

JOBSHEET 1

DATA WAREHOUSE

Tugas 1

1. Berapa jumlah kolom pada data tersebut?

Jawab : Jumlah keseluruhan kolom pada data tersebut adalah 7 kolom, meliputi kolom customer, product, region, date, item unit price, no items, dan total sales.

2. Apa arti atau isi data dari setiap kolom yang ada?

Jawab : Setiap kolom dalam tabel merepresentasikan jenis data tertentu sesuai dengan nama kolom

- Customer : Menunjukkan nama atau ID Pelanggan yang melakukan pembelian
- Product : Menunjukkan produk yang dibeli oleh pelanggan
- Region : Wilayah tempat transaksi dilakukan
- Date : Tanggal transaksi dilakukan
- Item unit price : Harga satuan produk yang dibeli
- No. Items : Jumlah produk yang dibeli dalam satu transaksi
- Total Sale : Total penjualan, hasil perkalian dari jumlah produk dan harga satuan

3. Adakah data yang memiliki nilai null/data yang tidak lengkap?

Jawab : Terdapat 5 baris data yang memiliki nilai null/data kosong antara lain:

- Baris 4 : kolom product bernilai null/kosong

Customer 3		MLG	7/24/2019	168.95	2	337.9
------------	--	-----	-----------	--------	---	-------

- Baris 8 : kolom region dan date bernilai null/kosong

Customer 2	Avanza AT			168.95	15	2534.25
------------	-----------	--	--	--------	----	---------

- Baris 13 : kolom product, region, dan date bernilai null/kosong

Customer 1				340.95	7	2386.65
------------	--	--	--	--------	---	---------

- Baris 17 : kolom product bernilai null/kosong

Customer 2		KDR	8/7/2020	799.95	14	11199.3
------------	--	-----	----------	--------	----	---------

- Baris 19 : kolom region bernilai null/kosong

Customer 1	Avanza AT		8/12/2020	168.95	4	675.8
------------	-----------	--	-----------	--------	---	-------

4. Adakah data yang memiliki tipe data yang berbeda dengan data lainnya pada kolom yang sama?

Jawab : Tidak ada tipe data yang berbeda dalam kolom yang sama

- **Customer** menggunakan tipe data String
- **Product** menggunakan tipe data String
- **Region** menggunakan tipe data String
- **Date** menggunakan tipe data date (format tanggal)
- **Item unit price, No.Items, dan Total Sale** menggunakan tipe data angka

A. Pengambilan Data (Extract)

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
1	Customer 1	Yaris AT	MLG	Wed Mar 13 00:00:00 ICT 2019	340.95	6	2045.7
2	Customer 1	Etios AT	SBY	Tue Apr 16 00:00:00 ICT 2019	799.95	1	799.95
3	Customer 3	Avanza AT	MLG	Tue Jun 18 00:00:00 ICT 2019	168.95	7	1182.65
4	Customer 3	<null>	MLG	Wed Jul 24 00:00:00 ICT 2019	168.95	2	337.9
5	Customer 1	Avanza AT	SBY	Wed Aug 21 00:00:00 ICT 2019	168.95	9	1520.55
6	Customer 2	Etios AT	KDR	Sat Sep 07 00:00:00 ICT 2019	799.95	7	5599.65
7	Customer 3	Etios AT	MLG	Sat Sep 07 00:00:00 ICT 2019	799.95	8	6399.6
8	Customer 2	Avanza AT	<null>	<null>	168.95	15	2534.25
9	Customer 1	Avanza AT	SBY	Sat Sep 14 00:00:00 ICT 2019	168.95	15	2534.25
10	Customer 1	Etios AT	SBY	Wed Sep 25 00:00:00 ICT 2019	799.95	2	1599.9
11	Customer 1	Avanza AT	SBY	Tue Jan 21 00:00:00 ICT 2020	168.95	14	2365.3
12	Customer 1	Yaris AT	MLG	Sat Feb 01 00:00:00 ICT 2020	340.95	2	681.9
13	Customer 1	<null>	<null>	<null>	340.95	7	2386.65
14	Customer 1	Etios AT	SBY	Sat May 30 00:00:00 ICT 2020	799.95	9	7199.55
15	Customer 3	Avanza AT	MLG	Sun Jun 07 00:00:00 ICT 2020	168.95	10	1689.5
16	Customer 3	Yaris AT	MLG	Tue Jun 09 00:00:00 ICT 2020	340.95	4	1363.8
17	Customer 2	<null>	KDR	Fri Aug 07 00:00:00 ICT 2020	799.95	14	11199.3
18	Customer 2	Avanza AT	KDR	Sat Aug 08 00:00:00 ICT 2020	168.95	12	2027.4
19	Customer 1	Avanza AT	<null>	Wed Aug 12 00:00:00 ICT 2020	168.95	4	675.8

Tugas 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

Jawab : Data hasil eksekusi sesuai dengan data aslinya, terdapat 19 baris data dan terdapat 5 baris data yang memiliki nilai null/kosong

2. PDI Spoon melakukan proses extract, perhatikan Tab Logging di Execution Results Area, langkah-langkah apa sajakah yang dilakukan PDI Spoon untuk melakukan extract data?

Jawab :

Langkah extract data memiliki beberapa tahap antara lain :

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2025/02/19 09:06:30	- Spoon - Save file as...
2025/02/19 09:06:32	- Spoon - Transformation opened.
2025/02/19 09:06:32	- Spoon - Launching transformation [Transformation 1]...
2025/02/19 09:06:32	- Spoon - Started the transformation execution.
2025/02/19 09:06:34	- Transformation 1 - Dispatching started for transformation [Transformation 1]
2025/02/19 09:06:35	- Input Data CSV.0 - Header row skipped in file 'F:\= KULIAH\SEMESTER 4\3. PRAK JOBSHEET\data warehouse\CarDistributionSales.csv'
2025/02/19 09:06:35	- Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)
2025/02/19 09:06:35	- Spoon - The transformation has finished!!

- Spoon – Save File as (menyimpan data) :** Spoon menyimpan data sebelum melakukan eksekusi
- Spoon – Transformation opened (membuka transformasi) :** Spoon membuka transformasi yang akan di jalankan
- Spoon – Launching Transformation :** Spoon menjalankan transformasi (Transformation 1)
- Spoon – Started the transformation execution :** Memproses transformasi
- Transformation 1 :** Menjalankan proses transformasi
- Input Data CSV.0 :** Spoon membaca CSV file dan melewati baris pertama jika terdapat header
- Input Data CSV.0 :** Spoon selesai membaca data dari file CSV dengan informasi jumlah baris yang diproses

h. **Spoon – The transformation has finished!** : Spoon menampilkan pesan bahwa proses extract data telah selesai dilakukan

3. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E?

2025/02/19 09:06:35 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)

Execution Results										
Logging Execution History Step Metrics Performance Graph Metrics Preview data										
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active
1	Input Data CSV	0	0	19	20	0	0	0	0	Finished

I : input

O : output

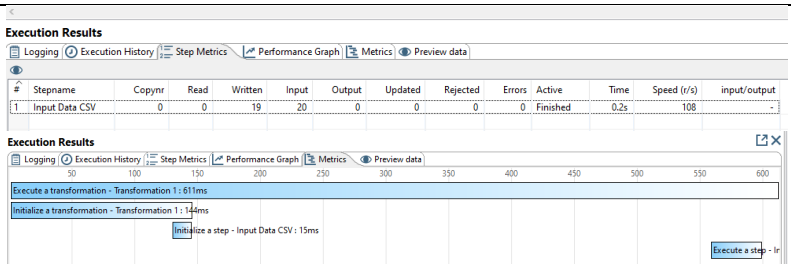
R : read

W : Written

U : updated

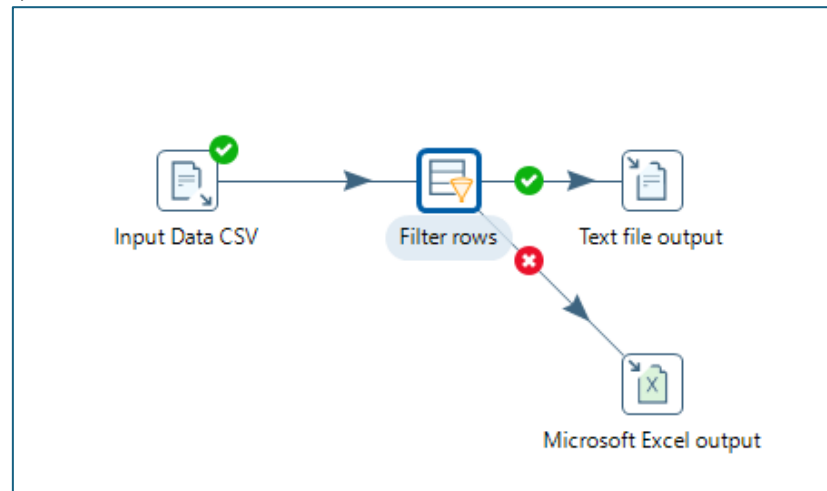
E : Errors

4. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini? Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU). Bandingkan dalam bentuk tabel

Waktu	Spesifikasi
	<ul style="list-style-type: none"> - Processor : Intel (R) Core (TM) i5-8250U CPU @1.60GHz (8 CPUs), ~1.8GHz - RAM : 8GB DDR3 L Memory - Storage : 1000 HDD - VGA : NVIDIA® GeForce® MX150 2 GB Intel® UHD Graphics 620 - CPU:

B. Filter Data (Transform) dan Pengemasan Data (Load)

1. Membuat transformation untuk melakukan filter data (transform) dan pengemasan data (load)



2. Melakukan setting kondisi pada filter rows, bertujuan untuk memisahkan baris yang memiliki data lengkap dan baris yang memiliki nilai null

Filter rows

Step name:

Send 'true' data to step:

Send 'false' data to step:

The condition:

☐ =

Filter rows

Step name:

Send 'true' data to step:

Send 'false' data to step:

The condition:

☐ IS NOT NULL

IS NOT NULL

The condition:

<input type="text"/>	^^ UP ^^	Level 3. Select UP to go up one level	+
Region	IS NOT NULL	<input type="text"/>	<input type="text"/>

The condition:

<input type="text"/>	^^ UP ^^	Level 4. Select UP to go up one level	+
Item unit price	IS NOT NULL	<input type="text"/>	<input type="text"/>

The condition:

<input type="text"/>	^^ UP ^^	Level 5. Select UP to go up one level	+
No.Items	IS NOT NULL	<input type="text"/>	<input type="text"/>

The condition:

<input type="text"/>	^^ UP ^^	Level 6. Select UP to go up one level	+
Total Sale	IS NOT NULL	<input type="text"/>	<input type="text"/>

3. Get fields

File	Content	Custom	Fields
#	Name	Type	Format
1	Customer	String	
2	Product	String	
3	Region	String	
4	Date	Date	
5	Item unit price	Number	0000.00
6	No.Items	Integer	
7	Total Sale	Number	00000.00

4. Menampilkan execution results kolom yang memiliki data null

Execution Results							
Logging Execution History Step Metrics Performance Graph Metrics Preview data							
<input checked="" type="radio"/> First rows <input type="radio"/> Last rows <input type="radio"/> Off							
#	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
1	Customer 3	<null>	MLG	Wed Jul 24 00:00:00 ICT 2019	168.95	2	337.9
2	Customer 2	Avanza AT	<null>	<null>	168.95	15	2534.25
3	Customer 1	<null>	<null>	<null>	340.95	7	2386.65
4	Customer 2	<null>	KDR	Fri Aug 07 00:00:00 ICT 2020	799.95	14	11199.3
5	Customer 1	Avanza AT	<null>	Wed Aug 12 00:00:00 ICT 2020	168.95	4	675.8

5. Menampilkan file output data excel yang memiliki nilai null

	A	B	C	D	E	F	G
1	Customer	Product	Region	Date	Item unit p	No.Items	Total Sale
2	Customer 3		MLG	Wednesday, July 24, 2019	0168,95	2,00	00337,90
3	Customer 2	Avanza AT			0168,95	15,00	02534,25
4	Customer 1				0340,95	7,00	02386,65
5	Customer 2		KDR	Friday, August 07, 2020	0799,95	14,00	11199,30
6	Customer 1	Avanza AT		Wednesday, August 12, 2020	0168,95	4,00	00675,80
7							
8							

6. Menampilkan file output data excel yang tidak memiliki nilai null

	A	B	C	D	E	F	G	H
1	Customer	Product	Region	Date	Item unit	No.Items	Total Sale	
2	Customer 1	Yaris AT	MLG	3/13/2019	340.9	6	2045.7	
3	Customer 1	Etios AT	SBY	4/16/2019	800	1	800	
4	Customer 3	Avanza AT	MLG	6/18/2019	168.9	7	1182.7	
5	Customer 1	Avanza AT	SBY	8/21/2019	168.9	9	1520.5	
6	Customer 2	Etios AT	KDR	09/07/2019	800	7	5599.6	
7	Customer 3	Etios AT	MLG	09/07/2019	800	8	6399.6	
8	Customer 1	Avanza AT	SBY	9/14/2019	168.9	15	2534.2	
9	Customer 1	Etios AT	SBY	9/25/2019	800	2	1599.9	
10	Customer 1	Avanza AT	SBY	1/21/2020	168.9	14	2365.3	
11	Customer 1	Yaris AT	MLG	02/01/2020	340.9	2	681.9	
12	Customer 1	Etios AT	SBY	5/30/2020	800	9	7199.6	
13	Customer 3	Avanza AT	MLG	06/07/2020	168.9	10	1689.5	
14	Customer 3	Yaris AT	MLG	06/09/2020	340.9	4	1363.8	
15	Customer 2	Avanza AT	KDR	08/08/2020	168.9	12	2027.4	
16								

Tugas 3

1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?

- File **CSV** (Comma-Separated Values) menyimpan data dalam format teks sederhana dan setiap nilai dipisahkan oleh koma atau delimiter lain seperti titik koma.
- File **Excel** (XLSX/XLS) menyimpan data dalam format yang lebih terstruktur untuk tabel, grafik, warna, formula, dan fitur lainnya
- **Perbedaan :**
 - Data dalam **CSV sulit dibaca** secara langsung karena tidak memiliki pemformatan tabel atau gaya visual seperti warna dan garis.
 - File **excel** lebih **mudah** dibaca dan dianalisis.

2. Jelaskan apa yang terjadi pada proses Filter rows!

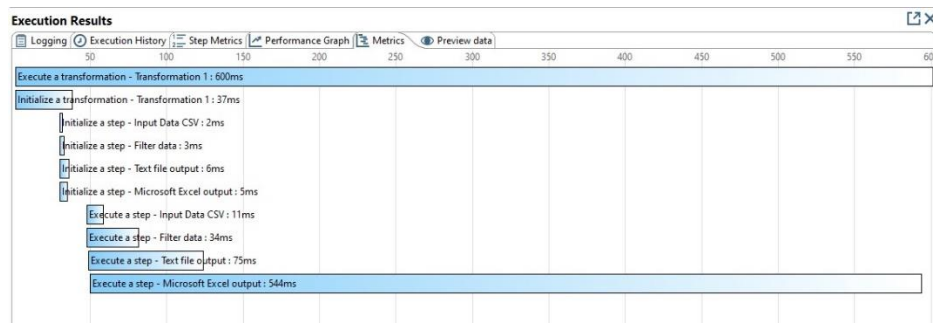
- Pada tahap **Filter Rows**, data yang diekstrak dari akan disaring sesuai kriteria
- **Setiap baris dalam dataset diperiksa** satu per satu.
 - Jika ada baris yang mengandung nilai NULL dalam kolom yang ditentukan (misalnya "Customer", "Product", "Region", dll.), maka baris tersebut akan dibuang.
 - Hanya baris yang memenuhi semua syarat ("IS NOT NULL") yang akan dipertahankan dalam dataset.

3. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract Transfer Load data ini? Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU). Bandingkan dalam bentuk tabel

Perbandingan 1

Execution Results

Execution Results													
Logging Execution History Step Metrics Performance Graph Metrics Preview data													
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Input Data CSV	0	0	19	20	0	0	0	0	Finished	0.0s	1,818	-
2	Filter data	0	19	19	0	0	0	0	0	Finished	0.0s	543	-
3	Microsoft Excel output	0	5	5	0	5	0	0	0	Finished	0.5s	9	-
4	Text file output	0	14	14	0	15	0	0	0	Finished	0.1s	195	-



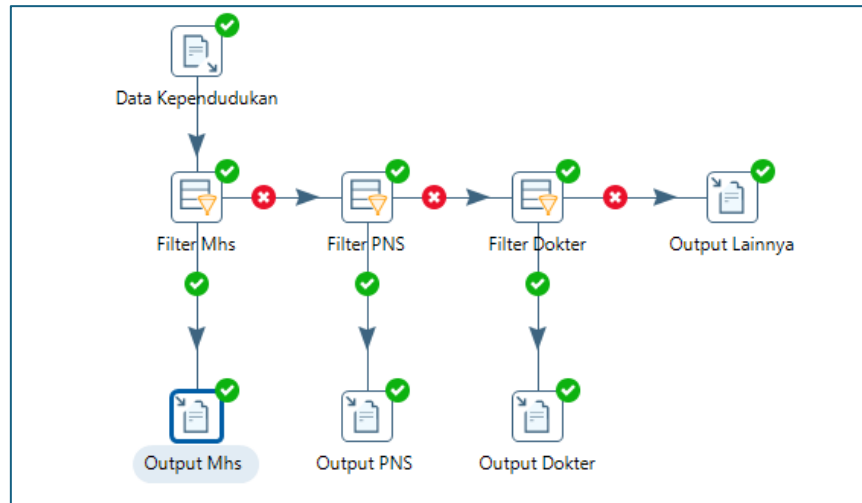
Perbandingan 2

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Input Data CSV	0	0	19	20	0	0	0	0	Finished	0.0s	6,667	-
2	Filter data	0	19	19	0	0	0	0	0	Finished	0.0s	3,167	-
3	Microsoft Excel output	0	5	5	0	5	0	0	0	Finished	0.1s	64	-
4	Text file output	0	14	14	0	15	0	0	0	Finished	0.0s	1,364	-

Perbandingan 1	Perbandingan 2
<ul style="list-style-type: none"> - Processor : Intel (R) Core (TM) i5-8250U CPU @1.60GHz (8 CPUs), ~1.8GHz - RAM : 8GB DDR3 L Memory - Storage : 1000 HDD - VGA : NVIDIA® GeForce® MX150 2 GB Intel® UHD Graphics 620 	<ul style="list-style-type: none"> - Processor : AMD Ryzen 7 7730U - RAM : 16GB - Storage : 512 GB

C. Studi Kasus

Filter Data Kependudukan Berdasarkan Pekerjaan



1. Input Data Dummy Kependudukan

CSV file input

Step name: Input File Data Kependudukan

Filename: F:\KULIAH\SEMESTER 4\3. PRAK JOBSHEET\data warehouse\29_Vita Eka Saraswati\Jobsheet\1\data

Delimiter: ,

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result? ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding: UTF-8

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	Nama	String		16		IDR	.	.	none
2	NIK	Integer	#	15	0	IDR	.	.	none
3	Tempat Lahir	String		11		IDR	.	.	none
4	Tanggal Lahir	Date	dd-MM-yyyy			IDR	.	.	none
5	Jenis Kelamin	String		9		IDR	.	.	none
6	Alamat	String		28		IDR	.	.	none
7	Kota	String		11		IDR	.	.	none
8	RT/RW	String		7		IDR	.	.	none
9	Kelurahan	String		18		IDR	.	.	none
10	Kecamatan	String		17		IDR	.	.	none

Help OK Get Fields Preview Cancel

Activate Windows
Go to Settings to activate Windows.

2. Melakukan Setting Filter Rows

○ Filter untuk Mahasiswa

Filter rows

Step name: Filter Mhs

Send 'true' data to step: Output Mhs

Send 'false' data to step: Filter PNS

The condition:

☐ Pekerjaan LIKE Mahasiswa (String)

- Filter untuk PNS

Filter rows

Step name: **Filter PNS**

Send 'true' data to step: **Output PNS**

Send 'false' data to step: **Filter Dokter**

The condition:

Pekerjaan **LIKE** **PNS** (String)

- Filter untuk Dokter

Filter rows

Step name: **Filter Dokter**

Send 'true' data to step: **Output Dokter**

Send 'false' data to step: **Output Lainnya**

The condition:

Pekerjaan **LIKE** **Dokter** (String)

3. Output File CSV berdasarkan filter pekerjaan

- Data Penduduk Mahasiswa

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan/Kecamatan	Agama	Status Per	Pekerjaan		
2	Siti Rahayu	3503010101010000	Bandung	15 May 1990	Perempuan	Jl. Cendrawasih	Bandung	003/004	Cikutra	Bandung	Kristen	Belum Menikah	Mahasiswa	
3	Adi Pratama	3503010101010000	Semarang	05 March 1995	Laki-laki	Jl. Pemuda	Semarang	009/010	Pandanan	Semarang	Buddha	Belum Menikah	Mahasiswa	
4	Maya Wulandari	3503010101010020	Jayapura	23 October 1993	Perempuan	Jl. Proklamasi	Jayapura	039/040	Kota Jayapura	Jayapura	Kristen	Belum Menikah	Mahasiswa	

- Data Penduduk PNS

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan/Kecamatan	Agama	Status Per	Pekerjaan		
2	Andi Wijaya	3503010101010000	Makassar	08-04-1993	Laki-laki	Jl. Jenderal Sudirman	Makassar	013/014	Barombong	Makassar	Islam	Menikah	PNS	
3	Agus Santia	3503010101010000	Padang	25-09-1980	Laki-laki	Jl. Sudirman	Padang	029/030	Padang Besar	Padang	Islam	Menikah	PNS	

- Data Penduduk Dokter

outputStudiKasusDokter - Notepad

File Edit Format View Help

Nama;NIK;Tempat Lahir;Tanggal Lahir;Jenis Kelamin;Alamat;Kota;RT/RW;Kelurahan;Kecamatan

Desi Purnama ;3503010101010008;Palembang ;08-04-1993;Perempuan;Jl. Demang Lebar Daun

Bambang Kusumo ;3503010101010017;Banjarmasin;25-09-1980;Laki-laki;Jl. A. Yani No. 60

- Data Penduduk Lainnya

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan/Kecamatan	Agama	Status Per	Pekerjaan		
2	Budi Santia	3503010101010000	Jakarta	01-01-1990	Laki-laki	Jl. Merdeka	Jakarta	001/002	Menteng	Jakarta	Islam	Menikah	Wiraswasta	
3	Joko Susilo	3503010101010000	Surabaya	05-05-1990	Laki-laki	Jl. Diponegoro	Surabaya	005/006	Gubeng	Surabaya	Katolik	Menikah	Pegawai Swasta	
4	Maya Dewi	3503010101010000	Yogyakarta	07-07-1990	Perempuan	Jl. Malioboro	Yogyakarta	007/008	Sosromen	Yogyakarta	Hindu	Belum Menikah	Guru	
5	Rini Setiawati	3503010101010000	Malang	01-11-1991	Perempuan	Jl. Simpar	Malang	011/012	Dinoyo	Malang	Kristen	Menikah	Wiraswasta	
6	Fajar Pratomo	3503010101010000	Medan	01-07-1991	Laki-laki	Jl. Sudirman	Medan	017/018	Sei Sikam	Medan	Hindu	Menikah	Wiraswasta	
7	Rina Sari	3503010101010010	Denpasar	01-09-1991	Perempuan	Jl. Gatot Subroto	Denpasar	019/020	Dauh Puri	Denpasar	Buddha	Belum Menikah	Freelancer	
8	Ahmad Suherman	3503010101010010	Surakarta	02-11-1991	Laki-laki	Jl. Pahlawan	Surakarta	021/022	Jebres	Surakarta	Islam	Belum Menikah	Wiraswasta	
9	Dian Pertiwi	3503010101010010	Maluku	03-02-1992	Perempuan	Jl. Pattimura	Ambon	023/024	Baguala	Ambon	Kristen	Menikah	Guru	
10	Rudi Hartono	3503010101010010	Pontianak	02-05-1992	Laki-laki	Jl. Ahmad Yani	Pontianak	025/026	Pontianak	Pontianak	Hindu	Menikah	Wiraswasta	
11	Lina Fitriani	3503010101010010	Manado	02-07-1992	Perempuan	Jl. Wolter Monginsidi	Manado	027/028	Mapanget	Manado	Kristen	Belum Menikah	Akuntan	
12	Rina Agus	3503010101010010	Bengkulu	03-11-1992	Perempuan	Jl. Imam Bonjol	Bengkulu	031/032	Ratu Sami	Bengkulu	Kristen	Menikah	Wiraswasta	
13	Nia Dewi Irena	3503010101010010	Manokwari	03-05-1993	Perempuan	Jl. Diponegoro	Manokwari	035/036	Manokwari	Manokwari	Kristen	Belum Menikah	Penulis	
14	Doni Setia	3503010101010010	Palu	03-07-1993	Laki-laki	Jl. Soekarno	Palu	037/038	Palu Bara	Palu	Islam	Menikah	Wiraswasta	