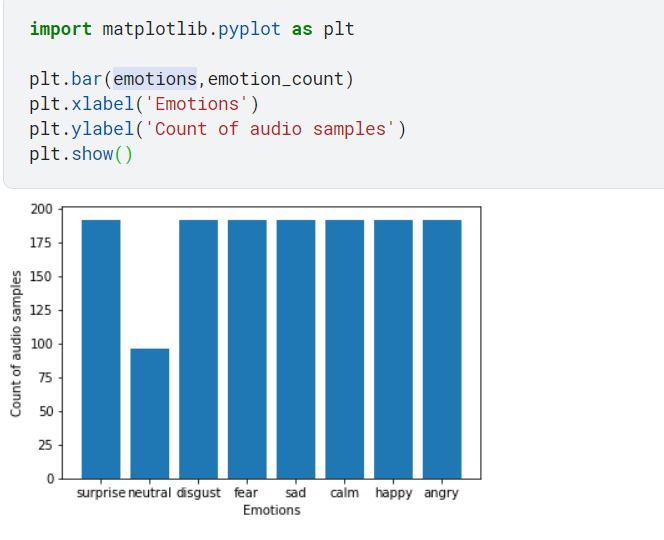
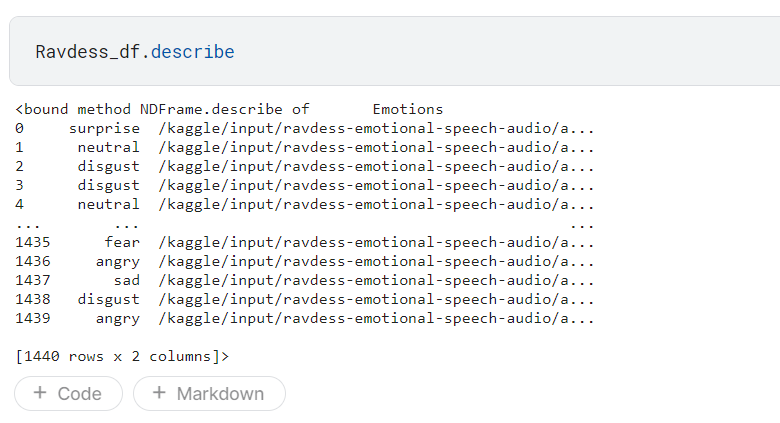
**Week-1**

**1) EDA of the Speech Data(Ravdess dataset):**

* The dataset consists of audio samples from 24 actors , 12 are male and 12 are female
* Audio samples consist of 3-4 seconds of audio.
* The emotions addressed in the dataset are 'surprised', 'neutral', 'disgust', 'fear', 'sad', 'calm', 'happy' ,'angry'.
* Each emotion is produced at two levels of emotional intensity (normal, strong), with an additional neutral expression.
* The actors vocalize two statements/sentences which are: 1. Kids are talking by the door, 2. Dogs are sitting by the door.
* I plotted emotions versus the number of audio samples in the dataset which reflect those emotions.

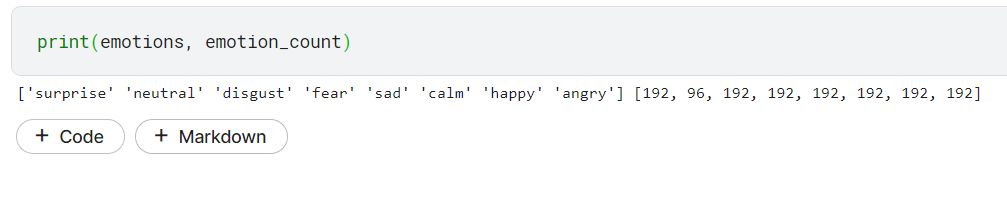


* There are a total of 1440 audio samples in the dataset. 60 trials per actor x 24 actors = 1440 samples audio samples are present in the dataset.



Other Screenshots:

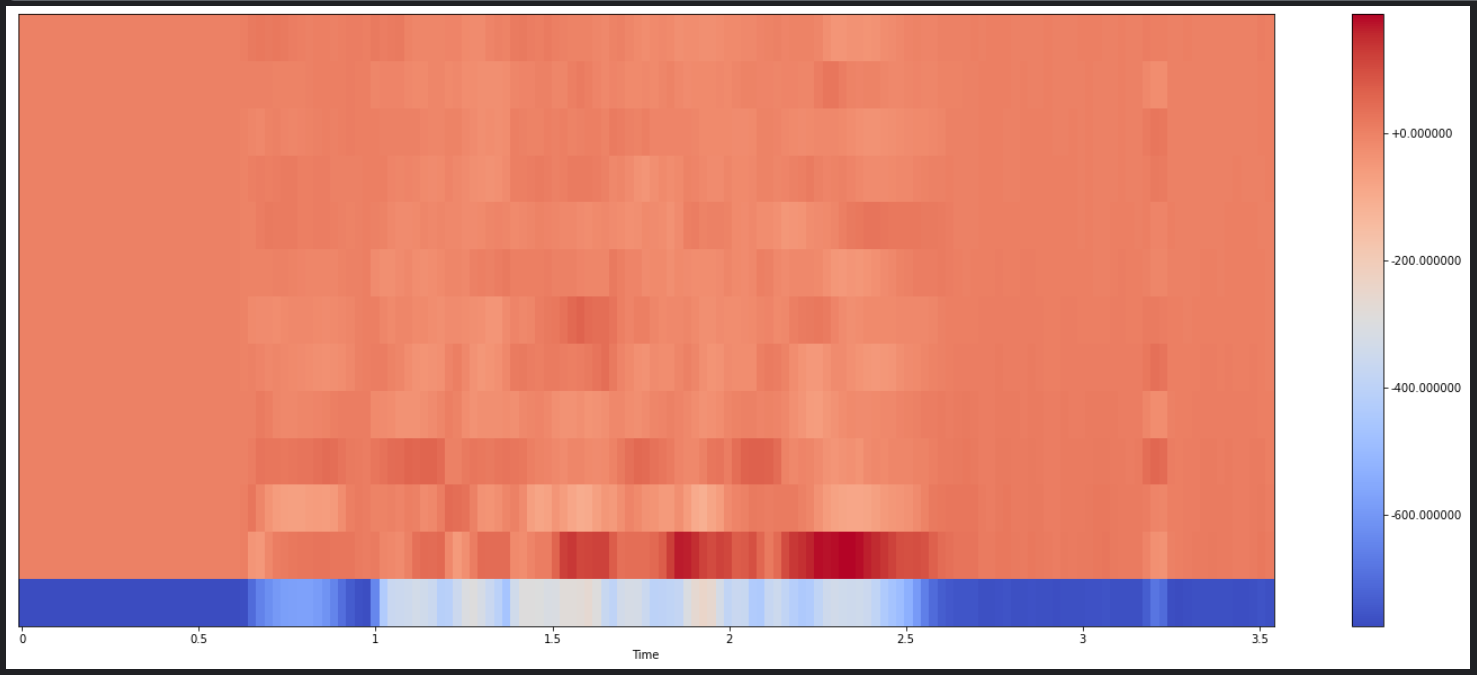


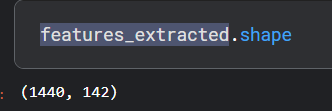


**2) Research:**

* Spent time on researching Speech Emotion Recognition projects and the different methods used to tackle the problem.
* Most Speech Emotion Recognition (SER) projects comprised of the following components :
* **Collection of required speech data**
* **Feature Extraction:** Extracting features from the speech data that can help the classifier distinguish between different emotions. Consists of extracting features that reflect the emotional characteristics in the speech sample. Some of the most common features extracted are: MFCC (Mel Frequency Cepstrum Coefficient) , Mel Spectrogram and other modulation spectrum features.
* **Feature Selection:** Reduce the number of features used to characterize a dataset to improve the models performance. The most common method used for this was Recursive Feature Elimination where recursively subsets of the best features are selected using a model like Linear Regression.
* **Classification:** Learn from the training samples and then use this learning to classify a new observation. Some of the common models used are SVM,RNN(LSTM),CNN.
* **Data Augmentation:** This technique is usually used to improve the accuracy of the base model by creating perturbations in existing audio data to make predictions more accurate.

**3) Feature Extraction**

* Feature Selection is considered the most important part while developing a SER.
* Used the python module librosa for feature extraction.
* Experimented a lot with MFCC’s and plotted the output values to get a clear understanding of what is done and how to implement it. 
* Performed MFCC and Mel Spectrogram extraction on the RAVDESS dataset and stacked them horizontally to form a single concise dataset with the following dimensions.



**3) Pre-processing and Splitting**

* Converted the extracted features into a single data frame.
* Performed One Hot Encoding on the dependent variable(emotions).
* Performed Train test Splitting of data (70-30 split) .
* Scaled the extracted features in the train and test set (independent variable).

**4) Feature Selection**

* Used the train set to train a Linear Regression model and got a score of --1.20.
* Performed Recursive Feature Elimination using Cross Validation(RFECV) on the dataset to find the optimal number of features to extract from the dataset.
* Using the Optimal Number of features I performed RFE and validated the updated dataset with a Linear Regression model and got an improved score of 0.0179.

**5) Tasks for next week:**

* Experiment more with RFE to get better results.
* Use the extracted dataset to construct base ML Models(CNN,SVM).
* Work on improving the accuracy of said ML Models by changing the features extracted and performing better feature selection.