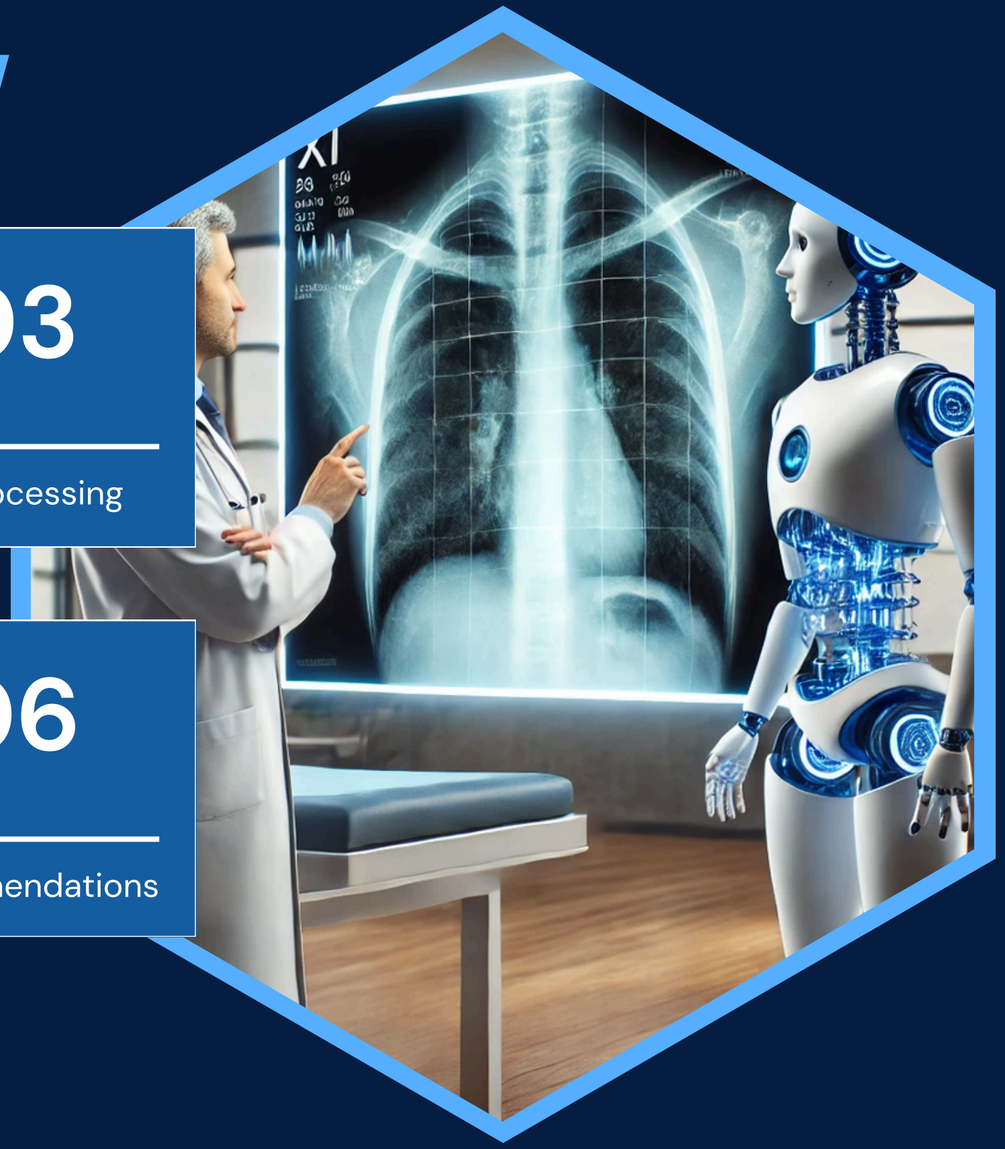
DIAGNOSING PNUEMONIA USING AI

CAPSTONE 3

Francisco J Torres

OVERVIEW

01	02	03
The Data	EDA	Preprocessing
04	05	06
Modeling	Evaluation	Recommendations



GOALS AND OBJECTIVES

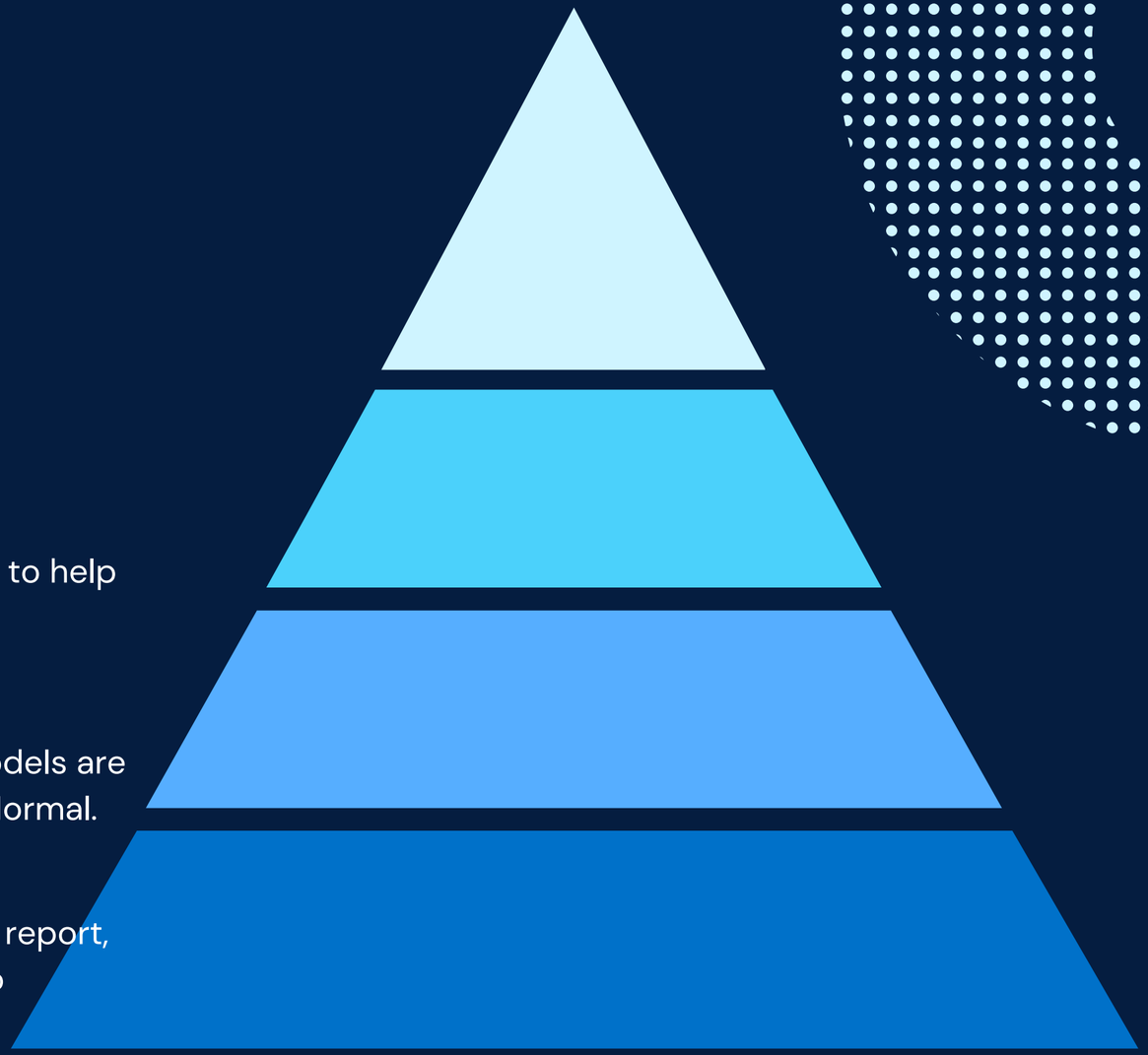
Goals

The goal of this project is to build a Convolutional Neural Network that can be used to aid radiologists in diagnosing patients with Pneumonia.

Target

In order to be consisdered useful I want the model to achieve at least a Recall score of 80%.

CONTENT



THE DATASET

- This data set is available on Kaggle.com. The data set includes over 5,000 X-ray images. It is split into training images, testing images and validation imges.
- The dataset has images of healthy lungs and images of lungs with Pneumonia from bacterial and viral infections.

1. LOAD DATA AND INSTALL NECESSARY LIBRARIES

LIBRARIES:

- MATPLOTLIB
- SEABORN
- KERAS
- CV2
- OS
- NUMPY
- PANDAS
- TENSORFLOW

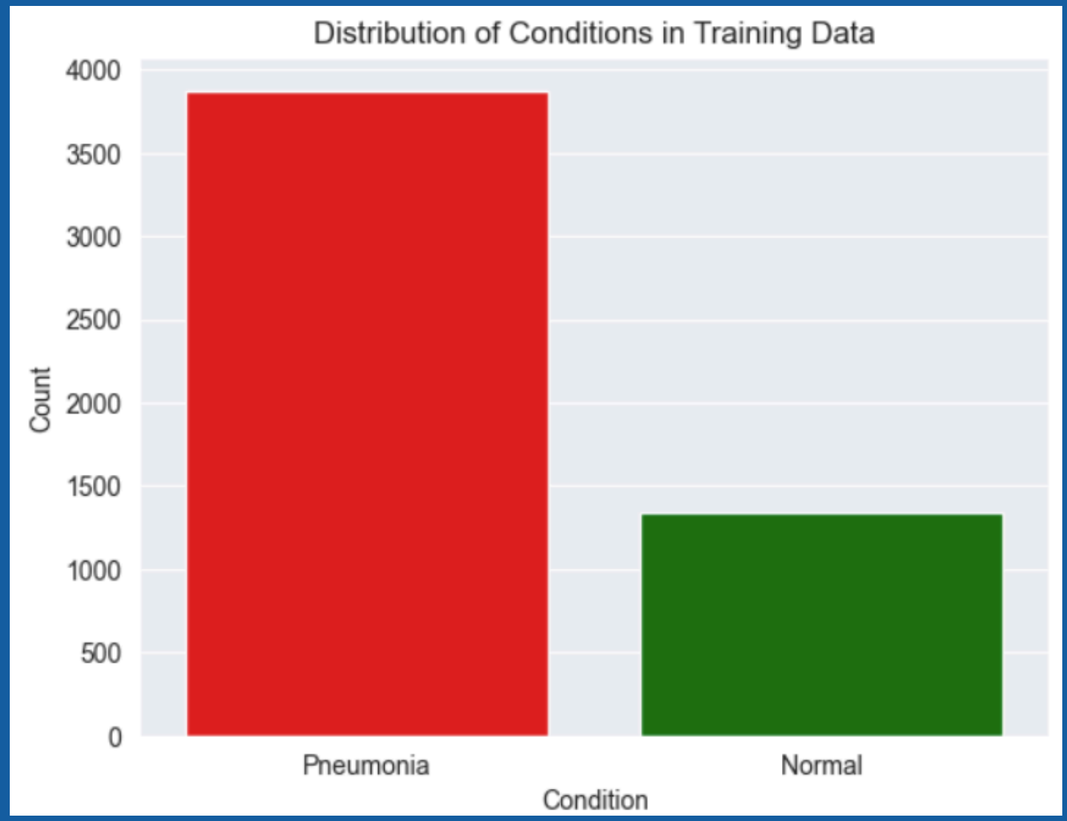
LOADING DATA:

During this part of the project I experience an issue.
This was my first time loading data into a notebook that is visual data (images).
I had to do some exploration and look for resources online.
After some time I learned that I could write a python function (**get_training_data**) that helped me load my X-ray images.

2. EXPLORATORY DATA ANALYSIS

LOADING DATA:

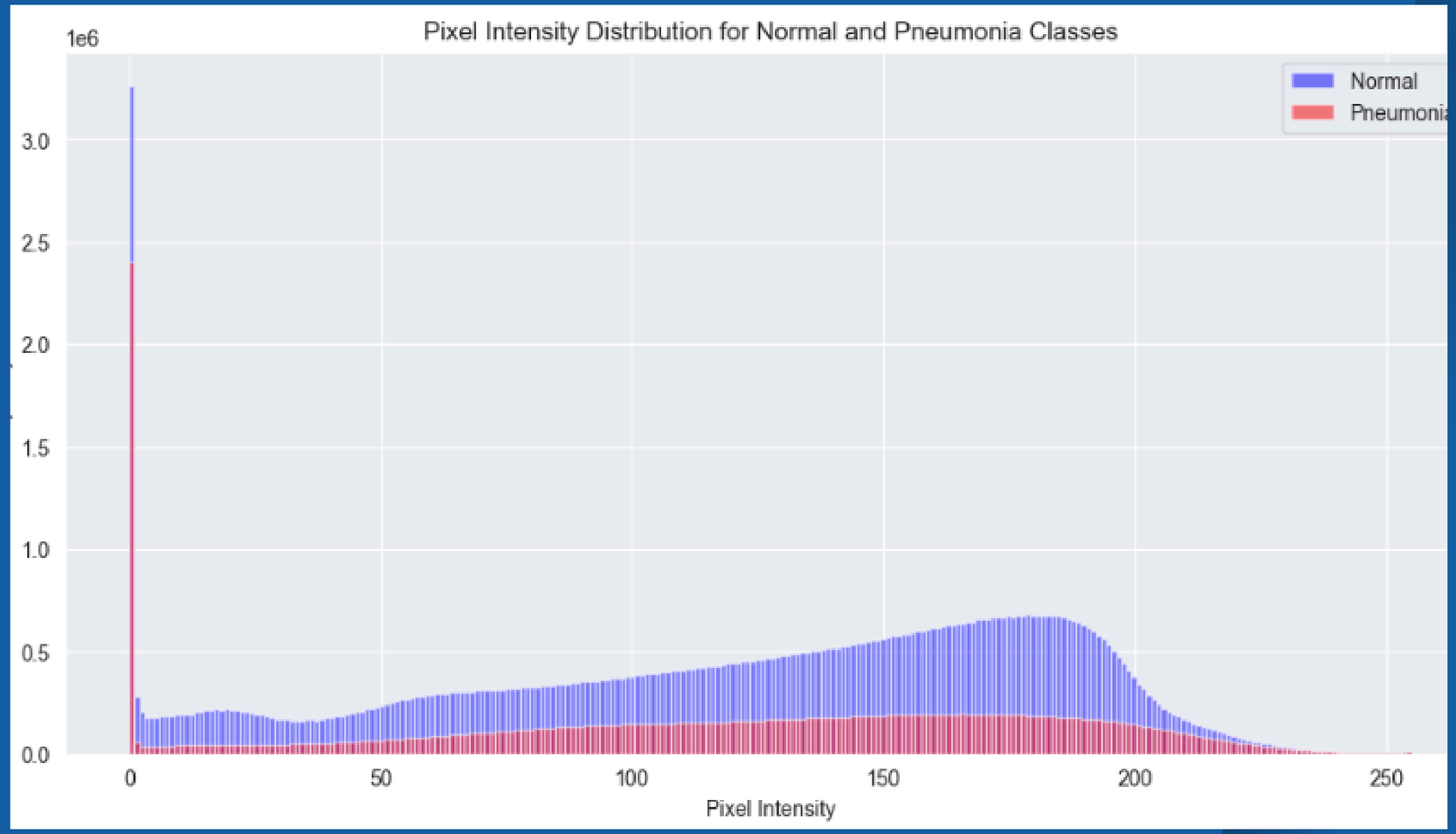
- Once I succefully loaded the images into my notebook I began EDA.
- During EDA I discovered that the data set was unbalanced and the majority of images were of lungs with Pneumonia..
- This was an imbalanced dataset that could lead to overfitting.
- EDA helped me keep this in mind during model training.
- Training data was 5216 images.
- Testing data: 624 imgages.



2. EXPLORATORY DATA ANALYSIS

EDA CONTINUED:

During EDA I also managed to discover that the pixel intensity of the images with health lungs were on average greater than the images with Pneumonia.
This seemed to be an important feature that would have to be dealt with during preprocessing.



2. EXPLORATORY DATA ANALYSIS

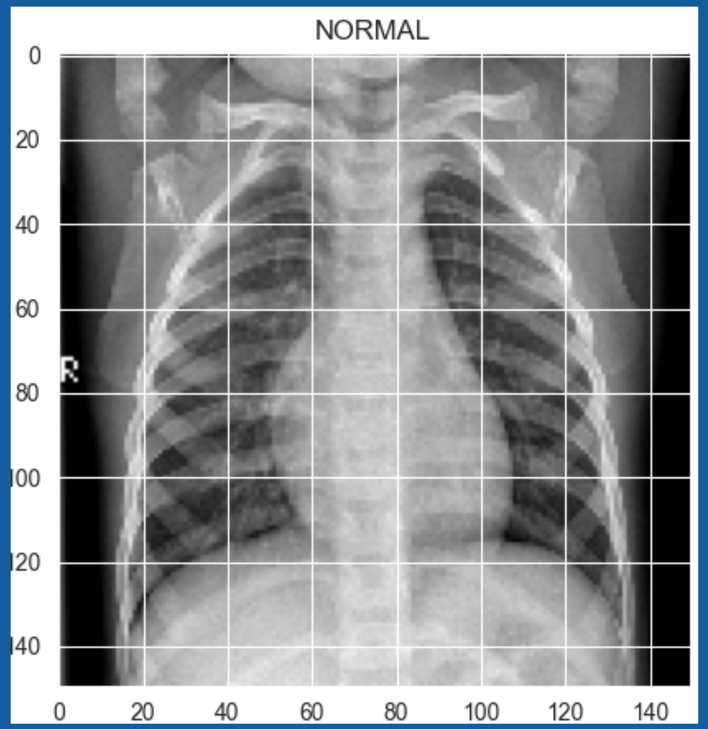
COMPARING IMAGES BY EYE

Exploring intensity of Normal vs. Pneumonia images.

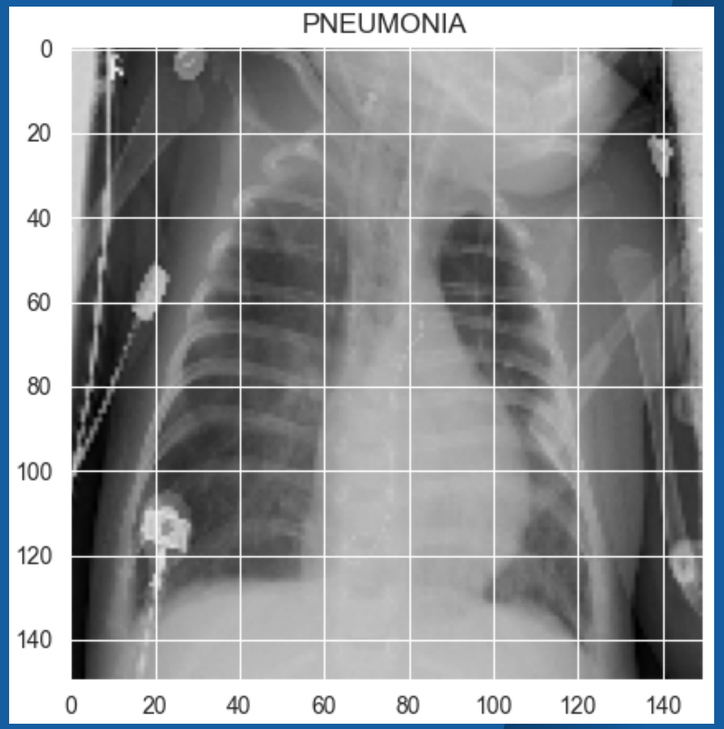
Normal image appears to be brighter and image quality is sharper.

Pneumonia image is more opaque and not as sharp.

NORMAL



PNEUMONIA



3. PREPROCESSING

DATA AUGMENTATION:

Important tool:
ImageDataGenerator.
I was able to utilize this tool to

- rescale: normalizes pixel values.
- rotating_range: Randomly rotates images up to 20% to help make model more robust.
- width_shift_range: Helps increase robustness which helps in generalization.
- height_shift_range: Allows vertical shift which also increases models ability to generalize.
- zoom_range: Helps model generalize with medical images since scale may vary.

```
train_datagen = ImageDataGenerator(rescale=1/255.,
                                   rotation_range=0.2,
                                   width_shift_range=0.2,
                                   height_shift_range=0.2,
                                   zoom_range=0.2,
                                   horizontal_flip=True
                                   )
```

4. MODELING

MODEL 1

Convolutional Neural Network using Keras. I used Sequential model which adds layers one after the other. 9 layers were added which included filtering layers and max pooling layers. The activation functions used were ReLU and sigmoid.
The model was trained for 20 epochs.

MODEL 1 LAYERS:

1. Conv2D - Layer 1
2. MaxPooling2D - Layer 2
3. Conv2D - Layer 3
4. MaxPooling2D - Layer 4
5. Conv2D - Layer 5
6. MaxPooling2D - Layer 6
7. Flatten - Layer 7
8. Dense (100 neurons) - Layer 8
9. Dense (1 neuron) - Layer 9

MODEL 2

Model 2 was built as an upgrade of model 1 with some additional layers to help reduce potential overfitting and improve generalization.
The updates included:

- BatchNormalization: improve robustness of model.
- Dropout: Helps prevent overfitting.
- Kernel regularization with L2: Adds penalty to loss function to help reduce overfitting.

5. MODEL EVALUATION METRICS

Evaluation Metrics:

- Test Accuracy
- Confusion Matrix
- Classification Report

MODEL 1 EVALUATION METRICS

Recall Score **87%**

Test accuracy **77%**

	precision	recall	f1-score	support
0	0.49	0.21	0.29	234
1	0.65	0.87	0.74	390
accuracy			0.62	624
macro avg	0.57	0.54	0.52	624
weighted avg	0.59	0.62	0.57	624

MODEL 2 EVALUATION METRICS

Recall Score **83%**

Test accuracy **73%**

	precision	recall	f1-score	support
0	0.36	0.16	0.22	234
1	0.62	0.83	0.71	390
accuracy			0.58	624
macro avg	0.49	0.49	0.47	624
weighted avg	0.52	0.58	0.53	624

6. RECCOMENTDATION

Improve efficency

Models such as these can be used to improve speed and efficiency in detection of Pneumonia. These models can help increase productivity of Radiologists.

Prescreening Tool

Models such as this one can be used to detect early signes of Pneumonia that might not be immediatley apparent to the human eye. Early diagnosis can be vital in treatment of illness.

Address inequality

These models can be used in areas where Radiologists are scares or unavalable. These models can be made avalable online and images of X-rays can be uploaded to help underdeveloped areas improve diagnosis of Pneumonia.

NEXT STEPS AND FUTURE PROJECTS

Improve Model Recall

I would like to keep working with this dataset and continue to improve my model to achieve a Recall score of at least 90%.

Project 2

I would also like to create a model that can detect the difference between Pneumonia caused by a bacterial infection vs. a viral. Pneumonia caused by bacterial infection can be severe.

Project 3

Anomaly Detection
Train a CNN on normal images to identify abnormal structures in CT, MRI, or X-ray scans.

ABOUT ME

Francisco J Torres

Datascience Career Track, Springboard

I AM A LIFELONG LEARNER WITH AN INSATIABLE CURIOSITY. MY DIVERSE EXPERIENCES INCLUDE WORKING AS A FOREST FIREFIGHTER, A HIGH SCHOOL TUTOR,AND AN ENGLISH TEACHER IN SPAIN. I HAVE EARNED A MASTER OF ARTS IN TEACHING WITH A CONCENTRATION IN SECONDARY MATHEMATICS. DURING THE PANDEMIC, I DISCOVERED MY PASSION FOR PROGRAMMING AND DATA SCIENCE, WHICH I PURSUED AS A HOBBY FOR FOUR YEARS BEFORE DECIDING TO TRANSITION INTO THE FIELD PROFESSIONALLY.