Abstract

This project aimed to analyze cardiovascular disease to find which variables are related to the disease. The used data in this project are provided by Kaggle. The dataset contains information about patients doing cardiovascular disease examination.

Design

This project is one of the T5 Data Science BootCamp requirements. Data provided by Kaggle has been used in this project. This project to discover reasons behind the increased rate in cardiovascular disease in people based on some variables.

Data

The dataset is provided in .csv format. It contains 70000 patients, each patient has 13 features. The most features of this project are age, systolic blood pressure, diastolic blood pressure, Height, and Weight for the patient. I added a new feature by using the Height,and Weight for the patient and the name for it is bmi(body mass index), also I added a new feature by using the systolic blood pressure and diastolic blood pressure of the patient and the name for it is bp_c(blood_pressure_categories).

Algorithms

Feature Engineering
- Cleaning the dataset by dropping in a useful column for doing analysis, which is the id column.
- Cleaning the systolic blood pressure and diastolic blood pressure columns by dropping the negative values or impossible values.
- Cleaning the bmi column by dropping the outlier values.

Tools

Pandas for data manipulation.
Seaborn for plotting.

Communication

The slides are provided here.