Dipersiapkan untuk

**Final Project**

**User Guideline**

Dipersiapkan oleh Kelompok 8A

**2023**

# Daftar Isi

[**Daftar Isi**](#_jww94eyyqdqw) **2**

[**A. Data Pipeline**](#_8exmmu16248d) **3**

[A.1. Model Data](#_z2b06r5jcyvb) 4

[A.2. Akun Penting](#_ht4lx2k5nvn2) 5

[A.3. Data Preprocessing](#_9yphaxwciw8f) 6

[A.4. Staging Area, Data Warehouse, Data Mart](#_nenmynjkx5xq) 6

[A.5. Transformasi Data Menggunakan dbt](#_a55e5lommo0y) 7

[**B. Prediksi Menggunakan Machine Learning**](#_vtx83zznhiu) **8**

[**C. Penggunaan Dashboard**](#_yypj8uy5834v) **9**

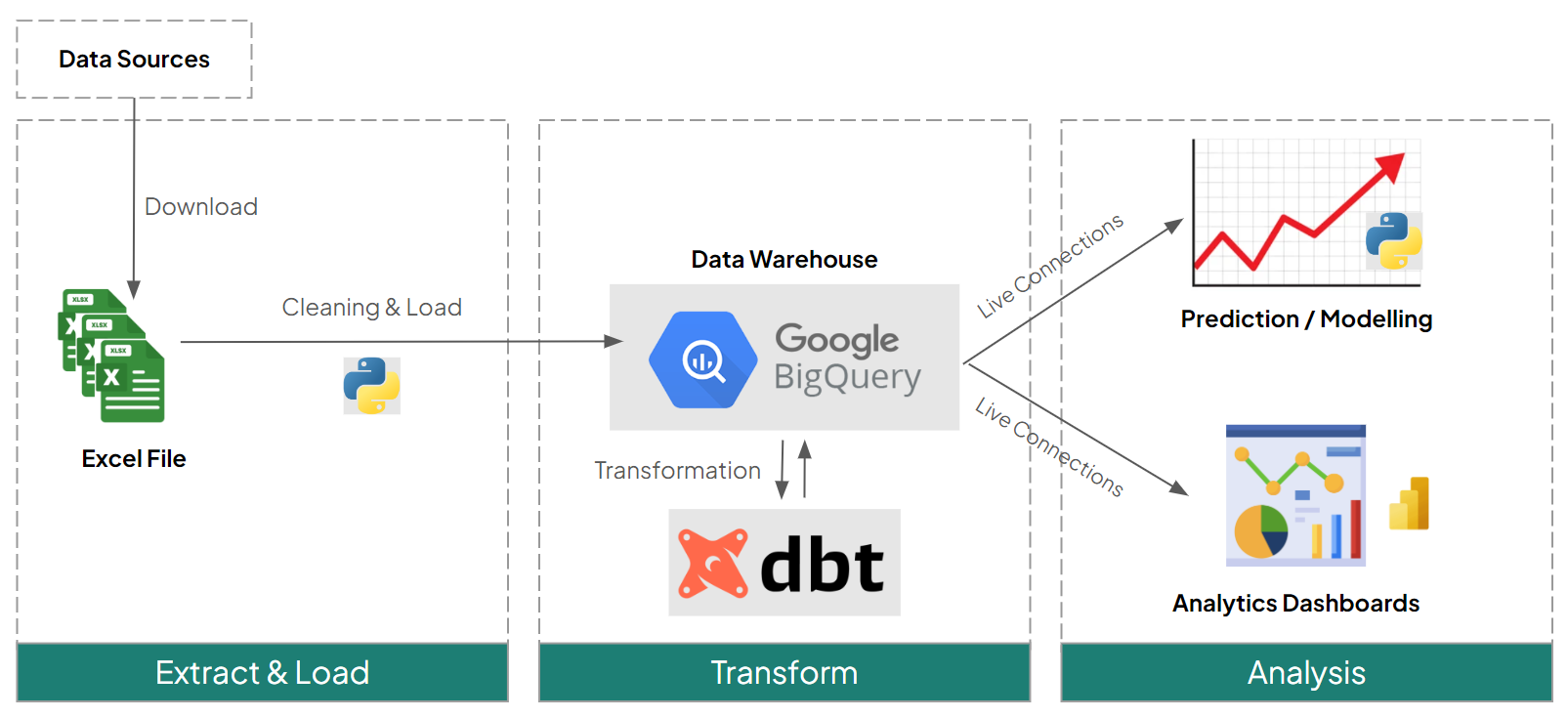
[C.1. Ketentuan Login](#_sg3fwdxm1oqq) 9

[C.2. Unduh Workbook](#_ut0hjzapawni) 9

[C.3. Publish Dashboard](#_ocnglhn6ycyf) 10

[C.4. Anatomi Dashboard](#_khro3e9x6lkt) 11

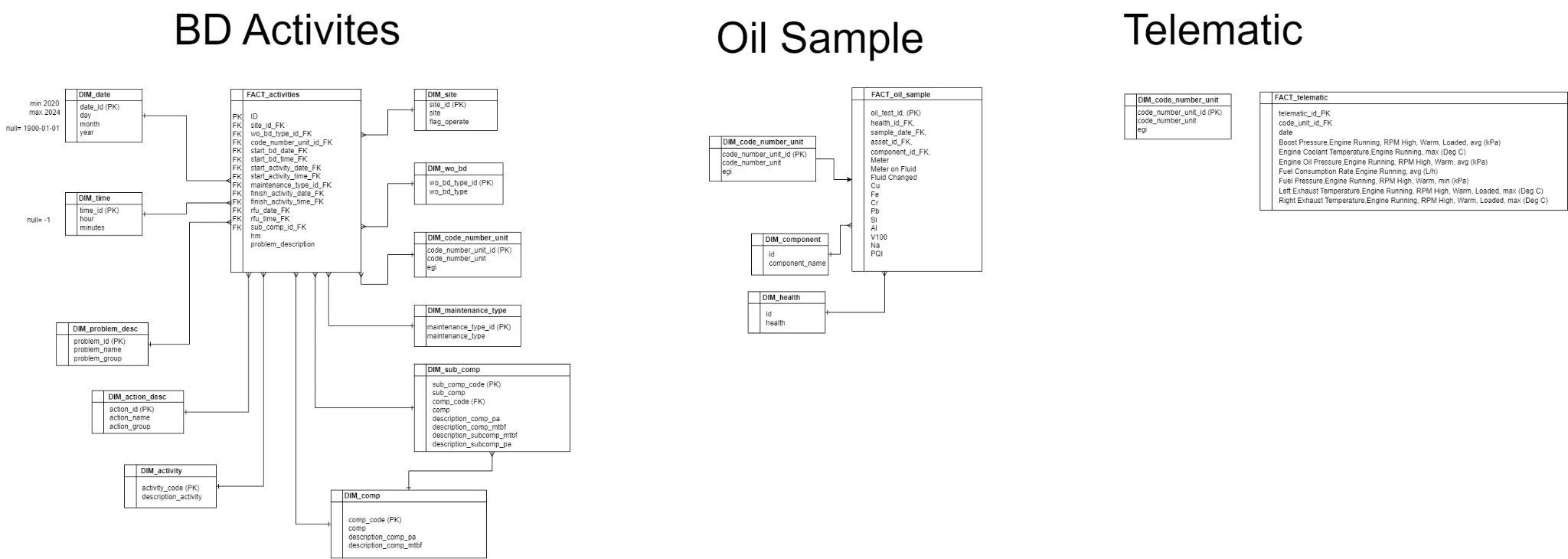
# A. Data Pipeline



Data pipeline dari solusi project ditunjukkan pada gambar diatas. Berawal dari data mentah dan pada titik akhir akan disajikan dalam model prediksi dan dashboard operasional. Berikut merupakan penjelasan proses pipeline.

1. Mula-mula data mentah diproses terlebih dahulu dengan script python. Hal ini dilakukan agar dataset dapat di load ke BigQuery tanpa error yang muncul.
2. Data mentah yang telah di load ke dalam BigQuery masuuk ke tabel staging area.
3. Data di dalam staging area akan ditransformasi secara langsung dengan dbt.
4. Data yang telah ditransformasi akan di load kembali ke tabel data warehouse.
5. Data dalam data warehouse akan dilakukan transformasi lanjutan sesuai dengan kebutuhan dari Data Analyst dan Data Scientist guna memenuhi kebutuhan dashboard dan modeling.

## A.1. Model Data



Gambar diatas merupakan data model dari dataset BD Activities. Data model menggunakan skema star dengan tabel di tengah sebagai fact tabel. Schema diatas menjadi acuan dalam desain dari data warehouse yang ditempatkan pada BigQuery.

Proses pengembangan data model:

1. Mula-mula dari dataset BD Activities, dilakukan identifikasi fact variabel dan dimension variable.
2. Mulai dengan mengembangkan tabel-tabel dimension yang akan menjadi pendukung/penjelas/pemberi konteks dari fact tabel
3. Mengembangkan fact tabel
4. Mengidentifikasi PK dan FK dari tabel-tabel yang telah disusun.

# 

Gambar diatas merupakan data model dari dataset Oil tabel. Data model menggunakan skema star dengan tabel di tengah sebagai fact tabel. Schema diatas menjadi acuan dalam desain dari data warehouse yang ditempatkan pada BigQuery.

Proses pengembangan data model:

1. Mula-mula dari dataset Oil tabel, dilakukan identifikasi fact variabel dan dimension variable.
2. Mulai dengan mengembangkan tabel-tabel dimension yang akan menjadi pendukung/penjelas/pemberi konteks dari fact tabel
3. Mengembangkan fact tabel
4. Mengidentifikasi PK dan FK dari tabel-tabel yang telah disusun.

## A.2. Akun Penting

Dalam menyelesaikan project ini berikut merupakan akun-akun penting yang digunakan, diantaranya akun BigQuery, akun dbt, dan akun Github. Informasi kredensial dari akun terdapat pada [sheet berikut](https://docs.google.com/spreadsheets/d/1gCuArpgGJdwq4e0wzfYU25SECHf6VESQEl-4TkF8aQU/edit?usp=share_link).

* BigQuery digunakan sebagai pengembangan data warehouse dan data mart.
* Dbt digunakan sebagai tools untuk transformasi data.
* Github digunakan sebagai repository penyimpanan code.

# 

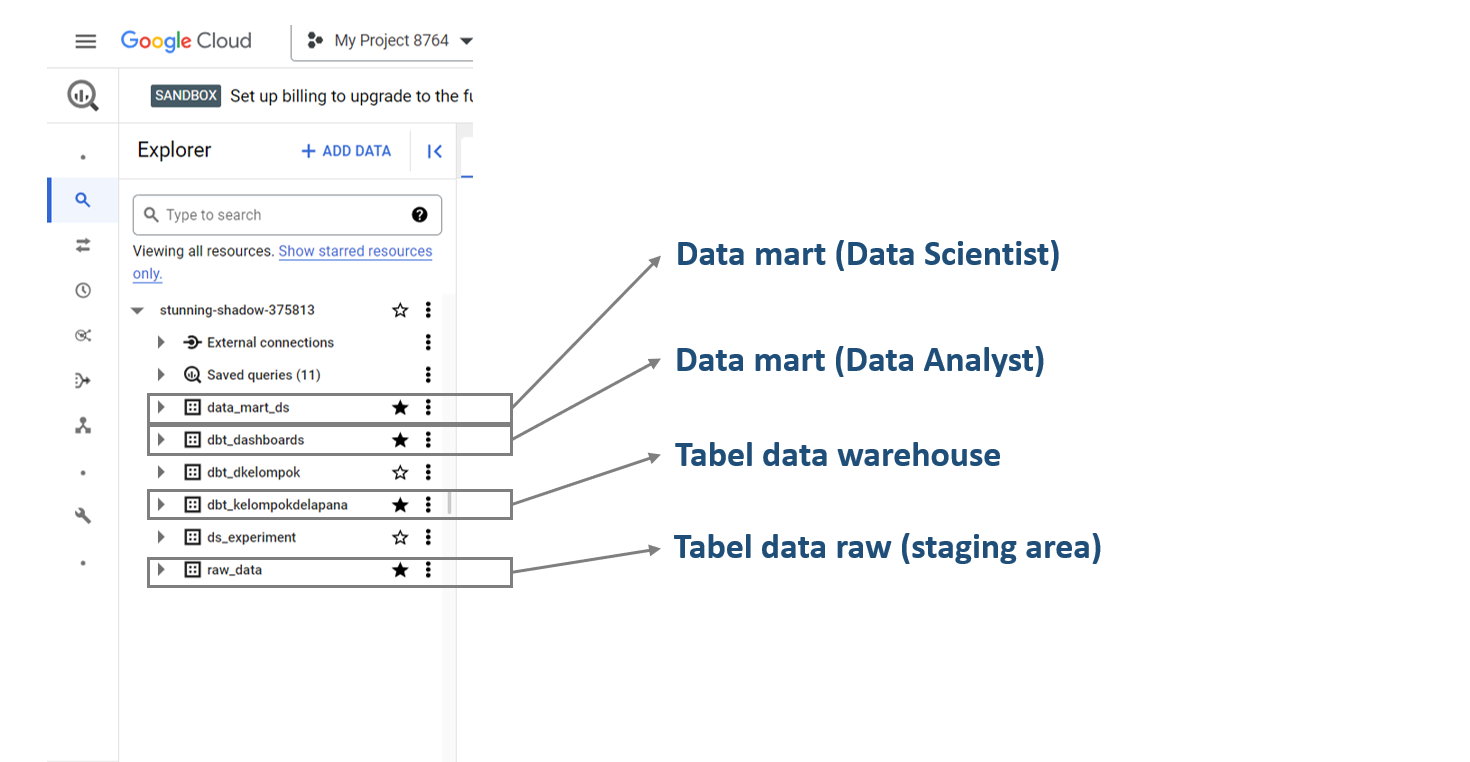
## A.3. Data Preprocessing

Data raw yang didapatkan pada project ini masih bersifat mentah dan kotor. Sedemikian sehingga untuk dapat di load ke BigQuery harus perlu adanya proses preprocess sebelum dapat ditransformasikan lebih lanjutan.

Tools yang dilakukan untuk melakukan preprocess dan load data raw ke BigQuery adalah dengan menggunakan Python dan Jupyter notebook. Source code dari proses ini dapat diakses di [link berikut](https://drive.google.com/file/d/1_1tQW7kDLuoVK5qEzE-qlWTOACyukZeP/view?usp=share_link).

## A.4. Staging Area, Data Warehouse, Data Mart

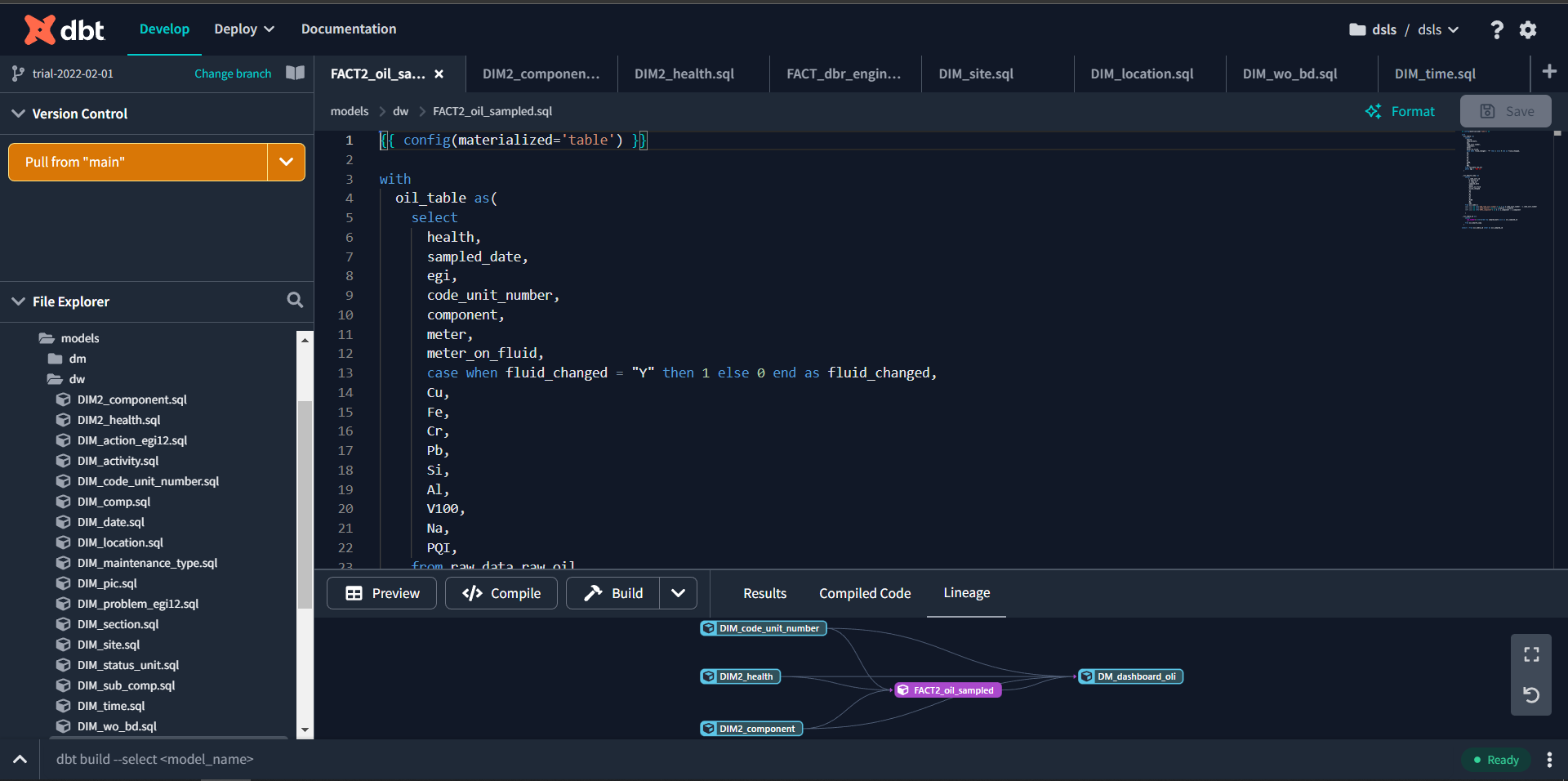
Setelah melakukan login menggunakan akun Google dan masuk pada Google BigQuery dari akun penting yang disebutkan diatas, berikut merupakan penjelasan bagian dari tabel-tabel di dalam BigQuery.

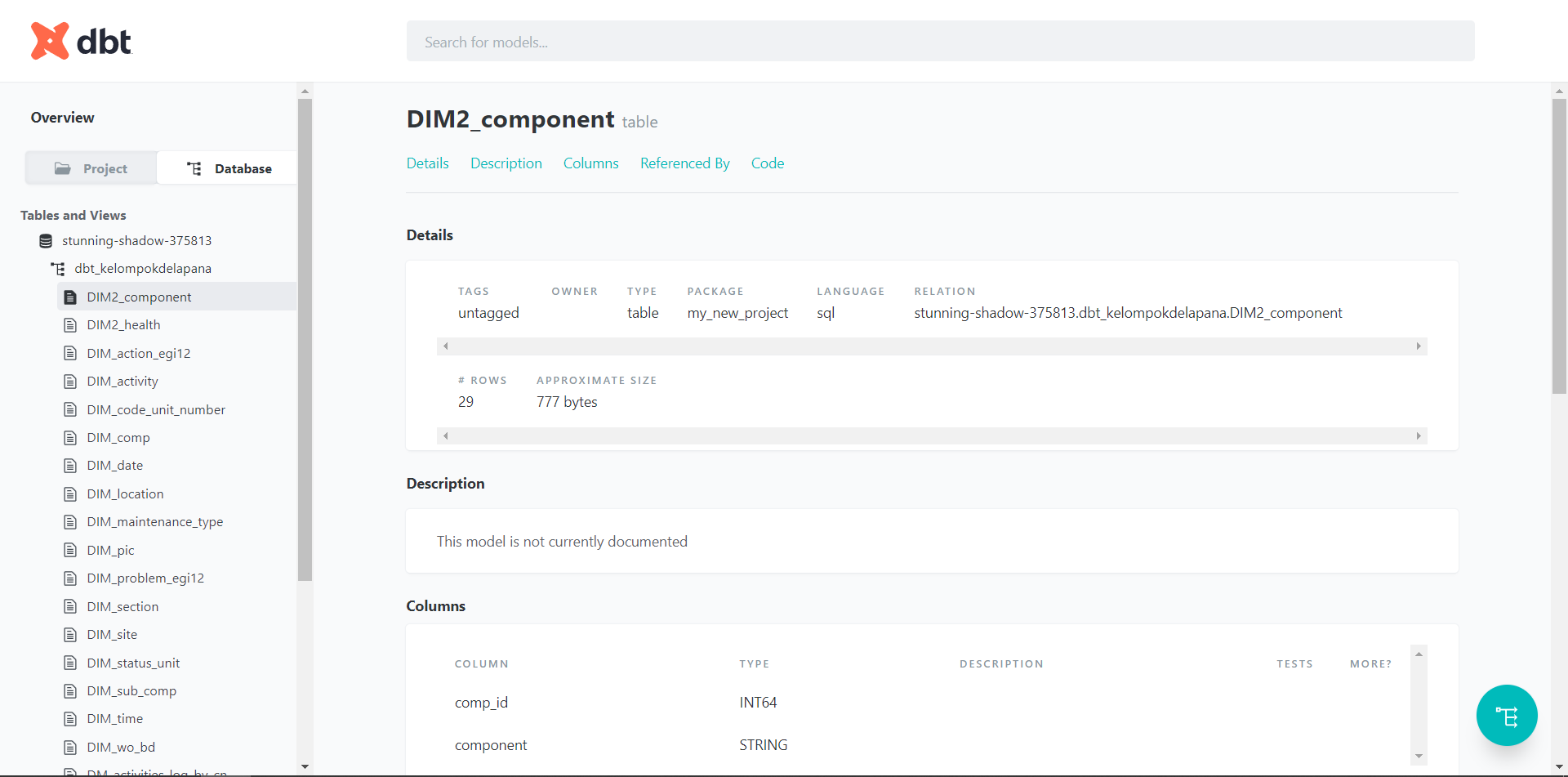


1. Tabel data raw (staging area): merupakan schema berisi tabel-tabel dari raw dataset yang telah di preprocess awal terlebih dahulu.
2. Tabel data warehouse: merupakan schema berisi tabel-tabel dari data model yang telah didesain pada bagian data model. Analyst ataupun BI dapat mengambil informasi dari schema ini dengan menggunakan query SQL.
3. Data mart: schema ini terdiri dari 2 bagian, yaitu data mart untuk kebutuhan modeling dan untuk kebutuhan dashboard operasional.

# 

## A.5. Transformasi Data Menggunakan dbt



Data yang telah di load pada staging area akan dilakukan transformasi dengan menggunakan dbt. Query yang digunakan untuk transformasi terdapat pada bagian repository model. Query yang telah tersimpan di dbt dapat disimpan pada Github. Selain dapat dilakukan transformasi data, juga dapat didapatkan dokumentasi sebagaimana terlihat pada Gambar di bawah ini.  


# B. Prediksi Menggunakan Machine Learning

* [Model Folder](https://drive.google.com/drive/folders/10RE-r2JG8m1_Fa2mcVHGGupRPDRj9dgC?usp=share_link)
  + Hasil pemodelan yang dapat digunakan
* [Inference Notebook](https://colab.research.google.com/drive/1pxUFemHhKF_RmKpJWuBmk2R7f220VFkB)
  + Aplikasi untuk melakukan prediksi
* [Retraining Notebook](https://colab.research.google.com/drive/1a42fkF7hai6Cuc2nQsXK8yUSrPCzSXXW?usp=share_link)
  + Dapat digunakan apabila dirasa model yang digunakan sudah bersifat tidak akurat

**Catatan**: Dokumentasi terkait ketiga hal tersebut dapat dilihat pada link yang dituju

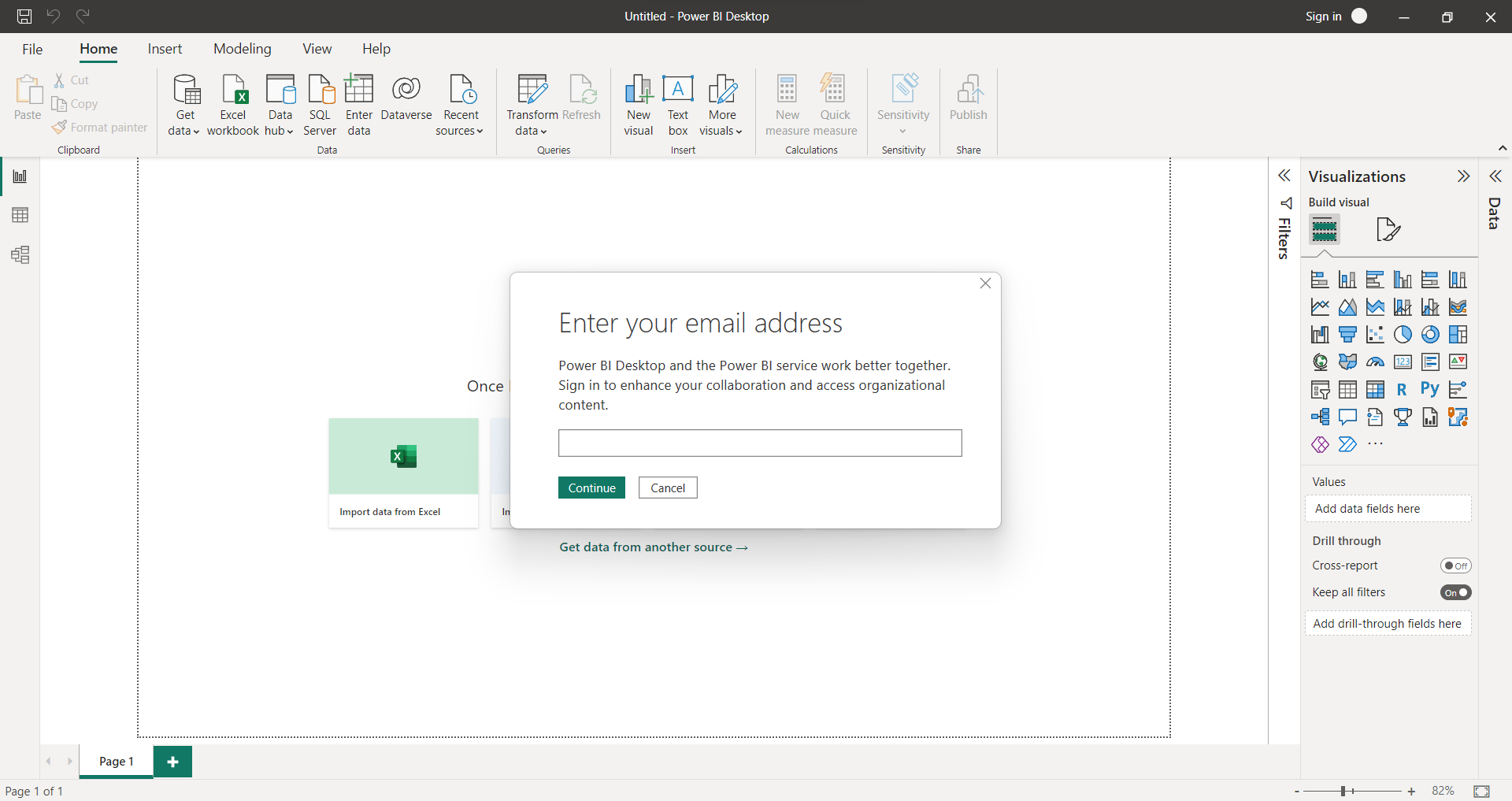
# C. Penggunaan Dashboard

Dalam proyek ini, dashboard yang dibuat ditujukan untuk pengambilan keputusan berdasarkan hasil analisis dan automasi yang telah dilakukan oleh sistem. Dashboard dibangun menggunakan aplikasi [Power BI Desktop](https://www.microsoft.com/en-us/download/details.aspx?id=58494) (berbasis lokal) yang ditampilkan hasilnya dalam [Power BI Service](https://app.powerbi.com/home) (berbasis web atau cloud). Instalasi Power BI Desktop dapat dibaca rinciannya [disini](https://learn.microsoft.com/en-us/power-bi/fundamentals/desktop-get-the-desktop). Berikut kami sampaikan petunjuk penggunaan dashboard.

## C.1. Ketentuan Login

Buka aplikasi Power BI Desktop lalu lakukan login dengan klik ikon Sign in, lalu akan muncul pop up untuk memasukkan alamat email dan password. Berikut ini adalah beberapa ketentuan dalam login aplikasi Power BI.

* Jika perusahaan sudah mempunyai akun microsoft, maka silakan login dengan akun tersebut.
* Jika perusahaan belum mempunyai akun microsoft, maka silakan membuat akun microsoft terlebih dahulu.
* Power BI hanya menerima email perusahaan atau sekolah, sehingga tidak bisa menggunakan domain email yang umum seperti (@gmail, @outlook, @hotmail). [sumber](https://help.agworld.com/en/articles/3495404-signing-up-for-power-bi)
* Masukkan password untuk melanjutkan proses login.
* Ketentuan lengkapnya dapat dibaca [disini.](https://learn.microsoft.com/en-us/power-bi/fundamentals/service-self-service-signup-for-power-bi)

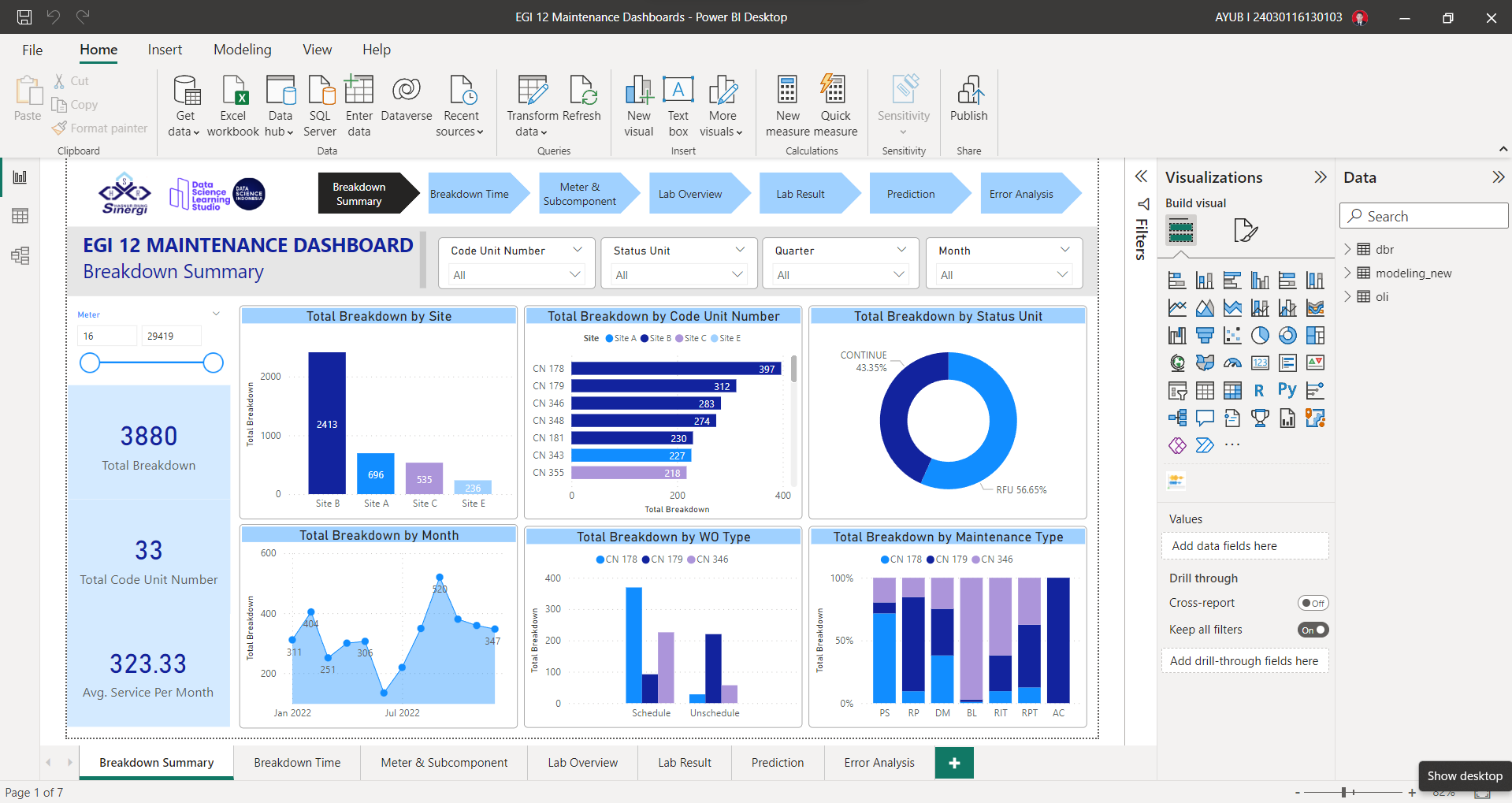


## C.2. Unduh Workbook

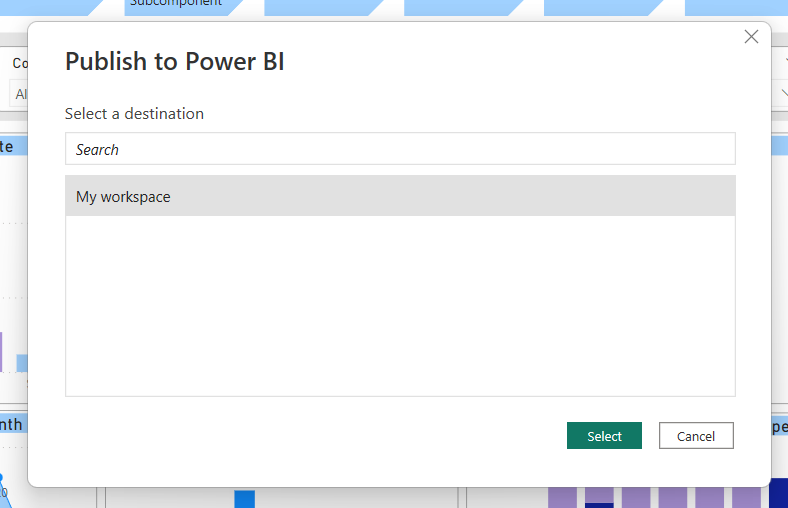
Untuk menjalankan dashboard, lakukan pengunduhan workbook Power BI dalam bentuk format .pbix yang dapat diunduh pada [link berikut](https://drive.google.com/drive/u/2/folders/1IGCVPvL_iKFqyru9Ae0u54N1IJCnXI5H). Lalu buka workbook pada aplikasi Power BI Desktop.

## C.3. Publish Dashboard

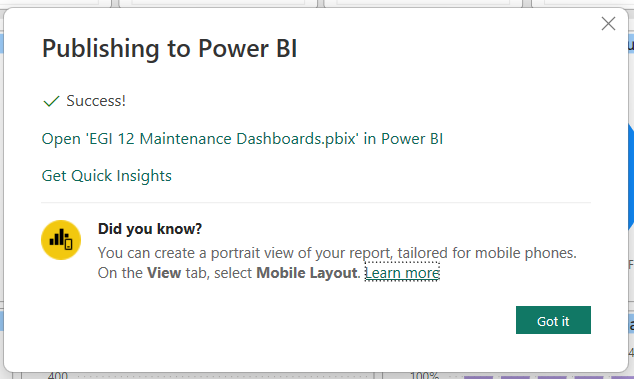
Untuk menampilkan dashboard dalam mode presentasi (full screen), hanya diakomodasi oleh Power BI Service, oleh karena itu harus dilakukan publish dari Power BI Desktop ke Power BI Service.



Selanjutnya akan muncul pop up untuk memilih destinasi file di Power BI Service, pilih My workspace.

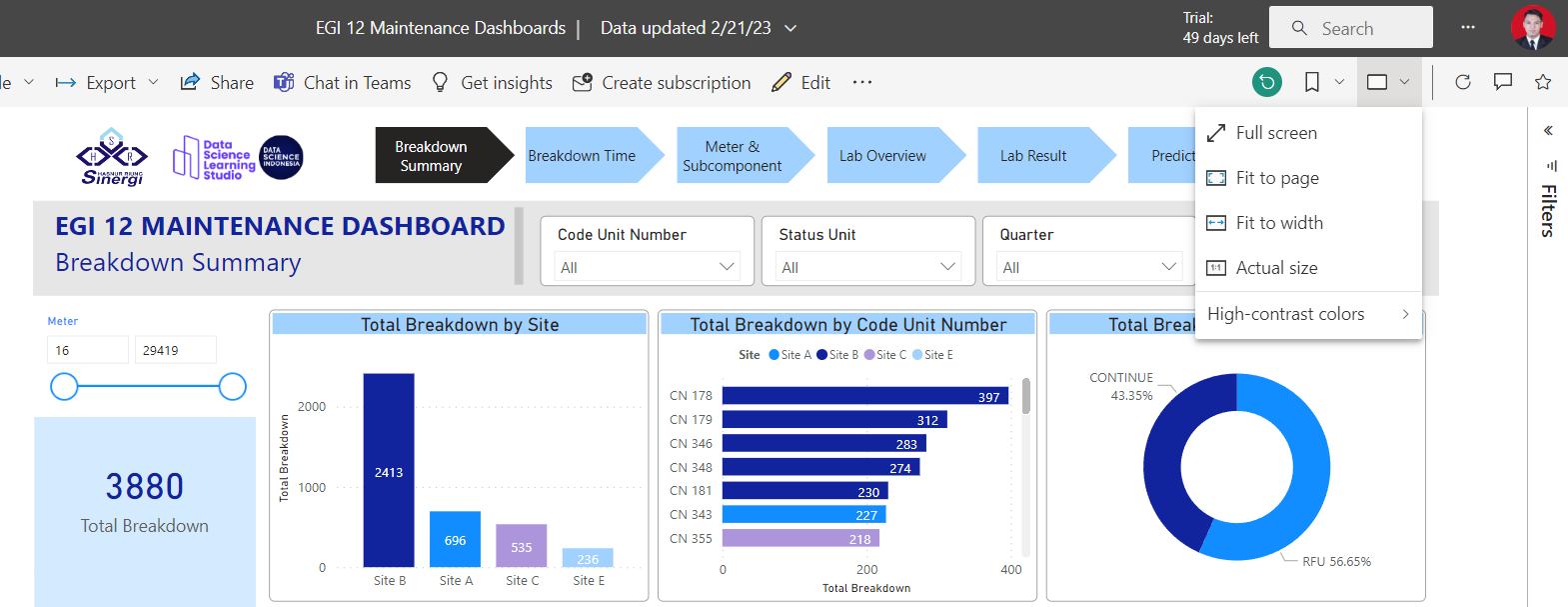


Jika proses publish berhasil, maka akan muncul pop up ‘success’, selanjutnya buka file dengan klik Open ‘EGI 12 Maintenance Dashboard.pbix’ in Power BI.

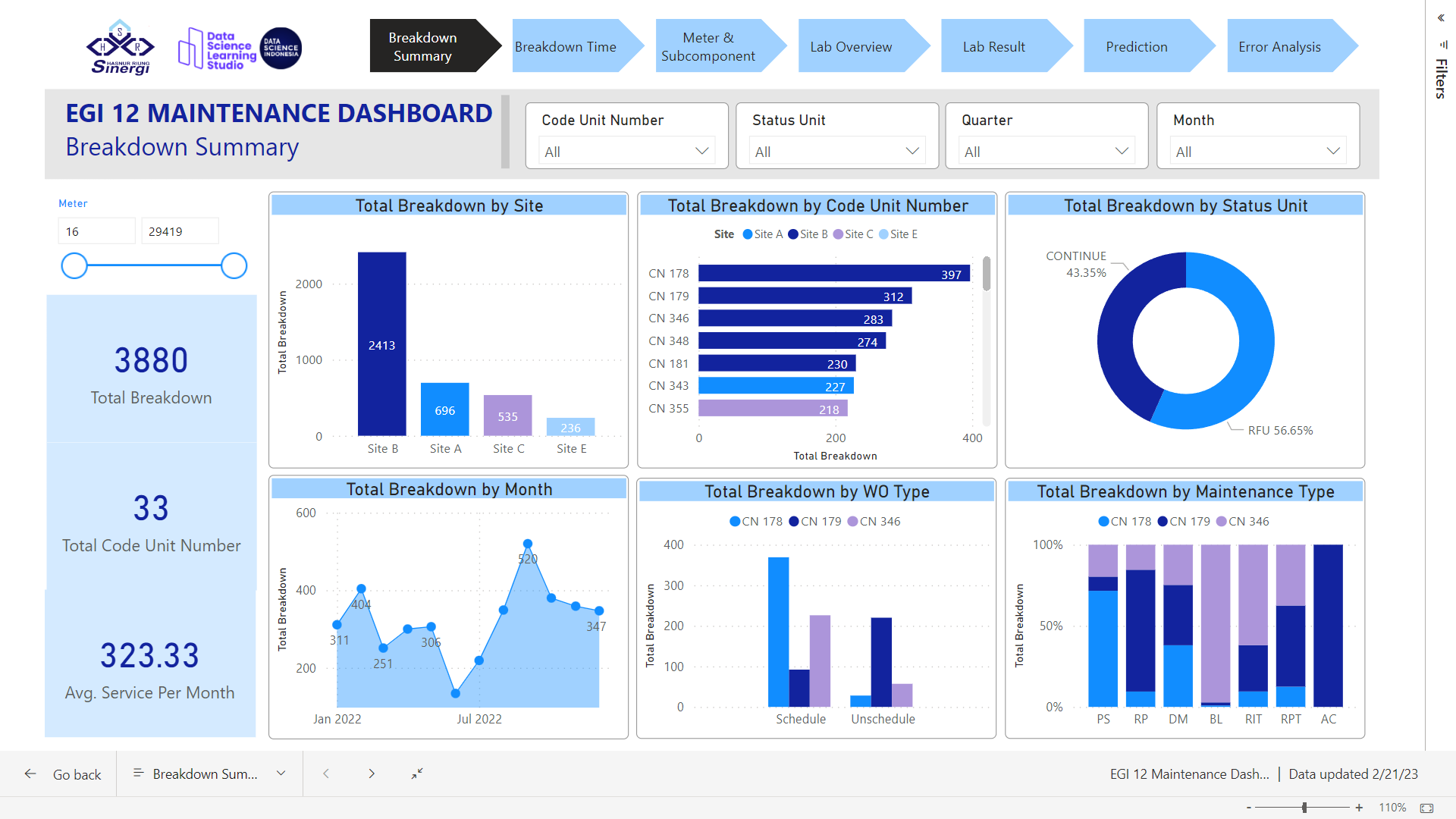


## C.4. Anatomi Dashboard

Untuk menampilkan dashboard dalam mode presentasi (full screen), silakan klik ikon persegi panjang dan pilih full screen.



Berikut adalah tampilan dashboard dan rincian fiturnya.



1. **Page navigation**

Terdapat 7 (tujuh) halaman yang dapat diakses, lakukan klik pada page navigation lalu dashboard akan otomatis berpindah ke halaman berikutnya.

1. **Dropdown filter**

Di setiap halaman terdapat dropdown filter yang berada pada area yang sama. Filter yang dihasilkan adalah bersifat *one-page* (hanya berlaku pada satu halaman saja).

1. **Slider**

Slider ini dapat digunakan untuk melakukan filter data yang bersifat nilai *continuous* atau *date*. Filter yang dihasilkan adalah bersifat *one-page*.

Selain menggunakan dropdown filter atau slider, filter juga dapat dilakukan dengan langsung melakukan klik pada bagian dari chart di dalam dashboard.

# D. Penutup

Kami sampaikan terima kasih sebesar-besarnya kepada Data Science Learning Studio dan Data Science Indonesia yang telah menyediakan wadah pembelajaran selama bootcamp dengan tema Data Consultant. Kami sampaikan terima kasih juga kepada PT. Hasnur Riung Sinergi (HRS) yang telah menjadi partner dataset dan use case yang telah memberikan ruang kepada kami untuk belajar langsung dalam *real case* pemecahan masalah dengan solusi data science secara langsung. Semoga karya yang kami hasilkan dapat bermanfaat bagi PT. Hasnur Riung Sinergi (HRS) dalam menunjang kebutuhannya guna mengefisiensikan proses operasionalnya. Kami berharap produk yang kami hasilkan dapat dikembangkan lebih lanjut agar lebih sempurna. Kami sangat terbuka untuk diskusi selanjutnya khususnya dengan PT. Hasnur Riung Sinergi (HRS).

Hormat Kami,  
**Kelompok 8A**

Bootcamp Data Science Learning Studio 2023

**Anggota tim:**

1. Gibran Brahmanta (Data Scientist)
2. Ayub Indra (Data Analyst)
3. Diana Afifah (Data Analyst)
4. Ilham Kurniawan (Data Analyst)
5. Luthfi Muhammad Irsyad (Data Engineer)
6. Ichfan Kurniawan (Data Engineer)