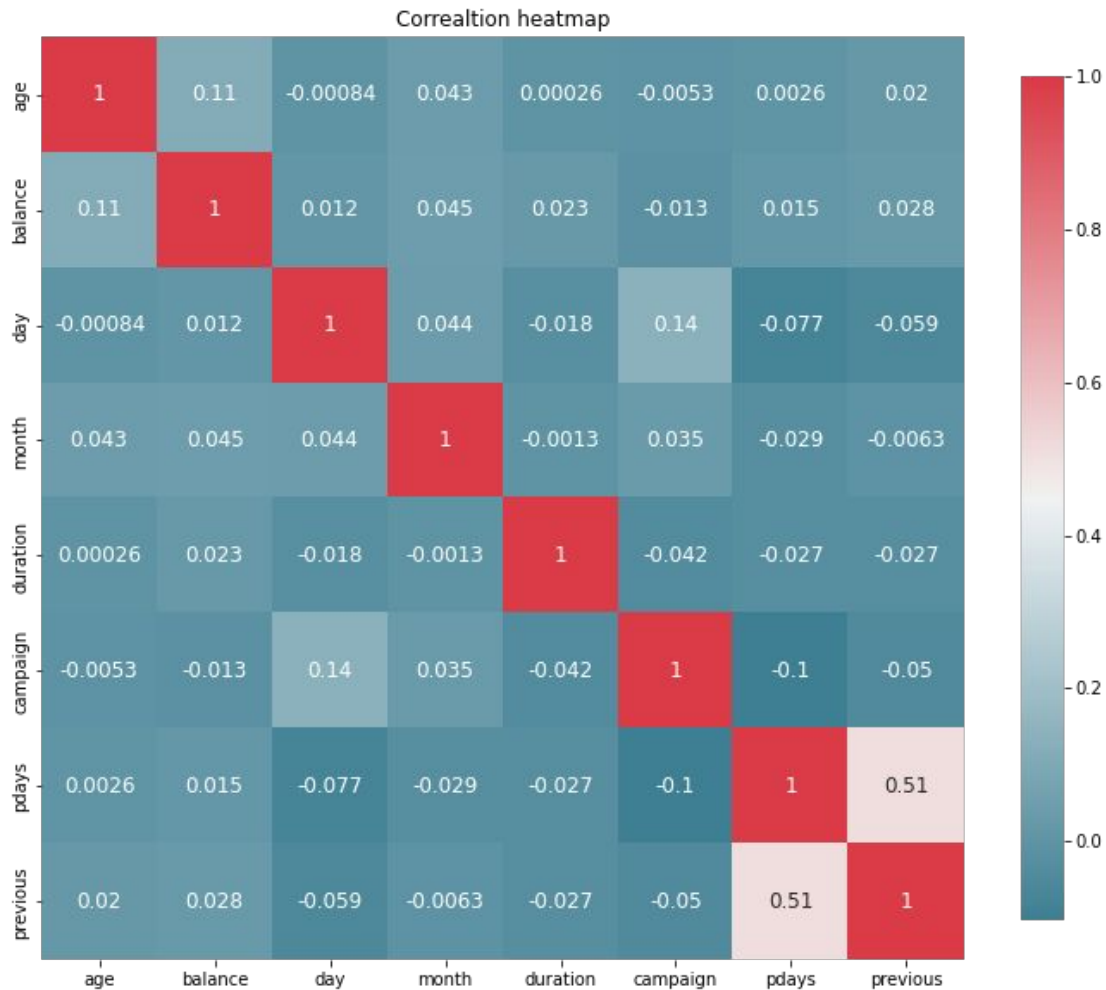
Understanding The Problem

The data is related to direct marketing campaigns of a financial institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to assess if the product (bank term deposit) would be ('yes') or not ('no') subscribed. You will have to analyze the dataset in order to find ways to look for future strategies in order to improve future marketing campaigns for the bank.

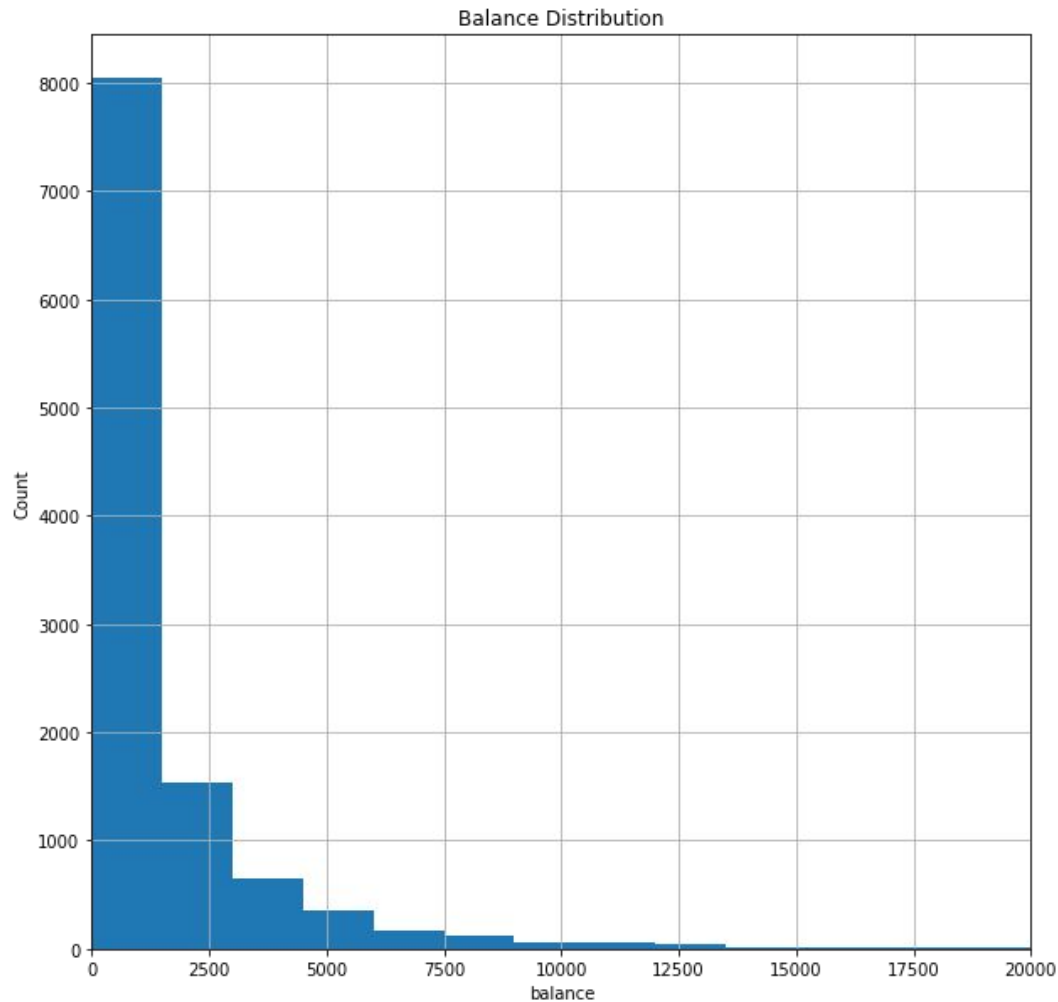
Data Preparation

- *Dataset Description (View Colab Notebook)*
- *Dataset Preprocessing : Assessing data, Cleaning Null values and imputing the missing ones.*

Visualization

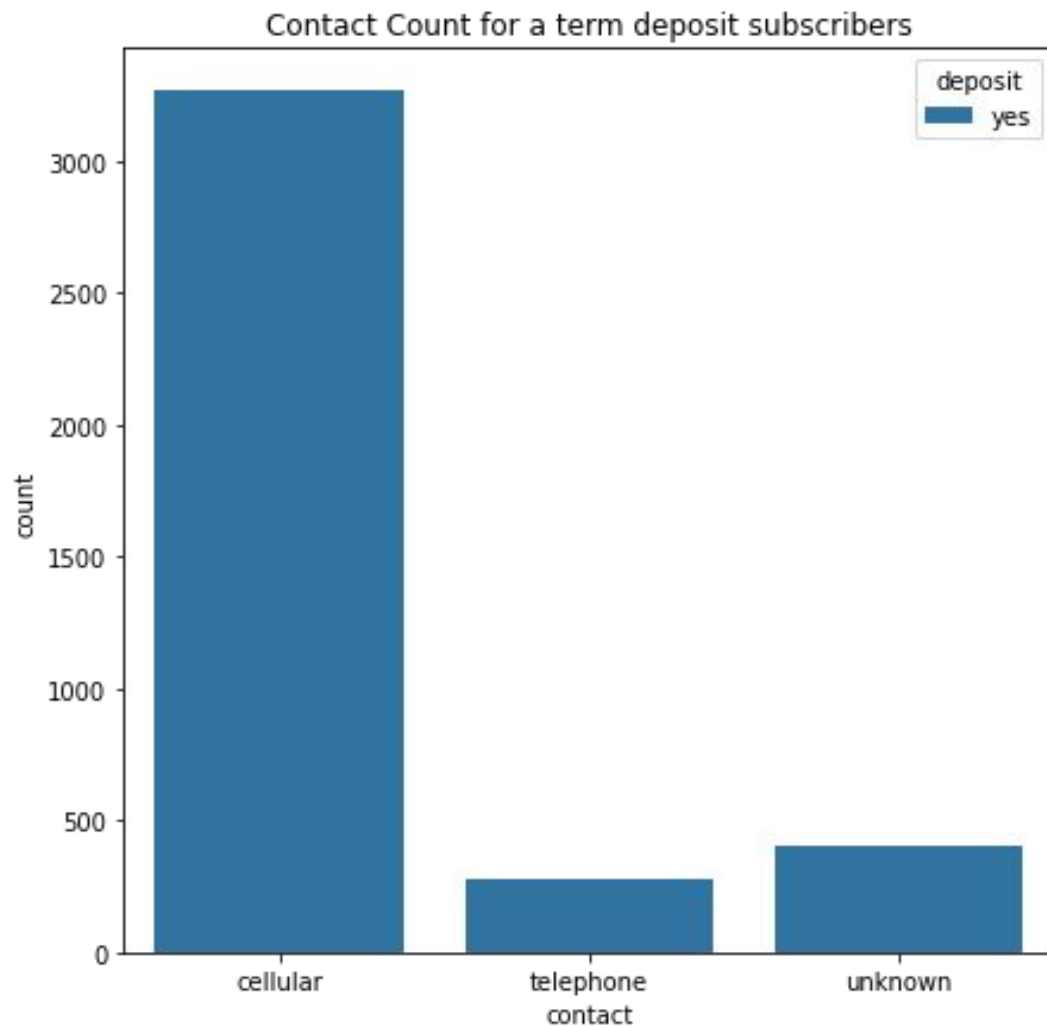


Correlation Heat Map



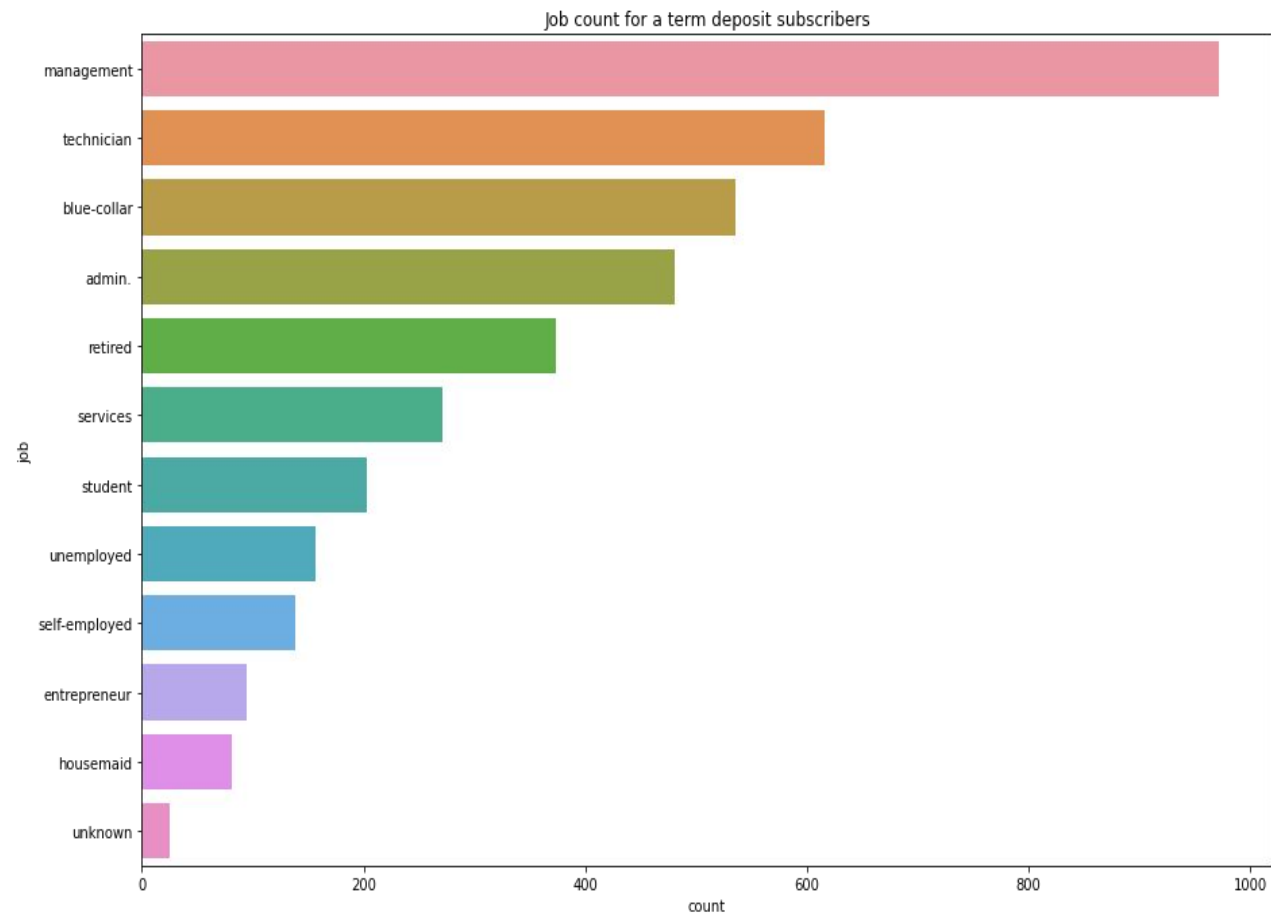
Balance Distribution

→ We see a right skewed histogram with a peak from 0 to 1500 with 8000 count.



Contact count for a term deposit subscribers

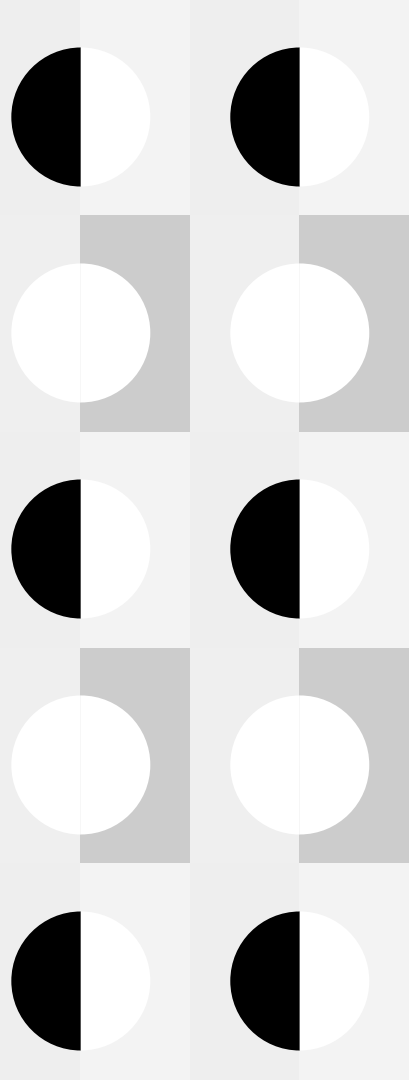
→ We see that people who've been contacted through cellular are often to subscribe a term deposit.



Job count for a term deposit subscribers

→ We observe that management are the job having the most term deposit subscribers

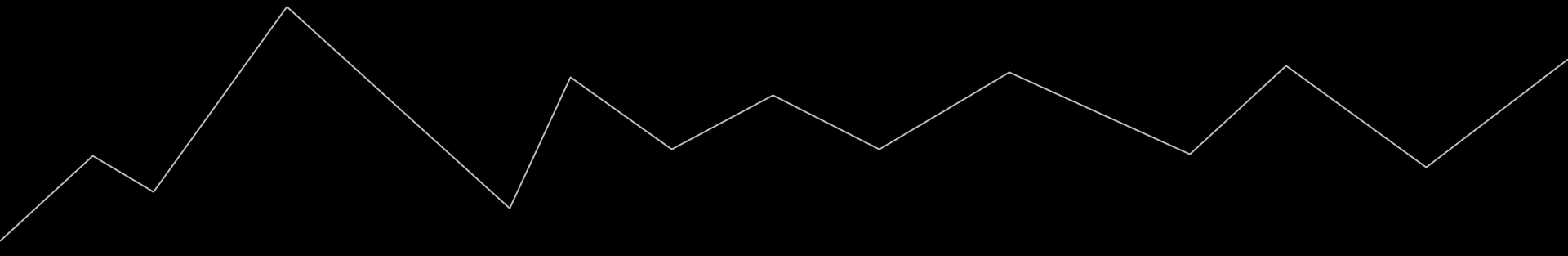
Modeling



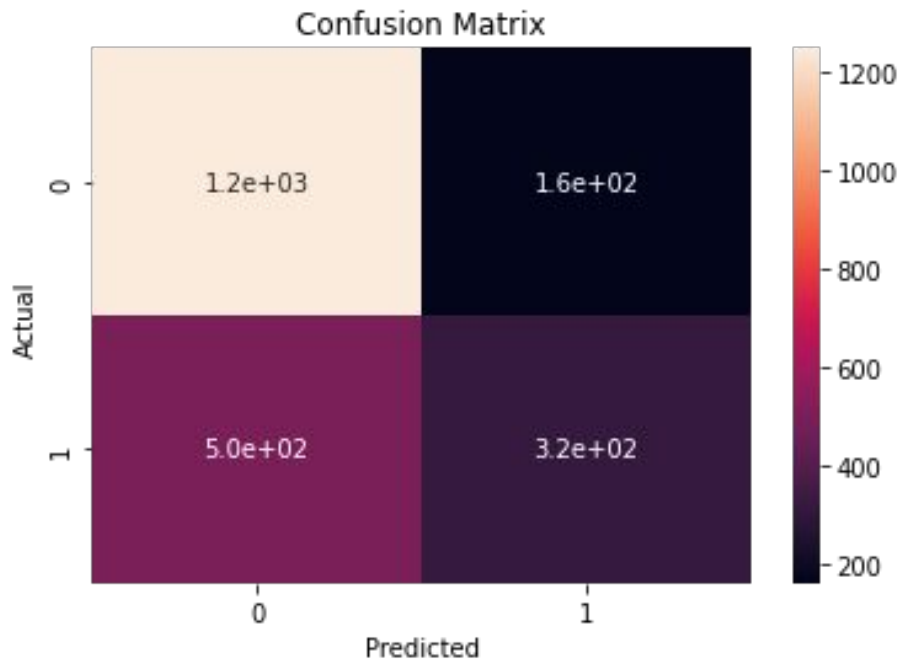
- ***Logistic Regression***
- ***KNN***
- ***Random Forest (train test split Method + Cross Validation Method)***
- ***Agglomerative Clustering***
- ***K-Means***
- ***(Multi) Linear Regression (train test split Method + Cross Validation Method)***
- ***(Multi) Polynomial Regression (train test split Method + Cross Validation Method)***

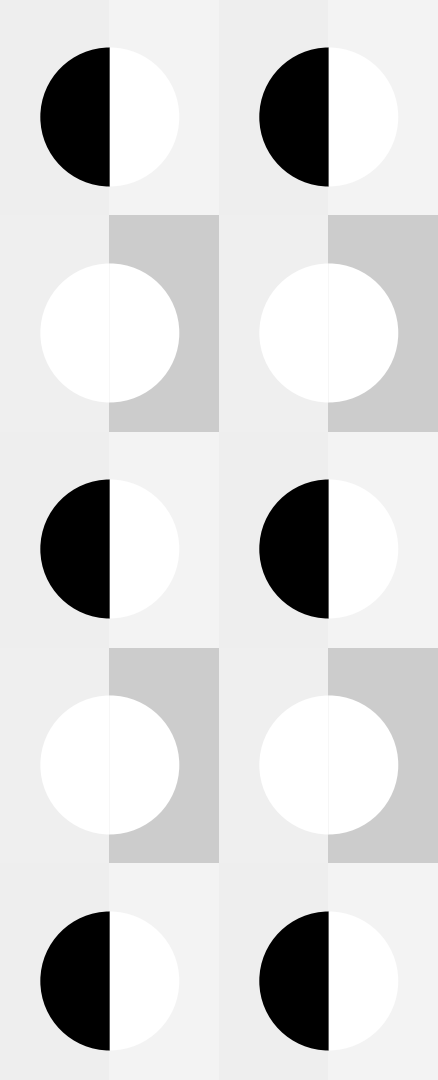
Metrics

Supervised Learning

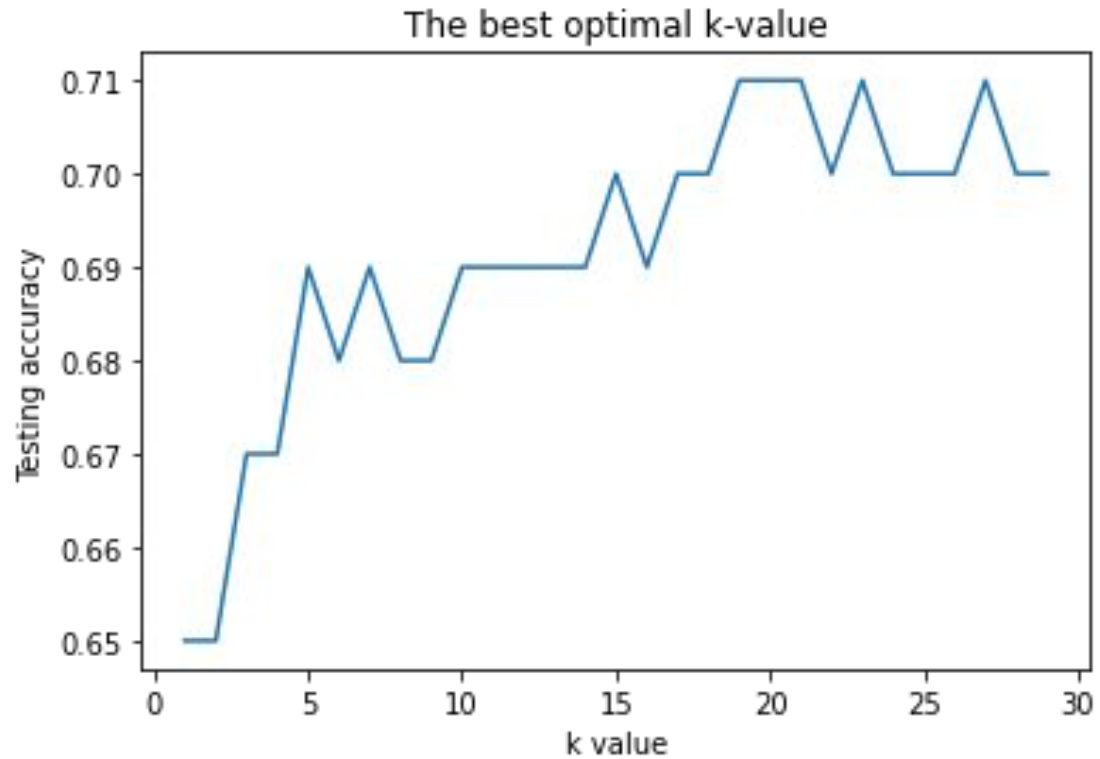


Logistic Regression : 70%



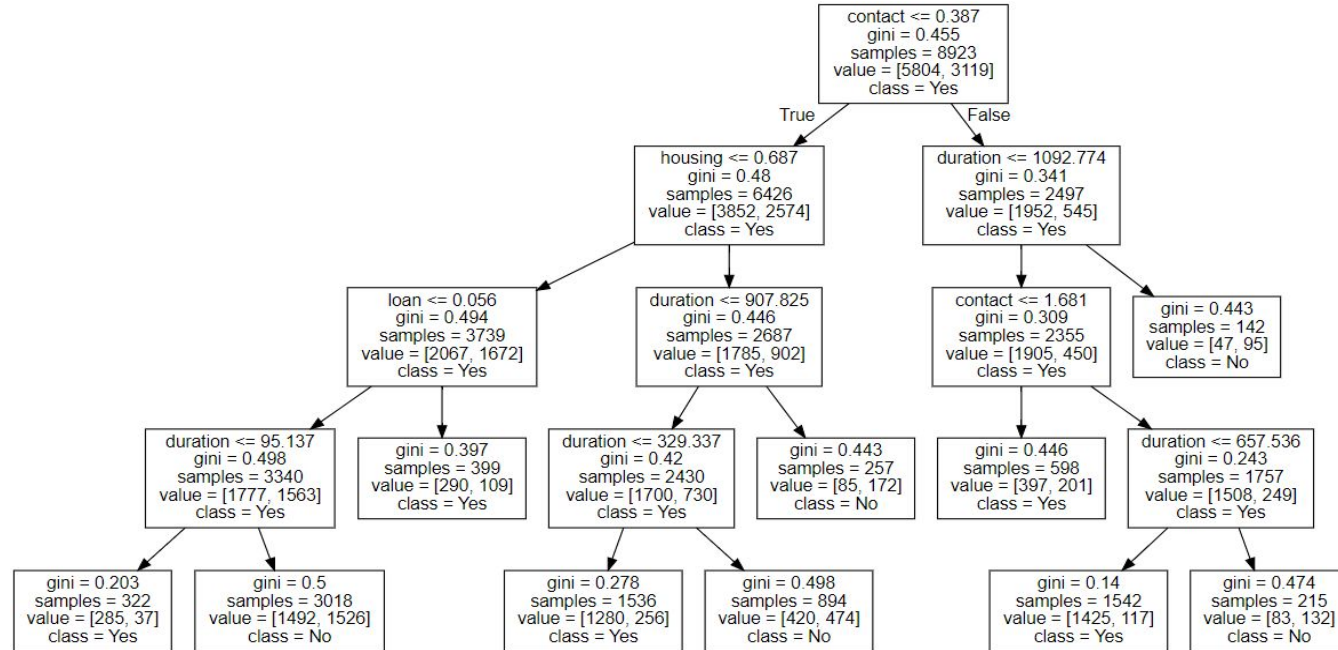


KNN : 71%

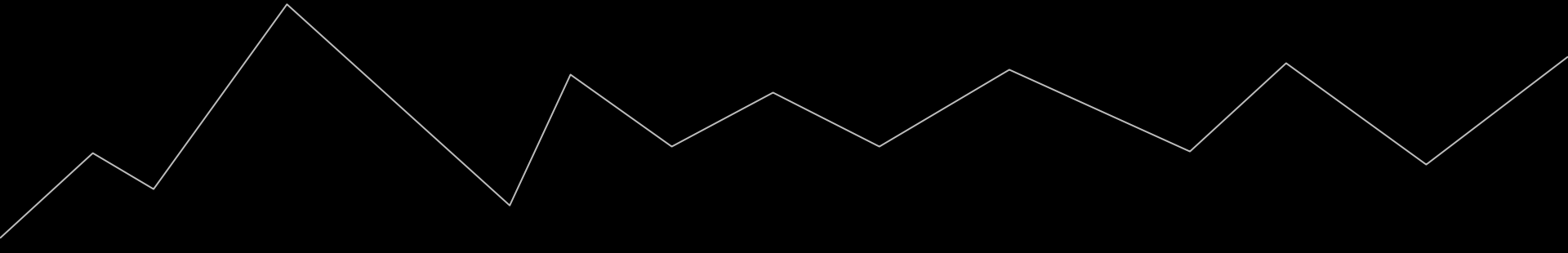


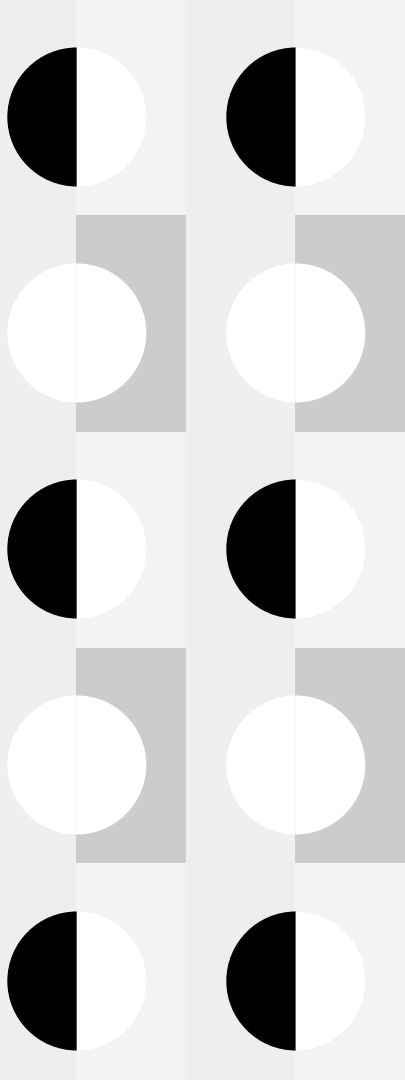
Decision Tree : 69%

Random Forest : 72%

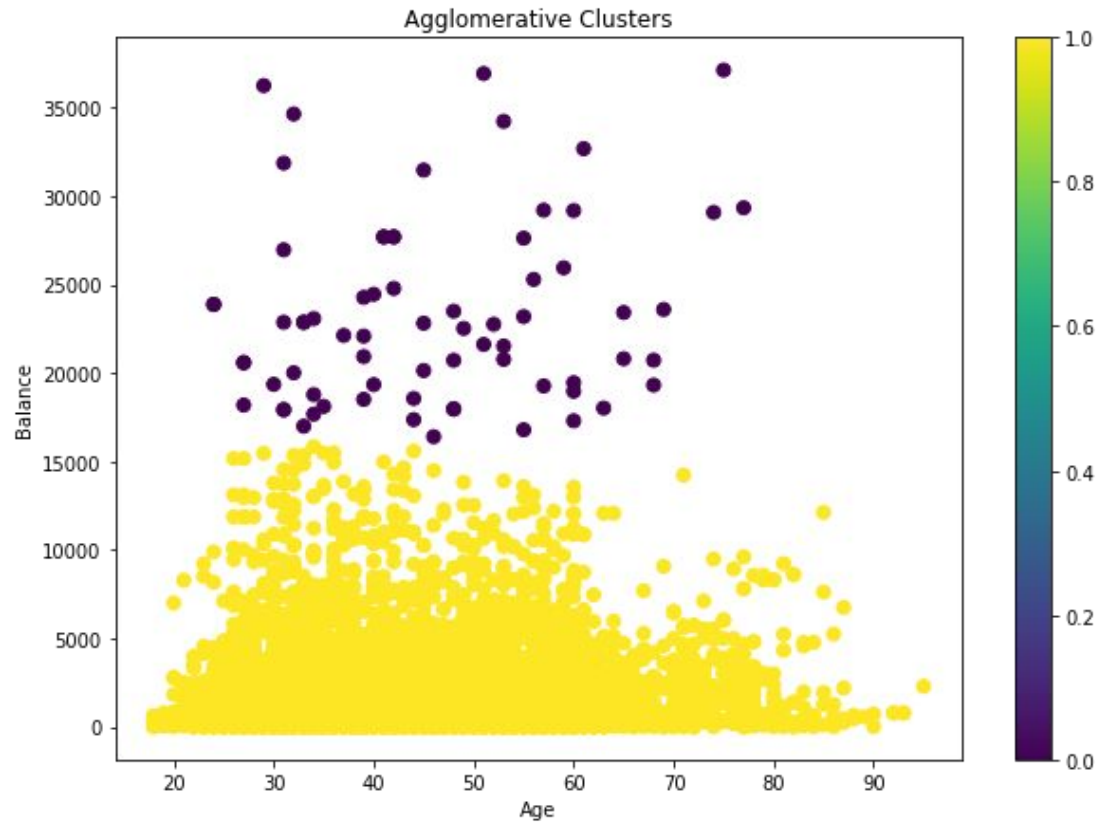


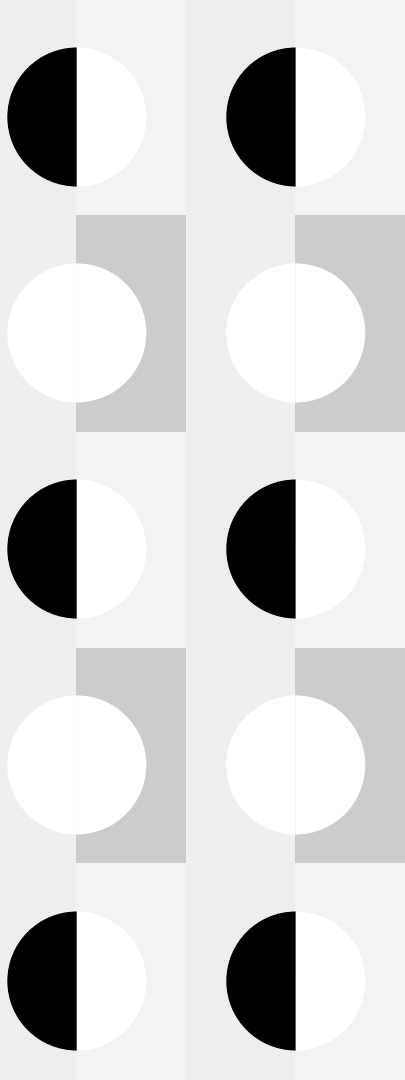
Unsupervised Learning



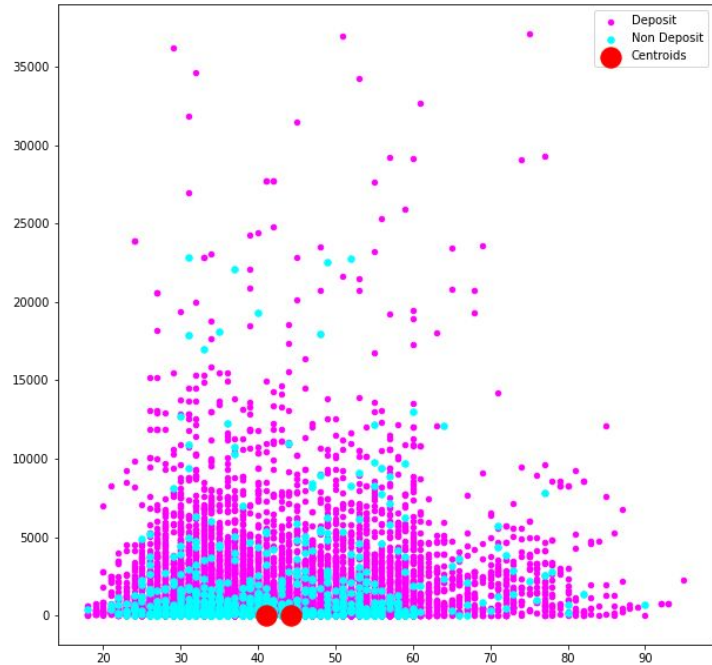
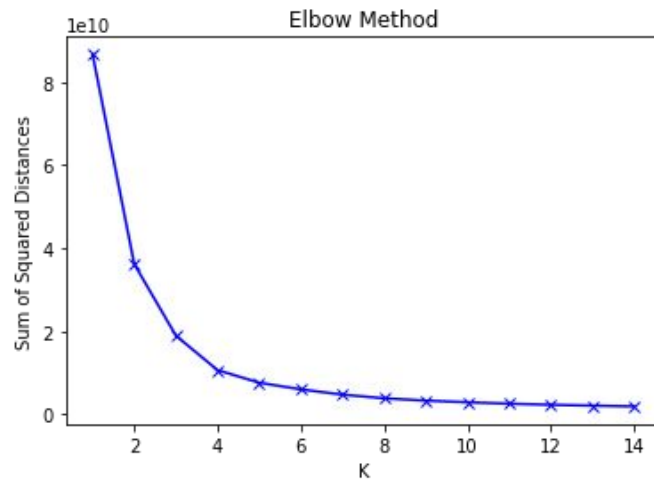


Agglomerative Clustering :



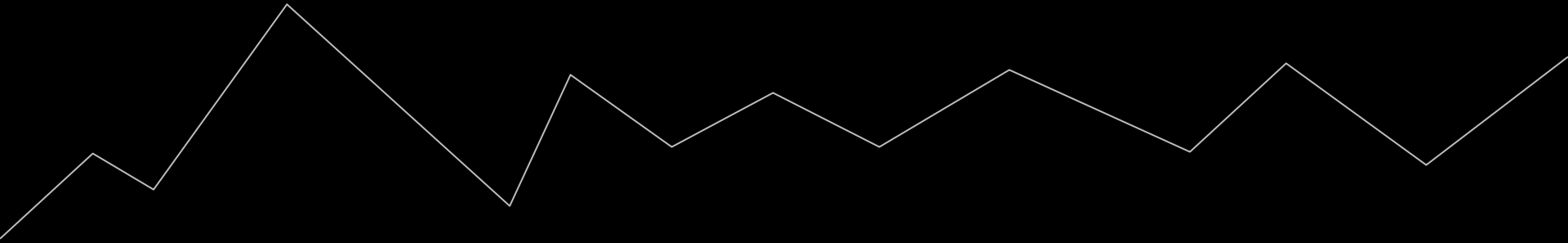


K-Means :

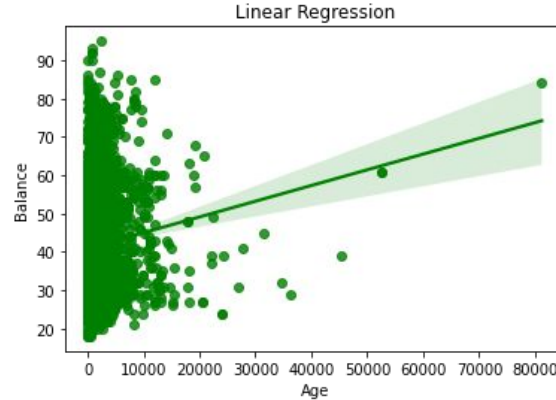


Linear Regression and Polynomial Regression

Predicting the age of the subscribers



Linear Regression : 0.9 % (Very Bad Model)



Polynomial Regression : 37% (Not Good Enough)

- MSE 110.61286492206503
- R Squared 0.3725496121655779



Multi-Linear Regression : 30%

MSE 122.73144088220644

R Squared 0.30380710927905497

Multi-Polynomial Regression : 33%

MSE 112.79487406259484

R Squared 0.328111495519408

**Thanks for your
attention**
