

BAB III

PELAKSANAAN KERJA MAGANG

3.1 Kedudukan dan Koordinasi

Selama pelaksanaan program magang di PT Eka Satya Puspita (ESP), saya ditempatkan sebagai Data Analyst Intern di bawah koordinasi langsung Divisi Operational IT. Posisi ini menempatkan saya dalam tim yang bertanggung jawab terhadap pengelolaan data logistik perusahaan, yang mencakup tahapan pengumpulan, pembersihan, analisis, hingga visualisasi data. Peran yang saya jalankan tidak hanya bersifat teknis, tetapi juga memiliki dimensi strategis, karena hasil kerja yang saya lakukan memberikan kontribusi langsung dalam mendukung proses pengambilan keputusan berbasis data di lingkungan perusahaan.

Sejak hari pertama, saya terlibat aktif dalam proses kerja yang sistematis dan berorientasi pada target. Alur kerja dimulai dari pembaruan data logistik yang diterima dari Supervisor IT maupun tim operasional. Setelah data diterima, saya mengikuti diskusi rutin bersama Supervisor IT dan tim Operational IT untuk memahami konteks bisnis dari data yang akan dianalisis. Diskusi ini mencakup penjelasan mengenai ekspektasi hasil analisis, tantangan yang dihadapi oleh tim operasional, serta penentuan metrik-metrik utama yang menjadi fokus analisis.

Setelah sesi diskusi, saya berkoordinasi dengan tim IT lapangan untuk melakukan validasi data yang telah dikumpulkan. Apabila ditemukan ketidaksesuaian atau kesalahan, saya melakukan perbaikan data berdasarkan standar yang ditetapkan perusahaan. Jika data telah dinyatakan valid, saya melanjutkan ke tahap pengumpulan dan pembersihan data (data cleaning), yang meliputi penghapusan duplikasi, penyesuaian format, dan penanganan data yang hilang (missing values).

Data yang telah dibersihkan kemudian dianalisis untuk mengidentifikasi pola-pola penting, seperti tren keterlambatan pengiriman, distribusi beban armada yang tidak merata, dan wilayah dengan volume pengiriman tertinggi. Hasil analisis tersebut saya visualisasikan menggunakan Microsoft Excel agar informasi yang dihasilkan lebih mudah dipahami oleh tim dan manajemen.

Setelah visualisasi selesai, saya berkoordinasi kembali dengan senior data analyst untuk memperoleh masukan dan melakukan revisi apabila diperlukan. Setelah mendapat persetujuan, saya menyusun laporan akhir yang memuat hasil analisis dan visualisasi data. Laporan yang telah disetujui kemudian saya presentasikan kepada manajemen. Pada tahap akhir, dilakukan evaluasi hasil bersama supervisor, dan temuan yang diperoleh dijadikan dasar untuk pengambilan keputusan strategis oleh manajemen. Proses kerja ini berlangsung secara berulang

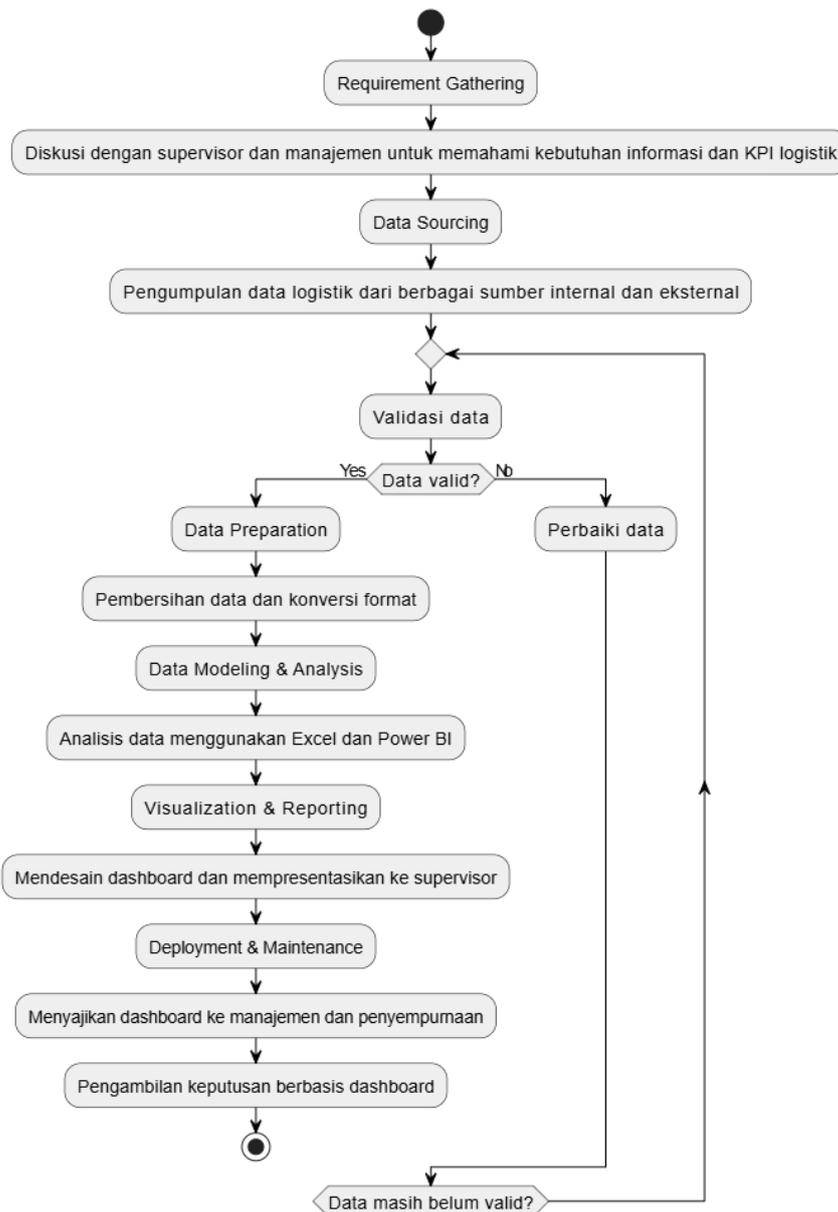
setiap minggu, dengan jenis data dan kebutuhan analisis yang dapat berubah sesuai prioritas serta tantangan yang dihadapi oleh tim logistik.

Dalam melaksanakan tugas tersebut, saya menerapkan prinsip-prinsip pengolahan data yang sistematis menggunakan pendekatan Business Intelligence (BI) Life Cycle. Pendekatan ini mencakup tahapan pengumpulan kebutuhan bisnis (*requirement gathering*), pengumpulan data (*data sourcing*), persiapan dan pengolahan data (*data preparation*), analisis dan pemodelan data (*data modeling & analysis*), visualisasi dan pelaporan (*visualization & reporting*), serta implementasi dan pemeliharaan (*deployment & maintenance*). Kerangka kerja ini membantu saya dalam memastikan setiap proses pengolahan data dilakukan dengan efektif, efisien, dan selaras dengan kebutuhan manajemen.

Selama proses tersebut, kolaborasi menjadi elemen penting yang mendukung keberhasilan pekerjaan. Saya secara rutin bekerja sama dengan berbagai divisi, termasuk tim IT lapangan, supervisor IT, dan senior data analyst. Kolaborasi ini memungkinkan saya untuk memahami kebutuhan operasional yang dinamis serta memastikan bahwa solusi yang dikembangkan benar-benar sesuai dengan ekspektasi dan tantangan yang dihadapi oleh perusahaan.

Selain memberikan kontribusi kepada perusahaan, pengalaman ini juga memberikan manfaat yang signifikan bagi pengembangan kompetensi pribadi saya. Saya tidak hanya meningkatkan kemampuan teknis dalam pengelolaan dan analisis data, tetapi juga mengasah keterampilan komunikasi, problem solving, dan kerja tim. Keterampilan ini menjadi bekal yang sangat berharga untuk mendukung karier profesional saya di bidang data analytics di masa depan.

Alur kerja dan koordinasi selama masa magang digambarkan pada Gambar 3.1 berikut:



Gambar 3. 1 Alur Koordinasi Magang Sebagai Data Analyst Intern di PT Eka Satya Puspita.

Gambar ini menunjukkan tahapan koordinasi kerja yang dilakukan selama masa magang, dimulai dari proses **pembaruan data logistik**, **diskusi dengan Supervisor IT**, **validasi data**, hingga tahap **pengumpulan dan pembersihan data**. Data yang telah diproses kemudian dianalisis dan divisualisasikan menggunakan Excel sebelum dilakukan koordinasi lanjutan dengan senior data analyst. Hasil analisis kemudian disusun dalam laporan akhir yang disampaikan kepada manajemen untuk dievaluasi dan dijadikan dasar dalam pengambilan

keputusan. Seluruh proses ini berlangsung secara berulang setiap minggu selama periode magang.

3.2 Tugas dan Uraian Kerja Magang

Selama pelaksanaan program kerja magang yang berlangsung selama 4 bulan, mahasiswa melakukan berbagai pekerjaan magang terkait bidang **Data Analyst Intern**. Pekerjaan magang tersebut dilakukan dalam proses implementasi dashboard logistik pada **PT Eka Satya Puspita (ESP)** yang sedang berjalan. Berikut ini merupakan tabel realisasi dan uraian kerja magang yang telah dikerjakan selama menempati posisi **Data Analyst Intern**.

Table 2.1 Tabel Realisasi Pelaksanaan Magang

No	Kegiatan	Tanggal Mulai	Tanggal Selesai
Requirement Gathering (Pengumpulan Kebutuhan Bisnis)			
1	Memahami kebijakan, etika, dan regulasi perusahaan.	10 Februari 2025	14 Februari 2025
2	Memahami proses bisnis dan alur distribusi logistik di perusahaan.	17 Februari 2025	21 Februari 2025
3	Mengidentifikasi kebutuhan informasi dan metrik kinerja yang dibutuhkan manajemen logistik.	24 Februari 2025	28 Februari 2025
4	Menganalisis kinerja pesaing utama di industri logistik (seperti CKL Cargo).	3 Maret 2025	7 Maret 2025
Data Sourcing (Pengumpulan Data)			
5	Mengumpulkan data dari berbagai sumber internal dan eksternal.	10 Maret 2025	14 Maret 2025
6	Memvalidasi dan memastikan keakuratan serta kelengkapan data.	17 Maret 2025	21 Maret 2025
Data Preparation (Persiapan dan Pengolahan Data)			
7	Membersihkan data dari kesalahan dan duplikasi.	24 Maret 2025	31 Maret 2025
8	Mengonversi data ke format yang sesuai untuk analisis dan visualisasi.	1 April 2025	4 April 2025

Data Modeling & Analysis (Pemodelan dan Analisis Data)			
9	Melakukan analisis data menggunakan Excel dan Power BI.	7 April 2025	18 April 2025
10	Mendesain dan membangun dashboard informasi untuk mendukung pengambilan keputusan logistik.	21 April 2025	9 Mei 2025
11	Menyempurnakan desain dashboard berdasarkan umpan balik supervisor.	12 Mei 2025	23 Mei 2025
Visualization & Reporting (Visualisasi dan Pelaporan)			
12	Mengevaluasi akurasi dan kejelasan informasi pada dashboard.	26 Mei 2025	30 Mei 2025
13	Menguji fungsionalitas dan kegunaan dashboard bersama supervisor dan tim.	2 Juni 2025	6 Juni 2025
Deployment & Maintenance (Implementasi dan Pemeliharaan)			
14	Menyusun laporan hasil analisis dan pengembangan dashboard.	9 Juni 2025	13 Juni 2025
15	Menyajikan dashboard dan laporan kepada manajemen.	16 Juni 2025	20 Juni 2025
16	Berkolaborasi dengan tim untuk penerapan dashboard di operasional harian dan melakukan pemeliharaan awal.	23 Juni 2025	24 Juni 2025

Kegiatan magang yang dilakukan mahasiswa mengikuti tahapan yang sistematis berdasarkan Business Intelligence (BI) Life Cycle yang terdiri dari requirement gathering, data sourcing, data preparation, data modeling & analysis, visualization & reporting, serta deployment & maintenance. Setiap tahapan diselesaikan dalam rentang waktu tertentu dan didokumentasikan dengan baik untuk memastikan keberhasilan pengembangan dashboard informasi yang dapat digunakan oleh manajemen PT Eka Satya Puspita (ESP) dalam mendukung pengambilan keputusan logistik berbasis data.

Selama proses tersebut, mahasiswa berkolaborasi secara aktif dengan supervisor serta berbagai divisi terkait untuk memastikan keakuratan data, relevansi metrik yang digunakan, dan efektivitas solusi visualisasi yang dihasilkan. Kolaborasi ini juga memungkinkan mahasiswa untuk memahami kebutuhan operasional perusahaan secara lebih mendalam serta mengadaptasikan dashboard

yang dikembangkan agar sesuai dengan dinamika bisnis dan kebutuhan manajerial di PT ESP.

3.2.1 Deskripsi Umum Tugas

Selama menjalani program magang di PT Eka Satya Puspita (ESP), saya ditempatkan sebagai Data Analyst Intern di bawah koordinasi langsung Divisi Operational IT. Posisi ini menempatkan saya dalam tim yang bertanggung jawab untuk mendukung pengelolaan data logistik perusahaan. Tugas utama saya meliputi berbagai proses kerja, mulai dari pengumpulan data pengiriman barang, validasi dan pembersihan data, analisis data, hingga penyajian hasil analisis dalam bentuk visualisasi interaktif yang dapat digunakan oleh tim operasional dan manajemen sebagai dasar pengambilan keputusan.

Lingkup kerja saya tidak hanya terbatas pada kegiatan teknis pengolahan data, tetapi juga mencakup pemahaman konteks bisnis yang mendasari kebutuhan analisis. Sejak hari pertama magang, saya terlibat dalam diskusi rutin bersama Supervisor IT dan tim operasional untuk memahami proses bisnis, ekspektasi manajemen terhadap hasil analisis, serta tantangan yang dihadapi oleh tim operasional di lapangan. Diskusi ini menjadi fondasi penting dalam menentukan metrik-metrik kinerja utama (Key Performance Indicators/KPI) yang perlu dianalisis dan divisualisasikan.

Tugas saya dimulai dengan proses pengumpulan data pengiriman yang diperoleh dari berbagai sumber internal, termasuk laporan pengiriman harian, data keterlambatan pengiriman, serta data performa vendor maskapai penerbangan yang bekerja sama dengan perusahaan. Setelah data dikumpulkan, saya melakukan validasi data untuk memastikan bahwa data yang digunakan akurat, lengkap, dan sesuai dengan kebutuhan analisis. Apabila ditemukan ketidaksesuaian atau kesalahan dalam data, saya melakukan koordinasi dengan tim IT lapangan dan tim operasional untuk melakukan perbaikan sesuai dengan standar yang telah ditetapkan perusahaan.

Tahapan berikutnya adalah pembersihan data (data cleaning), yang mencakup penghapusan data duplikasi, penyesuaian format data, serta penanganan data yang hilang (missing values). Proses ini penting untuk memastikan bahwa data yang dianalisis benar-benar dapat merepresentasikan kondisi operasional perusahaan secara akurat. Setelah data dinyatakan bersih, saya melanjutkan ke tahap analisis data, yang bertujuan untuk mengidentifikasi pola-pola penting seperti tren keterlambatan pengiriman, distribusi beban armada, serta performa vendor maskapai.

Hasil dari analisis data tersebut kemudian saya visualisasikan menggunakan Microsoft Excel untuk tahap awal, dan kemudian dikembangkan lebih lanjut dalam bentuk dashboard interaktif menggunakan Power BI. Visualisasi ini dirancang agar mudah dipahami oleh tim operasional dan manajemen, sehingga informasi yang dihasilkan dapat langsung digunakan untuk pengambilan keputusan yang lebih cepat dan akurat.

Selain menjalankan tugas rutin tersebut, saya juga mendapatkan tanggung jawab tambahan berupa pengembangan dashboard informasi baru yang dirancang untuk menggantikan dashboard lama berbasis Google Looker Studio. Dashboard lama memiliki keterbatasan dalam hal fleksibilitas, kelengkapan visualisasi, serta kemampuan analisis data yang mendalam. Oleh karena itu, pengembangan dashboard baru difokuskan pada peningkatan interaktivitas, penyajian metrik kinerja utama yang lebih komprehensif, serta integrasi fitur analisis tren waktu yang memungkinkan manajemen untuk memantau performa pengiriman barang dari waktu ke waktu.

Pengembangan dashboard baru dilakukan melalui pendekatan kerja yang sistematis dengan mengikuti kerangka kerja Business Intelligence (BI) Life Cycle. Pendekatan ini mencakup tahapan pengumpulan kebutuhan (requirement gathering), pengumpulan data (data sourcing), persiapan dan pembersihan data (data preparation), pemodelan dan analisis data (data modeling & analysis), visualisasi dan pelaporan (visualization & reporting), serta implementasi dan

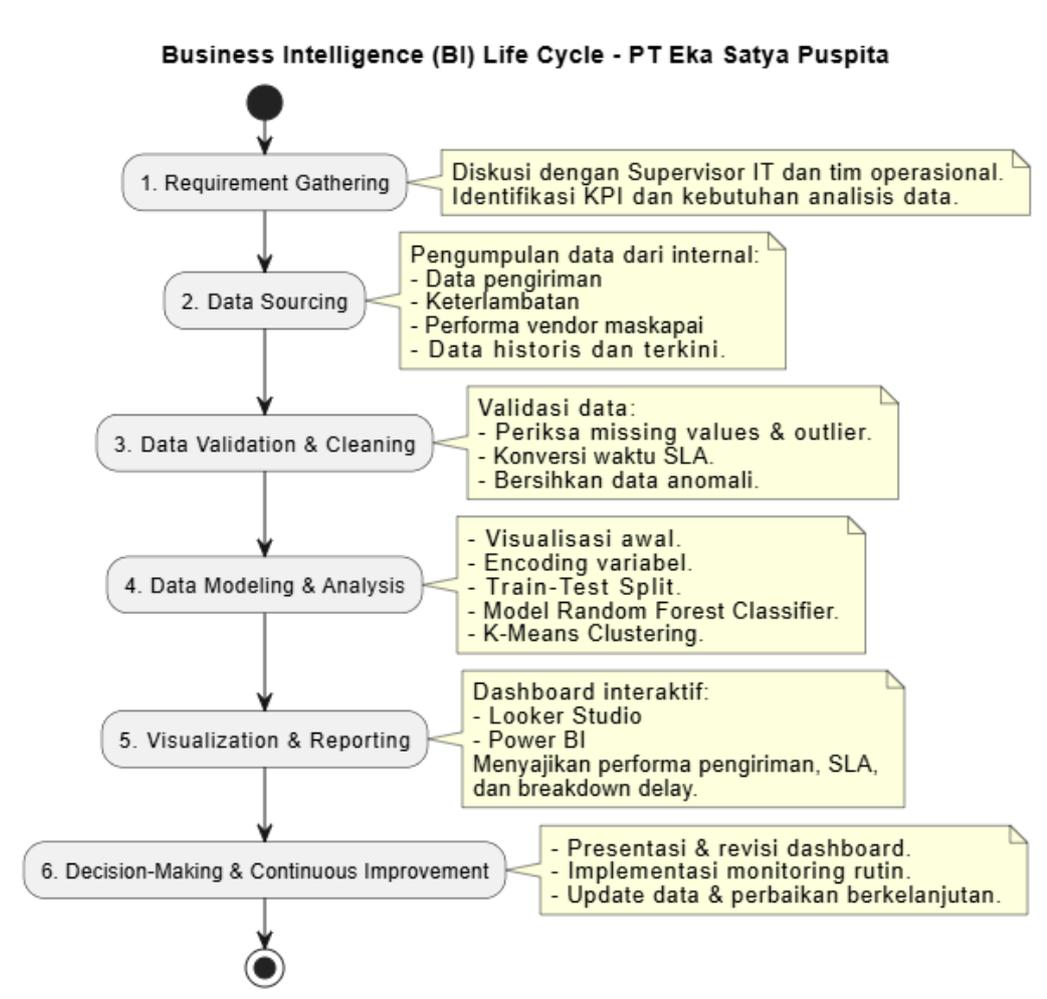
pemeliharaan (deployment & maintenance). Setiap tahapan dilaksanakan dengan mempertimbangkan kebutuhan bisnis yang telah disepakati bersama tim operasional dan manajemen.

Selama proses pengembangan dashboard, saya juga aktif berkoordinasi dengan senior data analyst, Supervisor IT, serta tim operasional untuk mendapatkan masukan dan melakukan revisi terhadap visualisasi yang telah dibuat. Kolaborasi lintas divisi ini tidak hanya memastikan bahwa solusi yang dikembangkan benar-benar sesuai dengan kebutuhan pengguna, tetapi juga memberikan saya kesempatan untuk memahami dinamika kerja di lingkungan profesional serta mengembangkan keterampilan komunikasi dan kerja tim.

Secara keseluruhan, tugas yang saya laksanakan selama program magang tidak hanya memberikan kontribusi nyata bagi perusahaan dalam bentuk peningkatan efisiensi pelaporan dan analisis data, tetapi juga memberikan pengalaman berharga bagi pengembangan kompetensi saya di bidang data analytics. Pengalaman ini memperkuat pemahaman saya tentang pentingnya pengolahan data yang akurat dan penyajian informasi yang efektif dalam mendukung pengambilan keputusan berbasis data di dunia bisnis modern.

3.2.2 Tahapan Pengerjaan Tugas

Dalam melaksanakan tugas sebagai Data Analyst Intern di PT Eka Satya Puspita (ESP), saya mengikuti tahapan kerja yang sistematis agar setiap proses pengolahan data dapat dilakukan secara efektif dan menghasilkan output yang sesuai dengan kebutuhan manajemen. Tahapan-tahapan tersebut disusun berdasarkan pendekatan Business Intelligence (BI) Life Cycle yang terdiri dari enam fase utama, yaitu Requirement Gathering, Data Sourcing, Data Validation & Cleaning, Data Modeling & Analysis, Visualization & Reporting, serta Decision-Making & Continuous Improvement. Setiap tahapan menggambarkan alur kerja mulai dari pengumpulan kebutuhan dan data, proses validasi serta pembersihan data, analisis dan pembuatan model prediktif, pembuatan visualisasi interaktif, hingga implementasi dashboard untuk mendukung pengambilan keputusan berbasis data dan perbaikan berkelanjutan.



Gambar 3. 2 Diagram BI Life Cycle PT Eka Satya Puspita

Gambar 3.2 Diagram BI Life Cycle PT Eka Satya Puspita menampilkan alur proses tersebut yang telah diterapkan selama periode magang, mulai dari diskusi awal dengan tim operasional dan Supervisor IT, pengumpulan dan pembersihan data pengiriman, analisis dan pemodelan data menggunakan algoritma machine learning, hingga pengembangan dashboard interaktif dengan Looker Studio dan Power BI yang digunakan untuk mendukung monitoring operasional serta pengambilan keputusan strategis di perusahaan.

1. Requirement Gathering

Tahapan pertama adalah mengidentifikasi kebutuhan informasi yang dibutuhkan oleh tim operasional dan manajemen. Saya melakukan diskusi bersama Supervisor IT dan tim operasional untuk memahami masalah yang sedang dihadapi,

menentukan indikator kinerja utama (Key Performance Indicators/KPI), serta merumuskan tujuan analisis data yang akan dilakukan. Proses ini juga mencakup penentuan data apa saja yang relevan untuk dianalisis.

2. Data Sourcing

Setelah kebutuhan informasi ditentukan, saya mengumpulkan data dari berbagai sumber internal perusahaan. Data tersebut mencakup informasi pengiriman barang, keterlambatan pengiriman, performa vendor maskapai penerbangan, serta laporan-laporan operasional lainnya. Data yang dikumpulkan terdiri dari data historis maupun data terkini yang dibutuhkan untuk keperluan analisis tren dan performa.

Date	Vendor	SC Origin	SC Destination	LT Number	TO Number	Gross Weight	AWB	Actual Airline Name	Flight No	ATDFlight	ATAFlight	ATAPOD Warehouse	ATDP
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202412312IOPK	30.00	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202412312IOPK	21.63	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202412312IOPK	18.26	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO2024122910M9C	0.35	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202501013AQUU	22.61	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202501013AQUU	19.76	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202412312IOQN	24.15	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202501013BYVI	24.31	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202412312PVYN	25.55	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202412291AWHS	22.77	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202501013BW19	10.46	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202501013BW17	24.24	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202412312VYXZ	18.88	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202412291S4GE	20.73	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202501013B599	15.55	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202412312IUIHZ	10.27	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202501013BW16	2.97	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202412312UIHD	20.35	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202501013BYVI	16.67	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202501013B5FD	11.76	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202412312PVYH	8.86	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202501013BW18	22.22	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202412312V4WA	0.62	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO20241225R8M9C	8.70	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO2024123125W54	18.92	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202501013B5FZ	13.35	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO202412312394Z	31.40	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00
1-Jan-2025	ESP	Kosambi DC	Padang DC	LT0P1110M7P1	TO20241219NDNSN	22.92	888-8166250	Citilink CGK	QG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 12:00

Gambar 3. 3 Gambar Data Source

Data pengiriman dikumpulkan dari sheet Merge Performance yang memuat berbagai kolom penting seperti tanggal pengiriman, vendor, origin, destinasi, berat barang (Gross Weight), nama maskapai, serta informasi penerbangan seperti nomor penerbangan dan waktu keberangkatan. Data ini menjadi sumber utama untuk proses validasi, pembersihan, analisis, hingga pembuatan visualisasi dashboard.

3. Data Validation & Cleaning

Data yang telah dikumpulkan dari berbagai sumber internal perusahaan tidak selalu dalam kondisi siap pakai untuk dianalisis. Oleh karena itu, saya melakukan proses validasi dan pembersihan data (data cleaning) agar kualitas data dapat mendukung analisis yang akurat dan relevan dengan kebutuhan perusahaan. Proses ini dimulai dengan pemeriksaan dimensi data dan tipe data yang tersedia. Berdasarkan hasil pemeriksaan pada dataset yang dinamakan Merge Performance, data yang digunakan memiliki 79.173 baris dan 26 kolom (Gambar 3.4). Setiap kolom memiliki tipe data yang sesuai, seperti object untuk teks dan float untuk data numerik seperti berat barang (Gross Weight).

STEP 1: CEK DIMENSI DATA & TIPE DATA

```
In [2]: import pandas as pd

# Membaca data
file_path = "C:/Users/Lenovo/Downloads/DATASET/Performance ESP MODELING.xlsx"
df = pd.read_excel(file_path, sheet_name="Merge Performance")

# Cek dimensi data
print("Jumlah baris dan kolom:", df.shape)

# Tampilkan nama kolom dan tipe datanya
print("\nTipe data per kolom:")
print(df.dtypes)

# Lihat 5 baris pertama sebagai sampel
print("\nContoh data:")
print(df.head())

Jumlah baris dan kolom: (79173, 26)

Tipe data per kolom:
Date                object
Vendor              object
SC Origin           object
SC Destination      object
LT Number           object
TO Number           object
Gross Weight        float64
AWB                 object
Actual\nAirline Name object
Flight No           object
ATD\nFlight         object
ATA\nFlight         object
ATA\nPOD Warehouse object
ATD\nPOD Warehouse object
```

Gambar 3. 4

Langkah selanjutnya adalah mengecek keberadaan data yang hilang (missing values) di setiap kolom. Berdasarkan hasil analisis, ditemukan bahwa sebagian besar kolom tidak memiliki data yang hilang. Namun, terdapat dua kolom, yaitu Issue dan Remarks, yang memiliki jumlah missing values cukup besar, masing-masing 68.141 dan 68.165 data yang hilang (Gambar 3.5). Selain itu, kolom Target SLA Hour juga ditemukan memiliki sejumlah data kosong. Meskipun demikian, kolom-kolom ini lebih bersifat deskriptif dan tidak menjadi variabel utama dalam

analisis performa SLA sehingga keberadaan missing values di kolom tersebut dapat ditangani dengan strategi tertentu tanpa mengganggu kualitas keseluruhan analisis.

STEP 2: CEK MISSING VALUES & DISTRIBUSI Arrival Status

```
In [5]: # Cek missing values per kolom
print("Jumlah Missing Values per Kolom:\n")
print(df.isnull().sum())

# Cek distribusi Arrival Status
print("\nDistribusi Arrival Status:\n")
print(df['Arrival Status Vendor'].value_counts(dropna=False))
```

Jumlah Missing Values per Kolom:

Date	0
Vendor	0
SC Origin	0
SC Destination	0
LT Number	0
TO Number	0
Gross Weight	0
AWB	0
Actual\nAirline Name	0
Flight No	0
ATD\nFlight	0
ATA\nFlight	1
ATA\nPOD Warehouse	0
ATD\nPOD Warehouse	0
ATA\nVendor WH Destination\n(Sertakan Link Evidence)	0
Dooring Activity Vendor	0
Arrival Status Vendor	0
Issue	68141
REMARKS (Mandatory)	68165
Concat	0
Target SLA Hour	1882
Actual SLA Hour	0
Station	0
Concat Origin-Station	0
Rate	0

Gambar 3. 5

Setelah proses identifikasi missing values, langkah berikutnya adalah memastikan bahwa semua data waktu pada kolom Target SLA Hour dan Actual SLA Hour dikonversi menjadi format jam desimal (float). Proses konversi ini penting agar data waktu dapat dihitung dan dianalisis lebih lanjut, misalnya untuk menghitung rata-rata SLA dan mengidentifikasi outlier. Dengan menggunakan fungsi custom di Python, waktu dalam format HH:MM:SS berhasil diubah menjadi angka desimal (Gambar 3.6). Ini memungkinkan perhitungan lebih lanjut yang diperlukan untuk analisis SLA.

STEP 3: KONVERSI Target SLA & Actual SLA MENJADI JAM (float)

```
In [6]: import numpy as np

# Fungsi untuk mengubah format "HH:MM:SS" ke jam float
def time_to_hours(time_str):
    try:
        h, m, s = map(int, str(time_str).split(':'))
        return h + m / 60 + s / 3600
    except:
        return np.nan

# Terapkan ke kolom
df['Target_SLA_Hours'] = df['Target SLA Hour'].apply(time_to_hours)
df['Actual_SLA_Hours'] = df['Actual SLA Hour'].apply(time_to_hours)

# Cek hasil konversi
print(df[['Target SLA Hour', 'Target_SLA_Hours', 'Actual SLA Hour', 'Actual_SLA_Hours']].head(10))
```

	Target SLA Hour	Target_SLA_Hours	Actual SLA Hour	Actual_SLA_Hours
0	32:00:00	32.0	17:59:00	17.983333
1	32:00:00	32.0	17:59:00	17.983333
2	32:00:00	32.0	17:59:00	17.983333
3	32:00:00	32.0	17:59:00	17.983333
4	32:00:00	32.0	17:59:00	17.983333
5	32:00:00	32.0	17:59:00	17.983333
6	32:00:00	32.0	17:59:00	17.983333
7	32:00:00	32.0	17:59:00	17.983333
8	32:00:00	32.0	17:59:00	17.983333
9	32:00:00	32.0	17:59:00	17.983333

Gambar 3. 6

Setelah data SLA berhasil dikonversi, saya menghitung rata-rata SLA aktual (Actual SLA Hour) dan SLA target (Target SLA Hour) berdasarkan status kedatangan pengiriman (Arrival Status Vendor). Hasilnya menunjukkan perbedaan yang signifikan antara pengiriman yang tepat waktu (On Time) dan yang terlambat (Late). Rata-rata SLA aktual untuk pengiriman yang terlambat jauh lebih tinggi dibandingkan yang tepat waktu (Gambar 3.7). Informasi ini menjadi dasar untuk memahami pola keterlambatan dan faktor-faktor penyebabnya.

STEP 4: CEK RATA-RATA SLA UNTUK SETIAP Arrival Status

```
In [7]: print("Rata-rata Actual SLA Hour berdasarkan Arrival Status:\n")
print(df.groupby('Arrival Status Vendor')['Actual_SLA_Hours'].mean())

print("\nRata-rata Target SLA Hour berdasarkan Arrival Status:\n")
print(df.groupby('Arrival Status Vendor')['Target_SLA_Hours'].mean())
```

Rata-rata Actual SLA Hour berdasarkan Arrival Status:

```
Arrival Status Vendor
Late      9257.604303
On Time  1854.644061
Name: Actual_SLA_Hours, dtype: float64
```

Rata-rata Target SLA Hour berdasarkan Arrival Status:

```
Arrival Status Vendor
Late      29.042057
On Time   29.804920
Name: Target_SLA_Hours, dtype: float64
```

Gambar 3. 7

Selanjutnya, dilakukan pemeriksaan terhadap nilai maksimum SLA aktual untuk mengidentifikasi kemungkinan adanya data pencilan (outlier). Dari hasil analisis, ditemukan bahwa beberapa data memiliki nilai SLA yang sangat besar, bahkan ada yang mencapai 1.577.892 jam atau sekitar 180 tahun. Hal ini jelas menunjukkan adanya anomali data yang kemungkinan disebabkan oleh kesalahan

pencatatan atau proses ETL yang tidak sempurna (Gambar 3.8). Oleh karena itu, saya menetapkan ambang batas outlier sebesar 500 jam. Data yang melebihi nilai ini dianggap tidak wajar untuk konteks logistik dan dihapus dari dataset.

```
STEP 5: CEK NILAI MAKSIMUM Actual SLA Hour

In [8]: print("Nilai maksimum Actual SLA Hour:", df['Actual_SLA_Hours'].max())

# Tentukan batas outlier (misal Lebih dari 500 jam dianggap outlier)
outlier_threshold = 500

# Hitung berapa banyak yang di atas batas
outlier_count = (df['Actual_SLA_Hours'] > outlier_threshold).sum()
print("Jumlah data di atas", outlier_threshold, "jam:", outlier_count)

Nilai maksimum Actual SLA Hour: 1577892.5833333333
Jumlah data di atas 500 jam: 175

Nilai maksimum 1.577.892 jam -- sekitar 180 tahun (!)

Ada 175 baris yang melebihi 500 jam -- jelas tidak masuk akal untuk pengiriman logistik.

Ini memang harus dibersihkan, karena bisa merusak analisis dan model prediksi.
```

Gambar 3. 8

Langkah terakhir dalam proses data cleaning adalah menghapus seluruh data yang melebihi ambang batas SLA tersebut. Setelah proses penyaringan dilakukan, jumlah data berkurang menjadi 78.998 baris (Gambar 3.9). Pembersihan ini memastikan bahwa data yang digunakan untuk analisis selanjutnya telah bebas dari anomali yang dapat mengganggu akurasi hasil analisis. Dengan dilakukannya proses validasi dan pembersihan data secara sistematis ini, kualitas data meningkat secara signifikan. Hal ini memberikan dasar yang kuat untuk tahap analisis selanjutnya, termasuk pembuatan model prediktif dan pengembangan dashboard performa logistik yang akan membantu manajemen dalam mengambil keputusan strategis.

```
STEP 6: FILTER DATA - HAPUS OUTLIER

In [10]: # Buat salinan data tanpa outlier
df_clean = df[df['Actual_SLA_Hours'] <= 500].copy()

print("Jumlah data setelah hapus outlier:", df_clean.shape[0])

Jumlah data setelah hapus outlier: 78998
```

Gambar 3. 9

4. Data Modeling & Analysis

Setelah seluruh proses validasi dan pembersihan data selesai dilakukan, tahap berikutnya yang saya kerjakan adalah *Data Modeling & Analysis*. Tahapan ini bertujuan untuk menggali pola-pola tersembunyi dalam data pengiriman,

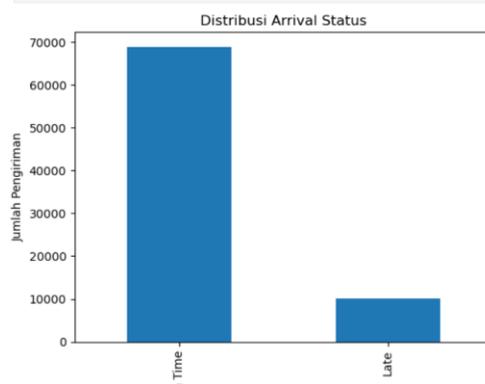
membangun model prediktif untuk memproyeksikan potensi keterlambatan pengiriman, serta melakukan segmentasi data agar perusahaan dapat memahami karakteristik pengiriman dengan lebih baik. Pada tahap ini, dilakukan serangkaian proses mulai dari visualisasi data, transformasi variabel, pembangunan model klasifikasi, hingga segmentasi pengiriman menggunakan teknik clustering. Setiap langkah dilakukan dengan hati-hati dan disesuaikan dengan kebutuhan analisis yang telah dirumuskan sebelumnya bersama tim IT dan supervisor.

Langkah awal pada tahap ini adalah melakukan visualisasi data untuk mendapatkan pemahaman menyeluruh terkait distribusi variabel-variabel penting yang akan dianalisis lebih lanjut. Visualisasi membantu dalam mengidentifikasi pola umum, distribusi nilai, serta potensi outlier yang dapat mempengaruhi hasil modeling. Setelah visualisasi, dilakukan transformasi data termasuk encoding variabel kategorikal dan pembagian data menjadi data latih dan data uji. Data yang telah disiapkan ini kemudian digunakan untuk membangun model klasifikasi menggunakan algoritma *Random Forest Classifier*. Selain itu, dilakukan juga proses clustering dengan metode *K-Means* untuk mengelompokkan pengiriman berdasarkan karakteristik tertentu.

Berikut adalah penjelasan detail setiap tahapan modeling dan analisis yang telah dilakukan.

STEP 7: VISUALISASI — DISTRIBUSI Arrival Status

```
In [11]: import matplotlib.pyplot as plt
# Plot distribusi Arrival Status
df_clean['Arrival Status Vendor'].value_counts().plot(kind='bar')
plt.title('Distribusi Arrival Status')
plt.xlabel('Arrival Status')
plt.ylabel('Jumlah Pengiriman')
plt.show()
```

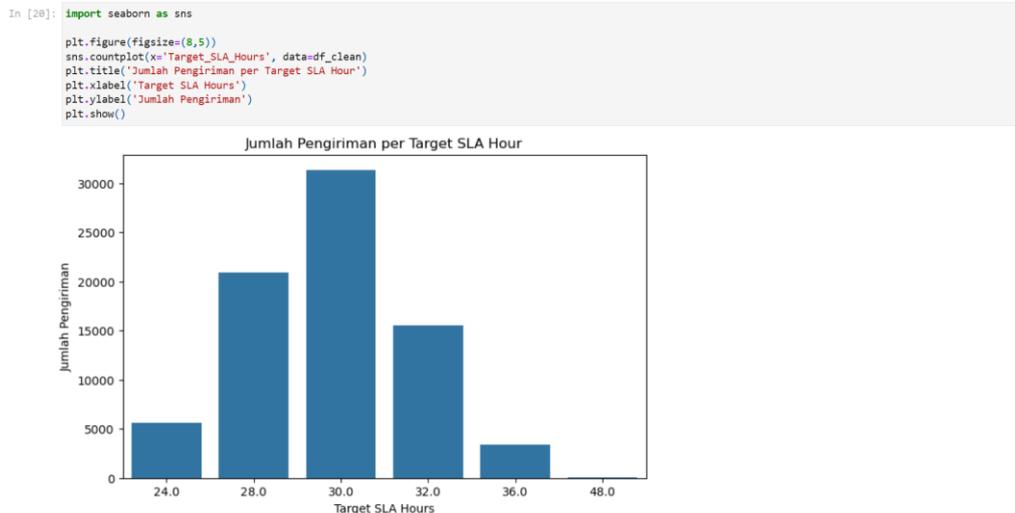


Gambar 3. 10 Distribusi Arrival Status

Langkah awal dalam proses analisis data adalah memahami distribusi dari *Arrival Status*, yaitu status kedatangan pengiriman apakah tepat waktu (*On Time*) atau terlambat (*Late*). Untuk mendapatkan gambaran yang jelas, saya membuat visualisasi berupa diagram batang (*bar chart*) yang menunjukkan jumlah pengiriman pada masing-masing kategori *Arrival Status*.

Hasil visualisasi yang ditampilkan pada Gambar 3.10 menunjukkan bahwa sebagian besar pengiriman berhasil tiba tepat waktu, dengan jumlah pengiriman mencapai lebih dari 70.000 transaksi. Sementara itu, pengiriman yang mengalami keterlambatan relatif lebih sedikit, hanya sekitar 10.000 transaksi. Distribusi ini mencerminkan bahwa secara umum performa pengiriman berjalan dengan baik, meskipun masih terdapat sejumlah pengiriman yang tidak memenuhi target waktu (*SLA*).

Visualisasi ini sangat penting sebagai dasar analisis lebih lanjut. Dengan mengetahui proporsi pengiriman yang berhasil dan yang terlambat, saya dapat menentukan apakah perlu dilakukan investigasi lebih lanjut terhadap faktor-faktor yang menyebabkan keterlambatan. Selain itu, informasi ini juga menjadi acuan untuk proses segmentasi data dan pembangunan model klasifikasi pada tahap berikutnya.



Gambar 3. 11 Jumlah Pengiriman per Target SLA Hour

Setelah memahami distribusi *Arrival Status*, langkah berikutnya adalah menganalisis distribusi pengiriman berdasarkan *Target SLA Hour*. SLA (Service Level Agreement) adalah waktu target pengiriman yang ditetapkan agar suatu pengiriman dianggap berhasil tepat waktu. Pada Gambar 3.11, ditampilkan visualisasi distribusi jumlah pengiriman terhadap nilai *Target SLA Hour* yang tersedia di dataset.

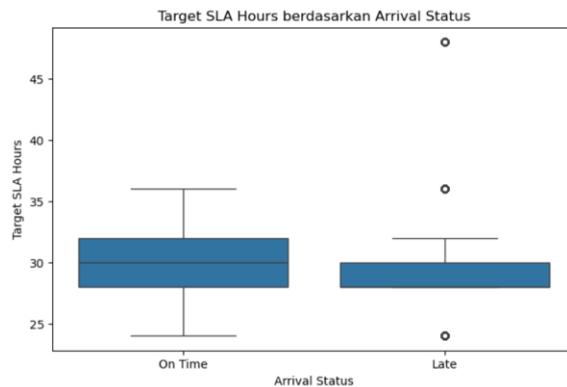
Hasil visualisasi menunjukkan bahwa sebagian besar pengiriman memiliki *Target SLA Hour* di kisaran 28 hingga 32 jam. Khususnya, puncak frekuensi pengiriman terlihat pada target SLA sekitar 30 jam dengan lebih dari 30.000 transaksi. Ini menunjukkan bahwa perusahaan menetapkan waktu target pengiriman sekitar 30 jam sebagai standar operasionalnya untuk mayoritas pengiriman. Nilai SLA yang terlalu rendah atau terlalu tinggi relatif jarang terjadi, seperti terlihat pada kategori 24 jam dan di atas 36 jam yang hanya memiliki sedikit jumlah pengiriman.

Distribusi ini memberikan wawasan penting bahwa sebagian besar pengiriman memang difokuskan untuk memenuhi target yang realistis dan dapat dicapai. Selain itu, pola distribusi ini membantu dalam proses analisis lanjutan seperti segmentasi dan pembuatan model prediktif, karena mengetahui nilai SLA dominan dapat mempengaruhi bagaimana algoritma memahami performa pengiriman yang

diharapkan. Pemahaman distribusi ini juga bermanfaat untuk mengidentifikasi SLA yang perlu dievaluasi kembali jika ada kecenderungan keterlambatan pada rentang waktu tertentu.

STEP 9: BOX PLOT — Target SLA Hours vs Arrival Status

```
In [21]: plt.figure(figsize=(8,5))
sns.boxplot(x='Arrival Status Vendor', y='Target_SLA_Hours', data=df_clean)
plt.title('Target SLA Hours berdasarkan Arrival Status')
plt.xlabel('Arrival Status')
plt.ylabel('Target SLA Hours')
plt.show()
```



Gambar 3.12 Box Plot — Target SLA Hours vs Arrival Status

Gambar 3.12 menampilkan *box plot* yang menggambarkan distribusi *Target SLA Hours* berdasarkan kategori *Arrival Status*, yaitu "On Time" dan "Late". Visualisasi ini bertujuan untuk memahami bagaimana variasi nilai *Target SLA Hours* berhubungan dengan status ketepatan waktu pengiriman.

Pada box plot tersebut, dapat diamati bahwa median *Target SLA Hours* untuk pengiriman yang *On Time* dan yang *Late* memiliki nilai yang relatif serupa, yaitu sekitar 30 jam. Namun, distribusi datanya menunjukkan perbedaan yang cukup jelas. Untuk pengiriman yang *On Time*, rentang interkuartil (IQR) atau jarak antara kuartil pertama dan ketiga sedikit lebih lebar, yang menunjukkan adanya variasi target SLA yang lebih besar dalam kategori ini. Selain itu, terdapat beberapa outlier dengan nilai SLA yang lebih tinggi, meskipun jumlahnya tidak signifikan.

Sebaliknya, pada kategori *Late*, distribusi *Target SLA Hours* cenderung lebih sempit dengan sedikit variasi, namun tetap terdapat outlier. Hal ini menunjukkan bahwa keterlambatan pengiriman tidak selalu disebabkan oleh target SLA yang

terlalu ketat, melainkan kemungkinan besar oleh faktor eksternal lain seperti masalah operasional, cuaca, atau ketidaksesuaian proses di lapangan.

Analisis ini penting untuk memahami apakah target SLA yang ditetapkan realistis untuk dicapai dalam kondisi operasional saat ini. Jika nilai median SLA yang *Late* terlalu rendah dibandingkan SLA yang *On Time*, maka bisa jadi SLA perlu disesuaikan agar lebih sesuai dengan kenyataan operasional. Namun, pada hasil ini, kesamaan median menunjukkan bahwa faktor keterlambatan lebih kompleks dan tidak hanya bergantung pada target SLA yang ditetapkan.

Dengan menggunakan *box plot* ini, perusahaan dapat mengevaluasi dan, bila perlu, menyesuaikan target SLA yang lebih realistis untuk meminimalkan keterlambatan tanpa mengorbankan efisiensi dan produktivitas pengiriman.

STEP 10.1: Encoding Target (Arrival Status)

```
In [23]: # Buat kolom target biner
df_clean['Arrival_Status_Binary'] = df_clean['Arrival Status Vendor'].map({'On Time': 0, 'Late': 1})

# Cek distribusi target biner
print("Distribusi Arrival_Status_Binary:\n")
print(df_clean['Arrival_Status_Binary'].value_counts())

Distribusi Arrival_Status_Binary:

Arrival_Status_Binary
0    68666
1     10049
Name: count, dtype: int64
```

STEP 10.2: Encoding Fitur Kategorikal

```
In [24]: from sklearn.preprocessing import LabelEncoder

# Copy dataframe supaya aman
df_model = df_clean.copy()

# List kolom kategorikal
kategori_cols = ['Vendor', 'SC Origin', 'SC Destination']

# Label Encoding
le = LabelEncoder()

for col in kategori_cols:
    df_model[col] = le.fit_transform(df_model[col])
    print(f'Label encoding selesai untuk kolom: {col}')

# Cek hasil encoding
print("\nContoh data setelah encoding:")
print(df_model[['kategori_cols']].head())

Label encoding selesai untuk kolom: Vendor
Label encoding selesai untuk kolom: SC Origin
Label encoding selesai untuk kolom: SC Destination

Contoh data setelah encoding:
  Vendor  SC Origin  SC Destination
0      0           1                11
1      0           1                11
2      0           1                11
3      0           1                11
4      0           1                11
```

STEP 10.3: Split Data Train-Test

```
In [25]: from sklearn.model_selection import train_test_split

# Fitur (X) dan Target (y)
X = df_model[['Gross Weight', 'Target_SLA_Hours', 'Rate', 'Vendor', 'SC Origin', 'SC Destination']]
y = df_model['Arrival_Status_Binary']

# Split data: 88% training, 28% testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)

print("Jumlah data train:", X_train.shape[0])
print("Jumlah data test:", X_test.shape[0])

Jumlah data train: 62972
Jumlah data test: 15743
```

Gambar 3. 13 Proses Encoding Target dan Fitur Kategorikal serta Pembagian Data Train-Test

Setelah proses eksplorasi dan pembersihan data selesai, langkah selanjutnya adalah mempersiapkan data untuk proses pemodelan dan analisis prediktif. Tahap ini dimulai dengan melakukan encoding pada variabel target dan fitur kategorikal agar data dapat diproses oleh algoritma machine learning yang memerlukan input numerik.

Encoding Target (Arrival Status)

Pertama-tama, dilakukan encoding terhadap kolom *Arrival Status Vendor* yang sebelumnya memiliki nilai kategorikal "On Time" dan "Late". Kategori ini dikonversikan menjadi nilai biner dengan menggunakan aturan: *On Time* diberikan label 0 dan *Late* diberikan label 1. Proses ini menghasilkan sebuah fitur baru bernama *Arrival_Status_Binary*. Hasil distribusi menunjukkan bahwa mayoritas pengiriman tergolong *On Time*, sedangkan sebagian kecil lainnya mengalami keterlambatan.

Encoding Fitur Kategorikal

Selanjutnya, dilakukan encoding pada fitur kategorikal lainnya, yaitu *Vendor*, *SC Origin*, dan *SC Destination*. Label encoding digunakan untuk mengubah nilai kategorikal menjadi angka, sehingga fitur-fitur tersebut dapat diinterpretasikan oleh algoritma machine learning. Proses ini memastikan bahwa semua fitur kategorikal diubah menjadi nilai numerik tanpa mengubah makna dari kategori aslinya.

Split Data Train-Test

Setelah seluruh fitur dikonversikan menjadi numerik, dataset dibagi menjadi dua bagian: data pelatihan (training) dan data pengujian (testing). Pembagian dilakukan

dengan proporsi 80% untuk training dan 20% untuk testing menggunakan fungsi `train_test_split` dari `sklearn`. Stratifikasi dilakukan berdasarkan target `Arrival_Status_Binary` untuk memastikan distribusi kelas yang seimbang antara data pelatihan dan pengujian. Hasil split menunjukkan bahwa sebanyak 62.972 baris data digunakan untuk training dan 15.743 baris data untuk testing.

Langkah-langkah ini sangat penting untuk memastikan bahwa data yang digunakan dalam proses pelatihan model machine learning sudah bersih, terstruktur, dan siap untuk digunakan dalam pembuatan model prediktif yang akurat dan andal.

STEP 11: Train Random Forest Classifier

```
In [26]: from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix

# Buat model Random Forest
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)

# Latih model
rf_model.fit(X_train, y_train)

# Prediksi data test
y_pred = rf_model.predict(X_test)

# Evaluasi hasil
print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

print("\nClassification Report:")
print(classification_report(y_test, y_pred))

Confusion Matrix:
[[12917  816]
 [ 1208  802]]

Classification Report:
              precision    recall  f1-score   support

     0       0.91      0.94      0.93    13733
     1       0.50      0.40      0.44     2010

 accuracy          0.71      0.67      0.69    15743
 macro avg          0.71      0.67      0.69    15743
 weighted avg          0.66      0.67      0.67    15743

In [29]: #OPSI LAIN:

In [29]: #OPSI LAIN:
# Random Forest dengan class_weight balanced
rf_model_bal = RandomForestClassifier(n_estimators=100, random_state=42, class_weight='balanced')

# Latih model
rf_model_bal.fit(X_train, y_train)

# Prediksi data test
y_pred_bal = rf_model_bal.predict(X_test)

# Evaluasi hasil
print("Confusion Matrix (Balanced):")
print(confusion_matrix(y_test, y_pred_bal))

print("\nClassification Report (Balanced):")
print(classification_report(y_test, y_pred_bal))

Confusion Matrix (Balanced):
[[12222 1511]
 [ 1040  970]]

Classification Report (Balanced):
              precision    recall  f1-score   support

     0       0.92      0.89      0.91    13733
     1       0.39      0.48      0.43     2010

 accuracy          0.66      0.69      0.67    15743
 macro avg          0.66      0.69      0.67    15743
 weighted avg          0.85      0.84      0.85    15743
```

Gambar 3. 14

Setelah data melalui proses encoding dan pemisahan train-test, tahap selanjutnya adalah membangun model klasifikasi untuk memprediksi *Arrival Status* berdasarkan fitur-fitur yang telah dipersiapkan. Algoritma yang dipilih dalam proses ini adalah **Random Forest Classifier**, karena algoritma ini memiliki keunggulan dalam menangani data dengan fitur kategorikal dan numerikal, serta mampu mengelola masalah *imbalanced class* dengan baik. Proses pelatihan model dilakukan menggunakan data pelatihan (*training data*) dengan proporsi 80% dari total data, sementara sisanya 20% digunakan untuk pengujian (*testing data*).

Pada awalnya, model Random Forest dilatih tanpa penyesuaian bobot kelas (*class weight*). Hasil evaluasi dari model ini menunjukkan bahwa akurasi yang diperoleh mencapai **86%**, dengan nilai *precision* sebesar **0.91** untuk kelas "On Time" dan **0.60** untuk kelas "Late". *Recall* untuk kelas "On Time" juga tinggi di angka **0.94**, namun *recall* untuk kelas "Late" hanya sebesar **0.44**. Hal ini menunjukkan bahwa model cenderung lebih baik dalam mengidentifikasi pengiriman yang tepat waktu (*On Time*), tetapi kurang optimal dalam mendeteksi pengiriman yang terlambat (*Late*). Ketidakseimbangan *recall* antar kelas ini menandakan adanya tantangan dalam menangani ketidakseimbangan data (*class imbalance*) pada target variabel.

Untuk mengatasi permasalahan tersebut, dilakukan pendekatan kedua dengan melatih model Random Forest menggunakan parameter **class_weight='balanced'**. Dengan penyesuaian ini, model memberikan bobot lebih besar kepada kelas minoritas (*Late*) untuk meningkatkan sensitivitas (*recall*) dalam mendeteksi keterlambatan pengiriman. Hasil evaluasi model yang telah disesuaikan menunjukkan peningkatan *recall* untuk kelas "Late" dari **0.44** menjadi **0.48**, meskipun terdapat sedikit penurunan pada metrik *precision* dan *f1-score* secara keseluruhan. Akurasi model setelah penyesuaian menjadi **84%**, dengan *weighted average f1-score* sebesar **0.85**, yang mencerminkan performa model yang cukup stabil dalam menghadapi ketidakseimbangan data.

Secara umum, model Random Forest yang telah diterapkan berhasil menunjukkan kinerja yang baik untuk baseline analisis. Walaupun *recall* untuk kelas keterlambatan masih dapat ditingkatkan, model ini memberikan gambaran awal yang cukup kuat untuk memprediksi status kedatangan pengiriman berdasarkan fitur operasional seperti berat barang (*Gross Weight*), SLA target (*Target SLA Hours*), biaya pengiriman (*Rate*), dan variabel lain yang relevan. Model ini menjadi landasan penting untuk pengembangan lebih lanjut, termasuk penerapan teknik *hyperparameter tuning* dan eksplorasi algoritma lain seperti XGBoost atau LightGBM untuk meningkatkan performa klasifikasi.

STEP 12.1: Jalankan K-Means (3 Cluster)

```
In [31]: from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans

# Salin data dan buang baris yang mengandung NaN di fitur clustering
df_cluster = df_model.dropna(subset=['Gross Weight', 'Target_SLA_Hours', 'Rate']).copy()

# Fitur numerik untuk clustering
X_cluster = df_cluster[['Gross Weight', 'Target_SLA_Hours', 'Rate']]

# Scaling
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X_cluster)

# Jalankan K-Means
kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
df_cluster['Cluster'] = kmeans.fit_predict(X_scaled)

# Lihat jumlah data per cluster
print("Jumlah data per cluster:")
print(df_cluster['Cluster'].value_counts())

Jumlah data per cluster:
Cluster
2    37689
1    34285
0     4911
Name: count, dtype: int64
```

STEP 12.3: Visualisasi Cluster (PCA)

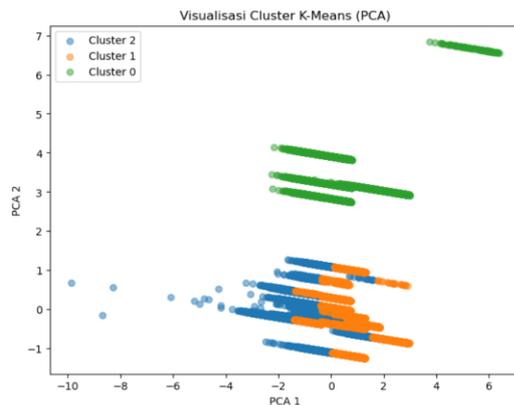
```
In [32]: from sklearn.decomposition import PCA
import matplotlib.pyplot as plt

# Reduksi dimensi ke 2D dengan PCA
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_scaled)

# Tambahkan hasil PCA ke data
df_cluster['PCA1'] = X_pca[:, 0]
df_cluster['PCA2'] = X_pca[:, 1]

# Plot cluster
plt.figure(figsize=(8,6))
for cluster in df_cluster['Cluster'].unique():
    subset = df_cluster[df_cluster['Cluster'] == cluster]
    plt.scatter(subset['PCA1'], subset['PCA2'], label=f'Cluster {cluster}', alpha=0.5)

plt.title('Visualisasi Cluster K-Means (PCA)')
plt.xlabel('PCA 1')
plt.ylabel('PCA 2')
plt.legend()
plt.show()
```



Dari grafik PCA kita bisa lihat:

Cluster 0 (hijau) → jelas terpisah di bagian kanan atas. Ini kemungkinan cluster yang unik, mungkin berat besar atau biaya tinggi.

Cluster 1 & 2 (oranye dan biru) → posisinya lebih rapat, artinya mirip-mirip secara fitur Gross Weight, SLA, dan Rate.

Artinya: K-Means berhasil membagi shipment menjadi 3 kelompok yang secara karakteristik cukup berbeda.

STEP 12.4: Analisis Karakteristik Cluster

```
In [34]: print("Rata-rata fitur per cluster:")
print(df_cluster.groupby("cluster")[["Gross Weight", "Target_SLA_Hours", "Rate"]].mean())
```

Rata-rata Fitur per cluster:

Cluster	Gross Weight	Target_SLA_Hours	Rate
0	14.824884	31.407860	47364.971696
1	7.331830	29.833980	19419.382325
2	22.282233	29.390857	19848.542758

Insight:

Cluster 0 → Paket berat menengah (14 kg), SLA agak tinggi, tapi Rate sangat tinggi → mungkin jalur khusus atau barang bernilai tinggi.

Cluster 1 → Paket ringan dengan SLA standar dan biaya rendah → mayoritas shipment regulier.

Cluster 2 → Paket berat terbesar (22 kg) dengan SLA standar dan biaya mirip Cluster 1 → mungkin shipment besar tapi rute mudah/umum.

Summary Modeling & Clustering

Model	Hasil
Random Forest (balanced)	Accuracy 84%, recall Late naik dari 40% ke 48%. Sudah cukup baik untuk baseline.
K-Means Clustering (3 cluster)	Segmentasi shipment berhasil, cluster dibedakan berdasarkan berat, SLA, dan biaya pengiriman.

Gambar 3. 15

Setelah tahap prediksi dengan Random Forest dilakukan, langkah selanjutnya dalam proses analisis data ini adalah melakukan segmentasi menggunakan algoritma **K-Means Clustering**. Segmentasi ini bertujuan untuk mengelompokkan pengiriman berdasarkan karakteristik tertentu sehingga dapat diidentifikasi pola-pola khusus yang tidak langsung terlihat pada analisis sebelumnya.

Pada tahap **Step 12.1**, proses K-Means dilakukan dengan menentukan jumlah cluster sebanyak tiga (3), berdasarkan pertimbangan eksplorasi awal dan untuk mempermudah interpretasi. Data yang digunakan untuk clustering terdiri dari fitur numerik seperti *Gross Weight*, *Target SLA Hours*, dan *Rate*. Sebelum dilakukan proses klasterisasi, data di-*scaling* menggunakan *StandardScaler* agar setiap fitur

memiliki skala yang seragam, sehingga tidak ada fitur yang mendominasi hasil clustering karena perbedaan satuan atau rentang nilai.

Hasil clustering menunjukkan distribusi jumlah data pada masing-masing cluster yang cukup berimbang. Cluster 2 memiliki jumlah data terbanyak, diikuti oleh Cluster 1 dan Cluster 0. Ini menunjukkan bahwa data pengiriman cenderung tersebar pada tiga kelompok yang relatif jelas.

Selanjutnya, pada **Step 12.3**, dilakukan visualisasi cluster menggunakan teknik *Principal Component Analysis (PCA)* untuk mereduksi dimensi data ke dalam dua dimensi yang dapat divisualisasikan dengan baik. Plot PCA memperlihatkan distribusi masing-masing cluster, di mana:

1. **Cluster 0** (hijau) tampak terpisah di bagian kanan atas. Ini mengindikasikan bahwa cluster ini memiliki karakteristik unik, kemungkinan mewakili pengiriman dengan berat sedang namun dengan SLA dan biaya yang tinggi, mungkin karena jalur khusus atau barang bernilai tinggi.
2. **Cluster 1** dan **Cluster 2** (oranye dan biru) lebih rapat satu sama lain. Ini menunjukkan bahwa pengiriman pada cluster tersebut memiliki karakteristik yang mirip dari segi *Gross Weight*, SLA, dan *Rate*, meskipun terdapat perbedaan pada rata-rata nilai fitur yang dianalisis lebih lanjut pada langkah berikutnya.

Pada **Step 12.4**, dilakukan analisis karakteristik tiap cluster dengan menghitung rata-rata fitur di masing-masing kelompok. Hasilnya:

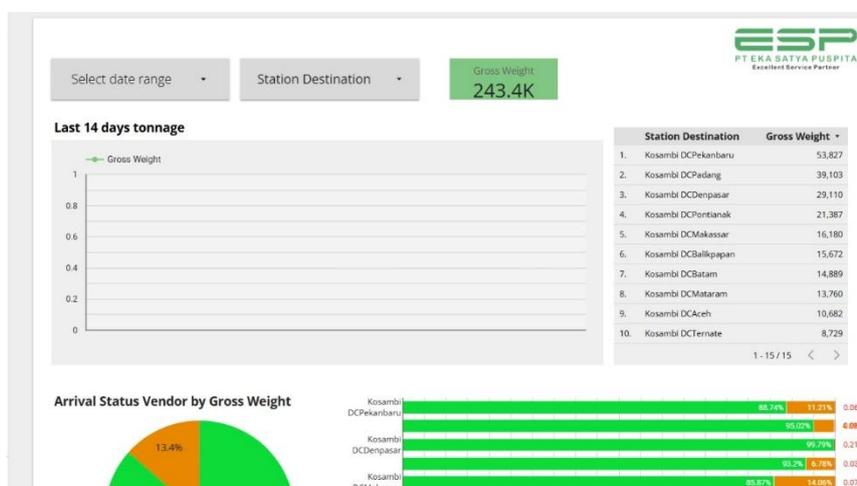
1. **Cluster 0** → Paket dengan berat menengah (14 kg), SLA tinggi (47 jam), dan biaya pengiriman tertinggi. Kemungkinan besar adalah pengiriman khusus atau barang bernilai tinggi.
2. **Cluster 1** → Paket ringan dengan SLA standar dan biaya rendah, menunjukkan mayoritas pengiriman reguler.

3. **Cluster 2** → Paket berat (22 kg) dengan SLA standar dan biaya yang sedikit lebih tinggi, kemungkinan besar mewakili pengiriman besar tetapi melalui rute yang umum atau mudah.

Terakhir, dilakukan ringkasan hasil modeling dan clustering. Model Random Forest yang dilatih sebelumnya berhasil memberikan akurasi yang cukup baik (84%), dengan recall untuk pengiriman *Late* sekitar 48%, yang sudah cukup baik untuk baseline model. Di sisi lain, proses K-Means Clustering berhasil membagi shipment menjadi tiga kelompok yang secara karakteristik berbeda satu sama lain, memberikan wawasan tambahan terkait segmentasi pengiriman yang dapat dimanfaatkan untuk pengambilan keputusan bisnis, seperti penyesuaian SLA, biaya pengiriman, atau strategi layanan pelanggan.

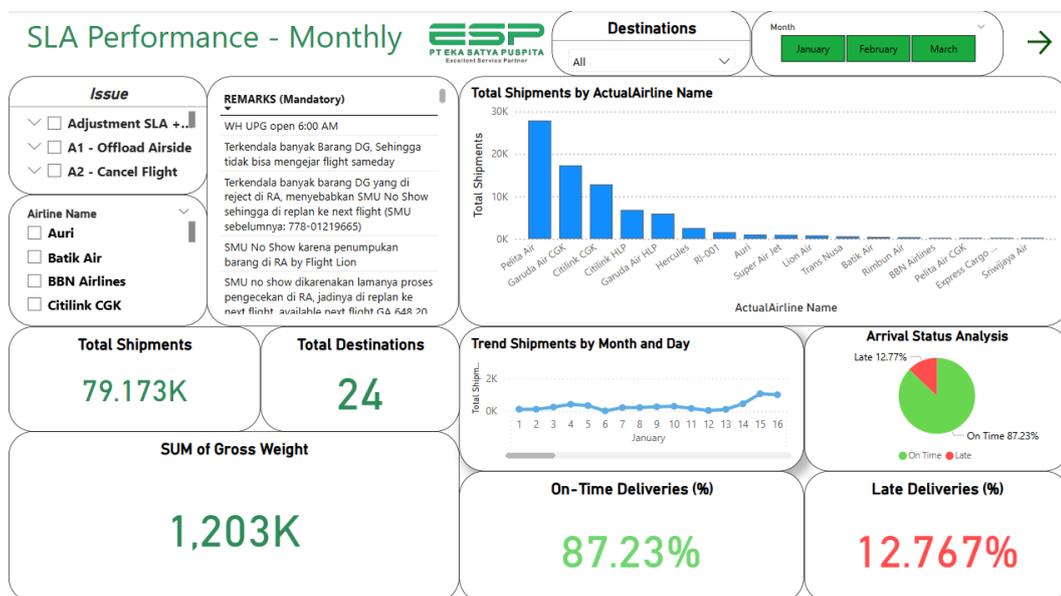
5. Visualization & Reporting

Sebagai tahap akhir dari keseluruhan proses analisis data, penulis mengembangkan serangkaian dashboard interaktif yang bertujuan untuk menyajikan hasil temuan dan mempermudah proses monitoring operasional logistik di PT Eka Satya Puspita. Dashboard ini dibangun menggunakan *Looker Studio* dan *Power BI*, dengan fokus utama pada penyajian informasi yang relevan dan mudah dipahami oleh pihak manajemen maupun divisi operasional.



Gambar 3. 16 — Dashboard awal (*Looker Studio - Kosambi DC*).

Pada tahap awal, dashboard yang dikembangkan (Gambar 3.16) memvisualisasikan data pengiriman dari Distribution Center (DC) Kosambi. Visualisasi ini mencakup total berat pengiriman (*gross weight*) selama 14 hari terakhir, distribusi pengiriman berdasarkan stasiun tujuan, serta status kedatangan pengiriman yang dibedakan antara *on-time* dan *late*. Selain itu, dashboard ini juga menyediakan filter untuk mempermudah pengguna memilih rentang tanggal dan destinasi tertentu. Walaupun dashboard awal ini sudah memberikan gambaran umum, namun cakupan analisis masih terbatas dan belum mencakup dimensi-dimensi yang lebih kompleks seperti kinerja SLA atau penyebab keterlambatan.



Gambar 3. 17 Dashboard SLA Performance — Monthly.

Seiring meningkatnya kebutuhan analisis dan kompleksitas data, penulis kemudian membangun dashboard yang lebih komprehensif menggunakan Power BI (Gambar 3.17). Dashboard ini dinamakan *SLA Performance – Monthly*, yang menyediakan informasi lebih mendalam, termasuk total pengiriman dan destinasi, distribusi pengiriman berdasarkan maskapai penerbangan, tren pengiriman harian, serta metrik utama seperti persentase pengiriman tepat waktu (*On-Time Deliveries*) dan keterlambatan (*Late Deliveries*). Selain itu, tersedia pula filter yang

memungkinkan pengguna untuk memilih destinasi dan bulan tertentu untuk analisis lebih terperinci. Dashboard ini tidak hanya membantu memahami performa pengiriman secara umum, tetapi juga menjadi alat penting untuk mengidentifikasi masalah operasional yang dapat mempengaruhi tingkat layanan (*service level agreement* atau SLA).



Gambar 3. 18 Dashboard Breakdown Delay Berdasarkan Maskapai dan Status Kedatangan.

Lebih lanjut, untuk mendukung analisis penyebab keterlambatan secara spesifik, penulis mengembangkan dashboard tambahan yang dinamakan *Breakdown Delay* (Gambar 3.18). Dashboard ini menyajikan analisis mendalam terkait keterlambatan pengiriman berdasarkan maskapai penerbangan (*Airline Analysis*) dan status kedatangan (*Arrival Status Based on Airline*). Selain itu, ditampilkan pula detail isu (*Issue*), keterangan permasalahan (*Remarks*), dan total berat pengiriman yang terdampak. Dengan adanya dashboard ini, pihak manajemen dapat dengan mudah mengidentifikasi maskapai atau jalur distribusi yang memiliki tingkat keterlambatan tinggi, serta memahami akar penyebabnya berdasarkan kategori permasalahan yang telah ditentukan.

Pengembangan dashboard yang berlapis ini mencerminkan penerapan prinsip *data-driven decision making*, di mana data tidak hanya digunakan sebagai bahan

laporan pasif, tetapi menjadi alat aktif untuk monitoring, diagnosis, dan pengambilan keputusan strategis. Visualisasi yang intuitif dan informatif membantu mempermudah komunikasi antar divisi dan mempercepat respons terhadap permasalahan yang terjadi di lapangan. Dengan demikian, inisiatif ini berhasil meningkatkan efisiensi operasional dan kualitas layanan di PT Eka Satya Puspita.



Gambar 3. 19 Shipment Cost & Efficiency Overview Dashboard

Gambar 3.19 Shipment Cost & Efficiency Overview Dashboard menampilkan analisis mendalam tentang efisiensi biaya pengiriman menggunakan data operasional dari Januari hingga Maret 2025. Dashboard ini terdiri dari empat visualisasi utama, yaitu scatter plot yang menggambarkan hubungan antara berat barang (gross weight) dengan biaya pengiriman (shipment cost) berdasarkan vendor, sehingga membantu mengidentifikasi pola efisiensi biaya terkait berat barang. Selanjutnya, terdapat diagram Pareto yang memperlihatkan biaya rata-rata per kilogram untuk masing-masing vendor dan rute pengiriman, yang berguna dalam menentukan area fokus penghematan biaya. Grafik batang juga disajikan untuk menunjukkan sepuluh rute dengan biaya pengiriman rata-rata paling mahal, yang memungkinkan perusahaan untuk mengevaluasi serta mengoptimalkan rute-rute dengan biaya operasional tinggi. Terakhir, terdapat heatmap berupa matriks efisiensi biaya rata-rata per kilogram berdasarkan vendor dan rute, dengan skala warna yang menunjukkan tingkat efisiensi biaya, di mana warna hijau mengindikasikan biaya yang efisien, sedangkan merah menunjukkan inefisiensi. Dengan demikian, dashboard ini membantu manajemen perusahaan dalam

mengevaluasi dan mengambil keputusan strategis secara cepat terkait efisiensi operasional logistik.

6. Decision-Making & Continuous Improvement

Setelah seluruh proses pengolahan data dan pengembangan dashboard selesai dilaksanakan, saya melanjutkan ke tahap akhir, yaitu *Decision-Making & Continuous Improvement*. Tahapan ini memiliki peran yang sangat penting dalam memastikan bahwa hasil kerja yang telah dikembangkan benar-benar memberikan manfaat nyata bagi perusahaan dan dapat mendukung proses pengambilan keputusan yang berbasis data (*data-driven decision making*).

Langkah pertama dalam tahap ini adalah mengajukan dan mempresentasikan dashboard yang telah dikembangkan kepada **Senior Data Analyst** dan **Supervisor IT**. Presentasi ini bertujuan untuk mendapatkan masukan terkait keakuratan data, kelengkapan metrik yang ditampilkan, serta kejelasan visualisasi. Umpan balik yang diberikan mencakup aspek teknis, seperti validitas data dan kesesuaian filter analisis, serta aspek non-teknis seperti kemudahan penggunaan (*usability*) dan kejelasan informasi bagi pengguna non-teknis di manajemen.

Berdasarkan umpan balik tersebut, saya melakukan beberapa revisi pada dashboard. Revisi yang dilakukan meliputi penyesuaian visualisasi agar lebih mudah dipahami oleh pengguna, penambahan metrik yang dianggap relevan oleh manajemen, serta pengoptimalan filter untuk memungkinkan pengguna melakukan analisis mendalam berdasarkan destinasi pengiriman, maskapai, dan periode waktu tertentu. Proses revisi ini tidak hanya bersifat memperbaiki tampilan, tetapi juga meningkatkan fungsionalitas dashboard agar mampu menjawab kebutuhan bisnis yang dinamis.

Setelah revisi diselesaikan dan dashboard mendapat persetujuan akhir dari Supervisor IT dan tim manajemen, dashboard tersebut **diimplementasikan** sebagai alat bantu monitoring operasional yang digunakan secara rutin. Dashboard ini berfungsi untuk mendukung pengambilan keputusan harian oleh tim operasional

maupun keputusan strategis yang melibatkan level manajerial. Beberapa contoh keputusan yang didukung oleh dashboard ini antara lain penyesuaian rute pengiriman, evaluasi performa maskapai, serta identifikasi dan penanganan isu operasional yang berulang.

Selain implementasi, saya juga bertanggung jawab melakukan **pemeliharaan dan pembaruan data** pada dashboard secara berkala. Proses pemeliharaan ini mencakup update data terbaru, pengawasan terhadap potensi anomali atau kesalahan data baru, serta penyesuaian visualisasi apabila terdapat perubahan kebutuhan informasi dari manajemen. Hal ini penting untuk memastikan bahwa dashboard tetap relevan dan akurat seiring berjalannya waktu dan berkembangnya kebutuhan perusahaan.

Proses *continuous improvement* tidak berhenti pada satu iterasi saja. Selama periode magang, saya secara aktif mengumpulkan umpan balik tambahan dari pengguna dashboard dan melakukan perbaikan berkelanjutan (*iterative improvement*). Pendekatan ini memungkinkan dashboard menjadi semakin efektif dalam memberikan wawasan dan mendukung keputusan yang cepat serta berbasis data.

Dengan demikian, tahapan *Decision-Making & Continuous Improvement* menjadi kunci dalam mewujudkan pengambilan keputusan yang lebih responsif, berbasis data, dan mampu beradaptasi terhadap dinamika operasional di PT Eka Satya Puspita. Pendekatan ini sekaligus mencerminkan penerapan prinsip *Business Intelligence* yang tidak hanya fokus pada pengumpulan dan analisis data, tetapi juga pada penyajian informasi yang dapat langsung diterapkan untuk meningkatkan kinerja dan efektivitas operasional perusahaan

3.2.3 Transformasi Sistem Logistik dari Perspektif Orang, Teknologi, dan Proses

Sebagai bagian dari implementasi Business Intelligence Life Cycle, pengembangan dashboard logistik di PT Eka Satya Puspita memberikan dampak nyata yang menyentuh tiga pilar utama: **orang**, **teknologi**, dan **proses**. Transformasi ini tidak hanya dapat diamati dari sisi praktik kerja di lapangan, tetapi juga selaras dengan pendekatan ilmiah dalam pengelolaan data dan sistem

pendukung keputusan. Tabel berikut menyajikan ringkasan kondisi sebelum dan sesudah pengembangan dashboard berdasarkan ketiga perspektif tersebut, serta implikasi teoritis yang menyertainya:

Table 3.1 Transformasi Sistem Logistik dari Perspektif Orang, Teknologi, dan Proses

Perspektif	Kondisi Sebelum	Kondisi Setelah	Implikasi Ilmiah
Orang	Staf operasional dan administrasi mencatat data pengiriman secara manual menggunakan Excel atau dokumen fisik. Tidak ada akses cepat ke data performa, dan komunikasi antar divisi sering mengalami miskomunikasi.	Staf kini memiliki akses dashboard interaktif yang menampilkan SLA, keterlambatan, dan performa vendor secara real-time. Komunikasi lebih terarah karena seluruh tim mengacu pada sumber data yang sama.	Terjadi peningkatan literasi data pada staf non-teknis serta penerapan sistem pendukung keputusan (DSS) berbasis visualisasi interaktif yang mendorong kolaborasi berbasis data.
Teknologi	ESP hanya menggunakan Excel sebagai alat utama pelaporan. Tidak ada sistem analitik yang bisa mendukung visualisasi, analisis SLA, atau deteksi tren secara otomatis. Data tersebar dan tidak terintegrasi.	Teknologi Power BI diintegrasikan dengan Python untuk proses cleaning dan modeling. Visualisasi data kini mencakup klasifikasi pengiriman, segmentasi shipment berdasarkan performa, dan analisis akar masalah keterlambatan.	ESP memasuki fase maturity BI dengan implementasi analisis prediktif menggunakan machine learning (Random Forest dan K-Means). Sistem ini selaras dengan praktik data science modern di industri logistik.
Proses	Proses pelaporan dilakukan secara bulanan, dan keterlambatan pengiriman hanya diketahui setelah laporan dikompilasi. Tidak ada sistem monitoring harian untuk mendeteksi masalah secara langsung.	Dashboard digunakan setiap hari oleh tim operasional untuk memantau pengiriman, SLA, dan delay. Manajemen dapat mengambil keputusan berbasis data harian tanpa menunggu laporan bulanan.	Perusahaan mulai menerapkan prinsip continuous improvement dan real-time monitoring dalam kerangka BI Life Cycle, memungkinkan evaluasi dan perbaikan strategi secara cepat dan berulang.

Transformasi sistem logistik di PT Eka Satya Puspita tidak hanya mencakup penerapan teknologi dashboard, tetapi juga menyentuh perubahan menyeluruh pada dimensi manusia, teknologi, dan proses bisnis. Dari **perspektif orang**, perubahan signifikan terlihat pada peningkatan kapasitas staf operasional dalam memahami dan menggunakan data. Jika sebelumnya mereka bekerja secara terpisah dan mengandalkan laporan manual, kini mereka dapat melihat performa secara langsung dan real-time, yang mendorong kolaborasi lebih efektif. Implikasi

akademis dari hal ini adalah meningkatnya literasi data dan pemanfaatan *decision support system* (DSS) secara langsung dalam workflow harian. Dari **perspektif teknologi**, PT ESP beralih dari alat konvensional seperti Excel menuju platform analitik modern yang menggabungkan Power BI, Python, dan machine learning. Pendekatan ini memungkinkan perusahaan untuk tidak hanya melihat apa yang terjadi, tetapi juga memahami *mengapa* keterlambatan terjadi dan *bagaimana* mengurangnya melalui analisis prediktif. Implementasi model seperti Random Forest untuk klasifikasi dan K-Means untuk segmentasi menunjukkan integrasi teori *predictive analytics* dan *data-driven strategy* dalam praktik nyata.

Sedangkan dari **perspektif proses**, alur kerja logistik yang sebelumnya bersifat reaktif kini menjadi proaktif. Dengan adanya dashboard harian, manajemen tidak perlu menunggu laporan akhir bulan untuk mengetahui kendala operasional. Perubahan ini mencerminkan prinsip *continuous monitoring* dan *real-time evaluation* dalam siklus Business Intelligence Life Cycle, yang menjadi kunci dalam menciptakan proses logistik yang adaptif dan efisien. Secara keseluruhan, transformasi ini bukan hanya perubahan alat, melainkan pergeseran paradigma kerja di PT ESP — dari berbasis intuisi dan laporan manual menjadi berbasis data dan otomatisasi visual. Ini menunjukkan keberhasilan integrasi antara teori akademik dan kebutuhan industri secara konkret.

3.3 Kendala yang Ditemukan

Dalam menjalankan tugas sebagai Data Analyst Intern di PT Eka Satya Puspita (ESP), saya menghadapi berbagai kendala dan tantangan yang mempengaruhi kelancaran proses kerja. Kendala-kendala tersebut terjadi di berbagai tahapan, mulai dari pengumpulan data, pembersihan data, analisis, hingga pengembangan visualisasi. Meskipun demikian, setiap kendala yang dihadapi menjadi pembelajaran berharga dan membantu saya untuk mengembangkan solusi yang kreatif serta meningkatkan keterampilan teknis dan manajerial.

1. Keterbatasan Kualitas Data

Salah satu kendala utama yang saya hadapi adalah kualitas data yang tidak konsisten. Data yang dikumpulkan dari berbagai sumber internal seringkali memiliki masalah seperti duplikasi, format yang tidak seragam, dan nilai yang hilang (*missing values*). Selain itu, terdapat beberapa data pencilan (*outlier*) yang ekstrem, seperti nilai SLA yang tidak masuk akal (misalnya, SLA lebih dari 1000 jam). Hal ini menyebabkan proses validasi dan pembersihan data menjadi lebih kompleks dan memakan waktu. Untuk mengatasi tantangan ini, saya melakukan pemeriksaan data secara menyeluruh menggunakan teknik data cleaning yang mencakup penghapusan duplikasi, penyesuaian format, konversi data waktu menjadi format numerik, serta penghapusan *outlier* yang tidak relevan untuk analisis.

2. Keterbatasan Dokumentasi dan Metadata

Data yang tersedia seringkali tidak dilengkapi dengan dokumentasi yang memadai atau metadata yang menjelaskan makna setiap kolom dan nilai. Hal ini menyebabkan kesulitan dalam memahami konteks data, terutama saat mencoba menginterpretasikan kolom-kolom yang memiliki nama teknis atau singkatan yang tidak familiar. Untuk mengatasi hal ini, saya melakukan diskusi rutin dengan Supervisor IT dan tim operasional untuk mendapatkan penjelasan lebih lanjut mengenai data yang digunakan dan arti dari setiap variabel.

3. Kendala Teknis dalam Pemodelan

Dalam tahap data modeling, khususnya saat membangun model klasifikasi menggunakan algoritma Random Forest, saya menghadapi tantangan terkait ketidakseimbangan kelas (class imbalance). Data pengiriman yang terlambat jauh lebih sedikit dibandingkan pengiriman yang tepat waktu, sehingga model cenderung bias terhadap kelas mayoritas (On Time). Untuk mengatasi masalah ini, saya menerapkan parameter `class_weight='balanced'` pada algoritma Random Forest agar model memberikan perhatian lebih besar pada kelas minoritas (Late). Selain itu, saya juga melakukan eksperimen dengan metode segmentasi (clustering) menggunakan K-Means untuk memahami karakteristik pengiriman yang berbeda.

4. Keterbatasan Waktu dan Sumber Daya

Sebagai mahasiswa magang dengan durasi waktu yang terbatas, saya harus mengelola waktu secara efektif agar semua tahapan pengerjaan — mulai dari pengumpulan data hingga pembuatan dashboard — dapat diselesaikan sesuai jadwal. Proses iterative improvement pada dashboard yang memerlukan revisi berdasarkan umpan balik dari pengguna juga menjadi tantangan tersendiri, terutama ketika ada permintaan perubahan visualisasi atau penambahan fitur yang kompleks.

5. Integrasi Data ke Dashboard

Dalam proses pembuatan dashboard menggunakan Power BI, saya menemui beberapa kendala terkait penggabungan data dari berbagai sumber serta pengaturan filter dan hubungan antar tabel. Beberapa kali terjadi kesalahan dalam data relationship yang menyebabkan visualisasi tidak tampil sesuai harapan. Untuk mengatasi masalah ini, saya melakukan pengaturan ulang data model dan memanfaatkan fitur data transformation di Power BI untuk memastikan data terintegrasi dengan baik sebelum divisualisasikan.

6. Perubahan Kebutuhan Informasi

Kebutuhan informasi dari pihak manajemen dan tim operasional bersifat dinamis. Pada beberapa kesempatan, metrik atau indikator yang awalnya tidak

diminta kemudian menjadi prioritas baru. Perubahan kebutuhan ini mengharuskan saya untuk menyesuaikan model analisis dan dashboard secara fleksibel agar tetap relevan dengan tujuan bisnis yang berkembang.

7. Kendala dalam Data Biaya Pengiriman

Selain data performa SLA, saya juga menghadapi tantangan dalam pengolahan data biaya pengiriman. Data dari berbagai vendor dan rute sering kali tidak dicatat secara seragam, baik dari segi satuan berat maupun format nilai biaya. Beberapa entri memiliki nilai ekstrem atau kosong, yang memengaruhi akurasi penghitungan efisiensi biaya per kilogram. Untuk mengatasi hal ini, saya melakukan validasi dan pembersihan tambahan, termasuk konversi satuan ke kilogram dan filterisasi data tidak valid, sebelum digunakan dalam visualisasi dan analisis biaya yang ditampilkan pada dashboard efisiensi (Gambar 3.19).

Sebanyak apapun tantangan yang dihadapi selama pelaksanaan magang, saya berupaya untuk menghadapinya secara sistematis dan kolaboratif. Diskusi rutin dengan supervisor, eksplorasi mandiri terhadap solusi teknis, serta pengalaman dalam menangani dinamika kebutuhan bisnis telah membentuk pemahaman saya terhadap kompleksitas dunia kerja nyata. Pengalaman ini juga memperkuat kemampuan analitis, adaptif, dan problem-solving saya sebagai calon profesional di bidang data.

3.4 Solusi atas Kendala yang Ditemukan

Untuk mengatasi berbagai kendala yang dihadapi selama pelaksanaan magang di PT Eka Satya Puspita (ESP), saya menerapkan sejumlah solusi yang disesuaikan dengan permasalahan yang muncul di setiap tahapan proses pengolahan data dan pengembangan dashboard. Solusi-solusi ini tidak hanya membantu mengatasi hambatan teknis dan non-teknis, tetapi juga mendukung kelancaran pembuatan dashboard yang menjadi produk akhir dari proyek analitik yang saya kerjakan.

1. Pendalaman Proses Bisnis dan Kebutuhan Informasi

Kendala dalam memahami proses bisnis dan kebutuhan manajemen diselesaikan dengan melakukan diskusi rutin bersama Supervisor IT dan tim operasional. Diskusi ini membantu saya mengidentifikasi Key Performance Indicators (KPI) yang relevan serta menentukan metrik visualisasi yang dibutuhkan oleh manajemen. Pendekatan komunikasi aktif ini memastikan bahwa dashboard yang dikembangkan benar-benar sesuai dengan kebutuhan pengguna (user requirements).

2. Validasi dan Integrasi Data yang Akurat

Untuk mengatasi permasalahan data yang tidak konsisten dan tidak terstruktur, saya melakukan proses validasi data dan pembersihan (data cleaning) secara menyeluruh. Data yang tidak relevan, duplikat, atau mengandung nilai ekstrem (outlier) diidentifikasi dan disaring agar tidak mempengaruhi hasil analisis. Selain itu, saya membangun model relasi data yang solid di Power BI agar integrasi antar tabel menjadi lebih efektif dan memudahkan proses visualisasi data.

3. Desain Dashboard Modular dan Fleksibel

Menghadapi permintaan revisi atau perubahan analisis dari manajemen yang cukup dinamis, saya menerapkan desain dashboard yang modular. Dengan cara ini, perubahan atau penambahan fitur visualisasi dapat dilakukan dengan mudah tanpa perlu membangun ulang seluruh dashboard. Fleksibilitas desain ini memungkinkan dashboard beradaptasi sesuai kebutuhan analisis yang berkembang.

4. Optimalisasi Pengelolaan Waktu

Keterbatasan waktu pengerjaan diatasi dengan menerapkan prinsip iterative improvement. Saya memulai pengembangan dengan membuat dashboard sederhana (Looker Studio) yang kemudian dikembangkan menjadi dashboard yang lebih kompleks dan komprehensif di Power BI. Pendekatan bertahap ini membantu mengelola waktu dengan baik sekaligus memungkinkan penyempurnaan dashboard seiring proses evaluasi berjalan.

5. Pengembangan Kompetensi Teknis

Menghadapi keterbatasan kemampuan teknis di awal magang, saya secara aktif mengikuti pelatihan mandiri dan memanfaatkan berbagai sumber pembelajaran daring untuk meningkatkan keterampilan di bidang analisis data, pembersihan data menggunakan Python, serta pembuatan dashboard menggunakan Power BI. Peningkatan kompetensi ini berkontribusi langsung pada keberhasilan dalam membangun dashboard analitik yang sesuai standar profesional.

6. Peningkatan Kolaborasi dan Komunikasi Tim

Untuk menghindari miskomunikasi yang dapat menyebabkan kesalahan data atau keterlambatan revisi dashboard, saya menggunakan platform kolaborasi seperti Google Sheets dan Microsoft Teams. Semua perkembangan data, hasil analisis, serta feedback dari supervisor dan tim didokumentasikan dengan baik sehingga proses kerja menjadi lebih transparan dan terkoordinasi.

7. Validasi dan Transformasi Data Biaya Pengiriman

Untuk mengatasi kendala dalam data biaya pengiriman, saya melakukan proses validasi tambahan yang mencakup penghapusan entri tidak valid, normalisasi satuan berat ke dalam kilogram, serta pengecekan kelengkapan data pada setiap rute dan vendor. Langkah ini bertujuan agar data biaya dapat dianalisis secara akurat dan dapat mendukung visualisasi efisiensi biaya logistik. Hasil dari proses ini kemudian digunakan dalam pengembangan dashboard Shipment Cost & Efficiency Overview (Gambar 3.19), yang menjadi pelengkap dari analisis performa SLA.

Solusi-solusi yang diterapkan tersebut berkontribusi secara langsung terhadap keberhasilan pengembangan dashboard analitik yang tidak hanya memenuhi kebutuhan informasi manajemen tetapi juga memberikan landasan kuat bagi pengambilan keputusan berbasis data (data-driven decision making) di PT Eka Satya Puspita.