1. Dataset names
   - Parkinson's disease data set

2. Number of records in the dataset
   - 195 records

3. List of features in the dataset
   - Name
   - MDVP:Fo(Hz) - Average vocal fundamental frequency
   - MDVP:Fhi(Hz) - Maximum vocal fundamental frequency
   - MDVP:Flo(Hz) - Minimum vocal fundamental frequency
   - MDVP: Jitter(%), MDVP: Jitter(Abs), MDVP: RAP, MDVP: PPQ, Jitter: DDP - Several measures of variation in fundamental freq.
   - MDVP: Shimmer, MDVP: Shimmer(dB), Shimmer: APQ3, Shimmer: APQ5, MDVP: APQ, Shimmer: DDA - Several measures of variation in amplitude
   - NHR, HNR - Two measures of the ratio of noise to tonal components in the voice
   - Status - The health status of the subject (one) - Parkinson's, (zero) - healthy
   - RPDE, D2 - Two nonlinear dynamical complexity measures
   - DFA - Signal fractal scaling exponent
   - spread1, spread2, PPE - Three nonlinear measures of fundamental frequency variation

4. A short description of the dataset
   This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds to one of 195 voice recordings from these individuals ("name" column).

5. Public or not

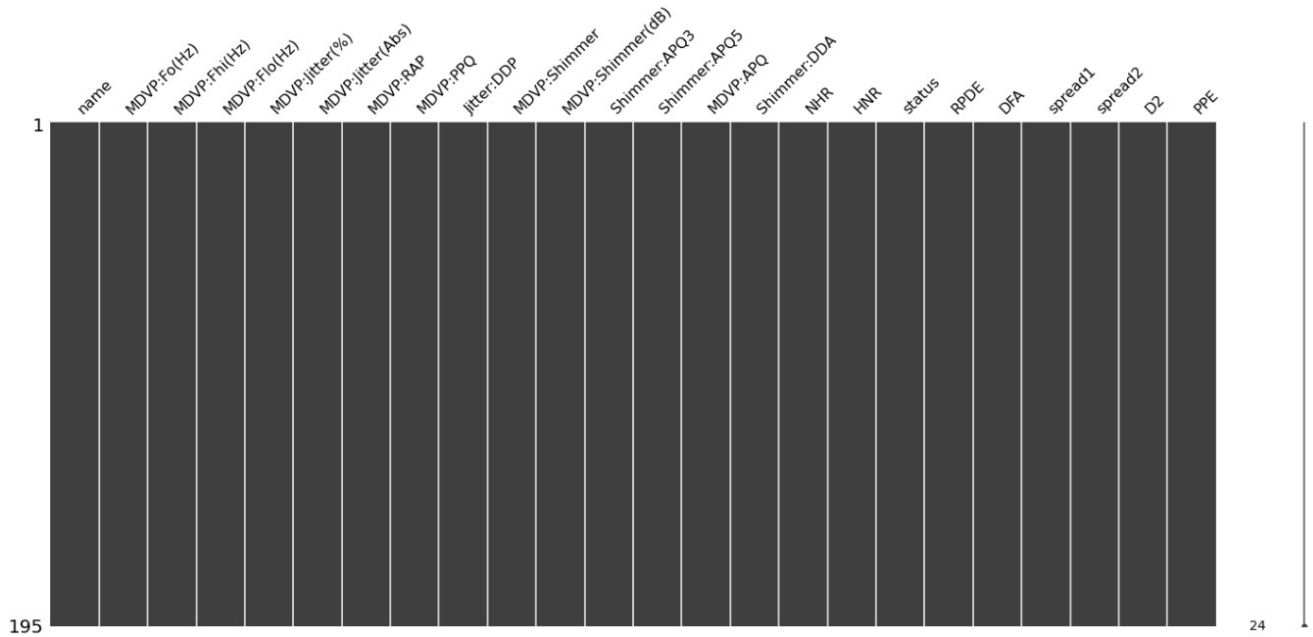
   It’s public

6. Dataset link

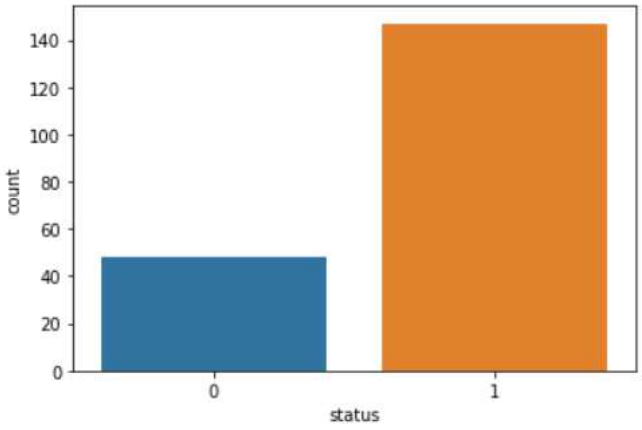
   https://www.kaggle.com/vikasukani/parkinsons-disease-data-set

7. Data

   Matrix shows us that all data is clear. There is null value nowhere.

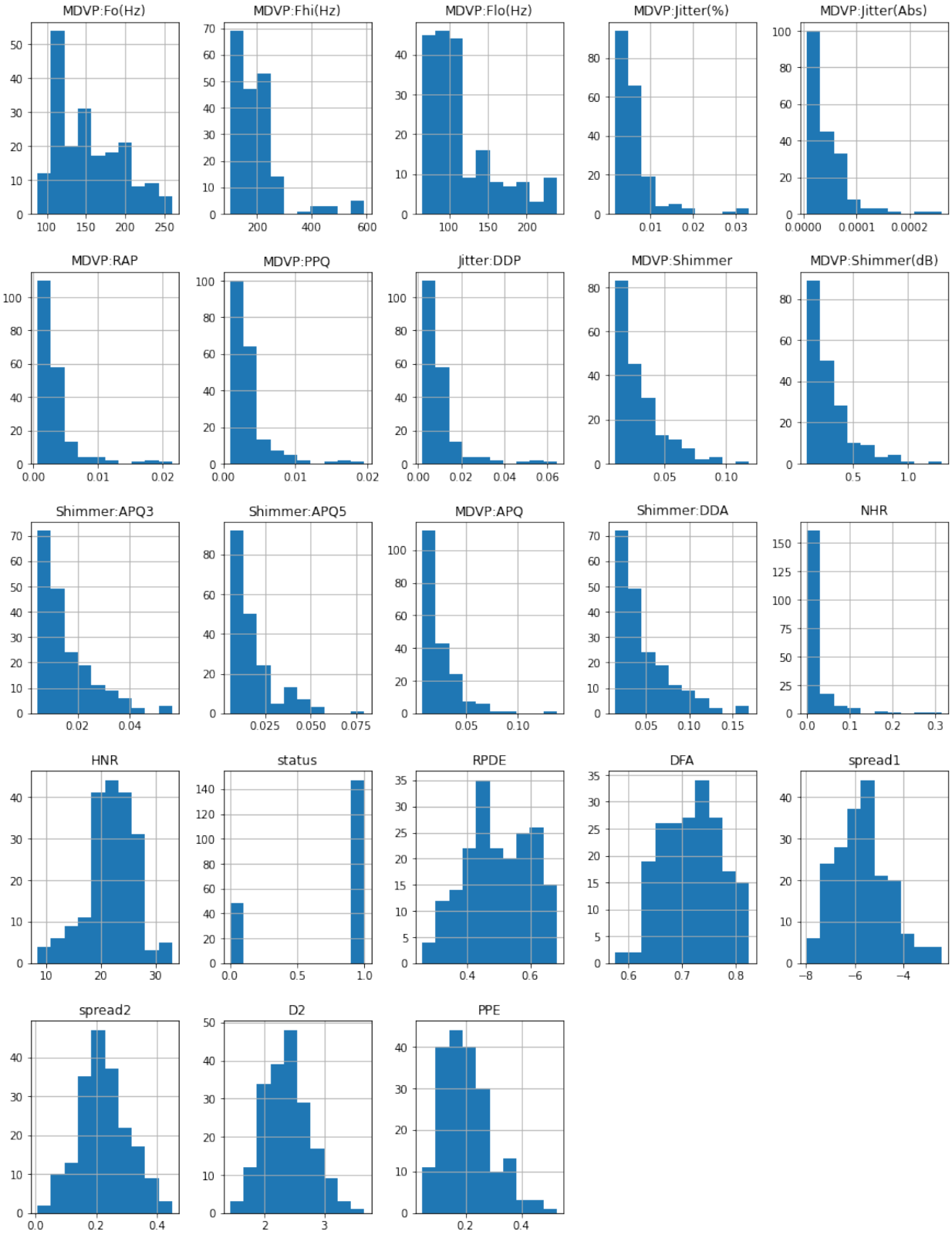


   In accordance with the bar chart we can say the number of ill people significantly exceeds the amount of healthy patients.

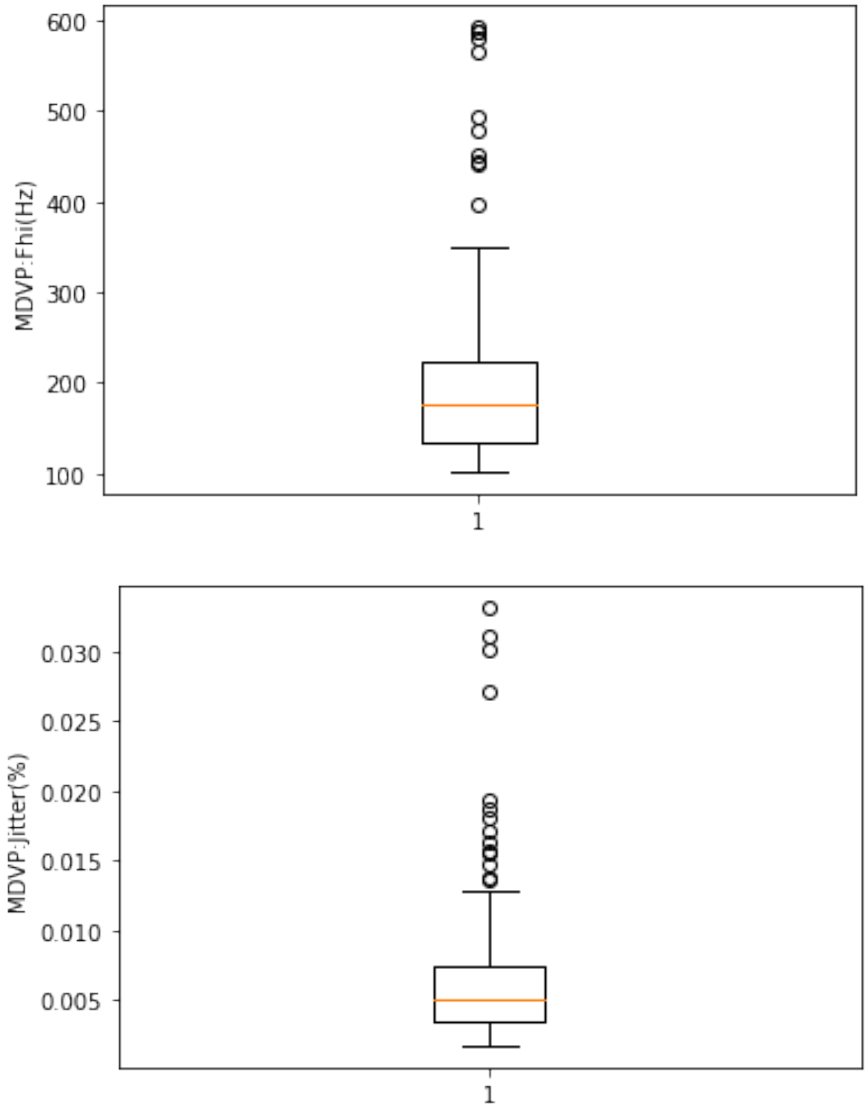


   We can see some of the data is normally distributed and most of the attributes are right skewed

Histogram for each numeric input variable



There are some outliers as we can see some attributes have huge difference in their 75 percentile value and maximum value.



8. Prediction

Based on details of patients' voice I built machine learning model with XGB Classifier algorithm. My test shows the accuracy of the model is over 87 %.