TABLE OF CONTENT

- ✓Problem statement
- ✓Data description
- ✓Gender count
- ✓Spectral entropy, flatness and gender
- ✓Correlation
- ✓Classifier model

PROBLEM STATEMENT

Gender Recognition by Voice and Speech Analysis

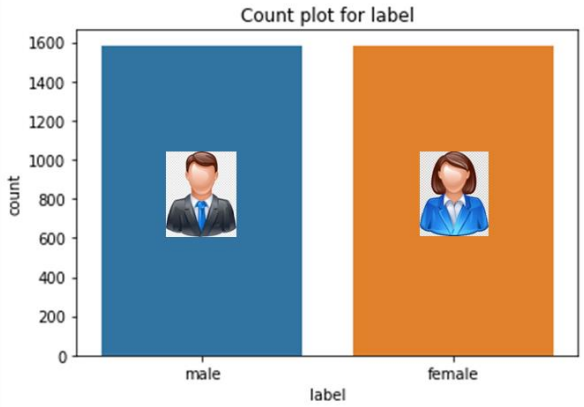
Goal is to identify a voice as male or female, based upon acoustic properties of the voice and speech.

DATA DESCRIPTION

This database was created to identify a voice as male or female, based upon acoustic properties of the voice and speech. The dataset consists of recorded voice samples, collected from male and female speakers. The voice samples are pre-processed by acoustic analysis in R using the seewave and tuneR packages, with an analyzed frequency range of 0hz-280hz (human vocal range).

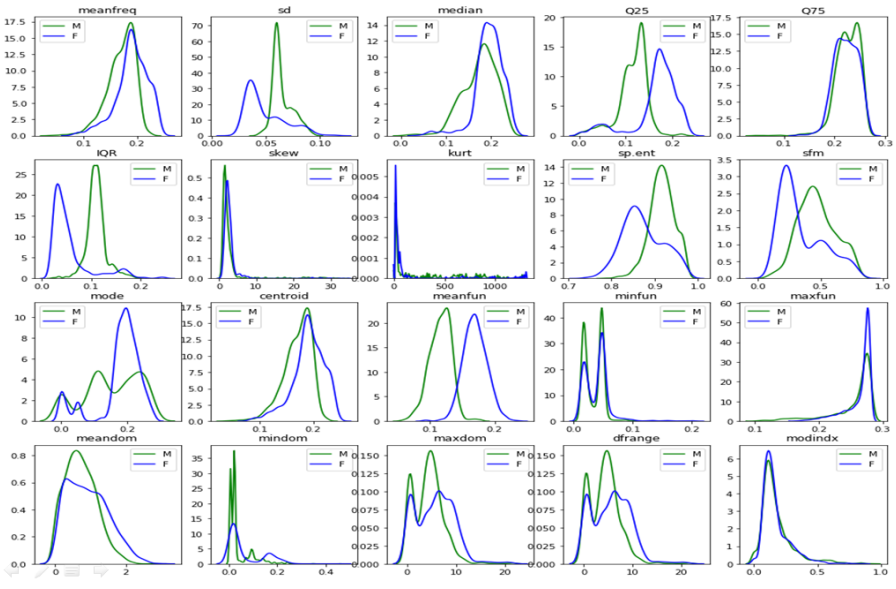
The dataset consists of 3168 samples and 21 features.

GENDER COUNT



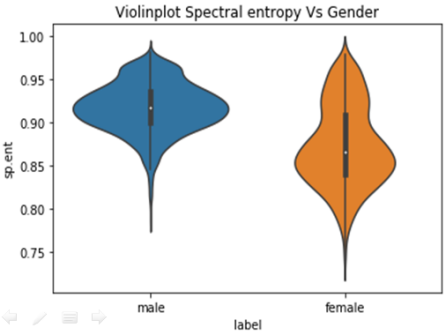
There are equal of male(1584) and female(1584) voice samples in the dataset.

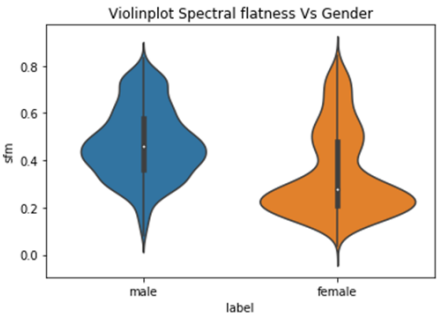
We observe that **Q25,IQR and meanfun** are the major features distinguishing between male and female voice, followed by **sd, sp.ent and sfm**



SPECTRAL ENTROPY, FLATNESS & GENDER

Most of male voice samples fall under Spectral entropy range between 0.92 and 0.93. Where as many female voice samples have got 0.85 spectral entropy.

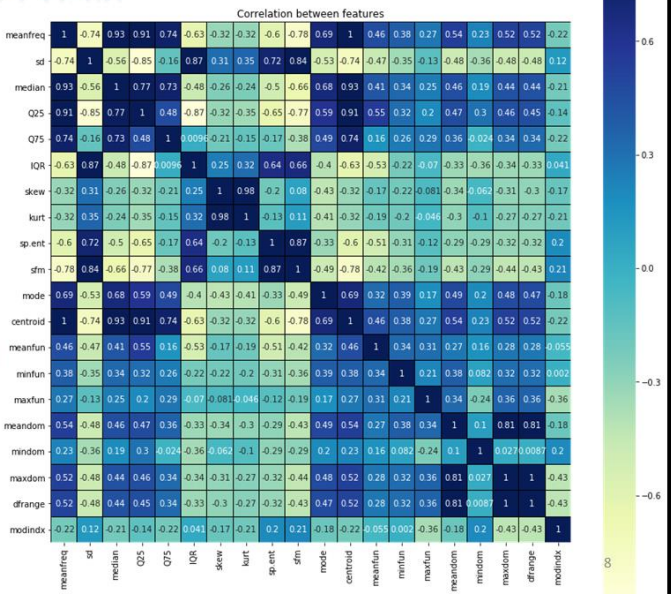




Many **Female** voice samples have got **low Spectral flatness** compared to male.

CORRELATION

meanfreq, median, Q25, Q75, centroid are positively correlated.
sfm-centroid, IQR-Q25, sd-centroid, sd-meanfreq are negatively correlated with each other.



CLASSIFIER MODEL

Various classifier models used to predict the gender using the voice dataset. These models are evaluated on different metrics to know the level of performance. The results are listed here:

Index	Algorithm	Accuracy	Precision	Recall	f1_Score	AUC_ROC_Score
1	LogisticRegression	0.897476	0.858333	0.956656	0.904832	0.896335
2	DecisionTree	0.970032	0.984076	0.956656	0.970173	0.97029
3	RandomForest	0.974763	0.981191	0.96904	0.975078	0.974874
4	KNN	0.976341	0.98125	0.972136	0.976672	0.976422
5	SVM	0.960568	0.954268	0.96904	0.961598	0.960404
6	NaiveBayes	0.869085	0.877358	0.863777	0.870515	0.869188

From the above, we observe that KNN model has performed well with highest scores in all the metrics.

THANK YOU!