
SHARK tank

Agenda

SHARK TANK PITCH PREDICTION

1 - SharkTank introduction

2 - Exploratory Data Analysis

3 - Machine learning Models

4 - Recommendations

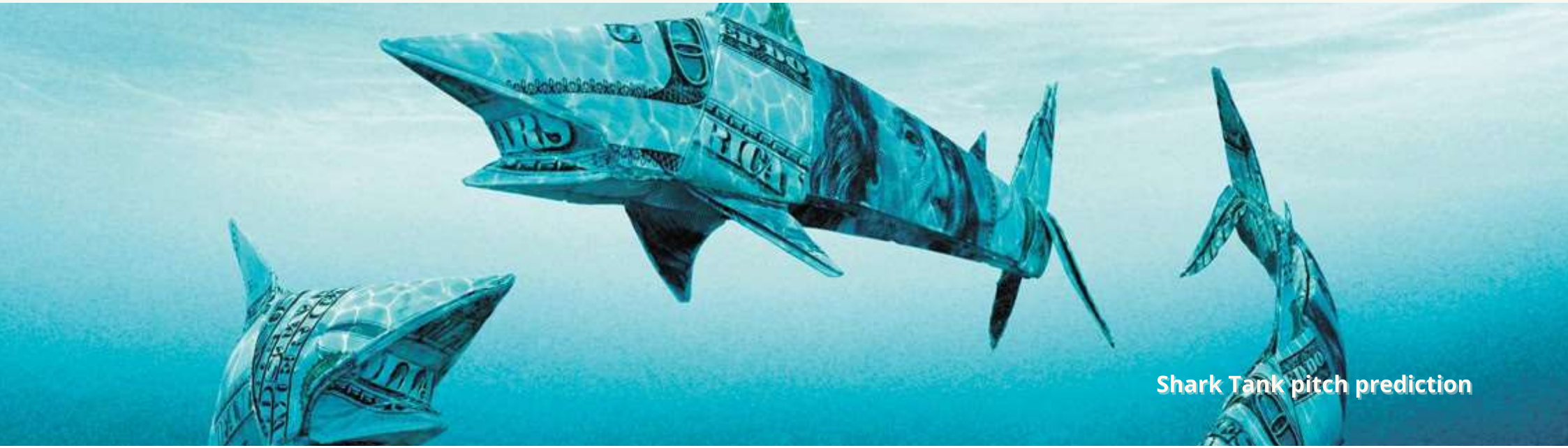
Introduction

ABOUT SHARK TANK

Shark Tank is a reality show for entrepreneurs. On the one hand, entrepreneurs who have a start-up project, or an existing business that they want to take to a new level. On the other hand, investors, all "self-made" entrepreneurs who have earned tens or hundreds of millions of dollars with their business and who are on the lookout for innovative concepts that have the power to change the game... and make them earn a lot of money.

Entrepreneurs present their project and try to convince at least one "shark" to invest in their company... in exchange for a percentage of the shares. They are often looking for the skills and network of the sharks as much as the money.

In this project, we'll build models to predict whether or not a pitch on the TV show Shark Tank made a deal with one or more investors on the show.



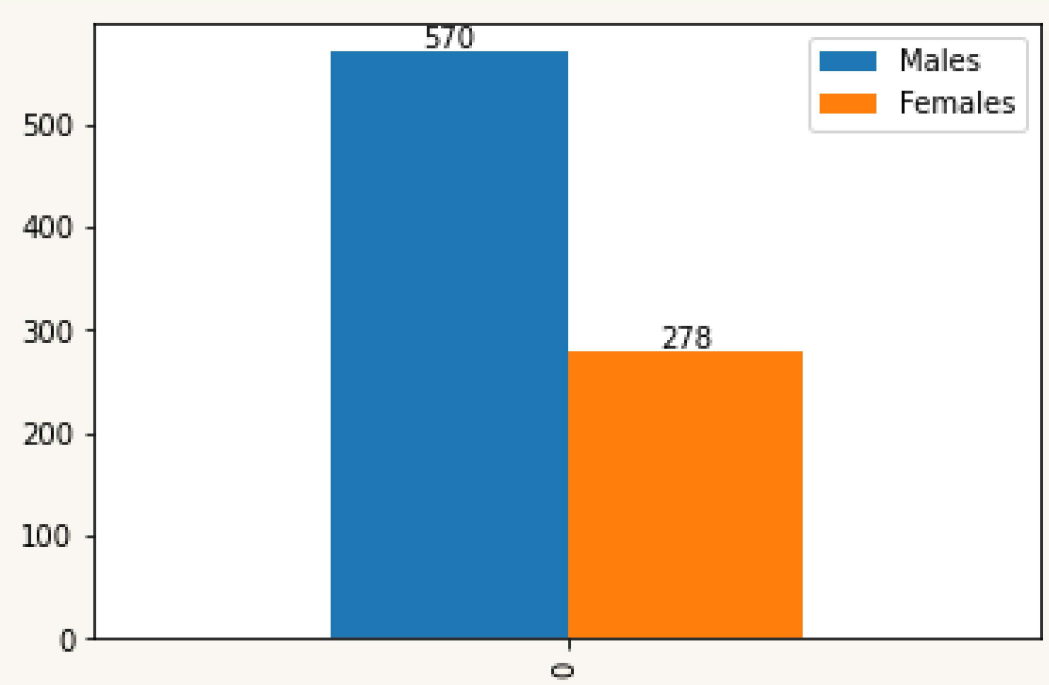
Shark Tank pitch prediction

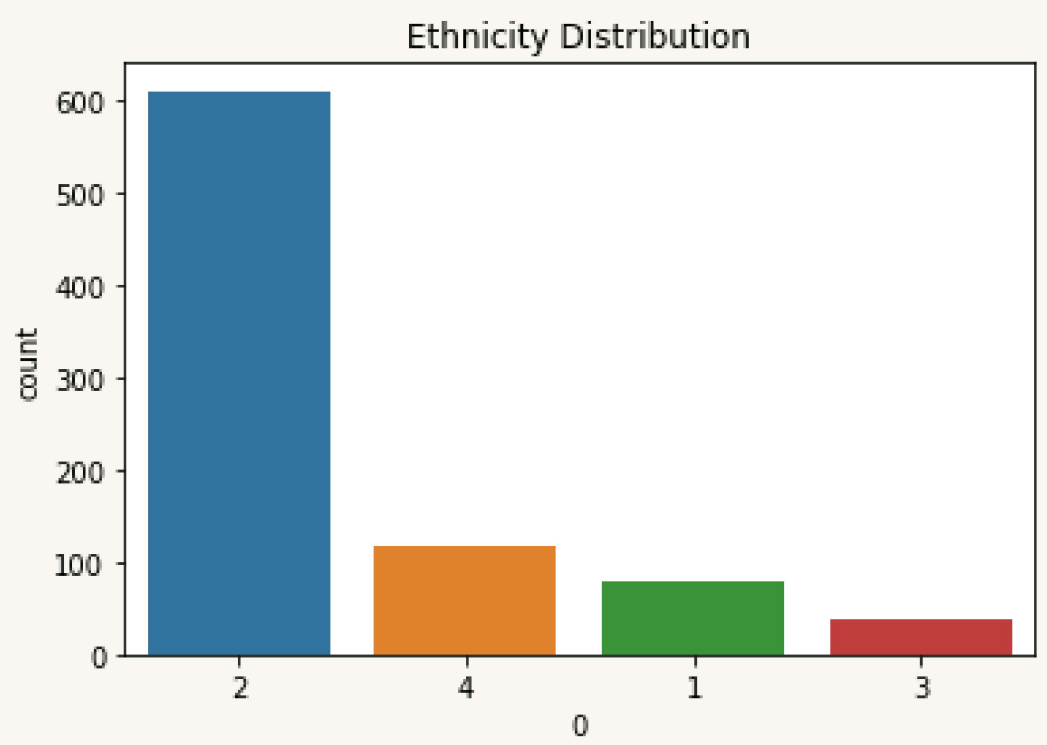
SHARK TANK PITCH PREDICTION

Exploratory Data Analysis

Gender and Ethnies Distribution

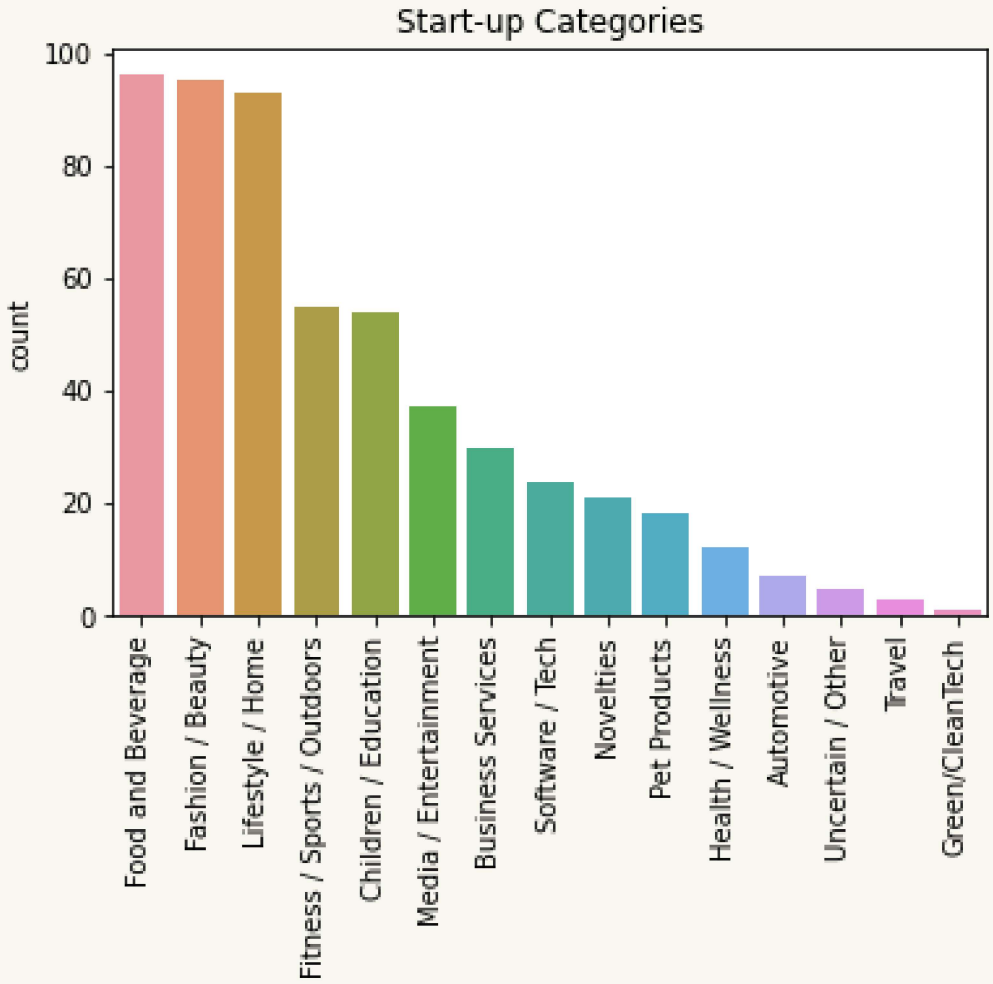
- As we can see, there are almost twice as many entrepreneurs who are men than women
- The 'ethnic group 2' is the majority among the entrepreneurs

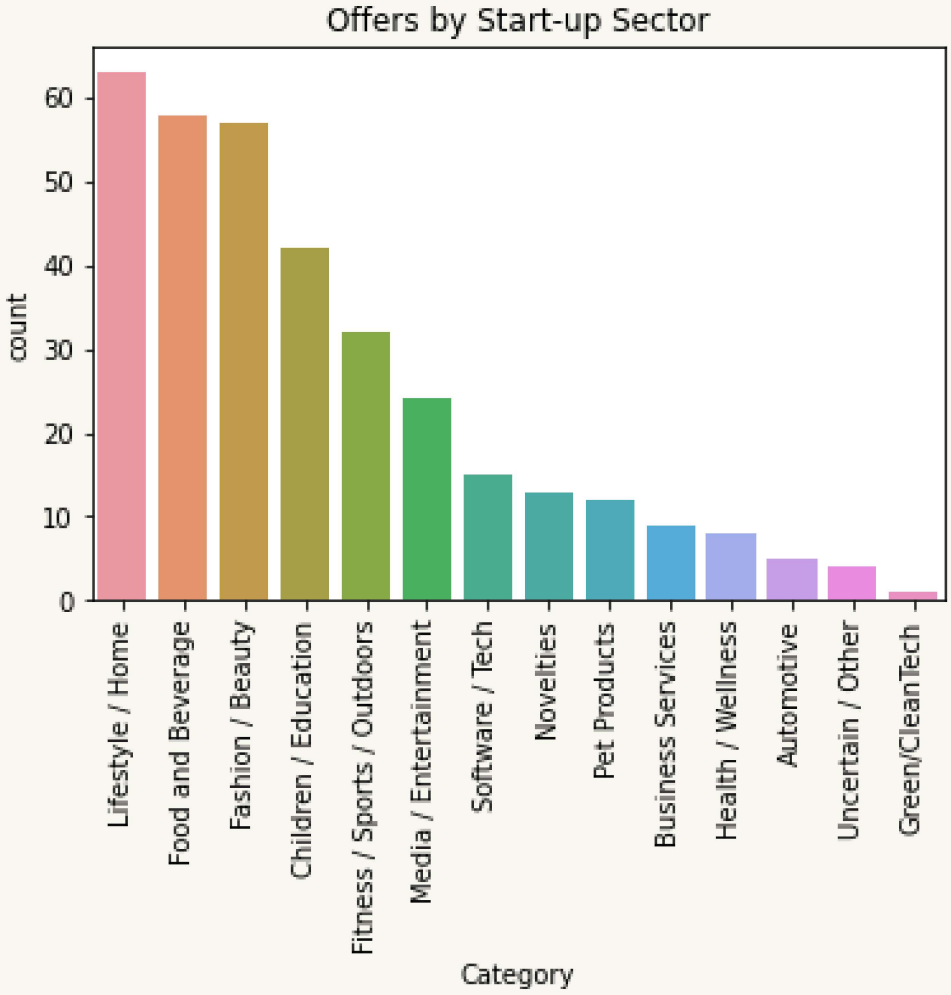




Start-Up Industries

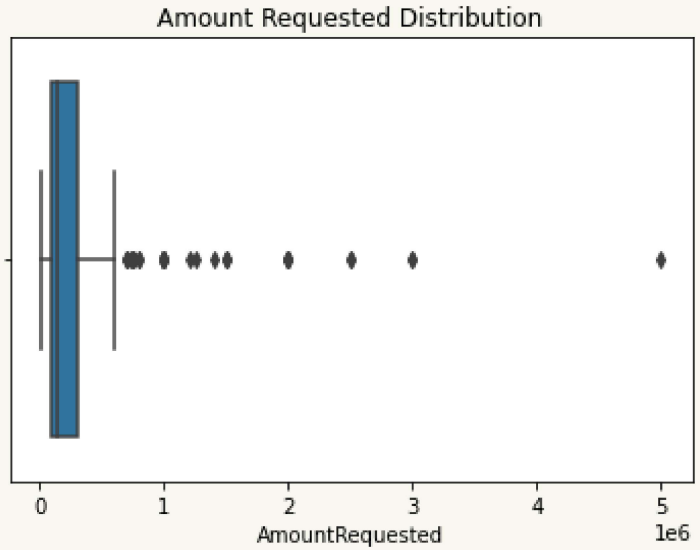
- Most of the start-ups are in the 'food and beverage', 'fashion/beauty' and 'lifestyle/home industries'

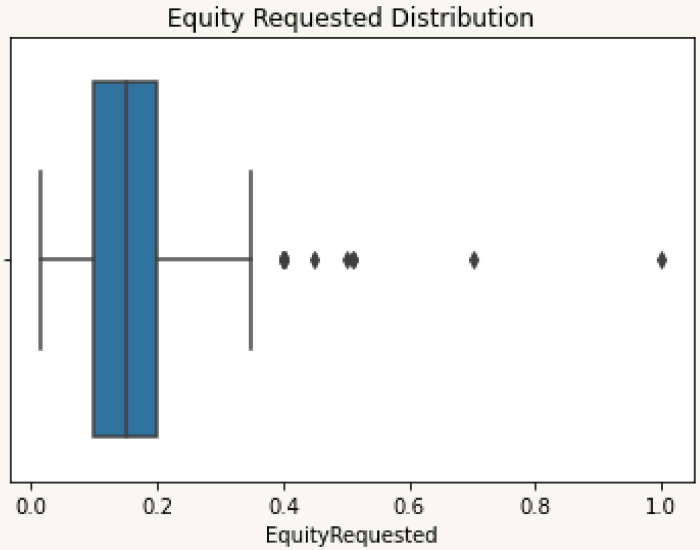


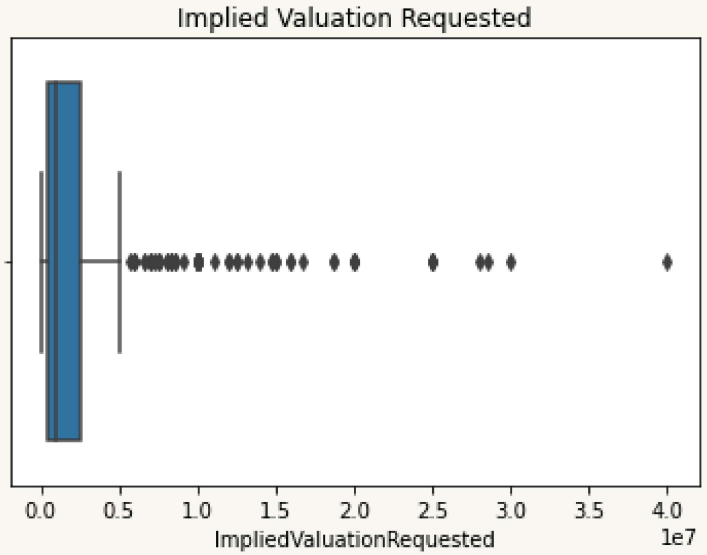


Start-ups 'Amount request', 'Equity request' and 'Implied valuation Analysis'

As we can see, most of the start-ups request approximately the same amount and the same equity.
Furthermore, their implied valuation is similar.
(Even though there are several outliers)

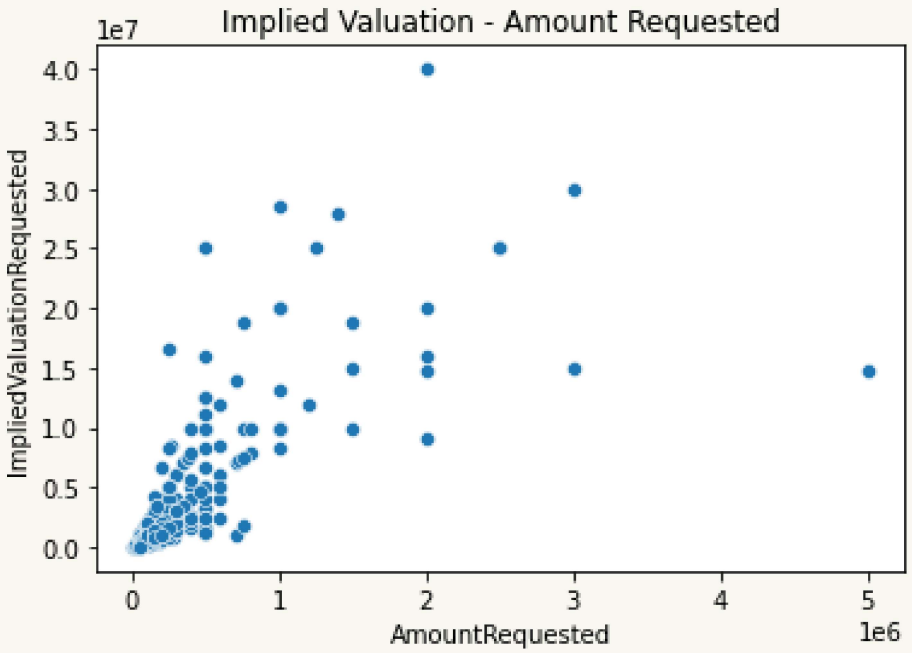


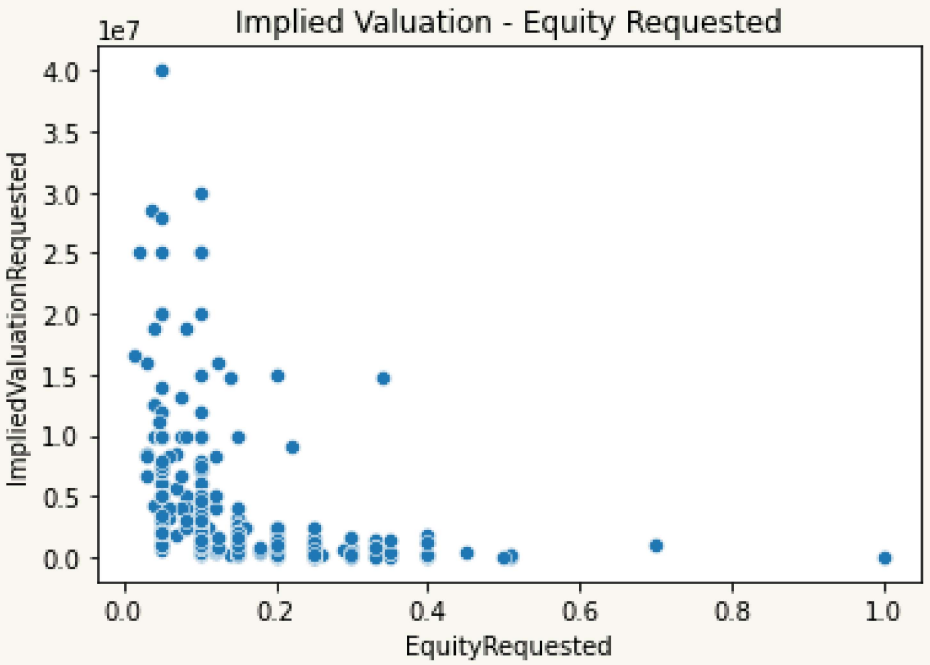




'Implied Valuation' and 'Amount Requested' / 'Equity Requested Analysis'

We can see that the amount requested by start up is linearly correlated with their implied valuation.
However, the equity requested by start ups is not correlated with their implied valuation.





SHARK TANK PITCH PREDICTION

Machine learning Models

ML Model Analysis

LOGISTIC REGRESSION
PREDICTION SCORE

Besides the accuracy of 41% being lower than a coinflip. We note that the recall, which tells us how many of the actual positive cases we were able to predict correctly with our model, is significantly low for the class 1 (Deal struck)

Logistic Regression :

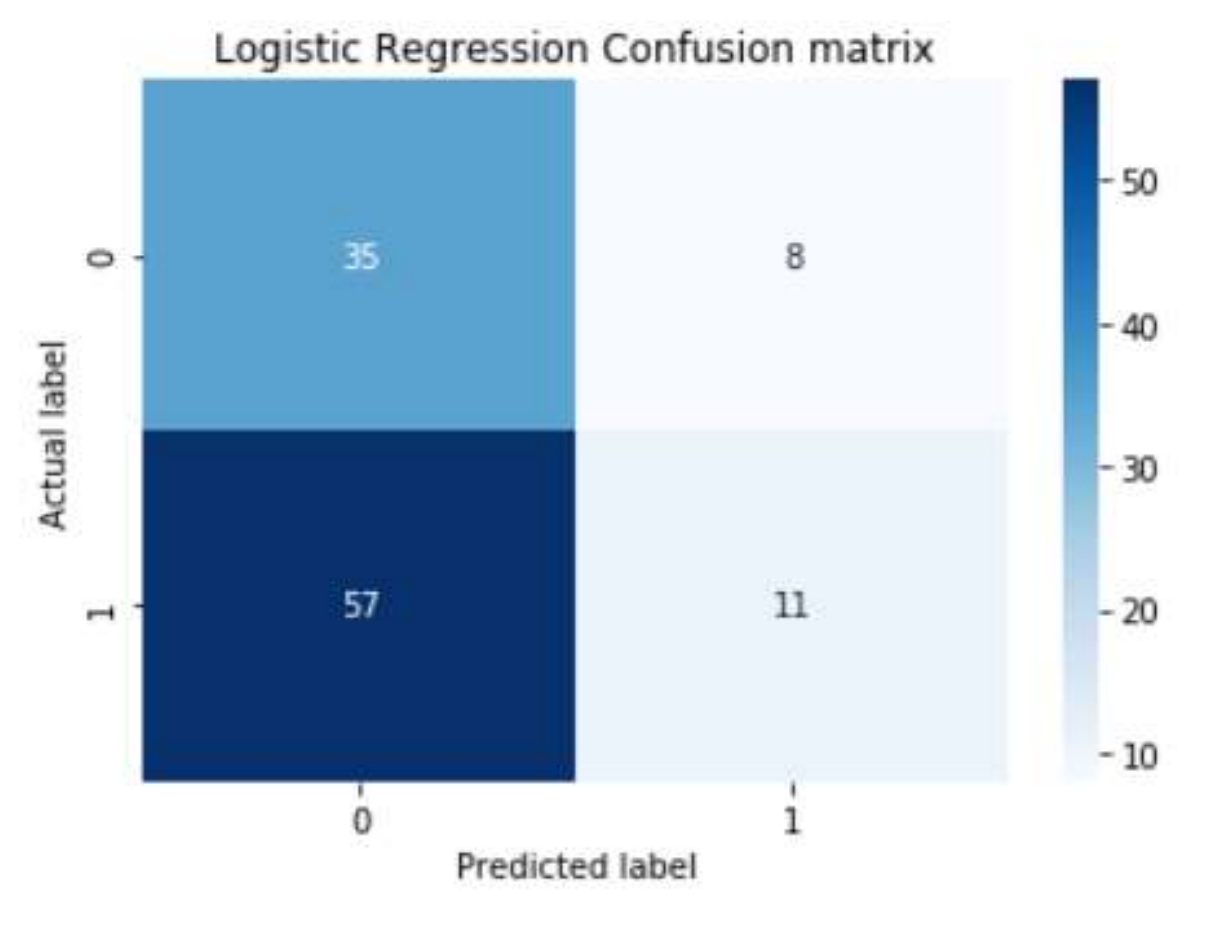
	precision	recall	f1-score	support
0	0.38	0.81	0.52	43
1	0.58	0.16	0.25	68
accuracy			0.41	111
macro avg	0.48	0.49	0.39	111
weighted avg	0.50	0.41	0.36	111

Score : 41.44144144144144

ML Model Analysis

LOGISTIC REGRESSION
CONFUSION MATRIX

The matrix shows indeed that **Type 2 error** is an issue with the logistic regression.
False negatives are too present, when the actual value is positive the model predicts it as negative.



ML Model Analysis

K-NEAREST NEIGHBORS
PREDICTION SCORE

This model seems better even though the **accuracy** of 55% is still **low**.

K-Nearest Neighbors :

	precision	recall	f1-score	support
0	0.43	0.53	0.48	43
1	0.66	0.56	0.60	68
accuracy			0.55	111
macro avg	0.54	0.55	0.54	111
weighted avg	0.57	0.55	0.56	111

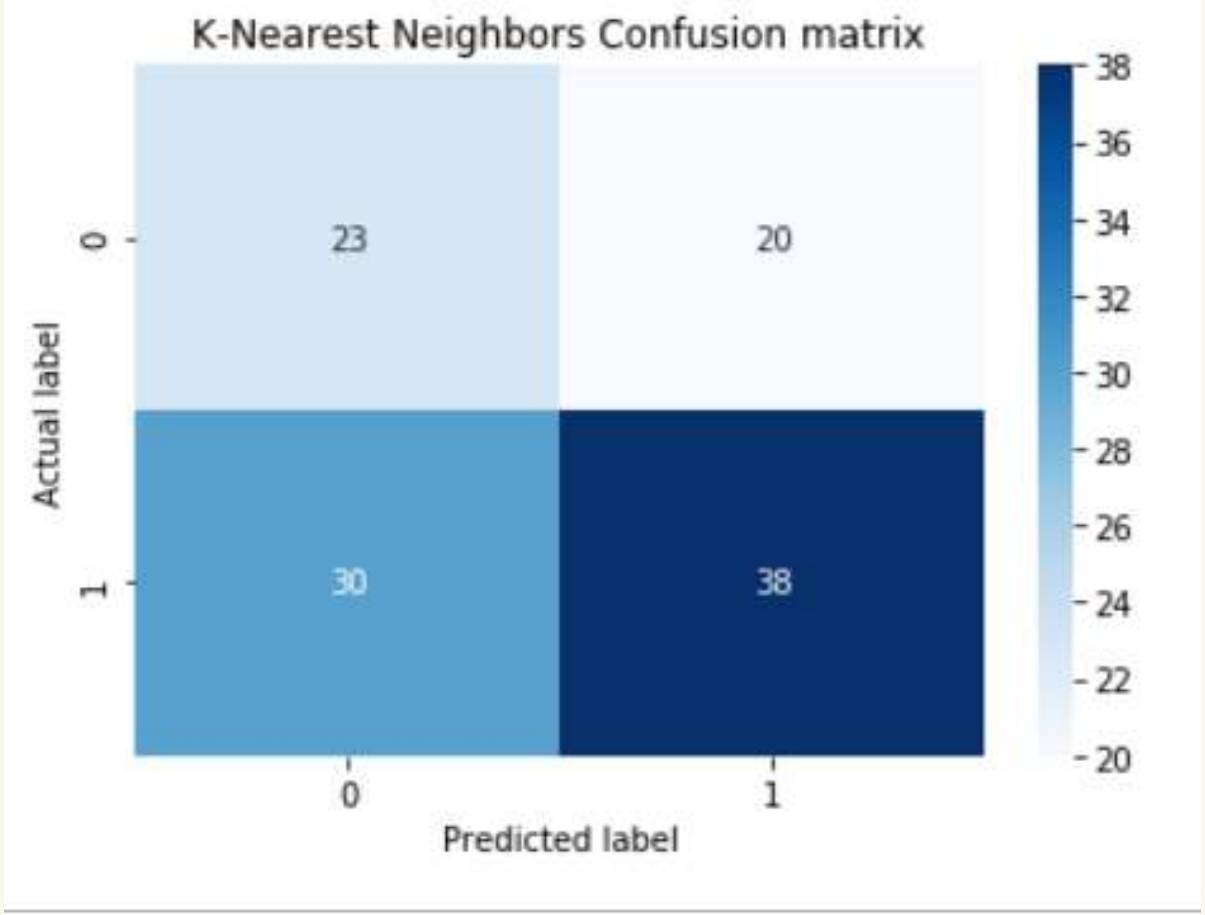
Score : 54.95495495495496

ML Model Analysis

K-NEAREST NEIGHBORS
CONFUSION MATRIX

We still face the issue of **Type 2 error** although not as heavy as with the logistic regression.

However, this model cannot be considered satisfactory due to its prediction being closer to a coin toss



ML Model Analysis

MLP

PREDICTION SCORE

The accuracy of 61% is admittedly better but we face the issue of a low recall (21%) on the class 0 this time (Deal not struck)

The precision, which tells us how many of the correctly predicted cases actually turned out to be positive, is also too low on the class 0

MLP :

	precision	recall	f1-score	support
0	0.39	0.21	0.27	43
1	0.61	0.79	0.69	68
accuracy			0.57	111
macro avg	0.50	0.50	0.48	111
weighted avg	0.53	0.57	0.53	111

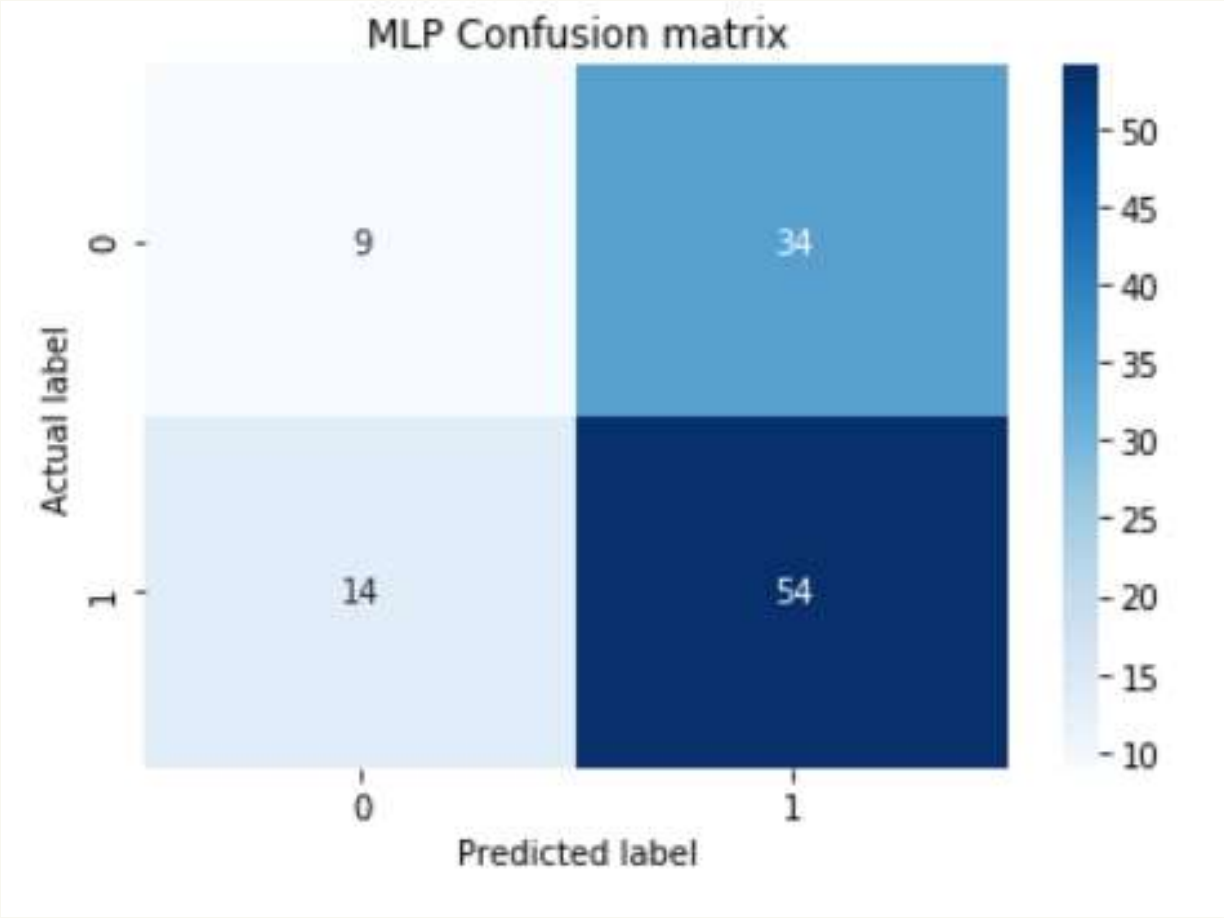
Score : 61.261261261261254

ML Model Analysis

MLP

CONFUSION MATRIX

The matrix shows indeed that **Type 1 error** is an issue. **False positives** are too present, when the actual value is negative the model predicts it as positive.




Fun fact : Unlike with confusion matrices, exposure to underwater settings have a positive impact on people's well-being
(reductions in blood pressure and heart rate).
Take a break and look at some sharks.

ML Model Analysis

DECISION TREE

PREDICTION SCORE

The **accuracy** of **90%** is satisfactory and the other scores do not raise any red flags

Decision Tree :

	precision	recall	f1-score	support
0	0.82	0.95	0.88	43
1	0.97	0.87	0.91	68
accuracy			0.90	111
macro avg	0.89	0.91	0.90	111
weighted avg	0.91	0.90	0.90	111

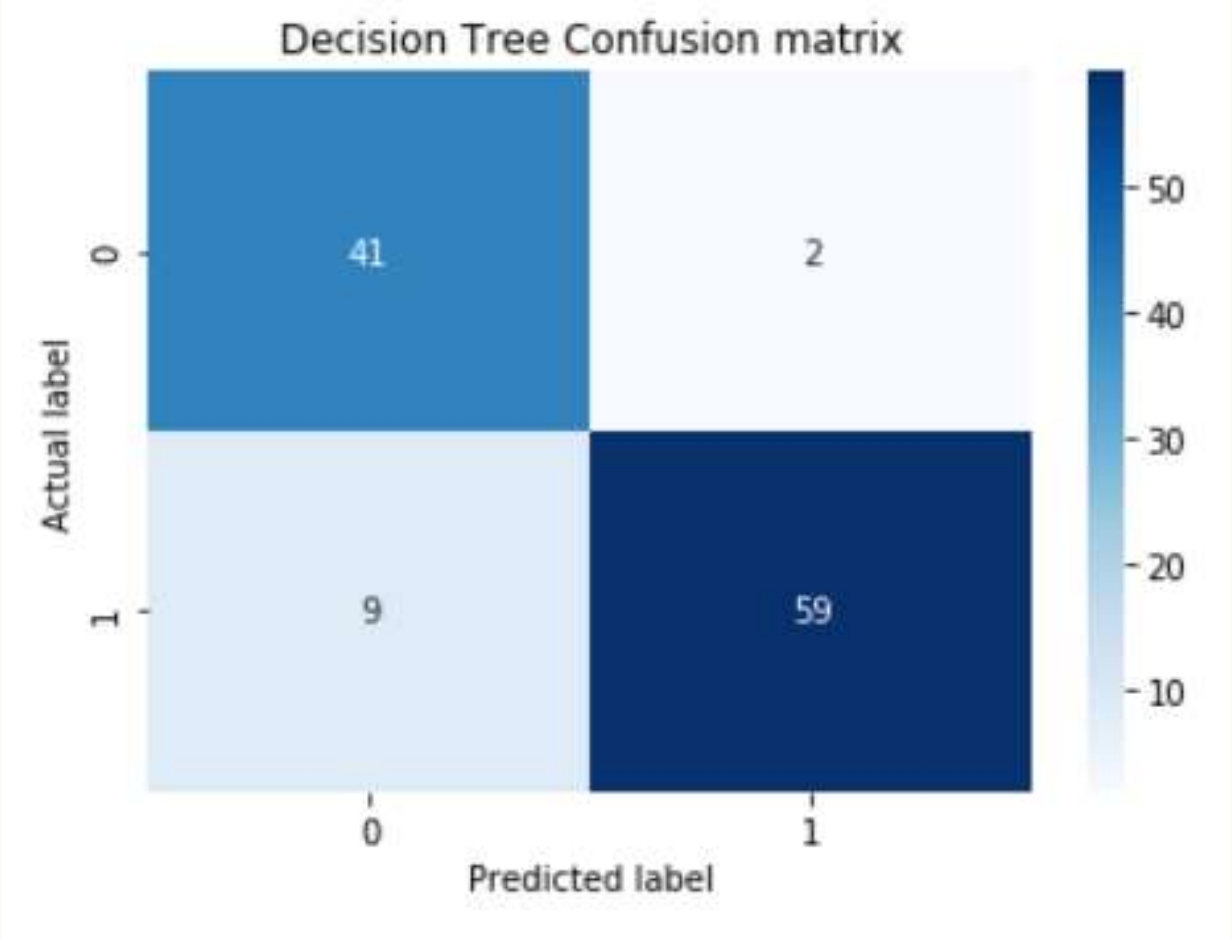
Score : 90.09009009009009

ML Model Analysis

DECISION TREE

CONFUSION MATRIX

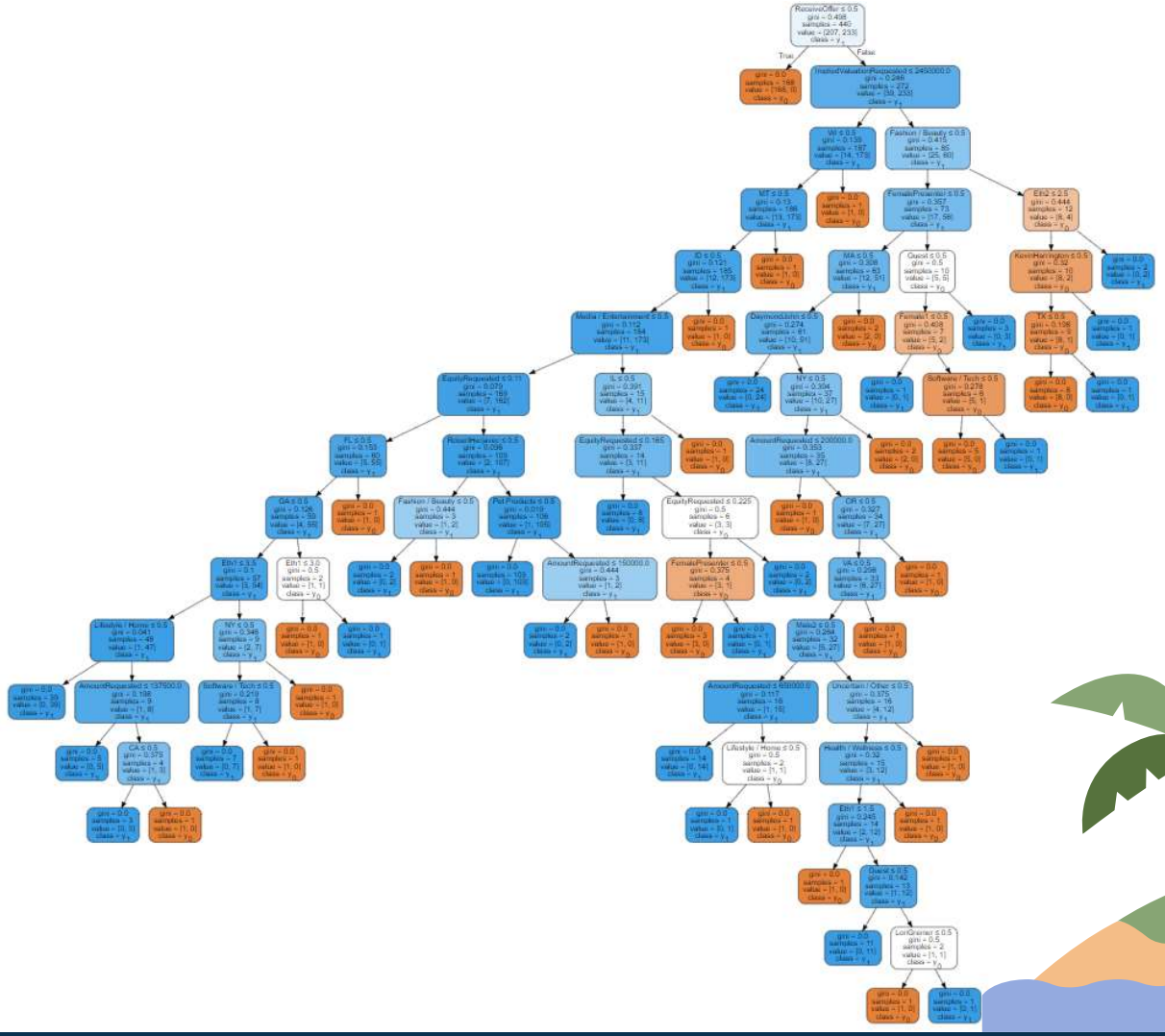
This model finally succeeds in predicting **True negatives** and **True positives.**



ML Model Analysis

DECISION TREE

See code for a clearer decision tree + Hyperparameter tuning of this model.



ML Model Analysis

RANDOM FOREST

PREDICTION SCORE

Same conclusions as for the decision tree model but we slightly better results of 93,7%

Random Forest :

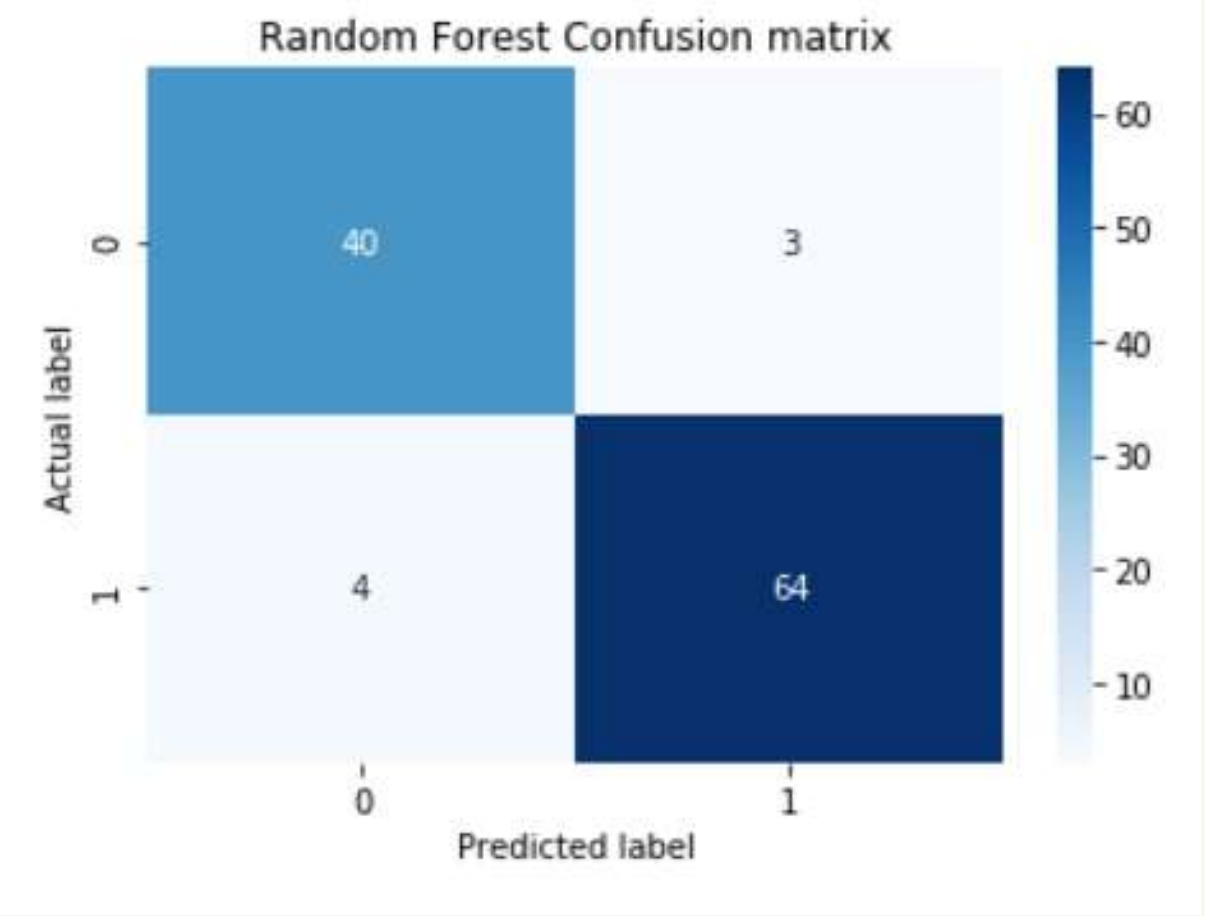
	precision	recall	f1-score	support
0	0.91	0.93	0.92	43
1	0.96	0.94	0.95	68
accuracy			0.94	111
macro avg	0.93	0.94	0.93	111
weighted avg	0.94	0.94	0.94	111

Score : 93.69369369369369

ML Model Analysis

RANDOM FOREST
CONFUSION MATRIX

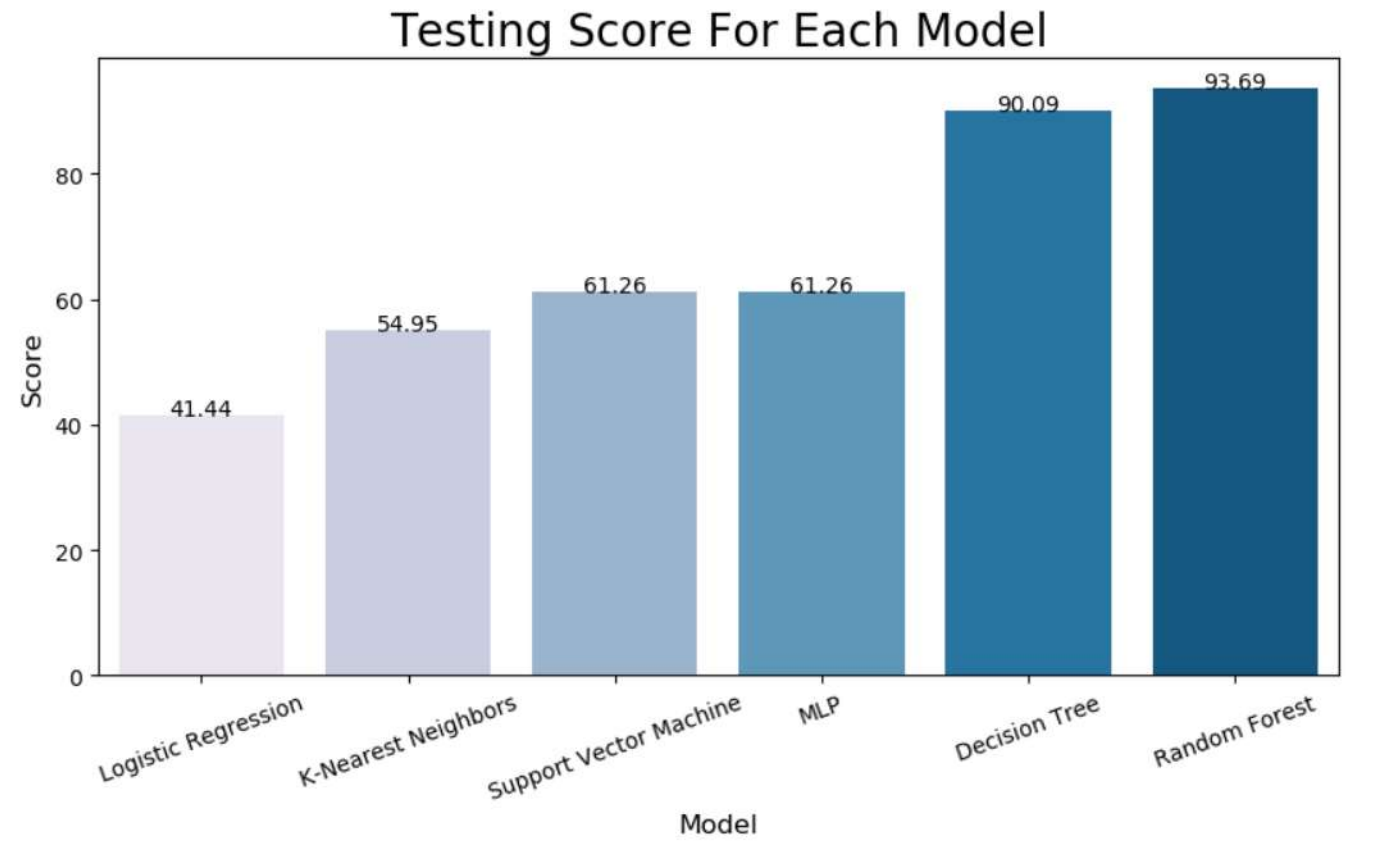
Same conclusions as for the decision tree model but we slightly better results.



Results

MODELS COMPARAISON

ML MODELS ACCURACIES



SHARK TANK PITCH PREDICTION

Recommendations

Thank You