PRACTICUM DATA WAREHOUSE

'Jobsheet 1 – Input Data'

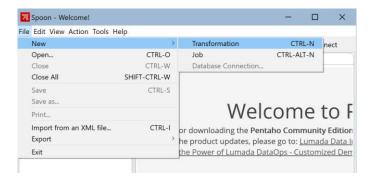


Name: Sabrina Rahmadini

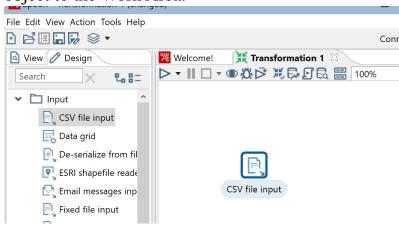
Student ID: 2341760155

Class: SIB 2G

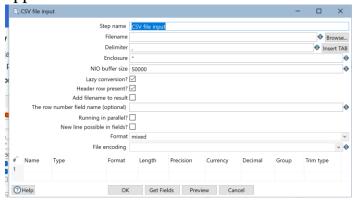
• Open the Transformation worksheet via the File toolbar.



 Locate the CSV input file object in the Design Area. Drag and drop the object to the Work Area.

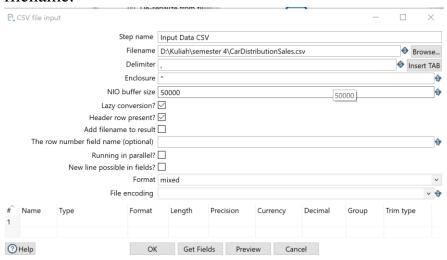


 Double-click on the CSV input file object until the configuration window appears.

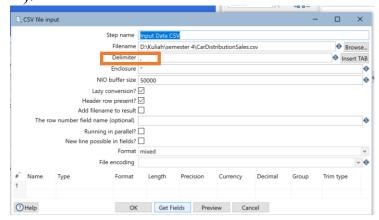


• Change the Step name to CSV Input Data, (naming is used for identification only, further naming can be adjusted).

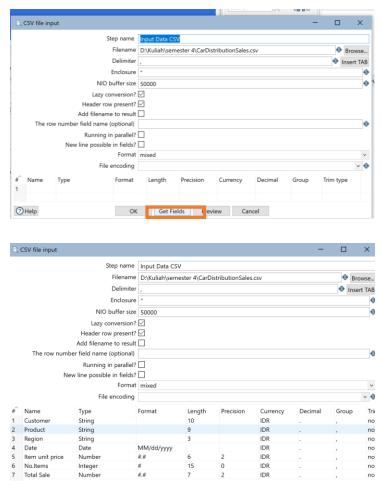
• Select the location of the CarDistributionSales.csv file through Browse on filename.



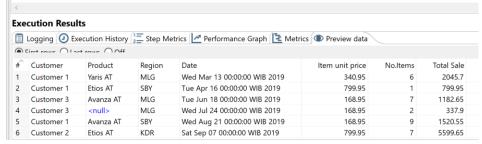
• If the data has been selected, make sure that the Delimiter section matches the data separator conditions in the selected file. (In this case it is comma ", ").



 Press Get Fields to retrieve the title of each column, make sure the column names are appropriate.



- Close the configuration window.
- Press the "Run" button on the top right corner of the Work Area, and the Execution Result Area will appear. Select the Prieview Data tab in the Execution Result Area. If the process is done correctly, the data will appear according to the original.



Jika langkah-langkah diatas telah berhasil, maka proses extract telah berhasil dilakukan.

TUGAS 2

- 1. Apakah data hasil eksekusi sesuai dengan data aslinya?
 - → Ya, sudah sesuai
- 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?
 - → Membuka koneksi ke sumber data
 - → Mengeksekusi atau membaca data
 - → Menjalankan transformasi
 - → Transformasi
 - → Mengirim data
 - → Menampilkan data
- 3. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E?

```
2024/02/17 06:51:48 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)
```

- → I (Input): Jumlah baris data yang diterima sebagai input oleh langkah tersebut.
- → O (Output): Jumlah baris data yang dikeluarkan oleh langkah tersebut ke langkah berikutnya.
- → R (Read): Jumlah baris data yang dibaca dari sumber data.
- → W (Written): Jumlah baris data yang berhasil ditulis atau diproses oleh langkah tersebut.
- → U (Updated): Jumlah baris data yang diperbarui.
- → E (Error): Jumlah baris data yang mengalami kesalahan selama pemrosesan.
- 4. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini?

```
Execution Results

| Logging | Execution History | Logging | Execution History | Logging | Step Metrics | Step Metrics | Logging | Step Metrics | Logging | Logging | County | Logging | County | Logging | County | Logging | Log
```

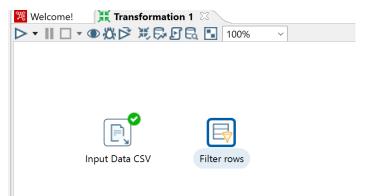
→ 1 detik

5. Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU). Bandingkan dalam bentuk tabel.

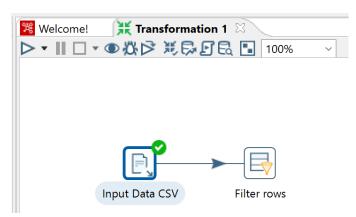
No	Nama	Waktu	Prosessor	RAM	Storage	VGA	CPU
1	Cindy	6s	Intel i5- 10300H	12 GB	HDD 1 TB	NVIDIA	60%
						GTX 1050	
2	Sabrina	1s	intel i5-7300U	8192	Samsung	Intel ®	50%
				MB(8GB)	239 GB	graphics	
						620	
3	nada	1s	11 th gen	16 GB	SSD	Intel®Iris	23%
			intel@core TM i5-		NVMe	® Xe	
					477GB	graphics	
4	zahra	1s	Intel®core	16 GB	SSD	Intel ®	7%
			тмі7-		477GB	UHD	
			13650HX(13 th			graphics	
			Gen, 20 Cores)			+	
						NVIDIA	
						GeForce	
						RTX 4050	

Filter Data (Transform) dan Pengemasan data (Load)

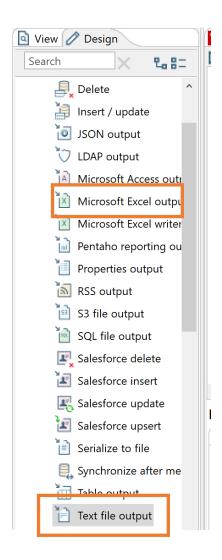
• Carilah objek **Filter rows** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.



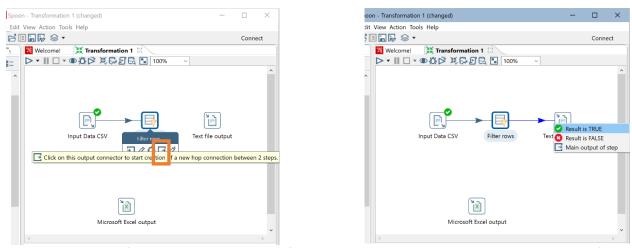
• Hubungkan output pada **Input Data CSV** menuju **Filter rows**. Pilih koneksi **Main output of step** saat menghubungkan kedua objek tersebut.



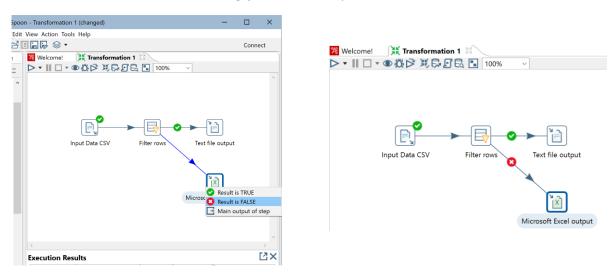
Carilah objek Text file output dan Microsoft excel output pada Design
Area. Drag and drop objek tersebut pada Work Area.



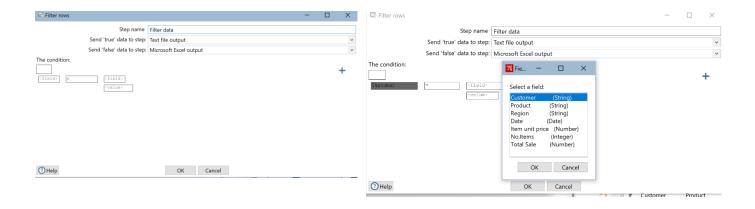
Hubungkan filter rows dengan text file output menggunakan konektor output dari filter rows.
 Pilih Result is true sebagi jenis konektornya.

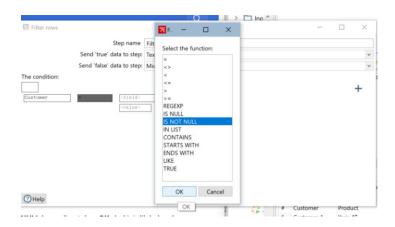


• Hubungkan filter rows dengan microsoft excel output menggunakan konektor output dari filter rows. Pilih Result is false sebagi jenis konektornya.

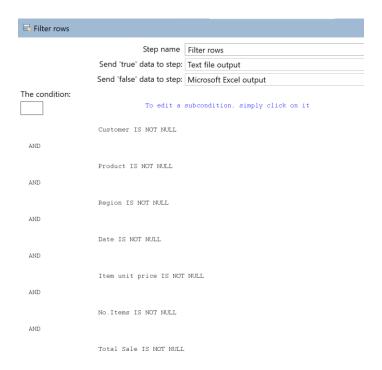


- Double click pada filter rows hingga muncul jendela konfigurasi. ubah step name menjadi Filter
 data
 - Cek bagian **send 'true' data to step** menjadi **Text file output** dan cek bagian **Send 'false' data to step** menjadi **Microsoft Excel output**.
- Klik <Field> pada bagian condition, Pilih field yang diinginkan untuk digunakan sebagai kondisi.
 Pada kesempatan ini pilih Customer, kemudian klik OK.

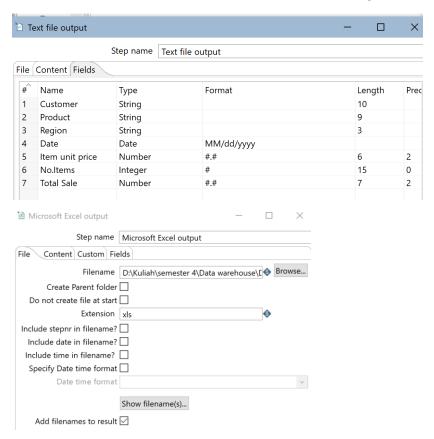


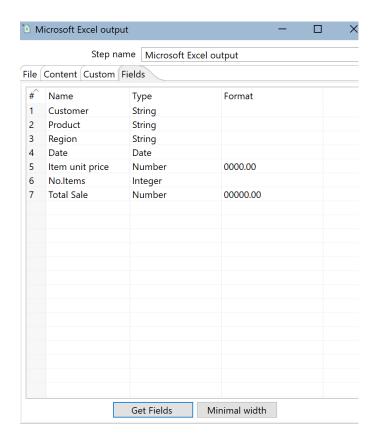


- Tekan tombol + untuk menambah kondisi. Pilih **statement** yang masih null untuk menambah kondisi. Ulangi proses nomor 9,10,11 dengan mengganti semua **field** yang ada. hingga statement menjadi seperti pada gambar dibawah ini.
- Gunakan operator AND untuk menghubung

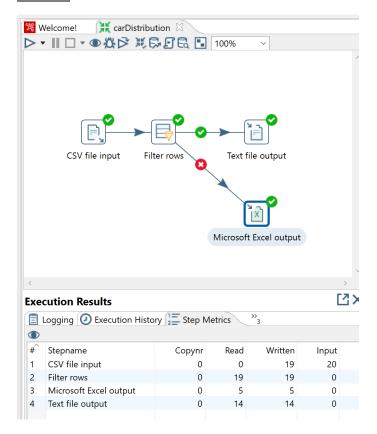


• Pada tab Fields, tekan tombol Get fields untuk mengambil field data.

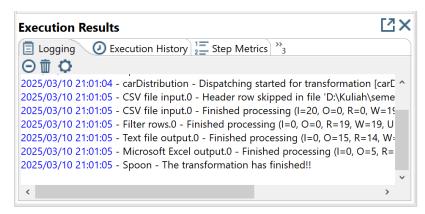




TUGAS 3



Apabila sudah berhasil di running



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

 \rightarrow csv:

. q : t	A1 \checkmark : $(\times \checkmark f_x)$ Customer									
0		А	В	С	D	Е	F	G	Н	
	1	Customer	Product	Region	Date	ltem unit p	No.ltems	Total Sale		
٥	2	Customer 1	Yaris AT	MLG	3/13/2019	340.9	6	2045.7		
þ	3	Customer 1	Etios AT	SBY	4/16/2019	800	1	800		
a	4	Customer 3	Avanza AT	MLG	6/18/2019	168.9	7	1182.7		
a	5	Customer 1	Avanza AT	SBY	8/21/2019	168.9	9	1520.5		
1	6	Customer 2	Etios AT	KDR	########	800	7	5599.6		
Р	7	Customer 3	Etios AT	MLG	#########	800	8	6399.6		
d	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		
2	11	Customer 1	Yaris AT	MLG	########	340.9	2	681.9		
٨	12	Customer 1	Etios AT	SBY	5/30/2020	800	9	7199.6		
s	13	Customer 3	Avanza AT	MLG	#########	168.9	10	1689.5		
]	14	Customer 3	Yaris AT	MLG	#########	340.9	4	1363.8		
	15	Customer 2	Avanza AT	KDR	#########	168.9	12	2027.4		
۵	4.0									

XIs

→ File xls berisi data yang ada nilai null atau tidak lengkap

A1	A1 \checkmark : \times \checkmark f_x Customer										
	Α	В	С	D	Е	F	G				
1	Customer	Product	Region	Date	Item unit p	No.Items	Total Sale				
2	Customer 3		MLG	########	0168,95	2,00	00337,90				
3	Customer 2	Avanza AT			0168,95	15,00	02534,25				
4	Customer '				0340,95	7,00	02386,65				
5	Customer 2		KDR	########	0799,95	14,00	11199,30				
6	Customer '	Avanza AT		#########	0168,95	4,00	00675,80				
7											

- 2. Jelaskan apa yang terjadi pada proses Filter rows\XDX!
 - → Langkah transformasi yang bertujuna untuk menyaring data berdasrkan kondisi tertentu. Jika true akan diteruskan ke langkah selanjutnya, jika false maka dapat dihapus atau diteruskan ke langkah yang lainnya.
- 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 table

No	Nama	Waktu	Prosessor	RAM	Storage	VGA	CPU
1	Cindy	6s	Intel i5- 10300H	12 GB	HDD 1 TB	NVIDIA	60%
						GTX 1050	
2	Sabrina	1s	intel i5-7300U	8192	Samsung	Intel ®	50%
				MB(8GB)	239 GB	graphics	
						620	
3	nada	1s	11 th gen	16 GB	SSD	Intel®Iris	23%
			intel®core TM i5-		NVMe	® Xe	

					477GB	graphics	
4	zahra	1s	Intel®core	16 GB	SSD	Intel ®	7%
			тмі7-		477GB	UHD	
			13650HX(13 th			graphics	
			Gen, 20 Cores)			+	
						NVIDIA	
						GeForce	
						RTX 4050	

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,

Data masyarakat berprofesi mahasiswa akan diberikan ke Pak Emir Makarena selaku ketua RT wilayah pendidikan.

Data masyarakat berprofesi PNS akan diberikan ke Pak Yasana Laili selaku ketua RT wilayah pemerintahan

Data masyarakat berprofesi Dokter akan diberikan ke Pak Budi Pekerti selaku ketua RT wilayah kesehatan

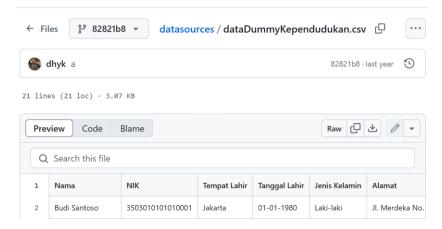
data-data tersbut akan digunakan untuk analisa di masing-masing wilayah dan juga Dashboard untuk emmbantu mengambil keputusan.

Pak Pascanowo mendapatakan data penduduk yang akan pindah ke wilayahnya melalui link: https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aead01 https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aead01 https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aead01

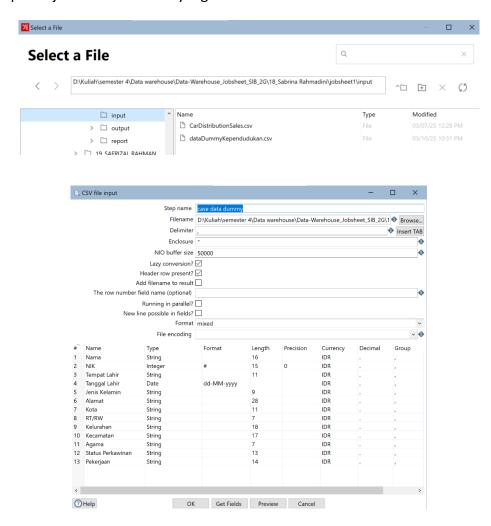
Bantulah Pak Pascanowo untuk memishkan data tersebut menjadi 4 file: MasyarakatMahasiswa, MasyarakatPNS, MasyarakatDokter, MasyarakatLainnya.

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

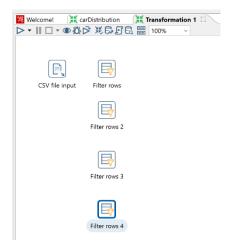
1) Download terlebih dahulu data dummy kependudukannya yang akan kita pakai nanti



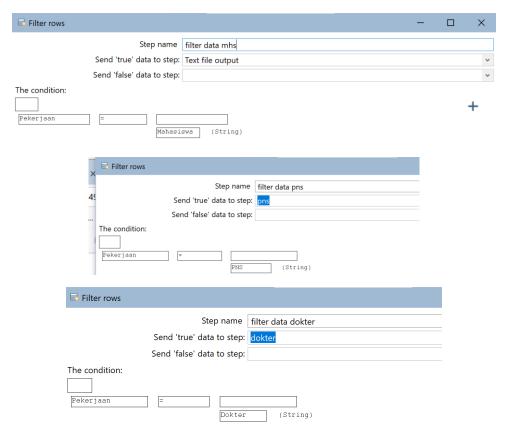
2) Lalu selanjutnya drag and drop csv file input dari input, lalu extract data yang sudah kita download dari github tadi ke kolom filename, setelah itu get file untuk membaca dan menampilkan jumlah kolom data yang sudah kita extract tadi



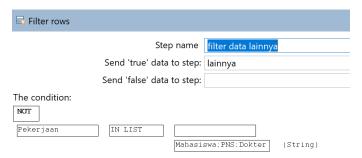
3) Selanjutnya drag and drop filter row nya dari flow sebanyak 4x yang masing masing nanti akan dibuat untuk mahasiswa, pns, dokter, dan Masyarakat dengan pekerjaan lainnya.



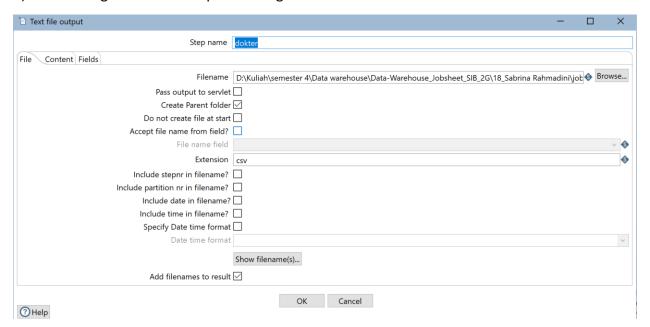
4) Atur di bagian filter row seperti gambar di bawah

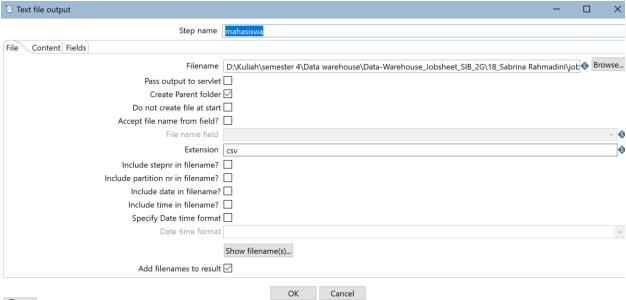


5) Pada bagian pengaturan file rows Masyarakat lainnya agak sedikit berbeda dengan filter data mahasiswa, pns dan dokter, dikarenakan adanya pengecualian data mahasiswa, pns, dokter sehingga pengaturannya seperti di bawah ini.

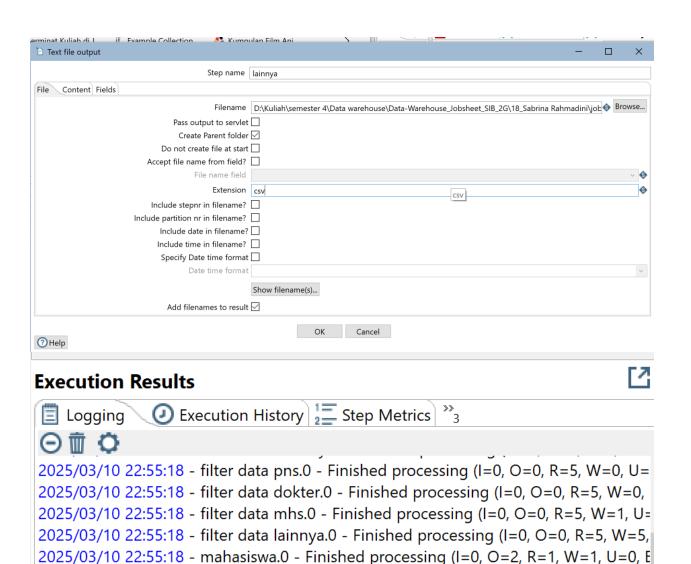


6) Atur di bagian text file output sesuai gambar di bawah ini





? Help

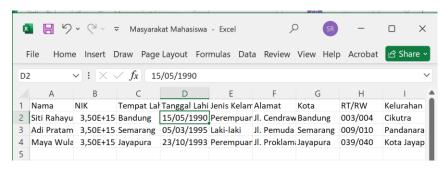


2025/03/10 22:55:18 - lainnya.0 - Finished processing (I=0, O=6, R=5, W=5, U=0, E=0)

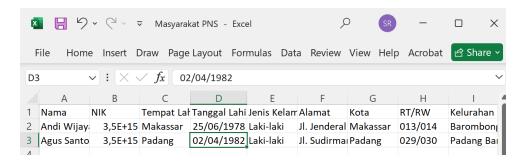
2025/03/10 22:55:18 - Spoon - The transformation has finished!!

Output:

Masyarakat mahasiswa:



Masyarakat PNS:



Masyarakat dokter:



Masyarakat lainnya:

