

# Praktikum Mengumpulkan Data

## Tahap 1: Menentukan Kebutuhan Data

Langkah-langkah praktikum dalam menentukan kebutuhan data

- Identifikasi kebutuhan data berdasarkan tujuan teknis data science dengan cara:

- Buat daftar fitur-fitur data yang akan digunakan dalam proyek, dan jelaskan alasan pemilihan setiap fitur tersebut.

**Contoh 1:** Anda bekerja pada proyek untuk mengidentifikasi pelanggan loyal di sebuah bank. Untuk melakukan ini, Anda membutuhkan data seperti umur, pendapatan, jumlah transaksi, dan skor kredit dengan deskripsi sebagai berikut:

- *umur*: untuk melihat distribusi usia nasabah.
- *pendapatan*: untuk memahami kemampuan finansial nasabah.
- *jumlah transaksi*: untuk menilai seberapa sering nasabah menggunakan layanan bank.
- *skor kredit*: untuk mengevaluasi risiko kredit nasabah.

**Contoh 2:** Anda mengerjakan proyek untuk memprediksi churn pelanggan di perusahaan telekomunikasi. Anda memerlukan data seperti lama berlangganan, frekuensi panggilan ke pusat layanan pelanggan, paket data yang digunakan, dan riwayat panggilan dengan deskripsi sebagai berikut:

- *lama berlangganan*: untuk menilai seberapa lama pelanggan telah menggunakan layanan.
- *frekuensi panggilan ke pusat layanan pelanggan*: untuk melihat masalah atau ketidakpuasan pelanggan.
- *paket data yang digunakan*: untuk memahami penggunaan layanan.
- *riwayat tagihan*: untuk menganalisis pola pembayaran pelanggan.

- Periksa ketersediaan data dari sumber yang relevan, baik internal maupun eksternal.

**Proyek:** Analisis Faktor yang Mempengaruhi Biaya Hidup Mahasiswa

### Fitur yang Dipilih dan Alasan Pemilihan

1. Usia – Mengetahui apakah ada pola gaya hidup yang berubah berdasarkan usia.
2. Jenis Kelamin – Menganalisis perbedaan gaya hidup antara laki-laki dan perempuan.
3. Alamat & Kota Tempat Tinggal – Melihat pengaruh lokasi terhadap preferensi gaya hidup.
4. Jenis Kendaraan – Mengetahui apakah kepemilikan kendaraan mempengaruhi kebiasaan mobilitas mahasiswa.
5. Pengeluaran BBM – Mengukur pola mobilitas dan pengaruhnya terhadap gaya hidup.
6. IPK – Meneliti hubungan antara akademik dan kebiasaan sosial mahasiswa.
7. Hobi – Mengetahui preferensi aktivitas mahasiswa dalam waktu luang.
8. Tinggi dan Berat Badan – Melihat keterkaitan antara gaya hidup dengan kebiasaan kesehatan dan pola makan.

### Ketersediaan Data

Semua fitur tersedia dalam dataset DataSIB2E.csv, namun beberapa data mungkin perlu dibersihkan sebelum analisis lebih lanjut.

## Tahap 2: Pengambilan Data dengan Tools yang Tepat

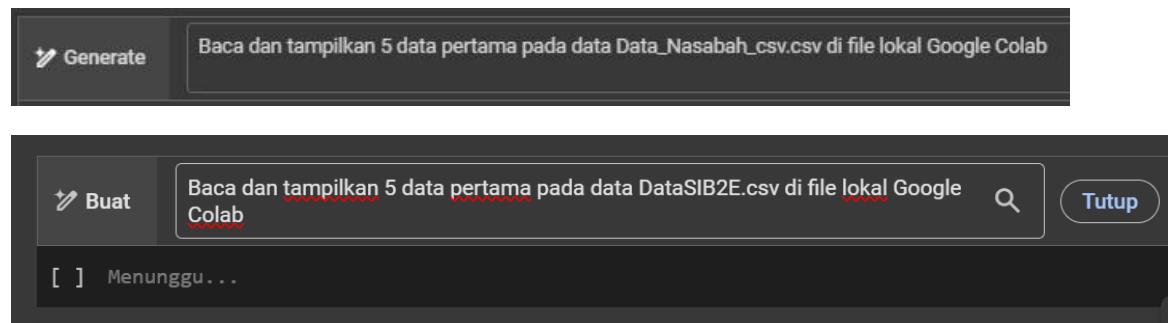
Langkah-langkah praktikum dalam pengambilan data dengan tools yang tepat

- Pilih tools dan metode pengambilan data yang sesuai. Gunakan tools seperti Python dengan pustaka BeautifulSoup untuk web scraping, atau API untuk mengambil data dari sumber yang tersedia secara online. Kaggle juga merupakan salah satu pilihan untuk pengambilan data.

Berikut beberapa contoh cara mengerjakan sesuai dengan tools yang dipilih:

- Pengambilan data menggunakan *pandas* untuk file CSV**

Ketikan prompt pada *generate with AI* di Google Colab:



Prompt tersebut akan menghasilkan kode seperti berikut:

```
import pandas as pd  
  
# Baca file CSV lokal  
  
data = pd.read_csv('/content/Data_Nasabah_csv.csv',  
delimiter=';')  
  
# Tampilkan 5 data pertama  
  
print(data.head(5))
```

The code cell has a status message: 'Gunakan kode dengan hati-hati' (Use the code with care).

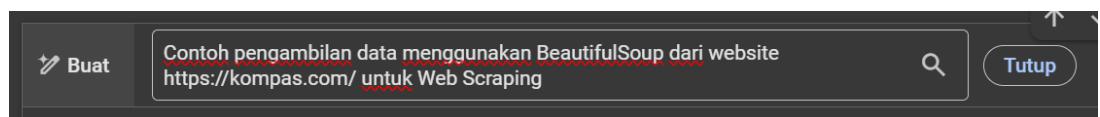
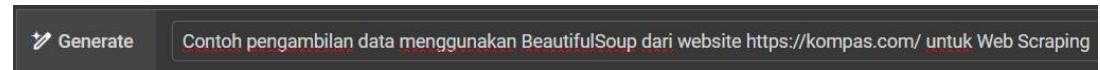
Output dari kode yang dihasilkan akan tampak seperti di bawah ini:

nasabah_id	umur	jenis_kelamin	pendapatan	saldo_rata_rata	\
0	N001	22	Perempuan	5800000	1508000
1	N002	64	Perempuan	5700000	1254000
2	N003	27	Perempuan	2950000	590000
3	N004	34	Perempuan	3100000	186000
4	N005	45	Laki-Laki	6700000	1474000
jumlah_transaksi	jenis_produk	frekuensi_kunjungi_cabang	\		
0	19	tabungan		1	
1	9	kartu_kredit		2	
2	12	tabungan		1	
3	16	deposito		5	
4	15	kartu_kredit		3	
pengguna_mobile_banking	skor_kredit				
0	TIDAK	900			
1	TIDAK	900			
2	YA	500			
3	TIDAK	700			
4	YA	800			

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	Unnamed: 9	Unnamed: 10	Unnamed: 11	Unnamed: 12
0	Absensi	Nama Lengkap	Tempat Tanggal Lahir	Usia	Jenis Kelamin	Alamat	X Kota Tempat Tinggal	Jenis Kendaraan	Pengeluaran BBM	IPK	Hobi	Tinggi dan Berat Badan	Data Lain
1	1	Abhinaya Nuzuluzzuhdi	Malang, 31 Oktober 2004	20	Laki-laki	Jl. Gajayana	Malang	Sepeda Motor	Rp. 90.000	NaN	Maen Game	173/50kg	NaN
2	2	Alvi Choirinnikmah	Bililar, 9 September 2004	20	Perempuan	Jln. Kembang Turi	Malang	NaN	NaN	NaN	Menonton film	NaN	NaN
3	3	Alya Ajeng Ayu	Malang, 18 November 2004	20	Perempuan	Jl. Parkit Selatan no. 2	Malang	Sepeda Motor	Rp. 25.000 - 30.000	3.74	Baking	155cm 49kg	NaN
4	4	Ardhelia Putri Maharani	Malang, 11 Oktober 2004	20	Perempuan	Serenia Garden Regency B9	Malang	Sepeda Motor	Rp. 30.000	NaN	Live Tiktok	164 - 46	NaN

## 0 Pengambilan data menggunakan BeautifulSoup untuk Web Scraping

Ketikan prompt pada generate with AI di Google Colab:



Prompt tersebut akan menghasilkan kode seperti berikut:

```
import pandas as pd

import requests

from bs4 import BeautifulSoup


# URL yang ingin di-scrape
url = 'https://www.kompas.com/'


# Mengirim permintaan GET ke URL
response = requests.get(url)


# Memeriksa apakah permintaan berhasil
if response.status_code == 200:

    # Membuat objek BeautifulSoup dari konten HTML
    soup = BeautifulSoup(response.content,
    'html.parser')


    # Mencari semua berita utama (misalnya, dengan tag
    <h3 class="most_title">)

    berita_utama = soup.find_all('h3',
    class_='most_title')

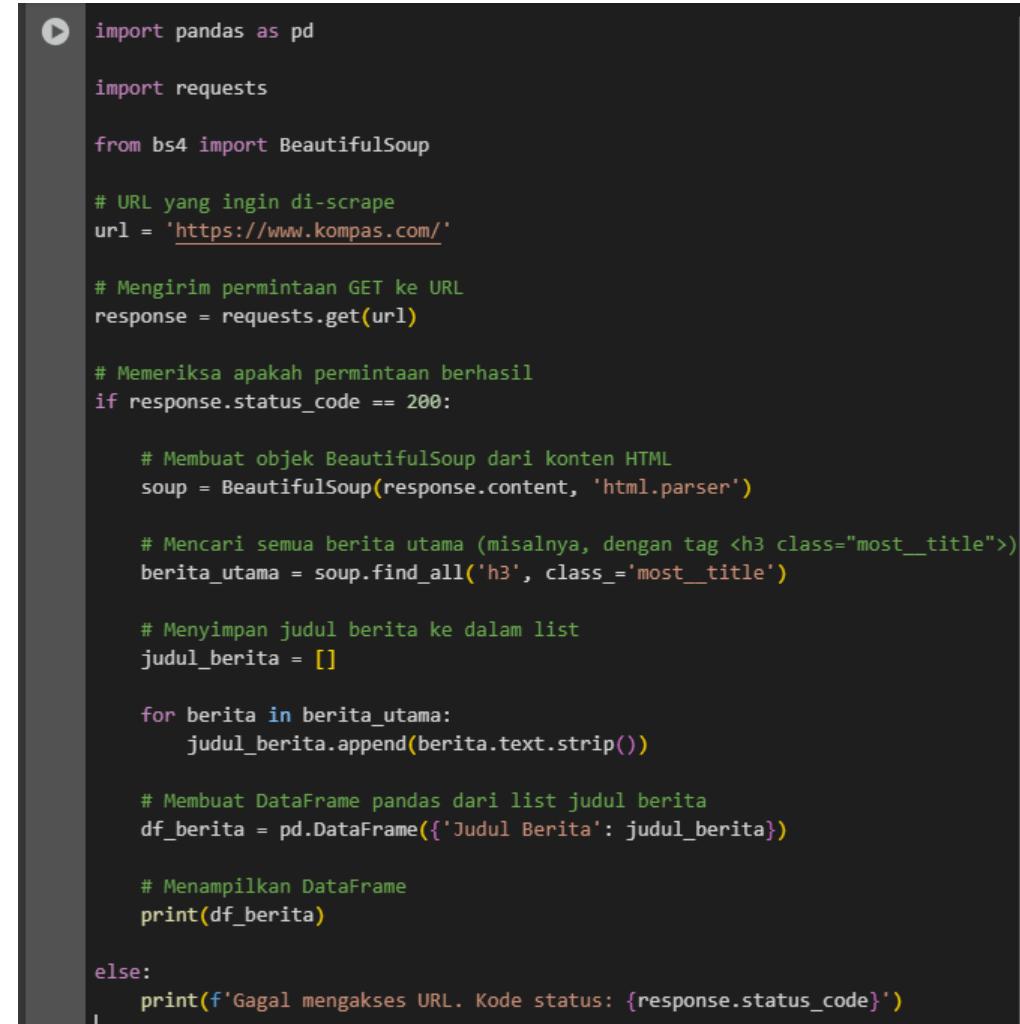

    # Menyimpan judul berita ke dalam list
    judul_berita = []

    for berita in berita_utama:
        judul_berita.append(berita.text.strip())


    # Membuat DataFrame pandas dari list judul berita
    df_berita = pd.DataFrame({'Judul': judul_berita})
```

```
# Menampilkan DataFrame
print(df_berita)

else:
    print(f'Gagal mengakses URL. Kode status: {response.status_code}'')
```



```
import pandas as pd
import requests
from bs4 import BeautifulSoup
url = 'https://www.kompas.com/'
response = requests.get(url)
if response.status_code == 200:
    soup = BeautifulSoup(response.content, 'html.parser')
    berita_utama = soup.find_all('h3', class_='most_title')
    judul_berita = []
    for berita in berita_utama:
        judul_berita.append(berita.text.strip())
    df_berita = pd.DataFrame({'Judul Berita': judul_berita})
    print(df_berita)
else:
    print(f'Gagal mengakses URL. Kode status: {response.status_code}')
```

Berikut untuk output kode di atas:

Empty DataFrame  
Columns: [Judul Berita]  
Index: []

Empty DataFrame  
Columns: [Judul Berita]  
Index: []

Oleh karena itu, butuh penyesuaian dengan melakukan inspect pada halaman website yang dipilih (dalam contoh ini [Kompas](#)) sebagai berikut:

```
import requests
from bs4 import BeautifulSoup
from datetime import datetime
import pytz

# URL yang ingin di-scrape
url = 'https://www.kompas.com/food/read/2024/10/08/103100275/resep-kukis-kacang-rendah-gula-cocok-untuk-pengidap-diabetes'

# Mengirim permintaan GET ke URL
response = requests.get(url)

# Memeriksa apakah permintaan berhasil
if response.status_code == 200:
    # Membuat objek BeautifulSoup dari konten HTML
    soup = BeautifulSoup(response.content, 'html.parser')
```

**URL lebih  
spesifik**

```
# Mengambil judul artikel
content_title = soup.find('h1').text      # Menyimpan judul
# Mencoba mengambil tanggal publikasi
# Mengambil konten dan mengonversi ke objek datetime
utc_time = datetime.fromisoformat(content_published_date['content'].replace('Z', '+00:00'))
# Mengonversi ke zona waktu WIB
wib_time = utc_time.astimezone(pytz.timezone('Asia/Jakarta'))
published_date = wib_time.strftime('%d/%m/%Y, %H:%M WIB') # Format yang diinginkan
else:
    published_date = "Tanggal publikasi tidak ditemukan"
# Mengambil tag
content_tags = soup.find('meta', {'name': 'keywords'})
if content_tags:
    tags = content_tags['content']      # Mengambil tag dari tag meta
else:
    tags = "Tag tidak ditemukan"
# Menampilkan hasil
```

```

print("Judul Artikel:", content_title)
print("Tanggal Publikasi:", published_date)
print("Tag:", tags)

else:
    print(f'Gagal mengakses URL. Kode status: {response.status_code}')

```

Data yang ditampilkan lebih spesifik

```

import requests
from bs4 import BeautifulSoup
from datetime import datetime
import pytz

# URL yang ingin di-scrape
url = "https://www.kompas.com/food/read/2024/10/08/103100275/resep-kukis-kacang-rendah-gula-cocok-untuk-pengidap-diabetes"

# Menambahkan headers agar tidak diblokir
headers = {
    "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36"
}

# Mengirim permintaan GET ke URL
response = requests.get(url, headers=headers)

# Memeriksa apakah permintaan berhasil
if response.status_code == 200:
    # Membuat objek BeautifulSoup dari konten HTML
    soup = BeautifulSoup(response.content, 'html.parser')

    # Mengambil judul artikel
    content_title = soup.find('h1').text.strip() if soup.find('h1') else "Judul tidak ditemukan"

    # Mencoba mengambil tanggal publikasi
    content_published_date = soup.find('meta', property='article:published_time')
    if content_published_date and 'content' in content_published_date.attrs:
        try:
            # Mengambil konten dan mengonversi ke objek datetime
            utc_time = datetime.fromisoformat(content_published_date['content'].replace('Z', '+00:00'))

            # Mengonversi ke zona waktu WIB
            wib_time = utc_time.astimezone(pytz.timezone('Asia/Jakarta'))
            published_date = wib_time.strftime('%d/%m/%Y, %H:%M WIB') # Format yang diinginkan
        except ValueError:
            published_date = "Format tanggal tidak valid"
    else:
        published_date = "Tanggal publikasi tidak ditemukan"

    # Mengambil tag
    content_tags = soup.find('meta', {'name': 'keywords'})
    tags = content_tags['content'] if content_tags and 'content' in content_tags.attrs else "Tag tidak ditemukan"

    # Menampilkan hasil
    print("Judul Artikel:", content_title)
    print("Tanggal Publikasi:", published_date)
    print("Tag:", tags)

else:
    print(f'Gagal mengakses URL. Kode status: {response.status_code}')

```

Output yang akan diberikan sebagai berikut:

```

→ Judul Artikel: Resep Kukis Kacang Rendah Gula, Cocok untuk Pengidap Diabetes
Tanggal Publikasi: 08/10/2024, 10:31 WIB
Tag: Resep kukis kacang rendah gula, Resep kukis kacang rendah gula untuk pengidap diabetes, Resep makanan

```

## 0 Pengambilan data melalui Kaggle

Setelah login melalui website [Kaggle](#), pilih bagian datasets dan ketik kata kunci data yang ingin dicari melalui search bar seperti pada gambar di bawah ini.

The screenshot shows the Kaggle Datasets search interface. At the top, there's a search bar with the query "house price". Below it are several filter buttons: "All datasets X", "Computer Science", "Education", "Classification", "Computer Vision", "NLP", and "Data Visualization". There's also a "Pre-Trained Model" button. The main search results are displayed below, with one result highlighted: "House Price" by Juhi Bhojani, updated a year ago, with 117 upvotes and a Silver badge. Other results include "House price prediction" by Shree, "House Price Prediction Challenge" by Anmol Kumar, and "California House Price" by Shibu Mohapatra.

Kemudian pilih datasets yang relevan untuk proyek data science.

This screenshot shows the search results for "House Price" on Kaggle. The results are sorted by Hotness. The top result is "House Price" by Juhi Bhojani, updated a year ago, with 117 upvotes and a Silver badge. Other results include "House price prediction" by Shree, "House Price Prediction Challenge" by Anmol Kumar, and "California House Price" by Shibu Mohapatra.

This screenshot shows the search results for "House Price" on Kaggle with filters applied. The filters are set to "Relevance" and "Last 90 days". The results are as follows:

- House Price** by Juhi Bhojani (2 years ago) - 151 upvotes, 13,045 downloads
- House price prediction** by Shree (6 years ago) - 686 upvotes, 94,421 downloads
- House Price Prediction Challenge** by Anmol Kumar (4 years ago) - 242 upvotes, 22,804 downloads

Lanjutkan dengan membaca deskripsi dan tekan tombol download untuk menyimpan dataset.

House Price

Data Card Code (29) Discussion (1) Suggestions (0)

house\_prices.csv (106.15 MB)

Detail Compact Column 10 of 21 columns

About this file Add Suggestion

A file containing 21 columns and 187531 rows

# Index	Title	Description	Amount(in rupees)	P
0	2 BHK Ready to Oc... 1% 4 BHK Ready to Oc... 1% Other (183778) 98%	[null] 2% Multistorey apartm... 1% Other (181776) 97%	Call for Price 5% 85 Lac 3% Other (172583) 92%	0
8	1 BHK Ready to Occupy Flat for sale in Srushti Siddhi Mangal Murti Complex	Bhiwandi, Thane has an attractive 1 BHK Flat for sale. The property is Ideally	42 Lac	6000

House Price

Data Card Code (36) Discussion (1) Suggestions (0)

house\_prices.csv (106.15 MB)

Detail Compact Column 10 of 21 columns

About this file Suggest Edits

A file containing 21 columns and 187531 rows

# Index	Title	Description	Amount(in rupees)	Price (in rupees)	location
0	2 BHK Ready to Oc... 1% 4 BHK Ready to Oc... 1% Other (183778) 98%	[null] 2% Multistorey apartm... 1% Other (181776) 97%	Call for Price 5% 85 Lac 3% Other (172583) 92%	0 6.70m	new-delhi bangalore Other (131)
8	1 BHK Ready to Occupy Flat for sale in Srushti Siddhi Mangal Murti Complex	Bhiwandi, Thane has an attractive 1 BHK Flat for sale. The property is Ideally	42 Lac	6000	thane

Ketikan prompt pada generate ai

Buat Baca dan tampilkan 5 data pertama pada data house\_prices.csv di file lokal Google Colab

Prompt akan menghasilkan kode

```
# prompt: Baca dan tampilkan 5 data pertama pada data house_prices.csv di file lokal Google Colab

import pandas as pd

# Assuming the file is named 'house_prices.csv' and is in the same directory as the notebook
try:
    df = pd.read_csv('house_prices.csv')
    print(df.head())
except FileNotFoundError:
    print("Error: 'house_prices.csv' not found. Please upload the file or provide the correct path.")
```

Unggah file dataset dan gunakan *pandas* untuk melihat isi dataset pada Google Colab seperti pada gambar di bawah ini.

## Output yang di berikan

```
▶ data = pd.read_csv('/content/house_prices.csv')

▶ print(data.head())

      Index          Title \
0      0 1 BHK Ready to Occupy Flat for sale in Srushti...
1      1 2 BHK Ready to Occupy Flat for sale in Dosti V...
2      2 2 BHK Ready to Occupy Flat for sale in Sunrise...
3      3       1 BHK Ready to Occupy Flat for sale Kasheli
4      4 2 BHK Ready to Occupy Flat for sale in TenX Ha...

                                         Description Amount(in rupees) \
0  Bhiwandi, Thane has an attractive 1 BHK Flat f...           42 Lac
1  One can find this stunning 2 BHK flat for sale...         98 Lac
2  Up for immediate sale is a 2 BHK apartment in ...        1.40 Cr
3  This beautiful 1 BHK Flat is available for sal...        25 Lac

      Index          Title \
0      0 1 BHK Ready to Occupy Flat for sale in Srushti...
1      1 2 BHK Ready to Occupy Flat for sale in Dosti V...
2      2 2 BHK Ready to Occupy Flat for sale in Sunrise...
3      3       1 BHK Ready to Occupy Flat for sale Kasheli
4      4 2 BHK Ready to Occupy Flat for sale in TenX Ha...

                                         Description Amount(in rupees) \
0  Bhiwandi, Thane has an attractive 1 BHK Flat f...           42 Lac
1  One can find this stunning 2 BHK flat for sale...         98 Lac
2  Up for immediate sale is a 2 BHK apartment in ...        1.40 Cr
3  This beautiful 1 BHK Flat is available for sal...        25 Lac
4  This lovely 2 BHK Flat in Pokhran Road, Thane ...        1.60 Cr

      Price (in rupees) location Carpet Area      Status      Floor \
0      6000.0    thane   500 sqft Ready to Move  10 out of 11
1     13799.0    thane   473 sqft Ready to Move   3 out of 22
2     17500.0    thane   779 sqft Ready to Move  10 out of 29
3        NaN    thane   530 sqft Ready to Move   1 out of 3
4     18824.0    thane   635 sqft Ready to Move  20 out of 42

      Transaction ... facing      overlooking \
0      Resale ...   NaN           NaN
1      Resale ...   East        Garden/Park
2      Resale ...   East        Garden/Park
3      Resale ...   NaN           NaN
4      Resale ...  West  Garden/Park, Main Road

      Society Bathroom Balcony Car Parking \
0  Srushti Siddhi Mangal Murti Complex      1      2      NaN
1            Dosti Vihar      2      NaN      1 Open
2      Sunrise by Kalpataru      2      NaN      1 Covered
3            NaN             1      1      NaN
4  TenX Habitat Raymond Realty      2      NaN      1 Covered

      Ownership Super Area Dimensions Plot Area
0            NaN      NaN      NaN      NaN
1      Freehold      NaN      NaN      NaN
2      Freehold      NaN      NaN      NaN
3            NaN      NaN      NaN      NaN
4  Co-operative Society      NaN      NaN      NaN

[5 rows x 21 columns]
```

### Tahap 3: Integrasi Data ke dalam Sistem

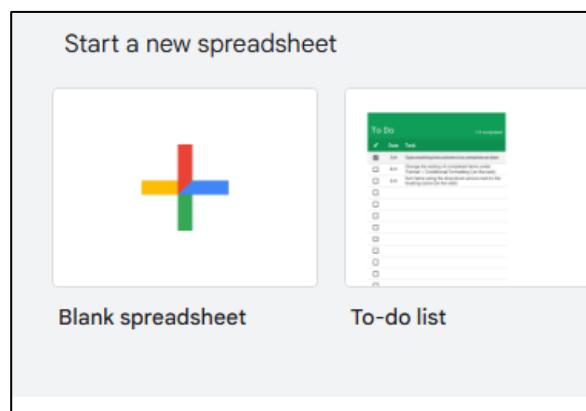
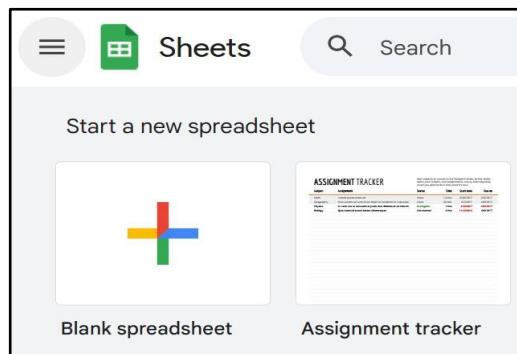
Langkah-langkah praktikum dalam integrasi data ke dalam sistem

- Integrasikan data yang telah dibersihkan ke dalam database atau sistem penyimpanan data dengan memperhatikan beberapa hal berikut:
  - Gunakan tools basis data untuk menyimpan data, memastikan data diatur dengan baik dan dapat diakses untuk analisis di masa depan.
  - Verifikasi bahwa data terintegrasi dengan benar dan tidak ada kesalahan selama proses transfer.

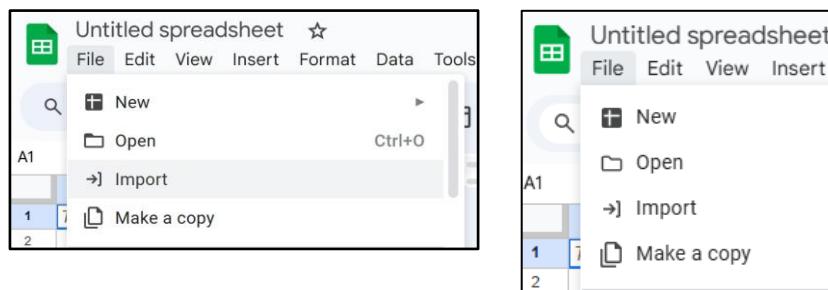
Berikut adalah contoh untuk integrasi data ke dalam sistem:

- Menggunakan Google Spreadsheets

Buka Google Spreadsheets dan pilih “Blank spreadsheet”



Pada menu File pilih “Import” dan pilih file dataset yang sudah di download.

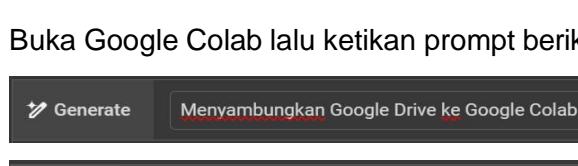


Beri nama file untuk dataset yang sudah diimport.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	sqft_above	sqft_basement	yr_built	yr_renovated	street		
2	2014-05-02 00:00	313000	3	1.5	1340	7912	1.5	0	0	3	1340	0	1955	2005	18811		
3	2014-05-02 00:00	2384000	5	2.5	3650	9050	2	0	4	5	3370	280	1921	0	709 V		
4	2014-05-02 00:00	342000	3	2	1930	11947	1	0	0	4	1930	0	1966	0	26201		
5	2014-05-02 00:00	420000	3	2.25	2000	8030	1	0	0	4	1000	1000	1963	0	857 1		
6	2014-05-02 00:00	550000	4	2.5	1940	10500	1	0	0	4	1140	800	1976	1992	9105		
7	2014-05-02 00:00	490000	2	1	880	6380	1	0	0	3	880	0	1938	1994	522 N		
8	2014-05-02 00:00	335000	2	2	1350	2560	1	0	0	3	1350	0	1976	0	2616		
9	2014-05-02 00:00	482000	4	2.5	2710	35888	2	0	0	3	2710	0	1989	0	2376;		
10	2014-05-02 00:00	452500	3	2.5	2430	88426	1	0	0	4	1570	860	1985	0	4661		
11	2014-05-02 00:00	640000	4	2	1520	6200	1.5	0	0	3	1520	0	1945	2010	6811		
12	2014-05-02 00:00	463000	3	1.75	1710	7320	1	0	0	3	1710	0	1948	1994	Burke		
13	2014-05-02 00:00	1400000	4	2.5	2920	4000	1.5	0	0	5	1910	1010	1909	1988	3838		
14	2014-05-02 00:00	588500	3	1.75	2330	14892	1	0	0	3	1970	360	1980	0	1833		
15	2014-05-02 00:00	365000	3	1	1090	6435	1	0	0	4	1090	0	1955	2009	2504		
16	2014-05-02 00:00	1200000	5	2.75	2910	9480	1.5	0	0	3	2910	0	1939	1969	3534		
17	2014-05-02 00:00	242500	3	1.5	1200	9720	1	0	0	4	1200	0	1965	0	1403;		
18	2014-05-02 00:00	419000	3	1.5	1570	6700	1	0	0	4	1570	0	1956	0	1542;		
19	2014-05-02 00:00	367500	4	3	3110	7231	2	0	0	3	3110	0	1997	0	1122		
20	2014-05-02 00:00	257950	3	1.75	1370	5858	1	0	0	3	1370	0	1987	2000	1605		
21	2014-05-02 00:00	275000	3	1.5	1180	10277	1	0	0	3	1180	0	1983	2009	1242		
22	2014-05-02 00:00	750000	3	1.75	2240	10578	2	0	0	5	1550	690	1923	0	3225		
23	2014-05-02 00:00	435000	4	1	1450	8800	1	0	0	4	1450	0	1954	1979	3922		
24	2014-05-02 00:00	626000	3	2.25	1750	1572	2.5	0	0	3	1470	280	2005	0	3140		

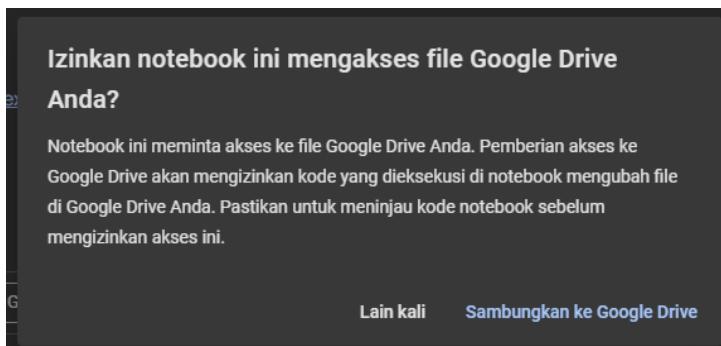
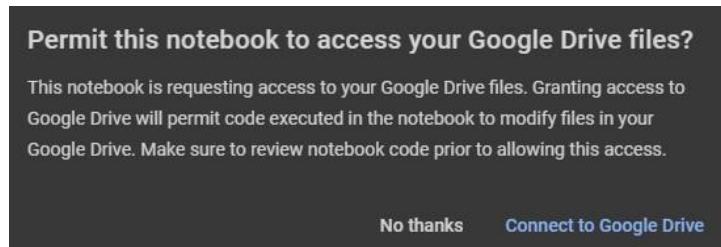
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Index	Title	Description	Amount(in rupees)	Price (in rupees)	location	Carpet Area	Status	Floor	Furnishing	facing	overlooking	Society	Bathroom	Balcony	Car Parking	Ownership
2	0.1	BHK Ready to Occupy	Flat for sale in Shruti Siddhi Mangal Murti Complex Bhivandi,"Bhivandi, Thane	has an attractive 1 BHK Flat for sale. The property is ideally located in a strategic location in Shruti Siddhi Mangal Murti Complex township. This flat for resale is a choice property. This apartment ready for move in is available at a reasonable price.													
3	1.2	BHK Ready to Occupy	Flat for sale in Dosti Vihar Pokhran Road,"One can find this stunning 2 BHK flat for sale in Pokhran Road, Thane. It enjoys an excellent location within the Dosti Vihar. This flat for resale is a choice property. This ready to move flat in Pokhran Road can be availed at a reasonable price.														
4	2.2	BHK Ready to Occupy	Flat for sale in Sunrise by Kalpataru Kolshet Road,"Up for immediate sale is a 2 BHK apartment in Kolshet Road, Thane. Don't miss this bargain flat for sale. Situated in the Sunrise By Kalpataru township, it has a prime location. This flat for resale has a desirable location. You can buy this flat for resale at a reasonable price.														
5	3.1	BHK Ready to Occupy	Flat for sale Kasheli,"This beautiful 1 BHK Flat is available for sale in Kasheli, Thane. This flat for resale has a desirable location. This ready to move flat is offered at an economic price of INR 25 Lac. You will find it unfurnished. This flat for resale is a choice property. This flat is an attractive property for resale. You can buy this flat for resale at a reasonable price.														
6	4.2	BHK Ready to Occupy	Flat for sale in Texn Habitat Raymond Realty Pokhran Road,"This lovely 2 BHK Flat in Pokhran Road, Thane is up for sale. This flat is situated in the Texn Habitat Raymond Realty township and is equipped with premium facilities. This flat is an attractive property for resale. You can buy this flat for resale at a reasonable price.														
7	5.1	BHK Ready to Occupy	Flat for sale in Virat Aangan Tittwala,"Creatively planned and constructed is a 1 BHK flat for sale in Virat Aangan Tittwala. This flat is available as a resale property. You can buy this ready to move flat in Virat Aangan Tittwala at a reasonable price.														
8	6.1	BHK Ready to Occupy	Flat for sale in Mumbra,"This magnificent 1 BHK Flat is available for sale in Mumbra, Thane. The premium flat is available for resale at an unbelievable price, so, grab it before it's gone! Located in Mumbra, this ready to move apartment is sold at a fair selling price of INR 16.5 Lac. This flat for resale is a choice property. This flat for resale is a choice property. You can buy this flat for resale at a reasonable price.														
9	7.1	BHK Ready to Occupy	Flat for sale in Kalwa,"Creatively planned and constructed is a 1 BHK flat for sale in Kalwa, Thane. This flat is an attractive property for resale. This ready to move flat located in Kalwa is available for purchase at a fair price of INR 60 Lac. This furnished flat is strategically designed with all modern amenities.														
10	8.1	BHK Ready to Occupy	Flat for sale in Kalwa,"Discover this immaculate 1 BHK flat for sale at the pristine Kalwa in Thane. This apartment is a property of choice for resale. This ready to move flat is offered at an economical price of INR 25 Lac. This flat for resale is a choice property. You can buy this flat for resale at a reasonable price.														
11	9.3	BHK Ready to Occupy	Flat for sale in Pride Palms Kolshet,"One can find this stunning 3 BHK flat for sale in Kolshet Road, Thane. Ideally situated in the Pride Palms township it enjoys a prime location. This flat for resale is the perfect property for you! This ready to move flat in Kolshet Road is available at a low-priced budget of INR 35 Lac. This flat for resale is a choice property. You can buy this flat for resale at a reasonable price.														
12	10.3	BHK Ready to Occupy	Flat for sale in Cosmos Lounge Manpada Thane West,"Up for immediate sale is a 3 BHK apartment in Manpada Thane West, Thane. Don't miss this bargain flat for sale. It is in a prime location within the Cosmos Lounge. Invest your valuable money in this flat that is for resale. This flat for resale is a choice property. You can buy this flat for resale at a reasonable price.														
13	11.2	BHK Ready to Occupy	Flat for sale in Regency Heights Kolshet Road,"2 BHK Multistorey Apartment is available for sale in Kolshet Road, Thane for 1.3 Crore(*).1.36 Cr.,11674,thane,,Ready to Move,16 out of 24,Resale,Semi-Furnished,,Regency Heights,2,,1165 sqft,														
14	12.2	BHK Ready to Occupy	Flat for sale in Regency Heights Kolshet Road,"2 BHK Multistorey Apartment is available for sale in Kolshet Road, Thane for 1.3 Crore(*).1.35 Cr.,15995,thane,,Ready to Move,8 out of 20,Resale,Semi-Furnished,,Regency Heights,2,,844 sqft,														
15	13.4	BHK Ready to Occupy	Flat for sale in Hirandani Estate,"Creatively planned and constructed is a 4 BHK flat for sale in Hirandani Estate, Thane. This flat is an attractive property for resale. This ready to move flat in Hirandani Estate comes at an affordable price of INR 4.25 Cr. It is a semi-furnished apartment.														
16	14.1	BHK Ready to Occupy	Flat for sale in Rabodi,"Discover this immaculate 1 BHK flat for sale at the pristine Rabodi in Thane. This premium flat is available for resale at an unbelievable price, so, grab it before it's gone! The ready to move flat in the prime area of Rabodi is available at a reasonable price of INR 16 Lac.														
17	15.2	BHK Ready to Occupy	Flat for sale in Puraniks Tokyo Bay Kasarvadavali,"Kasarvadavali, Thane has an appealing 2 BHK flat for sale with various amenities. It is housed in the well-planned Puraniks Tokyo Bay township in an advantageous location. This is a desirable apartment for sale. This apartment in Kasarvadavali is up for sale. Enjoy its prime location and modern amenities.														
18	16.1	BHK Ready to Occupy	Flat for sale in Green Woods Shilpata,"Have a look at this immaculate 2 BHK flat for sale in Shilpata, Thane. This flat for resale is a prime location. This premium flat is available for sale at an unbelievable price, so, grab it before it's gone! This ready to move flat in Shilpata is available at a low-priced budget of INR 35 Lac. This flat for resale is a choice property. You can buy this flat for resale at a reasonable price.														
19	17.1	BHK Ready to Occupy	Flat for sale in Dombivli West,"This magnificent 1 BHK flat is available for sale in Dombivli West, Thane. This flat for resale has a desirable location. The ready to move flat in Dombivli West is all ready for sale at a low-priced budget of INR 35 Lac. This flat is an ideal choice because it is located in a prime location.														
20	18.2	BHK Ready to Occupy	Flat for sale in Everest World Lavender Kolshet,"Discover this immaculate 2 BHK flat for sale at the pristine Kolshet in Thane. Take the experience of contemporary living to a new high, with exquisitely designed apartments at Everest World Lavender. This flat for resale is a choice property. You can buy this flat for resale at a reasonable price.														
21	19.1	BHK Ready to Occupy	Flat for sale in Shahad,"This lovely 1 BHK Flat in Shahad, Thane is up for sale. This apartment for resale is a great choice. The ready to move flat in the prime area of Shahad is available at a reasonable price of INR 35 Lac. It is a semi-furnished apartment.														
22	20.1	BHK Ready to Occupy	Flat for sale in Pride Residency Kasarvadavali,"Creatively planned and constructed is a 1 BHK flat for sale in Kasarvadavali, Thane. This flat is situated within the renowned township of Pride Residency. Your hunt for the perfect apartment for resale comes to an end here. This ready to move flat in Kasarvadavali is available at a reasonable price.														
23	21.2	BHK Ready to Occupy	Flat for sale in Ashar Edge Pokhran Road Number 2,"This gorgeous 2 BHK Flat is available for sale in Pokhran Road Number 2, Thane. Enjoying a prime location, this property is housed in the Ashar Edge society. This is one of the best properties available for resale. You can buy this flat for resale at a reasonable price.														
24	22.2	BHK Ready to Occupy	Flat for sale in Lohda Amara Kolshet Road,"Up for immediate sale is a 2 BHK apartment in Kolshet Road, Thane. Don't miss this bargain flat for sale. Ideally situated in the Lohda Amara township it enjoys a prime location. This flat for resale is a choice property. You can buy this flat for resale at a reasonable price.														
25	23.3	BHK Ready to Occupy	Flat for sale in Ghodbunder Road,"Creatively planned and constructed is a 3 BHK flat for sale in Ghodbunder Road, Thane. Situated in the excellent Tierra Cop Hsg Soc Ltd township. This flat enjoys a prime location. Your search ends here, because this flat for resale is a choice property. You can buy this flat for resale at a reasonable price.														
26	24.1	BHK Ready to Occupy	Flat for sale in Mumbra,"This is a desirable apartment for resale. The ready to move flat in Mumbra is all ready for sale at a low-priced budget of INR 7.5 Lac. It is unoccupied and available to accommodate you.														
27	25.1	BHK Ready to Occupy	Flat for sale in Kalwan,"A BHK flat available for sale in Thane in the prime location of Kalwan. Your search ends here, because this flat for resale is among the best bargains in town. The ready to move flat in Kalwan is all ready for sale at a low-priced budget of INR 4.2 Lac. This flat for resale is a choice property. You can buy this flat for resale at a reasonable price.														

Buka Google Colab lalu ketikan prompt berikut:



```
from google.colab import drive
drive.mount('/content/drive')
```

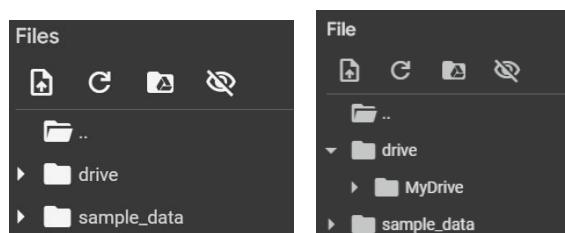
Setelah kode dijalankan akan muncul sebuah notifikasi, pilih “Connect to Google Drive”.



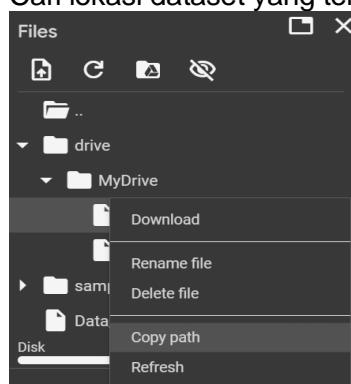
Berikut output jika sudah tersambung:

```
Mounted at /content/drive
1m  from google.colab import drive
      drive.mount('/content/drive')
Mounted at /content/drive
```

Pada menu Files akan muncul folder “drive”



Cari lokasi dataset yang telah diupload dan pilih “copy path”.



Baca data menggunakan *pandas*.

```
# data = pd.read_csv('/content/drive/MyDrive/data.csv')

print(data.head())

      date      price  bedrooms  bathrooms  sqft_living  sqft_lot
0  2014-05-02 00:00:00  313000.0        3.0       1.50      1340    7912
1  2014-05-02 00:00:00  2384000.0       5.0       2.50      3650   9050
2  2014-05-02 00:00:00  342000.0        3.0       2.00      1930  11947
3  2014-05-02 00:00:00  420000.0        3.0       2.25      2000   8030
4  2014-05-02 00:00:00  550000.0        4.0       2.50      1940  10500

      floors  waterfront  view  condition  sqft_above  sqft_basement  yr_built \
0      1.5          0     0           3        1340                  0      1955
1      2.0          0     4           5        3370                  280     1921
2      1.0          0     0           4        1930                  0     1966
3      1.0          0     0           4        1000                  1000     1963
```

Baca dan tampilkan 5 data pertama pada data house\_prices.csv di dalam Google Drive

Gunakan kode dengan hati-hati

```
# prompt: Baca dan tampilkan 5 data pertama pada data house_prices.csv di dalam Google Drive

import pandas as pd

# Path to your CSV file in Google Drive
file_path = '/content/drive/MyDrive/house_prices.csv' # Replace with your actual file path

try:
    # Read the CSV file into a pandas DataFrame
    df = pd.read_csv(file_path)

    # Display the first 5 rows of the DataFrame
    print(df.head())

except FileNotFoundError:
    print(f"Error: File not found at {file_path}")
except pd.errors.EmptyDataError:
    print(f"Error: The file at {file_path} is empty.")
except pd.errors.ParserError:
    print(f"Error: Unable to parse the file at {file_path}. Check the file format.")
except Exception as e:
    print(f"An unexpected error occurred: {e}")
```

Index	Title \
0	0 1 BHK Ready to Occupy Flat for sale in Srushti...
1	1 2 BHK Ready to Occupy Flat for sale in Dosti V...
2	2 2 BHK Ready to Occupy Flat for sale in Sunrise...
3	3 1 BHK Ready to Occupy Flat for sale Kasheli
4	4 2 BHK Ready to Occupy Flat for sale in TenX Ha...

	Description	Amount(in rupees) \
0	Bhiwandi, Thane has an attractive 1 BHK Flat f...	42 Lac
1	One can find this stunning 2 BHK flat for sale...	98 Lac
2	Up for immediate sale is a 2 BHK apartment in ...	1.40 Cr
3	This beautiful 1 BHK Flat is available for sal...	25 Lac
4	This lovely 2 BHK Flat in Pokhran Road, Thane ...	1.60 Cr

	Price (in rupees)	location	Carpet Area	Status	Floor	\
0	6000.0	thane	500 sqft	Ready to Move	10 out of 11	
1	13799.0	thane	473 sqft	Ready to Move	3 out of 22	
2	17500.0	thane	779 sqft	Ready to Move	10 out of 29	
3	Nan	thane	530 sqft	Ready to Move	1 out of 3	
4	18824.0	thane	635 sqft	Ready to Move	20 out of 42	

	Transaction	... facing	overlooking \
0	Resale	... Nan	Nan
1	Resale	... East	Garden/Park
2	Resale	... East	Garden/Park
3	Resale	... Nan	Nan
4	Resale	... West	Garden/Park, Main Road

	Society	Bathroom	Balcony	Car	Parking	\
0	Srushti Siddhi Mangal Murti Complex	1	2	Nan		
1	Dosti Vihar	2	Nan	1	Open	
2	Sunrise by Kalpataru	2	Nan	1	Covered	
3	Nan	1	1	Nan		
4	TenX Habitat Raymond Realty	2	Nan	1	Covered	

	Ownership	Super Area	Dimensions	Plot Area
0	Nan	Nan	Nan	Nan
1	Freehold	Nan	Nan	Nan
2	Freehold	Nan	Nan	Nan
3	Nan	Nan	Nan	Nan
4	Co-operative Society	Nan	Nan	Nan

[5 rows x 21 columns]

# Tugas Praktikum

## 1. Identifikasi Kebutuhan Data

- Buat daftar fitur yang diperlukan untuk proyek data science yang kalian pilih, jelaskan alasan pemilihan setiap fitur, dan periksa ketersediaan data tersebut.
- **Tugas:** Jelaskan secara singkat mengapa setiap fitur penting untuk tujuan proyek Anda.

## 2. Pengambilan Data Menggunakan Kaggle dan Tools Lain

- Lakukan pengambilan data menggunakan tools yang telah dipelajari (misalnya, Python, API data, Kaggle).
- **Tugas:** Serahkan skrip yang digunakan untuk mengambil data beserta hasilnya. Tulis deskripsi singkat tentang cara kerja skrip tersebut dan bagaimana data diambil.

## 3. Integrasi Data ke dalam Database atau Repository

- Unggah data ke dalam database atau sistem penyimpanan data.
- **Tugas:** Dokumentasikan proses integrasi data, termasuk tools yang digunakan dan verifikasi bahwa data terintegrasi dengan benar.

## Jawaban

### 1. Identifikasi Kebutuhan Data

Proyek: Analisis Penyebaran COVID-19 di Indonesia

Fitur yang Dibutuhkan:

1. Tanggal - Untuk melihat perkembangan kasus dari waktu ke waktu.
2. Provinsi - Untuk menganalisis distribusi penyebaran COVID-19 di berbagai daerah.
3. Jumlah Kasus Positif - Indikator utama untuk melihat tren penyebaran virus.
4. Jumlah Kasus Sembuh - Mengetahui tingkat pemulihan pasien COVID-19.
5. Jumlah Kasus Meninggal - Untuk menganalisis tingkat fatalitas akibat COVID-19.
6. Jumlah Tes COVID-19 - Melihat efektivitas dan cakupan pengujian di Indonesia

Ketersediaan Data:

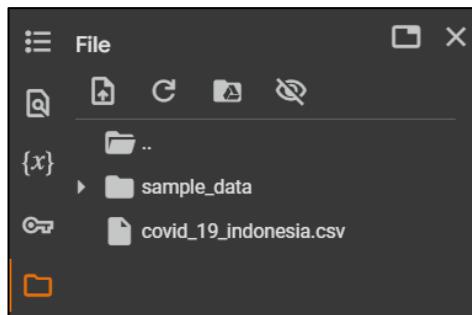
Dataset *covid\_19\_indonesia.csv* yang telah diunggah berisi informasi yang mencakup fitur-fitur di atas.

### 2. Pengambilan Data Menggunakan Kaggle

1. Masuk ke Kaggle dan cari dataset terkait COVID-19 di Indonesia.

2. Unduh dataset dalam format CSV dari halaman dataset yang dipilih.

3. Unggah dataset ke Google Colab dan melakukan pengolahan data lainnya untuk diproses lebih lanjut.



4. Gunakan Pandas untuk membaca dataset dalam format CSV dan melakukan eksplorasi awal data.

- Mengetikkan prompt pada generate with AI di Google Colab

- Menghasilkan kode berikut

```
# prompt: Baca dan tampilkan 5 data pertama pada data covid_19_indonesia.csv di file lokal Google Colab

import pandas as pd

# Assuming the file is named 'covid_19_indonesia.csv' and is in the Colab environment's local file system
try:
    df = pd.read_csv('covid_19_indonesia.csv')
    print(df.head())
except FileNotFoundError:
    print("Error: 'covid_19_indonesia.csv' not found in the current directory.")
except pd.errors.ParserError:
    print("Error: Could not parse the CSV file. Please check its format.")
except Exception as e:
    print(f"An unexpected error occurred: {e}")
```

- Menampilkan 5 data pertama menggunakan pandas untuk file CSV

The screenshot shows a Jupyter Notebook cell containing a Pandas DataFrame. The DataFrame has 5 rows and 38 columns. The columns are grouped by index and include:

- Date Location ISO Code Location**: Rows 0-4.
- New Cases New Deaths \**: Rows 0-4.
- New Recovered New Active Cases Total Cases Total Deaths \**: Rows 0-4.
- Total Recovered ... Latitude New Cases per Million \**: Rows 0-4.
- 75 ... -6.204699 0.18**
- 75 ... -6.204699 0.18**
- 0 ... -0.789275 0.01**
- 1 ... 0.511648 0.16**
- 75 ... -6.204699 0.18**
- Total Cases per Million New Deaths per Million Total Deaths per Million \**: Rows 0-4.
- 3.60 0.0 1.84**
- 3.78 0.0 1.84**
- 0.01 0.0 0.00**
- 0.16 0.0 0.00**
- 3.96 0.0 1.84**
- Total Deaths per 100rb Case Fatality Rate Case Recovered Rate \**: Rows 0-4.
- 0.18 51.28% 192.31%**
- 0.18 48.78% 182.93%**
- 0.00 0.00% 0.00%**
- 0.00 0.00% 100.00%**
- 0.18 46.51% 174.42%**
- Growth Factor of New Cases Growth Factor of New Deaths**: Rows 0-4.
- NaN NaN**
- 1.0 1.0**
- NaN NaN**
- NaN NaN**
- 1.0**

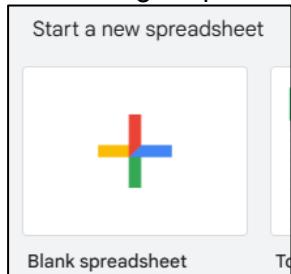
[5 rows x 38 columns]

covid\_19\_indonesia.csv

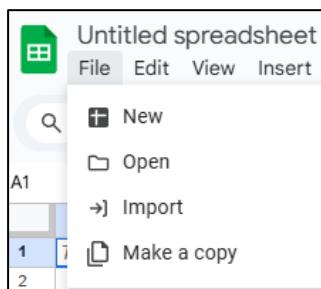
### 3. Integrasi Data ke dalam Sistem

#### 1. Menggunakan Spreadsheets

Buka Google Spreadsheets dan pilih “Blank spreadsheet”



#### 2. Mengimport data set

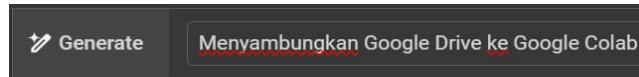


A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Date	Location ISO Co-Location	New Cases	New Deaths	New Recovered	New Active Cases	Total Cases	Total Deaths	Total Recovered	Total Active Cases	Location Level	City or Regency	Province
2	3/1/2020 ID-JK	DKI Jakarta	2	0	0	2	39	20	75	-56 Province		DKI Jakarta	
3	3/2/2020 ID-JK	DKI Jakarta	2	0	0	2	41	20	75	-54 Province		DKI Jakarta	
4	3/2/2020 IDN	Indonesia	2	0	0	2	42	0	0	2 Country			
5	3/2/2020 ID-RI	Riau	1	0	0	1	43	20	75	-52 Province		Riau	
6	3/3/2020 ID-JK	DKI Jakarta	2	0	0	2	45	20	75	-60 Province		DKI Jakarta	
7	3/3/2020 IDN	Indonesia	0	0	0	0	2	0	0	2 Country			
8	3/3/2020 ID-JB	Jawa Barat	1	1	0	0	1	1	60	-59 Province		Jawa Barat	
9	3/3/2020 ID-RI	Riau	0	0	0	0	1	0	1	0 Province		Riau	
10	3/4/2020 ID-JK	DKI Jakarta	2	0	0	2	45	20	75	-50 Province		DKI Jakarta	
11	3/4/2020 IDN	Indonesia	0	0	0	0	2	0	0	2 Country			
12	3/4/2020 ID-JB	Jawa Barat	1	0	0	1	2	1	60	-59 Province		Jawa Barat	
13	3/4/2020 ID-RI	Riau	0	0	0	0	1	0	1	0 Province		Riau	
14	3/5/2020 ID-JK	DKI Jakarta	0	1	0	-1	45	21	75	-51 Province		DKI Jakarta	
15	3/5/2020 IDN	Indonesia	0	0	0	0	2	0	0	2 Country			
16	3/5/2020 ID-JB	Jawa Barat	1	0	0	1	3	1	60	-58 Province		Jawa Barat	
17	3/5/2020 ID-RI	Riau	0	0	0	0	1	0	1	0 Province		Riau	
18	3/6/2020 ID-BT	Banten	1	0	1	0	1	5	111	-115 Province		Banten	
19	3/6/2020 ID-JK	DKI Jakarta	0	0	0	0	45	21	75	-51 Province		DKI Jakarta	
20	3/6/2020 IDN	Indonesia	2	0	0	2	4	0	0	4 Country			

### 3. Menggunakan Google Drive Mengaploud data set ke Google Drive



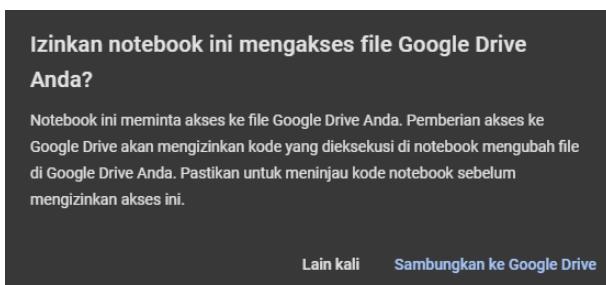
### 4. Buka Google Colab lalu ketikan prompt berikut



### 5. Lalu akan muncul kode berikut



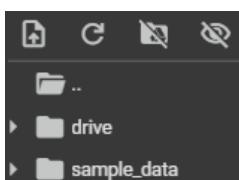
### 6. Setelah kode dijalankan akan muncul sebuah notifikasi, pilih “Connect to Google Drive”.



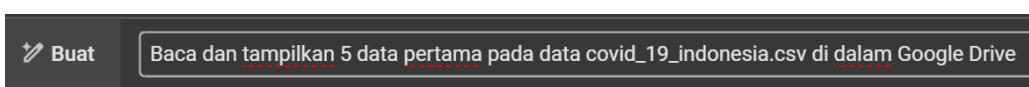
### 7. output jika sudah tersambung



### 8. Pada menu Files akan muncul folder “drive”



### 9. Mengetikkan prompt pada generate with AI di Google Colab



## 10. Menghasilkan kode berikut

```
▶ Kode yang dihasilkan mungkin tunduk pada lisensi |  
# prompt: Baca dan tampilkan 5 data pertama pada data covid_19_indonesia.csv di dalam Google Drive  
  
import pandas as pd  
  
# Mount Google Drive  
from google.colab import drive  
drive.mount('/content/drive')  
  
# Read the csv file from Google Drive  
file_path = '/content/drive/My Drive/covid_19_indonesia.csv' # Replace with the actual path  
try:  
    df = pd.read_csv(file_path)  
    # Display the first 5 rows  
    print(df.head())  
except FileNotFoundError:  
    print(f"Error: File not found at {file_path}")  
except pd.errors.ParserError:  
    print(f"Error: Unable to parse the CSV file at {file_path}. Please check the file format.")  
except Exception as e:  
    print(f>An unexpected error occurred: {e})
```

## 11. Menampilkan data awal

```
▶ Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).  
   Date Location ISO Code  Location New Cases New Deaths \  
0 3/1/2020           ID-JK  DKI Jakarta      2        0  
1 3/2/2020           ID-JK  DKI Jakarta      2        0  
2 3/2/2020            IDN  Indonesia      2        0  
3 3/2/2020           ID-RI       Riau       1        0  
4 3/3/2020           ID-JK  DKI Jakarta      2        0  
  
   New Recovered New Active Cases Total Cases Total Deaths \  
0          0         0        2       39        20  
1          0         0        2       41        20  
2          0         0        2       41        20  
3          0         0        1       41        20  
4          0         0        2       43        20  
  
   Total Recovered ... Latitude New Cases per Million \  
0           75     ... -6.204699        0.18  
1           75     ... -6.204699        0.18  
2            0     ... -0.789275        0.01  
3            1     ...  0.511648        0.16  
4           75     ... -6.204699        0.18  
  
   Total Cases per Million New Deaths per Million Total Deaths per Million \  
0             3.60            0.0            1.84  
1             3.78            0.0            1.84  
2             0.01            0.0            0.00  
3             0.16            0.0            0.00  
4             3.96            0.0            1.84  
  
   Total Deaths per 100rb Case Fatality Rate Case Recovered Rate \  
0            0.18          51.28%        192.31%  
1            0.18          48.78%        182.93%  
2            0.00          0.00%        0.00%  
3            0.00          0.00%        100.00%  
4            0.18          46.51%        174.42%  
  
   Growth Factor of New Cases Growth Factor of New Deaths  
0                  NaN            NaN  
1                  1.0            1.0  
2                  NaN            NaN  
3                  NaN            NaN  
4                  1.0            1.0  
  
[5 rows x 38 columns]
```

