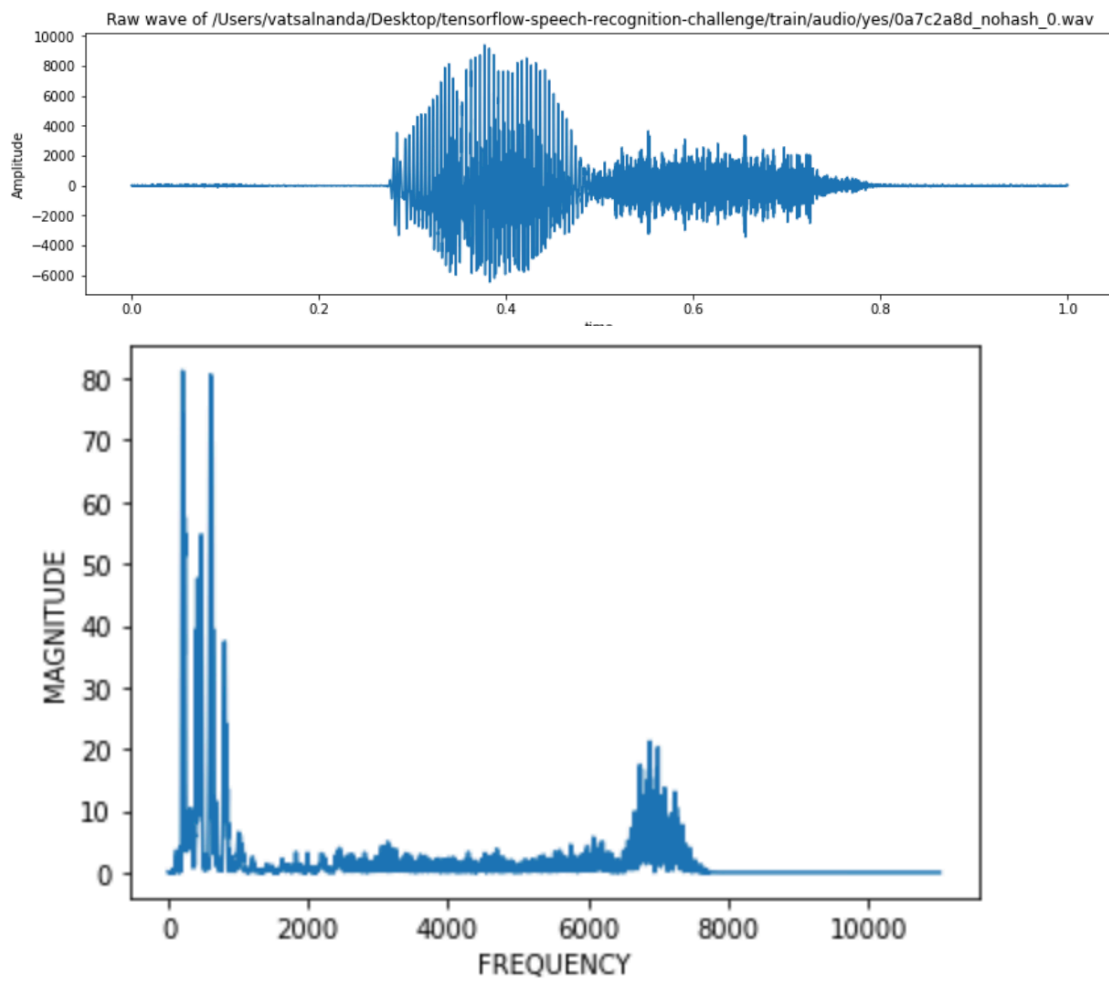# MIDAS TASK-1 DOCUMENTATION

- For the Tensorflow speech recognition challenge , I applied 2 models-
  - i) A basic neural network
  - ii) LSTM-RNN(LONG SHORT TERM MEMORY-RECURRENT NEURAL NETWORK)

- The entire code-base is divided into 5 parts-
  - i) IMPORTING LIBRARIES
  - ii) VISUALISING THE DATA AND PRE-PROCESSING
  - iii) TRAINING THE MODEL
  - iv) TESTING THE MODEL AND PLOTTING THE MODEL
  - v)PREDICTIONS AND SUBMITTING THE MODEL

- I will discuss a bit about every part one by one(The entire code-base is pushed on GitHub, can view there for further details)

  i) In this part I have added all the necessary libraries needed for 'Math/Scientific Math',   'Visualisation', 'Deep Learning', etc.

  ii) We use a particular file, yes/ 0a7c2a8d_nohash_0.wav', just to visualise the data.
 - First we found out the amplitude w.r.t time and plotted a fast Fourier transform (FFT) spectrogram

16000
16000

Raw wave of /Users/vatsalnanda/Desktop/tensorflow-speech-recognition-challenge/train/audio/yes/0a7c2a8d_nohash_0.wav

- An Short Time Fourier transform (STFT)
  spectrogram and a Mel-Frequency Cepstrum
  Coefficient MFCC
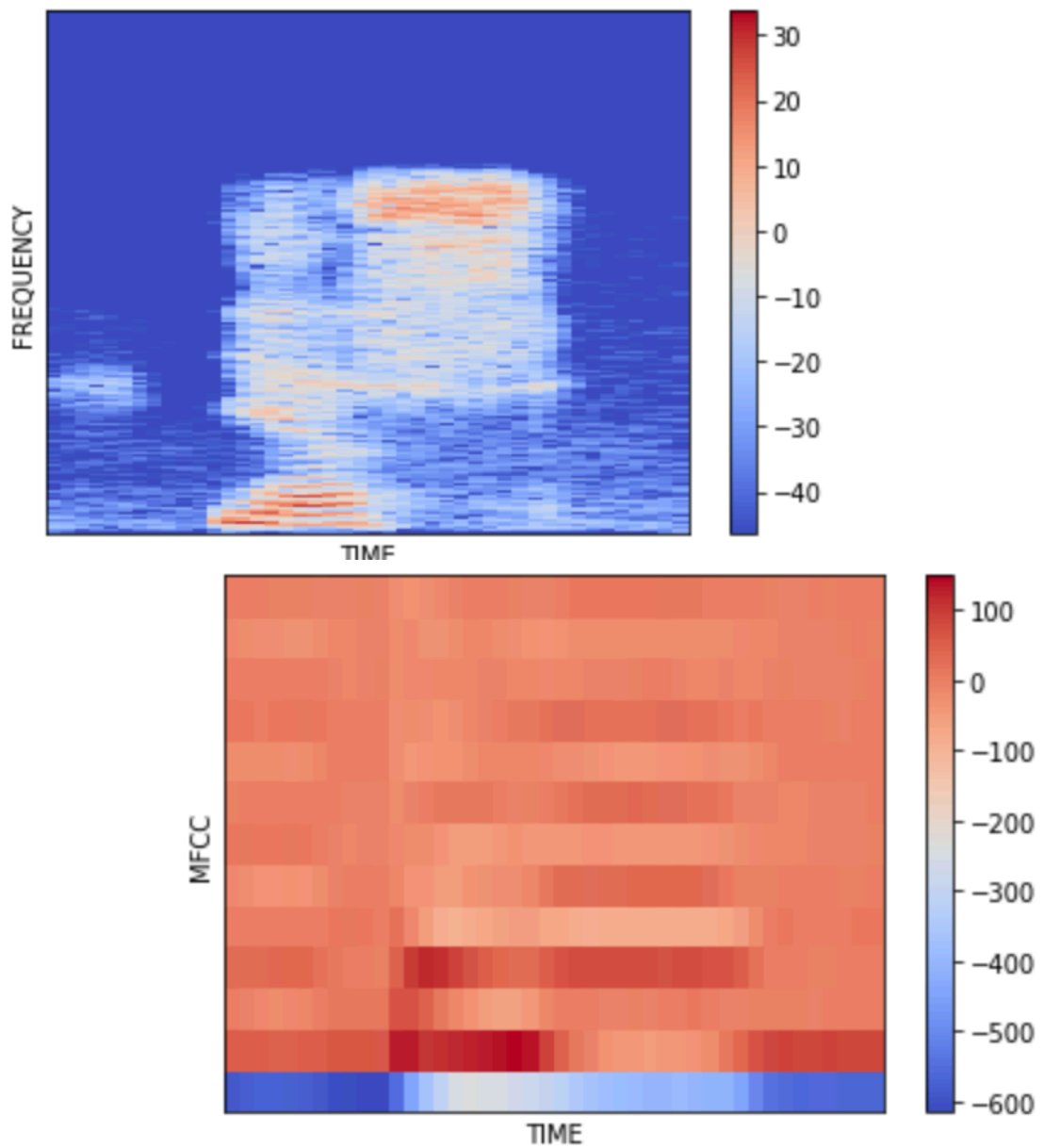
- A bar graph for the number of labels was plotted and duration of recordings-

No. of recordings for each command
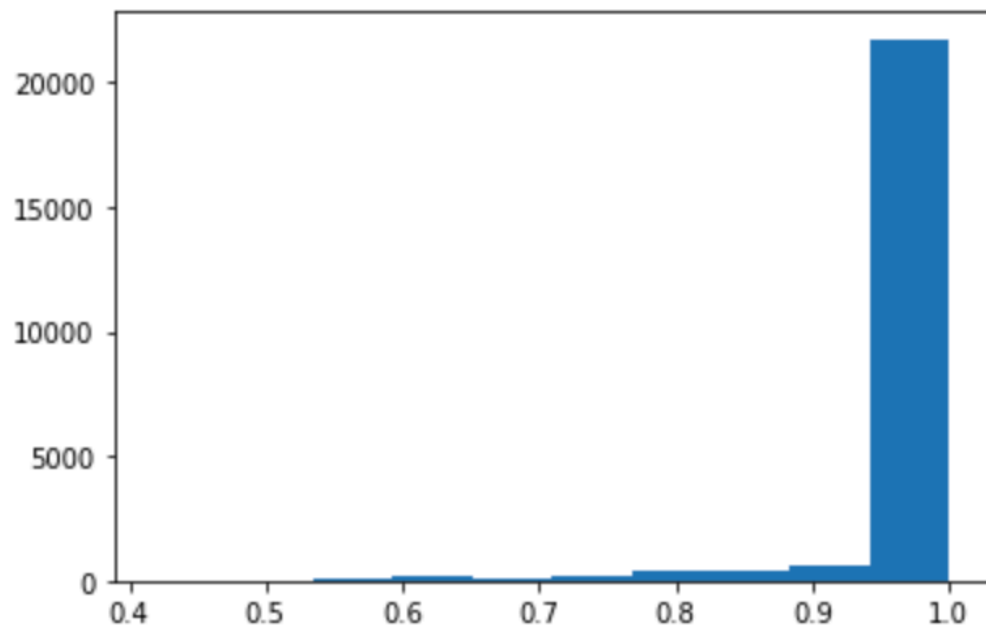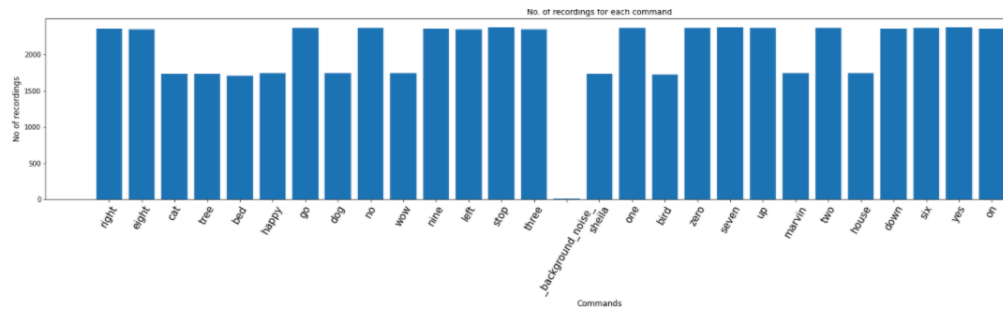
-iii) **EXPERIMENT-**
- The data was split as 80% for training and 20% for testing.
- In the basic neural network, we used one input layer, 2 hidden layers with 0.3 dropout ratio and l2 regularisation and ReLu activation function and for the output, softmax activation function was used .
- The other approach was to use an LSTM-RNN network with 2 lstm layers, 1 dense layer with a dropout ratio of 0.3, ReLu activation function and an output layer with softmax activation function .

OBSERVATIONS and RESULTS-

MODEL	PARAMETERS	VALUES
1) Basic CNN	Training Accuracy	60-65% (approx.) for 100 epochs
	Testing Accuracy	15-16%(approx.) for 100 epochs

	ETA	30 minutes approx. for 100 epochs
2) LSTM-RNN	Training Accuracy	9-11% (approx.) for 1 epoch
	Testing Accuracy	9-11% (approx.) for 1 epoch
	ETA	60 minutes approx. for 1 epoch

INFERENCE-
- The first model was chosen for predictions and testing seeing the results.

iv) The training and testing accuracies were respectively plotted which are clearly shown in the code-base.

v) The predictions were made and a csv file was made with 'fname' and 'label' as 2 headers.

REFERENCES-
i) https://www.kaggle.com/c/tensorflow-speech-recognition-challenge/overview
ii) https://www.youtube.com/user/krishnaik06
iii) https://www.youtube.com/channel/UCZPFjMe1uRSirmSpznqvJfQ
iv) https://jonathan-hui.medium.com/improve-deep-learning-models-performance-network-tuning-part-6-29bf90df6d2d (and many more medium articles)
v) https://www.tensorflow.org/tutorials/audio/simple_audio
vi) Github