Nama: Andrian Andi Prakasa

NIM: 210411100114

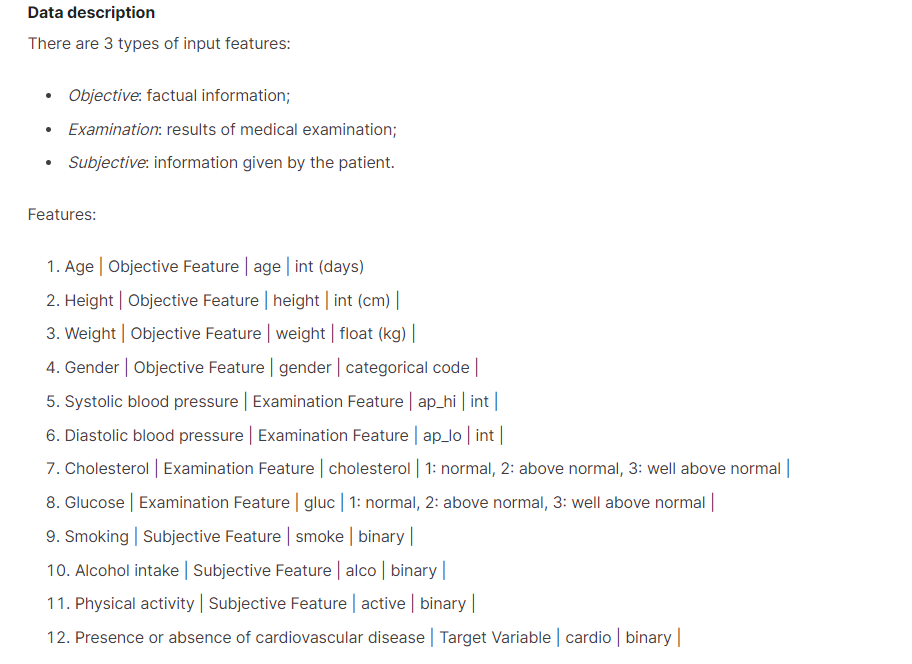
ANALISIS PENYAKIT CARDIOVASCULAR

Data yang terpakai merupakan data penyakit cardiovascular

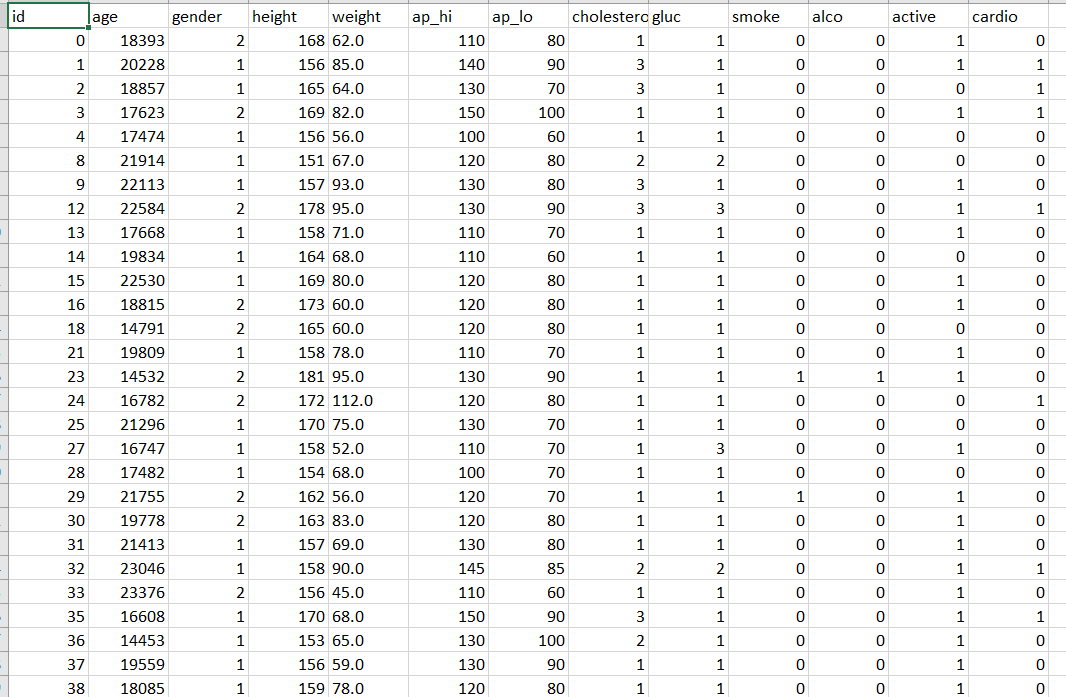
Penyakit kardiovaskular (CVD), juga dikenal sebagai penyakit jantung dan pembuluh darah, adalah kelompok kondisi medis yang memengaruhi jantung dan sistem pembuluh darah. Penyakit ini termasuk berbagai masalah yang dapat memengaruhi pembuluh darah seperti arteri koroner (arteri yang memasok darah ke jantung), pembuluh darah otak, arteri perifer (pembuluh darah yang memasok darah ke ekstremitas), dan jantung itu sendiri. Penyakit kardiovaskular adalah salah satu penyebab utama kematian di seluruh dunia.

Dataset yang peneliti gunakan adalah data pada kaggle dengan jumlah 70000 data. Langkah langkah yang akan peneliti lakukan adalah melakukan cleaning data, memindai dataset apakah terdapat missing value, melakukan balancing data, kemudian melakukan transformasi, lalu langkah terakhir melakukan analisa data

Berikut penjelasan dataset



*Penjelasan setiap fitur pada dataset*



*Dataset Cardiovascular*

**CLEANING DATA**

Dataset dilakukan cleaning data agar field duplikat dan field error dihapus terlebih dahulu



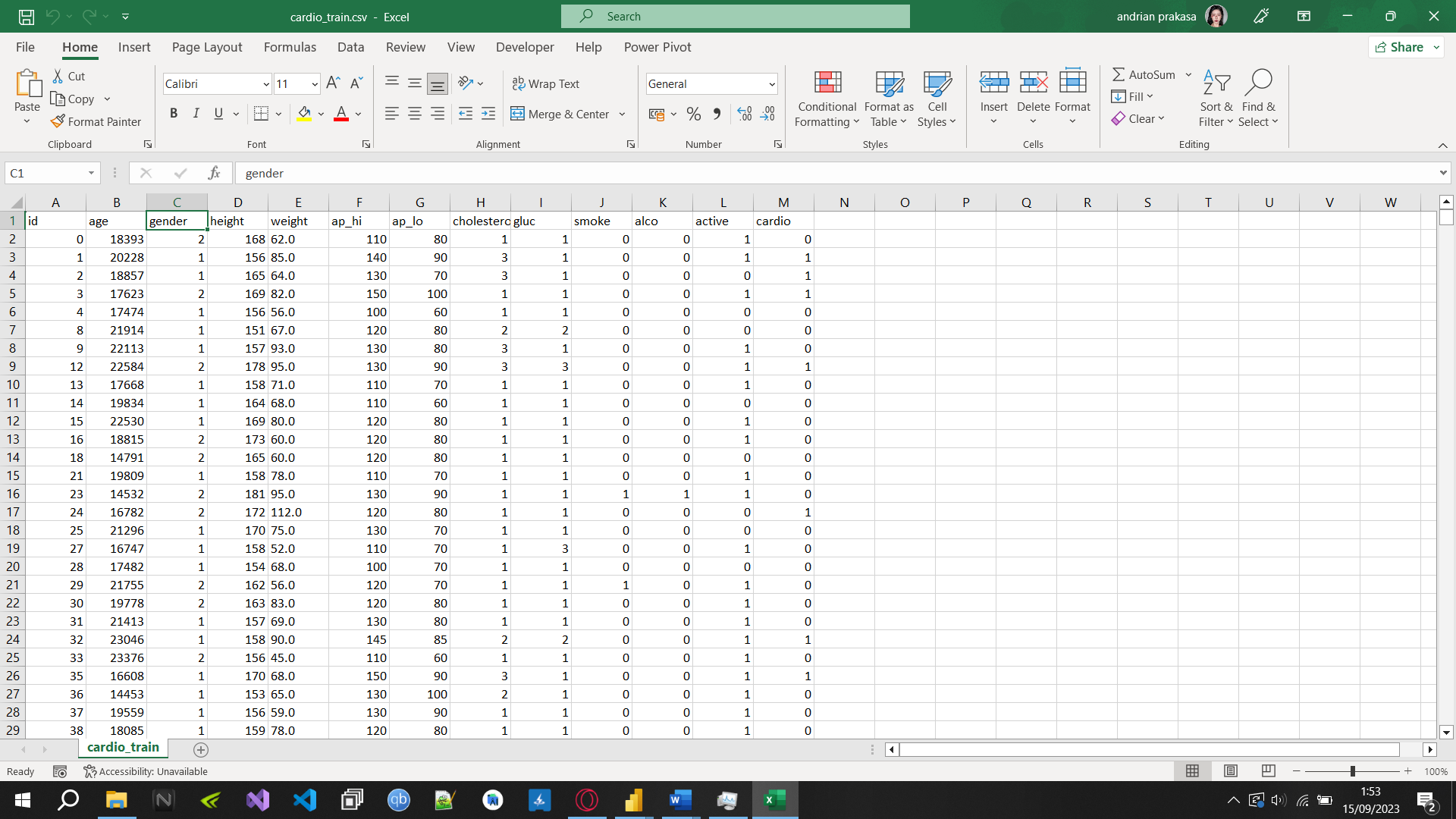
*Jumlah field sesudah dilakukan cleaning data menggunakan fitur dari POWERBI*

Pengurangan data setelah melakukan *remove duplicate,* dan *remove error* terdapat pengurangan sebanyak 24 field

**TRANSFORMASI DATA**

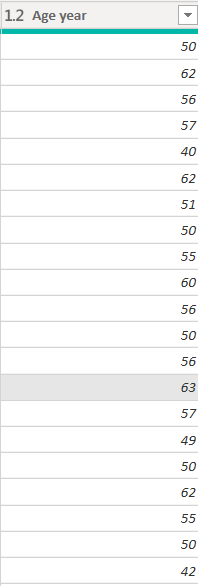
Transformasi data yang dilakukan peneliti adalah sebagai berikut:

1. Menghapus kolom yang tidak diperlukan
2. Melakukan perubahan pada hari menjadi tahun
3. Melakukan perubahan biner gender



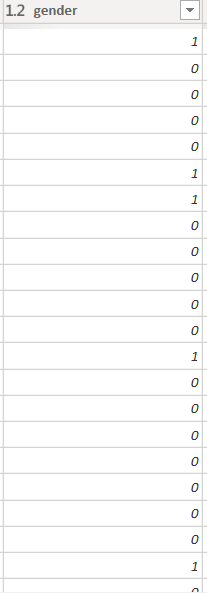
*File excel yang terunduh*

Terdapat kolom id pada file yang telah terunduh, peneliti menghapus kolom ini untuk melakukan transformasi data



*Kolom baru dengan nama Age year*

Peneliti juga melakukan perubahan umur menjadi tahun dan dibulatkan ke bawah menggunakan PowerBI dengan membuat kolom baru dengan menggunakan rumus



*Biner gender*

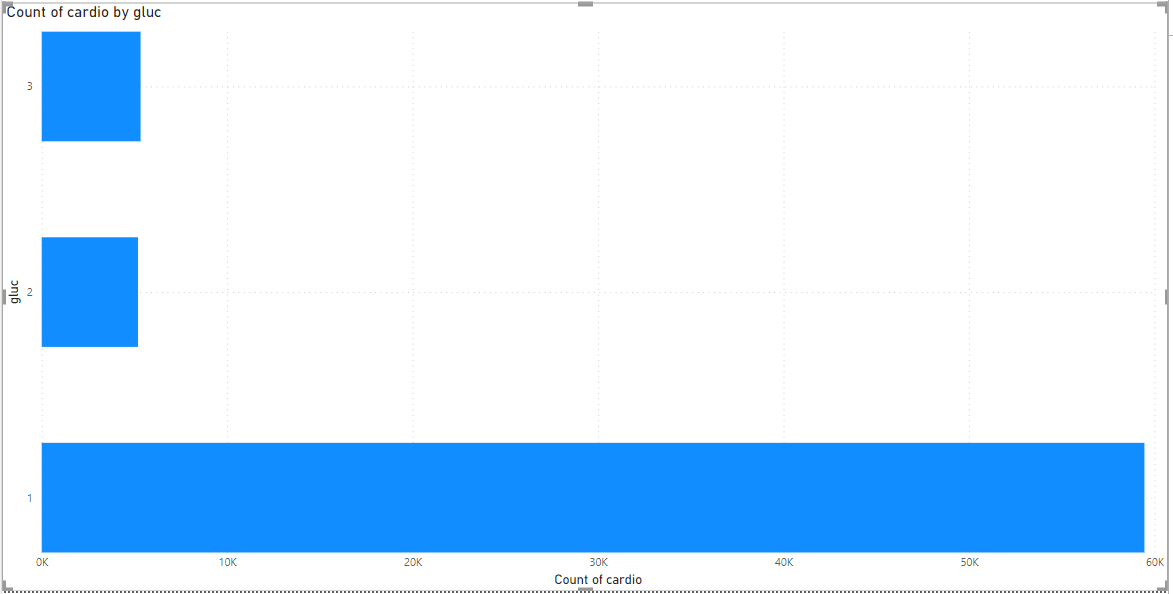
Biner gender pada awalnya adalah 1=female dan 2=male, peneliti merubah menjadi 0 dan 1 dikarenakan lebih umum

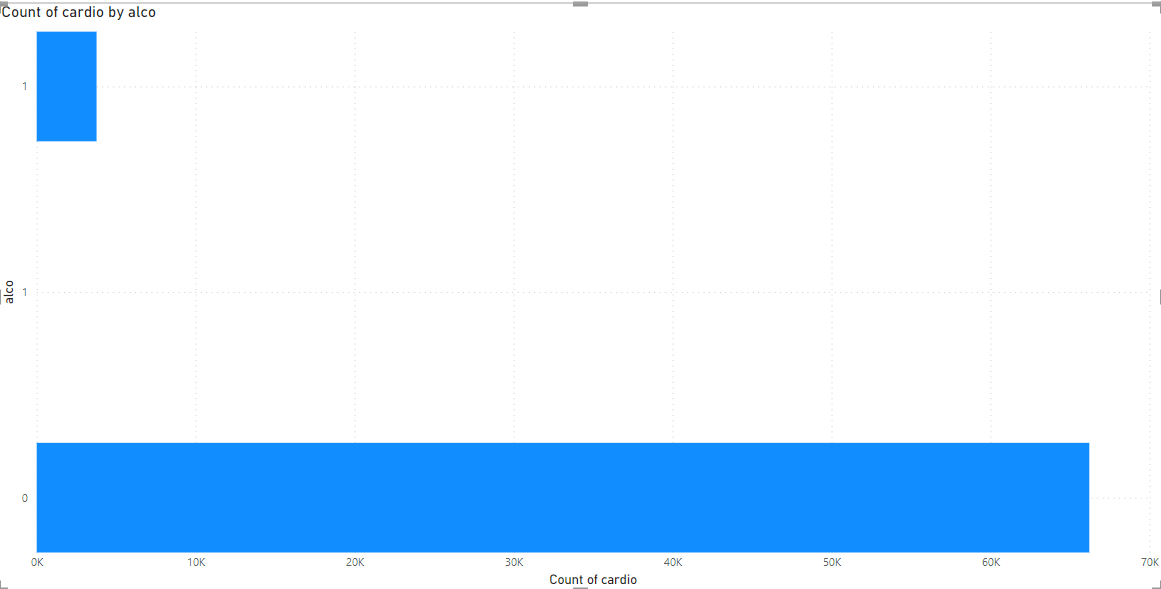
**Mengatasi missing value**

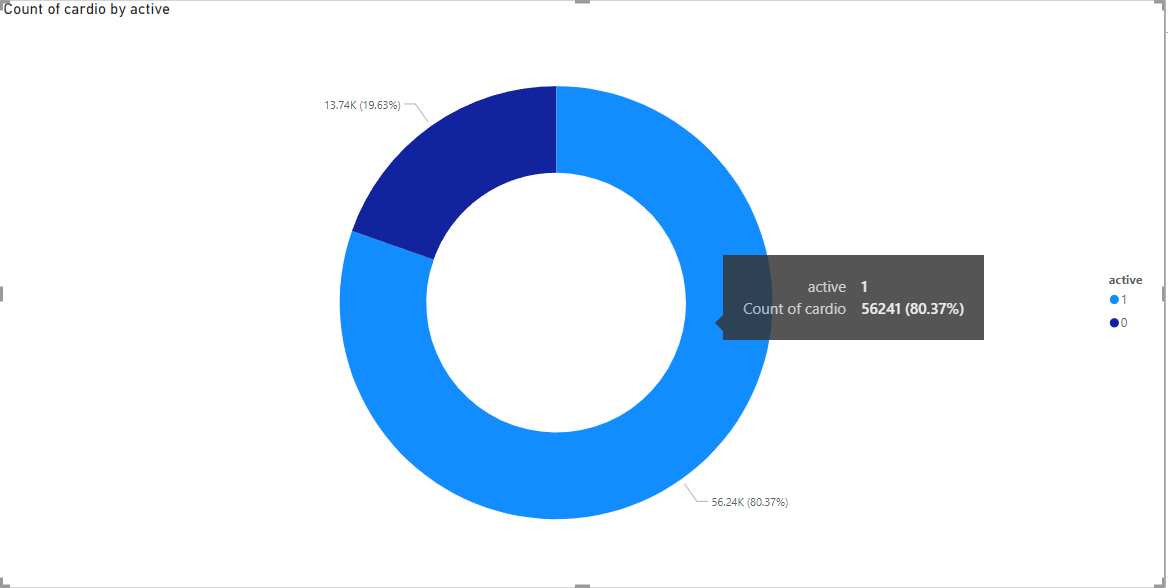
Missing value dapat diatasi pada PowerBI menggunakan fill down. Dikarenakan dataset sudah terisi semua dan tidak ditemukan adanya missing value, maka peneliti tidak melewati tahap ini

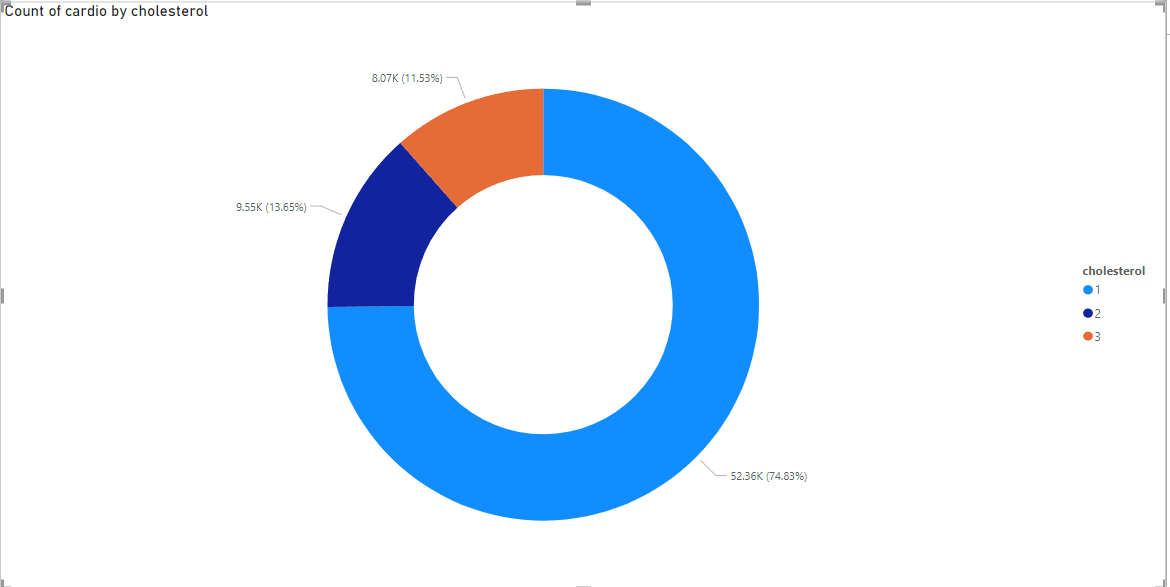
**Balancing data dan visualisasi**

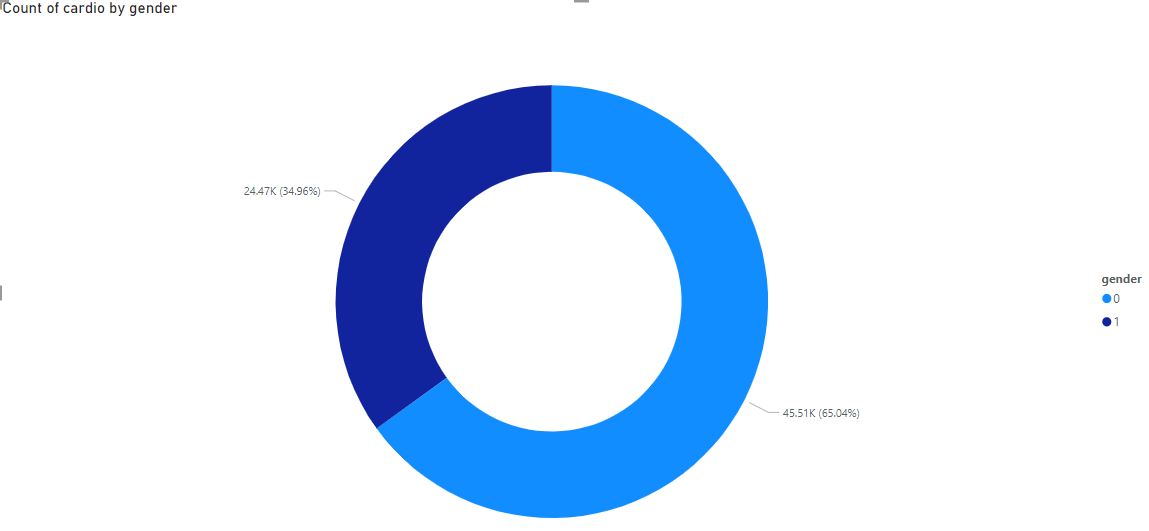
1. Glukosa



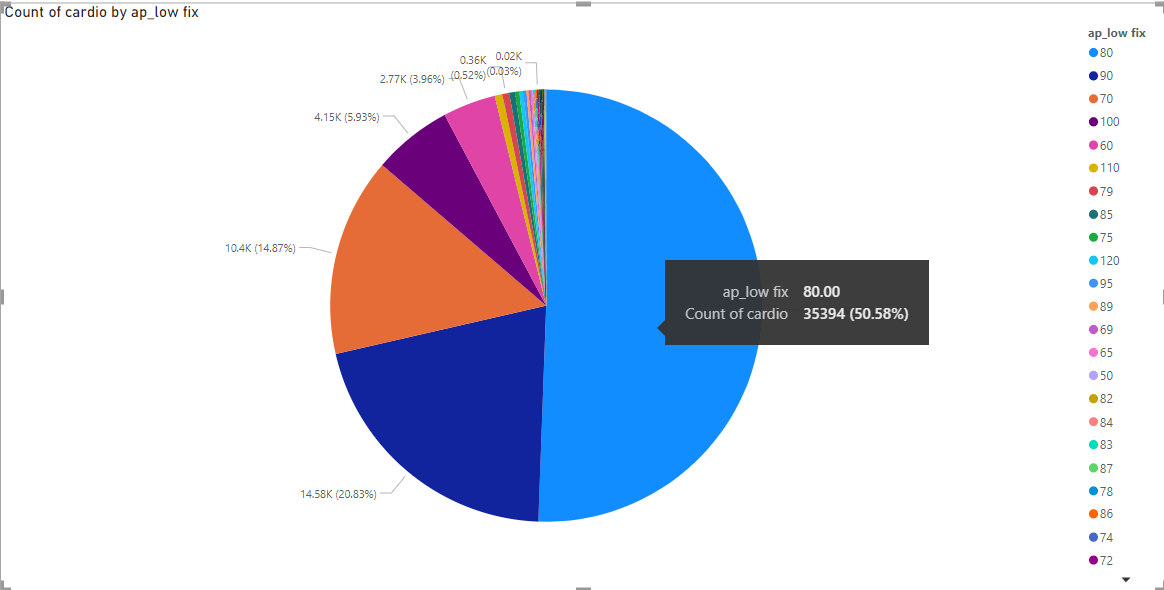
1. Alcohol
2. Physical active

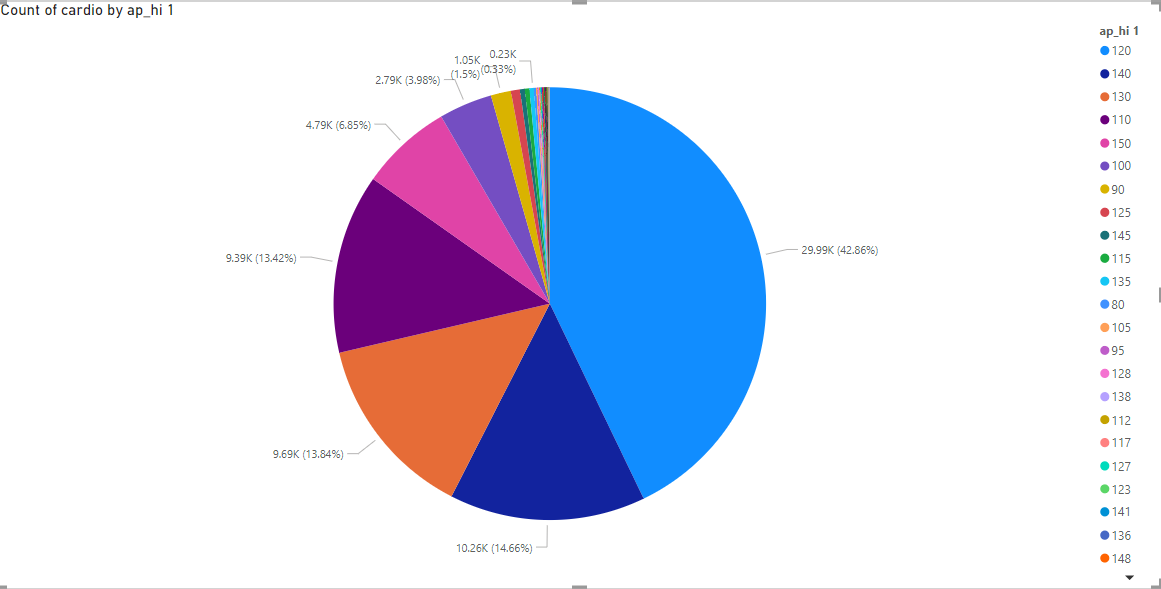


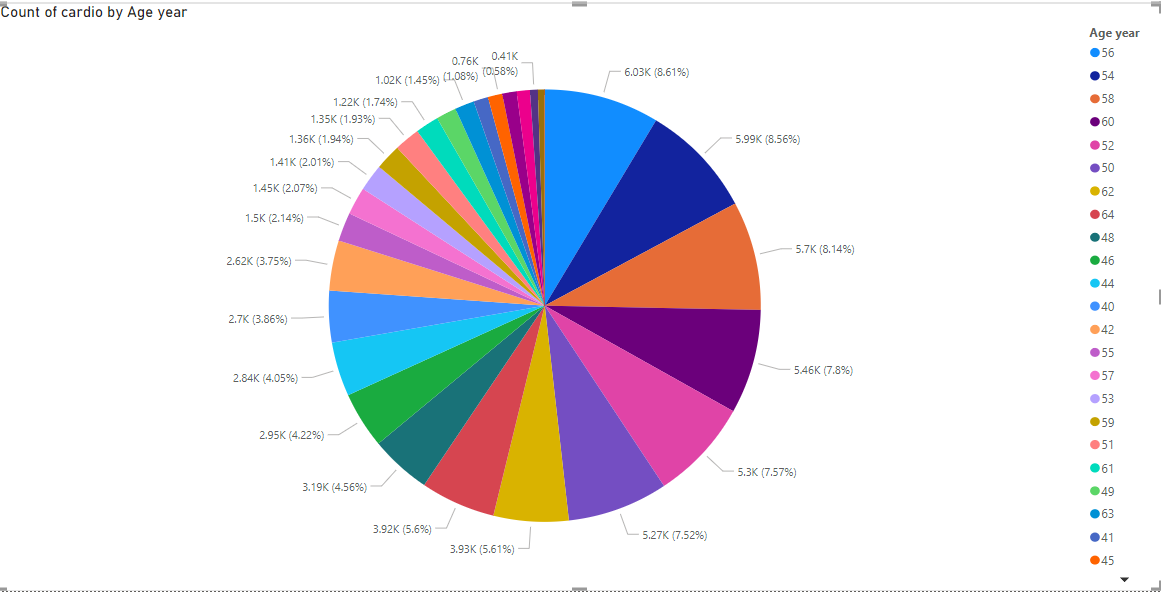
1. Cholesterol 
2. Gender



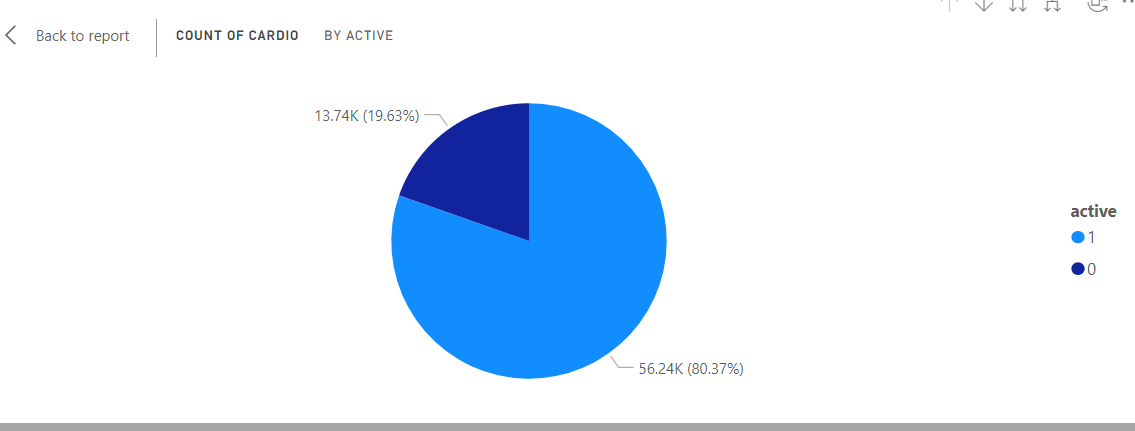
1. Systolic



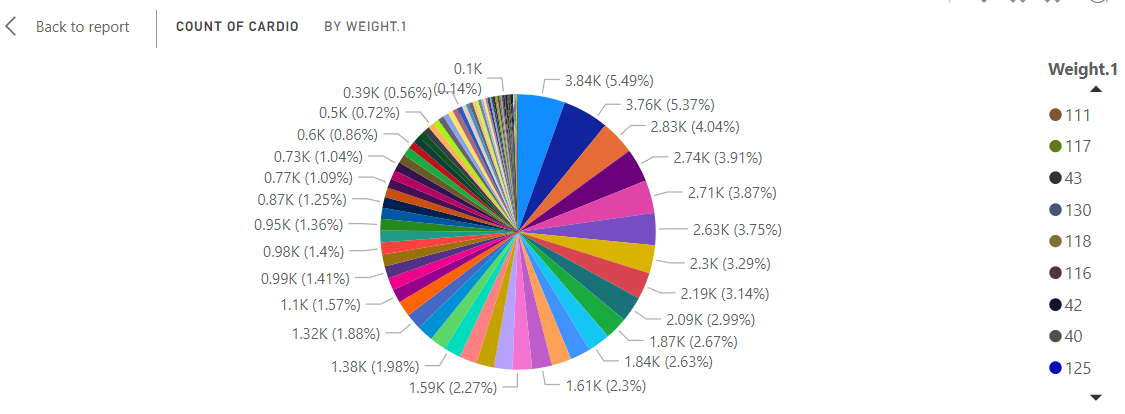
1. Diastolic   
   
2. Age



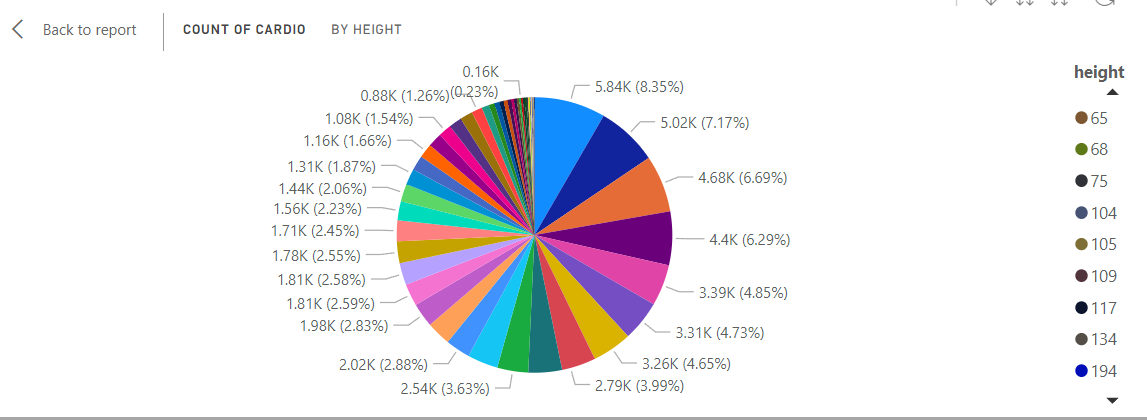
1. Smoker



1. Berat Badan



1. Tinggi badan



Daftar Pustaka

<https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset>

https://powerbi.microsoft.com/en-us/