



Implementasi *Data Warehouse* dan *Business Intelligence* Menggunakan Pentaho dan Metabase untuk Membuat *Dashboard* Visualisasi Kinerja Penjualan *E-Commerce Wish*

^{1*}Murtiwiwati, ²Hansel Agathon, ³Leli Safitri

^{1,2,3} Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma, Depok, Indonesia

Email: murtiwiwati@staff.gunadarma.ac.id^{1*}, hansellagathon@gmail.com², leli.s@staff.gunadarma.ac.id³

Alamat : Jl. Margonda Raya No. 100, 16421, Depok, Indonesia

Korespondensi penulis : murtiwiwati@staff.gunadarma.ac.id

Abstract As a business develops, the volume of data generated also increases. With the amount of data that continues to grow, it is necessary to process data that can support the improvement of the quality of business processes. Data processing on Wish can be done by implementing the Extract, Transform, Load process in the data warehouse using Pentaho and visualizing business intelligence with a dashboard using Metabase. The data used in this research is Wish transaction data for the period 2020-2023. The results of this study are a data warehouse using a PostgreSQL database and a dashboard to display Wish's sales performance. The results of the analysis of Wish's sales performance show that during 2020-2023 sales have increased, but most of these sales only come from one country, namely the United States. Therefore, Wish needs to make new innovations to attract customers from other countries.

Keywords: Business, Intelligence, Metabase, Pentaho

Abstrak Seiring dengan berkembangnya sebuah bisnis, volume data yang dihasilkan juga mengalami peningkatan. Dengan jumlah data yang terus bertambah, perlu dilakukan proses pengolahan data yang dapat menunjang peningkatan mutu proses bisnis. Pengolahan data pada Wish dapat dilakukan dengan mengimplementasikan proses *Extract, Transform, Load* pada *data warehouse* menggunakan Pentaho dan memvisualisasikan *business intelligence* dengan *dashboard* menggunakan Metabase. Data yang digunakan dalam penelitian ini berupa data transaksi Wish periode 2020-2023. Hasil penelitian ini berupa sebuah *data warehouse* menggunakan *database* PostgreSQL dan *dashboard* untuk menampilkan kinerja penjualan Wish. Hasil analisis kinerja penjualan Wish menunjukkan bahwa selama tahun 2020-2023 penjualan mengalami peningkatan, tetapi sebagian besar dari penjualan tersebut hanya berasal dari satu negara, yaitu Amerika Serikat. Oleh karena itu, Wish perlu melakukan inovasi baru untuk menarik minat pelanggan dari negara-negara lain.

Kata kunci: Intelijen, Bisnis, Metabase, Pentaho

1. PENDAHULUAN

Seiring dengan berkembangnya sebuah bisnis, volume data yang dihasilkan juga mengalami peningkatan. Hal ini mengakibatkan proses pengolahan data menjadi lebih sulit dilakukan. Proses pengolahan data yang efisien menjadi semakin penting karena data ini digunakan untuk mendapatkan informasi guna menunjang proses bisnis. Proses pengolahan data secara tradisional tidak efisien untuk mengolah data dalam jumlah yang besar, keterbatasan tersebut mengakibatkan keterlambatan dalam mendapatkan informasi.

Wish merupakan sebuah *e-commerce* asal Amerika Serikat yang menghasilkan data transaksi dalam jumlah besar dari proses bisnis yang dijalankannya. Data transaksi tersebut dipublikasikan di sebuah situs web penyedia data publik, yaitu Kaggle. Data yang dihasilkan dari proses bisnis Wish terus bertambah, sehingga diperlukan suatu solusi untuk meningkatkan efisiensi proses pengolahan data. Implementasi *data warehouse* dan *Extract, Transform, Load* (ETL) dapat menjadi solusi untuk mengatasi masalah tersebut. Dengan ETL, proses ekstraksi, transformasi, dan pemuatan data ke dalam *data warehouse* dapat dilakukan dengan lebih mudah. *Data warehouse* berfungsi sebagai tempat penyimpanan data yang tersentralisasi guna memudahkan penyimpanan, pencarian, maupun pengolahan data lebih lanjut. Informasi yang didapatkan dari data yang telah diolah dapat divisualisasikan dalam bentuk *dashboard* melalui proses *Business Intelligence* (BI) guna memudahkan penyediaan informasi dalam bentuk grafik.

ETL (*Extract, Transform, Load*) adalah serangkaian proses untuk mengintegrasikan data yang terdiri dari tiga tahap utama: ekstraksi data dari berbagai sumber, transformasi data ke dalam format yang sesuai, dan pemuatan data ke dalam sistem tujuan seperti *data warehouse*. Proses ETL digunakan untuk mengumpulkan, memproses, dan menyimpan data yang relevan sehingga bisa digunakan untuk analisis dan pengambilan keputusan (Puri, 2021).

Data warehouse adalah sistem yang mengintegrasikan data dari berbagai sumber operasional untuk menyediakan lingkungan yang stabil untuk pelaporan, kueri, dan analisis. Ini biasanya melibatkan proses ETL (*extract, transform, load*) untuk menyalin, memformat, dan memuat data ke dalam *data warehouse*. (Knight, 2023)

Arsitektur *data warehouse* biasanya terdiri dari tiga lapisan: lapisan data, lapisan aplikasi (OLAP *server*), dan lapisan presentasi yang dirancang untuk pengguna akhir dengan alat dan antarmuka pemrograman aplikasi untuk ekstraksi dan analisis data. (Coursera, 2023)

Online Analytical Processing (OLAP)

OLAP adalah teknologi yang memungkinkan analisis cepat dari data multidimensi yang disimpan dalam *data warehouse*. Ini memungkinkan pengguna untuk melakukan *query* dan menganalisis data dengan cara yang intuitif dan fleksibel, mendukung berbagai operasi seperti pengelompokan, pemecahan, dan agregasi data. (Ramya & Prasad, 2021)

BI melibatkan serangkaian proses termasuk pengumpulan data, penyimpanan data, analisis data, dan penyajian informasi. Tujuannya adalah untuk mendukung proses pengambilan keputusan di semua tingkat organisasi. BI mengintegrasikan data dari berbagai sumber untuk memberikan pandangan yang komprehensif tentang bisnis, membantu

mengidentifikasi tren, pola, dan wawasan yang dapat mendorong keputusan strategis. (Talaoui & Kohtamäki, 2021)

Visualisasi data merupakan proses kreatif dan teknis dari merancang dan mengembangkan tampilan visual yang efektif dan efisien dari data. Proses desain visualisasi data dimulai dari pemahaman atas data yang akan divisualisasikan. Hal ini meliputi pengumpulan data, analisis data, dan pemilihan jenis visualisasi yang paling sesuai untuk data yang ada. Setelah itu, desainer visualisasi data akan memulai proses merancang tampilan visual dari data yang telah dipilih. (Kurniawan et al., 2023)

Elemen utama dari *Pentaho Data Integration* (PDI) adalah *Transformation* dan *Job*. *Transformation* adalah sekumpulan instruksi untuk mengubah *input* menjadi *output* yang diinginkan (*input-process-output*). Sedangkan *Job* adalah kumpulan instruksi untuk menjalankan transformasi. *Pentaho Community Edition* atau *Pentaho BI Suite* adalah produk dari Pentaho yang sudah terintegrasi dengan kemampuan *reporting*, *dashboard*, *data mining*, *workflow*, dan ETL. (Saraswati & Martarini, 2020)

Visual Studio Code adalah aplikasi *code editor* untuk melakukan pengembangan aplikasi atau program. Visual Studio Code dibuat oleh Microsoft dan telah menjadi aplikasi *open source* yang memiliki banyak *extensions* yang membantu proses *Software Engineer* dalam melakukan pengembangan aplikasi. (ADITYO, 2022)

Metabase adalah sebuah aplikasi yang sering digunakan untuk keperluan *business intelligence*. Dengan menggunakan Metabase, dapat dilakukan visualisasi data melalui grafik. Metabase dapat dipasang secara *on-premise* di *server internal*, sehingga penggunaannya dapat menyesuaikan spesifikasi *server* yang dibutuhkan perusahaan dengan lebih mudah. Metabase juga memiliki fitur untuk ditanam di aplikasi masing-masing sebagai *iFrame*. Sehingga dalam praktiknya dapat diimplementasikan ke dalam *dashboard* klien dan *superadmin* bahkan dengan aplikasi eksternal milik klien dengan mudah. Metabase juga sudah mendukung beberapa bahasa pemrograman seperti PHP, Javascript, Golang dan Python. (ADITYO, 2022)

Menurut Muhammad Ariffudin (2021) PostgreSQL adalah merupakan sebuah sistem manajemen database relasional (RDBMS) yang bersifat *open source*. Manajemen database ini dapat mengolah data dalam tabel yang memiliki relasi satu sama lain dan dapat digunakan secara gratis serta bebas dimodifikasi. Dikembangkan oleh Berkeley Computer Science Department, PostgreSQL telah menjadi database yang andal dalam 30 tahun terakhir ini. Sistem manajemen database ini memiliki performa stabil, keamanan tinggi, serta fitur melimpah.

2. METODE PENELITIAN

Penelitian ini dilakukan menggunakan pendekatan kualitatif dengan dengan beberapa tahap, diantaranya:

1. Pengumpulan data

Pada tahap ini dilakukan proses pengumpulan data dengan cara melakukan dokumentasi data publik berupa data penjualan *e-commerce* Wish.

2. Analisis data

Pada tahap ini data dianalisis guna memahami dan mengidentifikasi kebutuhan transformasi data.

3. Pembersihan data

Pada tahap ini data mentah dibersihkan hingga menjadi data yang bersih, pembersihan data dilakukan dengan cara menghapus duplikasi pada data, penanganan *missing value*, standarisasi data, koreksi kesalahan pada data, dan penghapusan data yang tidak relevan.

4. Perancangan *data warehouse*

Pada tahap ini dilakukan perancangan *data warehouse* berdasarkan data yang sudah dibersihkan, *data warehouse* terdiri dari tabel fakta dan tabel dimensi dengan skema *snowflake*.

5. Penerapan ETL

Pada tahap ini dibuat alur proses ETL, dari data diekstraksi, ditransformasi, hingga dimuat ke dalam *data warehouse*.

6. Visualisasi BI

Pada tahap ini dilakukan visualisasi terhadap data yang sudah dimuat ke dalam *data warehouse*, visualisasi dibuat dalam bentuk *dashboard* dan menampilkan informasi kinerja penjualan Wish yang terbagi dalam beberapa area dalam bentuk grafik.

3. PEMBAHASAN

Analisis Data

Analisis data dilakukan dengan tujuan untuk memahami dan mengidentifikasi kebutuhan transformasi data. Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data penjualan pada *e-commerce* Wish dari tahun 2020-2023, dengan total lebih dari 20.000 data dalam bentuk *file* dengan format csv (*comma-separated values*). Data terbagi dalam lima *file*, yaitu *file* data pelanggan, *file* data produk, *file* data pesanan, *file* data rincian pesanan, dan *file* data wilayah. Variabel yang akan divisualisasikan adalah data

transaksi, keuntungan, asal negara pelanggan, tren penjualan bulanan, produk terlaris, dan kategori produk terlaris.

Pembersihan Data

Data mentah perlu dibersihkan terlebih dahulu sebelum diproses lebih lanjut, proses pembersihan data ini dilakukan untuk menghilangkan duplikasi pada data, menghilangkan atribut atau simbol yang tidak diperlukan, mengatasi *missing value*, dan sebagainya. Data mentah perlu dibersihkan agar data menjadi bersih dan terorganisir dengan baik, proses pembersihan data ini akan dilakukan dengan menggunakan sebuah *script* Python di mana data akan secara otomatis dibersihkan dengan dijalkannya *script* tersebut, semua data yang sudah dibersihkan akan disimpan kembali dalam bentuk file csv.

1. Pembersihan Data Pelanggan

```

1 import pandas as pd
2
3 cus_df = pd.read_csv('data/Customer.csv')
4 reg_df = pd.read_csv('data/Region.csv')
5
6 cus_df['Gender'] = cus_df['Gender'].replace({'M': 'Male', 'F': 'Female'})
7 mer_df = pd.merge(cus_df, reg_df[['City', 'State Code', 'State']], on=['City', 'State Code'], how='left')
8 mer_df = pd.merge(mer_df, reg_df[['City', 'State Code', 'Country']], on=['City', 'State Code'], how='left')
9 mer_df = mer_df.drop(columns=['State Code', 'Birthday'])
10 mer_df = mer_df[['CustomerKey', 'Gender', 'Name', 'City', 'State', 'Country', 'Zip Code']]
11
12 mer_df.to_csv('data/Customer.csv', index=False)

```

Gambar 1. *Script* Python pembersihan data pelanggan

Script Python pada gambar 1. dibuat untuk membersihkan data pelanggan akan menghasilkan data baru dengan data jenis kelamin yang sudah diubah, data negara dan negara bagian yang lebih jelas, dan penghapusan kolom kode negara bagian dan tanggal ulang tahun. Data pelanggan yang telah dibersihkan dapat dilihat pada gambar 2.

	CustomerKey	Gender	Name	City	State	Country	Zip Code
0	221614f7f59428ca5d8e7061abae25e	Female	Lilly Harding	WANDEARAH EAST	South Australia	Australia	5523
1	669668857d19473bab416b947461dff8	Female	Madison Hull	MOUNT BUDD	Western Australia	Australia	6522
2	9aae376ab5cd4398bf2cab81b007a824	Female	Claire Ferres	WINJALLOK	Victoria	Australia	3380
3	24c3657e08044abc9dd54ba9e7b13103	Male	Jai Poltpalingada	MIDDLE RIVER	South Australia	Australia	5223
4	c4c8803b717c47c19d28c048f1bd187c	NaN	Aidan Pankhurst	TAWONGA SOUTH	Victoria	Australia	3698

Gambar 2. Data pelanggan baru

2. Pemberihan Data Produk

```
1 import pandas as pd
2
3 df = pd.read_csv('data/Product.csv')
4
5 df['Unit Cost USD'] = df['Unit Cost USD'].replace({'r'\$': ''}, regex=True)
6 df['Unit Cost USD'] = df['Unit Cost USD'].replace({'r','': ''}, regex=True)
7
8 df['Unit Price USD'] = df['Unit Price USD'].replace({'r'\$': ''}, regex=True)
9 df['Unit Price USD'] = df['Unit Price USD'].replace({'r','': ''}, regex=True)
10
11 df['Unit Cost USD'] = df['Unit Cost USD'].astype(float)
12 df['Unit Price USD'] = df['Unit Price USD'].astype(float)
13
14 df['Profit'] = df['Unit Price USD'] - df['Unit Cost USD']
15
16 df = df[['ProductKey', 'Product Name', 'Brand', 'Unit Cost USD', 'Unit Price USD', 'Profit', 'Category', 'Subcategory']]
17
18 df.to_csv('data/Product.csv', index=False)
```

Gambar 3. Script Python pembersihan data produk

Script Python pada Gambar 3 dibuat untuk membersihkan data produk menghasilkan data baru dengan format dan tipe data dari kolom harga modal dan harga jual yang sudah diubah, dan terdapat kolom baru bernama profit yang merupakan profit dari tiap barang yang merupakan hasil selisih antara harga jual dan harga modal. Data produk yang telah dibersihkan dapat dilihat pada gambar 4.

	ProductKey	Product Name	Brand	Unit Cost USD	Unit Price USD	Profit	Category	Subcategory
0	1db89b615fed417ab505e453ef0dac63	Contoso 512MB MP3 Player E51	Contoso	6.62	12.99	6.37	Audio	MP4&MP3
1	f6459ea54dca4bf5adcc3a70dfafb7da	Contoso 1G MP3 Player E100	Contoso	7.40	14.52	7.12	Audio	MP4&MP3
2	e19bffb15d7c495992688f14cc8fc907	Contoso 2G MP3 Player E200	Contoso	11.00	21.57	10.57	Audio	MP4&MP3
3	771b14f940d8429a81eca52ec95b7a6d	Contoso 4G MP3 Player E400	Contoso	30.58	59.99	29.41	Audio	MP4&MP3
4	8646699ad89840f1ac253c992950fe4a	Contoso 4GB Flash MP3 Player E401	Contoso	35.72	77.68	41.96	Audio	MP4&MP3

Gambar 4. Data produk baru

3. Pemberihan Data Pesanan

```
1 import pandas as pd
2
3 df = pd.read_csv('data/Order.csv')
4
5 df['Order Date'] = pd.to_datetime(df['Order Date'])
6 df['Delivery Date'] = pd.to_datetime(df['Delivery Date'])
7 df['Received Date'] = pd.to_datetime(df['Received Date'])
8
9 df['Order Date'] = df['Order Date'].dt.strftime('%Y/%m/%d')
10 df['Delivery Date'] = df['Delivery Date'].dt.strftime('%Y/%m/%d')
11 df['Received Date'] = df['Received Date'].dt.strftime('%Y/%m/%d')
12
13 df.to_csv('data/Order.csv', index=False)
```

Gambar 5. Script Python pembersihan data pesanan

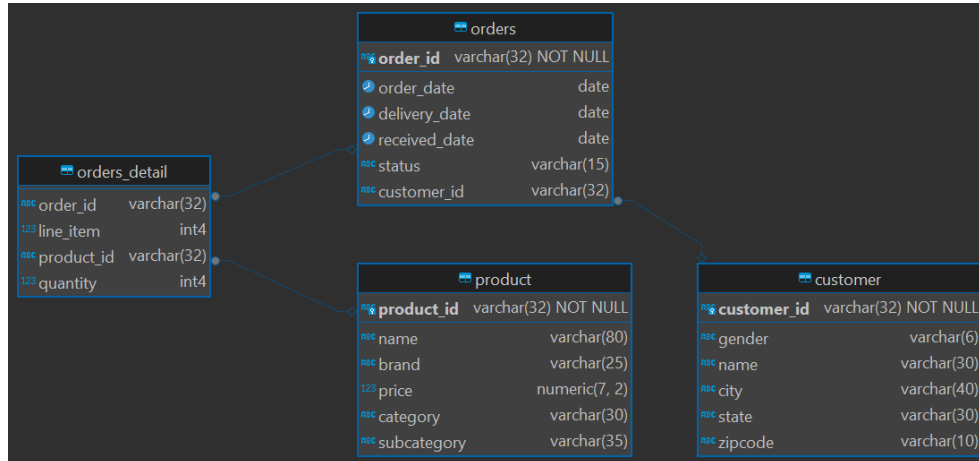
Script Python pada gambar 5. digunakan untuk membersihkan data pesanan menghasilkan data baru dengan format tanggal yang sudah diubah. Data pesanan yang telah dibersihkan dapat dilihat pada gambar 6.

	OrderKey	Order Date	Delivery Date	Received Date	Status	CustomerKey
0	395a8490a08347a8a92326f16846b79b	2020/01/01	2020/01/03	2020/02/05	Delivered	d1b18e1406454363b0fe46a13f5fd83c
1	8ac37c06340a4803bf8febf0763710c7	2020/01/01	2020/01/04	2020/03/02	Delivered	776047841720468aadfbf4ee3b4b7351
2	a62e6ce482bd46d8bc3ccb2529a3fe38	2020/01/01	2020/01/03	2020/01/22	Delivered	60f104e33caa47fcb95659907dca6f9c
3	ee9cbe5712a345348131103c2df7a762	2020/01/01	2020/01/01	2020/01/27	Delivered	016d60d1baeb4945b83bfb88ac79959a
4	55c818041cc547f98a38c2b7971d71f4	2020/01/01	2020/01/04	2020/02/22	Delivered	a680b6a0c6dc43439a8a18867d8aa8c2

Gambar 6. Data pesanan baru

Perancangan *Data Warehouse*

Data warehouse merupakan sebuah sistem yang bertugas untuk mengarsipkan dan melakukan analisis data historis untuk menunjang keperluan informasi pada sebuah bisnis ataupun organisasi. Berdasarkan analisa yang sudah dilakukan, tabel fakta akan menggunakan tabel pesanan, dan tabel dimensi akan menggunakan tabel pelanggan, detail pesanan, dan produk. Skema *data warehouse* dapat dilihat pada Gambar 7.



Gambar 7. Skema *data warehouse*

Penerapan ETL

Setelah melakukan pembersihan data dan membuat *data warehouse*, tahap berikutnya adalah menerapkan proses ETL untuk melakukan ekstraksi, transformasi, dan memuat data ke *data warehouse*. Proses ETL akan dilakukan dengan menggunakan *transformation* dan *database connection* pada Pentaho.

1. Extract

Proses ekstraksi dilakukan untuk mengambil data yang telah dibersihkan dengan menggunakan Python.

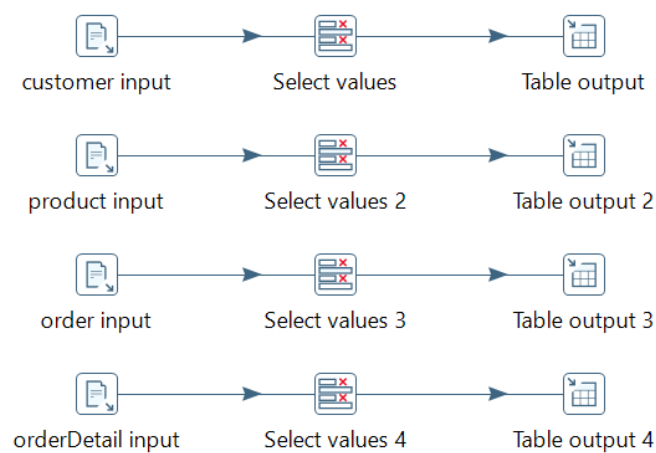
2. Transform

Proses transformasi dilakukan untuk mengubah struktur tabel seperti mengubah nama kolom agar lebih mudah dimengerti atau menghapus kolom yang tidak dipakai.

3. Load

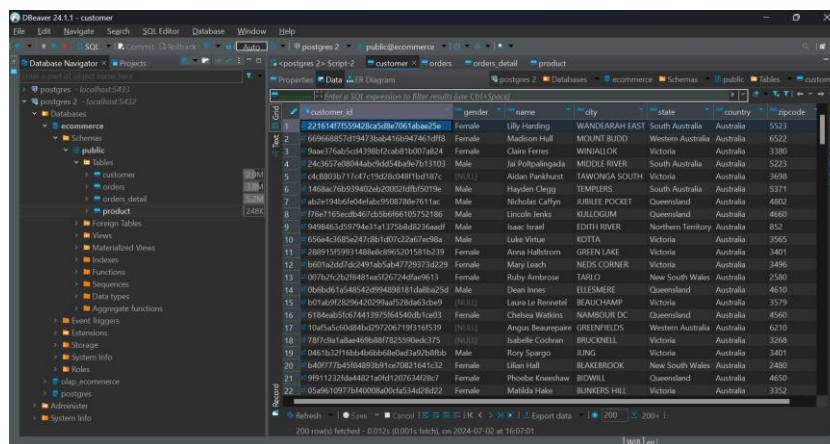
Proses *load* dilakukan untuk memuat data ke dalam *data warehouse*.

Seluruh rangkaian proses ETL dapat dilihat pada gambar 8.



Gambar 8. Rangkaian proses ETL

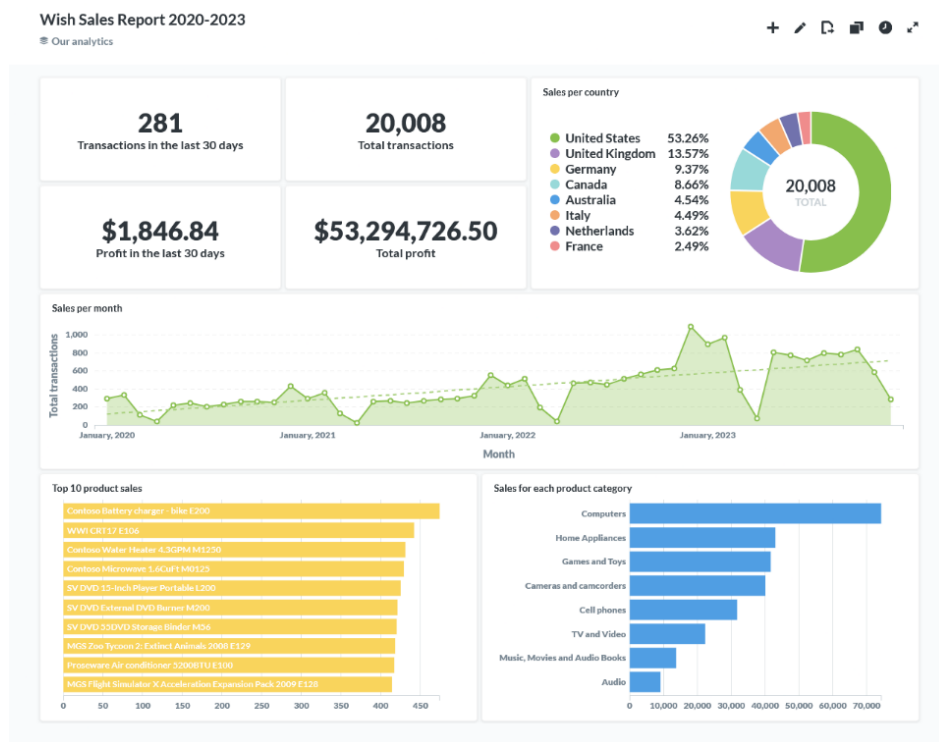
Hasil dari proses ETL dapat dilihat pada *data warehouse* menggunakan DBEaver seperti pada gambar 9.



Gambar 9. Tampilan data di dalam *data warehouse*

Pembuatan Dashboard

Tahap akhir dari penelitian ini yaitu melakukan pembuatan *dashboard* untuk menunjukkan kinerja penjualan pada *e-commerce* Wish. Terdapat beberapa segmentasi data yang dapat dijadikan laporan dalam menilai kinerja penjualan, yaitu segmentasi penjualan, segmentasi profit, segmentasi negara pelanggan, segmentasi produk, dan segmentasi kategori produk. Eksplorasi hasil laporan dan analisis atau dapat disebut dengan BI divisualisasikan dalam bentuk *dashboard* yang dibuat dengan menggunakan Metabase. Bentuk *dashboard* dapat dilihat pada gambar 10.



Gambar 10. *Dashboard e-commerce* Wish

4. KESIMPULAN

Implementasi *data warehouse* dan *Extract, Transform, Load* (ETL) dapat menjadi solusi untuk meningkatkan efisiensi proses pengolahan data pada *e-commerce* Wish. Dengan ETL, proses ekstraksi, transformasi, dan pemuatan data ke dalam *data warehouse* dapat dilakukan dengan lebih mudah. Informasi yang didapatkan dari data yang telah diolah dapat divisualisasikan dalam bentuk *dashboard* melalui proses *Business Intelligence* (BI) guna memudahkan penyediaan informasi dalam bentuk grafik. *Dashboard* yang dihasilkan memberikan informasi terkait kinerja penjualan *e-commerce* Wish dari tahun 2020-2023. *Dashboard* ini memiliki enam area yang memiliki kegunaan masing-masing antara lain

visualisasi data transaksi, keuntungan, negara pelanggan, tren penjualan bulanan, produk terlaris, dan kategori produk terlaris.

DAFTAR PUSTAKA

- Adityo, R. R. (2022). Pengembangan Aplikasi Qiscus Meet Menggunakan Jitsi Open Source Framework.
- Coursera. (2023, November 30). What Is a Data Warehouse? Definition, Concepts, and Benefits.
- Knight, M. (2023, June 23). What Is a Data Warehouse?
- Kurniawan, J., Hartoto, Fahmi, A. Z., Ahyani, H., Hikmah, Ridwan, M., Amane, A. P. O., Afnarius, S.,
- Priyanda, R., Arnita, Yudawisastra, H. G., Rosmawati, A., & Hozairi. (2023). Analisis dan Visualisasi Data.
- Muhammad Ariffudin (2021). Apa itu PostgreSQL? Mengenal Database PostgreSQL.
- Postgres. (n.d.). What is PostgreSQL?
- Puri, D. (2021, July 21). Mengenal apa itu ETL (Extract, transform, dan load) dan Toolsnya .
- Ramya, S., & Prasad, S. (2021). The Evolution and Impact of OLAP Technology in Data Analysis.
- Saraswati, N. W. S., & Martarini, N. M. L. (2020). Extract Transform Loading Data Absensi Stmik Stikom Indonesia Menggunakan Pentaho. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 19(2), 273–281. <https://doi.org/10.30812/matrik.v19i2.564>
- Talaoui, Y., & Kohtamäki, M. (2021). 35 years of research on business intelligence process: a synthesis of a fragmented literature.