Tugas 1  
PROYEK SAINS DATA

yang diampu oleh:

*Eka Mala Sari Rochman, S.Kom M.Kom*



**Oleh:**

Amril Viqrirachman (210411100065)

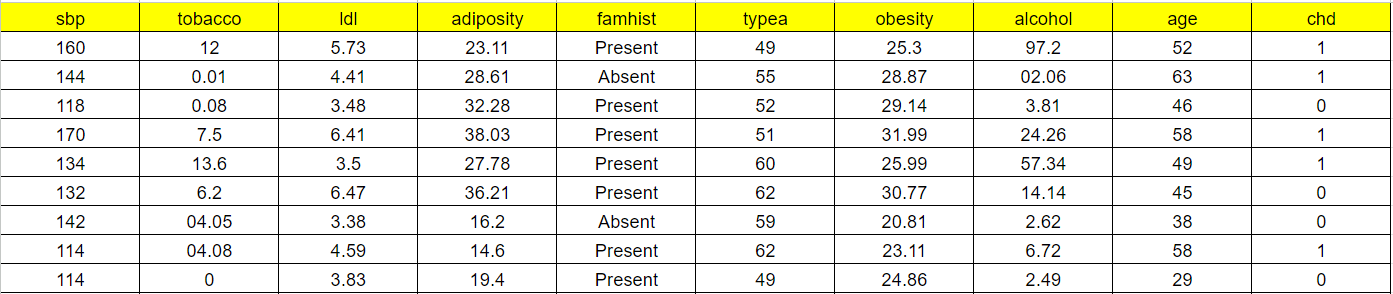
PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS TRUNOJOYO MADURA

**Coronary Heart Disease Prediction**

Dataset ini terdiri dari 10 variabel independen yaitu sbp(systolic blood pressure), tobacco, ldl(low-density lipoprotein), adiposity, famhist(family health history), typea(type A personality score), obesity, alcohol, age, dan chd(coronary heart disease). Terdiri atas 462 data. Parameter ini memainkan peran penting dalam analisis prediktif jantung koroner.



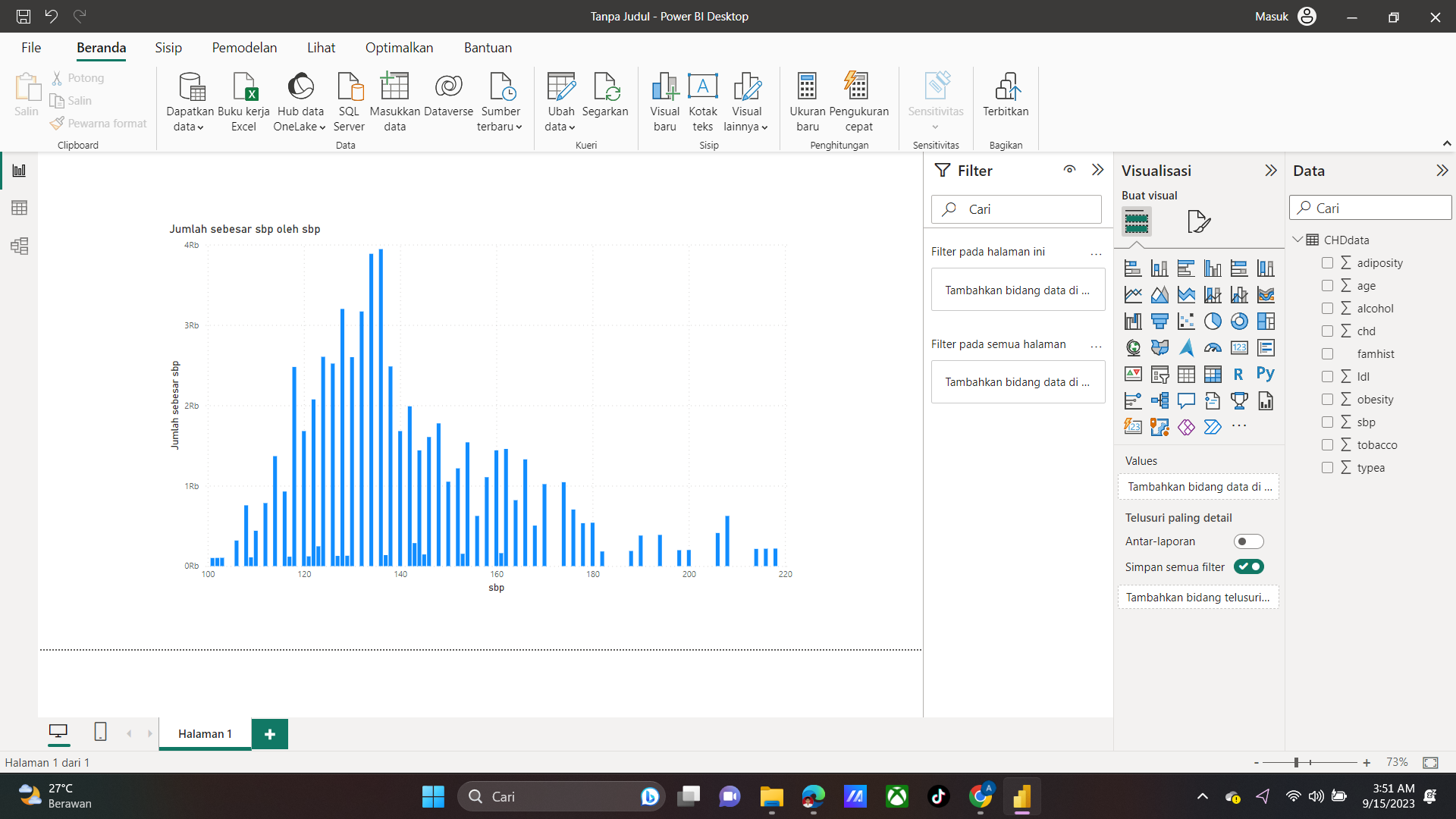
Variabel targetnya adalah diagnosa penyakit jantung koroner

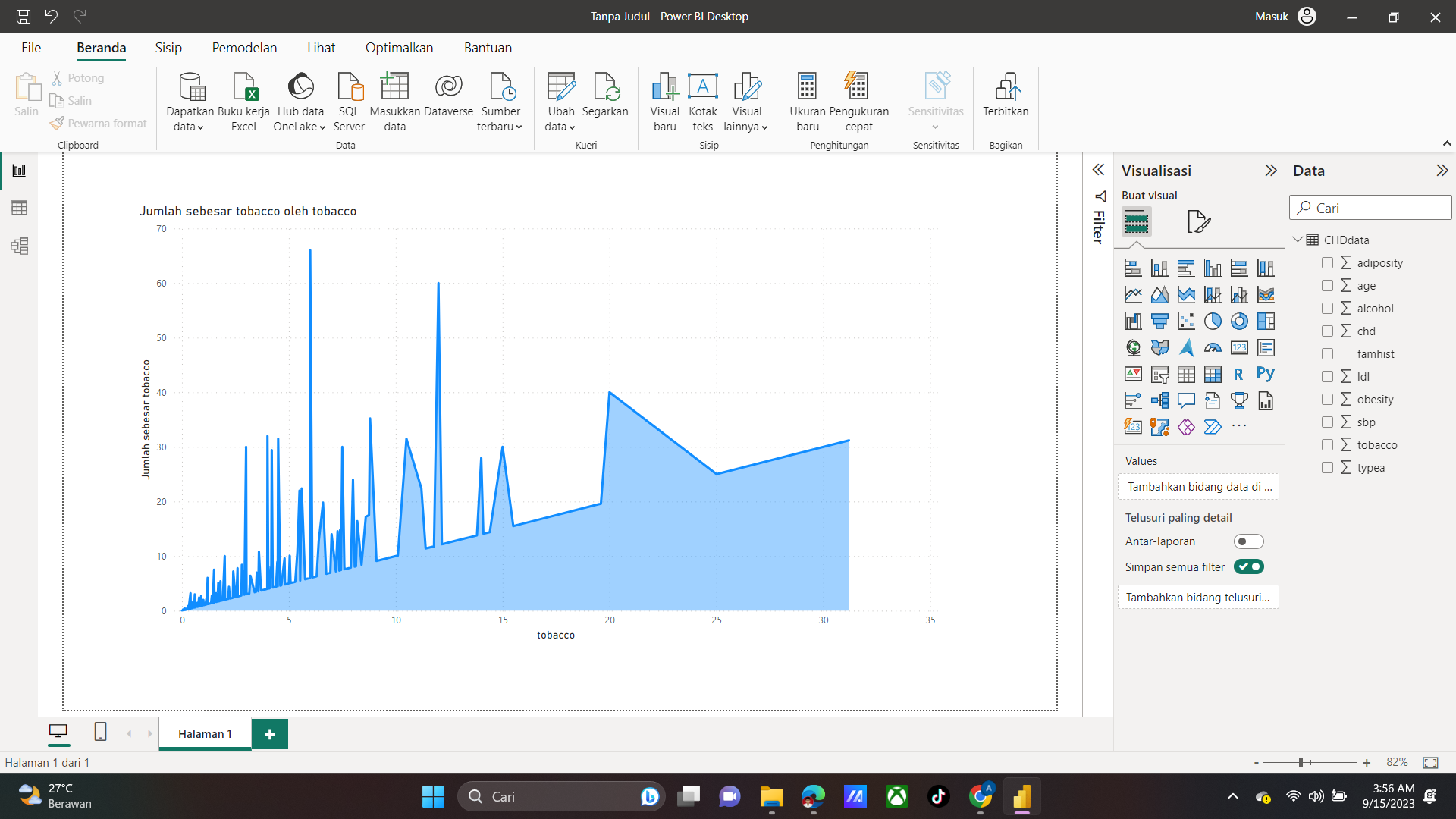
0 - Negative  
 1 - Positive

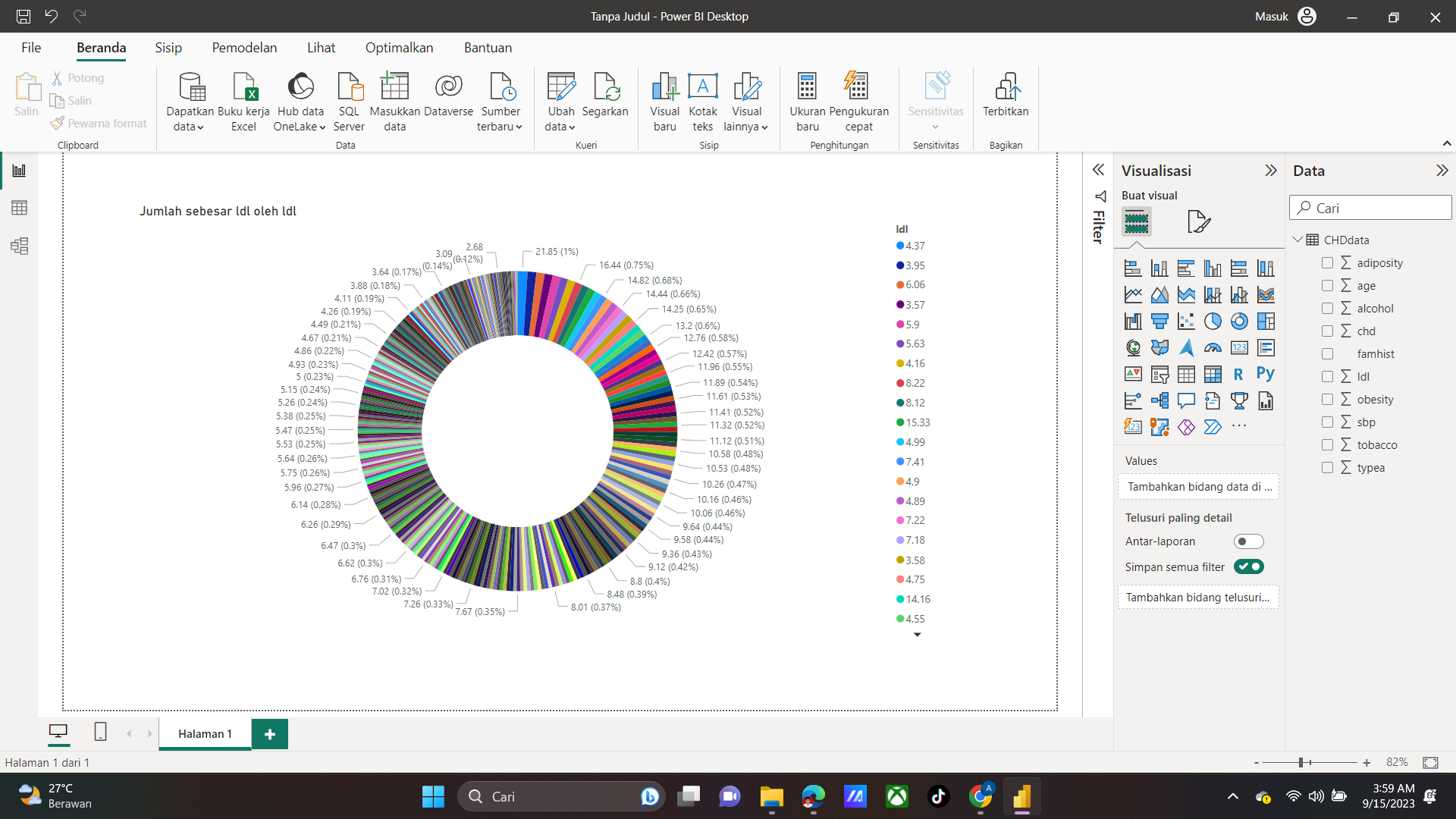
1. SBP  
   Systolic Blood Pressure, Tekanan darah merujuk kepada tekanan yang dialami darah pada pembuluh arteri darah ketika darah dipompa oleh jantung ke seluruh anggota tubuh manusia. Tekanan darah dibuat dengan mengambil dua ukuran dan biasanya diukur seperti berikut - 120 /80 mmHg.
2. Tobacco  
   Tembakau.
3. LDL  
   Low Density Lipoprotein (LDL) adalah salah satu jenis kolesterol. Adanya kolesterol dalam batas normal tidak berbahaya bagi tubuh. Sedangkan kadar LDL yang terlalu tinggi bisa menimbulkan masalah pada kesehatan. Itu sebabnya LDL dikenal juga sebagai kolesterol jahat.
4. Adiposity

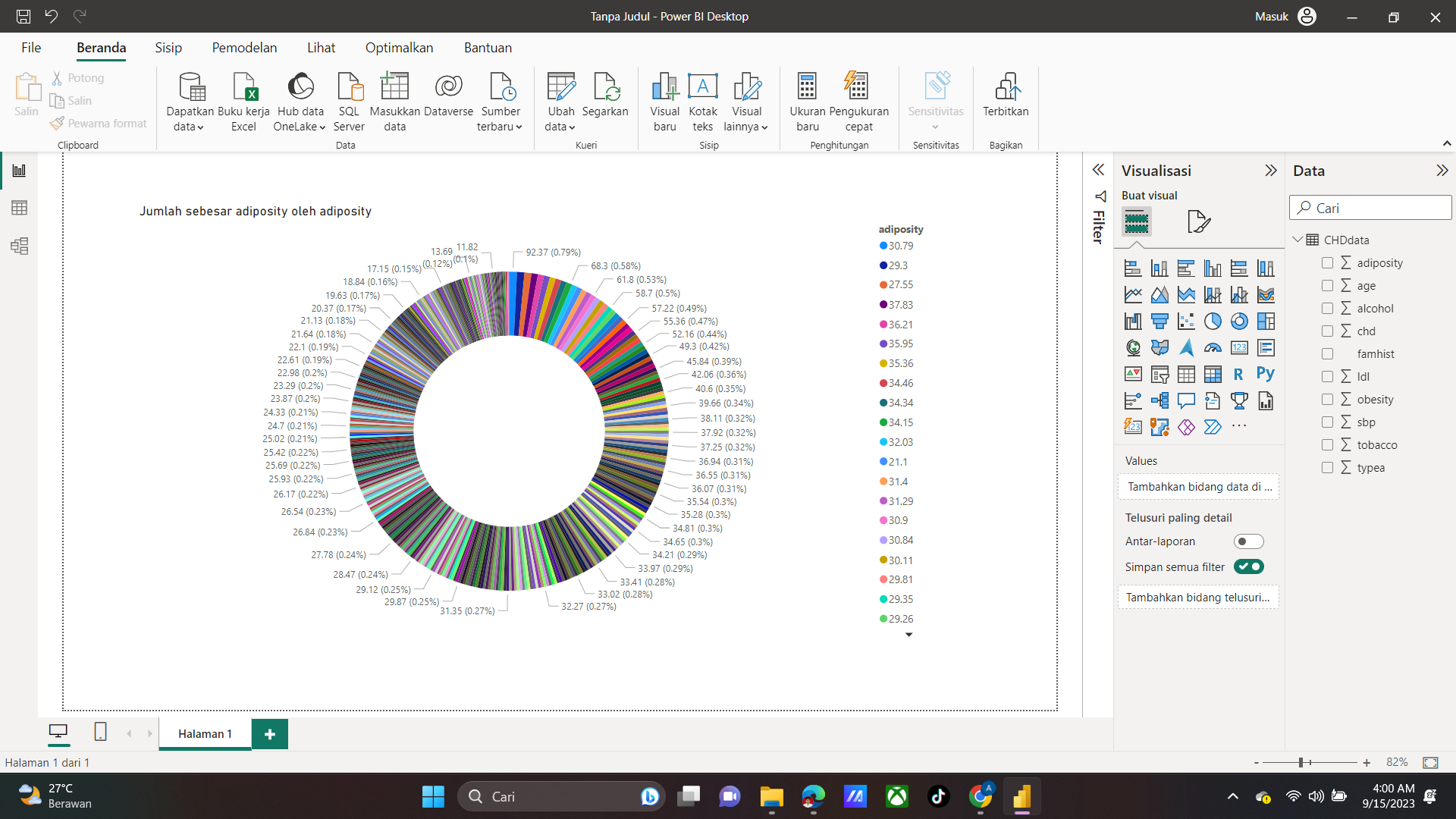
Keadaan gemuk, obesitas.

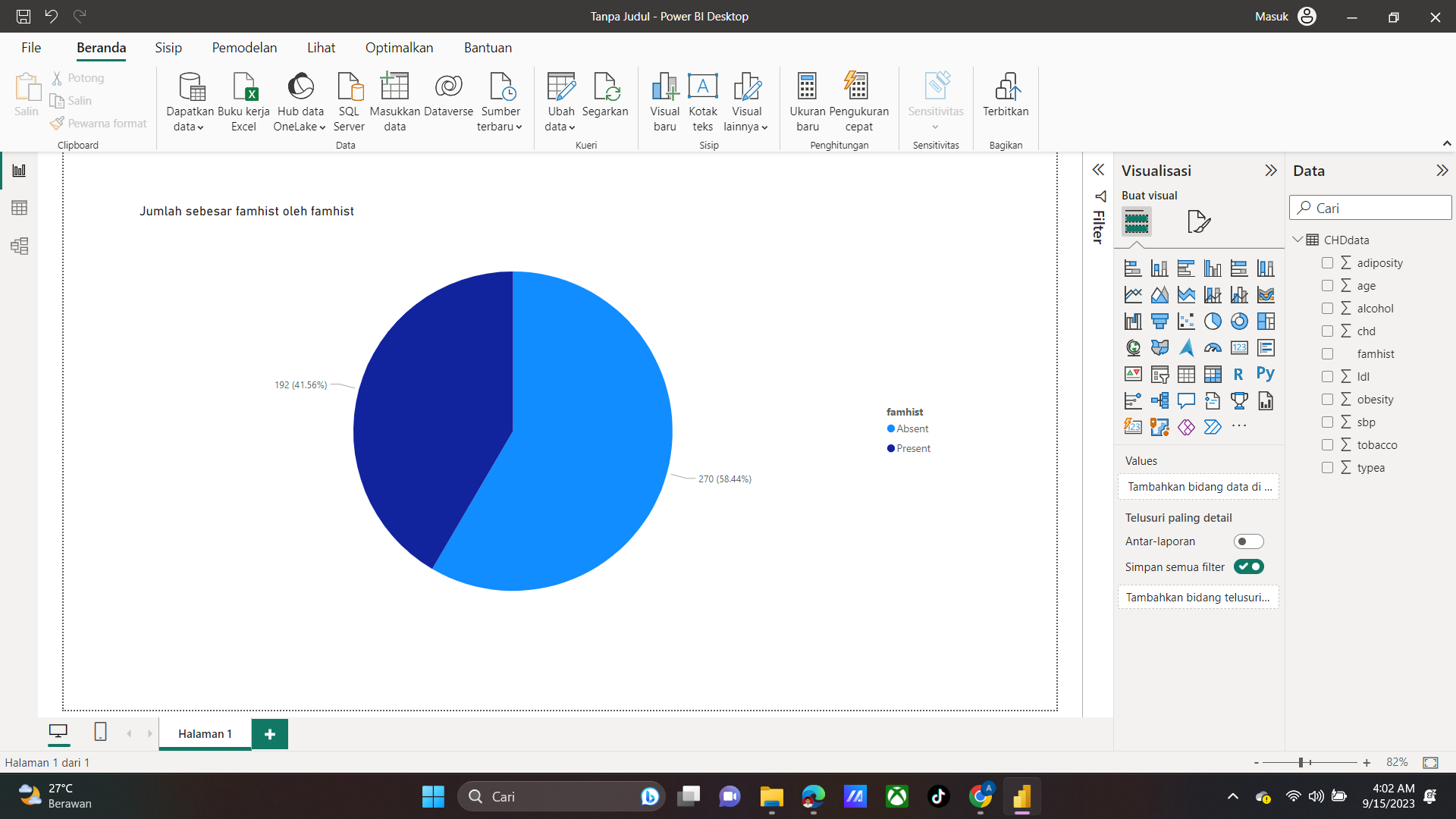
1. Famhist  
   Family health
2. Typea  
   Kepribadian tipe A cenderung lebih sukses terkena 90% serangan jantung.
3. Obesity  
   Kegemukan dan obesitas diartikan sebagai penumpukan lemak yang tidak normal atau berlebihan yang menimbulkan risiko terhadap kesehatan.
4. Alcohol
5. Age
6. CHD  
   Coronary Heart Disease
7. **Analisis data Menggunakan Power BI**

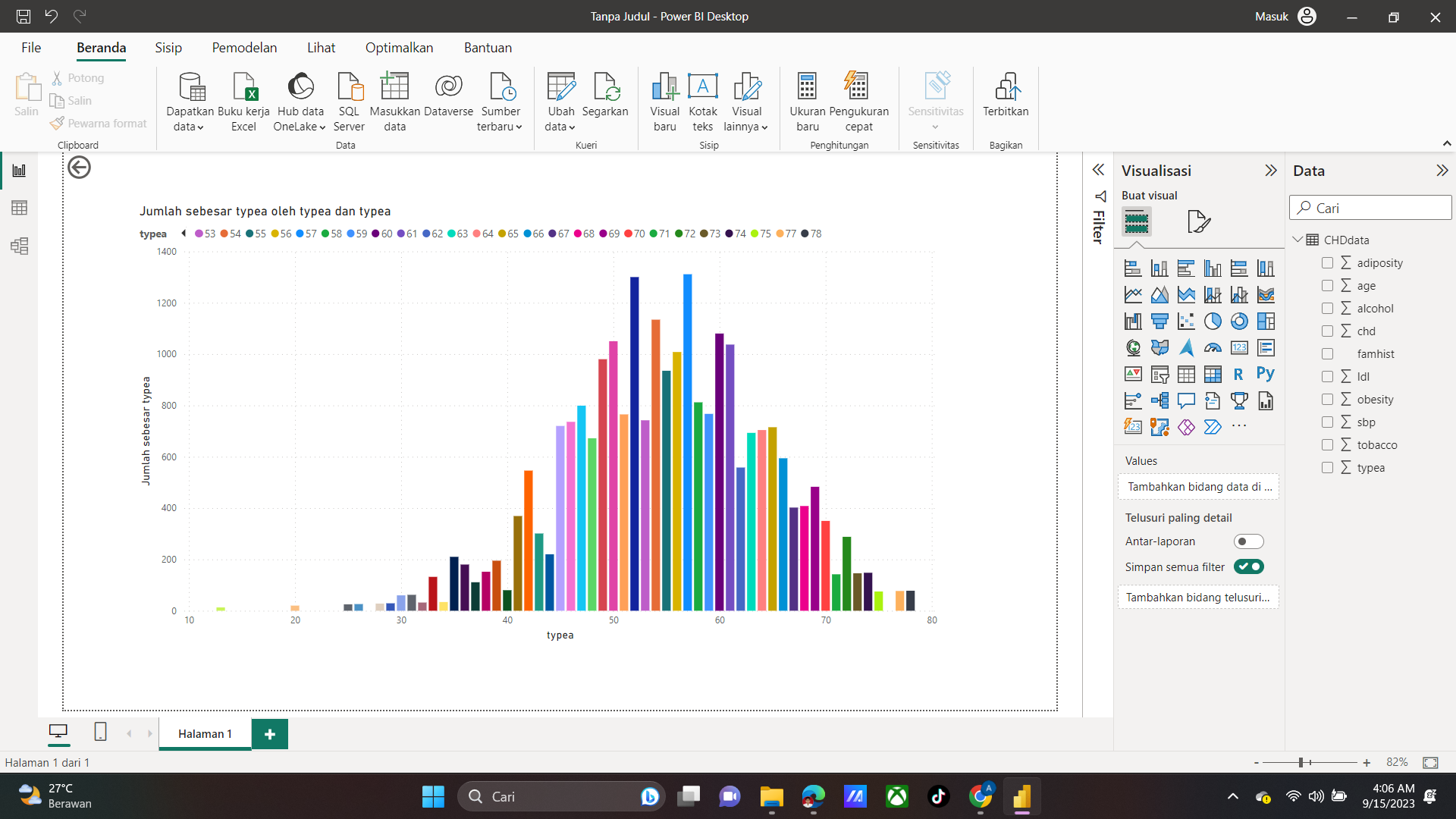
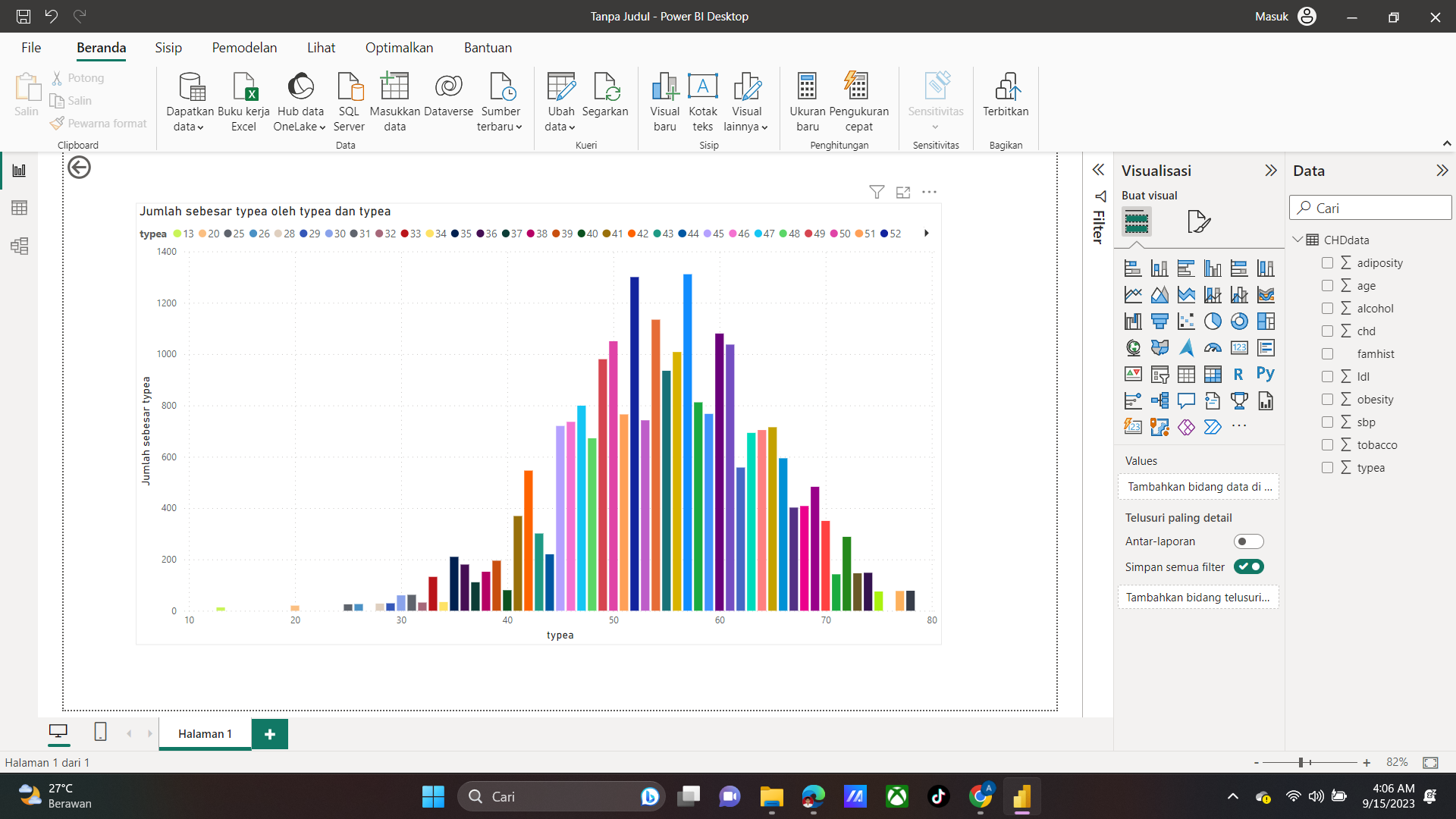


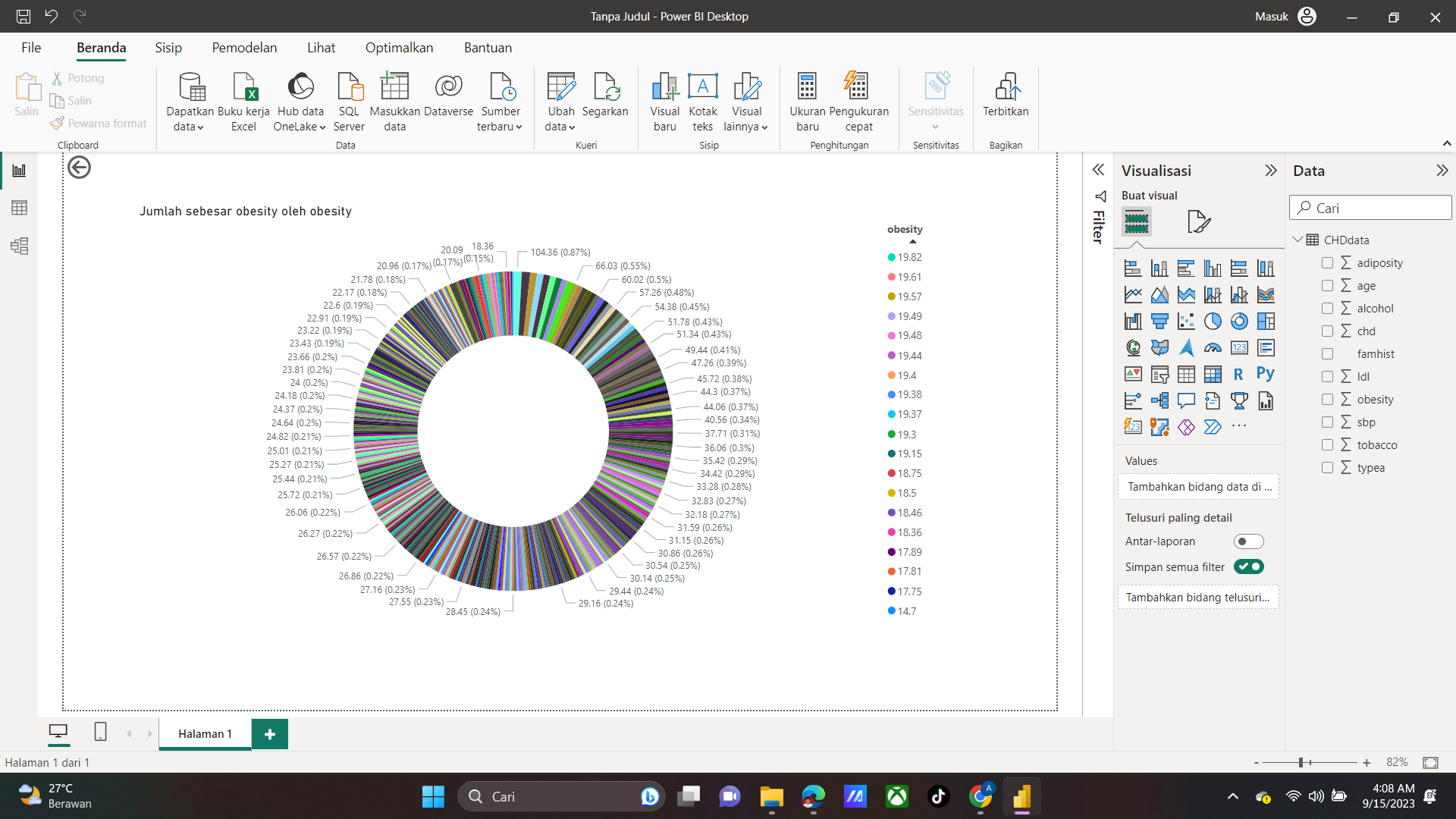
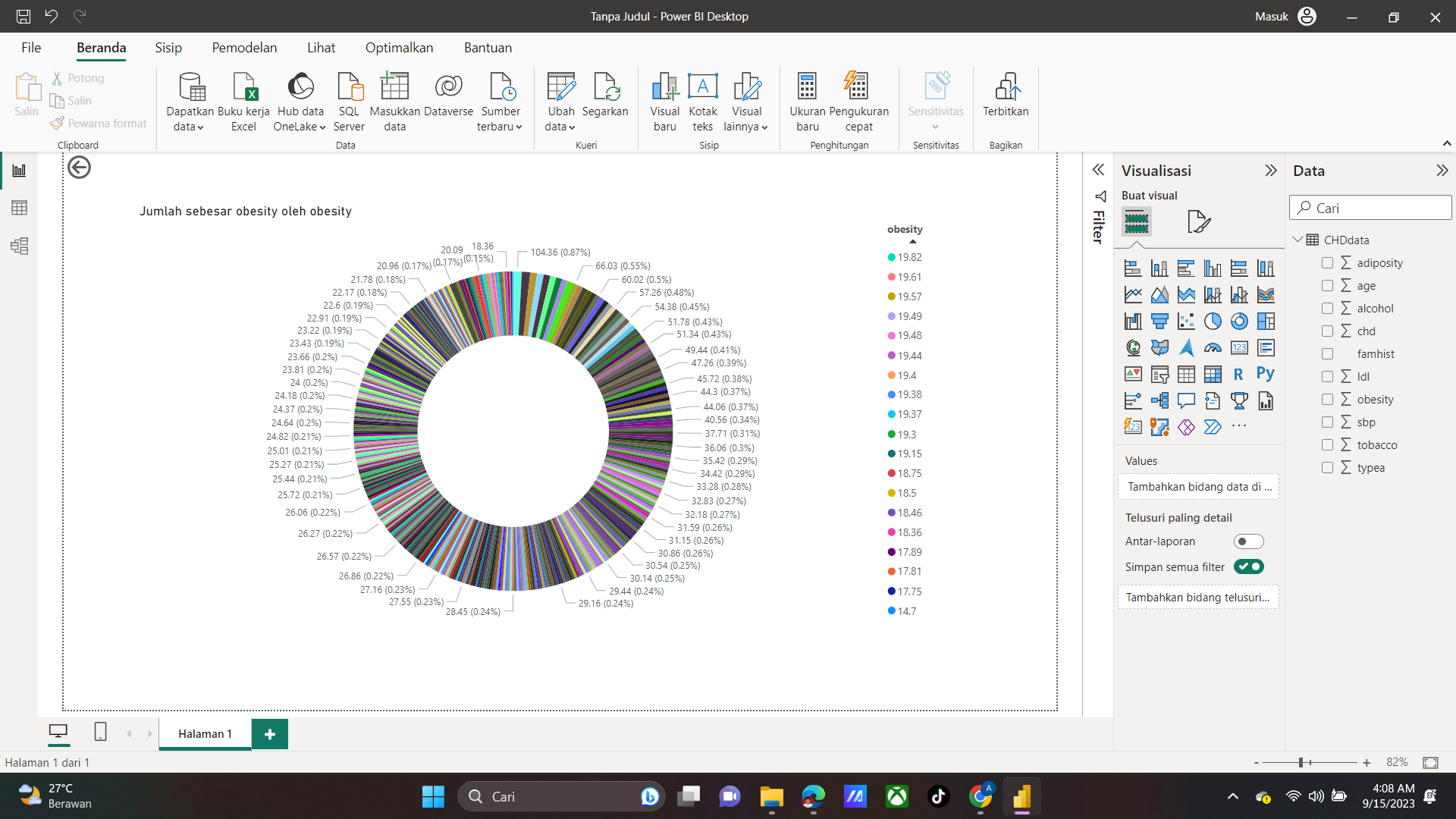


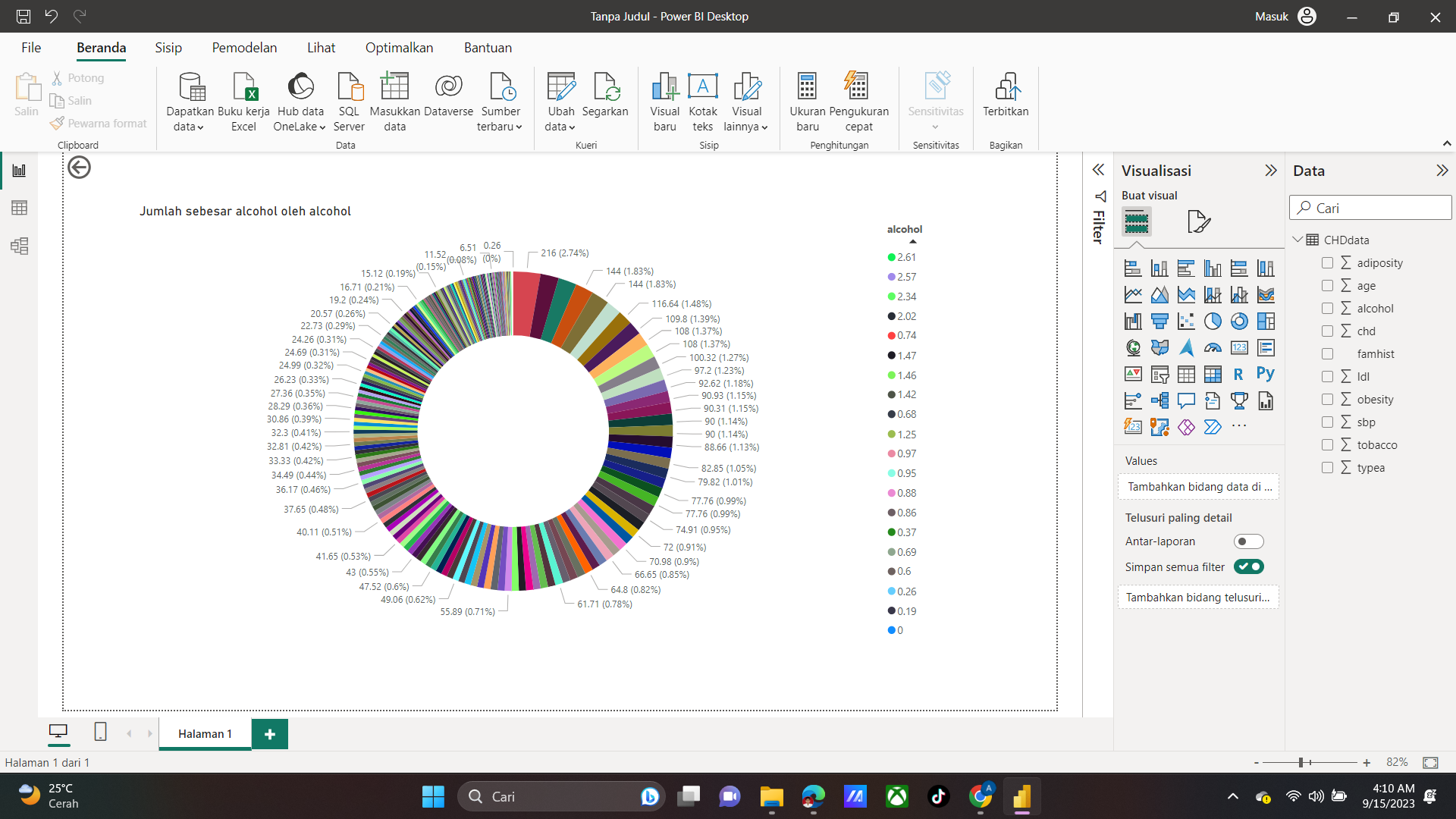


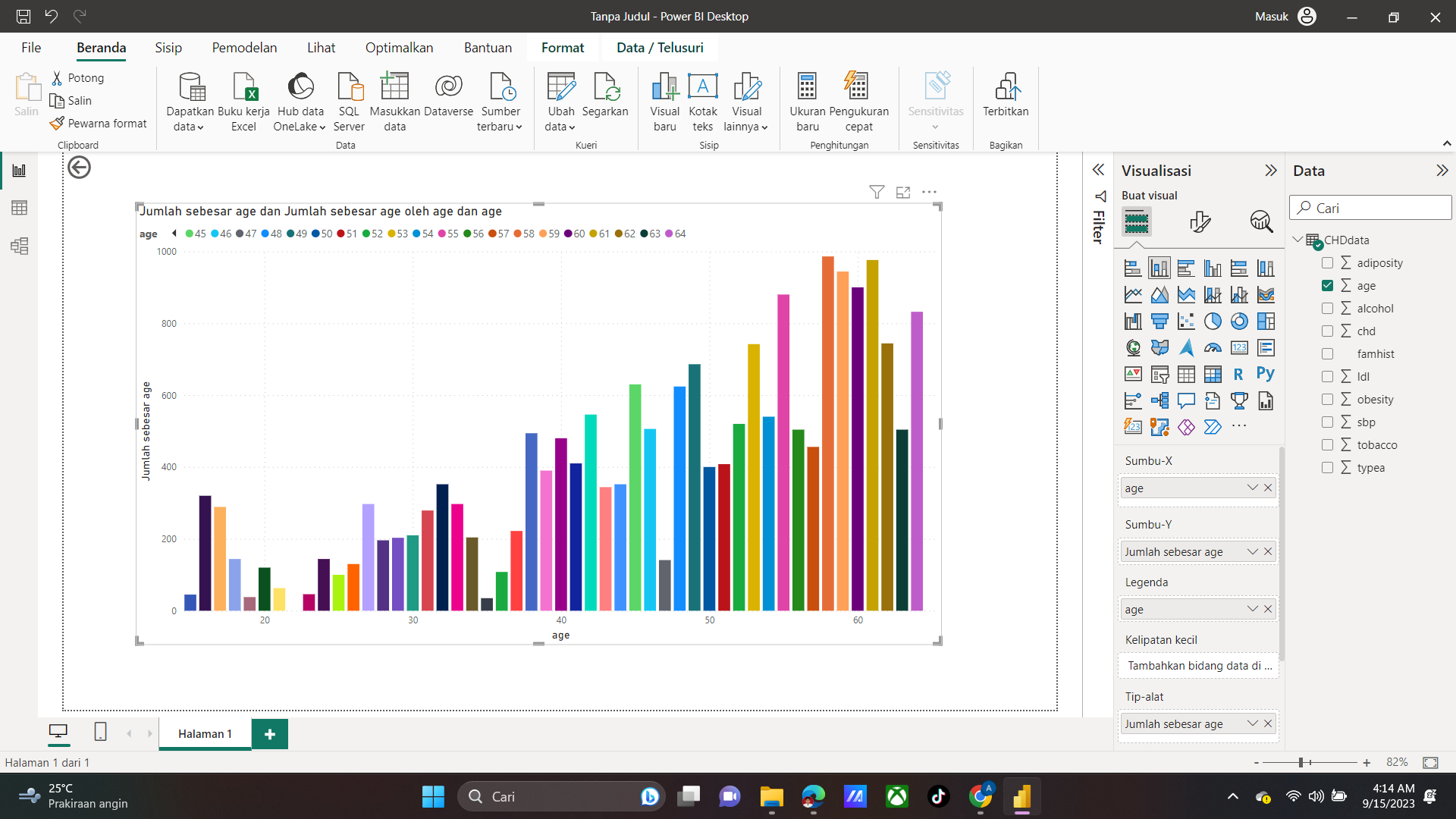
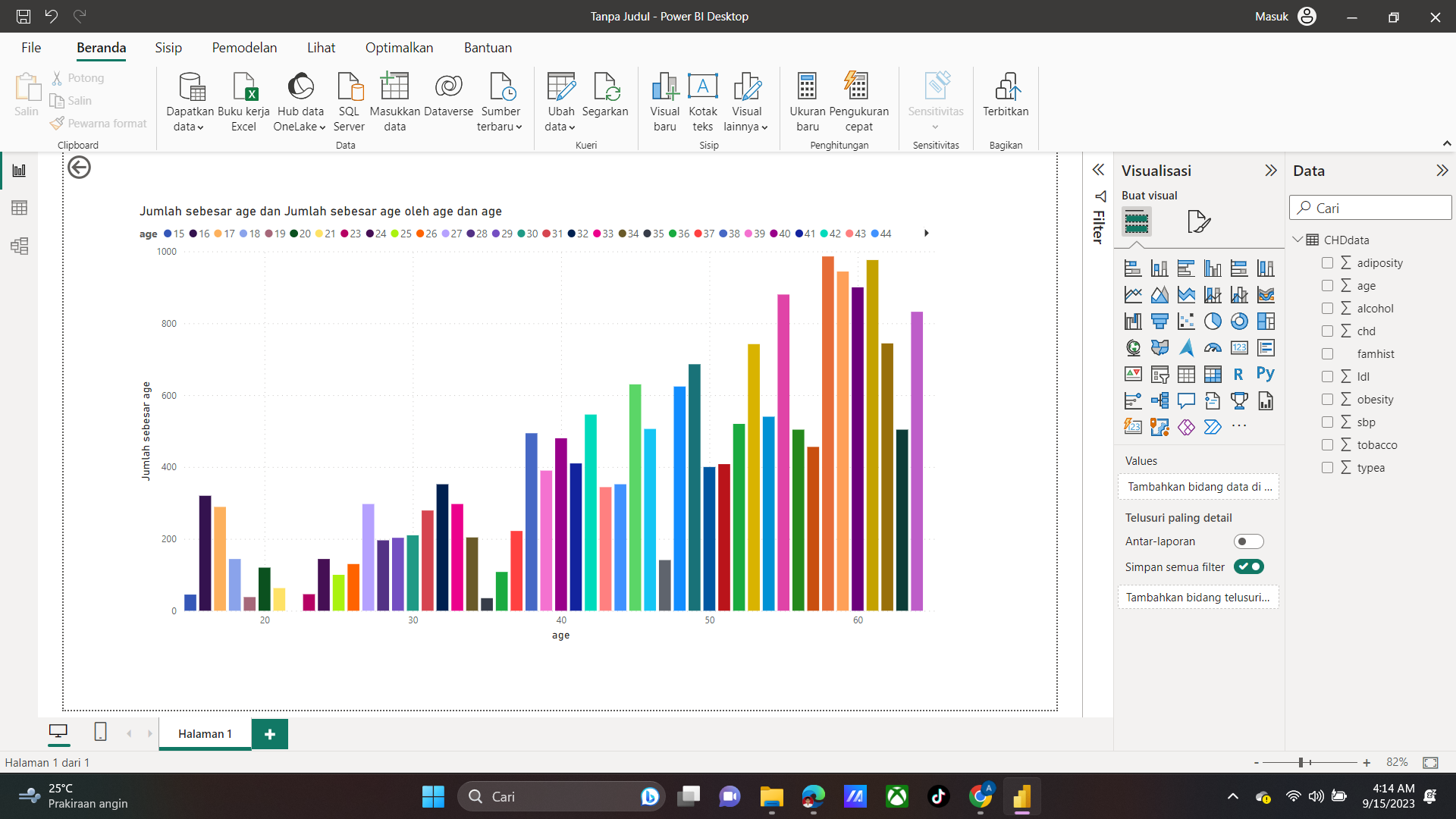






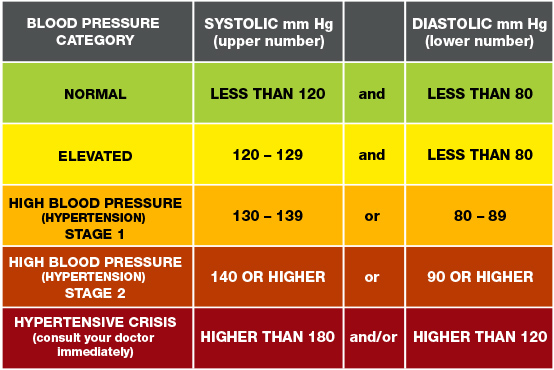






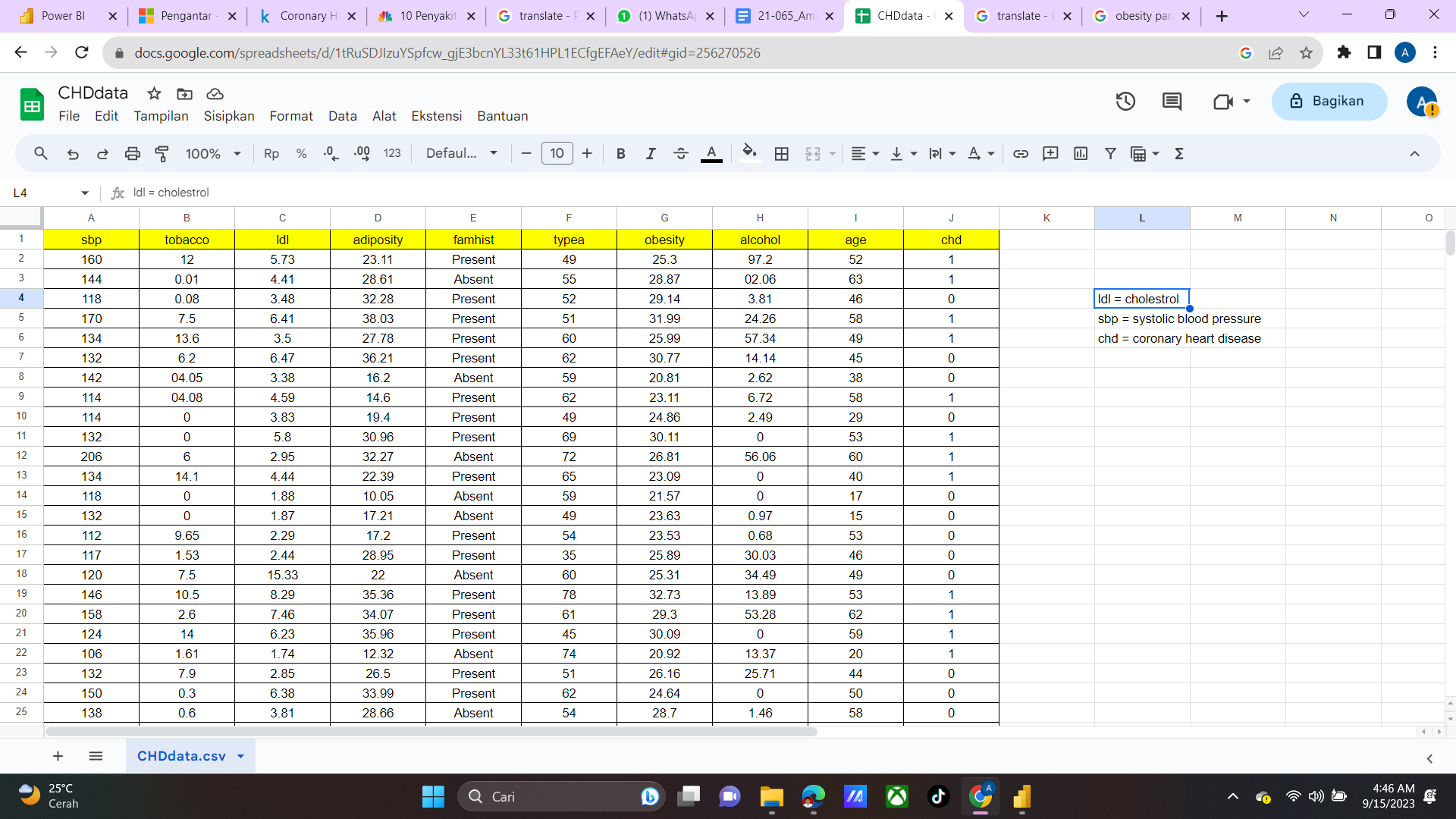


1. **Analisis Parameter**

* SBP  
  Berikut penjelasan tingkat SBP:  
  
* Tobacco  
  Penggunaan tembakau per-tahun dalam satuan kilogram(kg)
* Low Density Lipoprotein
* Adiposity
* Famhist  
  192 41,56% Absent  
  270 58,44% Present
* Typea  
  > 110 Type A1  
  80-109 Type A2  
  60-79 Type AB  
  < 59 Type B  
  30-50 Type B2  
  0-29 Type B1
* Obesity  
  18,5 - < 25 Sehat  
  25 - < 30 Overweight  
  30 - > Obesity
* Alcohol
* Age(umur)
* Coronary Heart Disease  
  160 34,63% Positive  
  302 65,37% Negative

1. **Cleaning Data**

Terdapat Parameter grade biner



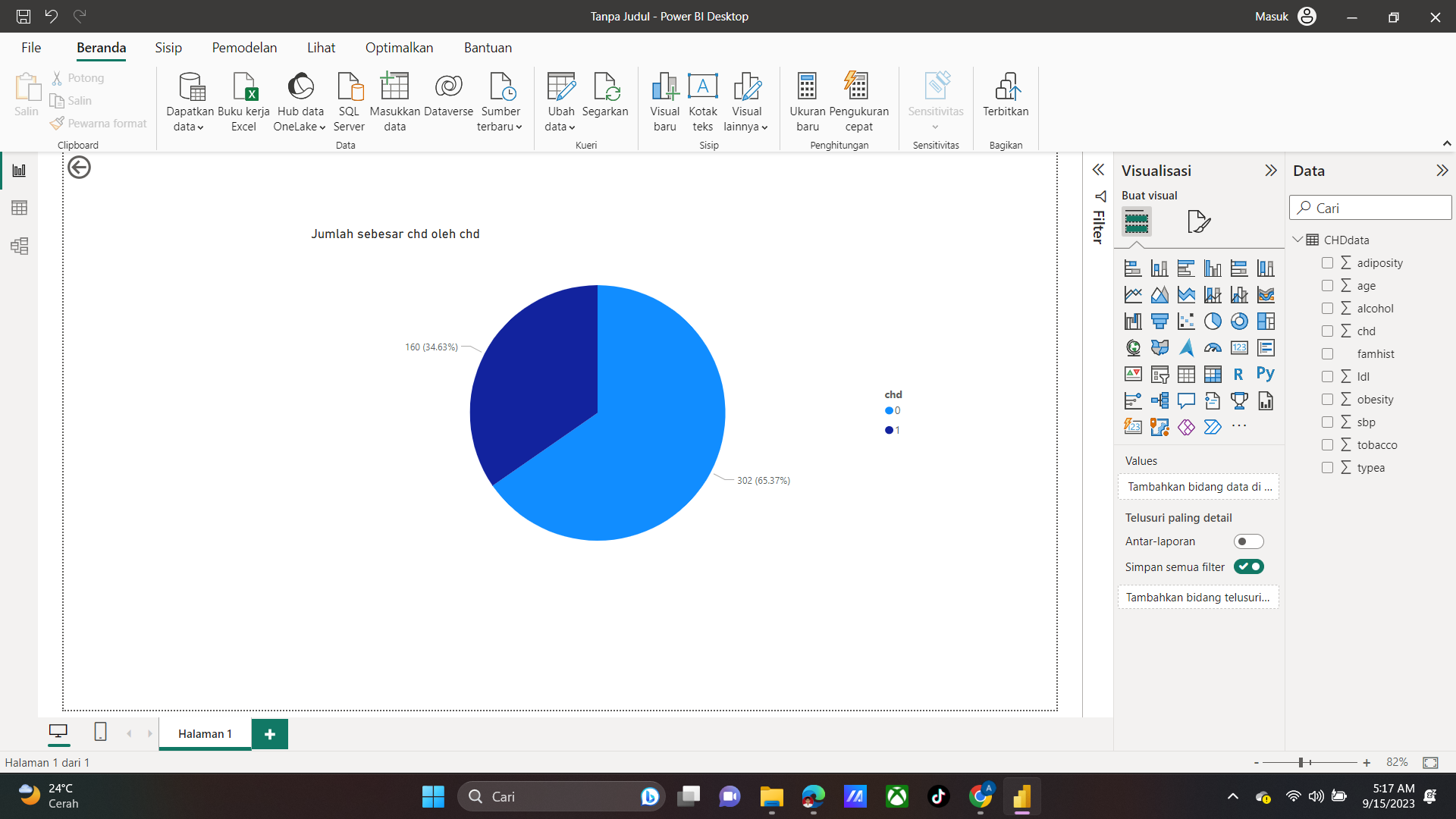
Sehingga dapat disimpulkan untuk grade

**Positive: 1**

**Negative : 0**

1. **Missing Data**

Tidak terdapat missing data dalam data Milk quality production

1. **Balancing Data**Imbalanced class dapat diartikan sebagai adanya rasio yang tidak proporsional di setiap kelas di dalam dataset. Macam-macam algoritma imbalanced class yaitu : a. Under sampling Menyeimbangkan dataset dengan mengurangi ukuran kelas mayor. Metode ini digunakan ketika jumlah data mencukupi.  
   Dataset yang digunakan pada analisis kali ini belum *balance* dikarenakan jumlah data diabetes yaitu 0 dan 1 tidak seimbang seperti pada gambar di bawah ini:  
     
   Teknik Oversampling Minoritas Sintetis (SMOTE) adalah teknik statistik untuk meningkatkan jumlah kasus dalam himpunan data dengan cara yang seimbang. Komponen bekerja dengan menghasilkan instans baru dari kasus minoritas yang ada yang Anda berikan sebagai input.  
   