



MWN CONSULTANCY



WWW.MWN.COM

TANZANIA WATER WELLS CLASSIFICATI ON





TABLE OF CONTENTS

01

OVERVIEW

03

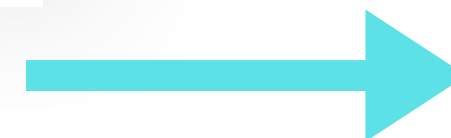
**METHODOLOGY
& RESULTS**

02

OBJECTIVES

04

EVALUATION





01

OVERVIEW

Water is a basic need for humans for health and sanitation. In remote areas, water wells become the source of this requirement. For a population of over 65 million people in Tanzania, being able to service and maintain this wells becomes paramount and therefore fulfils the sustainable development goal number 6: Clean Water and Sanitation

MWN Consultancy has been tasked to predict the condition of a well through the data provided. Through modeling, we should come up with the best model that classifies pumps into 3 categories; functional, non-functional, requires repair.

This model should be able to categorise the current installed base but also declare combined characteristics and features of attributes that are precursors to the conditions of the wells





**MWN
CONSULTANCY**



WELCOME TO PRESENTATION

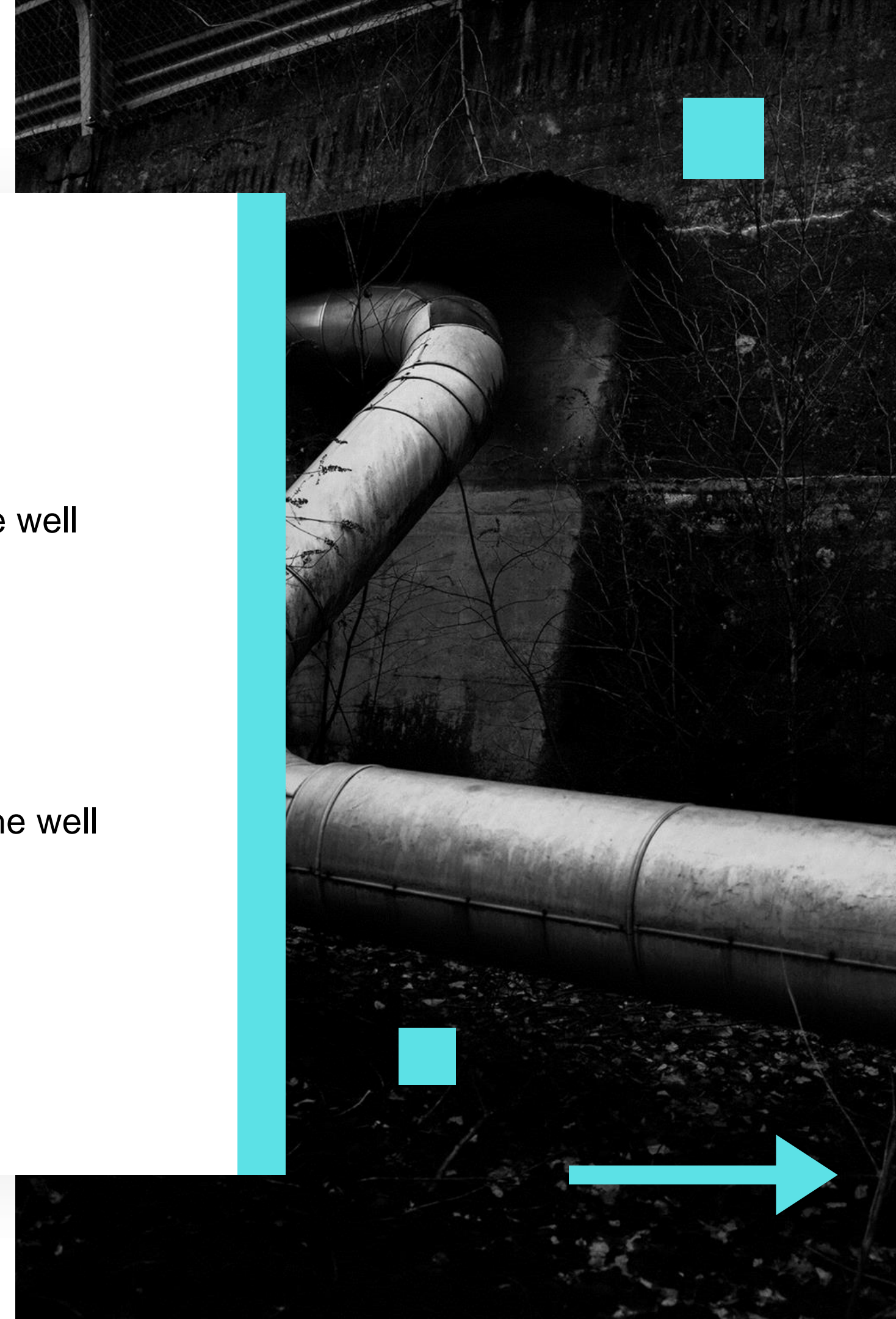
I'm Maureen, and I'll be sharing with you the outcome of the process and outcome of the classification.

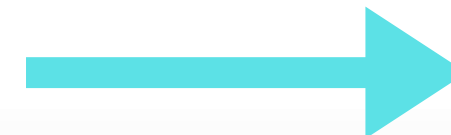


02

OBJECTIVES

1. Identify and present the best model that classifies the state of the well with highest accuracy
 - This will be the model used to identify the status of the well to aid in maintenance, repair and status update
1. 2. Identify which variables affect the state of the well
 - This will be the common features that will determine the status of the well and would be closely monitored or noted.





03

METHODOLOGY AND RESULTS

The dataset was imported and cleaned through removing null values and eliminating duplicates as well as unnecessary data

Encoding was done to enable classification of the variables which would help in data processing models such as Logistic regression, K Nearest neighbor and Support Vector Machine.

Scaling was also done to support the above models i.e. after identifying the skewness and p

The data was then passed through 6 models being ;

1. Logistic regression
2. K Nearest neighbor
3. Support Vector Machine.
4. Random Forest
5. Decision Trees
6. Gradient Boost



RESULTS

The models Logistic Regression, KNN, Support Vector Machine, had the least performance with below 50% levels on precision, recall and f1-score but with improved accuracy respectively.

Accuracy however is not a very reliable measure. The models performed poorly even after tuning because they are best suited for simpler and smaller datasets





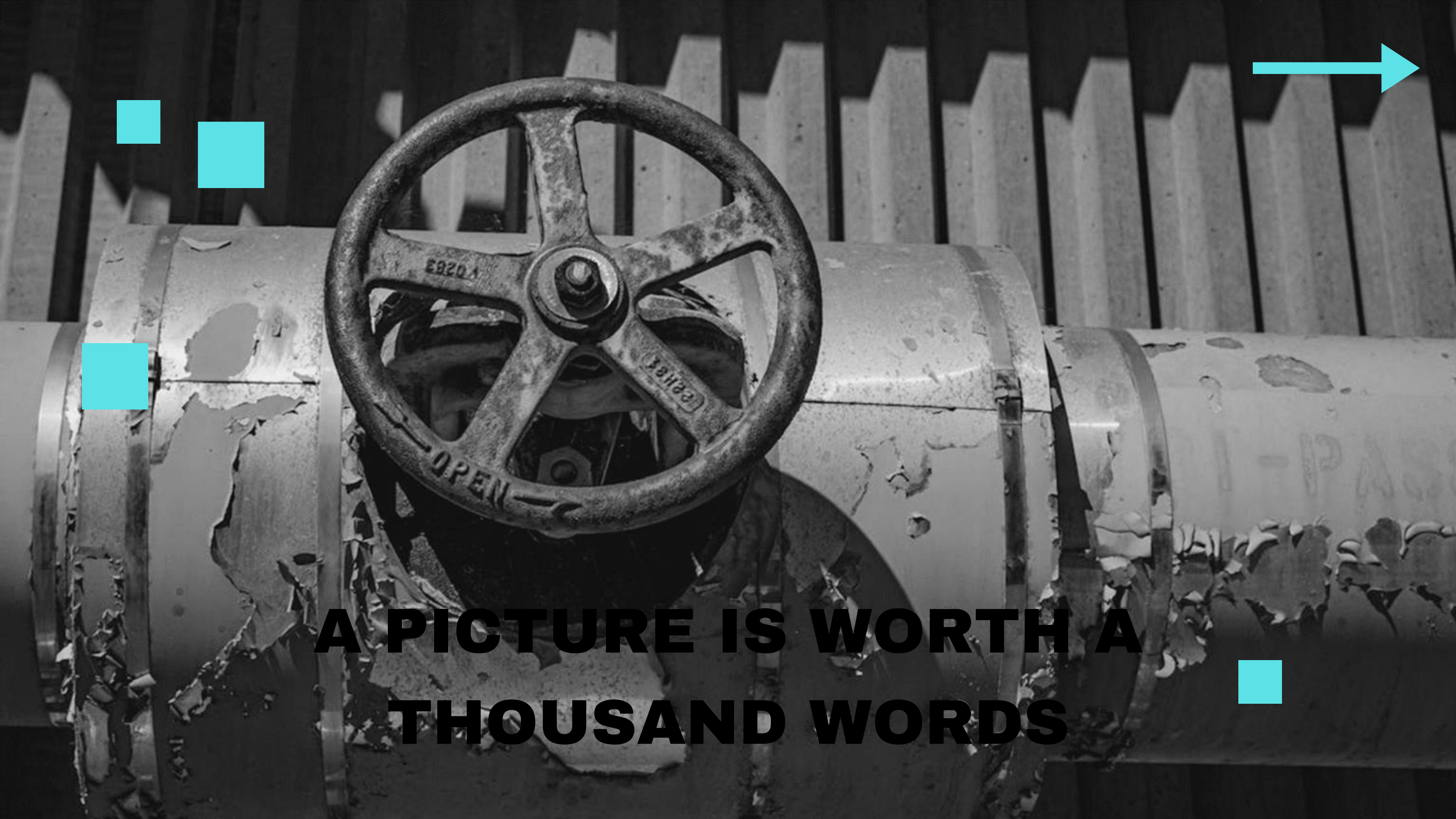
RESULTS

forming model was the tuned Random forest model with 75 % recall on class 1 which is the functional needs repair.Focus on this
to increase repairs on faulty wells by early detection.Better tuning will increase the overall performance.All other models had very
ce on accuracy and recall and will not be best suited for this dataset

result for Random Forest metrics

Report:

ision	recall	f1-score	support
79	0.74	0.76	5303
23	0.75	0.35	700
90	0.59	0.71	3654
		0.68	9657
0.64	0.69	0.61	9657
0.79	0.68	0.71	9657



**A PICTURE IS WORTH A
THOUSAND WORDS**



**MWN
CONSULTANCY**



**A BIRD IN HAND IS
WORTH TWO IN THE
BUSH**





MWN
CONSULTANCY

CONTACT US



Imara Daima, Itula Close



254705855005



ngahumaureen0@gmail.com



The slide features a white central rectangle with a subtle drop shadow, set against a light gray background. Several teal-colored squares of varying sizes are scattered around the white rectangle: one at the top right, one on the left side, one on the right side, one at the bottom left, and one at the bottom right.

**THANK YOU
QUESTIONS?**