PNS-2019 (National Health Survey 2019), Prediction of Heart Disease by LGBM Modeling

Exploratory Data Analysis (EDA)

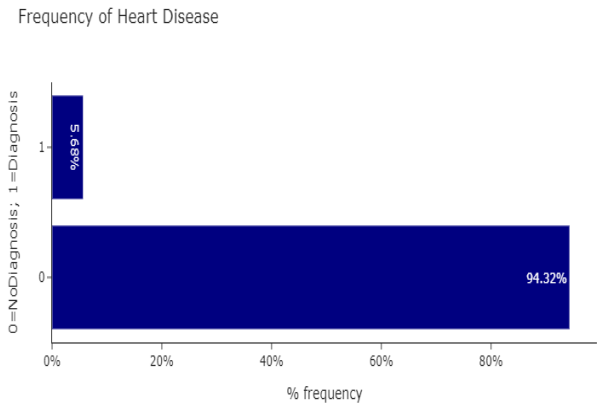
The idea behind the current EDA is to show some features within others in the current dataset associated with someone developing any Heart Disease (HD). According to WHO – World Health Organization https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1), Cardiovascular Diseases (CVD) are the leading cause of death globally, taking an estimated 17.9 million lives each year. CVDs include heart and blood vessel disorders, including coronary heart disease, cerebrovascular disease, rheumatic heart disease, and other conditions. More than four of five CVD deaths are due to heart attacks and strokes, and one-third of these deaths occur prematurely in people under 70.

The most important risk factors for heart disease are associated with an unhealthy diet, physical inactivity, tobacco use and harmful use of alcohol. In general, those risk factors might develop in individuals by increasing their blood pressure, blood glucose (Diabetes) and blood lipids (Dyslipidemia), overweight and obesity (higher BMI – body mass index).

To prevent and minimize the risk of developing cardiovascular diseases like heart disease and stroke. So, the recommendations have been to quit smoking, reduce salt consumption, add more fruit and vegetables to daily meals, do regular physical activity and avoid the harmful use of alcohol.

Regarding chronic diseases, the PNS-2019 (National Health Survey 2019) collected a considerable volume of information about Brazilian health conditions, including data about heart diseases and their risk factors that support the current project to develop a predictive heart model disease by LGBM Classifier.

Exploratory Data Analysis (EDA) gives an overview of the risk factors associated with heart disease. First, the graphic below shows the frequency of cardiovascular diseases (heart disease and stroke) in the Brazilian population.



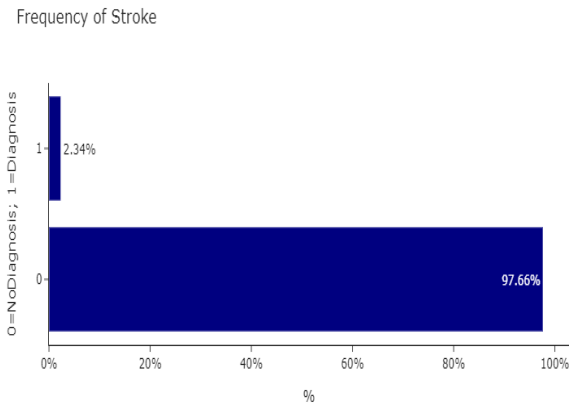


Fig 1 – Frequency of heart disease and stroke based on PNS-2019 through the question:
Have you ever received any diagnoses of heart disease and stroke?

Roughly 5.6% and 2.3% of the Brazilian population claimed in the current survey that they already had a cardiovascular disease diagnosis. It highlighted the urgency of public health issues that demand actions and changes in the population's behaviors.

In terms of modeling, the first hurdle just came up with a low frequency of people with heart disease diagnoses showing that the dataset is imbalanced. That will be considered later.

Figure 2 below shows a stratification of the frequency of heart disease. It seems a tendency that the way of increasing both income and education lower cases of heart diseases. At least three reasons might associate with this: 1) better access to the healthcare system; 3) quality of meals including more protein than carbohydrates, and 4) a better understanding of the risk factors, their consequences and actions to minimize them.

From the gender perspective, men are slightly higher than women in cases of heart disease. It might be because women are more concerned with family health and themselves than men. Graphic 2 also shows a higher frequency of heart disease lower 50 years old. A clue might be that the Brazilian population is higher in that age range; also, the current project grouped all types of heart diseases surveyed in PNS-2019 together.

.

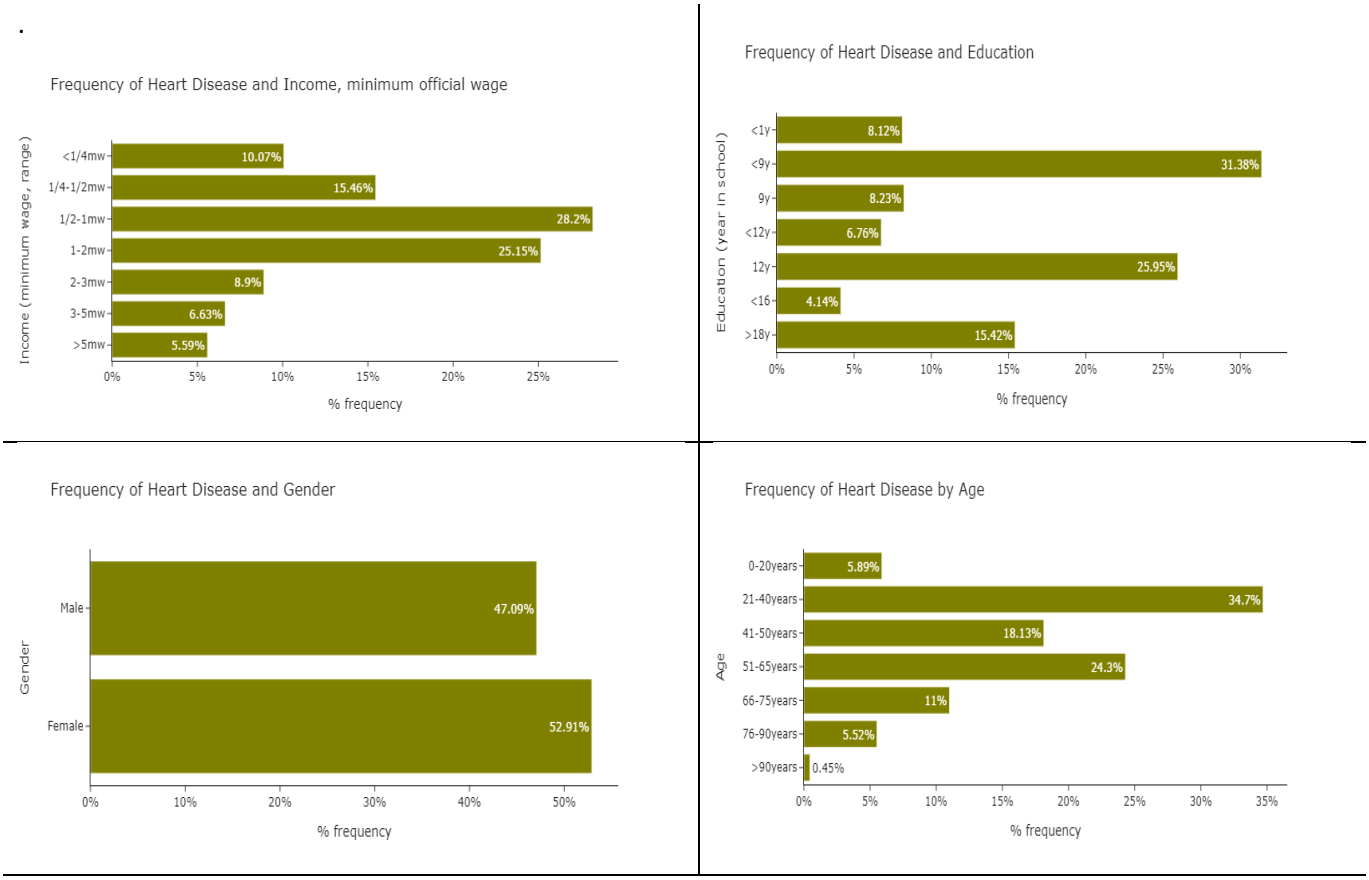
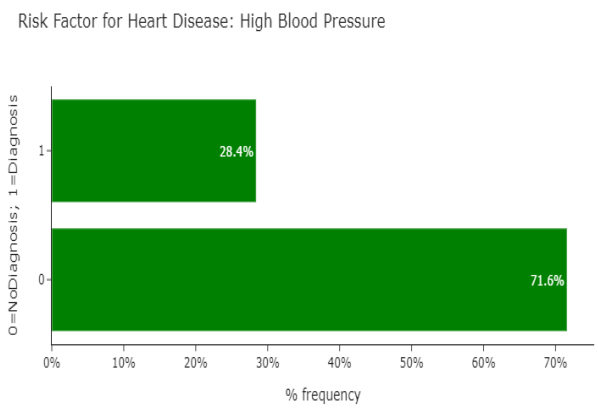
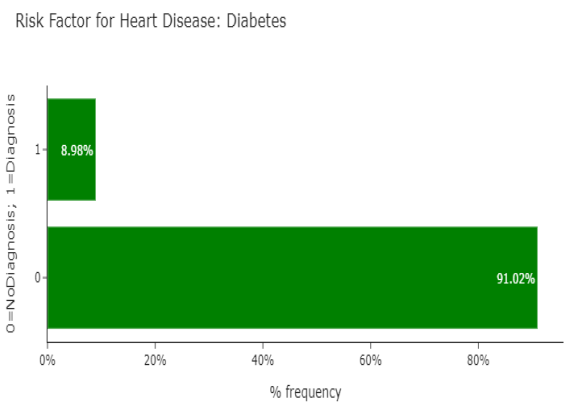
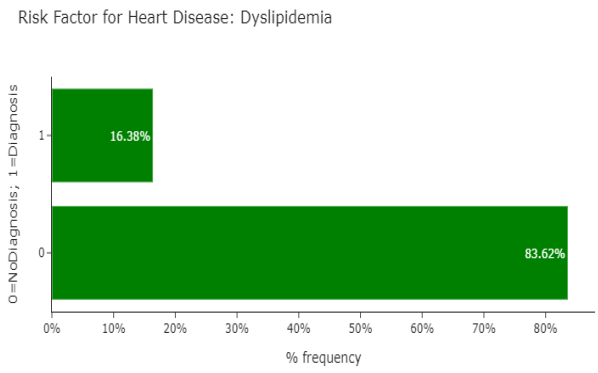


Fig 3 – The frequency of risk factors stratified by income range, education level, gender and age.

As mentioned earlier, heart disease is associated with different risk factors. The following graphics below show their frequencies in the Brazilian population through the PNS-2019







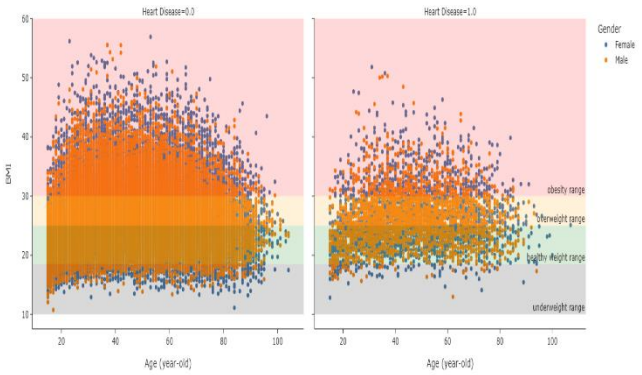
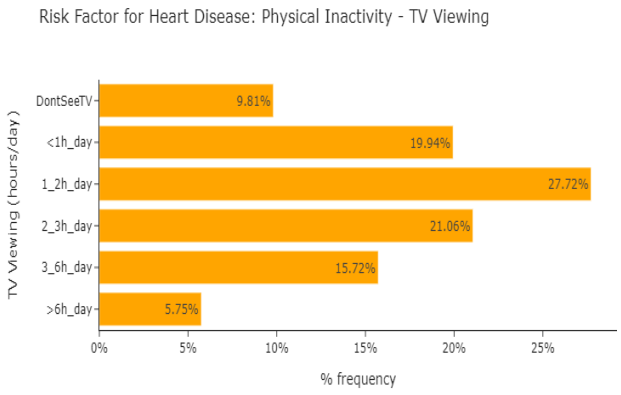
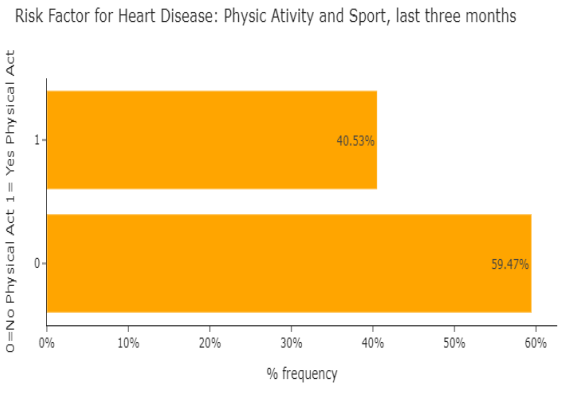
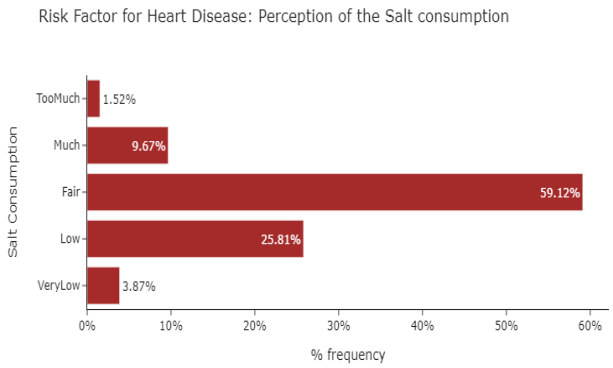


Fig 3 – The frequency of risk factors associated with heart disease is shown in the graphics below. For that PNS-2019 asked questions like Have you ever received any diagnoses of one of those diseases? BMI (body mass index) came from the weight and height of the interviewees.

The frequency of high blood pressure, dyslipidemia and diabetes should concern the entire population, mainly when associated with a large number of Brazilians with overweight and obese; once all of those risk factors feed more cases of cardiovascular disease in the middle and long-term.







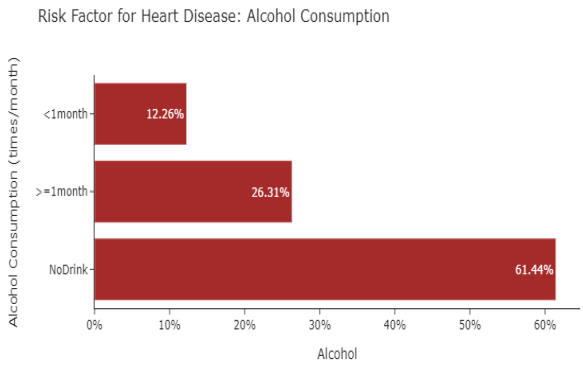


Fig 4 – The frequency of risk factors associated with heart disease is shown in the graphics below. PNS-2019 surveyed some habits and behaviors that might impact heart disease development.

Physical inactivity is represented here by the time spent viewing tv and practicing any physical activity or sport in the last three months. Think as a chain reaction when physical inactivity and unhealthy foods increase BMI, leading to high blood pressure, dyslipidemia, and diabetes, for instance. All of them potentialize the risk of heart disease and stroke development.

From the modeling perspective, those features and many others should significantly weigh the model's quality to predict health diseases.

Further info about the PNS-2019: methodologies, survey application, sampling program, etc.

https://doi.org/10.1590/S1679-49742020000500004 - Profile of National Health Databases • Epidemiol. Serv. Saúde 29 (5) 2020