

## TUGAS AKHIR

# ANALISA DAN PERAMALAN DATA PENUMPANG TRANSJAKARTA DENGAN ARIMA SERTA VISUALISASI DENGAN METODE FOUR-STEP KIMBALL MENGGUNAKAN POWER BI

065002100021 - ALDIS TAMARA PUTRI ISKANDAR

# DAFTAR ISI

## **Bab I:**

Latar Belakang  
Rumusan Masalah  
Batasan Masalah  
Tujuan Penelitian  
Manfaat Penelitian

## **Bab IV:**

Perolehan Raw Data  
Pre-processing  
Perancangan OLAP  
Proses ETL  
Analisa Prediksi  
Visualisasi Data

## **Bab II:**

Tinjauan Pustaka  
Penelitian Terdahulu

## **Bab V:**

Kesimpulan  
Saran

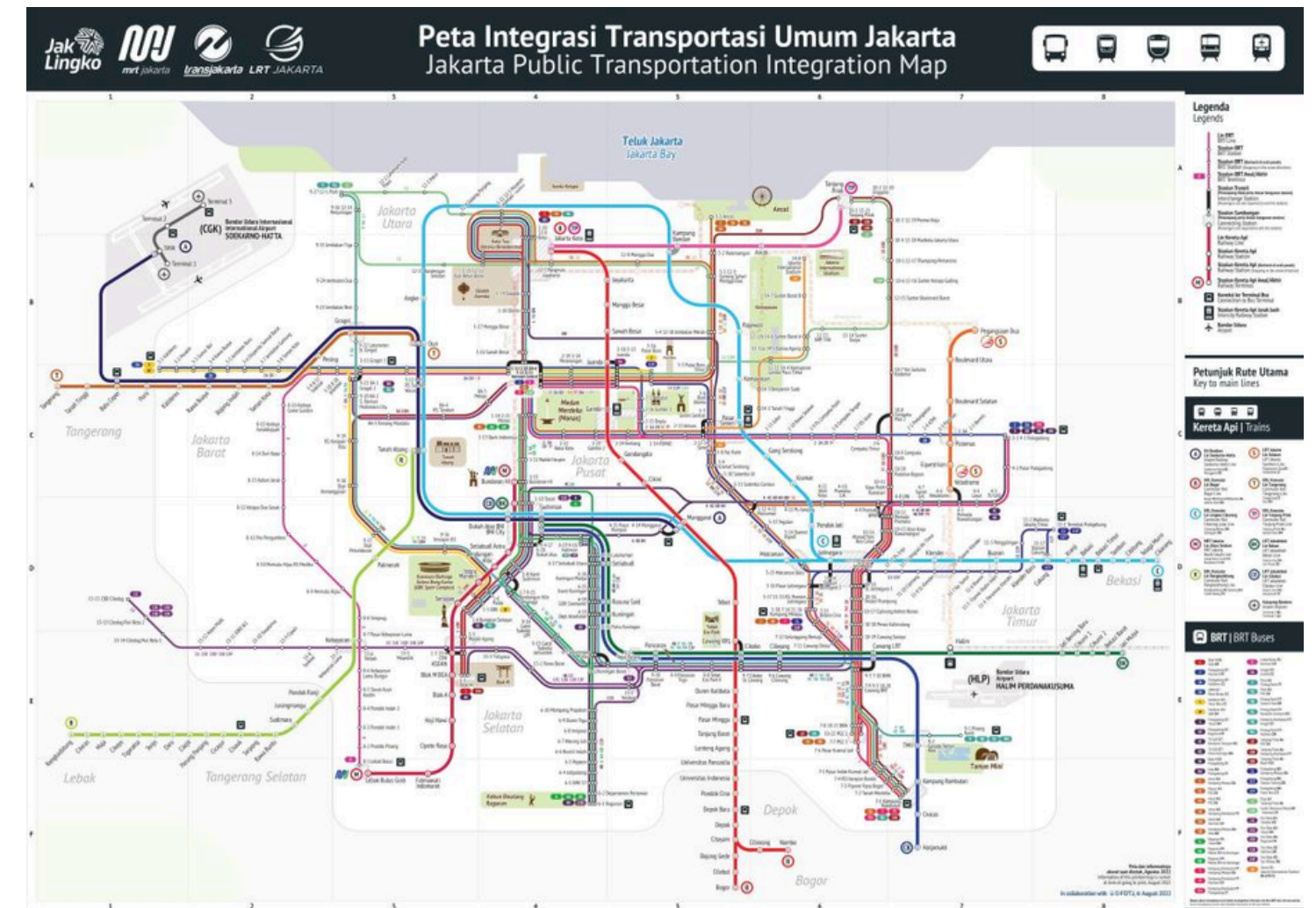
## **Bab III:**

Metodologi Penelitian  
Studi Literatur & Observasi  
Perolehan Raw Data  
Analisa Kebutuhan  
Pre-processing  
Perancangan OLAP  
Proses ETL  
Analisa Prediksi  
Pembuatan Visualisasi

# LATAR BELAKANG

DKI Jakarta, meskipun provinsi terkecil, memiliki tingkat kepadatan penduduk yang tinggi dan menjadi pusat ekonomi, sosial, dan pendidikan di Indonesia. Perkembangan pesat wilayah ini mendorong kebutuhan akan transportasi umum yang aman dan nyaman.

Transjakarta, sebagai penyedia layanan Bus Rapid Transit (BRT) dan non-BRT sejak 2004, kini melayani wilayah Jabodetabek dengan jaringan sepanjang 251,2 km, 273 halte, dan 14 koridor. Pada 2023, jumlah penumpang harian Transjakarta mencapai 1.174.098, dengan armada 4.348 bus, yang ditargetkan bertambah menjadi 4.728 pada 2024.



Sumber: Transjakarta

# LATAR BELAKANG

- Transjakarta merupakan salah satu moda **transportasi publik utama** di Jakarta.
- Setiap tahun, **jumlah penumpang** Transjakarta mengalami **fluktuasi** akibat berbagai faktor seperti kebijakan transportasi, pandemi, dan perubahan pola mobilitas masyarakat.
- Data penumpang Transjakarta dapat digunakan untuk menganalisis **tren penggunaan transportasi**.
- **Analisis data** yang efektif dapat membantu dalam pengambilan keputusan strategis, seperti **perencanaan rute** dan **penambahan armada di rute tertentu**.
- Memanfaatkan **Power BI** untuk menampilkan data dalam bentuk visualisasi interaktif, dan **ARIMA** untuk memprediksi penumpang sehingga memudahkan pengambilan keputusan berbasis data.
- Menghasilkan **dashboard interaktif** yang menampilkan tren jumlah penumpang, distribusi berdasarkan jenis trayek, serta analisis lainnya untuk mendukung evaluasi layanan Transjakarta.

# RUMUSAN MASALAH



1

Bagaimana pengelola Transjakarta menghadapi kendala dalam memperoleh informasi perkembangan jumlah penumpang per trayek dengan cepat untuk pengaturan rute armada?



2

Bagaimana mengatur data sehingga bisa dihasilkan analisis prediksi dan visualisasi interaktif?

# BATASAN MASALAH

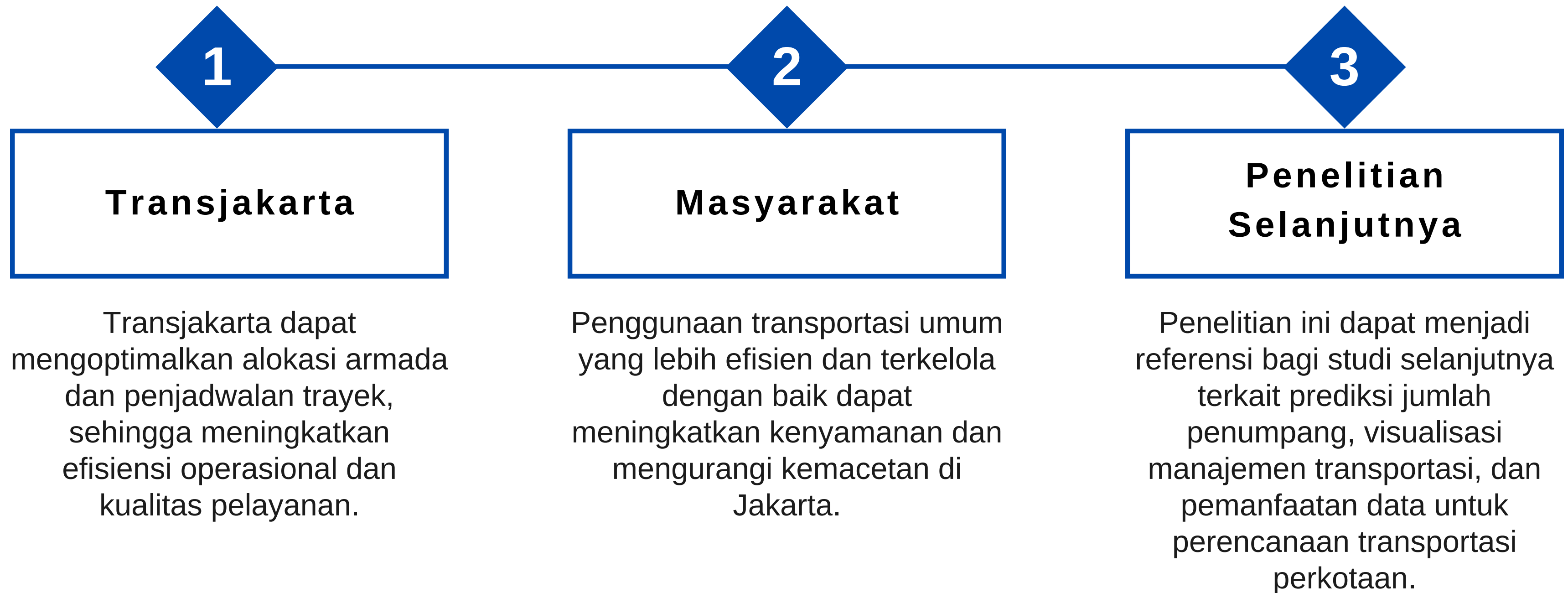
1. Penelitian ini hanya akan memfokuskan pada data penumpang Transjakarta pada tahun **2020-2024**.
2. Data yang digunakan adalah data yang berkaitan dengan Transjakarta seperti, **trayek, kode trayek, jenis dan jumlah penumpang**.
3. Data yang dianalisis berasal dari **Transjakarta** tanpa melibatkan moda transportasi umum lainnya.
4. Visualisasi yang dirancang akan dibatasi pada **data penumpang, trayek, dan jenis trayek**.
5. Penelitian ini **tidak** mencakup **analisis kinerja finansial** atau **operasional lainnya** dari Transjakarta.
6. Analisis Prediksi hanya menggunakan **1 sample** dari **kode trayek**.



# TUJUAN PENELITIAN

Penelitian ini bertujuan merancang visualisasi data untuk memprediksi jumlah penumpang Transjakarta dengan metode *Four Step Kimball*. Tujuannya adalah membantu manajemen memahami tren penggunaan, mengidentifikasi rute padat, dan memperkirakan kebutuhan armada. Dengan teknologi seperti OLAP, Pentaho, Power BI, dan ARIMA, penelitian ini mendukung pengambilan keputusan strategis di sektor transportasi umum DKI Jakarta.

# MANFAAT PENELITIAN





# BAB II

## PENELITI TERDAHULU

Penelitian	Metode Penelitian	Sumber Data	Hasil Pembahasan
"Analisis Business Intelligence Data Penjualan Elektronik 2014 - 2017"	Nine Step Kimball dan Power BI	Penjualan elektronik dari periode 2014 hingga 2017	Peningkatan penjualan produk elektronik antara 2014-2017 dengan pola musiman.
"The Implementation of Business Intelligence to Analyze Sales Trends in the Indofishing Online Store Using Power BI "	Kimball's Nine Steps dan Power BI	Indofishing Online Store	Dashboard visualisasi ini memungkinkan Indofishing untuk memahami tren permintaan produk, termasuk kota asal pembeli yang terbanyak.
"Implementation of Business Intelligence In Analyzing Data Using Tableau at PT Global Bintang Permata"	Nine-step Technology Ralph Kimball dan Tableau	Transaction data from January 2021 to December 2021 and marketing data on PT Global Bintang Permata.	Dashboard ini menyajikan data penjualan dan transaksi pemasaran pada tahun 2021.
"Perancangan Sistem Data Warehouse menggunakan Four-step Methodology Kimball Dengan Talend Open Studio Dan Posgre SQL"	Nine-step Methodology dan Tableau	Perusahaan Bank X	Merancang Data Warehouse terpusat memungkinkan bank mengatasi perbedaan data antar cabang.
"Penerapan Model Seasonal Autoregressive Integrated Moving Average (SARIMA) pada Jumlah Penumpang Kereta Api di Sumatera Barat"	Metode Seasonal Autoregressive Integrated Moving Average (SARIMA)	Jumlah Penumpang Kereta Api di Sumatera Barat	SARIMA(0,1,1)(0,1,0) menghasilkan model terbaik dengan AIC terkecil dan memenuhi uji signifikansi.

# TINJAUAN PUSTAKA

## ***DATA WAREHOUSE***

Data warehouse adalah gudang data yang mendukung pengambilan keputusan dengan menyimpan data terintegrasi, lengkap, dan cepat diakses.

## **ETL**

ETL (Extract, Transform, Load) adalah proses mengintegrasikan data dari berbagai sumber, mengubahnya sesuai kebutuhan, dan menyimpannya dalam format konsisten di data warehouse.

## ***OLAP***

OLAP adalah teknologi analisis data multidimensi untuk mendukung keputusan strategis dan memahami tren bisnis.

## ***FOUR-STEP DESIGN PROCESS***

### ***METHODOLOGY KIMBALL***

Metode Kimball adalah pendekatan bottom-up untuk Data Warehouse dengan pemodelan dimensi, meliputi langkah identifikasi proses bisnis, grain, dimensi, dan fakta.

# TINJAUAN PUSTAKA

## ***PENTAHO DATA INTEGRATION***

Pentaho Data Integration (PDI) adalah perangkat lunak ETL untuk mengumpulkan, membersihkan, dan memuat data ke sistem target, mendukung integrasi dan pemrosesan data besar.

## ***ANALISIS TIME SERIES***

Analisis Time Series memprediksi pola data waktu, seperti tren dan musiman, untuk mendukung proyeksi pasar, pendapatan, dan deteksi anomali. Model ARIMA dan SARIMA memanfaatkan data historis untuk perencanaan berbasis data.

## ***ARIMA***

Model ARIMA memadukan autoregressive (AR), differencing (I), dan moving average (MA) untuk menganalisis data time series. Cocok untuk data dengan tren sederhana, ARIMA digunakan dalam peramalan penjualan, analisis risiko, dan prediksi penumpang.

# TINJAUAN PUSTAKA

## POWER BI

Power BI mengubah data menjadi wawasan terintegrasi untuk analisis dan pelaporan dengan visualisasi dari berbagai sumber data.

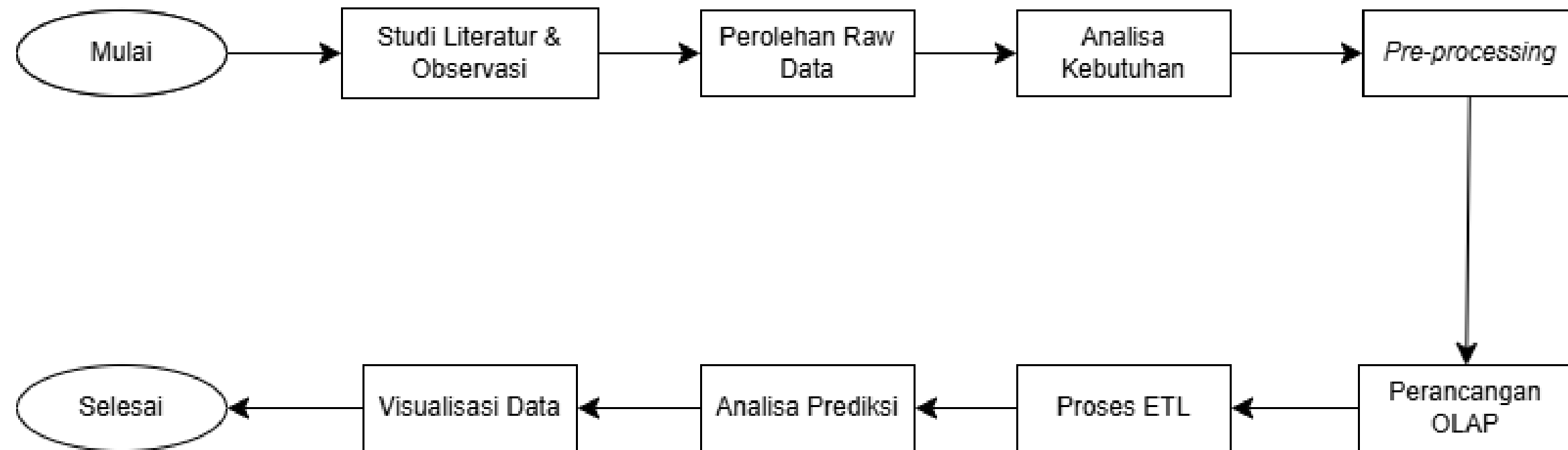
## ***VISUALISASI DATA***

Visualisasi data menyajikan data secara grafis untuk memudahkan pemahaman, mengungkap pola, dan membantu pengambilan keputusan.

## ***SARIMA***

SARIMA (Seasonal ARIMA) mengembangkan ARIMA dengan menambahkan komponen musiman: Seasonal AR, Differencing, dan MA (P, D, Q), serta periode musiman (s). Model ini unggul dalam menangkap pola musiman, seperti fluktuasi jumlah penumpang.

# METODOLOGI PENELITIAN



# ANALISA KEBUTUHAN

Kebutuhan Tampilan	Visualisasi yang digunakan	Tujuan Visualisasi
Trend Jumlah Penumpang per Bulan atau Tahun	Line chart	Menunjukkan pola perubahan jumlah penumpang dari waktu ke waktu. Membantu mengidentifikasi tren dan pola musiman.
Distribusi Penumpang Berdasarkan Jenis	Pie chart	Menampilkan proporsi atau persentase kontribusi dari masing-masing jenis penumpang terhadap total.
Distribusi Penumpang Berdasarkan Tahun	Donut chart	Menampilkan proporsi atau persentase kontribusi dari masing-masing tahun terhadap total.
Analisis Jenis Trayek per Bulan	Ribbon chart	Memvisualisasikan perubahan jumlah penumpang berdasarkan jenis trayek setiap bulannya, sekaligus menunjukkan dominasi atau perbandingan antar jenis trayek.
Analisis Quartal dan Jumlah Penumpang	Stacked column chart	Menunjukkan kontribusi jumlah penumpang dari berbagai kategori atau periode dalam setiap kuartal
Top 5 Rata-rata Penumpang tertinggi	Clustered bar chart	Mengidentifikasi trayek atau jenis penumpang dengan rata-rata tertinggi untuk fokus pada performa terbaik.



# ANALISA KEBUTUHAN

Kebutuhan Tampilan	Visualisasi yang digunakan	Tujuan Visualisasi
Top 5 Rata-rata Penumpang terendah	Clustered bar chart	Mengidentifikasi trayek atau jenis penumpang dengan performa terendah untuk mengevaluasi potensi perbaikan.
Filter Tanggal	Slicer	Memungkinkan untuk melihat data berdasarkan rentang tanggal tertentu untuk analisis waktu yang lebih spesifik.
Filter Tahun	Slicer	Mempermudah pemfilteran data berdasarkan tahun tertentu untuk analisis tahunan.
Filter Bulan	Slicer	Mempermudah pemfilteran data berdasarkan bulan tertentu untuk analisis bulanan.
Filter Jenis Trayek	Slicer	Memberikan fleksibilitas untuk memfilter data berdasarkan jenis trayek yang diinginkan.
Filter Trayek	Slicer	Memungkinkan analisis mendetail berdasarkan trayek tertentu.
Jumlah Penumpang	Card	Memberikan gambaran cepat tentang total jumlah penumpang.
Jenis Trayek	Card	Menyajikan informasi ringkas tentang jenis trayek yang tersedia atau sedang dianalisis.
Trayek	Card	Memberikan informasi mengenai trayek spesifik yang menjadi fokus analisis.

# PRE-PROCESSING

Pre-processing dimulai dengan menggabungkan data penumpang 2020-2023 dan 2024 menggunakan `pd.concat`. Selanjutnya, baris dengan jumlah penumpang 0 dihapus. Dataset hasil filter disimpan sebagai file CSV untuk analisis lanjutan.

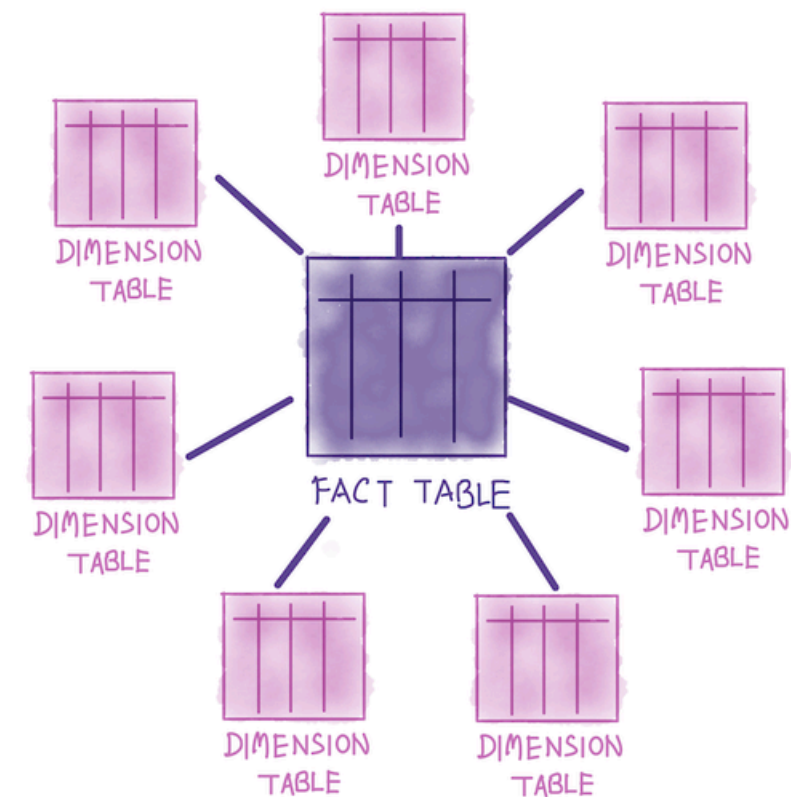
# PERANCANGAN OLAP

## FOUR-STEP PROCESS METHODOLOGY KIMBALL

Setelah menggabungkan data, struktur tabel data warehouse dirancang dengan model *Star Schema*, mencakup:

3 tabel dimensi: waktu, trayek, jenis

1 tabel fakta: jumlah penumpang



# PERANCANGAN OLAP

## *FOUR-STEP METHODOLOGY KIMBALL*

### ***IDENTIFY THE BUSINESS PROCESS***

Proses bisnis yang diidentifikasi adalah analisis jumlah penumpang Transjakarta dan kinerja trayek dari tahun 2020 hingga 2024. Tujuan dari proses ini adalah pada pola perjalanan penumpang, volume penumpang per trayek, serta penggunaan layanan yang paling padat berdasarkan waktu bulan.

### ***DECLARE THE GRAIN***

Grain atau tingkat granularitas data dalam penelitian ini ditetapkan pada tingkat jumlah penumpang yang tercatat untuk setiap trayek tertentu berdasarkan jenis transportasi tertentu pada periode waktu. Dengan demikian, setiap baris data pada tabel fakta merepresentasikan jumlah total penumpang yang menggunakan trayek tertentu dengan jenis transportasi spesifik selama satu bulan tertentu, yang memungkinkan analisis performa trayek dan pola penggunaan transportasi dalam rentang waktu tersebut.

- Informasi Jumlah Penumpang
- Informasi Jenis Trayek
- Informasi Kode Trayek

# PERANCANGAN OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL DIMENSI WAKTU

No	Nama	Tipe Data	Panjang
1	sk	Integer	11
2	tanggal	Date	
3	tahun	Integer	11
4	quarter	Char	2
5	bulan	Integer	11
6	bulan_nama	Varchar	50
7	hari	Integer	11

TABEL DIMENSI JENIS TRAYEK

No	Nama	Tipe Data	Panjang
1	jenis	Varchar	100
2	jenisID	Integer	100

# PERANCANGAN OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL DIMENSI TRAYEK

No	Nama	Tipe Data	Panjang
1	kode_trayek	Varchar	255
2	trayek	Varchar	255
3	trayekID	Integer	11

TABEL FAKTA PENUMPANG

No	Nama	Tipe Data	Panjang
1	sk	Integer	11
2	jenisID	Integer	100
3	trayekID	Integer	100
4	jumlah_penu mpang	Integer	255



# PROSES ETL

Pada tahap ETL, data yang telah dinormalisasi diolah menggunakan Spoon dari Pentaho Data Integration dan MySQL di XAMPP melalui proses *extract, transform, dan load*.

# ANALISA PREDIKSI

Analisa prediksi dilakukan dengan memproses data, mengubah kolom tanggal menjadi indeks, dan memvisualisasikan pola awal. **Stasioneritas** diuji menggunakan **Dickey-Fuller Test**, dengan **differencing** jika diperlukan. Model **ARIMA** dan **SARIMA** digunakan untuk memprediksi jumlah penumpang, diikuti **visualisasi hasil**. Prediksi digabungkan dengan data asli dan disimpan dalam file Excel untuk analisis lebih lanjut di Power BI.

# PEMBUATAN VISUALISASI

Pada tahap akhir, visualisasi data penumpang Transjakarta menggunakan Power BI menampilkan grafik tren jumlah penumpang berdasarkan bulan dan tahun, analisis jenis rute berdasarkan jumlah penumpang dan bulan, distribusi penumpang berdasarkan jenis trayek, analisis kuartal dan jumlah penumpang, top lima rata-rata penumpang tertinggi dan terendah berdasarkan trayek, serta card yang menampilkan jumlah total penumpang, jumlah jenis trayek, dan jumlah trayek.

# PEROLEHAN RAW DATA

tahun	bulan	kode_trayek	trayek	jenis	jumlah_penumpang
2020	Jan	1	Blok M - Kota	BRT	2539866
2020	Jan	2	Pulo Gadung 1 - Harmoni	BRT	815240
2020	Jan	3	Kalideres - Pasar Baru	BRT	1059250
2020	Jan	4	Pulo Gadung 2 - Tosari	BRT	774443
2020	Jan	5	Kampung Melayu - Ancol	BRT	
2020	Jan	6	Ragunan - Dukuh Atas 2	BRT	
2020	Jan	7	Kampung Rambutan - Kampung Melayu	BRT	
2020	Jan	8	Lebak Bulus - Harmoni	BRT	
2020	Jan	9	Pinang Ranti - Pluit	BRT	
2020	Jan	10	Tanjung Priok - PGC 2	BRT	
2020	Jan	11	Pulo Gebang - Kampung Melayu	BRT	
2020	Jan	12	Tanjung Priok - Pluit	BRT	
2020	Jan	13	Ciledug - Tendean	BRT	
2020	Jan	10A	Rusun Marunda - Tanjung Priok	Rusun	
2020	Jan	10B	Rusun Cipinang Besar Selatan - Penas Kalimalang	Rusun	
2020	Jan	10C	Pelabuhan Tanjung Priok - Tanjung Priok	Angkutan Umum Integrasi	
2020	Jan	10F	Sunter Kelapa Gading - Stasiun LRT Pegangsaan Dua	Angkutan Umum Integrasi	
2020	Jan	10K	Tanjung Priok - Senen via Taman BMW	Angkutan Umum Integrasi	
2020	Jan	11A	Pulo Gebang - Rawamangun	Angkutan Umum Integrasi	
2020	Jan	11B	Rusun Rawas Bebek - Penggilingan	Rusun	
2020	Jan	11C	Rusun Pinus Elok - Rusun Pulo Gebang	Rusun	
2020	Jan	11D	Pulo Gebang - Pulo Gadung 2 via PK	Angkutan Umum Integrasi	
2020	Jan	11K	Rusun Komarudin - Penggilingan	Rusun	
2020	Jan	11M	Rusun Rawas Bebek - Bukit Duri	Rusun	
2020	Jan	11N	Rusun Cipinang Muka - Jatenegara	Rusun	
2020	Jan	11P	Rusun Pondok Bambu - Walkota Jakarta Timur	Rusun	
2020	Jan	11Q	Kampung Melayu - Pulo Gebang via BKT	Angkutan Umum Integrasi	
2020	Jan	11R	Rusun Cakung KM2 - Bukit Duri	Rusun	
2020	Jan	11U	Pulo Gebang - Stasiun Cakung via Cakung Cilincing	Angkutan Umum Integrasi	
2020	Jan	12A	Pelabuhan Kaliadem - Kota	Angkutan Umum Integrasi	
2020	Jan	12B	Pluit - Senen	Angkutan Umum Integrasi	
2020	Jan	12C	Rusun Waduk Pluit - Penjarangan	Rusun	
2020	Jan	12D	Rusun Sukajaya - Sunter via Kelapa Gading	Rusun	
2020	Jan	12F	Rusun Marunda - Rusun Waduk Pluit	Rusun	
2020	Jan	12H	Rusun Penjarangan - Penjarangan	Rusun	
2020	Jan	1A	Pantai Maju - Balai Kota	Angkutan Umum Integrasi	
2020	Jan	1B	Stasiun Palmerah - Tosari	Angkutan Umum Integrasi	
2020	Jan	1C	Pesangrahan - Blok M	Angkutan Umum Integrasi	
2020	Jan	1E	Pondok Labu - Blok M	Angkutan Umum Integrasi	
2020	Jan	1F	Stasiun Palmerah - Bundaran Senayan	Angkutan Umum Integrasi	

DATA PENUMPANG TRANSJAKARTA

jenis	kode_trayek	trayek
BRT	1	BLOK M - KOTA
BRT	2	PULO GADUNG - MONUMEN NASIONAL
BRT	3	KALIDERES - MONUMEN NASIONAL VIA VETERAN
BRT	4	PULO GADUNG - GALUNGGUNG
BRT	5	KAMPUNG MELAYU - ANCOL
BRT	6	RAGUNAN - GALUNGGUNG
BRT	7	KAMPUNG MELAYU - KAMPUNG RAMBUTAN
BRT	8	LEBAK BULUS - PASAR BARU
BRT	9	PINANG RANTI - PLUIT
BRT	10	TANJUNG PRIOK - PGC
BRT	11	PULO GEBANG - KAMPUNG MELAYU
BRT	12	PLUIT - TANJUNG PRIOK
BRT	13	PURI BETA - TEGAL MAMPANG
BRT	14	JAKARTA INTERNATIONAL STADIUM - SENEN
Transjabodetabek	[14225]	HALTE NON BRT BEKASI BARAT
Transjabodetabek	[14226]	HALTE NON BRT BULAK KAPAL
Transjabodetabek	[14227]	HALTE NON BRT BEKASI TIMUR
Transjabodetabek	[14229]	HALTE NON BRT SUMMARECON BEKASI
Angkutan Umum Integrasi	[14230]	HALTE NON BRT PLAZA STASIUN MANGGARAI
Tidak Diketahui	[14246]	TANAH ABANG
Rusun	10A	RUSUN MARUNDA - TANJUNG PRIOK
Rusun	10B	RUSUN CIPINANG BESAR SELATAN - PENAS KALIMALANG
Angkutan Umum Integrasi	10C	PELABUHAN TANJUNG PRIOK - TANJUNG PRIOK
BRT	10D	TANJUNG PRIOK - KP RAMBUTAN
Angkutan Umum Integrasi	10E	PULO GEBANG - TANJUNG PRIUK
Angkutan Umum Integrasi	10F	SUNTER KELAPA GADING - ST.LRT PEGANGSAAN DUA

DATA RUTE TRANSJAKARTA

# PRE-PROCESSING

```
import pandas as pd
```

```
# Data from the two datasets
```

```
data = pd.read_excel('Data Penumpang_aldis.xlsx')
```

```
data_2024 = pd.read_excel('Penumpang 2024.xlsx')
```

```
# Creating dataframes
```

```
df_data = pd.DataFrame(data)
```

```
df_2024 = pd.DataFrame(data_2024)
```

```
# Merging both dataframes
```

```
merged_df = pd.concat([df_data, df_2024], ignore_index=True)
```

```
df_filtered = merged_df[merged_df["jumlah_penumpang"] != 0].reset_index(drop=True)
```

```
print(df_filtered)
```

```
df_filtered.info()
```

```
# Save to Excel file with proper filename
```

```
file_path_corrected = 'G:/My Drive/Universitas Trisakti/Skripsi/DATA BARU/Penumpang.csv'
```

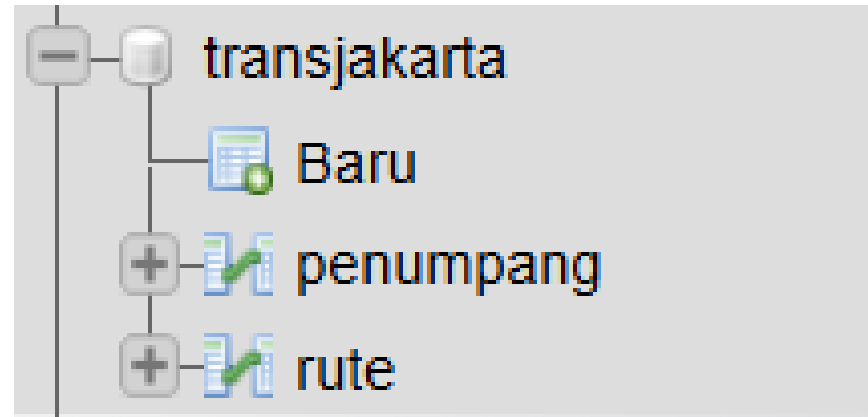
```
df_filtered.to_csv(file_path_corrected, index=False)
```

```
file_path_corrected
```



# PROSES ETL DATABASE OLAP

## RAW DATABASE



## STRUKTUR RAW DATABASE TABLE PENUMPANG

Server: 127.0.0.1 » Database: transjakarta » Tabel: penumpang

Jelajahi Struktur SQL Cari Tambahkan Ekspor Impor Hak Akses Operasi

Struktur tabel Tampilan hubungan

#	Nama	Jenis	Penyortiran	Atribut	Tak Ternilai	Bawaan	Komentar	Ekstra	Tindakan
<input type="checkbox"/> 1	tahun	int(4)			Ya	NULL			Ubah Hapus Lainnya
<input type="checkbox"/> 2	bulan	varchar(3)	utf8_general_ci		Ya	NULL			Ubah Hapus Lainnya
<input type="checkbox"/> 3	kode_trayek	varchar(12)	utf8_general_ci		Ya	NULL			Ubah Hapus Lainnya
<input type="checkbox"/> 4	trayek	varchar(53)	utf8_general_ci		Ya	NULL			Ubah Hapus Lainnya
<input type="checkbox"/> 5	jenis	varchar(23)	utf8_general_ci		Ya	NULL			Ubah Hapus Lainnya
<input type="checkbox"/> 6	jumlah_penumpang	int(7)			Ya	NULL			Ubah Hapus Lainnya

## STRUKTUR RAW DATABASE TABLE RUTE

Server: 127.0.0.1 » Database: transjakarta » Tabel: rute

Jelajahi Struktur SQL Cari Tambahkan Ekspor Impor Hak Akses O

Struktur tabel Tampilan hubungan

#	Nama	Jenis	Penyortiran	Atribut	Tak Ternilai	Bawaan	Komentar	Ekstra	Tindakan
<input type="checkbox"/> 1	jenis	varchar(27)	utf8_general_ci		Ya	NULL			Ubah Hapus Lainnya
<input type="checkbox"/> 2	kode_trayek	varchar(12)	utf8_general_ci		Ya	NULL			Ubah Hapus Lainnya
<input type="checkbox"/> 3	trayek	varchar(52)	utf8_general_ci		Ya	NULL			Ubah Hapus Lainnya



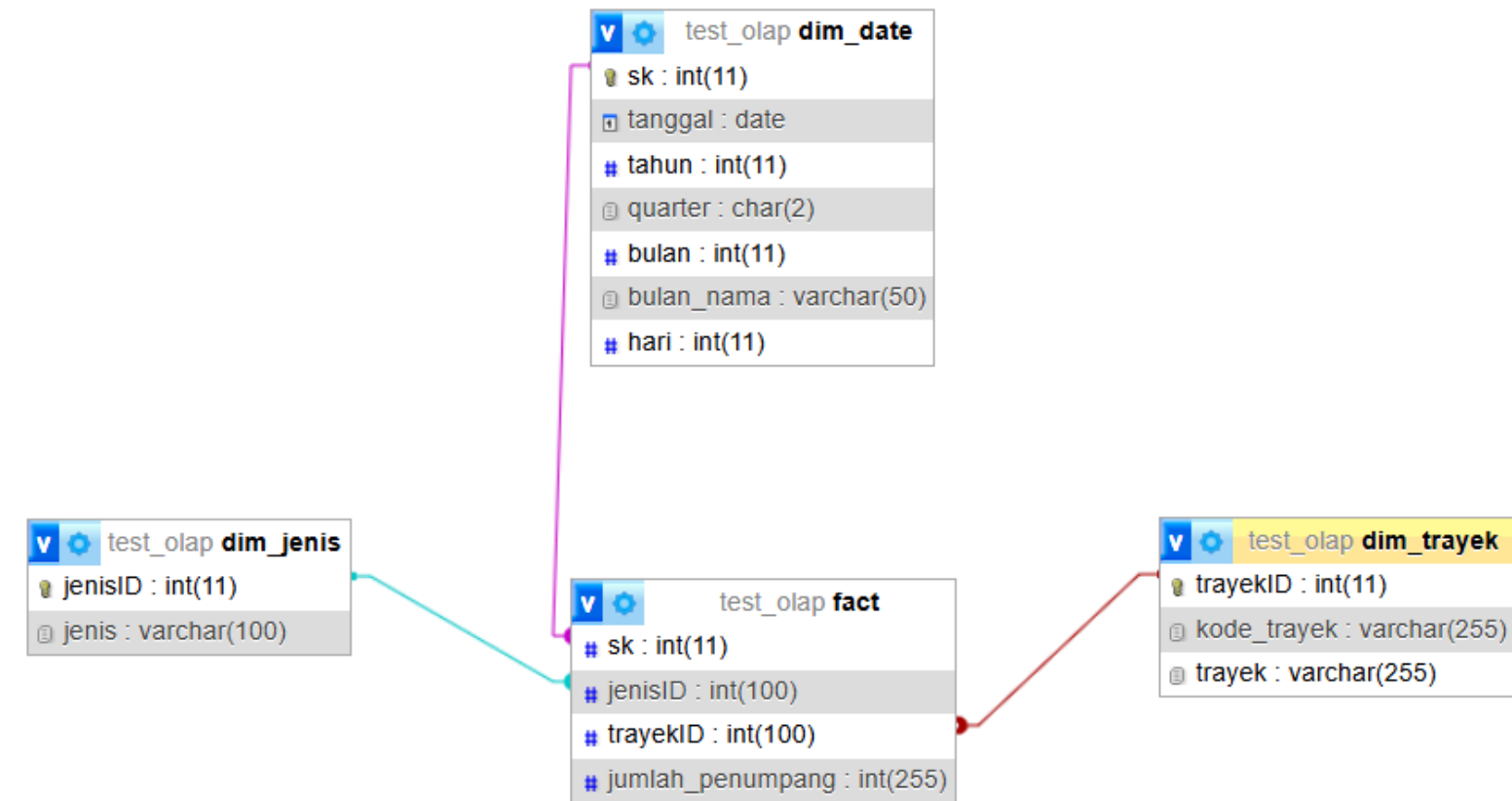
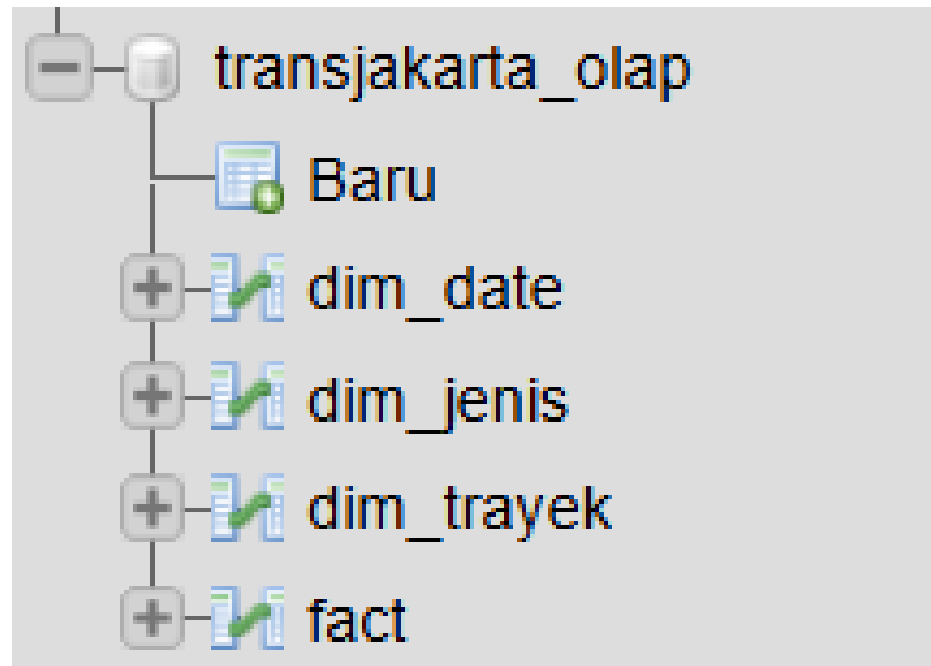
# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

MYSQL DATABASE

## STAR SCHEMA

### DATABASE OLAP



# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL DIMENSI - DIM DATE

MYSQL DATABASE

## STRUKTUR TABEL DIM DATE

Server: 127.0.0.1 » Database: transjakarta_olap » Tabel: dim_date									
Jelajahi Struktur SQL Cari Tambahkan Ekspor Impor Hak Akses Operasi Pelacak									
Struktur tabel Tampilan hubungan									
#	Nama	Jenis	Penyortiran	Atribut	Tak Ternilai	Bawaan	Komentar	Ekstra	Tindakan
<input type="checkbox"/>	1 sk	int(11)			Tidak	Tidak ada		AUTO_INCREMENT	Ubah Hapus Lainnya
<input type="checkbox"/>	2 tanggal	date			Ya	NULL			Ubah Hapus Lainnya
<input type="checkbox"/>	3 tahun	int(11)			Ya	NULL			Ubah Hapus Lainnya
<input type="checkbox"/>	4 quarter	char(2)	utf8mb4_general_ci		Ya	NULL			Ubah Hapus Lainnya
<input type="checkbox"/>	5 bulan	int(11)			Ya	NULL			Ubah Hapus Lainnya
<input type="checkbox"/>	6 bulan_nama	varchar(50)	utf8mb4_general_ci		Ya	NULL			Ubah Hapus Lainnya
<input type="checkbox"/>	7 hari	int(11)			Ya	NULL			Ubah Hapus Lainnya

## INDEKS TABEL DIM DATE

Indeks									
Tindakan	Nama kunci	Jenis	Unik	Dipadatkan	Kolom	Kardinalitas	Penyortiran	Tak Ternilai	Komentar
Ubah Rename Hapus	PRIMARY	BTREE	Ya	Tidak	sk	2000	A	Tidak	
Ubah Rename Hapus	tahun_key	BTREE	Tidak	Tidak	tahun	12	A	Ya	
Ubah Rename Hapus	quarter_key	BTREE	Tidak	Tidak	quarter	8	A	Ya	
Ubah Rename Hapus	bulan_key	BTREE	Tidak	Tidak	bulan	24	A	Ya	
Ubah Rename Hapus	bulan_nama_key	BTREE	Tidak	Tidak	bulan_nama	24	A	Ya	

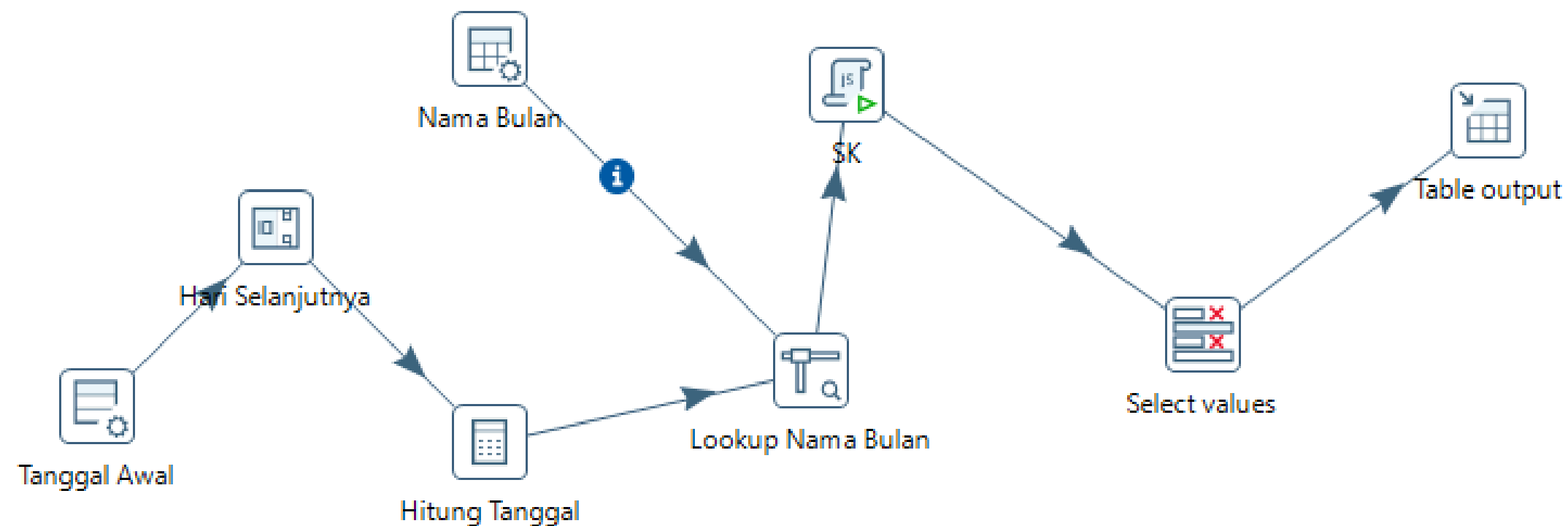
# PROSES ETL DATABASE OLAP

*FOUR-STEP METHODOLOGY KIMBALL*

*TABEL DIMENSI - DIM DATE*

*PENTAHO DATA INTEGRATION*

## *TRANSFORMASI PEMBUATAN TABEL DIMENSI WAKTU*



# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL DIMENSI - DIM DATE

PENTAHO DATA INTEGRATION

## 1.STEP GENERATE ROWS

Generate rows

Step name: Tanggal Awal

Limit: 1827

Never stop generating rows: ☐

Interval in ms (delay): 5000

Current row time field name: now

Previous row time field name: FiveSecondsAgo

Fields:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Value	Set empty string?
1	tanggal_awal	Date	yyyy-MM-dd						2019-12-31	N

Help OK Preview Cancel

## 2.STEP ADD SEQUENCE

Add sequence

Step name: Hari Selanjutnya

Name of value: hari\_selanjutnya

Use a database to generate the sequence

Use DB to get sequence? ☐

Connection: Transjakarta Edit... New... Wizard...

Schema name:  Schemas...

Sequence name: SEQ Sequences...

Use a transformation counter to generate the sequence

Use counter to calculate sequence? ☒

Counter name (optional):

Start at value: 1

Increment by: 1

Maximum value: 999999999

Help OK Cancel

# PROSES ETL DATABASE OLAP

## FOUR-STEP METHODOLOGY KIMBALL

**TABEL DIMENSI - DIM DATE**

## PENTHO DATA INTEGRATION

### 3.STEP CALCULATOR

**Calculator**

Step name

☐ Throw an error on non existing files

Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask	Decimal symbol	Grouping
1	tanggal	Date A + B Days	tanggal_awal	hari_selanjutnya		Date			N	yyyy-MM-dd		
2	tahun	Year of date A	tanggal			Integer			N			
3	quarter_int	Quarter of date A	tanggal			Integer			Y			
4	quarter_constant	Set field to constant value A	Q			String			Y			
5	quarter	A + B	quarter_constant	quarter_int		String			N			
6	bulan	Month of date A	tanggal			Integer			N			
7	hari	Day of month of date A	tanggal			Integer			N			

[Help](#) OK Cancel

# PROSES ETL DATABASE OLAP

## FOUR-STEP METHODOLOGY KIMBALL

### ***TABEL DIMENSI - DIM DATE***

## PENTAHO DATA INTEGRATION

## 4.STEP DATA GRID META

The screenshot shows a 'Data grid' window with the 'Meta' tab selected. The 'Step name' is 'Nama Bulan'. The table contains the following data:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Null if	Set empty string?
1	bulan_angka	Integer								N
2	bulan_nama	String								N

A 'Cancel' button is located at the bottom right of the window.

## 4.STEP DATA GRID DATA

Data grid

Step name

Nama Bulan

Meta

Data

#	bulan_angka	bulan_nama
1	1	Jan
2	2	Feb
3	3	Mar
4	4	Apr
5	5	May
6	6	Jun
7	7	Jul
8	8	Aug
9	9	Sep
10	10	Oct
11	11	Nov
12	12	Dec

Help

OK

Preview

Cancel



# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL DIMENSI - DIM DATE

PENTAHO DATA INTEGRATION

## 5.STEP STREAM LOOKUP

Stream lookup

Step name: Lookup Nama Bulan

Lookup step: Nama Bulan

The key(s) to look up the value(s):

#	Field	LookupField
1	bulan	bulan_angka

Specify the fields to retrieve:

#	Field	New name	Default	Type
1	bulan_nama			String

☒ Preserve memory (costs CPU)  
☐ Key and value are exactly one integer field  
☒ Use sorted list (i.s.o. hashtable)

? Help OK Cancel Get Fields Get lookup fields

## 6.STEP MODIFIED JAVASCRIPT VALUE

Modified JavaScript value

Step name: SK

Java script functions:

- Transform Scripts
- Transform Constant
- Transform Function
- Input fields
  - tanggal\_awal
  - hari\_selanjutnya
  - tanggal
  - tahun
  - quarter
  - bulan
  - hari
  - bulan\_nama
- Output fields

Java script:

```
//Script here
var sk;
sk = (tahun * 10000) + (bulan * 100) + hari;
```

Linenr: 0  
 Compatibility mode? ☐ Optimization level 9

Fields

#	Fieldname	Rename to	Type	Length	Precision	Replace value 'Fieldname' or 'Rename to'
1	sk		Integer			N

? Help OK Cancel Get variables Test script

# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL DIMENSI - DIM DATE

PENTAHO DATA INTEGRATION

## 7.STEP SELECT VALUES

Step name: **Select values**

Select & Alter Remove Meta-data

Fields:

#	Fieldname	Rename to	Length
1	tanggal		
2	tahun		
3	quarter		
4	bulan		
5	hari		
6	bulan_nama		
7	sk		

Get fields to select Edit Mapping

Include unspecified fields, ☐

Help OK Cancel

## 8.STEP TABLE OUTPUT

Step name: **Table output**

Connection: **Test** Edit... New... Wizard...

Target schema: **test\_olap** Browse...

Target table: **dim\_date** Browse...

Commit size: **1827**

Truncate table ☒

Ignore insert errors ☐

Specify database fields ☐

Main options Database fields

Partition data over tables ☐

Partitioning field:

Partition data per month ☒

Partition data per day ☐

Use batch update for inserts ☒

Is the name of the table defined in a field? ☐

Field that contains name of table:

Store the tablename field ☒

Return auto-generated key ☐

Name of auto-generated key field:

Help OK Cancel SQL

# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL DIMENSI - DIM DATE

MYSQL DATABASE

DIM DATE

←T→						sk	tanggal	tahun	quarter	bulan	bulan_nama	hari
<input type="checkbox"/>		Ubah		Salin		Hapus	20200101	2020-01-01	2020	Q1	1 Jan	1
<input type="checkbox"/>		Ubah		Salin		Hapus	20200102	2020-01-02	2020	Q1	1 Jan	2
<input type="checkbox"/>		Ubah		Salin		Hapus	20200103	2020-01-03	2020	Q1	1 Jan	3
<input type="checkbox"/>		Ubah		Salin		Hapus	20200104	2020-01-04	2020	Q1	1 Jan	4
<input type="checkbox"/>		Ubah		Salin		Hapus	20200105	2020-01-05	2020	Q1	1 Jan	5
<input type="checkbox"/>		Ubah		Salin		Hapus	20200106	2020-01-06	2020	Q1	1 Jan	6
<input type="checkbox"/>		Ubah		Salin		Hapus	20200107	2020-01-07	2020	Q1	1 Jan	7
<input type="checkbox"/>		Ubah		Salin		Hapus	20200108	2020-01-08	2020	Q1	1 Jan	8
<input type="checkbox"/>		Ubah		Salin		Hapus	20200109	2020-01-09	2020	Q1	1 Jan	9
<input type="checkbox"/>		Ubah		Salin		Hapus	20200110	2020-01-10	2020	Q1	1 Jan	10
<input type="checkbox"/>		Ubah		Salin		Hapus	20200111	2020-01-11	2020	Q1	1 Jan	11
<input type="checkbox"/>		Ubah		Salin		Hapus	20200112	2020-01-12	2020	Q1	1 Jan	12
<input type="checkbox"/>		Ubah		Salin		Hapus	20200113	2020-01-13	2020	Q1	1 Jan	13
<input type="checkbox"/>		Ubah		Salin		Hapus	20200114	2020-01-14	2020	Q1	1 Jan	14
<input type="checkbox"/>		Ubah		Salin		Hapus	20200115	2020-01-15	2020	Q1	1 Jan	15
<input type="checkbox"/>		Ubah		Salin		Hapus	20200116	2020-01-16	2020	Q1	1 Jan	16
<input type="checkbox"/>		Ubah		Salin		Hapus	20200117	2020-01-17	2020	Q1	1 Jan	17
<input type="checkbox"/>		Ubah		Salin		Hapus	20200118	2020-01-18	2020	Q1	1 Jan	18
<input type="checkbox"/>		Ubah		Salin		Hapus	20200119	2020-01-19	2020	Q1	1 Jan	19

# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL DIMENSI - DIM JENIS

MYSQL DATABASE

## STRUKTUR TABEL DIM JENIS

Server: 127.0.0.1 » Database: transjakarta\_olap » Tabel: dim\_jenis

JelajahiStrukturSQLCariTambahkanEksporImporHak AksesOperasiPela

Struktur tabelTampilan hubungan

#	Nama	Jenis	Penyortiran	Atribut	Tak Ternilai	Bawaan	Komentar	Ekstra	Tindakan
<input type="checkbox"/>	1 jenis	varchar(100)	utf8mb4_general_ci		Tidak	Tidak ada			Ubah  Hapus Lainnya
<input type="checkbox"/>	2 jenisID	int(11)			Tidak	Tidak ada		AUTO_INCREMENT	Ubah  Hapus Lainnya

## INDEKS TABEL DIM JENIS

Indeks ?

Tindakan	Nama kunci	Jenis	Unik	Dipadatkan	Kolom	Kardinalitas	Penyortiran	Tak Ternilai	Komentar
Ubah  Rename  Hapus	PRIMARY	BTREE	Ya	Tidak	jenisID	16	A	Tidak	
Ubah  Rename  Hapus	jenis_key	BTREE	Tidak	Tidak	jenis	16	A	Tidak	

# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL DIMENSI - DIM TRAYEK

MYSQL DATABASE

## STRUKTUR TABEL DIM TRAYEK

Server: 127.0.0.1 » Database: transjakarta\_olap » Tabel: dim\_trayek

JelajahiStrukturSQLCariTambahkanEksporImporHak AksesOperasiPelacak

Struktur tabelTampilan hubungan

	#	Nama	Jenis	Penyortiran	Atribut	Tak Ternilai	Bawaan	Komentar	Ekstra	Tindakan
<input type="checkbox"/>	1	kode_trayek	varchar(100)	utf8mb4_general_ci		Tidak	Tidak ada			Ubah  Hapus Lainnya
<input type="checkbox"/>	2	trayek	varchar(100)	utf8mb4_general_ci		Tidak	Tidak ada			Ubah  Hapus Lainnya
<input type="checkbox"/>	3	trayekID	int(11)			Tidak	Tidak ada		AUTO_INCREMENT	Ubah  Hapus Lainnya

## INDEKS TABEL DIM TRAYEK

Indeks

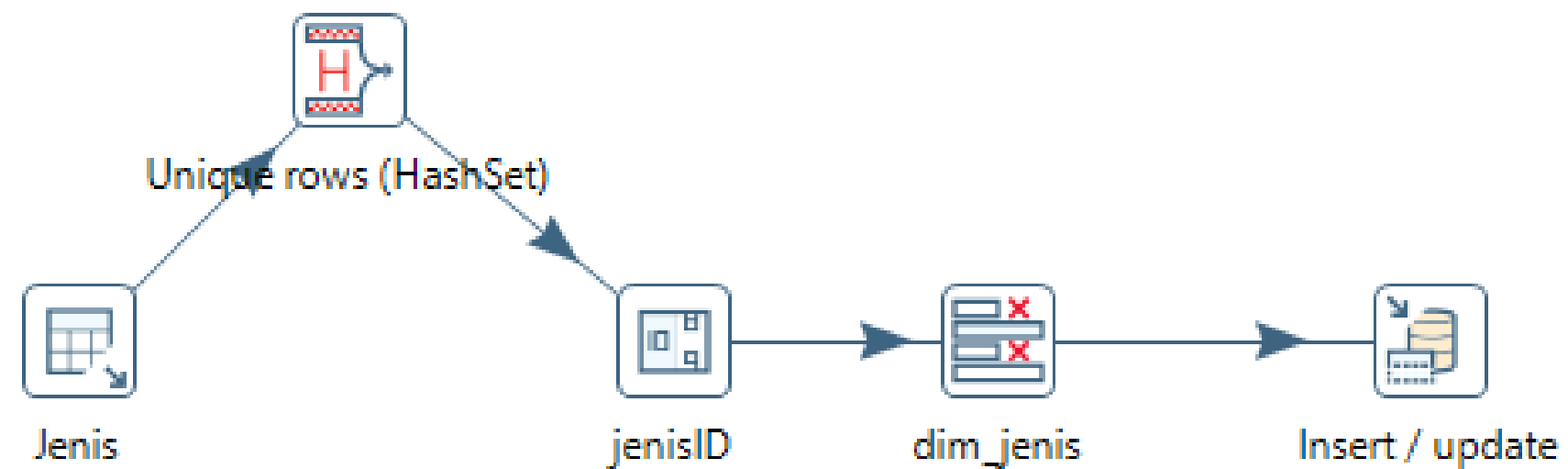
Tindakan	Nama kunci	Jenis	Unik	Dipadatkan	Kolom	Kardinalitas	Penyortiran	Tak Ternilai	Komentar
Ubah  Rename  Hapus	PRIMARY	BTREE	Ya	Tidak	trayekID	457	A	Tidak	
Ubah  Rename  Hapus	kode_trayek_key	BTREE	Tidak	Tidak	kode_trayek	457	A	Tidak	

# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

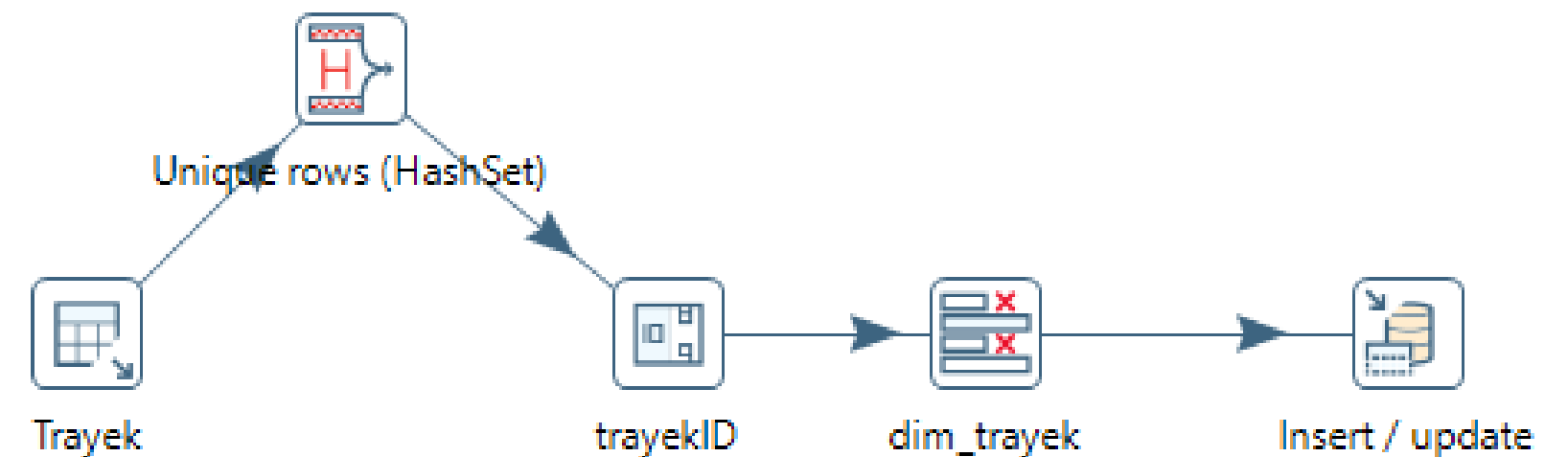
**TABEL DIMENSI - DIM JENIS & TRAYEK**

**PENTAHO DATA INTEGRATION**



## TRANSFORMASI PEMBUATAN TABEL DIMENSI JENIS

## TRANSFORMASI PEMBUATAN TABEL DIMENSI TRAYEK



# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL DIMENSI - DIM JENIS & TRAYEK

PENTAHO DATA INTEGRATION

## 1.STEP TABLE INPUT

The 'Table input' dialog box is shown with the following details:

- Step name:** Trayek
- Connection:** Test
- SQL:**

```
SELECT
jenis
, kode_trayek
, trayek
FROM test.rute
```
- Line 1 Column 0:**
  - Store column info in step meta data: ☐
  - Enable lazy conversion: ☐
  - Replace variables in script: ☒
  - Insert data from step:
  - Execute for each row?: ☐
  - Limit size: 0

Buttons: ? Help, OK, Preview, Cancel

## 2.STEP UNIQUE ROWS (HASHSET)

The 'Unique rows (HashSet)' dialog box is shown with the following details:

- Step name:** Unique rows (HashSet)
- Settings:**
  - Compare using stored row values?: ☐
  - Redirect duplicate row: ☒ Error description:
- Fields to compare (no entries means: compare complete row)**  
Note: values for the selected fields are stored in memory for every record.
- | # | Fieldname   |
|---|-------------|
| 1 | kode_trayek |

Buttons: ? Help, OK, Cancel, Get



**FOUR-STEP METHODOLOGY KIMBALL**  
**TABEL DIMENSI - DIM JENIS & TRAYEK**  
**PENTAHO DATA INTEGRATION**

## 4.STEP SELECT VALUES

Select values

Step name

Select & Alter Remove Meta-data

Fields:

#	Fieldname	Rename to	Length	Precision
1	kode_trayek			
2	trayek			
3	trayekID			

Get fields to select  
Edit Mapping

Include unspecified fields, ordered by name ☐

OK Cancel

? Help



# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL DIMENSI - DIM JENIS & TRAYEK

PENTAHO DATA INTEGRATION

## 5. STEP INSERT/UPDATE

Insert / update

Step name:

Connection:

Target schema:

Target table:

Commit size:

Don't perform any updates: ☐

The key(s) to look up the value(s):

#	Table field	Comparator	Stream field1	Stream field2
1	trayekID	=	trayekID	

Update fields:

#	Table field	Stream field	Update
1	kode_trayek	kode_trayek	Y
2	trayek	trayek	Y
3	trayekID	trayekID	Y

# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL DIMENSI - DIM JENIS & TRAYEK

MYSQL DATABASE

**DIM JENIS**

	jenis	jenisID
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	BRT	1
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	Transjabodetabek	2
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	Angkutan Umum Integrasi	3
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	Tidak Diketahui	4
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	Rusun	5
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	Royaltrans	6
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	Bus Wisata	7
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	Angkutan Umum Lainnya	8
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	Mikrotrans	9
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	Penugasan	10
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	EKSPRES	11
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	AMARI	12
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	MOBILE APPS	13
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	Angkutan Pekan Raya Jakarta	14
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	Khusus	15
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	Transcare	16

**DIM TRAYEK**

	kode_trayek	trayek	trayekID
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	1	BLOK M - KOTA	1
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	2	PULO GADUNG - MONUMEN NASIONAL	2
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	3	KALIDERES - MONUMEN NASIONAL VIA VETERAN	3
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	4	PULO GADUNG - GALUNGGUNG	4
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	5	KAMPUNG MELAYU - ANCOL	5
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	6	RAGUNAN - GALUNGGUNG	6
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	7	KAMPUNG MELAYU - KAMPUNG RAMBUTAN	7
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	8	LEBAK BULUS - PASAR BARU	8
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	9	PINANG RANTI - PLUIT	9
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	10	TANJUNG PRIOK - PGC	10
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	11	PULO GEBANG - KAMPUNG MELAYU	11
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	12	PLUIT - TANJUNG PRIOK	12
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	13	PURI BETA - TEGAL MAMPANG	13
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	14	JAKARTA INTERNATIONAL STADIUM - SENEN	14
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	[14225]	HALTE NON BRT BEKASI BARAT	15
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	[14226]	HALTE NON BRT BULAK KAPAL	16
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	[14227]	HALTE NON BRT BEKASI TIMUR	17
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	[14229]	HALTE NON BRT SUMMARECON BEKASI	18
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	[14230]	HALTE NON BRT PLAZA STASIUN MANGGARAI	19
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	[14246]	TANAH ABANG	20
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	10A	RUSUN MARUNDA - TANJUNG PRIOK	21
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	10B	RUSUN CIPINANG BESAR SELATAN - PENAS KALIMALANG	22
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	10C	PELABUHAN TANJUNG PRIOK - TANJUNG PRIOK	23
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	10D	TANJUNG PRIOK - KP RAMBUTAN	24
<input type="checkbox"/> Ubah <input type="checkbox"/> Salin <input type="checkbox"/> Hapus	10E	PULO GEBANG - TANJUNG PRIUK	25

# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL FAKTA - FACT PENUMPANG

MYSQL DATABASE

## STRUKTUR TABEL FAKTA

Server: 127.0.0.1 » Database: transjakarta\_olap » Tabel: fact

JelajahiStrukturSQLCariTambahkanEksporImporHak AksesOpe

Struktur tabelTampilan hubungan

#	Nama	Jenis	Penyortiran	Atribut	Tak Ternilai	Bawaan	Komentar	Ekstra	Tindakan
<input type="checkbox"/>	1 sk 🔑	int(11)			Ya	NULL			Ubah  Hapus Lainnya
<input type="checkbox"/>	2 jenisID 🔑	int(100)			Tidak	Tidak ada			Ubah  Hapus Lainnya
<input type="checkbox"/>	3 trayekID 🔑	int(100)			Tidak	Tidak ada			Ubah  Hapus Lainnya
<input type="checkbox"/>	4 jumlah_penumpang	int(255)			Tidak	Tidak ada			Ubah  Hapus Lainnya

## INDEKS TABEL FAKTA

Indeks ⓘ

Tindakan	Nama kunci	Jenis	Unik	Dipadatkan	Kolom	Kardinalitas	Penyortiran	Tak Ternilai	Komentar
Ubah  Rename  Hapus	jenisID_key	BTREE	Tidak	Tidak	jenisID	22	A	Tidak	
Ubah  Rename  Hapus	trayekID_key	BTREE	Tidak	Tidak	trayekID	622	A	Tidak	
Ubah  Rename  Hapus	sk_key	BTREE	Tidak	Tidak	sk	120	A	Ya	

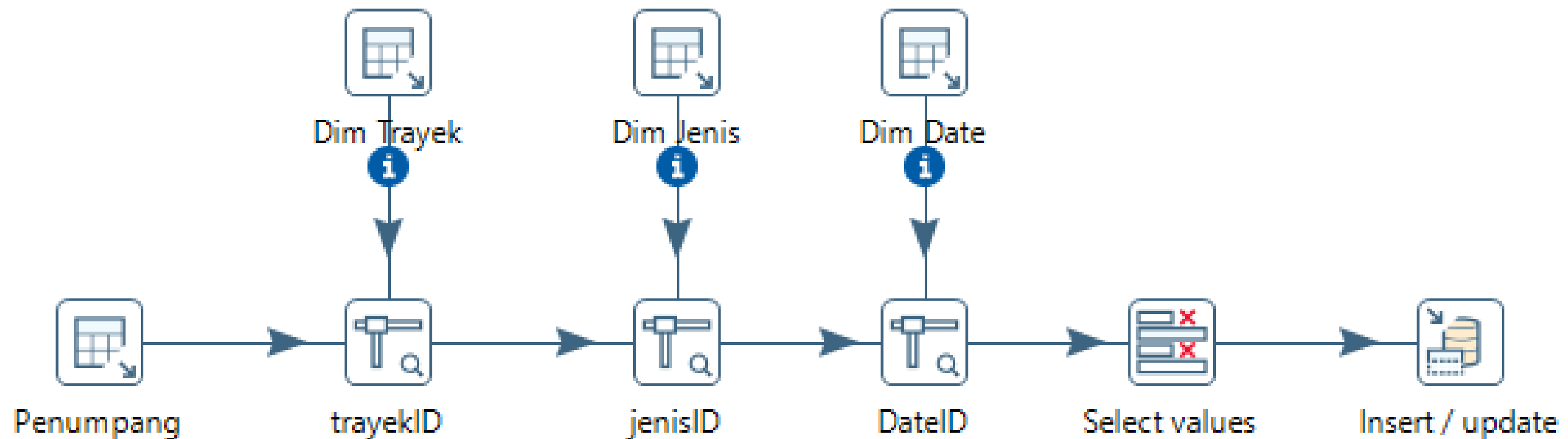
# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL FAKTA - FACT PENUMPANG

PENTAHO DATA INTEGRATION

## TRANSFORMASI PEMBUATAN TABEL FACT



# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL FAKTA - FACT PENUMPANG

PENTAHO DATA INTEGRATION

## 1.STEP TABLE INPUT

Table input

Step name:

Connection:

SQL

```
SELECT
  tahun
, bulan
, kode_trayek
, trayek
, jenis
, jumlah_penumpang
FROM test.penumpang
```

Line 1 Column 0

Store column info in step meta data ☐

Enable lazy conversion ☐

Replace variables in script? ☐

Insert data from step

Execute for each row? ☐

Limit size:

## 2.STEP STREAM LOOKUP

Stream lookup

Step name:

Lookup step:

The key(s) to look up the value(s):

#	Field	LookupField
1	kode_trayek	kode_trayek

Specify the fields to retrieve :

#	Field	New name	Default	Type
1	trayekID			Integer

Preserve memory (costs CPU) ☒

Key and value are exactly one integer field ☐

Use sorted list (i.s.o. hashtable) ☐

# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL FAKTA - FACT PENUMPANG

PENTAHO DATA INTEGRATION

## 3.STEP TABLE INPUT

Select values

Step name: Select values

Select & Alter Remove Meta-data

Fields to alter the meta-data for:

#	Fieldname	Rename to	Type	Length	Precision	Binary
1	jumlah_penumpang		None	9	0	N
2	trayekID		None	9	0	N
3	jenisID		None	9	0	N
4	sk		None			N

Get fields to change

OK Cancel

Help

## 4.STEP INSERT/UPDATE

Insert / update

Step name: Insert / update

Connection: Test Edit... New... Wizard...

Target schema: test\_olap Browse...

Target table: fact Browse...

Commit size: 100

Don't perform any updates: ☐

The key(s) to look up the value(s):

#	Table field	Comparator	Stream field1	Stream
1	jumlah_penumpang	=	jumlah_penumpang	
2	trayekID	=	trayekID	
3	jenisID	=	jenisID	
4	sk	=	sk	

Get fields

Update fields:

#	Table field	Stream field	Update
1	jumlah_penumpang	jumlah_penumpang	Y
2	trayekID	trayekID	Y
3	jenisID	jenisID	Y
4	sk	sk	Y

Get update fields

Edit mapping

Help OK Cancel SQL

# PROSES ETL DATABASE OLAP

FOUR-STEP METHODOLOGY KIMBALL

TABEL FAKTA - FACT PENUMPANG

MYSQL DATABASE

FACT PENUMPANG

sk	jenisID	trayekID	jumlah_penumpang
20200131	1	1	2539866
20200131	1	2	813240
20200131	1	3	1059350
20200131	1	4	774443
20200131	1	5	1092746
20200131	1	6	1079883
20200131	1	7	932231
20200131	1	8	1070869
20200131	1	9	1583719
20200131	1	10	989878
20200131	1	11	356512
20200131	1	12	300058
20200131	1	13	790517
20200131	5	21	44393
20200131	5	22	8780
20200131	3	23	748
20200131	3	26	3759
20200131	3	28	57716

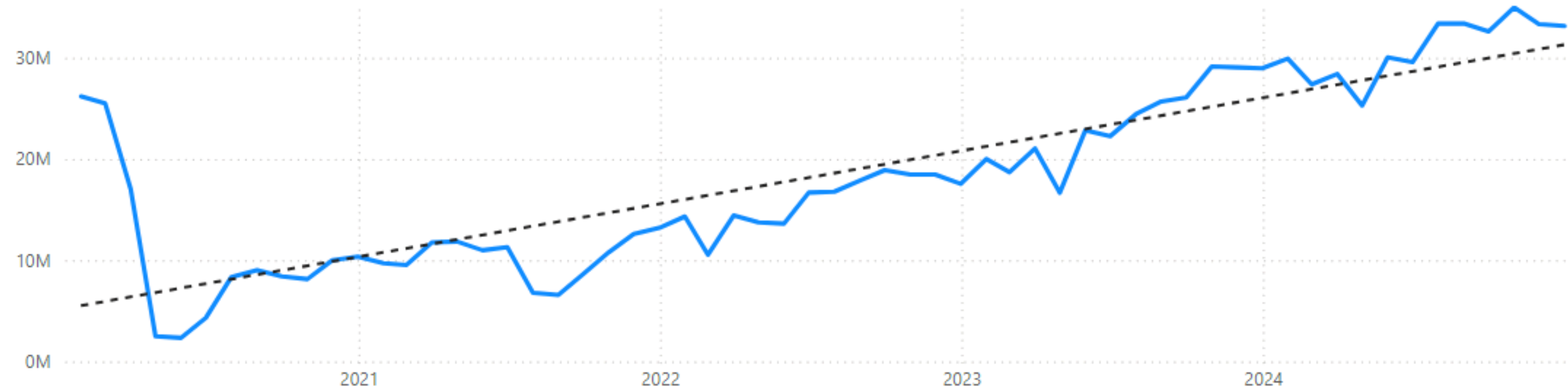


# VISUALISASI DATA

*DASHBOARD PENUMPANG TRANSJAKARTA*

## *TREN JUMLAH PENUMPANG BERDASARKAN BULAN DAN TAHUN*

Trend Jumlah Penumpang per Bulan atau Tahun



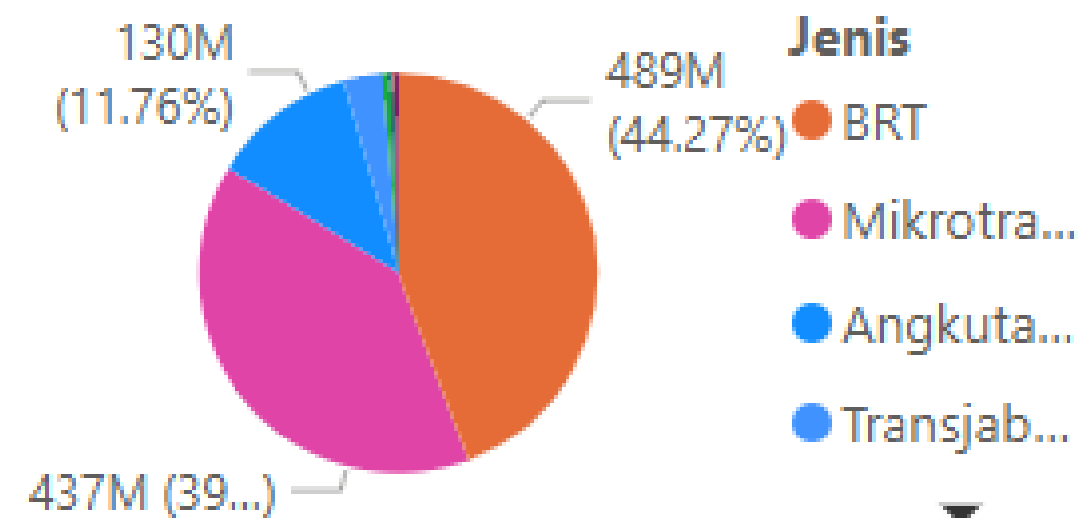


# VISUALISASI DATA

## DASHBOARD PENUMPANG TRANSJAKARTA

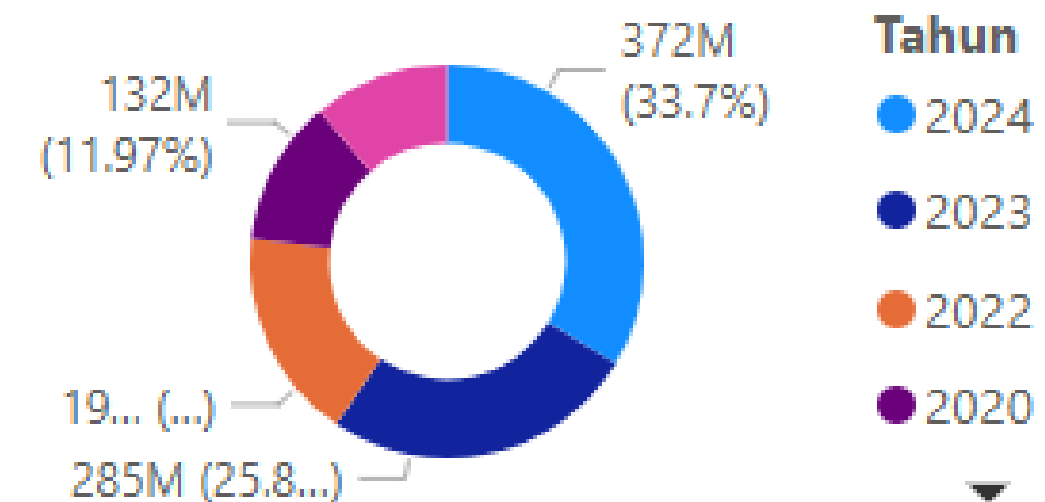
### DISTRIBUSI PENUMPANG BERDASARKAN JENIS TRAYEK

Distribusi Penumpang Berdasarkan Jenis Trayek



### DISTRIBUSI PENUMPANG BERDASARKAN TAHUN

Distribusi Penumpang Berdasarkan Tahun

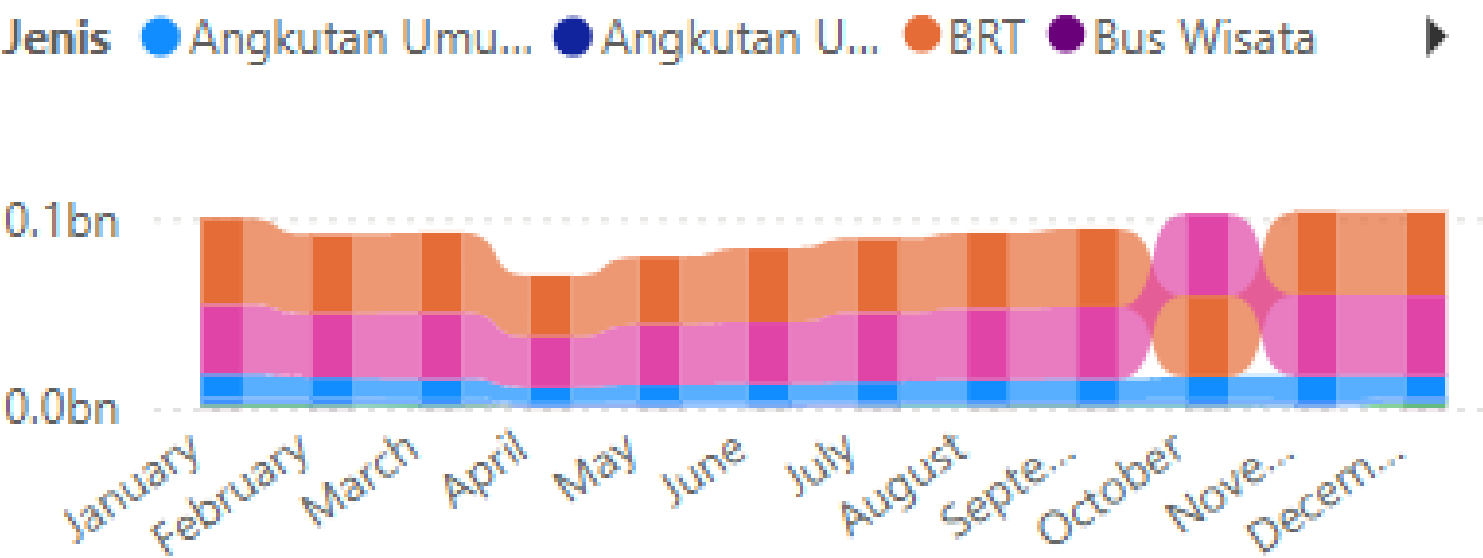


# VISUALISASI DATA

DASHBOARD PENUMPANG TRANSJAKARTA

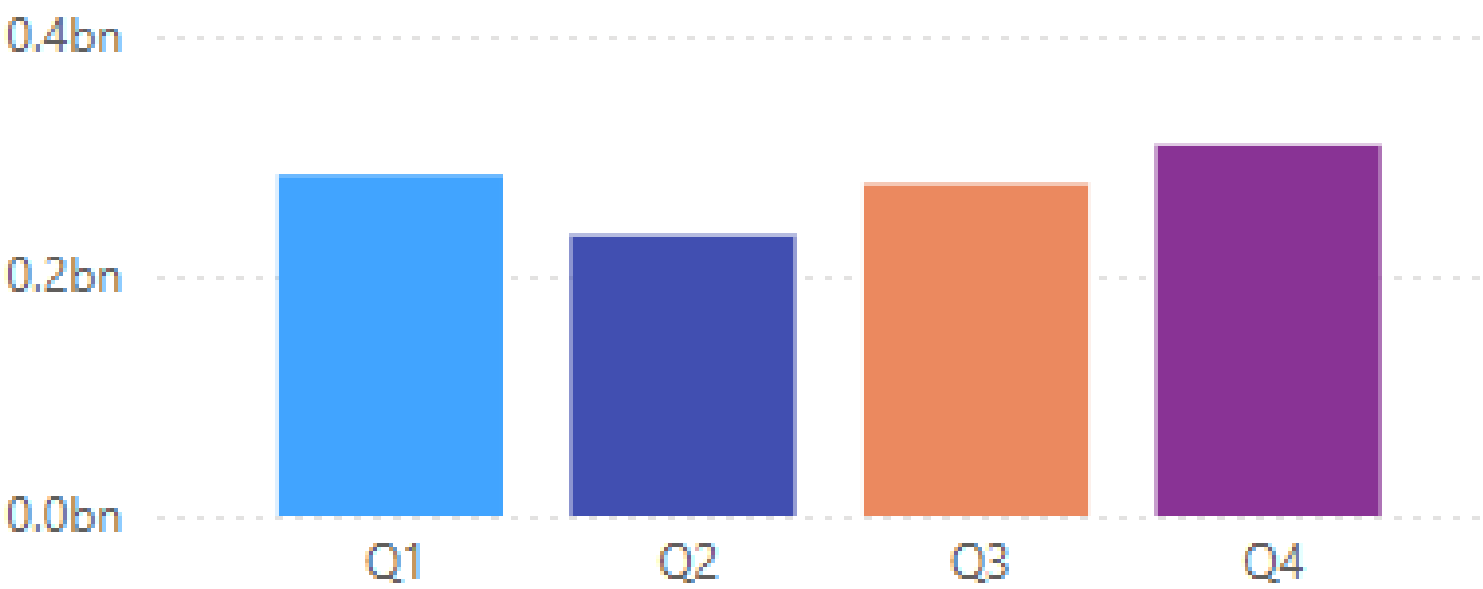
## ANALISIS JENIS RUTE BERDASARKAN JUMLAH PENUMPANG DAN BULAN

Analisis Jenis Trayek per Bulan



## ANALISIS KUARTAL DAN JUMLAH PENUMPANG

Analisis Quartal dan Jumlah Penumpang

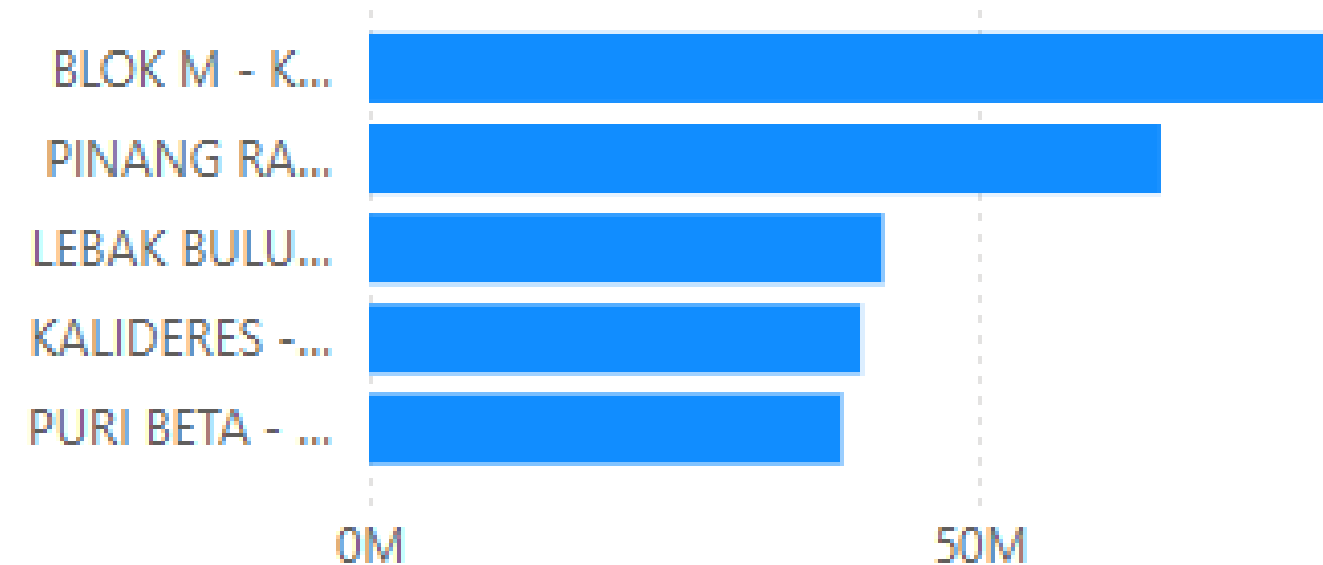


# VISUALISASI DATA

DASHBOARD PENUMPANG TRANSJAKARTA

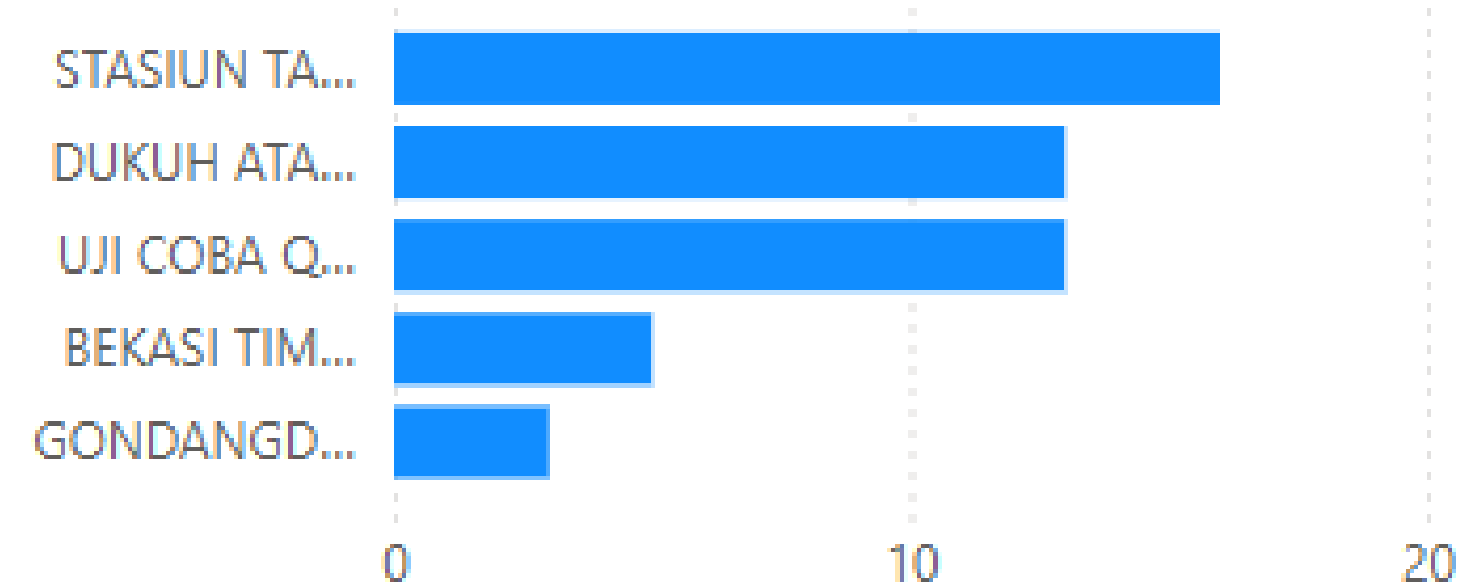
## TOP 5 RATA-RATA PENUMPANG TERTINGGI BERDASARKAN TRAYEK

Top 5 Rata-rata Penumpang Tertinggi



## TOP 5 RATA-RATA PENUMPANG TERENDAH BERDASARKAN TRAYEK

Top 5 Rata-rata Penumpang Terendah



# VISUALISASI DATA

*DASHBOARD PENUMPANG TRANSJAKARTA*

## *CARD JUMLAH TOTAL*

Jenis Trayek

16

Trayek

548

Penumpang

1103717345

# VISUALISASI DATA

## DASHBOARD PENUMPANG TRANSJAKARTA

### *FILTERING DASHBOARD*

Tanggal

1/1/2020

12/31/2024

Tahun

All

Bulan

All

Jenis

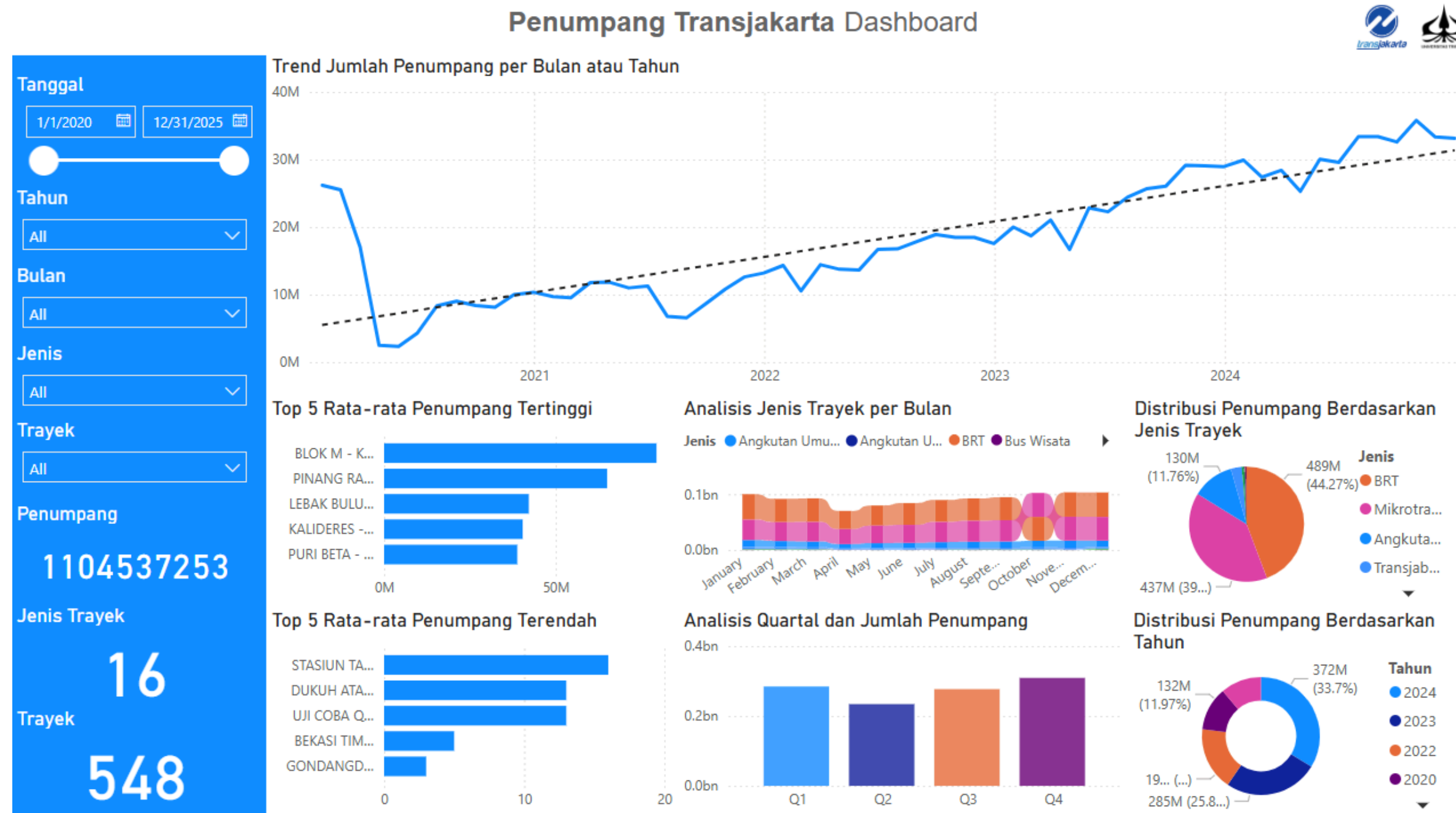
All

Trayek

All

# VISUALISASI DATA

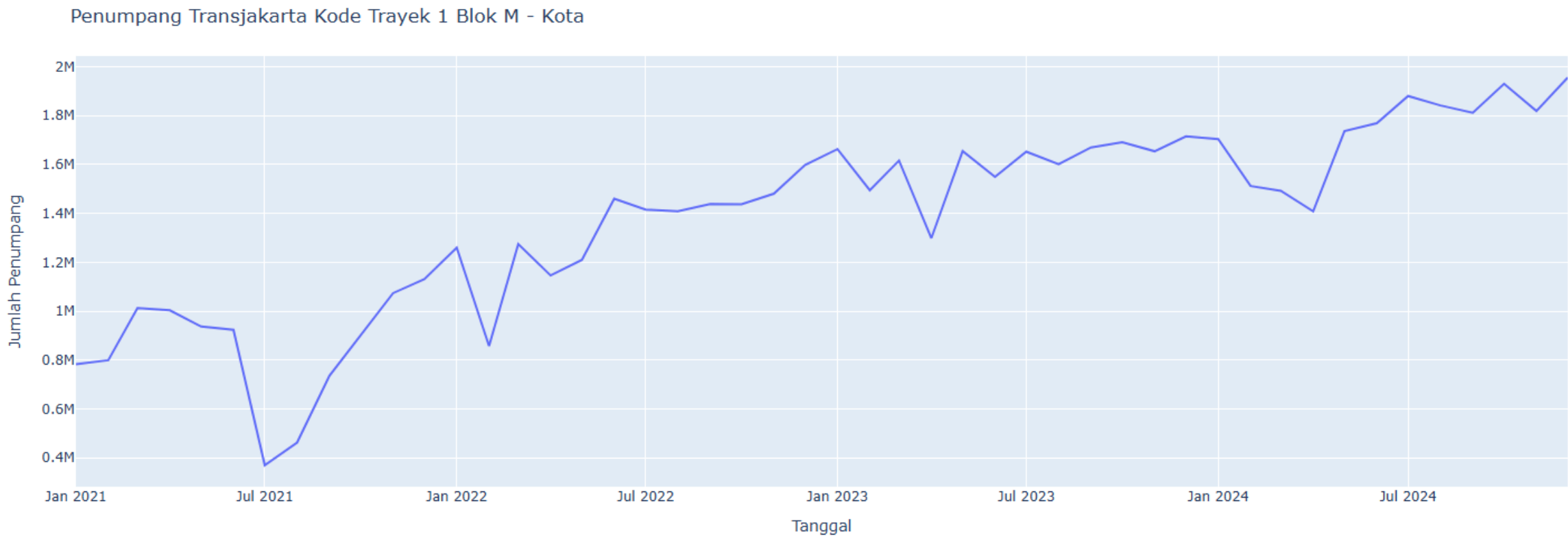
## DASHBOARD PENUMPANG TRANSJAKARTA



# ANALISA PREDIKSI

## SETELAH DATA CLEANING

jumlah_penumpang	
tanggal	
2021-01-01	782966
2021-02-01	799708
2021-03-01	1013221
2021-04-01	1003534
2021-05-01	937452





# ANALISA PREDIKSI

## UJI STASIONERITAS

Results of Dickey-Fuller Test:

```
Test Statistic      -1.234184
p-value             0.658721
#Lags Used          1.000000
Number of Observations Used 46.000000
Critical Value (1%)  -3.581258
Critical Value (5%)  -2.926785
Critical Value (10%) -2.601541
dtype: float64
result : Non stationary
```

## SETELAH DIFFERENCING

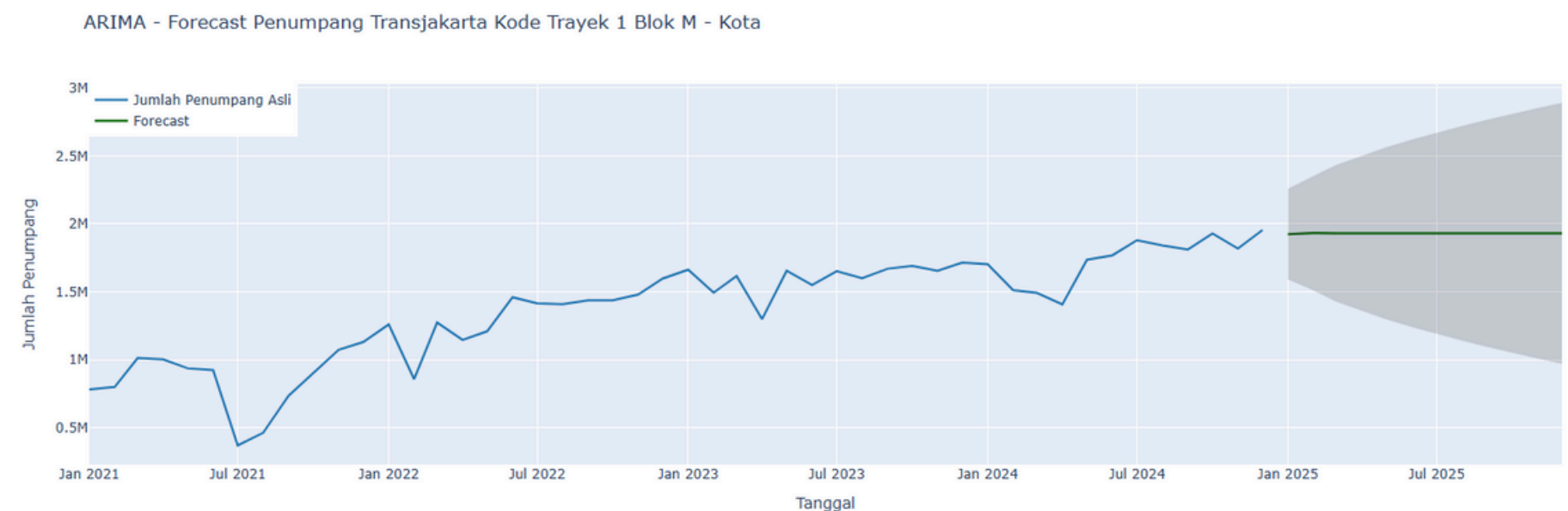
Results of Dickey-Fuller Test:

```
Test Statistic      -9.473117e+00
p-value             4.056047e-16
#Lags Used          0.000000e+00
Number of Observations Used 4.600000e+01
Critical Value (1%)  -3.581258e+00
Critical Value (5%)  -2.926785e+00
Critical Value (10%) -2.601541e+00
dtype: float64
result : Stationary
```

## MENENTUKAN MODEL ARIMA

```
SARIMAX Results
=====
Dep. Variable:          y      No. Observations:          48
Model:                 SARIMAX(1, 1, 0)      Log Likelihood      -633.056
Date:                 Mon, 27 Jan 2025      AIC                  1270.113
Time:                 15:48:56      BIC                  1273.813
Sample:               01-01-2021      HQIC                 1271.505
                  - 12-01-2024

Covariance Type:      opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ar.L1          -0.2211         0.116      -1.902      0.057      -0.449         0.007
sigma2         2.882e+10      5.71e-13      5.05e+22      0.000      2.88e+10      2.88e+10
=====
Ljung-Box (L1) (Q):          0.43      Jarque-Bera (JB):          16.75
Prob(Q):                    0.51      Prob(JB):              0.00
Heteroskedasticity (H):      0.25      Skew:                  -0.96
Prob(H) (two-sided):         0.01      Kurtosis:              5.20
=====
```



# ANALISA PREDIKSI

## MENENTUKAN MODEL SARIMA

Best model: ARIMA(0,1,0)(1,0,0)[12]  
Total fit time: 3.615 seconds

SARIMAX Results

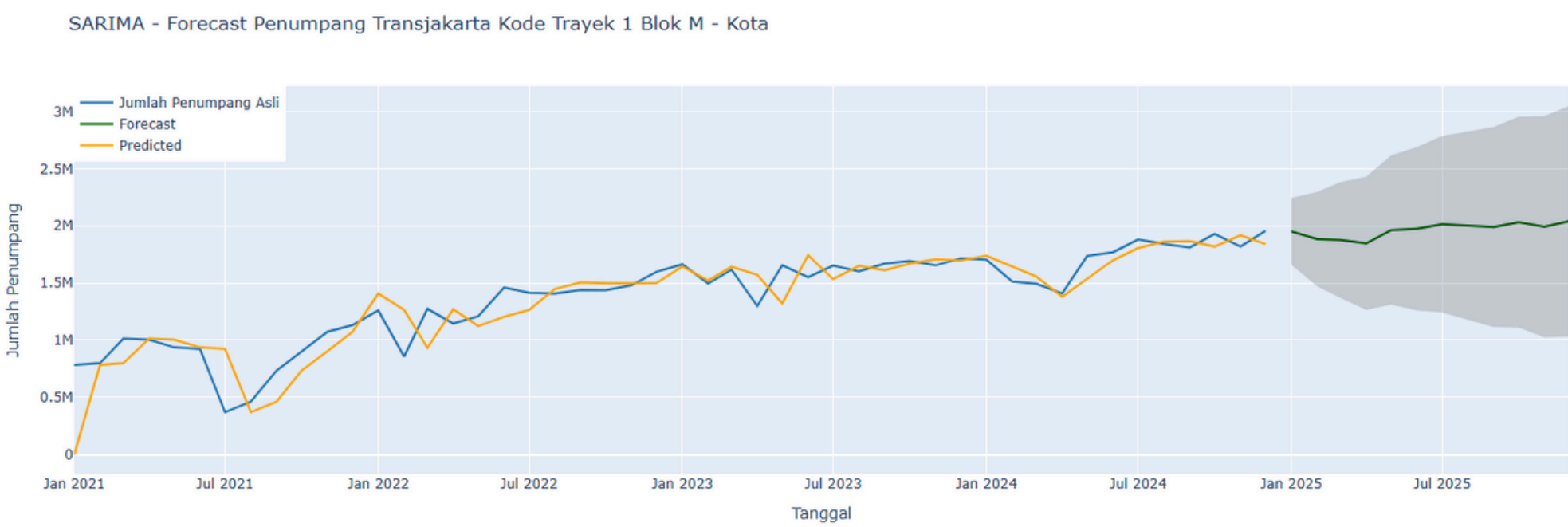
Dep. Variable: y  
Model: SARIMAX(0, 1, 0)x(1, 0, 0, 12)  
Date: Tue, 28 Jan 2025  
Time: 07:27:24  
Sample: 01-01-2021 - 12-01-2024

No. Observations: 48  
Log Likelihood: -631.617  
AIC: 1267.233  
BIC: 1270.934  
HQIC: 1268.626

Covariance Type: opg

	coef	std err	z	P> z	[0.025	0.975]
ar.S.L12	0.3542	0.102	3.486	0.000	0.155	0.553
sigma2	2.213e+10	1.8e-12	1.23e+22	0.000	2.21e+10	2.21e+10

Ljung-Box (L1) (Q): 1.28    Jarque-Bera (JB): 10.20  
Prob(Q): 0.26    Prob(JB): 0.01  
Heteroskedasticity (H): 0.16    Skew: -0.71  
Prob(H) (two-sided): 0.00    Kurtosis: 4.78



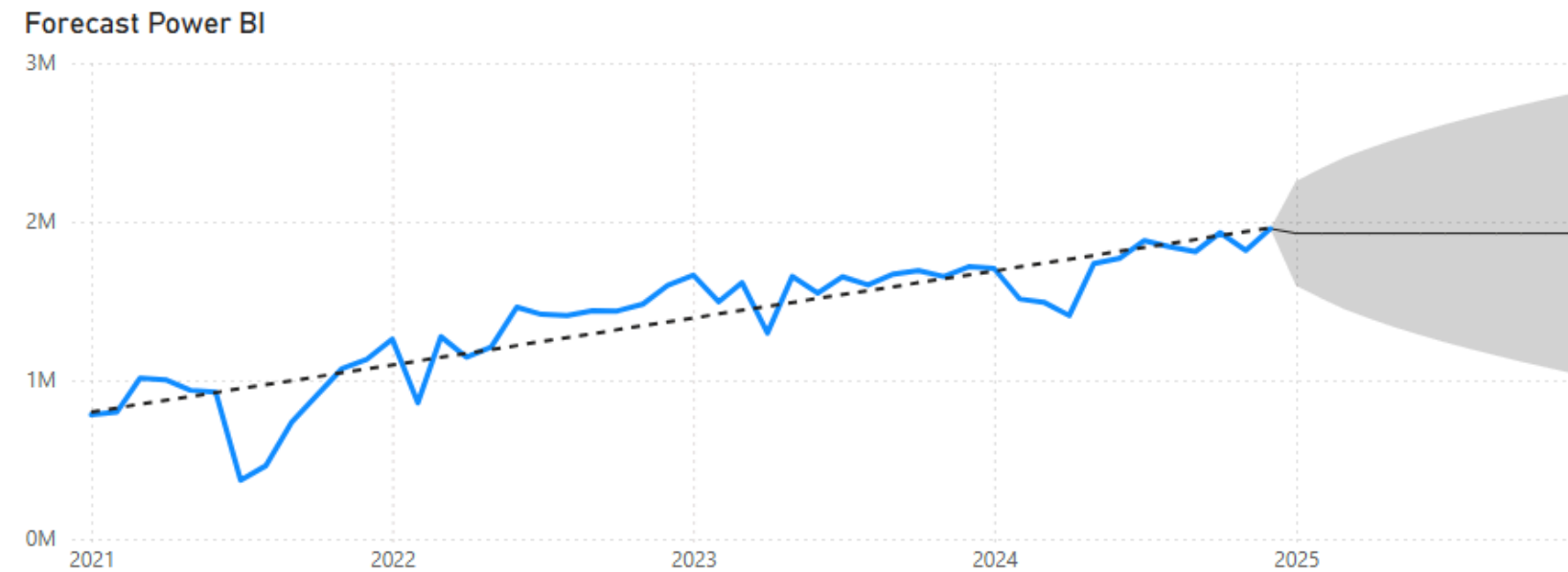
## HASIL PREDIKSI

	jumlah_penumpang	prediksi_arima	prediksi_sarima
2021-01-01	782966.0	NaN	NaN
2021-02-01	799708.0	NaN	NaN
2021-03-01	1013221.0	NaN	NaN
2021-04-01	1003534.0	NaN	NaN
2021-05-01	937452.0	NaN	NaN
2025-01-01	NaN	1.925489e+06	1.951691e+06
2025-02-01	NaN	1.932171e+06	1.883738e+06
2025-03-01	NaN	1.930693e+06	1.876709e+06
2025-04-01	NaN	1.931020e+06	1.846904e+06
2025-05-01	NaN	1.930948e+06	1.963159e+06
2025-06-01	NaN	1.930964e+06	1.974585e+06
2025-07-01	NaN	1.930960e+06	2.014089e+06
2025-08-01	NaN	1.930961e+06	2.000433e+06
2025-09-01	NaN	1.930961e+06	1.989707e+06

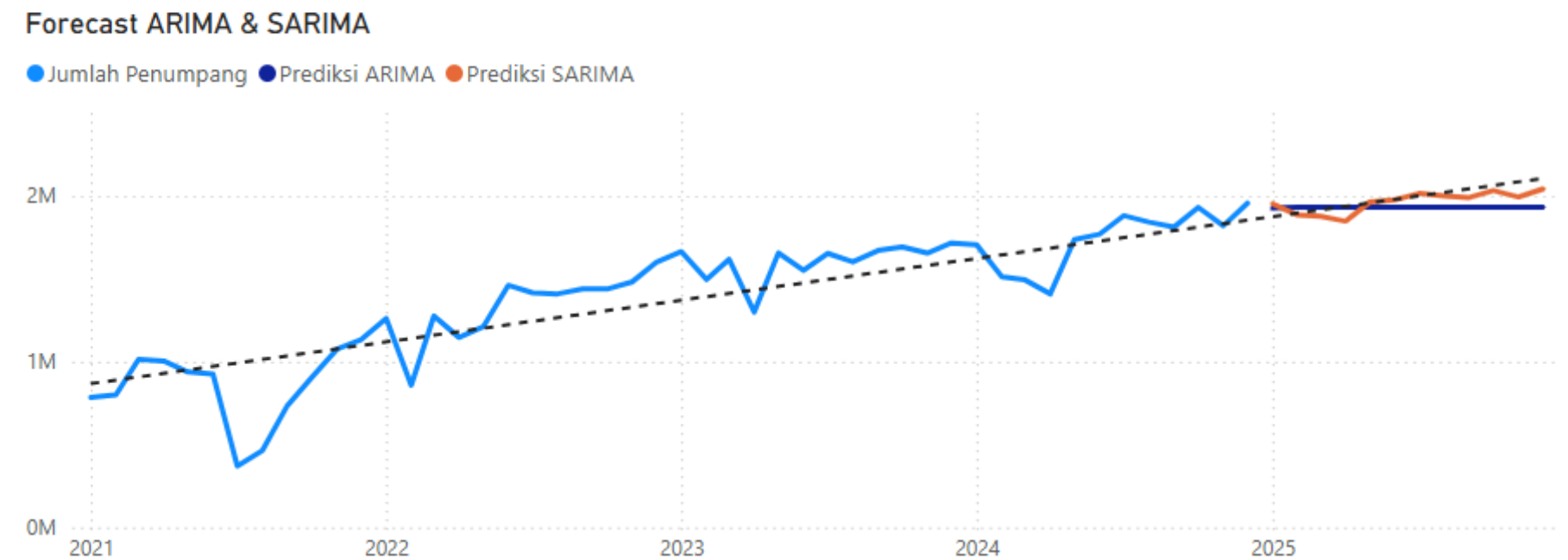
# VISUALISASI DATA

## DASHBOARD PREDIKSI PENUMPANG TRANSJAKARTA KODE TRAYEK 1 BLOK M - KOTA

### FORECAST POWER BI



### FORECAST ARIMA & SARIMA



# VISUALISASI DATA

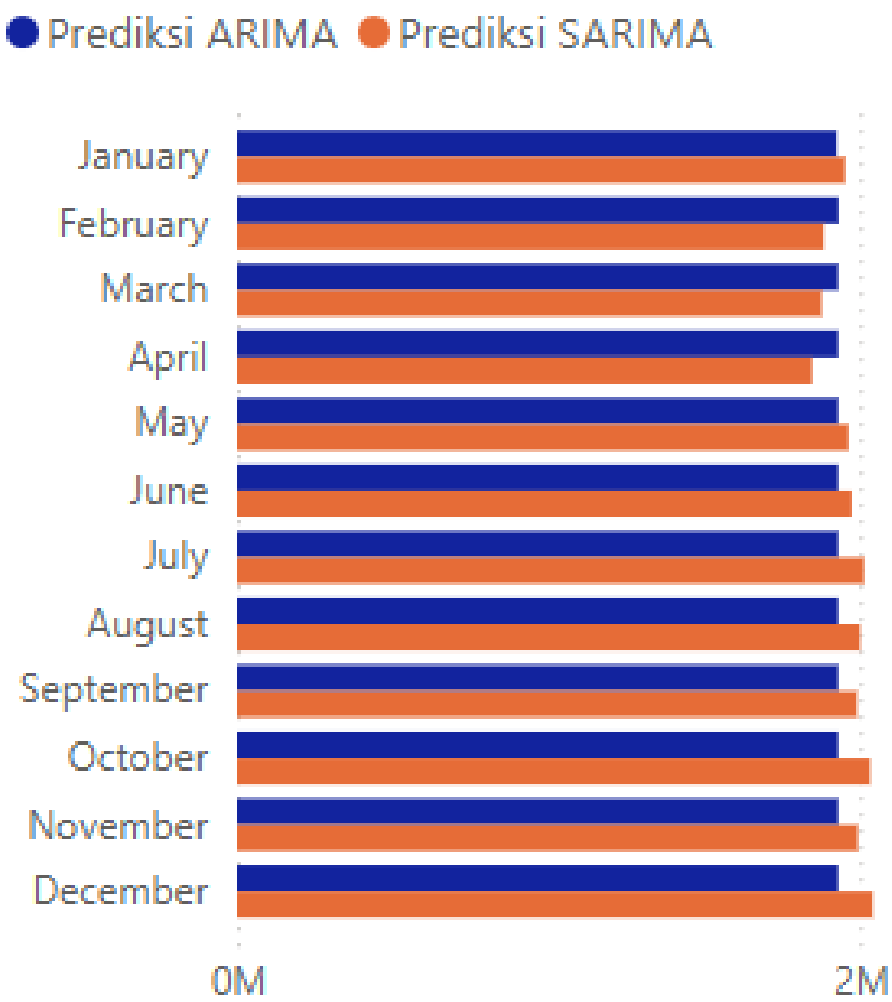
DASHBOARD PREDIKSI PENUMPANG TRANSJAKARTA KODE TRAYEK 1 BLOK M - KOTA

TABEL PREDIKSI ARIMA DAN SARIMA

Year	Month	Prediksi ARIMA	Prediksi SARIMA
2025	January	1925489	1951691
2025	February	1932171	1883738
2025	March	1930693	1876709
2025	April	1931020	1846904
2025	May	1930948	1963159
2025	June	1930964	1974585
2025	July	1930960	2014089
2025	August	1930961	2000433
2025	September	1930961	1989707
2025	October	1930961	2031773
2025	November	1930961	1992374
2025	December	1930961	2040764

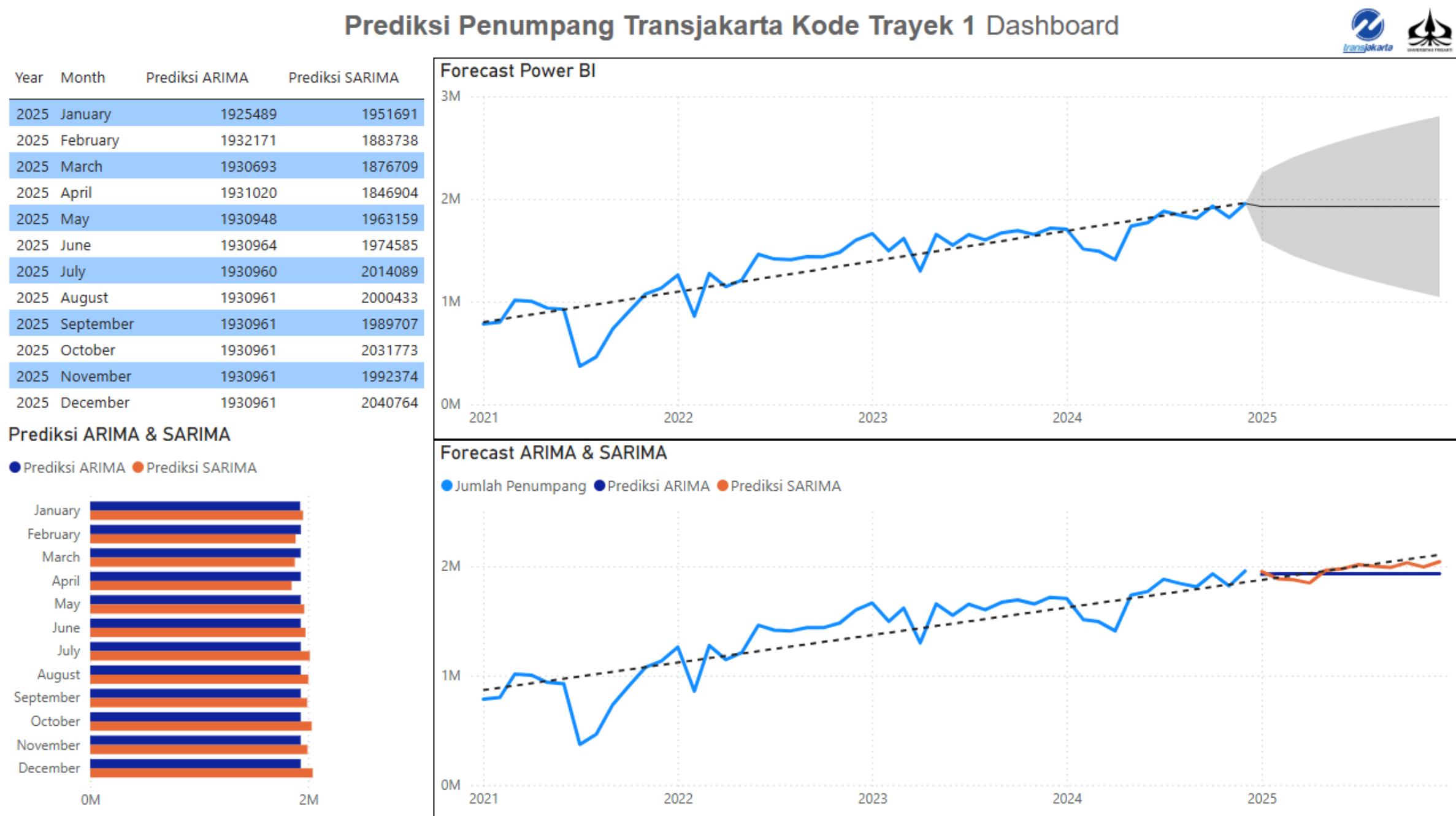
PREDIKSI ARIMA & SARIMA

Prediksi ARIMA & SARIMA



# VISUALISASI DATA

## DASHBOARD PREDIKSI PENUMPANG TRANSJAKARTA KODE TRAYEK 1 BLOK M - KOTA



# KESIMPULAN

- **Tren penumpang** Transjakarta **meningkat setiap tahun** setelah penurunan tajam pada awal 2020 akibat COVID-19 berdasarkan visualisasi data.
- Prediksi SARIMA untuk trayek "**Blok M - Kota**" menunjukkan penurunan dalam 4 bulan ke depan dan **kenaikan dalam 8 bulan** setelahnya. **MAPE** sebesar **13.86%** menunjukkan **model cukup akurat**.
- Rute dengan **jumlah penumpang tertinggi** adalah "**Blok M - Kota**", diikuti "Pinang Ranti - Pluit" dan "Lebak Bulus - Pasar Baru", sementara rute dengan penumpang terendah termasuk "Stasiun Tanah Abang - Stasiun Angke". Jenis trayek **BRT mendominasi (44,27%)**, diikuti oleh Mikrotrans dan Angkutan Umum Integrasi, sedangkan Transjabodetabek memiliki porsi terkecil. Tahun **2024** mencatat jumlah penumpang **tertinggi pascapandemi**. Distribusi kuartalan cukup merata, dengan Q3 dan Q4 sedikit lebih tinggi.

# SARAN

Penelitian selanjutnya dapat mengintegrasikan **data operasional** seperti **waktu kedatangan, keterlambatan armada, dan kepadatan halte** untuk analisis faktor-faktor yang mempengaruhi pola penumpang. Penggunaan **lisensi Power BI Pro** atau **Premium** juga disarankan agar memungkinkan akses laporan yang lebih fleksibel, berbagi dataset, dan penggunaan melalui aplikasi mobile. Selain itu, model **prediksi** lain seperti **LSTM, Prophet, Exponential Smoothing** dapat dikembangkan untuk meningkatkan akurasi peramalan jumlah penumpang dengan menangani data musiman yang lebih kompleks.



The background features a series of overlapping rectangles and lines in blue and orange. A large orange rectangle is centered behind the text. To its left, a blue rectangle is partially visible. Above the text, there are several smaller rectangles: a tall blue one on the left, a wide orange one in the center, and a blue one on the right. Below the text, there is a blue rectangle on the left, a small orange square, a wide blue rectangle, and a tall orange rectangle on the right. A long orange line runs horizontally across the middle, and another orange line runs horizontally below the text, connected by a vertical line on the right side.

**THANK YOU**