# LAPORAN TUGAS DATA MINING PERTEMUAN 2

Dosen Pengampu: Vipkas Al Hadid Firdaus, S.T., M.T.



Muhammad Afiq Firdaus / 14 SIB-2B / 2341760189

# PROGRAM STUDI D-IV SISTEM INFORMASI BISNIS JURUSAN TEKNOLOGI INFORMASI POLITEKNIK NEGERI MALANG 2025

# Jobsheet1 Data Warehouse sederhana dan pengenalan PDI Spoon

# Tujuan Praktikum

Setelah melakukan praktikum ini, mahasiswa diharapakan dapat lebih mengenal apa itu data warehouse berdasarkan uji coba membuat data warehouse sederhana

### Studi Kasus

Pak Amir adalah Manager Sales Astro Mobil, yang bergerak di bidang distributor mobil wilayah Jawa Timur. Pak Amir meminta data penjualan dari beberapa cabang kepada supervisor. Data tersebut akan digunakan untuk membuat suatu Dashboard pengambilan keputusan. Dikarenakan proses permintaan ini dilakukan secara berulang setiap hari setelah jam kantor maka, Pak Amir akan membuat proses untuk menarik data yang ada pada file milik supervisor tersebut secara otomatis. Namun, kadang data tersebut tidak lengkap. Sehingga Pak Amir akan mengambil data yang lengkap saja dan mengembalikan data yang tidak lengkap.

Dari studi kasus tersebut maka, akan dilakukan:

- 1. Cek dan Analisa data penjualan.
- 2. Import data dari file (Extract)
- 3. Identifikasi data yang tidak lengkap (missing data) dan meletakkan pada file yang berbeda (Transform)
- 4. Memindahkan data yang sudah lengkap ke file dashboard. (Load)
- 5. Mengumpulkan data yang belum lengkap untuk dikembalikan.

Tugas 1
Analisa lah data tersebut!

	Α	В	С	D	E	F	G	Н
1	Customer	,Product,R	egion,Date	e, Item uni	t price , No	.Items , To	otal Sale	
2	Customer	1,Yaris AT,	,MLG,3/13/	2019,340.9	5,6,2045.7			
3	Customer	1,Etios AT	,SBY,4/16/2	2019,799.9	5,1,799.95			
4	Customer	3,Avanza	AT,MLG,6/1	18/2019,16	8.95,7,1182	2.65		
5	Customer	3,,MLG,7/	24/2019,16	8.95,2,337.	.9			
6	Customer	1,Avanza	AT,SBY,8/2	1/2019,168	.95,9,1520.	.55		
7	Customer	2,Etios AT	KDR,9/7/2	019,799.95	,7,5599.65			
8	Customer	3,Etios AT	,MLG,9/7/2	2019,799.95	5,8,6399.6			
9	Customer	2,Avanza	AT,,,168.95	,15,2534.2	5			
10	Customer	1,Avanza	AT,SBY,9/1	4/2019,168	.95,15,253	4.25		
11	Customer	1,Etios AT	,SBY,9/25/2	2019,799.9	5,2,1599.9			
12	Customer	1,Avanza	AT,SBY,1/2	1/2020,168	.95,14,236	5.3		
13	Customer	1,Yaris AT,	MLG,2/1/2	020,340.95	,2,681.9			
14	Customer	1,,,,340.95	,7,2386.65					
15	Customer	1,Etios AT	,SBY,5/30/2	2020,799.9	5,9,7199.55	5		
16	Customer	3,Avanza	AT,MLG,6/7	7/2020,168	.95,10,1689	9.5		
17	Customer	3,Yaris AT,	MLG,6/9/2	020,340.95	,4,1363.8			
18	Customer	2,,KDR,8/7	7/2020,799.	95,14,1119	9.3			
19	Customer	2,Avanza	AT,KDR,8/8	/2020,168.	95,12,2027	.4		
20	Customer	1,Avanza	AT,,8/12/20	20,168.95,	4,675.8			

1. Berapa jumlah kolom pada data tersebut?

Jawaban: 7 kolom

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

### Jawaban:

Kolom  $1 \rightarrow$  Customer  $\rightarrow$  Nomor Customer

Kolom 2 → Product → Nama Produk Mobil

Kolom 3 → Region → Nama Daerah

Kolom  $4 \rightarrow Date \rightarrow Tanggal$ 

Kolom 5 → Item Unit Price → Harga per unit

Kolom  $6 \rightarrow \text{No Item} \rightarrow \text{Nomor Item}$ 

Kolom  $7 \rightarrow \text{Total Sale} \rightarrow \text{Total}$ 

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

# Jawaban:

Ada pada

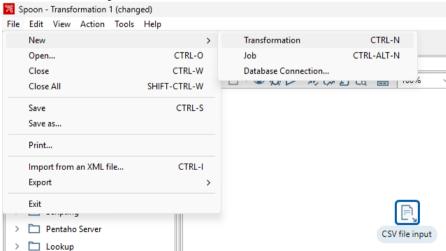
- Line 5 kolom Product
- Line 9 kolom Region, Date
- Line 14 kolom Product, Region, Date
- Line 18 kolom Product,
- Line 20 kolom Region
- 4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

# Jawaban:

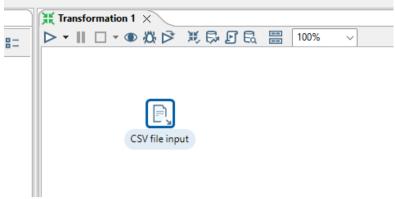
Dalam satu kolom setiap data memiliki tipe data yang sama dan memiliki data yang kembar.

# A. Pengambilan Data (Extract)

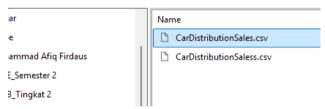
1. Buka lembar kerja **Transforamtion** melalui toolbar **File** 



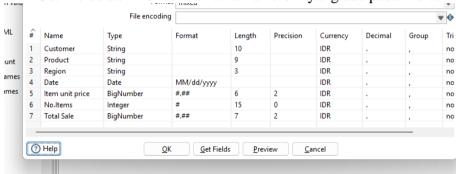
2. Cari lah objek CSV file input pada Design Area. Drag and drop objek tersebut menuju Work Area.



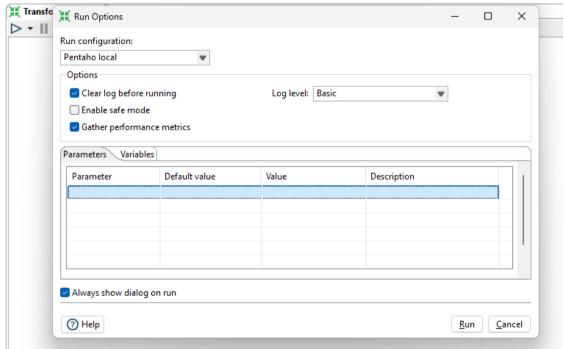
3. Memasukkan File CarDistributionSales.csv



4. Klik Get Fields dan muncul nama nama kolom yang ada pada file csv tersebut.



5. Run file



6. Hasil Execute berikut

### **Execution Results**

# Tugas 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

# Jawaban:

Sudah sesuai dengan file aslinya, data yang NULL tidak kosong tetapi terdapat keterangan NULL

2. PDI Spoon melakukan proses extract, perhatikan pada Tab Logging di Execution Results Area,

langkah-langkah apa sajakah yang dilakukan PDI Spoon untuk melakukan extract data?

# Jawaban:

```
2025-02-26 21:21:30.540 - Spoon - Save file as...
2025-02-26 21:21:31.007 - Spoon - Transformation opened.
2025-02-26 21:21:31.007 - Spoon - Launching transformation [Transformation 1]...
2025-02-26 21:21:31.009 - Spoon - Started the transformation execution.
2025-02-26 21:21:32.145 - Transformation 1 - Dispatching started for transformation [Transformation 1]
2025-02-26 21:21:32.226 - Carte - Installing timer to purge stale objects after 1440 minutes.
2025-02-26 21:21:32.338 - CSV file input.0 - Header row skipped in file 'D:\Muhammad Afiq
Firdaus\2B_Tingkat 2\Semester 2\Data Warehouse\Pertemuan 2\CarDistributionSales.csv'
2025-02-26 21:21:32.344 - CSV file input.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)
2025-02-26 21:21:32.351 - Spoon - The transformation has finished!!
```

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

# Jawaban:

 $I=20 \rightarrow Input 20$ 

 $O=0 \rightarrow Output 0$ 

 $R=0 \rightarrow Read 0$ 

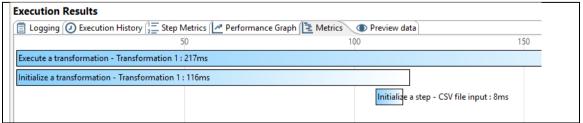
 $W=19 \rightarrow Written 19$ 

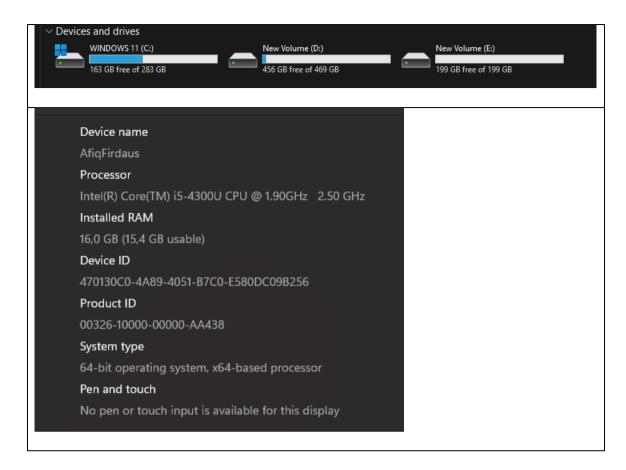
 $U=0 \rightarrow Updated 0$ 

 $E=0 \rightarrow Errors 0$ 

4. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini? (Processor, RAM, Storage, VGA, CPU).

# Jawaban:





# B. Filter Data (Transform) dan pengemasa data (Load)

# **Praktikum**

1. Masukkan Filter Rows kedalam file



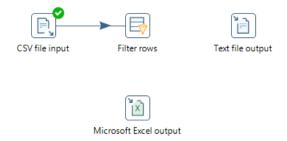
2. Sambungkan tombol output pada Input Data CSV ke Filter rows kemudian klik Main output of step



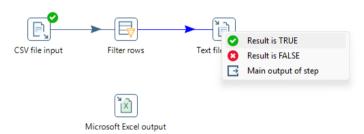
3. Tambahkan Text File Output



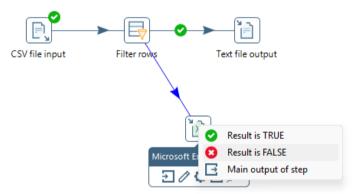
4. Tambahkan juga Microsoft Excel Output



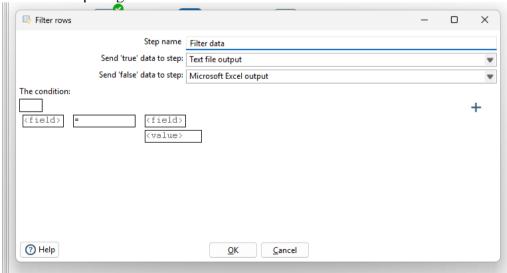
5. Sambungkan Output dari Filter Rows menuju Text File Output lalu klik TRUE



6. Sambungkan Output dari Filter Rows menuju Microsoft Excel Output lalu klik FALSE



7. Klik 2x pada Text File Output lalu ubah nama step name dan pastikan send true dan send false seperti gambar di bawah.

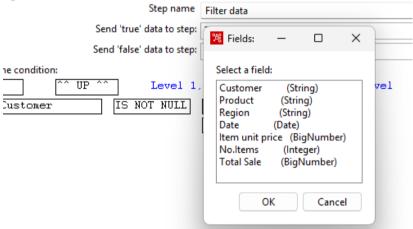


Klik <field> lalu ubah menjadi Customer kemudian kotak disebelah Customer diisi IS NOT NULL

. , ,	711022		
		Step name	Filter data
	Send 'true	data to step:	Text file output
	Send 'false	data to step:	Microsoft Excel output
	The condition:  Customer IS N	OT NULL	- -

8. Klik tombol Plus untuk membuat kondisi baru, tampilan akan seperti dibawah ini. Lakukan seperti Langkah sebelumnya akan tetapi <field> diisi berurutan ke urutan selanjutnya setelah Customer.

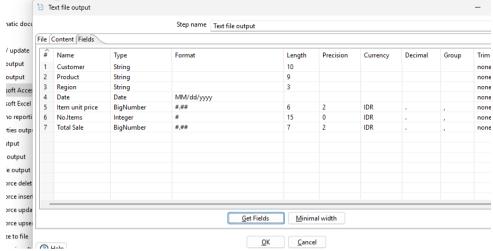
Ketika di klik OK akan terbentuk kode seperti berikut yang disambungkan dengan operator AND.



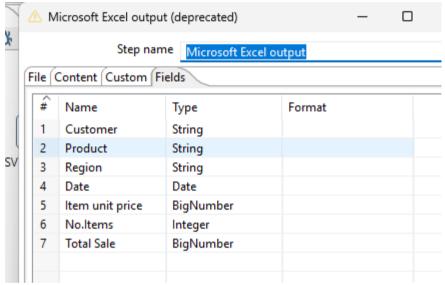
9. Klik Text File Output 2x, lalu ubah File Name seperti berikut.

i Text file output	
Step name	Text file output
File Content Fields	
Filename	file
Pass output to servlet	
Create Parent folder	
Do not create file at start	
Accept file name from field?	
File name field	
Extension	txt
Include stepnr in filename?	
Include partition nr in filename?	
Include date in filename?	
Include time in filename?	
Specify Date time format	
Date time format	
	Show filename(s)
Add filenames to result	
① Help	OK Cancel

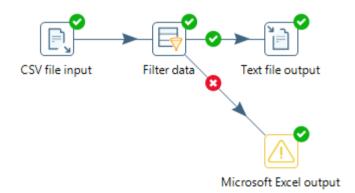
10. Klik bagian sheet Fields, lalu klik Get Fields.



11. Klik Microsoft Excel Output kemudian ubah seperti gambar dibawah ini dan klik Get Fields.



12. Klik RUN. Kemudian akan tergambar cetang hijau semua.



```
2025-02-27 08:45:42.628 - Spoon - Transformation opened.
2025-02-27 08:45:42.628 - Spoon - Launching transformation [Transformation 1]...
2025-02-27 08:45:42.628 - Spoon - Spoon - Started the transformation execution.
2025-02-27 08:45:42.628 - Spoon - Started the transformation execution.
2025-02-27 08:45:42.792 - CSV file input.0 - Header row skipped in file 'D:\Muhammad Afiq Firdaus\2B_Tingkat 2\Semester 2\Data Warehouse\Pertemuan 2\CarDistributionSales.csv'
2025-02-27 08:45:42.797 - CSV file input.0 - Finished processing (I=20, 0=0, R=0, W=19, U=0, E=0)
2025-02-27 08:45:42.848 - Filter data.0 - Finished processing (I=0, 0=19, W=19, U=0, E=0)
2025-02-27 08:45:42.901 - Text file output.0 - Finished processing (I=0, 0=15, R=14, W=14, U=0, E=0)
2025-02-27 08:45:43.022 - Microsoft Excel output.0 - Finished processing (I=0, 0=5, R=5, W=5, U=0, E=0)
2025-02-27 08:45:43.022 - Microsoft Excel output.0 - Finished processing (I=0, 0=5, R=5, W=5, U=0, E=0)
```

# 13. File outputSalesCarDistribution.csv

Cı	ısto	mer,Produ	ct,Region,	Date, Iten	n unit pric	e , No.Items	, Total Sa	le	
	4	Α	В	С	D	Е	F	G	
	1	Customer	Product,R	egion,Dat	e, Item ur	it price , No	.Items , To	otal Sale	
	2	Customer	1,Yaris AT,	,MLG,3/13	/2019,340	.95,6,2045.7			
	3	Customer	1,Etios AT	,SBY,4/16/	2019,799.	95,1,799.95			
	4	Customer	3,Avanza	AT,MLG,6/	18/2019,1	68.95,7,1182	2.65		
	5	Customer	3,,MLG,7/	24/2019,16	8.95,2,33	7.9			
	6	Customer	1,Avanza	AT,SBY,8/2	1/2019,16	8.95,9,1520.	.55		
	7	Customer	2,Etios AT	KDR,9/7/2	2019,799.9	5,7,5599.65			
	8	Customer	3,Etios AT	,MLG,9/7/	2019,799.9	95,8,6399.6			
	9	Customer	2,Avanza	AT,,,168.95	,15,2534.:	25			
	10	Customer	1,Avanza	AT,SBY,9/1	4/2019,16	8.95,15,253	4.25		
	11	Customer	1,Etios AT	,SBY,9/25/	2019,799.	95,2,1599.9			
	12	Customer	1,Avanza	AT,SBY,1/2	1/2020,16	8.95,14,236	5.3		
	13	Customer	1,Yaris AT	,MLG,2/1/	2020,340.9	5,2,681.9			
	14	Customer	1,,,,340.95	,7,2386.65					
	15	Customer	1,Etios AT	,SBY,5/30/	2020,799.	95,9,7199.55	i		
	16	Customer	3,Avanza	AT,MLG,6/	7/2020,16	8.95,10,1689	.5		
	17	Customer	3,Yaris AT,	,MLG,6/9/	2020,340.9	5,4,1363.8			
	18	Customer	2,,KDR,8/7	7/2020,799	.95,14,111	99.3			
	19	Customer	2,Avanza	AT,KDR,8/8	3/2020,16	3.95,12,2027	.4		
	20	Customer	1,Avanza	AT,,8/12/2	020,168.9	5,4,675.8			
	21								

# 14. Isi file failedSalesCarDistribution.xls

4	Α	В	С	D	Е	F	G	Н
1	Customer	Product	Region	Date	Item unit p	No.ltems	Total Sale	
2	Customer		MLG		168,95	2,	337,9	
3	Customer	Avanza AT			168,95	15,	2534,25	
4	Customer				340,95	7,	2386,65	
5	Customer		KDR		799,95	14,	11199,3	
6	Customer	Avanza AT			168,95	4,	675,8	
7								
^								

# Tugas 3

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

# Jawaban:

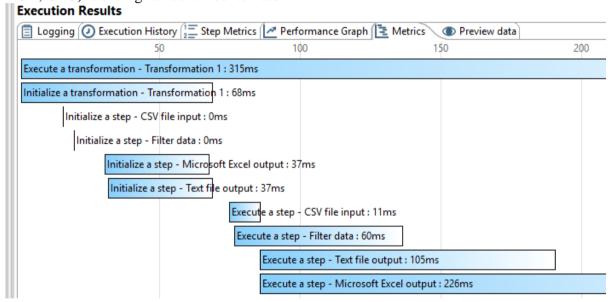
- outputSalesCarDistribution.csv Berisi hasil dari kondisi yang benar yaitu tidak ada satupun cell yang bernilai null
- failedSalesCarDistribution.xls
   Berisi hasil dari kondisi yang salah yaitu semua row memiliki nilai null didalamnya
- 2. Jelaskan apa yang terjadi pada proses Filter rows!

# Jawaban:

1) Filter rows disambungkan dengan Input CSV yang bernilai file raw asli dari

CarDistribution. File rows berada di Tengah input dan output sebagai alat untuk melakukan filterisasi hasil yang diinginkan

- 2) Setelah data dari input dikirimkan ke Filter Rows, filter rows melakukan tugasnya yaitu melakukan Filterisasi Row sesuai kondisi yang di masukkan. Pada kasus diatas kita membuat kondisi yaitu, Kondisi benilai true adalah semua row tanpa nilai NULL(terisi semua). Sedangkan kondisi false adalah semua row yang memiliki nilai NULL di dalamnya
- 3) Setelah melakukan proses filter, Filter row menyajikan hasil filterisasinya menjadi 2 file TRUE and FALSE
- 3. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract Transfer Load data ini? identifikasi pula spesifikasi perangkat yang anda gunakan (Processor, RAM, Storage, VGA, CPU). bandingkan dalam bentuk table



# Studi Kasus

Pak Pascanowo seorang ketua RW di Ibu Kota Nusabangsa (IKN). Terdapat 20 orang dari negara Konoha yang akan pindah dan berdomisili di IKN. Tata letak pemukiman pada IKN telah dikondisikan agar berkelompok disesuaikan dengan lokasi pekerjaannya agar akses terjangkau.

Sebagai contoh, mahasiswa akan dikumpulkan dan berdomisili di wilayah edukasi yang akses dekat dengan kampus-kampus perguruan tinggi. Sedangkan dokter akan berdomisili dekat dengan rumah sakit atau layanan kesehatan dan PNS akan berdomisili dekat dengan perkantoran dan layanan publik. Selain itu, orang yang bukan dari ketiga kategori tersebut bebas memilih domisili atau tempat tinggalnya di IKN.

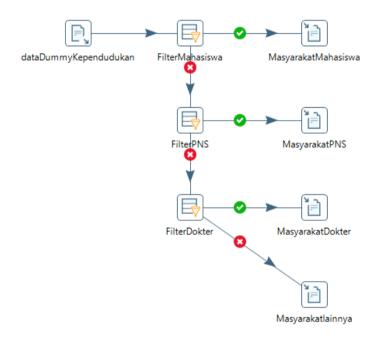
Pak Pascanowo akan mendapatkan data dari pemerintah pusat tentang Masyarakat Konoha yang akan pindah ke IKN. darai data tersebut Pak Pascanowo akan membagi menjadi 4 kelompok akan diberikan ke Pak Emir Makarena selaku ketua RT wilayah Pendidikan akan

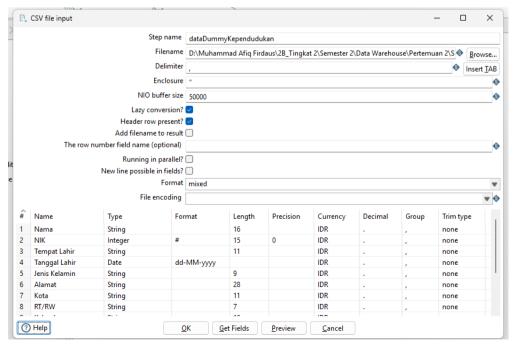
diberikan ke Pak Yasana Laili selaku ketua RT wilayah pemerintahan akan diberikan ke Pak Budi Pekerti selaku ketua RT wilayah kesehatan

data-data tersbut akan digunakan untuk analisa di masing-masing wilayah dan juga Dashboard untuk emmbantu mengambil keputusan.

Bantulah Pak Pascanowo untuk memishkan data tersebut menjadi 4 file: MasyarakatMahasiswa, MasyarakatPNS, MasyarakatDokter, MasyarakatLainnya.

Dikarenakan proses tersebut akan repetisi setiap ada masyarakat Konoha yang berpindah ke IKN maka buatlah sistem tersebut menggunakan PDI Spoon.





# 

# MasyarakatDokter

Α	В	С	D	E	F	G	Н	1	J	K	L	М	N
Nama	NIK	Tempat La	Tanggal La	Jenis Kela	Alamat	Kota	RT/RW	Kelurahan	Kecamata	Agama	Status Per	Pekerjaan	
Desi Purna	3,5E+15	Palemban	########	Perempua	Jl. Deman	Palemban	015/016	Alang-Ala	Palemban	Kristen	Belum Me	Dokter	
Bambang	3,5E+15	Banjarmas	**********	Laki-laki	Jl. A. Yani	Banjarmas	033/034	Banjarmas	Banjarmas	Hindu	Menikah	Dokter	

# MasyarakatLainnya



# MasyarakatMahasiswa



# **MasyarakatPNS**

4	А	В	С	D	E	F	G	н	1	J	K	L	М	N	
1	Nama	NIK	Tempat La	Tanggal La	Jenis Kela	Alamat	Kota	RT/RW	Kelurahar	Kecamata	Agama	Status Per	Pekerjaan		Τ
2	Andi Wija	3,5E+15	Makassar	########	Laki-laki	Jl. Jendera	Makassar	013/014	Barombor	Makassar	Islam	Menikah	PNS		
3	Agus Sant	3,5E+15	Padang	########	Laki-laki	Jl. Sudirm	Padang	029/030	Padang Ba	Padang	Islam	Menikah	PNS		
4															