

# Gender voice recognition - eksploracja danych

---

MICHAŁ PIASECKI, BARTOSZ SIŃSKI

# Opisy zmiennych

	name	type	description
0	meanfreq	float	mean frequency (in kHz)
1	sd	float	standard deviation of frequency
2	median	float	median frequency (in kHz)
3	Q25	float	first quantile (in kHz)
4	Q75	float	third quantile (in kHz)
5	IQR	float	interquantile range (in kHz)
6	skew	float	skewness (see note in specprop description)
7	kurt	float	kurtosis (see note in specprop description)
8	sp.ent	float	spectral entropy
9	sfm	float	spectral flatness
10	mode	float	mode frequency
11	centroid	float	frequency centroid (see specprop)
12	meanfun	float	average of fundamental frequency measured across acoustic signal
13	minfun	float	minimum fundamental frequency measured across acoustic signal
14	maxfun	float	maximum fundamental frequency measured across acoustic signal
15	meandom	float	average of dominant frequency measured across acoustic signal
16	mindom	float	minimum of dominant frequency measured across acoustic signal
17	maxdom	float	maximum of dominant frequency measured across acoustic signal
18	dfrange	float	range of dominant frequency measured across acoustic signal
19	modindx	float	modulation index. Calculated as the accumulated absolute difference between adjacent measurements of fundamental frequencies divided by the frequency range
20	label	string	Predictor class, male or female

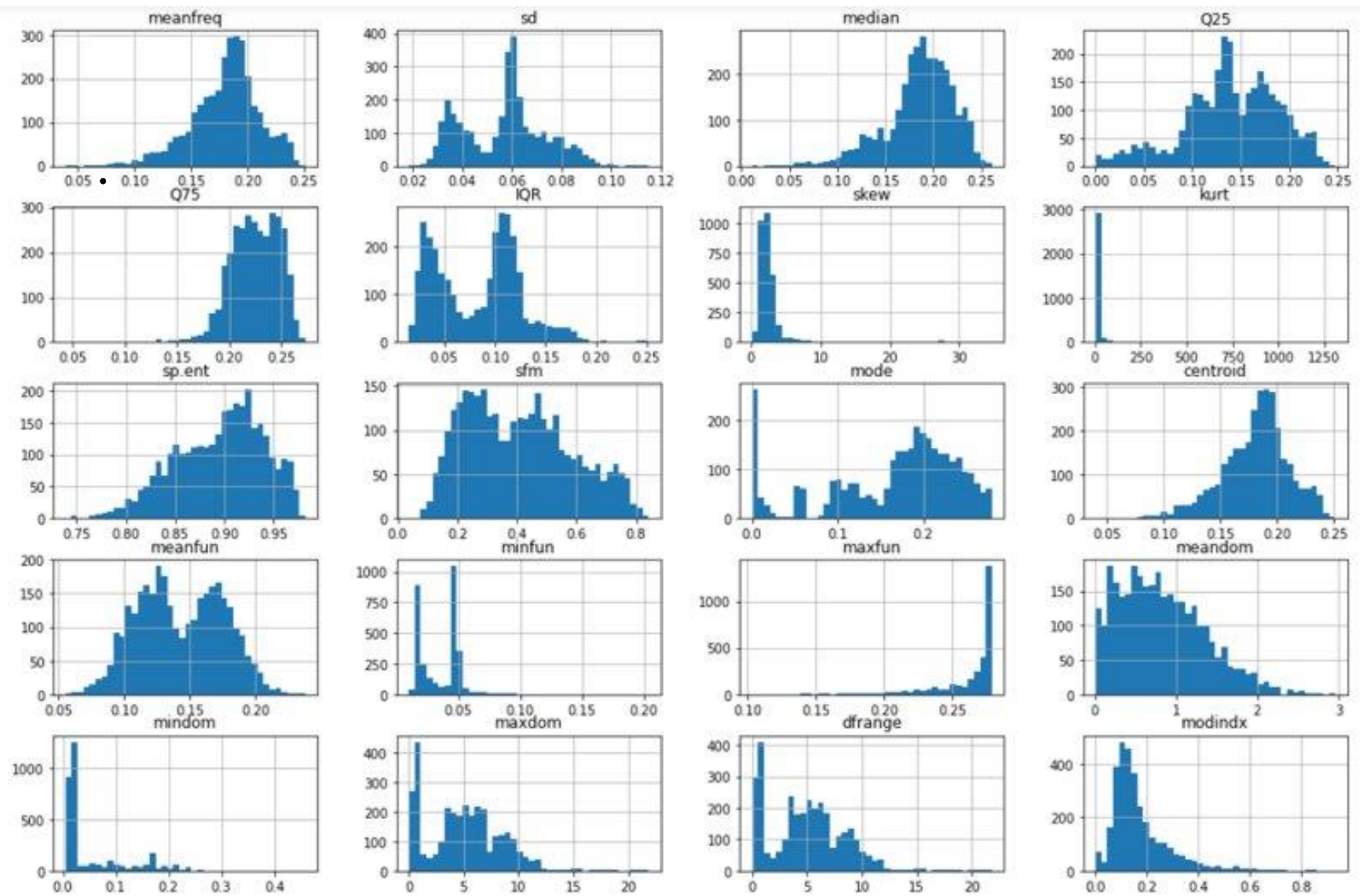
# Podstawowe informacje

	meanfreq	sd	median	Q25	Q75	IQR	skew	kurt	sp.ent	sfm	...	centroid	meanfun	minfun	maxfu
0	0.059781	0.064241	0.032027	0.015071	0.090193	0.075122	12.863462	274.402905	0.893369	0.491918	...	0.059781	0.084279	0.015702	0.27586
1	0.066009	0.067310	0.040229	0.019414	0.092666	0.073252	22.423285	634.613855	0.892193	0.513724	...	0.066009	0.107937	0.015826	0.25000
2	0.077316	0.083829	0.036718	0.008701	0.131908	0.123207	30.757155	1024.927705	0.846389	0.478905	...	0.077316	0.098706	0.015656	0.27118
3	0.151228	0.072111	0.158011	0.096582	0.207955	0.111374	1.232831	4.177296	0.963322	0.727232	...	0.151228	0.088965	0.017798	0.25000
4	0.135120	0.079146	0.124656	0.078720	0.206045	0.127325	1.101174	4.333713	0.971955	0.783568	...	0.135120	0.106398	0.016931	0.26666

```
label
female    1584
male      1584
Name: meanfreq, dtype: int64
```

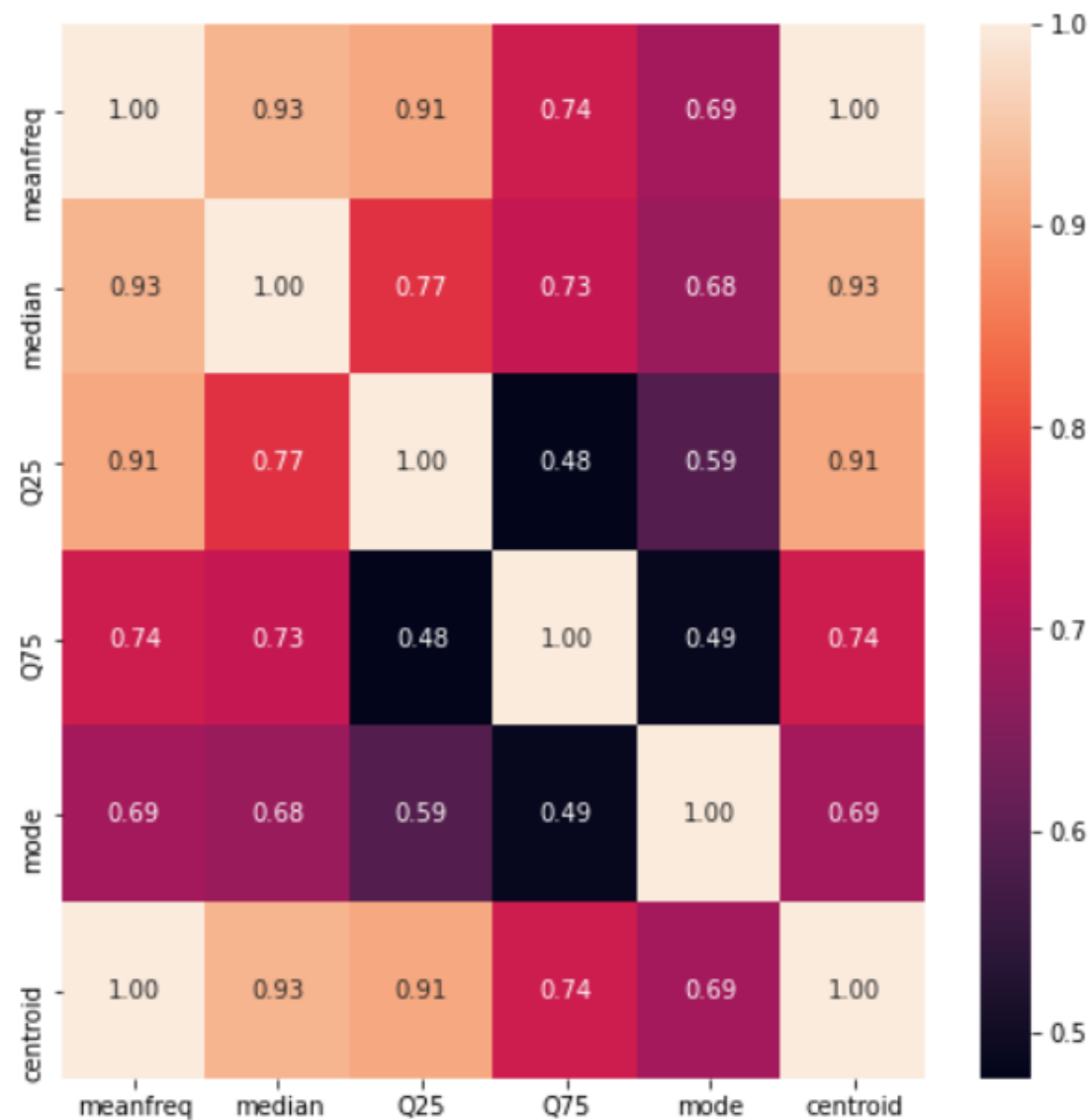
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3168 entries, 0 to 3167
Data columns (total 21 columns):
#   Column      Non-Null Count  Dtype
---  -
0   meanfreq    3168 non-null   float64
1   sd          3168 non-null   float64
2   median      3168 non-null   float64
3   Q25         3168 non-null   float64
4   Q75         3168 non-null   float64
5   IQR         3168 non-null   float64
6   skew        3168 non-null   float64
7   kurt        3168 non-null   float64
8   sp.ent      3168 non-null   float64
9   sfm         3168 non-null   float64
10  mode        3168 non-null   float64
11  centroid    3168 non-null   float64
12  meanfun     3168 non-null   float64
13  minfun      3168 non-null   float64
14  maxfun      3168 non-null   float64
15  meandom     3168 non-null   float64
16  mindom      3168 non-null   float64
17  maxdom      3168 non-null   float64
18  dfrange     3168 non-null   float64
19  modindx     3168 non-null   float64
20  label       3168 non-null   object
dtypes: float64(20), object(1)
memory usage: 519.9+ KB
```

# Rozkłady zmiennych



# Korelacje zmiennych

	label_male	label_female
label_male	1.000000	1.000000
label_female	-1.000000	1.000000
meanfun	-0.833921	0.833921
IQR	0.618916	-0.618916
Q25	-0.511455	0.511455
sp.ent	0.490552	-0.490552
sd	0.479539	-0.479539



# Różnice pomiędzy kobietami i mężczyznami

