

LAPORAN AKHIR
MAGANG & STUDI INDEPENDEN BERSERTIFIKAT
Modeling And Optimization Techniques In Data
Warehousing Bootcamp
Di PT.Dibimbing Digital Indonesia

Diajukan untuk memenuhi persyaratan kelulusan
Program MSIB MBKM

oleh :
Mhd.Fisabiluddin / 201111566



S1 TEKNIK INFORMATIKA
UNIVERSITAS MIKROSKIL
2023

Lembar Pengesahan

**Modeling And Optimization Techniques In Data
Warehousing Bootcamp**

Di PT.Dibimbing Digital Indonesia

Oleh:

Mhd.Fisabiluddin / 201111566

disetujui dan disahkan sebagai

Laporan Magang atau Studi Independen Bersertifikat Kampus Merdeka

Jakarta, 5 Januari 2024

Learning Design Manager Dibimbing.id



Tri Utami

NIP : D22009

Kata Pengantar

Puji syukur penulis panjatkan kepada Allah SWT atas segala limpahan rahmat, hidayah, dan karunia-Nya yang telah melimpahkan keberkahan dalam penulisan laporan akhir ini. Laporan ini disusun sebagai bagian dari penyelesaian Program Magang dan Studi Independen Bersertifikat (MSIB) dengan judul " Laporan Akhir Magang & Studi Independen Bersertifikat Modeling And Optimization Techniques In Data Warehousing Bootcamp ".

Penulis juga ingin menyampaikan rasa terima kasih dan penghargaan yang setinggi-tingginya kepada berbagai pihak yang telah memberikan dukungan, bimbingan, penghargaan, dan bantuan selama penulisan laporan ini. Ucapan terimakasih khusus disampaikan kepada:

1. Bapak Nadiem Makarim dan seluruh jajaran Kemendikbudristek atas program Magang dan Studi Independen Bersertifikat yang sangat bermanfaat khususnya mahasiswa yang ingin mencari pengalaman dan ilmu di luar kampus.
2. William, S.Kom., M.M. selaku wakil rektor III Universitas Mikroskil yang telah memberikan izin untuk mengikuti programMSIB
3. R. A. Fattah Adriansyah S.Kom., M.Kom. selaku PIC Kampus MerdekaUniversitas Mikroskil.
4. Frans Mikael Sinaga, S.Kom., M.Kom selaku dosen kaprodi penulis pada program MSIB di Mitra Data Academy.
5. Sio Jurnalis Pipin, S.Kom., M.Kom selaku dosen pembimbing penulis pada program MSIB di Mitra Data Academy.
6. Seluruh mentor dan pengurus Mitra Dibimbing.id yang telah membantu dan membimbing penulis dalam menyelesaikanprogram MSIB ini
7. Kak Thosan Girisona, selaku mentor Dibimbing.id , yang telah memberikan arahan yang berharga selama Program Studi Independen.
8. Kak Ardi, selaku mentor dan pembimbing Final Project dari Dibimbing.id, yang telah memberikan kontribusi aktif dalam memberikan arahan dan bimbingan menyelesaikan Final Project.

Daftar Isi

Lembar pengesahan	i
Kata Pengantar	ii
Daftar Isi	iii
Daftar Gambar	iv
Bab I Gambaran Umum	1
I.1 Profil Perusahaan	1
I.2 Struktur Organisasi	2
I.3 Deskripsi Kegiatan	2-3
I.4 Jadwal Kegiatan	3
Bab II Pelaksanaan Study Independen di Dibimbing	4
II.1 Kelas Teori dan Mentoring	4-9
II.2 Deskripsi Final Project	10
II.3 Final Project	11-16
II.4 Hasil Final Project	17
Bab III Aktivitas Mingguan	18-23
Bab IV Penutup	24
IV.1 Kesimpulan	24
IV.2 Saran	24
Referensi	
Lampiran	

Daftar Gambar

Gambar1. Struktur Organisasi Dibimbing	2
Gambar2. Tampilan Pada halaman LMS Dibimbing	5
Gambar3. Tampilan pada Kelas Teori	5
Gambar4. Tampilan pada Kelas Mentoring	6
Gambar5. Tampilan Interface Dbaver	7
Gambar6. Tampilan Interface Google colab	7
Gambar7. Tampilan Interface website Github	8
Gambar8. Tampilan Interface Postman	8
Gambar9. Tampilan Interface Vscode	9
Gambar10. Tampilan Interface Docker	9
Gambar11. Data Source Perusahaan Retail US	11
Gambar12. Tampilan Execute Dags pada Airflow	12
Gambar13. Tampilan Our Data postgres pada Metabase	13
Gambar14. Sketsa ERD	13
Gambar12. Bar chart rate product	15
Gambar12. Row chart rate category product	16
Gambar12. Bar chart gender pelanggan	16

Bab I . Gambaran Umum

I.1 Profil Perusahaan

Dibimbing.id merupakan sebuah platform pembelajaran yang menawarkan persiapan karier digital. Mereka berkomitmen penuh untuk mendukung pemilik growth mindset dalam pengembangan diri mereka menuju karier impian. Melalui layanan bootcamp, video course online, dan Digital Skill Fair, Dibimbing telah berhasil membantu 70% siswanya meraih sukses dalam dunia kerja. Mereka terus berinovasi dan memperbaiki kualitas pembelajaran demi membantu individu mendapatkan karier yang lebih baik secara digital.

Visi dan Misi Dibimbing:

Visi:

Digital skills merupakan salah satu kunci peningkatan daya saing, tidak hanya bagi angkatan kerja tetapi juga masyarakat umum. Dibimbing.id mempunyai ambisi untuk menjadi mesin pencetak talenta digital tanpa batas. Fun, inclusive, dan affordable adalah nilai yang selalu kami pegang untuk para student yang ingin belajar di Dibimbing.id

Misi:

1. Customer Obsession

Kami akan selalu memprioritaskanmu untuk selangkah lebih maju dari ekspektasimu

2. Innovation First

Menjadi media belajar yang inspiratif, interaktif, dan juga selalu up to date sesuai dengan kebutuhan industri saat ini

3. Inclusivity

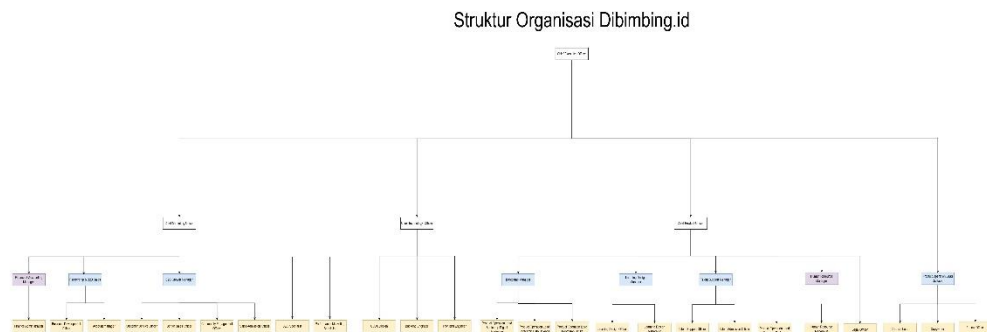
Semua orang dapat belajar dengan pelayanan yang sama tanpa terkecuali

4. People Empowering

Semua yang ada di lingkungan Dibimbing.id harus menjadi berdaya dan bermanfaat

I.2 Struktur Organisasi

Struktur Organisasi Dibimbing



Gambar 1. Struktur Organisasi Dibimbing

I.3 Deskripsi kegiatan

Sebagai peserta program Modeling And Optimization Techniques In Data Warehousing Bootcamp, peserta akan terlibat dalam latihan intensif terkait praktik data engineer. Merangkum materi serta akan mengerjakan proyek akhir yang fokus pada penggunaan ETL pada Airflow dalam kasus nyata. Sebelum memulai proyek akhir, peserta akan menjalani pembelajaran dalam penggunaan bahasa pemrograman Python dan implementasinya. Tujuan dari materi ini adalah agar peserta mampu memanfaatkan fitur dan fungsionalitas Python yang relevan dalam lingkup data warehousing. Selanjutnya, peserta akan belajar tentang kueri (query), penggunaan Docker, Airflow, serta dasar-dasar penggunaan Git, juga basic penggunaan postman. Selama tahapan proyek akhir, peserta akan bekerja secara kolaboratif untuk merancang dan menjalankan proses ETL menggunakan Airflow, melakukan pemodelan data pada layanan DBMS yang telah dipilih, serta menampilkan visualisasi data dalam lingkungan Docker. Mentor akan memberikan sesi mentoring untuk setiap tahapan proyek guna membahas dan menyelesaikan kendala yang mungkin muncul serta membantu meningkatkan kualitas proyek kelompok. Selain praktik teknis, peserta juga akan belajar berbagai teknik optimasi data warehousing, termasuk teknik pengoptimalan model, pemilihan metode optimasi yang sesuai untuk data yang digunakan, dan proses deployment dari model optimasi yang telah dikembangkan. Mereka akan diberi tugas-tugas, seperti

assignment mingguan, yang memperdalam pemahaman dan keterampilan. Setelah menyelesaikan serangkaian tugas dan pembelajaran intensif dari mentor, peserta akan menjalani proyek akhir selama 1 bulan ke depan. Proyek akhir dilakukan secara kelompok dengan topik menarik yang terfokus pada membangun infrastruktur data yang komprehensif untuk mengelola data dari berbagai pemasok global. Penggunaan Airflow untuk ETL/ELT, pemodelan data di Postgres, pembuatan dasbor visualisasi data, serta penyusunan presentasi yang menceritakan perjalanan proyek akan menjadi bagian penting dari proyek akhir ini. Mentor akan tetap memberikan bimbingan intensif selama proses pengerjaan proyek akhir untuk membantu setiap kelompok mengatasi tantangan dan mengoptimalkan hasil akhir proyek.

I.4 Jadwal kegiatan

Peserta akan mendapatkan materi intensif dari instruktur profesional yang ahli di bidangnya selama 4 bulan (2 sesi, 150 menit sesi teori dilakukan 3x seminggu dan 1 sesi mentoring seminggu). Setelah belajar mengenai teori dasar, peserta akan diberi tugas untuk membuat final project dan diberi waktu selama 1 bulan untuk menyelesaikan final project tersebut secara berkelompok. Hasil tugas akhir akan diperiksa oleh mentor pendamping. Selama pengerjaan final project, peserta akan mendapatkan sesi mentorship sehingga peserta bisa mendapatkan feedback dari mentor untuk menyempurnakan final projectnya.

Bab II. Pelaksanaan Study Independen di Dibimbing

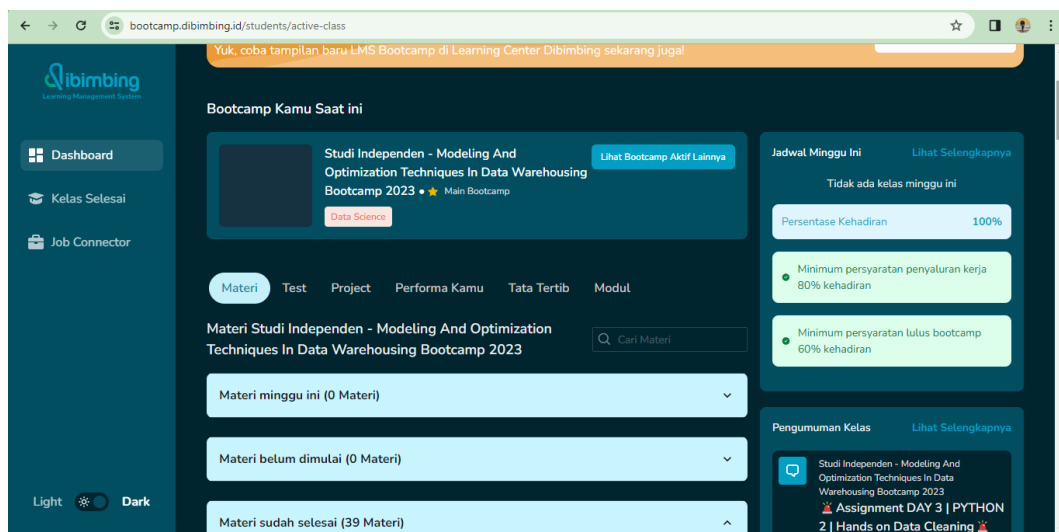
II.1 Kelas Teori dan Mentoring

Selama 4 bulan di program studi independen di Dibimbing, kami mengikuti pelatihan yang terdiri dari kelas teori dan sesi mentoring dengan fokus pada data warehousing dan data engineering. Dalam kelas teori, kami berinteraksi dengan mentor melalui ruang pertemuan online. Mentor kami memberikan penjelasan mendalam tentang berbagai teori yang relevan dan membimbing kami dalam mengaplikasikannya menggunakan alat dan perangkat lunak yang umum digunakan dalam industri.

Setiap minggu, kami menghadiri kelas teori sebanyak tiga kali, yaitu pada hari Selasa, Kamis, dan Sabtu dari pukul 18.30 hingga 21.00 WIB. Dalam total 39 sesi pertemuan kelas teori, setiap sesi membahas modul yang berbeda, mencakup berbagai topik penting dalam bidang data warehousing dan data engineering.

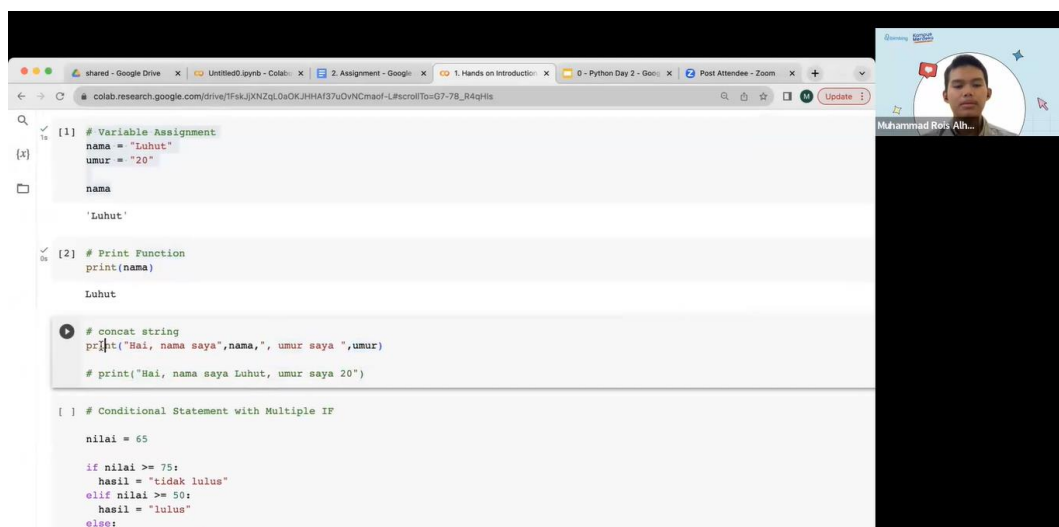
Platform yang kami gunakan untuk mengakses materi pelatihan adalah LMS (Learning Management System) yang disediakan oleh Dibimbing. Antarmuka dari platform ini sangat intuitif dan mudah digunakan, yang mempermudah akses dan navigasi ke berbagai modul pelatihan. Setiap modul menyajikan materi yang relevan dan terstruktur dengan baik, memungkinkan kami untuk memahami secara mendalam topik-topik seperti ETL, Dasar-dasar Data Warehousing, SQL, Python untuk Data Engineer, Web Scraping dan API, GIT, Docker, ETL dengan Python Airflow, Konsep Batch dan Stream, serta Cloud.

Melalui gabungan kelas teori yang intensif dan sesi mentoring yang difokuskan pada proyek akhir, program studi independen di Dibimbing memberikan pemahaman mendalam dan keterampilan praktis dalam bidang data warehousing dan data engineering. Untuk mengakses modul-modul tersebut, kami menggunakan platform bootcamp.dibimbing.id. Berikut adalah tampilan antarmuka dari platform tersebut.



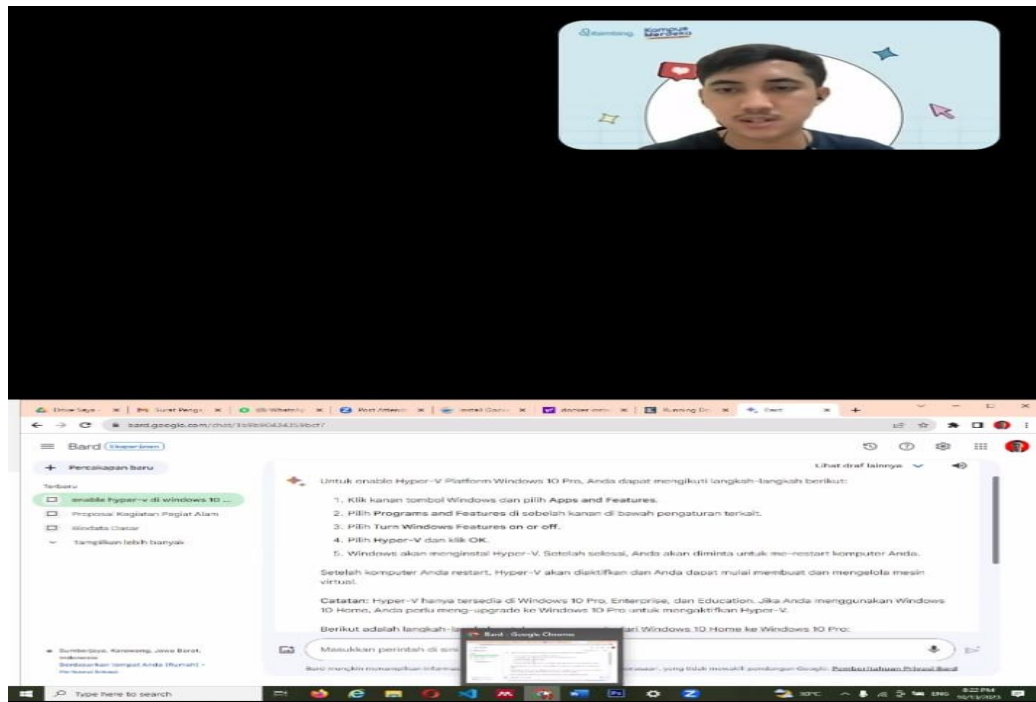
Gambar 2. Tampilan pada Platform LMS Dibimbing

Para peserta studi independen Modeling And Optimization Techniques In Data Warehousing Bootcamp program ini terdapat 1 kelas teori, kelas terdiri dari 67 peserta. Berikut adalah salah satu gambar yang menunjukkan kegiatan kelas teori Bersama mentor.



Gambar 3. Tampilan pada Kelas Teori

Selain kelas teori , terdapat juga metoring , para peserta akan diberikan materi tambahan dan ruang diskusi terkait semua hal tentang meteri untuk meningkatkan pemahaman para peserta dan juga mentor selalu membantu memberi masukan kepada peserta secara teori atau praktik tersebut Berikut adalah salah satu gambar yang menunjukkan kegiatan kelas mentoring Bersama mentor.



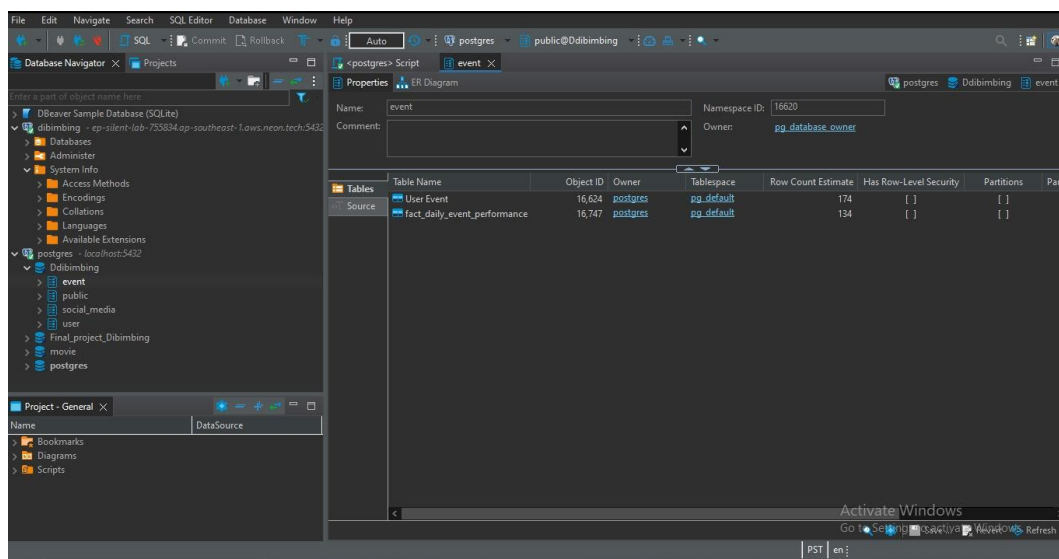
Gambar 4. Tampilan pada Kelas Mentoring

Dalam Modeling And Optimization Techniques In Data Warehousing Bootcamp, mitra Dibimbing menetapkan persyaratan bagi peserta untuk menggunakan beberapa alat dan platform yang relevan dengan peran Data Engineer. Ini termasuk Postgres dan DBAver untuk pembelajaran SQL, yang ditunjukkan pada Gambar 4. Kami juga menggunakan Google Colab untuk proses ETL dan manajemen data menggunakan Python, seperti yang diperlihatkan pada Gambar 5. Github berperan sebagai tempat penyimpanan untuk file-file proyek yang kami kerjakan, sebagaimana ditunjukkan pada Gambar 6. Selain itu, kami memanfaatkan Postman untuk mempelajari batch dan streaming, serta teknik web scraping, juga terlihat pada Gambar 7. Vscode untuk membuat dan mengatur code dags, seperti pada gambar 8. Docker digunakan untuk mengelola semua kontainer

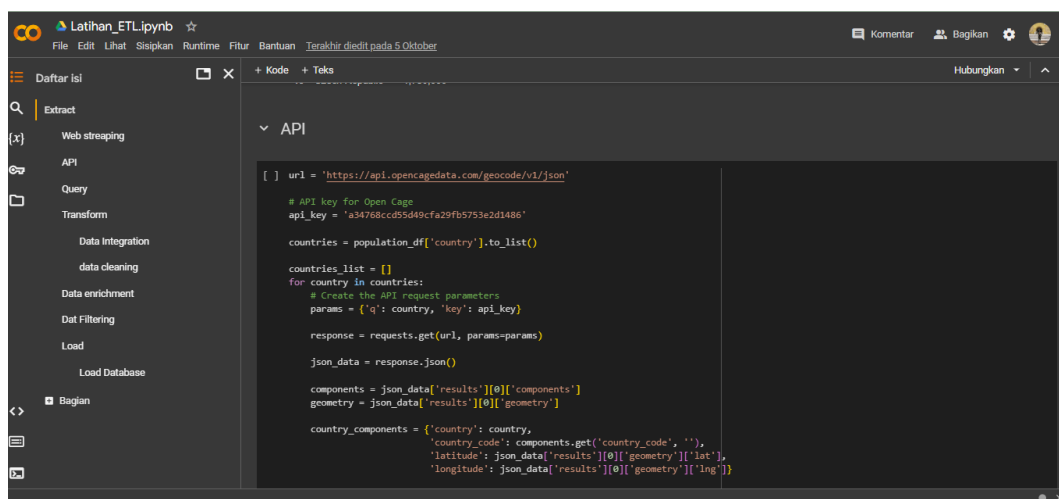
yang ingin kami jalankan pada Airflow, seperti yang terlihat pada Gambar 9. Kami juga

menggunakan bootcamp.dibimbing.id/LMS Dibimbing sebagai platform untuk mengumpulkan tugas-tugas yang telah diberikan kepada kami.

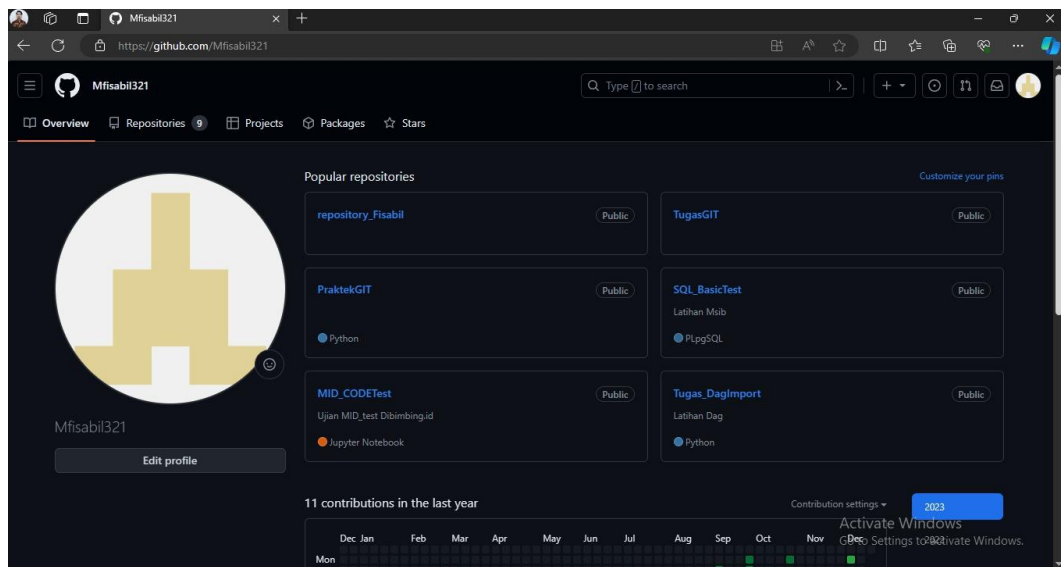
Kombinasi alat-alat ini diharapkan memberikan kami kesempatan untuk belajar dan berlatih dengan berbagai aspek penting dalam peran Data Engineer, mulai dari manajemen database hingga pemrosesan data, serta pemahaman mengenai aliran kerja yang berkelanjutan dalam lingkungan profesional.



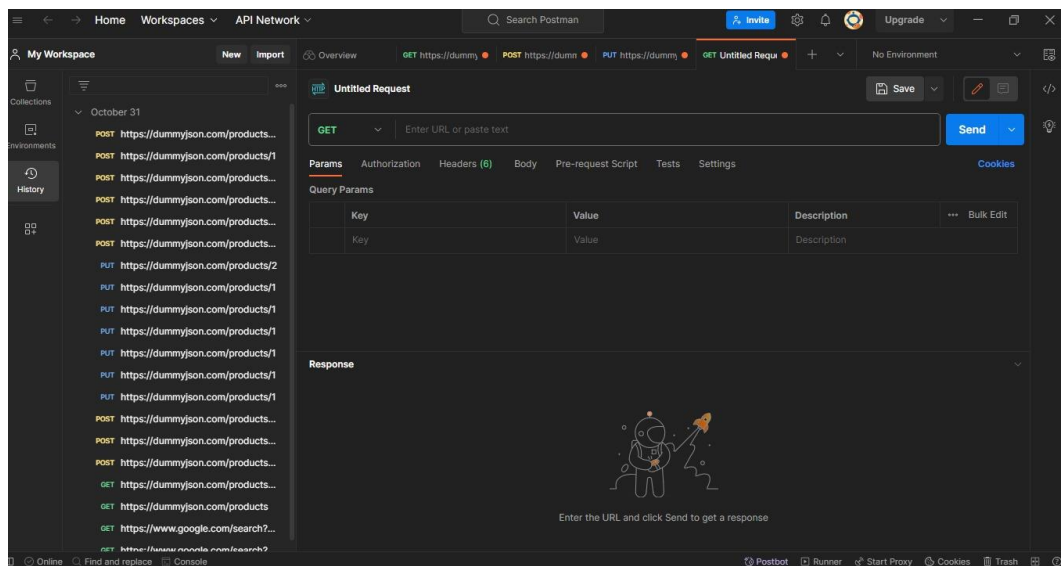
Gambar 5. Tampilan interface Dbaver



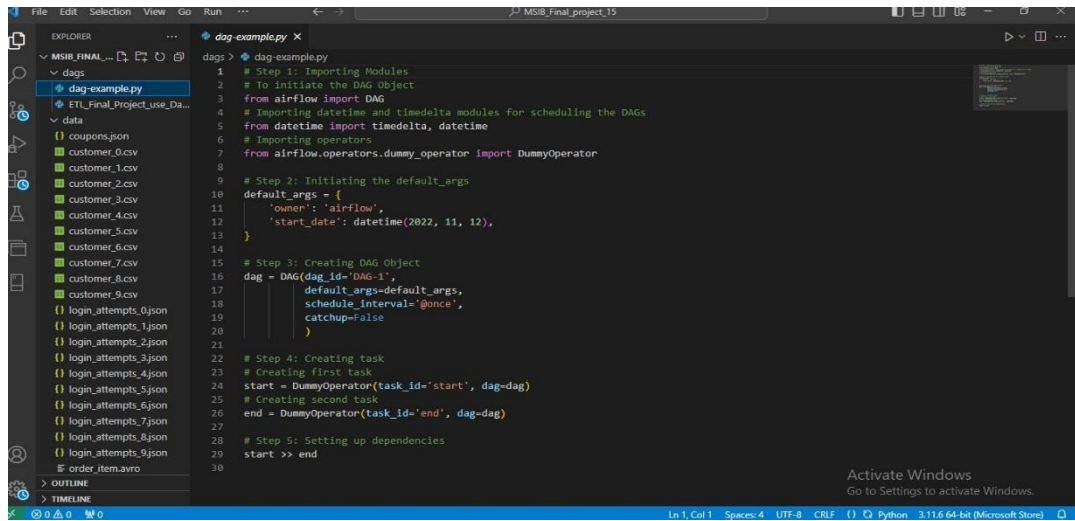
Gambar 6. Tampilan interface Google Colab



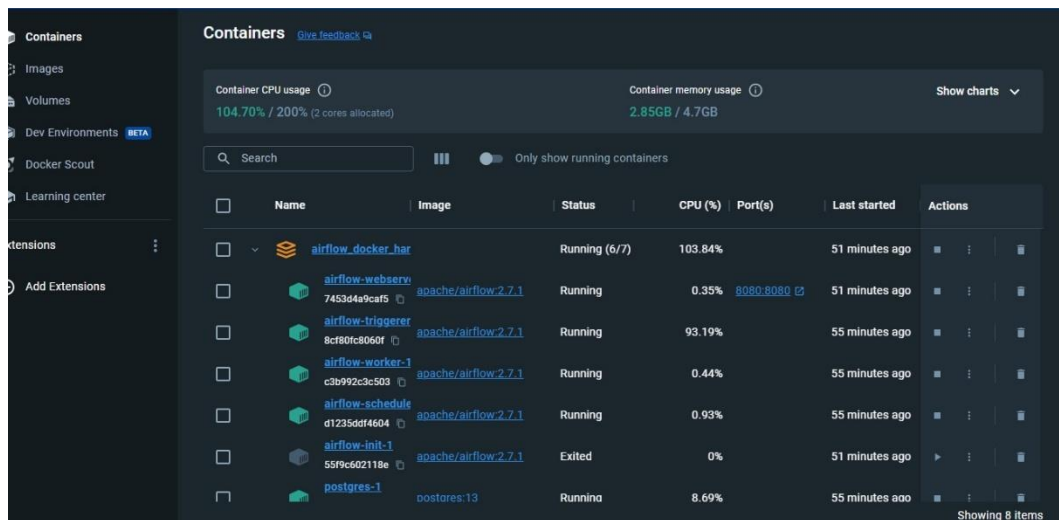
Gambar 7. Tampilan interface website Github



Gambar 8. Tampilan interface postman



Gambar 9. Tampilan interface Vscode



Gambar 10. Tampilan interface Docker

II.2 Deskripsi final project

Mitra Dibimbing memberikan deskripsi dan persyaratan final project yang harus dipenuhi oleh kelompok beranggotakan 4-5 peserta. Sebelum memulai proyek akhir, setiap peserta mengikuti sesi pembelajaran bersama mentor untuk memahami dasar-dasar yang diperlukan dalam menyelesaikan proyek tersebut, karena setiap peserta memiliki cara masing-masing dan pendekatan tersendiri untuk menyelesaikan proyek tersebut.

Final project ini berkaitan tugas sebagai seorang Data Engineer yang baru bergabung di perusahaan ritel daring di AS. Tantangannya adalah membangun infrastruktur data menggunakan data yang disusun untuk mencerminkan data dunia nyata dari perusahaan teknologi terkemuka. Kamu akan diberikan proyek template untuk membimbing pekerjaanmu. Berikut adalah tugas-tugasnya:

1. Pembuatan ETL/ELT dengan Airflow: Menarik data dari berbagai format ke dalam gudang data harian menggunakan Airflow, dengan memanfaatkan perpustakaan pemrosesan pilihanmu seperti Pandas, Polars, Spark, atau yang lain. Ada 8 tabel dengan 5 format yang berbeda, seperti Kupon (json), Pelanggan (csv), Percobaan Login (json), Item Pesanan (avro), Pesanan (parquet), Kategori Produk, Produk, dan Pemasok (xls). diharapkan saya bisa mendapatkan poin tambahan dengan menjelajahi dan menerapkan teknologi dan metodologi data lainnya seperti Datahub, dbt, polars, duckdb, dan alat data lainnya.
2. Pemodelan Data di Postgres: Membuat tabel informatif di gudang data Postgres, memanfaatkan data yang diberikan dan melakukan gabungan tabel yang kreatif sesuai prinsip pemodelan data Kimball (tabel fakta dan dimensi). Jumlah baris untuk masing-masing tabel juga disebutkan, dengan opsi untuk menambah atau memperkaya data yang ada untuk poin tambahan. Saya diharapkan membuat minimal 2 tabel baru dari gabungan tabel yang sudah diimpor sebelumnya.
3. Pembuatan Dasbor dengan Visualisasi Data: Setelah data diimpor dan dimodelkan, tugasmu adalah membuat visualisasi yang menarik dan informatif. Gunakan model data yang telah disusun untuk merancang dasbor informatif yang menunjukkan metrik dan tren bisnis kunci. Dapat menggunakan Metabase untuk visualisasi data, namun diperbolehkan juga untuk menggunakan alat lain di luar yang bisa memberikan poin tambahan.

Harapannya, proyek ini memungkinkanmu mempraktikkan keterampilan ETL, pemodelan data, dan visualisasi dalam konteks penggunaan nyata di industri ritel daring AS.

II.3 Final project

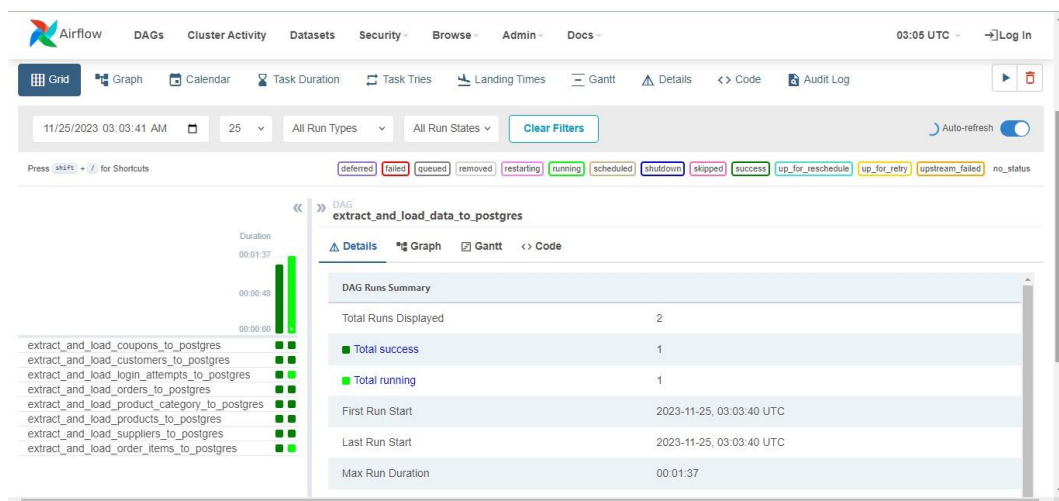
Final project yang diselesaikan berfokus pada ETL pada airflow , Pemodelan data dengan postgres dan visualisasi menggunakan metabase . Proyek ini bertujuan membangun infrastruktur data untuk mendukung operasi perusahaan ritel online AS dengan pembuatan job ETL/ELT, pemodelan data, serta pembuatan dashboard untuk visualisasi data. Hal ini akan meningkatkan pemahaman dan pengambilan keputusan berdasarkan informasi yang tersedia dari data ritel yang disimulasikan. Project Mulai dengan memahami isi data yang akan diolah, termasuk formatnya (JSON, CSV, Avro, Parquet, XLS/XLSX) dan sumber daya data yang berbeda.



Gambar 11. Data Source Perusahaan Retail US

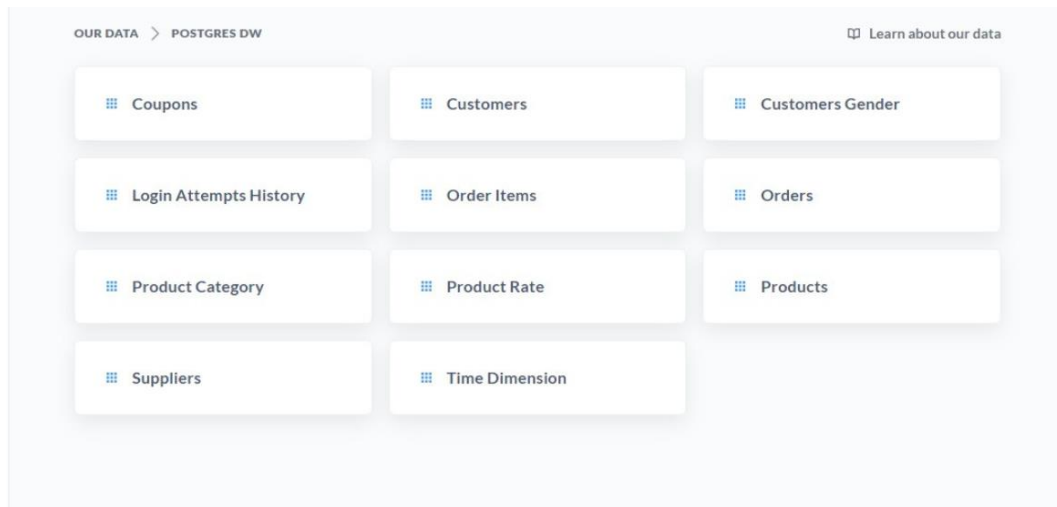
Selanjutnya membuat DAGS, DAG yang telah dibuat menjadi tulang punggung dalam menangani semua proses ETL/ELT dari berbagai format data ke dalam lingkungan PostgreSQL. Setiap tahap ekstraksi dirancang khusus untuk menangani file-file dengan format yang berbeda, mulai dari JSON, CSV, Avro, Parquet, hingga XLS/XLSX, memastikan bahwa data diproses ke dalam skema tabel yang telah ditentukan di dalam database PostgreSQL. Alur kerja DAG dipetakan dengan jelas, mengatur ketergantungan antar-tugas sehingga setiap langkah dalam proses ETL/ELT memanfaatkan data yang telah diproses sebelumnya, memastikan konsistensi dan keakuratan data.

Fungsi kunci dalam DAG, seperti pengiriman data dari Pandas DataFrame ke PostgreSQL, pembuatan dan manajemen tabel di PostgreSQL, serta serangkaian fungsi ekstraksi seperti `extract_and_load_coupons_to_postgres`, `extract_and_load_customers_to_postgres`, dirancang untuk menangani pembacaan dan pemrosesan data dari berbagai format file ke dalam Pandas DataFrame, dan selanjutnya dimasukkan ke dalam tabel dengan skema yang telah ditentukan. Penggunaan library seperti Pandas, Fastavro, SQLAlchemy, psycopg2-binary, xlrd, dan Python-snappy memberikan landasan yang kokoh dalam mengelola, memproses, dan memasukkan data dari berbagai sumber ke dalam database PostgreSQL, memungkinkan analisis data yang mendalam dan pengambilan keputusan bisnis yang tepat.



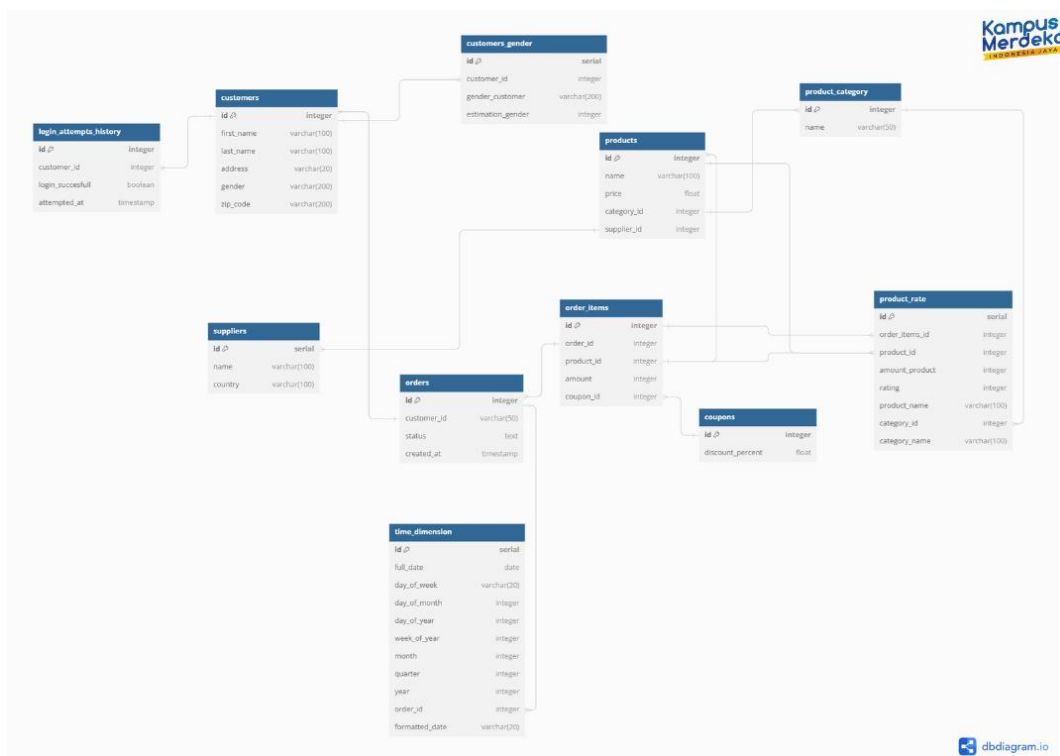
Gambar 12. Tampilan Execute Dags pada Ariflow

Hasil ETL yang telah berjalan dalam Airflow Telah menghasilkan data Coupons ,Customers, Login Attempts, Order Item, Order , Product Category, Product, Supplier ke dalam lingkungan postgres. Setelah itu melanjutkan menambahkan tabel baru (time_dimension, product_rate, dan customers_gender) berperan sebagai entitas tambahan yang memperkaya informasi dari data sebelumnya; time_dimension berfungsi sebagai dimensi waktu untuk analisis mendalam, product_rate memuat informasi rating produk untuk analisis kinerja produk, sementara customers_gender fokus pada data jenis kelamin pelanggan dari tabel customers dengan kemungkinan inklusi hasil prediksi atau estimasi.



Gambar 13. Tampilan Our Data Postgres pada Metabase

Data yang telah dihasilkan memiliki schema ERD yang sederhana dengan data yang telah di tambahkan dalam proses pemodelan data pada postgres. Desain ERD menampilkan hubungan antartabel dalam basis data.



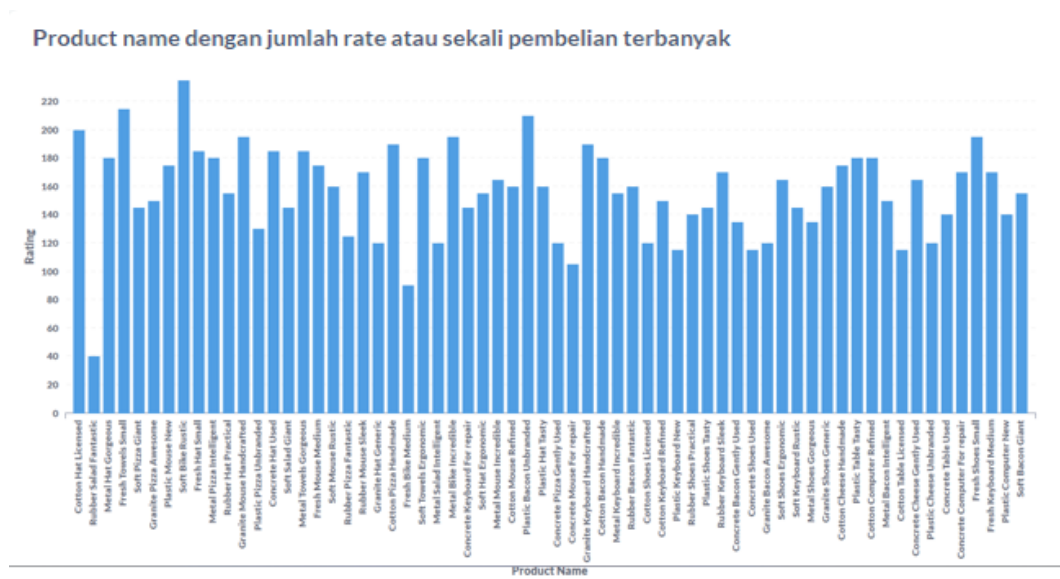
Gambar 14. Sketsa ERD

- **Coupons:** Tabel ini memuat informasi tentang kupon diskon yang dapat digunakan dalam pesanan. Mereka memiliki kolom id sebagai kunci utama dan discount_percent untuk menentukan jumlah diskon dalam kupon. Kolom coupon_id di tabel order_items berfungsi sebagai kunci asing yang mengacu pada kupon yang digunakan dalam pesanan tertentu.
- **Customers:** Berisi data pelanggan dengan kolom id sebagai kunci utama. Tabel ini menampung detail penting seperti first_name, last_name, address, gender, dan zip_code. Customer_id dalam tabel orders dan login_attempts_history merupakan kunci asing yang menghubungkan pesanan dengan pelanggan yang bersangkutan.
- **Login_Attempts_History:** Tabel ini merekam upaya login pengguna ke sistem dengan kolom id sebagai kunci utama. Menggunakan customer_id sebagai kunci asing untuk mengaitkan upaya login dengan pelanggan yang berusaha masuk.
- **Order_Items:** Tabel ini berfungsi untuk melacak detail produk dalam suatu pesanan. Id adalah kunci utama, order_id dan product_id adalah kunci asing yang menghubungkan pesanan dengan produk yang dibeli. Coupon_id adalah kunci asing lainnya yang mengacu pada kupon yang digunakan dalam pesanan.
- **Orders:** Tabel ini mencatat informasi penting tentang pesanan seperti status, waktu pembuatan, dan detail lainnya. Id sebagai kunci utama dan customer_id mengacu pada pelanggan yang melakukan pesanan. Terhubung dengan time_dimension untuk data tambahan terkait waktu.
- **Product_Category:** Menyimpan daftar kategori produk dengan kolom id sebagai kunci utama. Digunakan dalam tabel products dengan category_id sebagai kunci asing untuk mengkategorikan produk.
- **Products:** Tabel ini berisi detail produk dengan id sebagai kunci utama. Informasi seperti name, price, dan lainnya disimpan di sini.
- **Suppliers:** Tabel ini menyimpan informasi tentang pemasok produk, dengan id sebagai kunci utama.
- **Time_Dimension:** Berfungsi sebagai tabel dimensi waktu yang memiliki id sebagai kunci utama. Menyimpan informasi seperti tanggal, hari dalam minggu, bulan, tahun, dan detail waktu lainnya yang terkait dengan pesanan.
- **Product_Rate :** menghubungkan penilaian produk dengan detail pesanan dan produk melalui kunci asing order_items_id dan product_id. Ada kaitan dengan kategori produk yang diambil dari tabel Product_Category untuk memungkinkan analisis berdasarkan kategori produk dari penilaian yang dilakukan.
- **Customers_Gender:** Tabel ini menampung informasi jenis kelamin pelanggan. Id adalah kunci utama, dan customer_id digunakan sebagai kunci asing yang menghubungkan data jenis kelamin dengan data pelanggan pada tabel customers.

Desain ini diharapkan memfasilitasi struktur yang kokoh dan terorganisir, memungkinkan analisis data yang lebih dalam dan integrasi yang efisien antar berbagai entitas dalam sistem basis data.

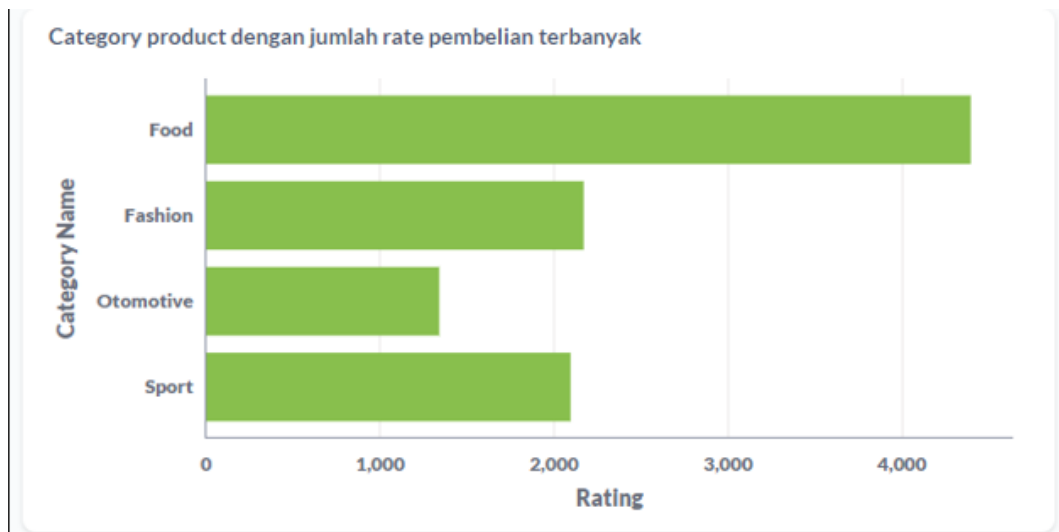
Setelah tahap pemodelan data telah selesai lanjut ke visualisasikan data yang telah ada di sini kami memvisualisasikan dengan 3 bentuk visualisasi.

1. Visualisasi bar chart untuk mengetahui nama product dengan banyak pembelian dalam sekali pesan (rate). bertujuan untuk menampilkan produk-produk yang paling sering dibeli dalam satu pesanan, memberikan gambaran langsung tentang produk dengan tingkat pembelian tertinggi. Ini membantu dalam identifikasi cepat produk-produk yang diminati pelanggan dalam satu transaksi, memungkinkan perusahaan untuk mengarahkan strategi penjualan, manajemen stok, dan pengembangan produk untuk meningkatkan efisiensi operasional serta kepuasan pelanggan.



Gambar 15. Bar chart rate product

2. Visualisasi row chart untuk mengetahui category dari product yang memiliki jumlah terbanyak pembelian juga jumlah terbanyak dalam sekali pembelian (rate). bertujuan untuk menampilkan kategori produk yang tidak hanya memiliki jumlah pembelian tertinggi secara umum tetapi juga banyak jumlah dibeli dalam satu transaksi. Ini membantu mengidentifikasi kategori produk yang paling diminati dalam setiap pembelian, memungkinkan perusahaan untuk merancang strategi bundling produk yang lebih efektif, mengelola stok dengan lebih baik, dan mengoptimalkan penjualan dengan fokus pada kategori produk yang sering dibeli secara bersamaan oleh pelanggan.



Gambar 16. Row chart rate category product

- Visualisasi bar chart untuk mengetahui jumlah mayoritas gender dari pelanggan bertujuan untuk menampilkan secara visual perbandingan jumlah pelanggan berdasarkan jenis kelamin, memudahkan perusahaan dalam mengidentifikasi gender mayoritas di antara pelanggan. Dengan informasi ini, perusahaan dapat menyesuaikan strategi pemasaran, pengembangan produk, serta pengalaman pelanggan untuk lebih sesuai dengan preferensi dan kebutuhan dari mayoritas gender pelanggan, mendukung peningkatan personalisasi layanan dan efektivitas kampanye pemasaran.



Gambar 17. Bar char Gender pelanggan.

II.4 Hasil Final Project

Project telah berhasil membangun infrastruktur ETL solid menggunakan Airflow untuk manajemen efisien siklus ekstraksi, transformasi, dan pemuatan data ke PostgreSQL. Penggunaan fungsi dalam DAG dan integrasi library kunci memastikan pengolahan data dari berbagai format dengan konsistensi. Penambahan tabel baru dalam model data meningkatkan informasi yang ada. Desain ERD yang terstruktur memfasilitasi integrasi antartabel. Melalui visualisasi data menggunakan Metabase, proyek ini memberikan wawasan mendalam tentang pola pembelian, preferensi produk, dan demografi pelanggan, memberikan landasan yang kuat bagi pengambilan keputusan bisnis yang tepat.

Bab III . Aktivitas Mingguan

Minggu	Kegiatan
1	Minggu ini, saya mengikuti onboarding Kampus Merdeka untuk memahami akses situs web dan aturan MSIB pada Senin. Pada Selasa, saya terlibat dalam pertemuan pengenalan platform pembelajaran dan diberi tugas untuk initial assessment sebelum kelas dimulai. Saya mencoba menyelesaikan tugas tersebut pada Rabu dan melanjutkan pekerjaannya pada Kamis sampai selesai. Pada Jumat, saya menghadiri onboarding Dibimbing.id yang membahas akses LSM Dibimbing, program, dan aturan, serta diharapkan berkenalan dengan peserta lainnya. Seminggu ini saya fokus pada pembelajaran dari Kampus merdeka juga dari mitra Dibimbing.id dan tugas yang diberikan.
2	Minggu ini saya membaca dan belajar tentang data warehouse sendiri serta modul dari mitra. lalu ikut pelajaran yang membahas data warehouse lebih dalam, seperti sejarahnya, cara kerjanya, dan alat yang digunakan. Saya juga mengerjakan tugas dan tes terkait itu. Saya juga mulai memahami dasar-dasar bahasa pemrograman Python, seperti menjumlahkan angka, mengambil keputusan, mengulang tindakan, dan mengatur kumpulan data. Saya juga mengevaluasi kesalahan post test yang saya buat sebelumnya dan mengulang pelajaran agar lebih paham. minggu ini membangun pengetahuan tentang data warehouse dan pemrograman Python.
3	Minggu ini saya memilih untuk mengambil inisiatif belajar mandiri dari modul Python yang disediakan oleh mitra. Selasa, saya mendalami Pandas, menguasai proses memuat, manipulasi data, dan penyimpanan hasil analisis dalam berbagai format file. Rabu, tugas Python tentang manipulasi data dan penyimpanan dalam format JSON sukses saya selesaikan. Kamis, fokus saya

	adalah memperdalam pemahaman tentang OOP, Parallel Computing, dan konsep class-object, sambil menyelesaikan tugas terkait OOP. Jumat, saya menyelesaikan tugas OOP dengan sukses, yaitu membuat program untuk menghitung luas berbagai bangun datar, sambil tetap mengikuti sesi mentoring di jumat malam .
4	Minggu ini, saya mengalokasikan waktu untuk memahami dasar-dasar Git melalui pembelajaran online. Pada Selasa, saya berhasil memahami konsep-konsep penting seperti pembuatan repository, commit, push, pull, cloning, dan branching. Kemudian, pada Kamis, saya mengikuti kelas perdana tentang SQL dan memperoleh pemahaman dasar, termasuk pembuatan tabel, penggunaan klausa WHERE, pernyataan agregat, dan pernyataan kondisional CASE WHEN. Di samping kelas, saya juga menghabiskan waktu membaca modul SQL yang disediakan. Pada Jumat, saya menyelesaikan tugas kolaborasi Git dengan anggota kelompok, mengelola repositori, dan memecahkan konflik yang muncul.
5	Minggu ini, saya melanjutkan pembelajaran SQL secara mandiri jika tidak ada jadwal kelas, menggali topik seperti "Join Statement" dan "Union Operation" pada Selasa, dan "Dasar-Dasar Database dan SQL" pada Rabu. Kemudian, pada Kamis, saya mengikuti kelas yang mencakup penggunaan ERD, schema SQL, dan operasi SQL penting. Jumat saya habiskan mengerjakan tugas SQL yang diberikan oleh mitra. Sementara itu, Sabtu saya mengikuti kelas softskill yang membahas pentingnya CV dan profil LinkedIn, personal branding, dan cara lolos dari sistem ATS. Semua ini merupakan langkah penting dalam persiapan menjadi seorang digital talent yang diinginkan oleh perusahaan
6	Selama seminggu ini, saya belajar tentang manajemen data dalam database. Pada hari Selasa, dalam pertemuan online, saya memahami Stored Procedures, Query Optimization, dan Window Functions dalam SQL. Pada Kamis, saya juga mempelajari NoSQL databases seperti MongoDB dan Redis, serta menyelesaikan posttest. Pada Jumat, saya melakukan tugas SQL Query practices yang melibatkan perancangan tabel dan implementasi fungsi dalam SQL. Meskipun tidak ada pertemuan online pada Senin dan Rabu, saya terus belajar sendiri.

7	Minggu ini, saya fokus pada persiapan untuk midtest dengan mempelajari SQL dan Python serta mengikuti ujian praktikum midtest secara mandiri. Kemudian, saya mengeksplorasi materi baru tentang Data Warehouse pada hari Rabu dan berpartisipasi dalam online meeting pada hari Kamis untuk memahami konsep utamanya, termasuk Arsitektur Data Warehouse dan Data Marts. Pada Sabtu, saya mengikuti online meeting tentang softskill untuk persiapan pencarian kerja, seperti pembuatan CV, profil LinkedIn, dan wawancara perilaku.
8	Minggu ini mempelajari fokus pada Data Modeling dan ETL (Extract Transform Load) dalam konteks Data Warehouse. Saya memahami dua pendekatan penting dalam pemodelan data, yaitu OLTP dan OLAP Data Modeling, serta konsep dasar ETL, termasuk pengambilan data, transformasi, dan pemuatan ke dalam Data Warehouse. Selain itu, saya meluangkan waktu untuk belajar mandiri melalui modul mitra. Saya juga berhasil menyelesaikan tugas pemodelan data yang diberikan oleh mitra. Melalui sesi mentoring mingguan, saya mendiskusikan masalah dan pemahaman saya dan menerima panduan lebih lanjut.
9	Minggu ini, perjalanan pembelajaran saya dimulai dengan pemahaman mendalam tentang Docker, termasuk konsep Docker Volume dan Docker Network serta kemampuan menyesuaikan gambar Docker dengan Dockerfile. Kami juga menjalankan Docker Compose untuk mengelola aplikasi berbasis beberapa kontainer dengan konfigurasi yang mudah diatur. Selanjutnya, saya memfokuskan diri pada pembelajaran mandiri melalui modul yang disediakan oleh mitra terkait Docker. Pada pertengahan minggu, beralih ke Apache Airflow, yang merupakan sistem manajemen aliran kerja yang kuat, dan mulai memahami bagaimana Docker dapat digunakan dalam konteks Airflow untuk memudahkan pengelolaan aliran pekerjaan. Di akhir pekan menghadiri kelas softskill untuk mempelajari pengembangan keterampilan komunikasi teknis dengan fokus pada penjelasan yang jelas dan relatable serta kemampuan merespons pertanyaan dan kekhawatiran anggota tim non-teknis.
10	Minggu ini, saya mengikuti serangkaian pelajaran yang mencakup konfigurasi Apache Airflow, penggunaan database terkait, pemahaman konsep Directed Acyclic Graphs (DAGs), sensor untuk pemantauan, serta implementasi DAGs dengan berbagai

	<p>tugas dan operator. Saya juga belajar menjadwalkan tugas dalam Airflow untuk otomatisasi. Pada hari Kamis, pelajaran terfokus pada pemeliharaan dan pemantauan alur kerja serta pembangunan pipa produksi dalam Airflow, memberikan pemahaman lebih mendalam tentang manajemen alur kerja, pemecahan masalah, dan integrasi alat eksternal. Semua pengetahuan ini akan berguna dalam kemampuan saya untuk mengelola alur kerja yang kritis dan kompleks dalam pekerjaan saya. Pada hari Jumat, saya mengerjakan tugas terkait ETL dengan Airflow DAGs sebagai penerapan langsung dari materi yang telah saya pelajari.</p>
11	<p>Minggu ini, saya belajar secara mandiri tentang batch, streaming, dan cloud computing. Pada Selasa, saya mengikuti kelas pengolahan data dalam batch dan streaming. Hari Rabu, berdiskusi dengan mentor tentang tugas DAGs. Kemudian, pada Kamis, saya memahami konsep cloud computing dan layanan terkait. Saya juga mengikuti kelas softskill pada Sabtu untuk memahami dampak data pada bisnis dan memahami lebih lanjut bagaimana data dapat membantu bisnis dan mengukur kesuksesannya.</p>
12	<p>Minggu ini, saya menghabiskan waktu untuk memahami modul web scraping dari mitra kami. Selanjutnya, saya mengikuti kelas tentang dasar API dan berlatih menggunakan Postman. Saya juga berpartisipasi dalam sesi mentoring untuk mendiskusikan topik seputar cloud computing. dan juga, saya mengikuti kelas mengenai teknik web scraping menggunakan Python. Terakhir, saya melakukan pencarian mandiri mengenai data governance. Semua ini dapat meningkatkan pemahaman dan keterampilan saya dalam mengelola dan mengumpulkan data dari berbagai sumber online.</p>
13	<p>Minggu ini, saya melakukan penelusuran sumber tentang Data Governance, menghadiri kelas konsep pengelolaan data, dan mempersiapkan diri untuk ujian final. Ujian tersebut menguji keterampilan coding dan pengetahuan praktis saya. Selain itu, saya menyelesaikan tugas artikel tentang Data Governance yang saya publikasikan, serta mengikuti kelas tentang "Storytelling with Data" untuk mengasah kemampuan visualisasi informasi bagi pengambil keputusan.</p>

14	Pada minggu ini, fokus saya adalah mempersiapkan final project secara mandiri. Saya mengikuti kelas mentoring untuk membahas detail pekerjaan terkait proyek, termasuk ETL menggunakan Airflow, pemodelan data di PostgreSQL, dan visualisasi dalam Docker. Saya juga menghabiskan waktu untuk memahami data proyek dan menjalankan contoh template menggunakan Airflow dan Metabase di Docker, yang diberikan oleh mitra. Selain itu, saya mengikuti kelas softskill yang membahas tanggung jawab dan karier sebagai seorang Data Engineer.
15	Minggu ini, fokus saya adalah memahami konsep ETL di Airflow, mencoba membuat contoh ekstraksi DAGs, dan menghadapi kendala saat memuat data ke PostgreSQL. Saya berdiskusi dengan mentor dan tim untuk mengatasi masalah ini, dan juga mengikuti kelas Project Management untuk memperdalam pengetahuan tentang manajemen proyek data.
16	Minggu ini saya mengalokasikan waktu untuk menyelesaikan permasalahan dalam proses ETL menggunakan SQLAlchemy, yang memungkinkan saya membangun tabel, mengeksekusi query, dan menyelesaikan proses ETL dalam satu dags. Saya juga hadir dalam kelas mentoring untuk memeriksa kanprogres tim saya . serta mempelajari ulang perintah SQL dan menambahkan tabel baru pada proyek saya. Setelah berhasil memuat data dari 8 file menjadi 8 tabel di PostgreSQL, saya fokus pada visualisasi data dari tabel-tabel tersebut, menghasilkan insight melalui berbagai grafik, seperti bar chart untuk melihat produk dengan pembelian terbanyak dan row chart untuk kategori produk dengan jumlah pesanan tertinggi. Saya juga mengikuti kelas softskill tentang kepemimpinan tim, menggali pentingnya komunikasi efektif, manajemen konflik, dan beragam strategi kepemimpinan. Mentoring bersama mentor juga membantu saya mengevaluasi progres dan mengatasi kendala dalam project saya.

17	Minggu ini fokus saya terutama pada penyelesaian proyek. Saya mengalokasikan waktu untuk membuat dashboard dengan tiga visualisasi yang telah saya buat, serta menyusun penjelasan laporan project dalam bentuk presentasi untuk final project. Setiap hari, saya terus bekerja pada penyelesaian laporan, dan berkesempatan untuk mendapatkan umpan balik dari mentor saat sesi mentoring. Puncaknya adalah pada hari Jumat, di mana saya menghadiri sesi presentasi final project untuk menyajikan hasil kerja tim kami.
18	Minggu ini, fokus menyelesaikan laporan akhir. Tanpa jadwal pembelajaran resmi dari mitra, saya menghabiskan waktu Senin hingga Jumat untuk menyelesaikan laporan tersebut dengan konsisten. Semua hari digunakan untuk progres dalam penyusunan laporan akhir.
19	minggu ini, fokus utama saya adalah membuat laporan akhir. Senin hingga Kamis, hari-hari tanpa jadwal pembelajaran resmi, saya tetap konsisten dalam proyek penyusunan laporan. Pada Jumat, saya menghadiri acara kelulusan mitra, memperingati perpisahan bagi peserta MSIB batch 5.
20	Minggu ini, fokus saya tetap pada penyusunan laporan akhir. Setiap hari, tanpa adanya agenda atau jadwal pembelajaran resmi, saya terus konsisten dalam proyek pembuatan laporan

Bab IV Penutup

IV.I Kesimpulan

Kesimpulan selama penulis mengikuti program Studi Independen Modeling And Optimization Techniques In Data Warehousing Bootcamp adalah:

1. Penulis banyak mendapatkan ilmu baru, wawasan, hard skill dan softskill, pengalaman, dan relasi baru selama menjalani program.
2. Dibimbing.id memfasilitasi banyak hal yang bisa membantu peserta mempelajari hal-hal mulai dari segi pembelajaran yang sangat membantu penulis untuk mendalami ilmu tentang Data Engineer, tugas-tugas dan proyek yang diberikan selama program ini, serta informasi lain yang sangat berguna untuk kehidupan karir kedepannya.

IV.II Saran

Saran yang diberikan selama penulis mengikuti program Studi Independen adalah :

1. Saran untuk Kampus Merdeka: Sederhanakan log book dan deskripsi bantuan akses seluruh fitur pada website kampus merdeka.
2. Saran untuk Dibimbing.id: Tingkatkan pendampingan peserta dengan menangani kendala yang mungkin timbul, periksa seluruh aktifitas pengerjaan tugas agar setiap anggota kelompok bisa berkontribusi semua dan sesuaikan penggunaan aplikasi dengan device peserta agar peserta bisa mengikuti pembelajaran dengan baik
3. Saran untuk Studi Independen: Berikan kesempatan kolaborasi dengan praktisi dibidang terkait melalui webinar, diskusi panel, atau kegiatan networking.

Referensi

- [1] Kampus Merdeka. (2023). *Activity Detail Modeling And Optimization Techniques In Data Warehousing Bootcamp* pada situs resmi kemdikbud.go.id.
<https://kampusmerdeka.kemdikbud.go.id/activity/active/detail/5898650>
- [2] Website Lms Dibimbing. (2023). *Modeling And Optimization Techniques In Data Warehousing Bootcamp* <https://bootcamp.dibimbing.id/students/active-class>

Lampiran

Berikut merupakan seluruh file dan dokumen pada final project:

https://github.com/Mfisabil321/MSIB_Final_project_15