

FULLSTACK INTENSIVE BOOTCAMP

Data Analyst

MySkill

Data Formatting & Cleansing

Oleh: Wildan Zakaria

JOIN THE BEST UPSKILLING COMMUNITY
WITH ME at myskill.id/bootcamp

MINI PORTOFOLIO

Course Summary



- **Data extraction**

Ekstraksi data yaitu membaca data(tabel) di excel dari berbagai sumber data yang berbeda seperti excel workbook, text/csv, website, pdf.

Steps : **data > get & transform data > pilih sumber data yang dituju**

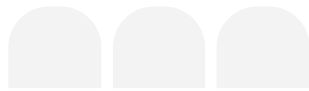
- **Data formatting**

Berfungsi untuk mempermudah membaca dalam menghitung dan menganalisis data.

Steps : **home > section number > pilih format sesuai kriteria**

Type of format data :

1. General (format cell tidak dispesifikan)
2. Number (bilangan)
3. Currency (mata uang)
4. Accounting (akuntansi)
5. Date (tanggal)
6. Time (waktu)
7. Percentage (persentase)
8. Fraction (pecahan)
9. Scientific (notasi ilmiah)
10. Text (teks)



Course Summary



- **Data duplicate**

Data duplikat atau data ganda biasanya terjadi karena kesalahan sistem atau kesalahan penginput. Data duplikat ditangani dengan cara menghapus data yang tidak diperlukan.

Steps : **blok tabel > data > remove duplicates**

- **Data validation**

Berfungsi untuk meminimalkan kesalahan input data yang menyebabkan data tidak konsisten atau data tidak sesuai peruntukannya..

Steps :

1. Data validation: setting

Blok sel/range untuk data validation > data > settings > pilih kriteria data validation

2. Data validation: input message (digunakan untuk memberikan petunjuk pengisian)

Blok sel/range untuk data validation > data > input message > masukan pesan yang ingin disampaikan pada sel/range data validation

3. Data validation: error alert (digunakan untuk memunculkan pesan jika user memasukkan data yang tidak sesuai dengan kriteria data validation)

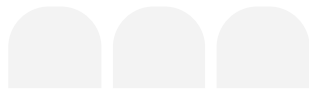
Blok sel/range untuk data validation > data > error alert > pilih jenis eror dan masukan pesan yang ingin disampaikan

Jenis error alert pada data validation :

1. Stop : memaksa user untuk mengisi data sesuai kriteria validasi dan menolak memasukkan data yang tidak sesuai

2. Warning : memberikan peringatan kepada user apabila data tidak sesuai dengan kriteria validasi. Memberikan pilihan apakah user yakin untuk menginput data yang tidak sesuai kriteria validasi

3. Information : memasukkan data walaupun tidak sesuai dengan kriteria validasi lalu memberikan informasi terkait hal tersebut



Jawaban Tugas

Case : Data Formatting

Angka	Tugas	Jawaban
9980000	Ubah ke dalam format mata uang rupiah	Rp9.980.000,00
0,53	Ubah ke dalam bentuk pecahan	53/100
04/01/1995	Ubah ke dalam format 'Hari, tanggal bulan, tahun'	Wednesday, 04 January 1995

Jawaban Tugas

Case : Data Duplicate

Perhatikan bahwa Penelope Jordan dan Madeline Walker memiliki EEID yang sama dikarenakan kesalahan pencatatan. Karena Penelope terlebih dulu muncul di dalam tabel maka hapus entri Madeline dengan metode remove duplicate

From this

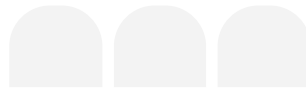
<i>EEID</i>	<i>Full Name</i>	<i>Job Title</i>	<i>Department</i>	<i>Business Unit</i>	<i>Gender</i>	<i>Ethnicity</i>
E02387	Emily Davis	Sr. Manger	IT	Research & D	Female	Black
E04105	Theodore Dinh	Technical Ar	IT	Manufacturi	Male	Asian
E02572	Luna Sanders	Director	Finance	Speciality Pro	Female	Caucasian
E02832	Penelope Jordan	Computer Sy	IT	Manufacturi	Female	Caucasian
E01639	Austin Vo	Sr. Analyst	Finance	Manufacturi	Male	Asian
E00644	Joshua Gupta	Account Rep	Sales	Corporate	Male	Asian
E01550	Ruby Barnes	Manager	IT	Corporate	Female	Caucasian
E04332	Luke Martin	Analyst	Finance	Manufacturi	Male	Black
E04533	Easton Bailey	Manager	Accounting	Manufacturi	Male	Caucasian
E02832	Madeline Walker	Sr. Analyst	Finance	Speciality Pro	Female	Caucasian

Jawaban Tugas



To this

EEID	Full Name	Job Title	Department	Business Unit	Gender	Ethnicity
E02387	Emily Davis	Sr. Manger	IT	Research & Development	Female	Black
E04105	Theodore Dinh	Technical Architect	IT	Manufacturing	Male	Asian
E02572	Luna Sanders	Director	Finance	Speciality Products	Female	Caucasian
E02832	Penelope Jordan	Computer Systems Manager	IT	Manufacturing	Female	Caucasian
E01639	Austin Vo	Sr. Analyst	Finance	Manufacturing	Male	Asian
E00644	Joshua Gupta	Account Representative	Sales	Corporate	Male	Asian
E01550	Ruby Barnes	Manager	IT	Corporate	Female	Caucasian
E04332	Luke Martin	Analyst	Finance	Manufacturing	Male	Black
E04533	Easton Bailey	Manager	Accounting	Manufacturing	Male	Caucasian



Jawaban Tugas

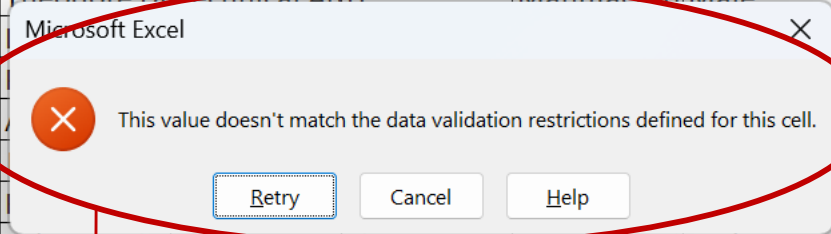
Case : Data Validation

Pada kolom ethnicity, nilai yang valid adalah asian, black, caucasian buat data validation dengan informasi di atas lalu beri tanda pada data yang tidak valid pada kolom age (kolom kosong) beri data validation dengan kriteria bilangan bulat dari 30 sampai 60.

<i>EEID</i>	<i>Full Name</i>	<i>Job Title</i>	<i>Department</i>	<i>Business Unit</i>	<i>Gender</i>	<i>Ethnicity</i>	<i>Age</i>
E02387	Emily Davis	Sr. Manger	IT	Research & Development	Female	Black	
E04105	Theodore Dinh	Technical Architect	IT	Manufacturing	Male	Asian	
E02572	Luna Sanders	Director	Finance	Speciality Products	Female	Caucasian	
E02832	Penelope Jordan	Computer Systems Manager	IT	Manufacturing	Female	Purple	
E01639	Austin Vo	Sr. Analyst	Finance	Manufacturing	Male	Asian	
E00644	Joshua Gupta	Account Representative	Sales	Corporate	Male	Asian	
E01550	Ruby Barnes	Manager	IT	Corporate	Female	Caucasian	
E04332	Luke Martin	Analyst	Finance	Manufacturing	Male	Green	
E04533	Easton Bailey	Manager	Accounting	Manufacturing	Male	Caucasian	
E02832	Madeline Walker	Sr. Analyst	Finance	Speciality Products	Female	Caucasian	

Jawaban Tugas

EEID	Full Name	Job Title	Department	Business Unit	Gender	Ethnicity	Age
E02387	Emily Davis	Sr. Manger	IT	Research & D	Female	Black	42
E04105	Theodore Di	Technical Ar	IT	Manufacturi	Male	Asian	30
E02572		Microsoft Excel				Caucasian	32
E02832						Purple	31
E01639						Asian	36
E00644						Asian	34
E01550						Caucasian	40
E04332	Luke Martin	Analyst	Finance	Manufacturi	Male	Green	31
E04533	Easton Baile	Manager	Accounting	Manufacturi	Male	Caucasian	40
E02832	Madeline W	Sr. Analyst	Finance	Speciality Pr	Female	Caucasian	130



Tanda circle = tanda data tidak valid pada kolom Ethnicity, karena nilai yang valid hanya Asian, Black, dan Caucasian

Tanda data tidak valid karena 130 tidak memenuhi kriteria data validation yang telah dibuat pada kolom age yaitu bilangan bulat dari 30 sampai 60.

Follow me!

Instagram : @Wildanzkk

Twitter : @rekfoop

LinkedIn : Wildan Zakaria

Bootcamp Data Analysis
by @myskill.id

