Nama : Aderisa Dyta Okvianti

NIM : 200411100013

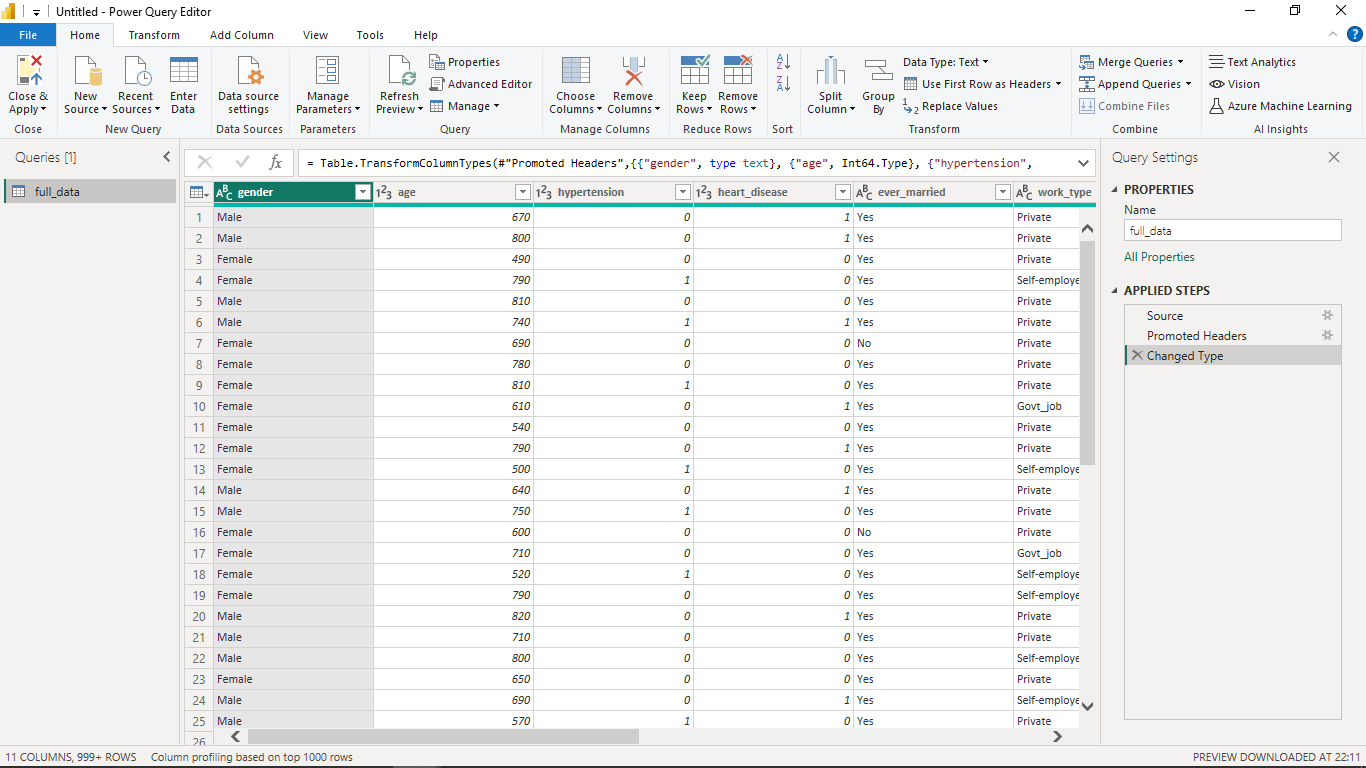
Kelas : PSD E

Tugas :

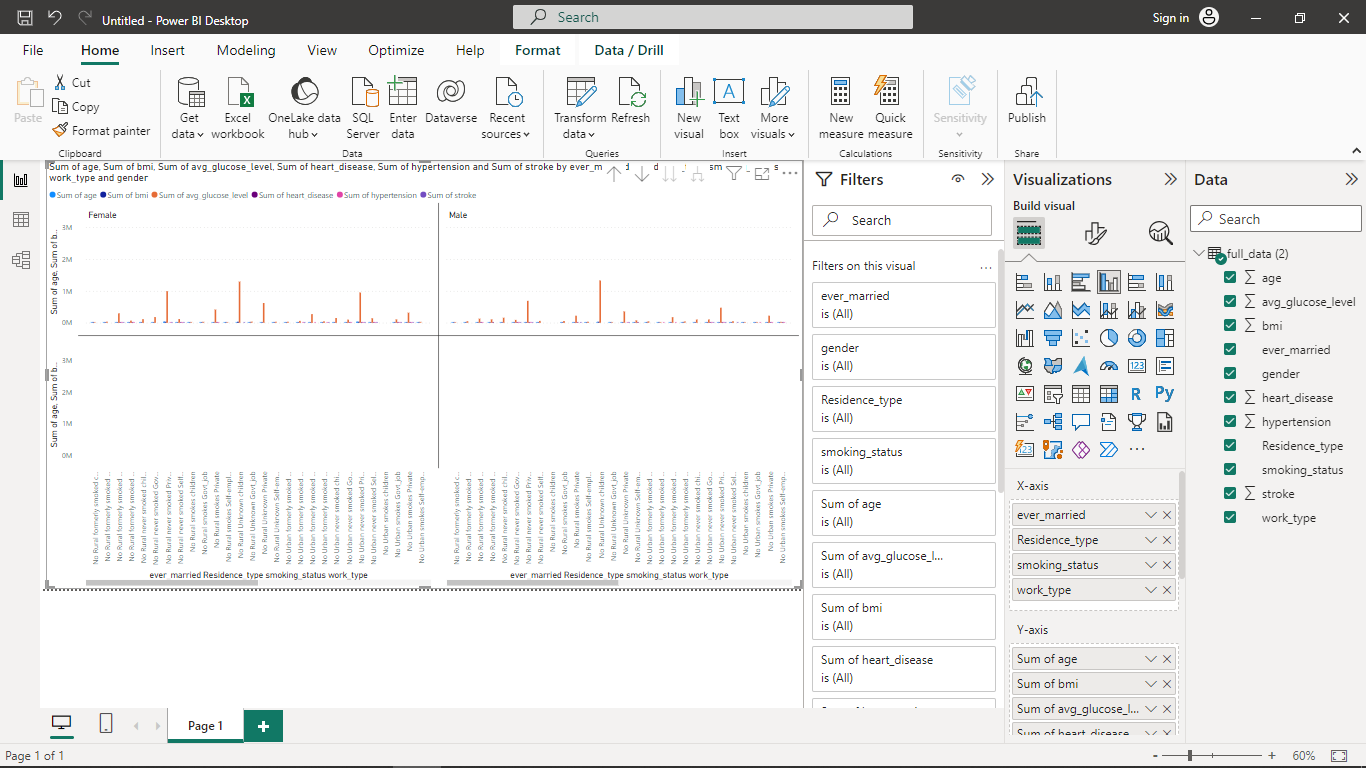
1. Analisis data

Data yang saya gunakan adalah data penyakit **Brain stroke prediction dataset** berasal dari situs kaggle <https://www.kaggle.com/datasets/zzettrkalpakbal/full-filled-brain-stroke-dataset>. Stroke adalah suatu kondisi medis di mana aliran darah yang buruk ke otak menyebabkan kematian sel. Dataset tersebut terdiri dari 9 fitur dan 4981 kolom.

1. Menganalisis data dengan mengimport data yang sudah kita download dari kaggle kedalam power BI



1. kita bisa menganalisis data mulai dari transform data, menampilkan diagram, visualisasi data dan banyak operasi lainnya.

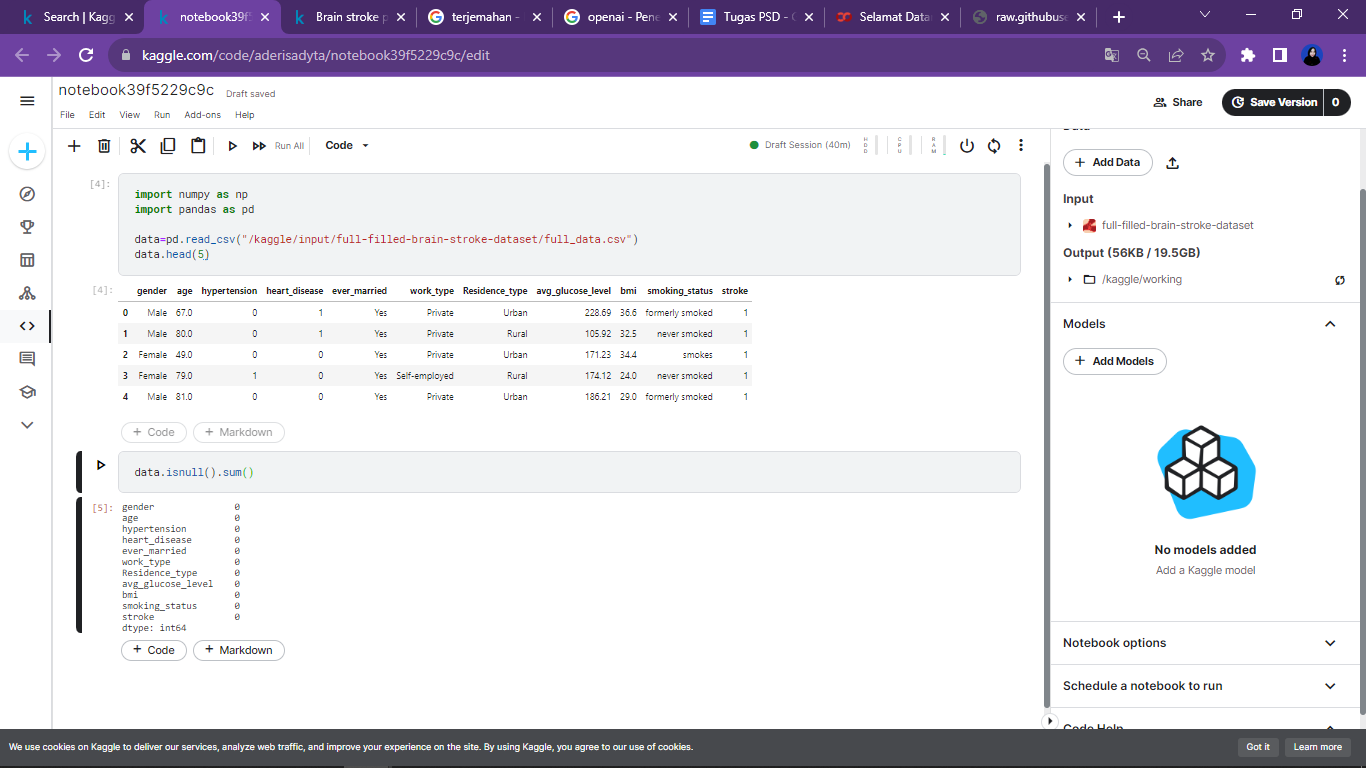


1. Analisis hubungan antar parameter yang ada
2. Jumlah fitur = 9

* Age = Fitur ini mengacu pada usia
* Gender = Fitur ini mengacu pada jenis kelamin (Male, Female)
* Hypertension = Fitur ini mengacu apakah peserta tersebut menderita hipertensi atau tidak (0 jika pasien tidak menderita hipertensi, 1 jika pasien menderita hipertensi)
* Heart\_disease = Fitur ini mengacu apakah peserta ini menderita penyakit jantung atau tidak (0 jika pasien tidak mempunyai penyakit jantung, 1 jika pasien mempunyai penyakit jantung)
* Ever married = Fitur ini mewakili status perkawinan peserta (No untuk peserta yang belum menikah, yes untuk peserta yang sudah menikah)
* Work type = Fitur ini mewakili status pekerjaan peserta(children, Govt\_jov, Never\_worked, Private, Self-employed)
* Residence type = Fitur ini mewakili status pekerjaan peserta (Urban, Rural)
* Avg glucose level = Fitur ini mengacu tingkat glukosa rata-rata peserta (kadar glukosa rata-rata dalam darah)
* BMI = Fitur ini mengacu indeks massa tubuh peserta
* Smoking Status = Fitur ini mengacu status merokok pada peserta (formerly smoked, never smoked,smokes,Unknown)
* Stroke = Fitur ini mengacu apakah peserta menderita stroke atau tidak (1 jika pasien terkena stroke atau 0 jika tidak)

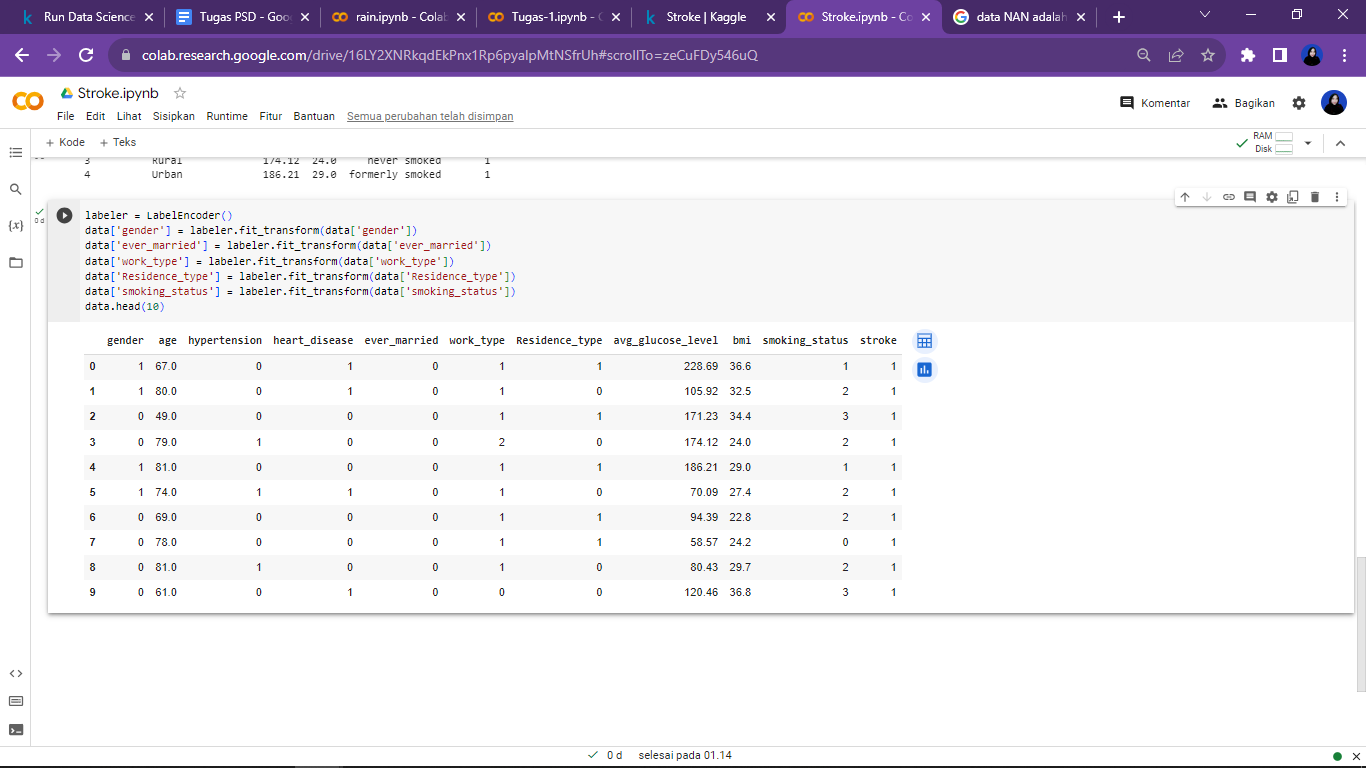
1. Lakukan proses cleaning data dan melihat apakah ada missing data atau tidak

cara mengatasi missing data bisa dengan nilai 1 atau 0, nilai mean atau modus, dengan machine learning. Dataset **Brain stroke prediction** sudah berbentuk data cleaning, karena tidak ada missing value yang terdeteksi setelah saya lakukan operasi mencari missing value pada kaggle.



1. Ubah data string/kategori menjadi numerik

Pada dataset **Brain stroke prediction** ada beberapa fitur yang masih dalam bentuk kategori yaitu gender, ever\_married, work\_type, Residence\_type, smoking\_status , saya melakukan perubahan dari data yang berbentuk kategori menjadi nominal



1. Balance

Pada dataset **Brain stroke prediction** di fitur stroke terdapat 2 class yaitu data pasien yang terdeteksi penyakit stroke (1) sebanyak 248 dan yang tidak terdeteksi penyakit stroke (0) sebanyak 4733 data. Berarti data ini belum seimbang maka perlu dilakukan proses balancing dengan menggunakan smoot

