**PROYEK SAINS DATA**

**“Implementasi Power BI Pada Data Breats Cancer”**

**Dosen Pengampu :**

EKA MALA SARI ROCHMAN, S.Kom.M.Kom

****

**Disusun Oleh :**

Jhon Filius Gultom\_210411100019

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS TRUNOJOYO MADURA**

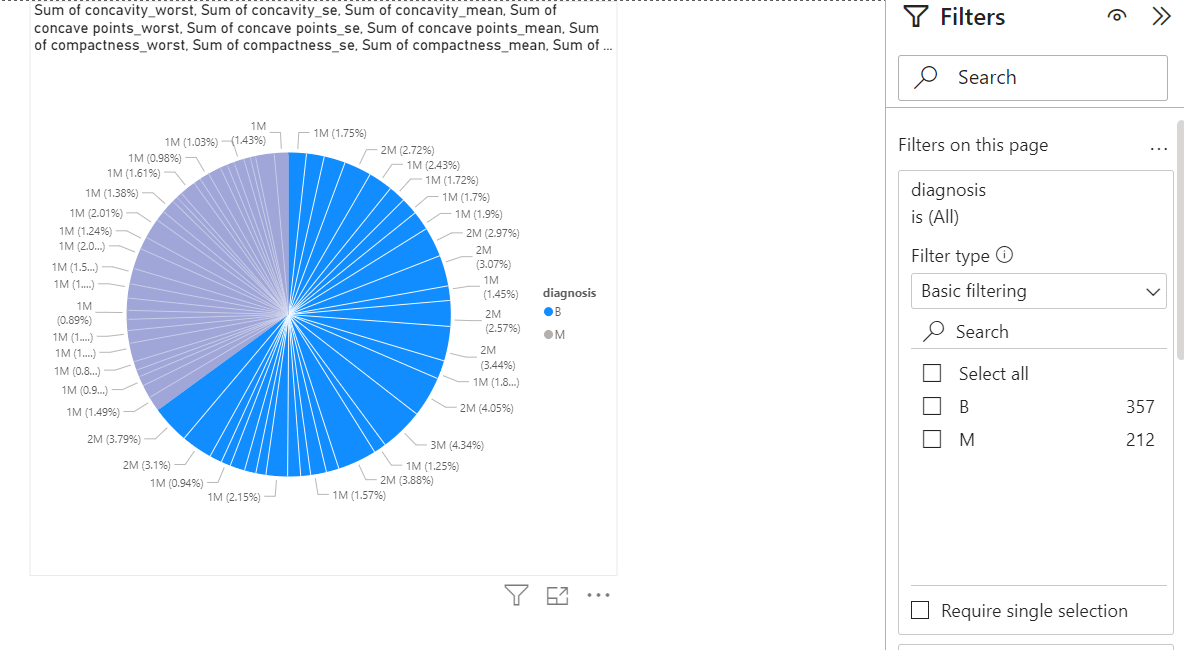
**2023**

***Implementasi Power BI Pada Data Breats Cancer***

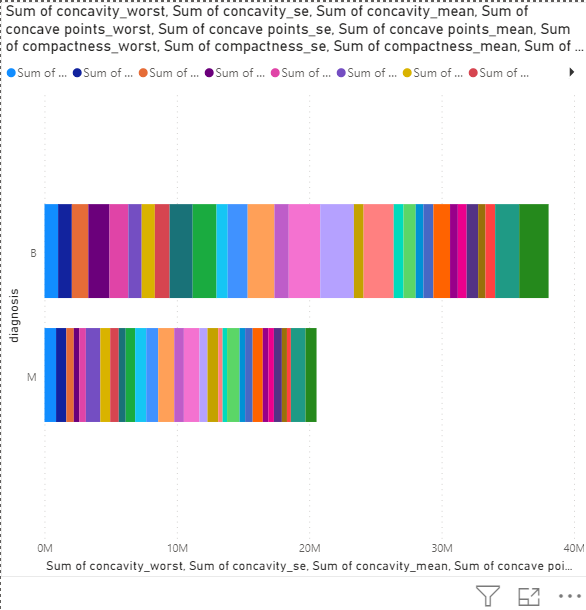
Kanker payudara (breast cancer) adalah jenis kanker yang terjadi ketika sel-sel ganas (kanker) berkembang di dalam jaringan payudara. Ini adalah salah satu jenis kanker yang paling umum terjadi pada wanita, meskipun jarang juga dapat terjadi pada pria. Kanker payudara dapat muncul di berbagai bagian payudara, termasuk dalam kelenjar susu atau dalam saluran susu.

1. **MELIHAT SEBARAN DATA PADA POWER BI**

Sebaran data (data distribution) adalah konsep dalam statistik yang merujuk kepada cara data terdistribusi atau tersebar di sekitar nilai-nilai tertentu dalam satu set data. Ini memberikan informasi tentang bagaimana data-data dalam kumpulan tersebut terdistribusi atau terletak di sepanjang berbagai nilai-nilai. Pemahaman sebaran data adalah penting dalam analisis statistik karena membantu kita memahami karakteristik dan pola data, serta dapat digunakan untuk membuat keputusan atau inferensi statistik.

  
*Gambar Sebaran Data menggunakan Pie chart*

Gambar Sebaran data menggunakan Pie Chart tersebut menampilkan hasil seluruh sebaran data, pada data yang saya gunakan, yang berarti menampilkan semua isi column, dimana target nya adalah hasil diagonis yaitu B yang artinya pasien mengidap penyakit kanker payudara jinak dan M yang artinya kanker payudara ganas , yang dimana terdapat hasil B sebanyak 357, dan M sebanyak 212. Data sebaran tersebut memiliki kaitan dengan data fitur lainnya yaitu sebagai data pendukung untuk menentukan hasil diagonisnya.

  
*Gambar sebaran data Menggunakan Stacked Bar chart*

Stacked bar chart (grafik batang bertumpuk) adalah jenis visualisasi data yang digunakan untuk menggambarkan sebaran data dalam kategori yang berbeda dengan cara yang memperlihatkan kontribusi relatif dari setiap sub-kategori ke dalam kategori yang lebih besar. Ini berguna untuk membandingkan komposisi kategori-kategori tersebut. Gambar ini menampilkan hasil seluruh sebaran data, pada data yang saya gunakan, yang berarti menampilkan semua isi column, dimana target nya adalah hasil diagonis yaitu B yang artinya pasien mengidap penyakit kanker payudara jinak dan M yang artinya kanker payudara ganas , yang dimana terdapat hasil B sebanyak 357, dan M sebanyak 212. Data sebaran tersebut memiliki kaitan dengan data fitur lainnya yaitu sebagai data pendukung untuk menentukan hasil diagonisnya.

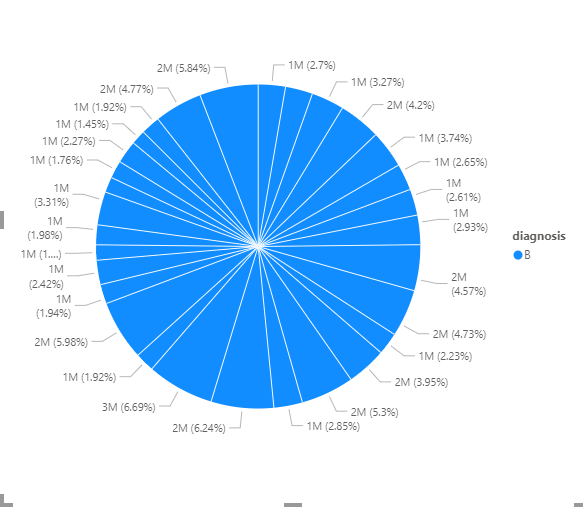
1. **ANALISIS DATA**

**HASIL ANALISIS DATA :**

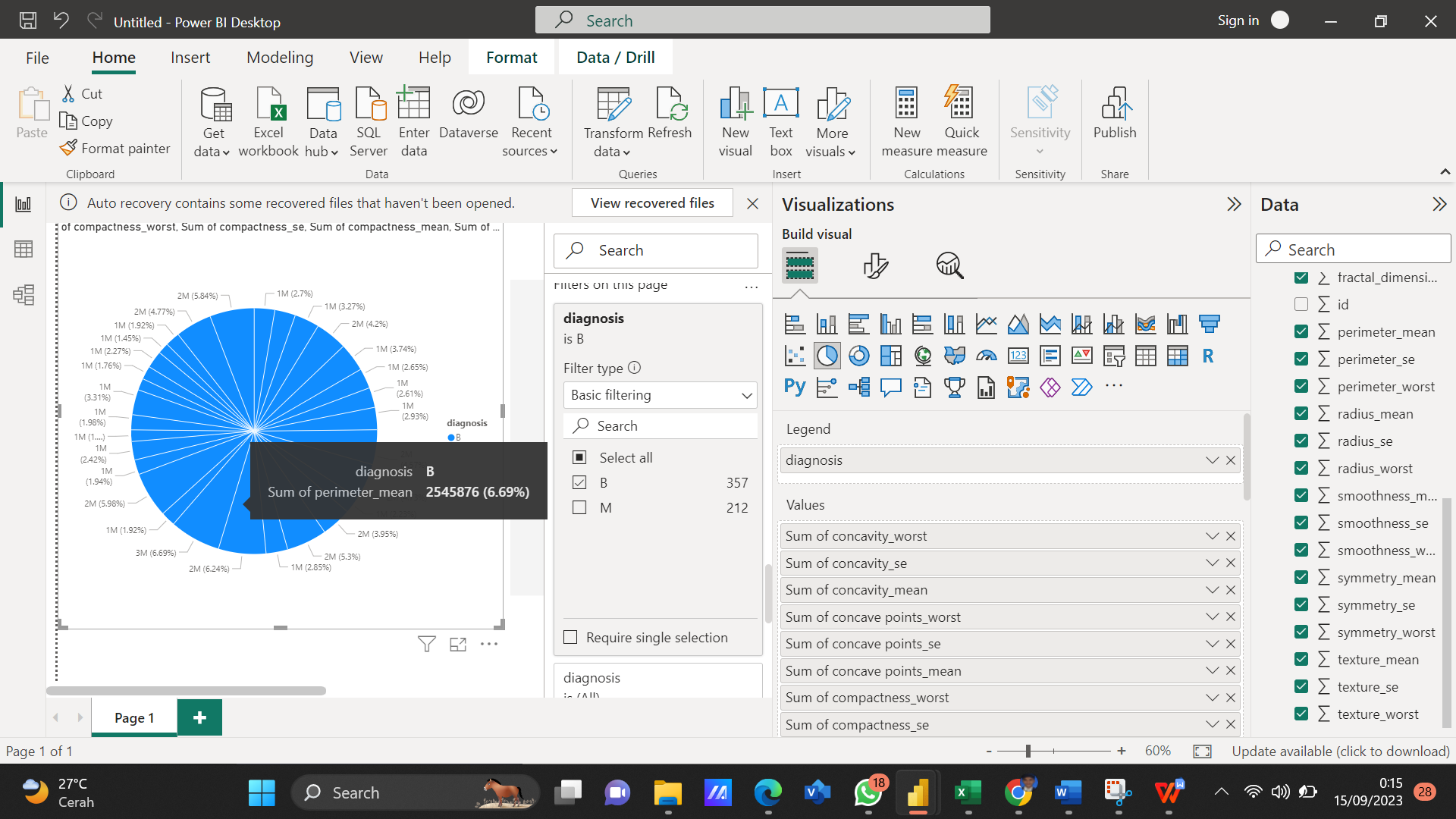
Fitur Data yang ada pada Breats Cancer (Data Kaglle), Dimana terdapat 32 kolom, dan 1 target yaitu M yang artinya diagnosisnya kanker payudara ganas dan B yang artinya kanker payudara jinak. Data yang didapat sebanyak 569 data. Dari data yang saya dapat tidak terdapat misiing value, untuk itu saya tidak melakukan proses missing value tersebut.

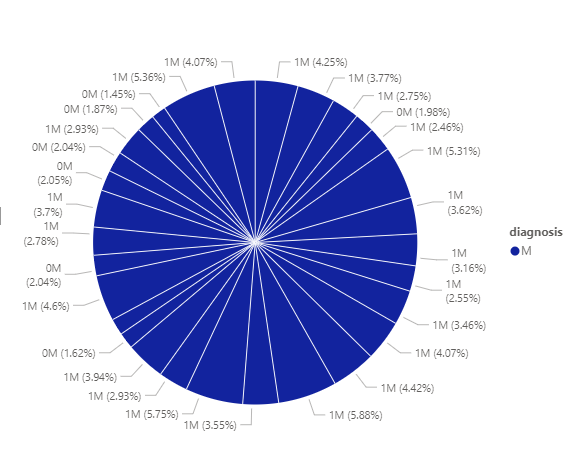
**Berikut isi datanya (Kolom):**

Unique ID, Target: M - Malignant B – Benign, Radius of Lobes, Mean of Surface Texture, Outer Perimeter of Lobes, Mean Area of Lobes, Mean of Smoothness Levels , Mean of, Compactness, Mean of Concavity, Mean of Cocave Points, Mean of Symmetry, Mean of, Fractal Dimension, SE of Radius, SE of Texture, Perimeter of SE, Are of SE, SE of Smoothness, SE of compactness, SEE of concavity, SE of concave points, SE of symmetry, SE of Fractal Dimension, Worst Radius, Worst Texture, Worst Permimeter, Worst Area, Worst Smoothness, Worse Compactness, Worst Concavity, Worst Concave Points, Worst Symmetry, Worst Fractal Dimension.

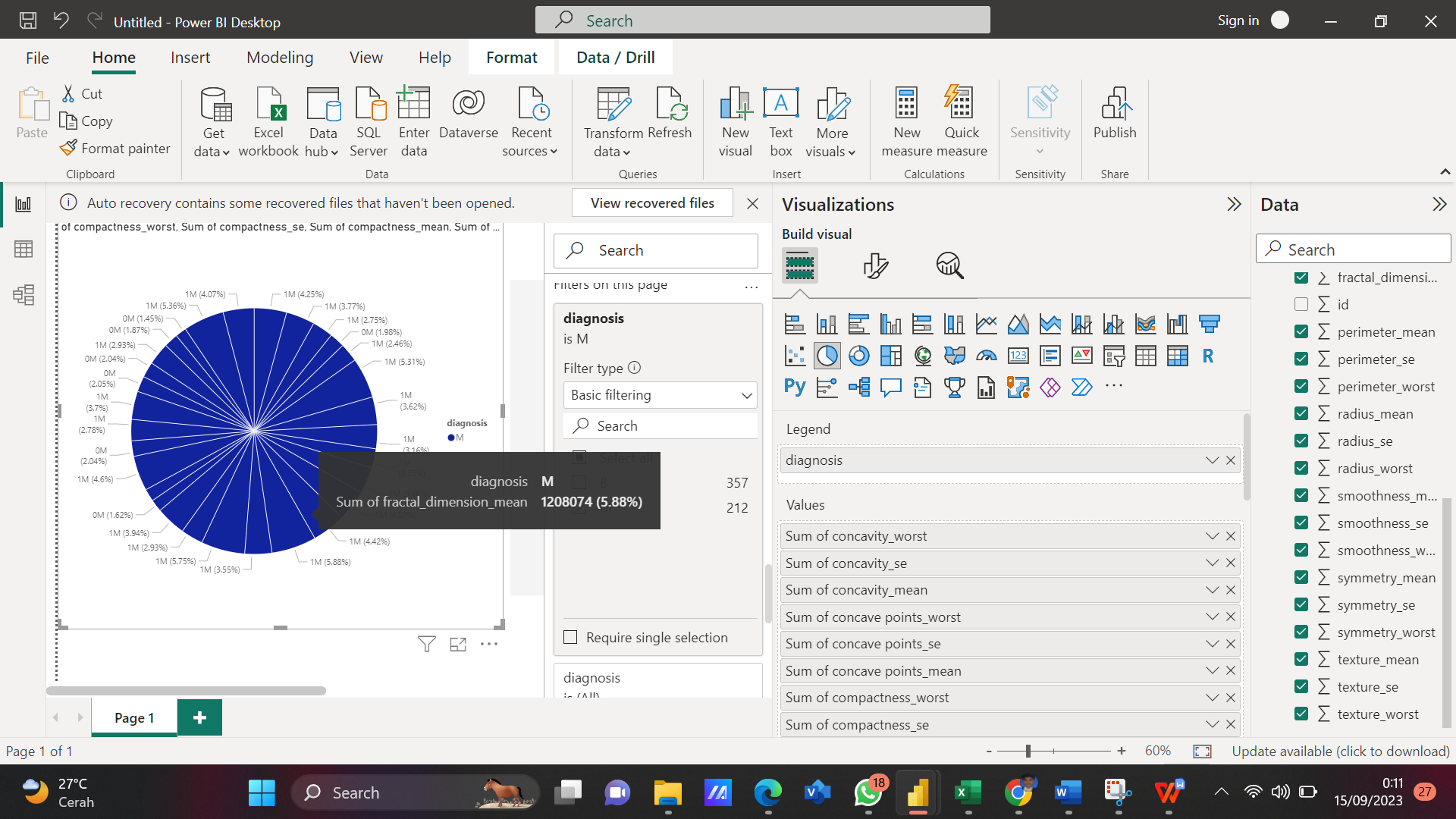
  
*Gambar Hasil analisis Diagnosis B (Kanker payudara Jinak) Berdasarkan Fitur pendukung lainnya*

Gambar tersebut menjukan hasil analisis data diagnosis B dengan menghilangkan fitur (kolom) Unique ID, (Dikarenakan tidak diperlukan), terdapat jumlah hasil diagnosisnya adalah 357 pasien. Terdapat data yang paling maxisimal pada fitur Sum OF perimeter\_mean Sebesar 2545876 (6.69%) pada diagnosis B



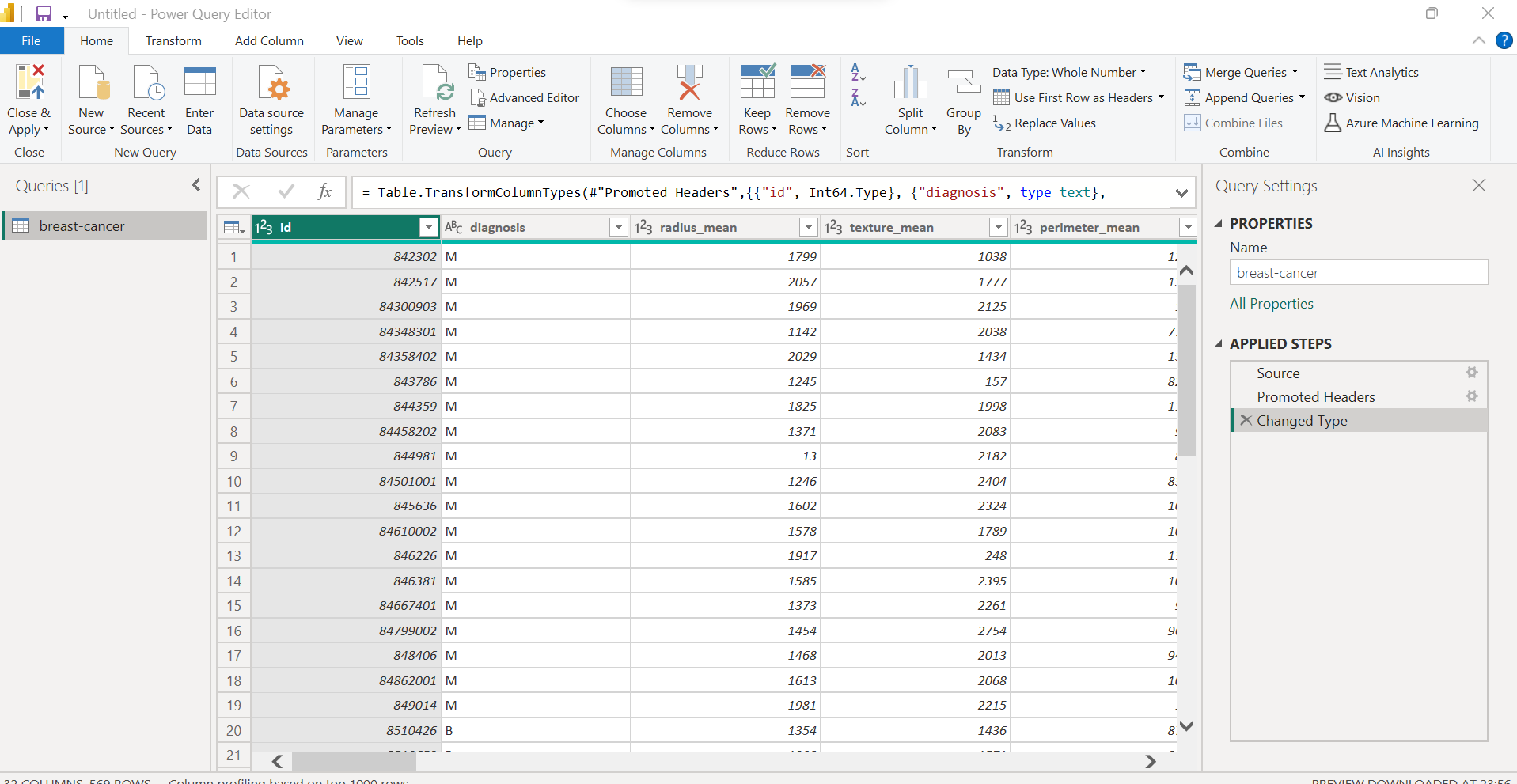
  
*Gambar Hasil analisis Diagnosis M (Kanker payudara Ganas) Berdasarkan Fitur pendukung lainnya*

Gambar tersebut menjukan hasil analisis data diagnosis B dengan menghilangkan fitur (kolom) Unique ID, (Dikarenakan tidak diperlukan), terdapat jumlah hasil diagnosisnya adalah 212 pasien. Terdapat data yang paling maxisimal pada fitur Sum OF fractal\_dimension\_Mean Sebesar 1208074 (5.88%) pada diagnosis M

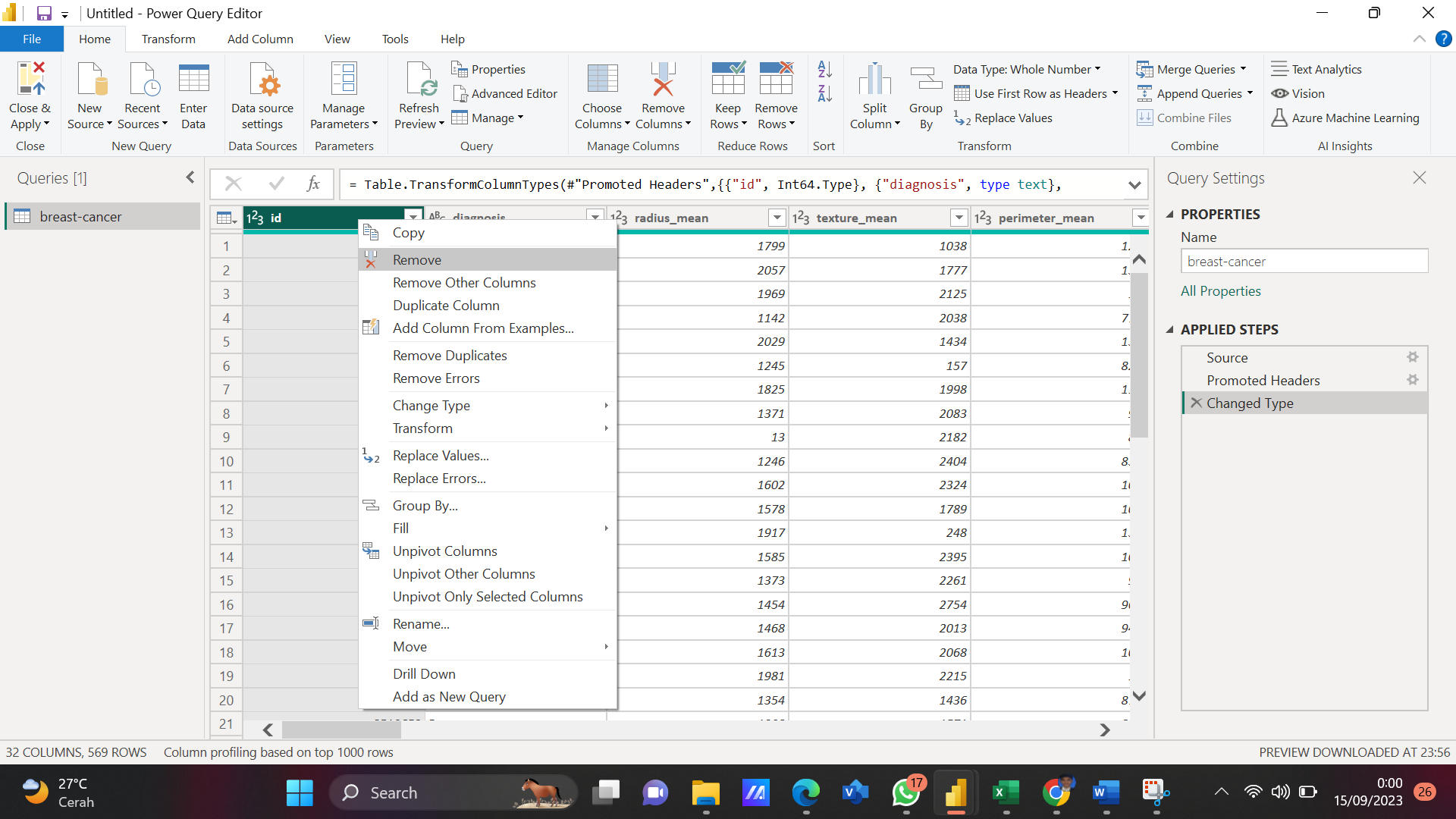


Jadi dapat dilihat data terbanyak yaitu data yang diagnosisnya B yaitu sebanyak 357 pasien. Untuk keterkaitan dalam data tersebut memiliki tingkat keterkaitan yang tinggi, hanya saja saya akan melakukan cleaning fitur (kolom yang tidak digunakan) yaitu Unique ID. Jadi adapun paremeter yang akan digunakan : Target: M - Malignant B – Benign, Radius of Lobes, Mean of Surface Texture, Outer Perimeter of Lobes, Mean Area of Lobes, Mean of Smoothness Levels , Mean of, Compactness, Mean of Concavity, Mean of Cocave Points, Mean of Symmetry, Mean of, Fractal Dimension, SE of Radius, SE of Texture, Perimeter of SE, Are of SE, SE of Smoothness, SE of compactness, SEE of concavity, SE of concave points, SE of symmetry, SE of Fractal Dimension, Worst Radius, Worst Texture, Worst Permimeter, Worst Area, Worst Smoothness, Worse Compactness, Worst Concavity, Worst Concave Points, Worst Symmetry, Worst Fractal Dimension.

1. **TRANFORMASI DATA PADA POWER BI**

  
*Gambar Proses tranformasi data*

transformasi data adalah bagian dari proses membersihkan data. Membersihkan data (data cleaning) adalah langkah awal dalam mempersiapkan data untuk analisis, dan transformasi data adalah langkah yang mengikuti membersihkan data. Pada tahap ini saya akan melakukan proses cleaning data. Membersihkan data adalah proses untuk mengidentifikasi, mengatasi, dan menghapus ketidakakuratan, ketidaklengkapan, atau ketidaksesuaian dalam satu set data, adapun parameter yang akan saya hapus adalah id.

  
*Gambar Tranformasi data yaitu cleaning data pada parameter ID*

Gambar Tersebut merupakan proses cleaning data pada satu parameter yaitu Id, karena menurut saya id tidak memiliki pengaruh besar dalam data yang saya analisis. Cleaning data tersebut dilakukan dengan cara, mengkik menu Tranform data > blok parameter yang ingin di cleaning > klik kanan dan pilih Remove > lalu klik Close & apply, selesai.