



Detecting Diabetes Early

Daniel Kim



Detecting Diabetes Early

DIABETES SYMPTOMS

Problem Statement & Goal - How to Detect it?



Blurry Eyesight



Feeling Hungry



Sudden Weight Loss



Feeling Thirsty



Frequent Urination



Data Wrangling

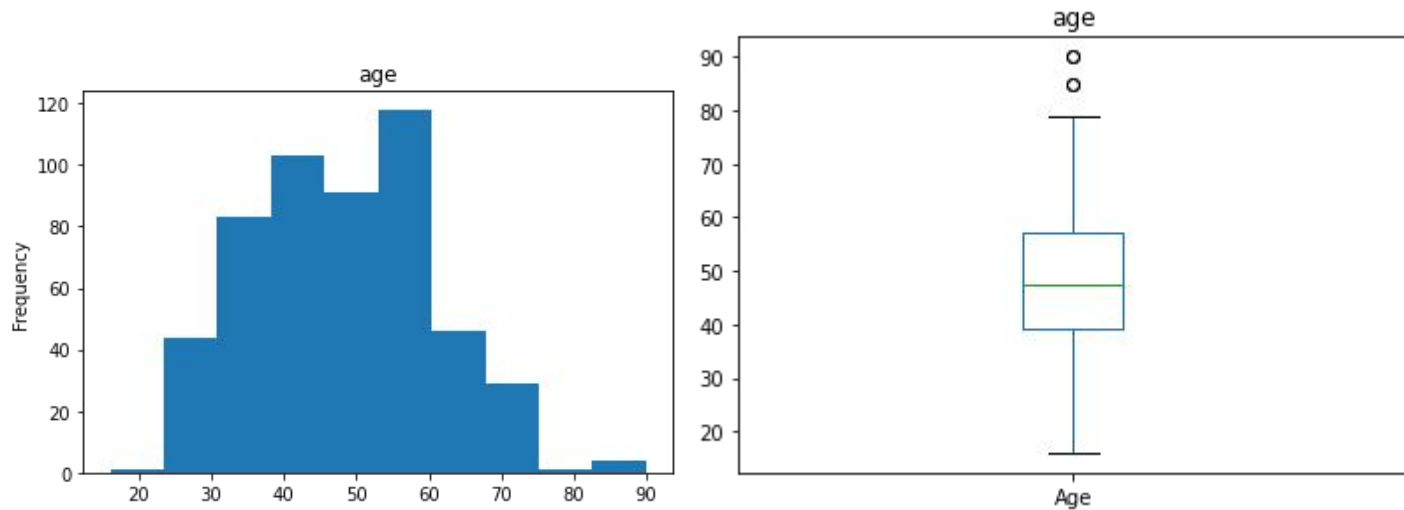
1. Detecting NaN & Redundancy

- a. There were no NaN
- b. Although there were redundancies, it is too early to decide to remove redundancies because the redundancies are not the same but different individuals having same symptoms.
- c. Except for Age, all other features are binary, so it was needed to do one hot encoding to make a machine learning understand the data better.

2. One Hot Encoding

After one-hot encoding, all the other binary columns ending with such as '_Female', '_Negative', or '_No' were removed.

Exploratory Data Analysis - Histogram & Boxplot



'Age' is the only feature showing normal distribution with a little right-skewness because all other features are binary. By its Boxplot, 'Age' seems having 2 outliers.