

Peningkatan Kualitas Data dalam Konsolidasi Data Karyawan melalui Data Wrangling

Joni*¹, Teguh Prasandy²

Bina Nusantara University

e-mail: ¹ joni@binus.ac.id, ² teguh.prasandy@binus.edu

*Penulis Korespondensi

Diterima: 01 November 2023; Direvisi: 12 Juli 2024; Disetujui: 15 November 2024

Abstrak

Data wrangling merupakan proses penting dalam persiapan data yang dapat secara signifikan meningkatkan kualitas data. Teknik data wrangling yang efektif, yang terdiri dari 6 langkah yaitu Discovery (Penemuan), Structuring (Penataan), Cleaning (Pembersihan), Enriching (Pengayaan), Validating (Validasi), dan Publishing (Publikasi). Data wrangling dapat membantu Divisi Sumber Daya Manusia untuk memastikan bahwa data mereka berkualitas tinggi dan siap untuk dianalisis. Dalam studi kasus ini, peneliti mengeksplorasi bagaimana teknik data wrangling yang efektif dapat digunakan untuk meningkatkan kualitas data dalam konsolidasi data karyawan. Hasil penelitian menemukan bahwa data karyawan yang diunduh dari berbagai sumber, tercatat tidak lengkap, tidak andal, atau salah sehingga dapat memengaruhi analisis data. Data wrangling berupaya menghilangkan risiko tersebut dengan memastikan data berada dalam keadaan andal sebelum dianalisis dan dimanfaatkan. Kumpulan data dari berbagai sumber data karyawan dianalisis dan menunjukkan bagaimana teknik data wrangling dapat digunakan untuk membersihkan dan mengubah data guna meningkatkan kualitas data dan siap untuk dianalisis. Hasil penelitian memberikan bukti empiris tentang dampak data wrangling terhadap kualitas data dan menyoroti pentingnya proses ini dalam konsolidasi data karyawan dan menyediakan analitik tenaga kerja.

Kata kunci: Data wrangling, Kualitas data, Persiapan data, Data pegawai

Abstract

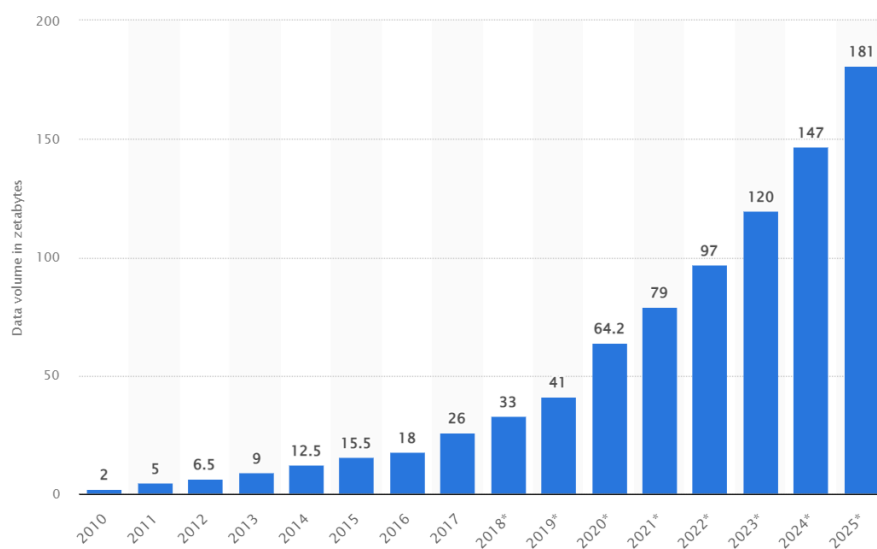
Data wrangling is a critical process in data preparation that can significantly improve data quality. Effective data wrangling techniques, which consists of 6 steps i.e. Discovery, Structuring, Cleaning, Enriching, Validating, Publishing, can help Corporate Human Resource Division to ensure that their data is of high quality and ready for analysis. In this case study, we explore how effective data wrangling techniques can be used to improve data quality in employee data consolidation. We found employee data downloaded from various sources, captured incomplete, unreliable, or incorrect so that it could affect data analysis. Data wrangling seeks to remove that risk by ensuring data is in a reliable state before it's analyzed and leveraged. We analyze a dataset from multiple sources of employee data and demonstrate how data wrangling techniques can be used to clean and transform the data to improve data quality and ready for analysis. Our study provides empirical evidence of the impact of data wrangling on data quality and highlights the importance of this process in employee data consolidation and provide workforce analytics.

Keywords: Data wrangling, Data quality, Data preparation, Employee data

1. PENDAHULUAN

Di era digital saat ini, jumlah data yang dihasilkan semakin meningkat dan bervariasi. Namun, sering kali data yang masih mentah ini mengandung banyak kesalahan, duplikasi, informasi yang tidak lengkap, dan tidak konsisten. Kesalahan data dapat terjadi dalam berbagai cara, seperti kesalahan input data, kesalahan pemrosesan data, dan kesalahan pada sistem basis data. Data mentah yang diperoleh dari berbagai sumber juga memiliki beragam format. Karena itu kebutuhan untuk memperbaiki dan mempersiapkan data sebelum digunakan dalam analisis juga semakin meningkat.

Seiring dengan peningkatan jumlah data, kebutuhan untuk memastikan kualitas data menjadi sangat penting, karena data adalah salah satu aset yang berharga dalam bisnis. Data digunakan untuk membuat keputusan yang tepat, mengembangkan strategi bisnis, dan mengidentifikasi tren pasar yang sedang berkembang. Kualitas data yang buruk dapat menyebabkan masalah dalam pengambilan keputusan dan analisis data. Tahun 2010 saat CEO Google Eric Schmidt berbicara pada konferensi Techonomy di Lake Tahoe California menyatakan bahwa “ada lima exabyte data yang diciptakan oleh seluruh dunia antara awal peradaban hingga 2003. Tetapi saat ini hanya dalam dua hari tercipta lima exabyte dan jumlah data ini akan terus meningkat dengan cepat” [7]. 1 exabyte (EB) = 10^{18} bytes = 1000 petabytes = 1 juta terabytes = 1 miliar gigabytes. Menurut statista.com, jumlah total data yang dibuat, disalin, dan dikonsumsi secara global juga menunjukkan peningkatan yang pesat dari tahun ke tahun, seperti terlihat pada gambar 1.



Gambar 1. Pertumbuhan volume data global (2010-2025) [9]

Selama pandemi Covid-19, membuat orang menghabiskan lebih banyak waktu di internet untuk bekerja, belajar, dan hiburan. Hal ini juga mencerminkan semakin banyak data yang dibuat setiap hari. Per April 2021, jumlah orang di Internet telah tumbuh sebesar 7,6%, berarti 60% populasi dunia sekarang online [1]. Pertumbuhan data yang sangat pesat ini hanya akan menjadi "data sampah" apabila tidak dapat memberikan hasil yang informatif. Oleh karena itu, diperlukan teknik data wrangling untuk mempersiapkan data agar siap digunakan dalam analisis data dan pembuatan dashboard yang akurat.

Tujuan penulisan ini adalah untuk mendemonstrasikan teknik data wrangling agar dapat meningkatkan kualitas data dan memberikan hasil yang lebih akurat dalam analisis data sebelum membuat sebuah dashboard. Untuk membandingkan penelitian sebelumnya yang menggunakan alat Trifacta®Wrangler [2], penelitian yang dilakukan ini menggunakan data primer yang diunduh dari beberapa sistem kepegawaian di perusahaan YXX.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan studi literatur melalui penelusuran terhadap jurnal, buku teks, dan internet. Studi literatur ini akan dibahas mengenai definisi data wrangling, perbedaan data cleaning dan data wrangling, teknik data wrangling, dan alat yang digunakan dalam proses data wrangling. Selain itu, juga akan dipaparkan studi kasus untuk mendemonstrasikan bagaimana teknik data wrangling diterapkan pada konsolidasi data karyawan di perusahaan XYZ dan bagaimana hasilnya digunakan untuk dashboard analisis.

Dikutip dari Harvard Business School Online (HBS.edu), data wrangling mengacu pada berbagai proses yang dirancang untuk mengubah data mentah menjadi format yang lebih mudah digunakan [8]. Metode yang tepat berbeda dari proyek ke proyek tergantung pada data yang akan dimanfaatkan dan tujuan yang ingin dicapai. Beberapa contoh proses data wrangling meliputi (1) menggabungkan beberapa sumber data menjadi satu set data untuk analisis, (2) mengidentifikasi kesenjangan dalam data (misalnya, sel kosong dalam spreadsheet) dan mengisi atau menghapusnya, (3) menghapus data yang tidak diperlukan atau tidak relevan dengan proyek yang sedang dikerjakan, dan (4) mengidentifikasi outlier ekstrim dalam data dan menjelaskan ketidaksesuaian atau menghilangkannya sehingga analisis dapat dilakukan.

Tujuan utama dari data cleaning (juga disebut sebagai data cleansing) adalah untuk mengidentifikasi dan menghapus ketidakkonsistenan, tanpa menghapus data yang diperlukan untuk menghasilkan wawasan [12]. Data wrangling dan data cleaning adalah dua proses yang berbeda. Penting untuk membedakan bahwa data cleaning merupakan langkah penting dalam proses data wrangling untuk menghapus data yang tidak akurat dan tidak konsisten. Sementara itu, data wrangling adalah keseluruhan proses mengubah data mentah menjadi bentuk yang lebih bermanfaat. Jadi data wrangling adalah proses mempersiapkan data mentah agar siap digunakan dalam analisis data. Kegiatan data wrangling berupaya memastikan data mentah tersebut berada dalam kondisi yang andal sebelum dianalisis dan dimanfaatkan. Ini menjadikannya bagian penting dari proses analisis, karena tujuan dari data wrangling adalah untuk menghasilkan data yang berkualitas dan akurat yang dapat digunakan dalam analisis data.

Data wrangling dapat dilakukan dengan proses manual atau otomatis. Dalam skenario di mana kumpulan data sangat besar, pembersihan data secara otomatis menjadi suatu keharusan. Dalam organisasi yang mempunyai tim data lengkap, seorang *data scientist* biasanya bertanggung jawab atas data wrangling, sedangkan di organisasi yang lebih kecil, spesialis non-data dapat diberikan tanggung jawab untuk membersihkan data sebelum memanfaatkannya. Data wrangling melibatkan serangkaian tindakan untuk membersihkan, memformat, dan mengubah data menjadi bentuk yang lebih mudah dianalisis. Teknik data wrangling meliputi penghapusan duplikasi, pemisahan dan penggabungan kolom, pengubahan format data, penghapusan nilai kosong atau data yang tidak relevan, dan penggabungan beberapa data dalam satu dataset. Tahapan data wrangling terdapat dalam gambar 2.



Gambar 2. Tahapan data wrangling [14]

Dikutip dari Harvard Business School Online [8] dan monkeylearn.com [14], terdapat enam langkah yang diperlukan untuk melakukan data wrangling.

Langkah 1: Penemuan (*Discovery*).

Langkah Penemuan (*Discovery*) mengacu pada proses membiasakan diri dengan data sehingga dapat mengkonseptualisasikan bagaimana menggunakan data tersebut. Dalam tahap penemuan ini dapat diidentifikasi tren atau pola dalam data, bersama dengan masalah yang jelas, seperti nilai yang hilang atau tidak lengkap yang perlu ditangani. Ini merupakan langkah penting, karena akan menginformasikan setiap aktivitas yang datang setelahnya.

Langkah 2: Penataan (*Structuring*).

Data mentah biasanya tidak dapat digunakan dalam keadaan apa adanya karena tidak lengkap atau salah format untuk aplikasi yang tertentu. Tahap penataan (*Structuring*) adalah proses dalam mengambil data mentah dan mengubahnya agar lebih mudah untuk dimanfaatkan. Bentuk data akan bergantung pada model analitik yang akan digunakan untuk menafsirkannya.

Langkah 3: Pembersihan (*Cleaning*).

Langkah pembersihan (*Cleaning*) adalah proses dalam menghilangkan kesalahan yang melekat pada data yang mungkin akan mendistorsi analisis data atau membuatnya kurang bernilai. Pembersihan dapat dilakukan dalam berbagai bentuk, termasuk menghapus sel atau baris kosong, menghapus outlier, dan menstandarisasi input. Tujuan data cleaning adalah untuk memastikan tidak ada kesalahan (atau sesedikit mungkin) yang dapat memengaruhi analisis akhir.

Langkah 4: Pengayaan (*Enriching*).

Setelah memahami data yang ada dan telah mengubahnya menjadi lebih berguna, maka harus menentukan apakah data yang sudah dimiliki tersebut diperlukan untuk proyek yang ada. Jika tidak, maka dapat memilih untuk memperkaya atau menambah data dengan memasukkan nilai dari kumpulan data lain. Untuk alasan ini, penting untuk memahami data lain yang tersedia untuk digunakan. Jika disepakati bahwa tahap pengayaan (*enrichment*) ini diperlukan, maka perlu mengulangi langkah-langkah di atas untuk setiap data yang baru.

Langkah 5: Pemvalidasian (*Validating*).

Validasi data mengacu pada proses verifikasi bahwa data tersebut konsisten dan memiliki kualitas yang cukup tinggi. Selama proses validasi, bisa saja ditemukan masalah yang perlu diselesaikan atau disimpulkan bahwa data tersebut sudah siap untuk dianalisis. Validasi biasanya dicapai melalui berbagai proses otomatisasi dan membutuhkan pemrograman.

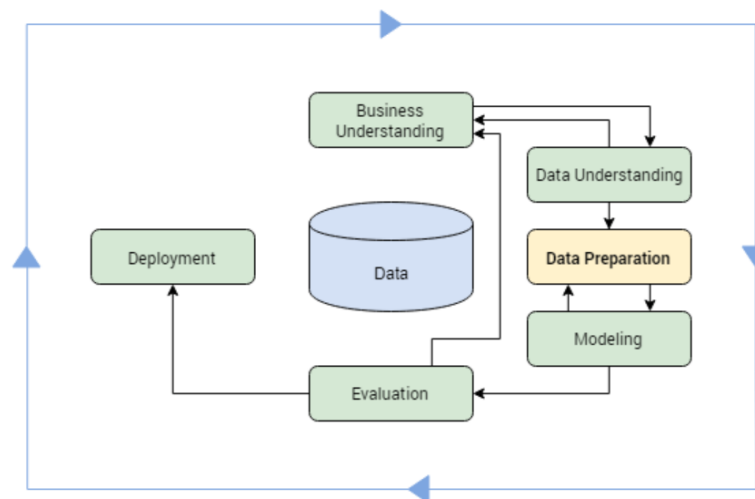
Langkah 6: Penerbitan (*Publishing*).

Setelah data divalidasi, maka dapat mempublikasikannya. Ini membuat data tersebut tersedia untuk orang lain dalam organisasi tersebut untuk dianalisis. Format yang digunakan untuk berbagi informasi akan bergantung pada data dan tujuan organisasi.

Dalam melakukan serangkaian langkah-langkah di atas, terdapat banyak alat yang dapat digunakan untuk membantu proses data wrangling, antara lain Excel Power Query / Spreadsheets, CSVKit, Google Cloud Dataprep, dan OpenRefine. Selain itu, juga dapat menggunakan Python atau R, apabila ada kebutuhan untuk melakukan data wrangling yang lebih kompleks. Setiap alat memiliki kelebihan dan kekurangan dalam melakukan data wrangling.

3. HASIL DAN PEMBAHASAN

Setiap analisis yang dilakukan oleh perusahaan, sangat tergantung pada data yang melengkapi informasinya. Jika data tidak lengkap, tidak bisa diandalkan, atau salah, maka besar kemungkinan hasil analisis pun akan salah. Di sinilah peran data wrangling sangat dibutuhkan. Data wrangling menghilangkan risiko itu dengan memastikan data dalam kondisi baik dan bisa dipercaya sebelum dianalisis dan dimanfaatkan. Kegiatan ini juga mengumpulkan data dari berbagai sumber ke dalam satu lokasi agar data tersebut bisa dipakai. Selain itu, data wrangling juga penting untuk dilakukan agar data mentah sudah tersusun dengan baik sehingga kesalahan dalam pengolahan data pun dapat diminimalisir. Proses data wrangling digambarkan sebagai "mencerna" data (sering disebut sebagai "munging") [3]. Untuk mencerna data ini diperlukan tahap persiapan untuk memahami bisnis dan setiap operasi bisnis yang terkait dengan data [13].



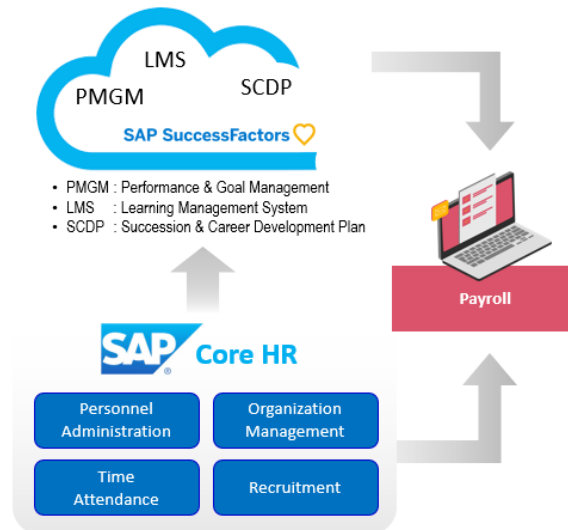
Gambar 3. Data understanding [13]

Data wrangling tidak hanya tentang mengubah data dan prosedur membersihkan data. Banyak aspek lain yang harus dipertimbangkan, antara lain: kualitas data, penggabungan sumber yang berbeda, dan pengelolaan sumber data. Dengan demikian, dapat dikatakan bahwa persiapan data, visualisasi data, validasi, standarisasi, pengayaan data, dan integrasi data telah tercakup semuanya dalam proses data wrangling [6]. Meskipun ada teknik dan platform Extract, Transform, dan Load (ETL) yang digunakan untuk melakukan ekstraksi dan pembersihan data, tetap dibutuhkan proses data wrangling ketika dihadapkan dengan volume data yang besar[4]. Data Wrangling merupakan tahap kritis dalam analisis data yang intinya melibatkan pembersihan data, penggabungan data dari sumber yang berbeda, pemilihan atribut yang relevan, dan transformasi data untuk memperoleh struktur yang sesuai. Beberapa tantangan yang sering dihadapi dalam Data Wrangling, seperti data yang tidak lengkap, format yang tidak konsisten, dan nilai yang hilang, menjadi bagian dari proses Data Wrangling yang harus dirapikan untuk mencegah bias, mengurangi kesalahan, dan meningkatkan kualitas data yang digunakan dalam analisis [5].

Dalam studi kasus ini, peneliti mengeksplorasi seberapa efektif teknik data wrangling dapat digunakan untuk meningkatkan kualitas data dari konsolidasi data karyawan. Peneliti menemukan data karyawan yang diunduh dari berbagai sumber, terdapat data-data yang belum lengkap ataupun salah, sehingga dapat memengaruhi analisis data. Karena itu, kumpulan data karyawan yang diperoleh dari berbagai sumber harus dirapikan kembali. Pada kasus ini, Corporate Human Resource Division dari perusahaan XYZ akan melakukan konsolidasi data karyawan dari beberapa grup perusahaan (*entity group*), yang data-datanya diperoleh dari tiga sumber seperti yang terdapat dalam gambar 4, yaitu:

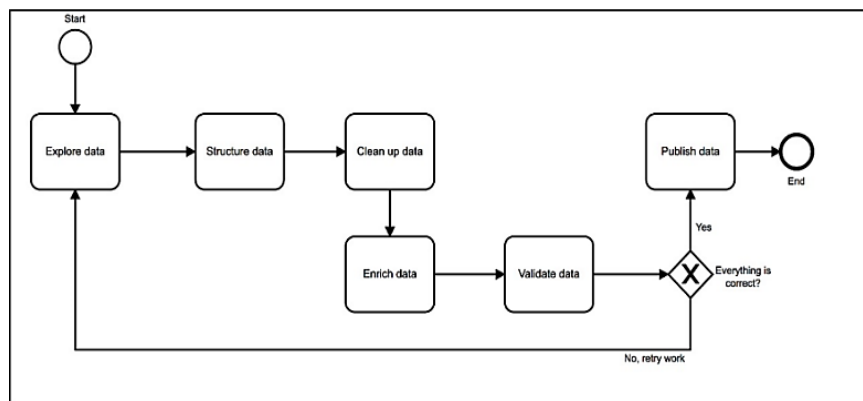
1. ERP SAP yang merupakan aplikasi utama penyimpanan master data karyawan. Data-data karyawan disimpan dalam bentuk infotype di modul SAP HR. Infotype ini merupakan unit informasi yang digunakan untuk menyimpan master data karyawan dimana infotype ini memiliki 4 digit kode dan nama yang terkait. Misalnya kode infotype 0001 adalah infotype “Organizational Assignment” yang menyimpan data organisasi karyawan, seperti entity group, personnel number, employee name, position. Kode infotype 0002 adalah infotype “Personal Data” yang menyimpan data pribadi karyawan, seperti personnel number, employee name, birthdate, religion, marital status, dan nationality.
2. SuccessFactors adalah perangkat lunak manajemen talenta berbasis cloud (software-as-a-service) yang merupakan perluasan dari modul inti SAP HR, untuk mengelola berbagai aspek Human Capital Management, termasuk kinerja karyawan, HR analitik, dan aktivitas pembelajaran [15].
3. Payroll System merupakan sistem administrasi penggajian karyawan yang dibuat terpisah dari sistem inti yang ada, dengan pertimbangan untuk menjaga kerahasiaan data

penggajian.



Gambar 4. Arsitektur informasi sumber data karyawan

Dalam studi kasus ini, akan diimplementasikan enam langkah teknik data wrangling untuk menghasilkan kualitas data karyawan yang diunduh dari tiga sumber sistem informasi di perusahaan XYZ. Langkah ini terdapat dalam gambar 5.



Gambar 5. Proses data wrangling [2]

Data yang diperoleh dari ketiga sumber tersebut disimpan menjadi *spreadsheet file* untuk diolah lebih lanjut. Sejumlah besar sumber data yang tersedia untuk memfasilitasi analisis kombinasi informasi heterogen, tidak dapat dicapai hanya melalui data "silo" yang diperoleh dari gudang data [10]. Karena itu, ketiga spreadsheet file ini kemudian akan dikonsolidasikan menjadi satu kesatuan data karyawan yang lengkap yang dapat digunakan untuk menganalisis data karyawan.

3.1. Penemuan (Discovery)

Setelah mengunduh data karyawan dari tiga sumber menjadi spreadsheet file, maka file yang berisi data mentah ini harus ditelaah terlebih dahulu agar mengerti informasi apa saja yang terkandung dalam data ini. Ketiga spreadsheet file yang berisi data-data mentah ini mencakup:

1. Spreadsheet file pertama dapat dilihat dalam gambar 6, berisi data mengenai Entity Group, Personnel Number, Employee Name, Position Name, Birthdate, Nationality, dan Employee status.

Entity Group	Personnel Number	Name	Position Name	Birthdate	Nationality	Employee Status
Head Office	1012424	Sutopo	Coordination Technician	31.01.1970	Indonesian	Local
BO Tangerang	1004245	Adji Cahyono	Operator, Finishing	02.12.1976	Indonesian	Local
BO Surabaya	1014251	Budianto	Technician, Electrical Maintenance	10.10.1976	Indonesian	Local
BO Surabaya	1024827	Steven Chow	Head, Finishing	101219.67	Chinese	Expatriate
BO Tangerang	1095945	Christian Stephen	Technician, QC	01.06.1975	Indonesian	Local
Head Office	1084752	Heryanti	Officer, Optimization Facility	09.09.1971	Indonesian	Local
Head Office	1084344	Huang Xiaoming	Head, Business Optimization	08.11.1968	Taiwanese	Expatriate
BO Tangerang	1420495	Alina Winata	Manager, Production	23.07.1965	Indonesian	Local
Head Office	1001596	Andika	Specialist, Industrial License	02.11.1967	Indonesian	Local
BO Surabaya	1042748	Lee Min Ho	Head, Cogen	22.01.1970	Taiwanese	Expatriate
BO Surabaya	1094904	Steve Widiono	Manager, IT	101019.8	Indonesian	Local
Head Office	1055207	John Black	Head, Market Research & Development	01.02.1951	American	Expatriate
Head Office	1001434	Jennifer Magdalena	Head, Log Opt (Sea Transportation)	29.01.1973	Indonesian	Local
Head Office	1070089	Wang Chang Su	Head, Product Research & Development	08.12.1959	Chinese	Expatriate
Head Office	1070089	Wang Chang Su	Head, Product Research & Development	08.12.1959	Chinese	Expatriate
BO Tangerang	1005295	Angela Veronica	Sales Executive	07.03.1987	Indonesian	Local
BO Surabaya	1985001	Bembi Eko	Officer, Accounting	02.01.1976	Indonesian	Local
BO Surabaya	1124952	Jacky Chan	Manager, QC	21119.81	Taiwanese	Expatriate
BO Surabaya	1412577	Rakesh Kumar	Expert, Material Supply	14.07.1984	Indian	Expatriate
Head Office	1002177	Ratna Sari	Specialist, Industrial License	21.04.1970	Indonesian	Local
BO Tangerang	1100499	Slamet Dodi Irawan	Operator, Field Boiler - Carton Box	15.05.1970	Indonesian	Local
BO Tangerang	1100499	Slamet Dodi Irawan	Operator, Field Boiler - Carton Box	15.05.1970	Indonesian	Local

Gambar 6. Spreadsheet file untuk Employee Data

2. Spreadsheet file kedua dapat dilihat dalam gambar 7, berisi data mengenai Entity Group, Personnel Number, Employee Name, Marital Status, dan Number of Children.

Entity Group	BO Tangerang	Head Office	BO Surabaya	BO Surabaya	BO Surabaya	Head Office
Personnel Number	1420495	1001596	1042748	1094904	1412577	1002177
Name	Alina Winata	Andika	Lee Min Ho	Steve Widiono	Rakesh Kumar	Ratna Sari
Marital Status	Married	Married	Married	Married	Single	Married
Number of Children	3	0	1	0	1	1

Entity Group	Head Office	BO Tangerang	BO Surabaya	BO Surabaya	BO Tangerang	Head Office
Personnel Number	1012424	1004245	1014251	1024827	1095945	1084752
Name	Sutopo	Adji Cahyono	Budianto	Steven Chow	Christian Stephen	Heryanti
Marital Status	Married	Married	Single	Married	Married	Married
Number of Children	3	1	3	3	1	1

Entity Group	Head Office	Head Office	Head Office	Head Office	BO Tangerang
Personnel Number	1055207	1001434	1070089	1070089	1005295
Name	John Black	Jennifer Magdalena	Wang Chang Su	Wang Chang Su	Angela Veronica
Marital Status	Married	Married	Married	Married	Single
Number of Children	1	2	2	2	2

Entity Group	BO Tangerang	BO Tangerang	Head Office	BO Surabaya	BO Surabaya
Personnel Number	1100499	1100499	1084344	1124952	1985001
Name	Slamet Dodi Irawan	Slamet Dodi Irawan	Huang Xiaoming	Jacky Chan	Bembi Eko
Marital Status	Married	Married	Married	Married	Married
Number of Children	2	2	2	2	0

Gambar 7. Spreadsheet file untuk Employee Family Data

3. Spreadsheet file ketiga dapat dilihat dalam gambar 8, berisi data mengenai Entity Group, Personnel Number, Employee Name, Education Level, dan Education Institution Name.

Entity Group	Head Office	BO Tangerang	BO Surabaya	BO Surabaya	BO Tangerang	Head Office
Personnel Number	1012424	1004245	1014251	1024827	1095945	1084752
Name	Sutopo	Adji Cahyono	Budianto	Steven Chow	Christian Stephen	Heryanti
Education Level	Senior High School	Senior High School	Senior High School	University/S2	Junior High School	Senior High School
Edu Institution Name	STM Pancasila 1		SMAN 01	Nanjing Tech University	Pemda	
Entity Group	Head Office	BO Tangerang	Head Office	BO Surabaya	BO Surabaya	Head Office
Personnel Number	1084344	1420495	1001596	1042748	1094904	1055207
Name	Huang Xiaoming	Alina Winata	Andika	Lee Min Ho	Steve Widiono	John Black
Education Level	University/S2	University/S1	Senior High School	University/S1	University/S2	University/S2
Edu Institution Name	Chinese Culture University	Universitas Bina Nusantara	SGPN IKIP	National Taipei University of Technology	Universitas Indonesia	
Entity Group	BO Surabaya	Head Office	BO Tangerang	BO Tangerang	Head Office	
Personnel Number	1412577	1002177	1100499	1100499	1001434	
Name	Rakesh Kumar	Ratna Sari	Slamet Dodi Irawan	Slamet Dodi Irawan	Jennifer Magdalena	
Education Level	University/S1	Senior High School	Senior High School	Senior High School	University/S1	
Edu Institution Name		SGON Jakarta	SMAN 78	SMAN 78	S1/Untar Jakarta	
Entity Group	Head Office	Head Office	BO Tangerang	BO Surabaya	BO Surabaya	
Personnel Number	1070089	1070089	1005295	1985001	1124952	
Name	Wang Chang Su	Wang Chang Su	Angela Veronica	Bembi Eko	Jacky Chan	
Education Level	University/S2	University/S2	University/S1	Senior High School	University/S2	
Edu Institution Name	Harvard University	Harvard University	Universitas Trisakti	SMAN 02	Chinese Culture University	

Gambar 8. Spreadsheet file untuk Employee Education Data

Pada tahap penemuan (Discovery) ini, data-data mentah ini masih belum terstruktur dengan baik. Tetapi semua spreadsheet file ini harus dipahami, dicermati, dan dilihat keterkaitannya, agar data tersebut dapat digabungkan menjadi satu tabel data karyawan yang lengkap dan terstruktur pada tahap penataan. Pemahaman dan pengaturan data agar dapat digunakan untuk analitik membutuhkan waktu yang lama [11]. Jika seseorang memiliki akun bisnis dan akun pribadi, apakah mereka dua pelanggan atau satu. Atau mungkin ada ID Pelanggan di tabel database, tetapi jika pelanggan pergi dan kembali, mereka mungkin mendapatkan ID pelanggan baru.

3.2. Penataan (Structuring)

Sebelumnya pada tahap penemuan, data karyawan yang diperoleh dari beberapa sumber, masih terdapat data yang berbentuk form dan ada yang dalam format tabular, sehingga data-data ini belum terstruktur dengan baik. Karena itu, dalam tahap penataan inilah dilakukan perbaikan atau perapian struktur data. Untuk melakukan penataan dalam studi kasus ini adalah dengan mengubah format dari ketiga tabel tersebut menjadi seragam yaitu dalam format tabular sehingga gambar 7 (family data) dan gambar 8 (education data) dilakukan proses *transpose*. Selanjutnya ketiga tabel tersebut digabungkan menjadi satu tabel berdasarkan primary key “Entity Group” dan “Personnel Number” sehingga dapat diperoleh satu tabel data karyawan yang lengkap dan terstruktur seperti terlihat pada gambar 9.

Entity Group	Personnel Number	Name	Position Name	Birthdate	Nationality	Employee Status	Marital Status	Number of Children	Education Level	Edu Institution Name
Head Office	1012424	Sutopo	Coordination Technician	31.01.1970	Indonesian	Local	Married	3	Senior High School	STM Pancasila 1
BO Tangerang	1004245	Adji Cahyono	Operator, Finishing	02.12.1976	Indonesian	Local	Married	1	Senior High School	
BO Surabaya	1014251	Budianto	Technician, Electrical Maintenance	10.10.1976	Indonesian	Local	Single	3	Senior High School	SMA N 01
BO Surabaya	1024827	Steven Chow	Head, Finishing	101219.67	Chinese	Expatriate	Married	3	University/S2	Nanjing Tech University
BO Tangerang	1095945	Christian Stephen	Technician, QC	01.06.1975	Indonesian	Local	Married	1	Junior High School	Pemda
Head Office	1084752	Heryanti	Officer, Optimization Facility	09.09.1971	Indonesian	Local	Married	1	Senior High School	
Head Office	1084344	Huang Xiaoming	Head, Business Optimization	08.11.1968	Taiwanese	Expatriate	Married	2	University/S2	Chinese Culture University
BO Tangerang	1420495	Alina Winata	Manager, Production	23.07.1965	Indonesian	Local	Married	3	University/S1	Universitas Bina Nusantara
Head Office	1001596	Andi Tandi	Specialist, Industrial License	02.11.1967	Indonesian	Local	Married	0	Senior High School	SGPN IKIP
BO Surabaya	1042748	Lee Min Ho	Head, Cogen	22.01.1970	Taiwanese	Expatriate	Married	1	University/S1	National Taipei University of Technology
BO Surabaya	1094904	Steve Widiono	Manager, IT	101019.8	Indonesian	Local	Married	0	University/S2	Universitas Indonesia
Head Office	1055207	John Black	Head, Market Research & Development	01.02.1951	American	Expatriate	Married	1	University/S2	
Head Office	1001434	Jennifer Magdalena	Head, Log Opt (Sea Transportation)	29.01.1973	Indonesian	Local	Married	2	University/S1	S1/Untar Jakarta
Head Office	1070089	Wang Chang Su	Head, Product Research & Development	08.12.1959	Chinese	Expatriate	Married	2	University/S2	Harvard University
Head Office	1070089	Wang Chang Su	Head, Product Research & Development	08.12.1959	Chinese	Expatriate	Married	2	University/S2	Harvard University
BO Tangerang	1005295	Angela Veronica	Sales Executive	07.03.1987	Indonesian	Local	Single	2	University/S1	Universitas Trisakti
BO Surabaya	1985001	Bembi Eko	Officer, Accounting	02.01.1976	Indonesian	Local	Married	0	Senior High School	SMAN 02
BO Surabaya	1124952	Jacky Chan	Manager, QC	21119.81	Taiwanese	Expatriate	Married	2	University/S2	Chinese Culture University
BO Surabaya	1412577	Rakesh Kumar	Expert, Material Supply	14.07.1984	Indian	Expatriate	Single	1	University/S1	
Head Office	1002177	Ratna Sari	Specialist, Industrial License	21.04.1970	Indonesian	Local	Married	1	Senior High School	SGON Jakarta
BO Tangerang	1100499	Slamet Dodi Irawan	Operator, Field Boiler - Carton Box	15.05.1970	Indonesian	Local	Married	2	Senior High School	SMAN 78
BO Tangerang	1100499	Slamet Dodi Irawan	Operator, Field Boiler - Carton Box	15.05.1970	Indonesian	Local	Married	2	Senior High School	SMAN 78

Gambar 9. Tabel Hasil Penataan (Structuring) dan Konsolidasi Data Karyawan

3.3. Pembersihan (Cleaning)

Pada studi kasus ini, setelah dilakukan penataan terhadap data karyawan dan mendapatkan hasil konsolidasi data karyawan dalam satu tabel data karyawan yang lengkap dan terstruktur, maka selanjutnya harus ditinjau setiap kolom dan baris dalam tahap pembersihan (*cleaning*) untuk memastikan data-data tersebut berada dalam kondisi yang andal sebelum dianalisis lebih lanjut. Data cleaning merupakan tahapan untuk membersihkan data yang masih belum sesuai format, masih terdapat duplikasi, maupun melengkapi data-data tertentu yang masih kosong. Data yang perlu diperbaiki terdapat dalam gambar 10, meliputi:

1. kekeliruan format data pada kolom birthdate
2. adanya duplikasi baris
3. adanya status anomali pada marital status dengan number of children karena terdapat data yang status pernikahan yang belum menikah (single) tetapi sudah mempunyai data anak.
4. adanya data yang kosong pada kolom edu institution name

Entity Group	Personnel Number	Name	Position Name	Birthdate	Nationality	Employee Status	Marital Status	Number of Children	Education Level	Edu Institution Name
Head Office	1012424	Sutopo	Coordination Technician	31.01.1970	Indonesian	Local	Married	3	Senior High School	STM Pancasila 1
BO Tangerang	1004245	Adji Cahyono	Operator, Finishing	02.12.1976	Indonesian	Local	Married	1	Senior High School	Missing Values
BO Surabaya	1014251	Budianto	Technician, Electrical Maintenance	10.10.1976	Indonesian	Local	Single	3	Senior High School	SMA N 01
BO Surabaya	1024827	Steven Chow	Head, Finishing	10121967%	Chinese	Expatriate	Married	3	University/S2	Nanjing Tech University
BO Tangerang	1095945	Christian Stephen	Technician, QC	01.06.1975	Indonesian	Local	Married	1	Junior High School	Pemda
Head Office	1084752	Heryanti	Officer, Optimization Facility	09.09.1971	Indonesian	Local	Married	1	Senior High School	Missing Values
Head Office	1084344	Huang Xiaoming	Head, Business Optimization	08.11.1968	Taiwanese	Expatriate	Married	2	University/S2	Chinese Culture University
BO Tangerang	1420495	Alina Winata	Manager, Production	23.07.1965	Indonesian	Local	Married	3	University/S1	Universitas Bina Nusantara
Head Office	1001596	Andi Tandi	Specialist, Industrial License	02.11.1967	Indonesian	Local	Married	0	Senior High School	SGPN IKIP
BO Surabaya	1042748	Lee Min Ho	Head, Cogen	22.01.1970	Taiwanese	Expatriate	Married	1	University/S1	National Taipei University of Technology
BO Surabaya	1094904	Steve Widiono	Manager, IT	10101980%	Indonesian	Local	Married	0	University/S2	Universitas Indonesia
Head Office	1055207	John Black	Head, Market Research & Development	01.02.1951	American	Expatriate	Married	1	University/S2	Missing Values
Head Office	1001434	Jennifer Magdalena	Head, Log Opt (Sea Transportation)	29.01.1973	Indonesian	Local	Married	2	University/S1	S1/Untar Jakarta
Head Office	1070089	Wang Chang Su	Head, Product Research & Development	08.12.1959	Chinese	Expatriate	Married	2	University/S2	Harvard University
Head Office	1070089	Wang Chang Su	Head, Product Research & Development	08.12.1959	Chinese	Expatriate	Married	2	University/S2	Harvard University
BO Tangerang	1005295	Angela Veronica	Sales Executive	07.03.1987	Indonesian	Local	Single	2	University/S1	Universitas Trisakti
BO Surabaya	1985001	Bembi Eko	Officer, Accounting	02.01.1976	Indonesian	Local	Married	0	Senior High School	SMAN 02
BO Surabaya	1124952	Jacky Chan	Manager, QC	2111981%	Taiwanese	Expatriate	Married	2	University/S2	Chinese Culture University
BO Surabaya	1412577	Rakesh Kumar	Expert, Material Supply	14.07.1984	Indian	Expatriate	Single	1	University/S1	Missing Values
Head Office	1002177	Ratna Sari	Specialist, Industrial License	21.04.1970	Indonesian	Local	Married	1	Senior High School	SGON Jakarta
BO Tangerang	1100499	Slamet Dodi Irawan	Operator, Field Boiler - Carton Box	15.05.1970	Indonesian	Local	Married	2	Senior High School	SMAN 78
BO Tangerang	1100499	Slamet Dodi Irawan	Operator, Field Boiler - Carton Box	15.05.1970	Indonesian	Local	Married	2	Senior High School	SMAN 78

Gambar 10. Tabel data karyawan sebelum proses pembersihan

Untuk data yang kolomnya kosong (*null*), maka dilakukan pemeriksaan kembali ke file yang tersimpan di SAP central repository storage, ataupun dilengkapi berdasarkan data yang ada di dokumen kepegawaian yang tersimpan di lemari arsip. Setelah data sudah dibersihkan dan dilengkapi, maka dapat dilihat pada gambar 11, bahwa data menjadi rapi tanpa ada duplikasi ataupun kekosongan nilai, sehingga ke depannya data-data ini akan lebih mudah untuk dianalisis.

Entity Group	Personnel Number	Name	Position Name	Birthdate	Nationality	Employee Status	Marital Status	Number of Children	Education Level	Edu Institution Name
Head Office	1012424	Sutopo	Coordination Technician	31.01.1970	Indonesian	Local	Married	3	Senior High School	STM Pancasila 1
BO Tangerang	1004245	Adji Cahyono	Operator, Finishing	02.12.1976	Indonesian	Local	Married	1	Senior High School	SMU Ipeka II
BO Surabaya	1014251	Budianto	Technician, Electrical Maintenance	10.10.1976	Indonesian	Local	Single	0	Senior High School	SMA N 01
BO Surabaya	1024827	Steven Chow	Head, Finishing	10.12.1967	Chinese	Expatriate	Married	3	University/S2	Nanjing Tech University
BO Tangerang	1095945	Christian Stephen	Technician, QC	01.06.1975	Indonesian	Local	Married	1	Junior High School	Pemda
Head Office	1084752	Heryanti	Officer, Optimization Facility	09.09.1971	Indonesian	Local	Married	1	Senior High School	SMU Ipeka II
Head Office	1084344	Huang Xiaoming	Head, Business Optimization	08.11.1968	Taiwanese	Expatriate	Married	2	University/S2	Chinese Culture University
BO Tangerang	1420495	Alina Winata	Manager, Production	23.07.1965	Indonesian	Local	Married	3	University/S1	Universitas Bina Nusantara
Head Office	1001596	Andi Tandi	Specialist, Industrial License	02.11.1967	Indonesian	Local	Married	0	Senior High School	SGPN IKIP
BO Surabaya	1042748	Lee Min Ho	Head, Cogen	22.01.1970	Taiwanese	Expatriate	Married	1	University/S1	National Taipei University of Technology
BO Surabaya	1094904	Steve Widiono	Manager, IT	10.10.1980	Indonesian	Local	Married	0	University/S2	Universitas Indonesia
Head Office	1055207	John Black	Head, Market Research & Development	01.02.1951	American	Expatriate	Married	1	University/S2	Harvard University
Head Office	1001434	Jennifer Magdalena	Head, Log Opt (Sea Transportation)	29.01.1973	Indonesian	Local	Married	2	University/S1	S1/Untar Jakarta
Head Office	1070089	Wang Chang Su	Head, Product Research & Development	08.12.1959	Chinese	Expatriate	Married	2	University/S2	Harvard University
BO Tangerang	1005295	Angela Veronica	Sales Executive	07.03.1987	Indonesian	Local	Single	0	University/S1	Universitas Trisakti
BO Surabaya	1985001	Bembi Eko	Officer, Accounting	02.01.1976	Indonesian	Local	Married	0	Senior High School	SMAN 02
BO Surabaya	1124952	Jacky Chan	Manager, QC	21.01.1981	Taiwanese	Expatriate	Married	2	University/S2	Chinese Culture University
BO Surabaya	1412577	Rakesh Kumar	Expert, Material Supply	14.07.1984	Indian	Expatriate	Single	0	University/S1	Harvard University
Head Office	1002177	Ratna Sari	Specialist, Industrial License	21.04.1970	Indonesian	Local	Married	1	Senior High School	SGON Jakarta
BO Tangerang	1100499	Slamet Dodi Irawan	Operator, Field Boiler - Carton Box	15.05.1970	Indonesian	Local	Married	2	Senior High School	SMAN 78

Gambar 11. Tabel data karyawan setelah proses pembersihan

3.4. Pengayaan (*Enriching*)

Setelah pembersihan, data harus diperkaya, dan ini dilakukan pada langkah pengayaan. Intinya di sini adalah untuk menambahkan lebih banyak data, baik dari sumber data internal maupun data eksternal lainnya, dengan mempertimbangkan data yang ada ini sudah cukup untuk mengakomodir proyek atau dashboard yang akan dikerjakan.

Pada studi kasus ini, data karyawan yang sudah rapi dan terstruktur akan diperkaya dengan menambahkan hasil penilaian kinerja (*Performance Rating*) selama 3 tahun terakhir dari tahun 2020 sampai dengan tahun 2022. Hal ini bertujuan agar bagian HRD dari Perusahaan XYZ dapat mengidentifikasi *talent* berdasarkan nilai *performance evaluation* dalam 3 tahun terakhir yang akan digunakan untuk melakukan program retensi bagi para karyawan yang berkinerja terbaik. Pengayaan data ini diperoleh dari sistem SuccessFactors modul Performance and Goal Management seperti yang ada pada gambar 12.

Dalam studi kasus ini, pengayaan data dilakukan dengan menggunakan formula *vlookup* dari spreadsheet data karyawan yang sudah dikonsolidasi terhadap spreadsheet data hasil penilaian kinerja karyawan. Hasilnya dapat dilihat dalam gambar 13.

Personnel Number	Name	Performance Rating 2020	Performance Rating 2021	Performance Rating 2022
1001434	Jennifer Magdalena	B	B	B
1001596	Andi Tandil	B	D	B
1002177	Ratna Sari	D	A	C
1004245	Adji Cahyono	B	B	B
1005295	Angela Veronica	B	A	B
1012424	Sutopo	B	A	B
1014251	Budianto	C	B	B
1024827	Steven Chow	B	D	B
1042748	Lee Min Ho	A	C	A
1055207	John Black	A	B	C
1070089	Wang Chang Su	A	B	B
1084344	Huang Xiaoming	B	B	A
1084752	Heryanti	C	A	B
1094904	Steve Widiono	C	D	B
1095945	Christian Stephen	C	A	A
1100499	Slamet Dodi Irawan	B	A	A
1124952	Jacky Chan	C	B	C
1412577	Rakesh Kumar	C	A	B
1420495	Alina Winata	C	D	B
1985001	Bembi Eko	A	B	B

Gambar 12. Data penilaian kinerja karyawan yang diunduh dari sistem SuccessFactors

Entity Group	Personnel Number	Name	Position Name	Birthdate	Nationality	Employee Status	Marital Status	Number of Children	Education Level	Edu Institution Name	Performance Rating 2020	Performance Rating 2021	Performance Rating 2022
Head Office	1012424	Sutopo	Coordination Technician	31.01.1970	Indonesian	Local	Married	3	Senior High School	STM Pancasila 1	B	A	B
BO Surabaya	1004245	Adji Cahyono	Operator, Finishing	02.12.1976	Indonesian	Local	Married	1	Senior High School	SMU Ipeka II	B	B	B
BO Surabaya	1014251	Budianto	Technician, Electrical Maintenance	10.10.1976	Indonesian	Local	Single	0	Senior High School	SMA N 01	C	B	B
BO Surabaya	1024827	Steven Chow	Head, Finishing	10.12.1967	Chinese	Expatriate	Married	3	University/S2	Nanjing Tech University	B	D	B
BO Tangerang	1095945	Christian Stephen	Technician, QC	01.06.1975	Indonesian	Local	Married	1	Junior High School	Pemda	C	A	A
Head Office	1084752	Heryanti	Officer, Optimization Facility	09.09.1971	Indonesian	Local	Married	1	Senior High School	SMU Ipeka II	C	A	B
Head Office	1084344	Huang Xiaoming	Head, Business Optimization	08.11.1968	Taiwanese	Expatriate	Married	2	University/S2	Chinese Culture University	B	B	A
BO Tangerang	1420495	Alina Winata	Manager, Production	23.07.1965	Indonesian	Local	Married	3	University/S1	Universitas Bina Nusantara	C	D	B
Head Office	1001596	Andi Tandil	Specialist, Industrial License	02.11.1967	Indonesian	Local	Married	0	Senior High School	SGPN IKIP	B	D	B
BO Surabaya	1042748	Lee Min Ho	Head, Cogen	22.01.1970	Taiwanese	Expatriate	Married	1	University/S1	National Taipei University of Technology	A	C	A
BO Surabaya	1094904	Steve Widiono	Manager, IT	10.10.1980	Indonesian	Local	Married	0	University/S2	Universitas Indonesia	C	D	B
Head Office	1055207	John Black	Head, Market Research & Development	01.02.1951	American	Expatriate	Married	1	University/S2	Harvard University	A	B	C
Head Office	1001434	Jennifer Magdalena	Head, Log Opt (Sea Transportation)	29.01.1973	Indonesian	Local	Married	2	University/S1	SI/Untar Jakarta	B	B	B
Head Office	1070089	Wang Chang Su	Head, Product Research & Development	08.12.1959	Chinese	Expatriate	Married	2	University/S2	Harvard University	A	B	B
BO Tangerang	1005295	Angela Veronica	Sales Executive	07.03.1987	Indonesian	Local	Single	0	University/S1	Universitas Trisakti	B	A	B
BO Surabaya	1985001	Bembi Eko	Officer, Accounting	02.01.1976	Indonesian	Local	Married	0	Senior High School	SMAN 02	A	B	B
BO Surabaya	1124952	Jacky Chan	Manager, QC	21.01.1981	Taiwanese	Expatriate	Married	2	University/S2	Chinese Culture University	C	B	C
BO Surabaya	1412577	Rakesh Kumar	Expert, Material Supply	14.07.1984	Indian	Expatriate	Single	0	University/S1	Harvard University	C	A	B
Head Office	1002177	Ratna Sari	Specialist, Industrial License	21.04.1970	Indonesian	Local	Married	1	Senior High School	SGON Jakarta	D	A	C
BO Tangerang	1100499	Slamet Dodi Irawan	Operator, Field Boiler - Carton Box	15.05.1970	Indonesian	Local	Married	2	Senior High School	SMAN 78	B	A	A

Gambar 13. Data karyawan yang sudah diperkaya dengan data penilaian kinerja karyawan yang diunduh dari sistem SuccessFactors

3.5. Pemvalidasian (Validating)

Langkah validasi (Validating) dilakukan untuk mengungkap masalah kualitas data dan konsistensi data. Apabila pada data karyawan tersebut akan ditambahkan dengan kolom / atribut tanggal perekrutan (*hiring date*), maka harus diperiksa konsistensi atribut "hiring date" ini sudah sesuai dengan format yang digunakan untuk atribut tanggal lahir (*birthdate*). Dalam studi kasus ini, tahap pemvalidasian lebih kepada memeriksa kembali kesesuaian data, dan meninjau kembali apakah data-data ini sudah dapat mengakomodir kebutuhan proyek dan memberikan hasil yang lebih akurat dalam analisis data. Hasil validasi dapat dilihat dalam gambar 14.

Entity Group	Personnel Number	Name	Position Name	Birthdate	Nationality	Employee Status	Marital Status	Number of Children	Education Level	Edu Institution Name	Performance Rating 2020	Performance Rating 2021	Performance Rating 2022
BO Surabaya	1014251	Budianto	Technician, Electrical Maintenance	10.10.1976	Indonesian	Local	Single	0	Senior High School	SMAN 01	C	B	B
BO Surabaya	1024827	Steven Chow	Head, Finishing	10.12.1967	Chinese	Expatriate	Married	3	University/S2	Nanjing Tech University	B	D	B
BO Surabaya	1042748	Lee Min Ho	Head, Cogen	22.01.1970	Taiwanese	Expatriate	Married	1	University/S1	National Taipei University of Technology	A	C	A
BO Surabaya	1094904	Steve Widiono	Manager, IT	10.10.1980	Indonesian	Local	Married	0	University/S2	Universitas Indonesia	C	D	B
BO Surabaya	1985001	Bembi Eko	Officer, Accounting	02.01.1976	Indonesian	Local	Married	0	Senior High School	SMAN 02	A	B	B
BO Surabaya	1124952	Jacky Chan	Manager, QC	21.01.1981	Taiwanese	Expatriate	Married	2	University/S2	Chinese Culture University	C	B	C
BO Surabaya	1412577	Rakesh Kumar	Expert, Material Supply	14.07.1984	Indian	Expatriate	Single	0	University/S1	Harvard University	C	A	B

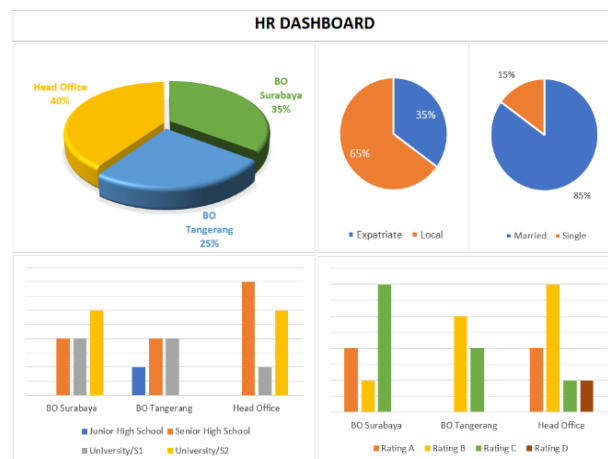
Gambar 14. Tabel yang sudah ditambahkan color formatting untuk mempermudah validasi

Pada tahap pemvalidasian ini, dapat dilakukan proses penyaringan (filtering) misalnya berdasarkan "entitiy group" dan menggunakan *color formatting* pada penilaian kinerja karyawan dengan warna yang berbeda-beda. Nilai A diberi format warna hijau, Nilai B diberi format warna kuning, Nilai C diberi format warna oranye dan Nilai D diberi format warna merah. Visualisasi data dengan *color formatting* ini akan membantu mempermudah proses validasi keakuratan data.

3.6. Penerbitan (Publishing)

Penerbitan merupakan langkah terakhir dari proses data wrangling untuk memastikan data sudah siap diterbitkan sehingga nantinya dapat digunakan untuk tujuan analisis. Penerbitan mengacu pada persiapan transfer data untuk digunakan dalam proyek HR Dashboard dengan mengirimkan atau memuat data ke dalam perangkat lunak analisis. Dashboard yang dibuat ini dapat digunakan analisis demografi karyawan, maupun untuk penentuan *career* dan *succession planning* karyawan berdasarkan evaluasi kinerja karyawan.

Melalui proses data wrangling yang tepat, data karyawan dari berbagai sumber di perusahaan XYZ dapat disiapkan sehingga menjadi lebih berkualitas, yang pada akhirnya dapat digunakan untuk membuat dashboard yang lebih akurat untuk pengambilan keputusan. Hasil dashboard terdapat dalam gambar 15.



Gambar 15. HR Dashboard di perusahaan XYZ

4. KESIMPULAN

Penelitian ini telah membahas dan mendemonstrasikan langkah demi langkah teknik data wrangling yang efektif untuk meningkatkan kualitas data karyawan (employee data), yang meliputi proses penemuan (discovery), penataan (structuring), pembersihan (cleaning), pengayaan (enriching), pemvalidasian (validating), dan penerbitan (publishing). Data karyawan dalam studi kasus ini menggunakan Excel Spreadsheets sebagai cara paling dasar untuk melakukan langkah demi langkah dari proses data wrangling.

Penelitian selanjutnya dapat mengevaluasi efektivitas teknik data wrangling dalam konteks yang berbeda, misalnya dalam data penjualan, data keuangan, atau data media sosial. Dengan konteks yang berbeda, mungkin diperlukan penyesuaian pada teknik data wrangling yang digunakan

DAFTAR PUSTAKA

- [1] Andre, L. (2021). 53 Important Statistics about How Much Data is Created Every Day. *FinancesOnline*.
- [2] Azeroual, O. (2020). Data Wrangling in Database Systems: Purging of Dirty Data. German Center for Higher Education Research and Science Studies (DZHW).
- [3] Endel, F., Piringer, H. (2015). Data Wrangling: Making data useful again. Elsevier, IFAC-PapersOnLine, Volume 48, Issue 1, 2015, Pages 111-112.
- [4] Furche, T., Gottlob, G., Libkin, L., Orsi, G., Paton, N. (2016). Data Wrangling for Big Data: Challenges and Opportunities. Published in Proc. 19th International Conference on Extending Database Technology (EDBT), March 15-18, 2016 - Bordeaux, France: ISBN 978-3-89318-070-7.
- [5] Mazilu, L., Paton, N., Konstantinou, N., Fernandes, A. (2020). Fairness in Data Wrangling. In 21st International Conference on Information Reuse and Integration for Data Science, IRI 2020, Las Vegas, NV, USA.
- [6] Patil, M., Hiremath, B. (2018). A Systematic Study of Data Wrangling. *I.J. Information Technology and Computer Science*, 2018, 1, 32-39 Published Online January 2018 in MECS (<http://www.mecspress.org/>) DOI: 10.5815/ijitcs.2018.01.04.
- [7] Siegler, M. (2010). Eric Schmidt: Every 2 Days We Create as Much Information as We Did Up To 2003.
- [8] Stobierski, T. (2021). Data Wrangling: What it is & why it's important. *Harvard Business School Online*.
- [9] Taylor, P. (2022). Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025. *Statista*.
- [10] Terrizzano, I., Schwarz, P., Roth, M., Colino, J. (2015). Data wrangling: The challenging journey from the wild to the lake. *CIDR Conference paper 2015*.
- [11] Thurber, M. (2018). What is Data Wrangling and Why Does it Take So Long? *Elder Research*.
- [12] Todd, S. (2020). Data wrangling vs. Data Cleaning: What's the difference? *Inzata Analytics*.
- [13] Vodovatova, E. (2019). What is Data Wrangling (Data Munging)? Steps, Solutions, and Tools. *Theappsolutions*.
- [14] Wolff, R. (2021). What Is Data Wrangling & Why Is It Necessary? *Monkeylearn*. <https://monkeylearn.com>
- [15] Yasar, K., Essex, D., Rebernik, D. (2022). SAP SuccessFactors. *TechTarget*.