

Working with Google Sheet:

Extract - Format - Clean

Rizki Teguh Kurniawan

Data Scientist, eFishery

Youtube: Solving with Akoyo

Senin, 24 Juli 2023

Agenda



- **Data Extraction:**
 - *Import data from CSV*
 - *Import data from another sheets*
- **Data Formatting**
 - *Number Format*
 - *Text Format*
 - *Date Format*
 - *Conditional Formatting*
- **Handling Duplicate Data**
 - *Various case of duplicate data*
 - *Handling duplicate data*
- **Data Validation**
 - *What is data validation*
 - *Creating validation settings*





Modul

https://drive.google.com/drive/folders/1raajBQWs8L-hdCAK0Q-pWWoAZi_gYQLy?usp=share_link

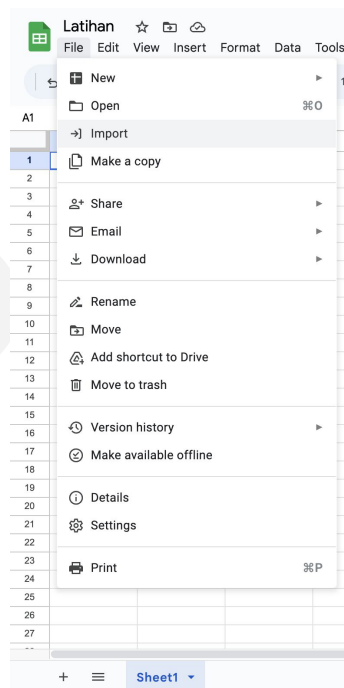




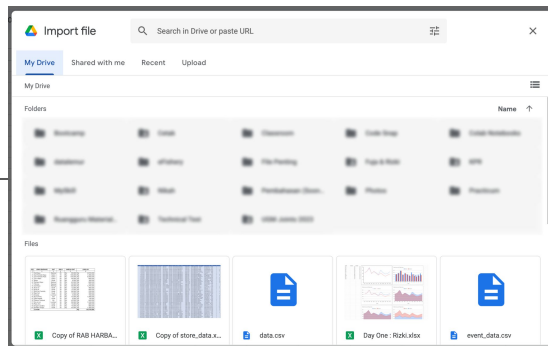
Data Extraction



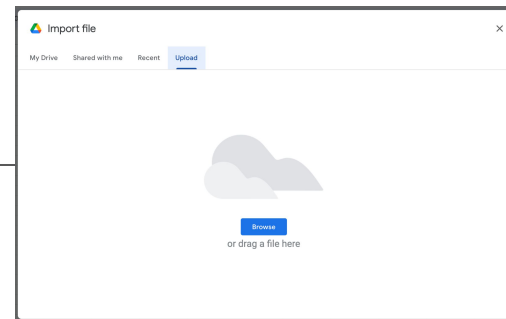
Data Extraction from External CSV file (GUI Method)



1. Klik file -> import



2. Jika file berada di gdrive, maka pilih tab My Drive lalu pilih file berakhiran .csv yang akan digunakan

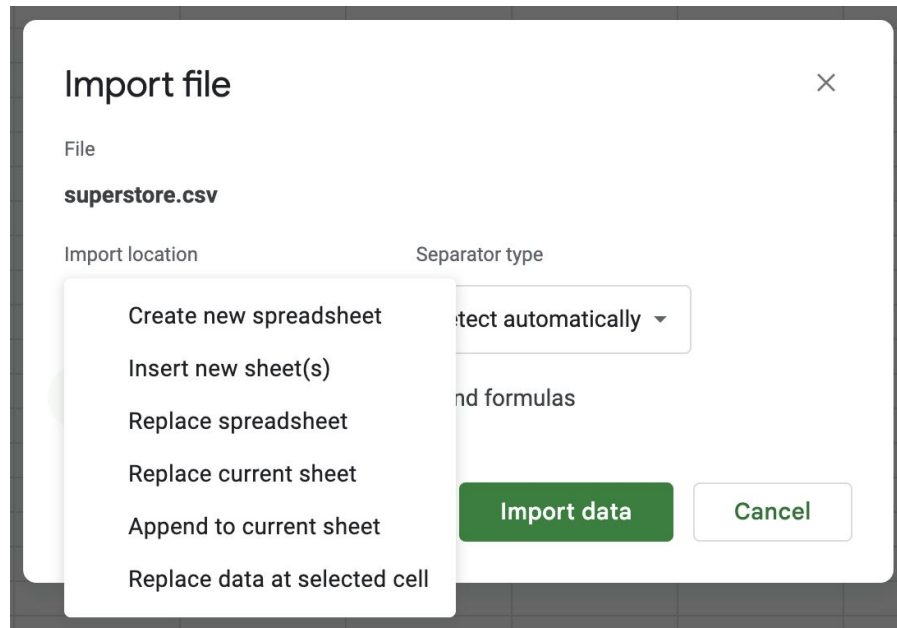


3. Jika file berada disimpan di perangkat lokal, maka pilih Upload -> Browse lalu pilih file berakhiran .csv yang akan digunakan

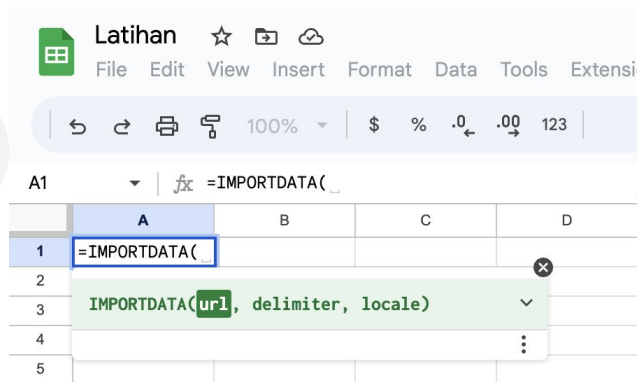
Data Extraction from External CSV file (GUI Method)

Setelah kamu memilih data csv yang akan digunakan, kamu akan diminta untuk memilih bagaimana file tersebut disajikan:

1. *Create new spreadsheet* -> Membuat file gsheet baru
2. *Insert new sheets* -> Menambahkan sheet baru pada gsheet yang sedang berjalan
3. *Replace spreadsheet* -> menggantikan spreadsheet yang sedang berjalan (dapat menghapus data pada gsheet)
4. *Replace current sheet* -> menggantikan sheet yang sedang berjalan (dapat menghapus data pada sheet)
5. *Append to current sheet* -> menyisipkan data pada sheet yang sedang berjalan
6. *Replace data at selected cell* -> menggantikan data yang sudah ditandai



Data Extraction from External CSV file (IMPORTDATA)



Selain itu, kita juga dapat menggunakan fungsi IMPORTDATA untuk mendapatkan data berdasarkan link

A1 `=IMPORTDATA("https://datahub.io/core/country-list/r/data.csv")`

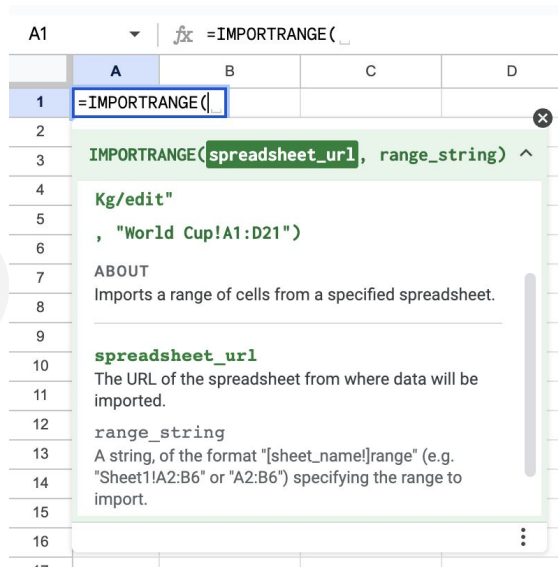
	A	B	C	D	E	F
1	Name	Code				
2	Afghanistan	AF				
3	Åland Islands	AX				
4	Albania	AL				
5	Algeria	DZ				
6	American Samoa	AS				
7	Andorra	AD				
8	Angola	AO				
9	Anguilla	AI				
10	Antarctica	AQ				
11	Antigua and Barbuda	AG				
12	Argentina	AR				
13	Armenia	AM				
14	Aruba	AW				
15	Australia	AU				
16	Austria	AT				
17	Azerbaijan	AZ				
18	Bahamas	BS				
19	Bahrain	BH				
20	Bangladesh	BD				
21	Barbados	BB				
22	Belarus	BY				
23	Belgium	BE				
24	Belize	BZ				
25	Benin	BJ				
26	Bermuda	BM				
27	Bhutan	BT				

Sheet1

Masukkan link dari file csv yang akan kita gunakan (pastikan berakhiran .csv) lalu tekan enter

Perhatikan bahwa data yang kita ekstrak ini tidak dapat diubah/akan menghasilkan error ketika diubah. Maka dari itu gunakan Copy & Paste Value Only terlebih dahulu

Data Extraction from Another Sheets (IMPORTRANGE)



Kita dapat menggunakan IMPORTRANGE untuk mengambil data dari sheet yang lain

The screenshot shows a Google Sheet with data imported from another spreadsheet. The formula bar at the top shows the formula: `=IMPORTRANGE("https://docs.google.com/spreadsheets/d/177TxHrI-n_CpL5670KsKjmp2FMfqvKvj5j0Y5B-1rUQ/edit?usp=shoring", "Sheet1!A1:G13")`. The data is organized in columns: Order_ID, Customer_ID, Postal_Code, Product_ID, Sales, Quantity, and Discount. An arrow points from the formula bar to the data table.

	A	B	C	D	E	F	G	H	I	J	K
1	Order_ID	Customer_ID	Postal_Code	Product_ID	Sales	Quantity	Discount				
2	CA-2019-	CG-12520	2016-02-20	FUR-BO-100017	261.96	2	0				
3	CA-2019-	CG-12520	2016-02-20	FUR-CH-100004	731.94	3	0				
4	CA-2019-	DV-13045	2146-07-04	OFF-LA-100002	14.62	2	0				
5	US-2018-	SO-20335	1991-03-14	FUR-TA-100005	957.5775	5	0.45				
6	US-2018-	SO-20335	1991-03-14	OFF-ST-100007	22.368	2	0.2				
7	CA-2017-	BH-11710	2146-06-30	FUR-FU-100014	48.86	7	0				
8	CA-2017-	BH-11710	2146-06-30	OFF-AR-100028	7.28	4	0				
9	CA-2017-	BH-11710	2146-06-30	TEC-PH-100022	907.152	6	0.2				
10	CA-2017-	BH-11710	2146-06-30	OFF-BI-1000391	18.504	3	0.2				
11	CA-2017-	BH-11710	2146-06-30	OFF-AP-100028	114.9	5	0				
12	CA-2017-	BH-11710	2146-06-30	FUR-TA-100015	1706.184	9	0.2				
13	CA-2017-	BH-11710	2146-06-30	TEC-PH-100020	911.424	4	0.2				
14											

pada IMPORTRANGE yang kita butuhkan adalah url dari sheet yang akan kita ambil datanya lalu range dari data.

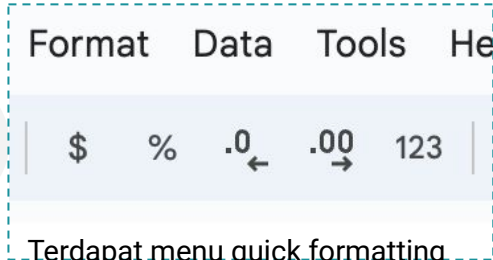
Perhatikan bahwa data yang kita ekstrak ini tidak dapat diubah/akan menghasilkan error ketika diubah. Maka dari itu gunakan Copy & Paste Value Only terlebih dahulu



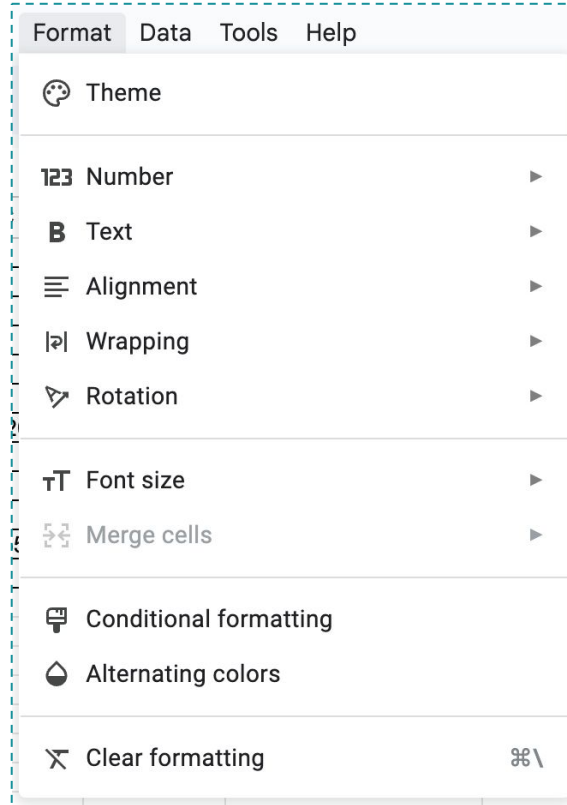
Data Formatting



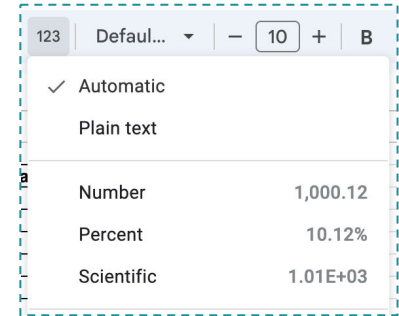
Number Formatting



Terdapat menu quick formatting untuk mengatur formatting dengan cepat seperti: currency, percentage, menambahkan dan mengurangi angka desimal

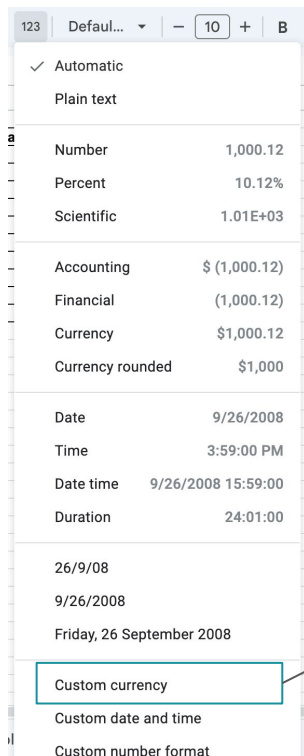


Serta pilihan yang khusus untuk memformat data numerik seperti di bawah ini



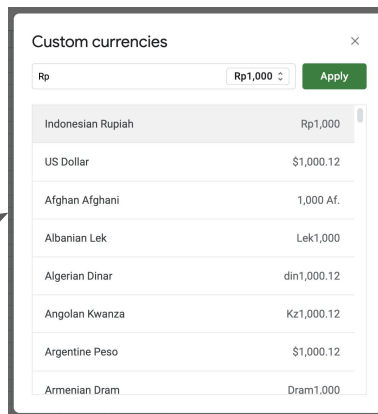
Atau menggunakan menu Format untuk mendapatkan menu formatting yang lebih luas konteksnya

Number Formatting



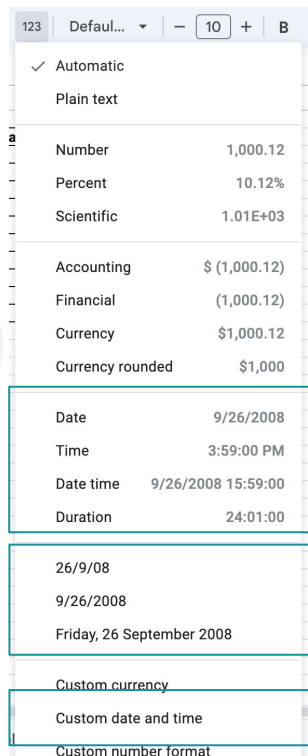
Pilihan default number formatting

Untuk menerapkan format tertentu pada data cukup select cell atau range lalu pilih format yang dikehendaki



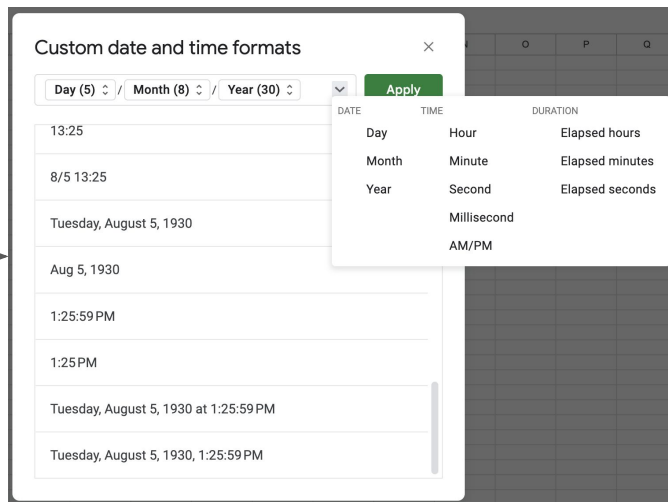
Karena simbol currency secara default adalah \$, untuk mengubah ke mata uang lain gunakan custom currency

Date & Time Formatting



Pilihan default
format date &
time

format yang
terakhir
digunakan



Kita juga dapat membuat format sesuai
kebutuhan menggunakan custom date and time

Text Formatting

Fungsi	Deskripsi
=LOWER()	Mengubah teks menjadi huruf kecil
=UPPER()	Mengubah teks menjadi huruf besar
=PROPER()	Mengubah huruf di setiap awal kata menjadi huruf kapital
=CONCATENATE()	Menyambungkan teks
=JOIN()	Menggabungkan setiap cell dan dipisahkan dengan delimiter
=LEFT()	Mengambil sebagian teks dari kiri
=RIGHT()	Mengambil sebagian teks dari kanan
=TRIM()	Menghapus space berlebih

Conditional Formatting

Conditional format rules

Single color Color scale

Apply to range

J2

Format rules

Format cells if...

Is not empty

Formatting style

Default

B I U A

Cancel Done

+ Add another rule

Conditional formatting digunakan untuk mewarnai cell berdasarkan kondisi tertentu.

Fungsi utama dari conditional formatting adalah untuk memudahkan user dalam menemukan data berdasarkan kriteria yang dicari.

Langkah langkah membuat conditional formatting:

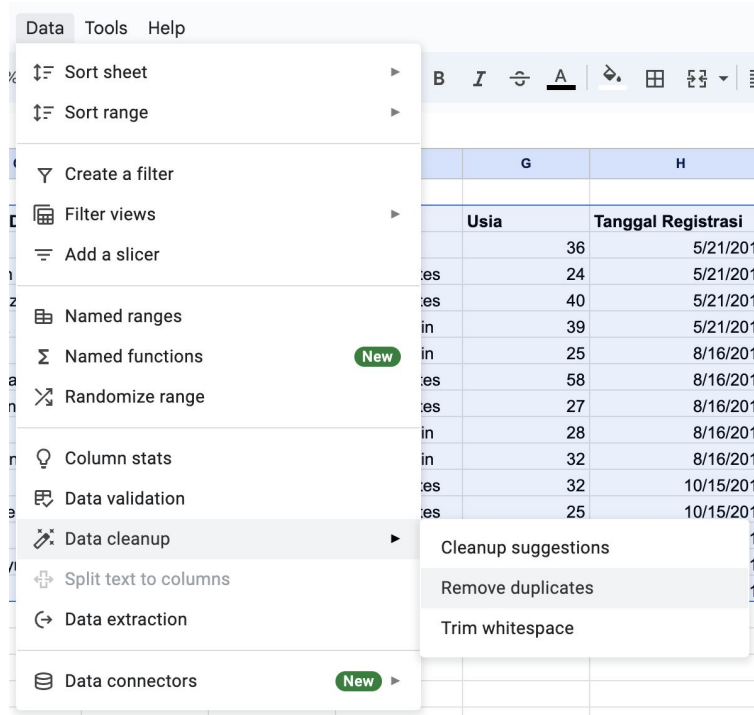
1. Masukkan range yang akan diterapkan conditional formatting pada kolom Apply to range
2. Masukkan aturan formatting pada Format rules, ikuti petunjuk yang disediakan
3. Tentukan Formatting styles



Duplicate Data



Duplicate Data



Untuk menerapkan remove duplicates, blok terlebih dahulu tabelnya

Untuk menampilkan dialog handling data duplicate gunakan:

Data -> Data cleanup -> Remove duplicates

Duplicate Data

NIP	Nama Depan	Nama Belakang	Jenis Kelamin	Negara	Usia	Tanggal Registrasi
2587	Philip	Gent	Male	France	36	5/21/2015
2554	Gaston	Brumm	Male	United States	24	5/21/2015
6548	Vincenza	Weiland	Female	United States	40	5/21/2015
1258	Arcelia	Bouska	Female	Great Britain	39	5/21/2015
1582	Mara	Hashimoto	Female	Great Britain	25	8/16/2016
2468	Nereida	Magwood	Female	United States	58	8/16/2016
2456	Earlean	Melgar	Female	United States	27	8/16/2016
5486	Fallon	Winward	Female	Great Britain	28	8/16/2016
3256	Sherron	Ascencio	Female	Great Britain	32	8/16/2016
1562	Dulce	Abril	Female	United States	32	10/15/2017
3549	Kathleen	Hanner	Female	United States	25	10/15/2017
3598	Etta	Hurn	Female	Great Britain	56	10/15/2017
2579	Franklyn	Unknow	Male	France	38	10/15/2017
4148	Philip	Gent	Male	France	38	5/21/2017

Remove duplicates

15 rows and 7 columns selected

☒ Data has header row

Columns to analyze

- ☒ Select all
- ☒ Column B - NIP
- ☒ Column C - Nama Depan
- ☒ Column D - Nama Belakang
- ☒ Column E - Jenis Kelamin
- ☒ Column F - Negara
- ☒ Column G - Usia
- ☒ Column H - Tanggal Registrasi

Cancel Remove duplicates

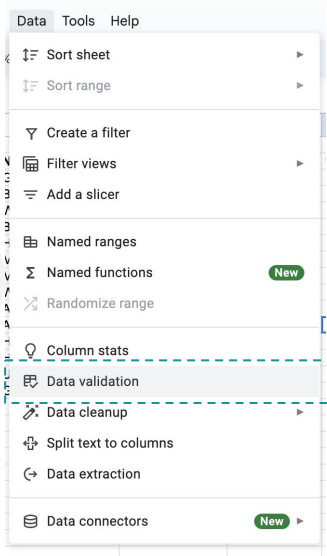
1. Centang pada bagian 'Data has header row' untuk menampilkan nama-nama kolom pada daftar kolom di bawahnya
2. Pilih kombinasi kolom yang akan diperiksa. Data yang memiliki elemen yang sama pada kolom-kolom yang dipilih akan dianggap duplikat
3. Data yang dipertahankan adalah data yang muncul paling atas



Data Validation

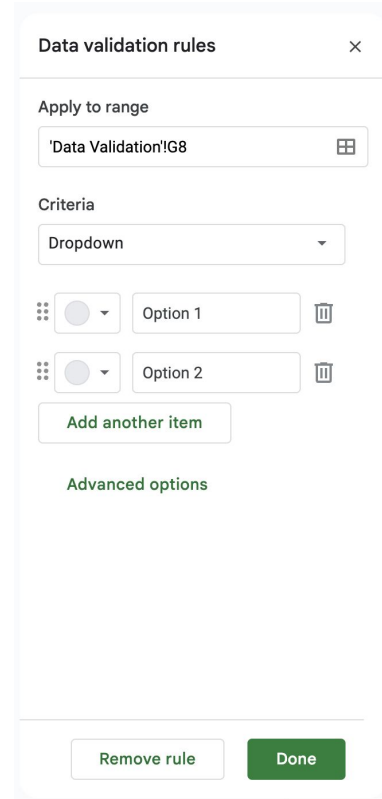


Data Validation



Data validation berfungsi untuk meminimalkan kesalahan input data yang menyebabkan data tidak konsisten atau data tidak sesuai peruntukannya. Data validation dapat diakses dengan:

Data -> Data validation



Langkah-langkah untuk menerapkan data validation

1. Pilih isi nama cell atau range yang akan diterapkan data validation pada bagian Apply to range
2. Pilih kriteria data validation pada Criteria
3. Ikuti aturan yang diminta oleh Criteria. Setiap Criteria memiliki aturan yang berbeda-beda
4. Pada menu Advanced options kita dapat mengatur seperti: memberikan warning atau menolak input yang salah, menampilkan help text, memilih display style dari data.



Mini Task



Task

1. Buat worksheet baru lalu extract data raw superstore pada folder modul ke dalam sheet
2. Ubah format data pada kolom diskon menjadi percentage
3. Ubah format data pada kolom sales dan profit menjadi currency dengan mata uang dollar dan maksimal dua angka di belakang koma
4. Pada kolom profit tandai profit yang negatif dengan warna merah
5. Ubah data pada kolom Ship_Mode sampai Region dengan Proper Case
6. Periksa apakah data superstore memiliki transaksi duplikat
7. Pada bagian Segment terapkan data validation sehingga data yang masuk adalah satu dari pilihan berikut: Corporate, Consumer, Home Office