



# Introduction to Python with Google Colab

**HARIS ARIEF KAMIS**

**Data Analyst**  
**FULLSTACK INTENSIVE BOOTCAMP**

harisariefkamis





# Final Project

LEARN PYTHON BASIC USING GOOGLE COLAB





## PELAJARI DASAR PYTHON MENGGUNAKAN GOOGLE COLAB

Silahkan untuk “Save a Copy in Drive” ke google drive masing-masing

1. Buka [google colab](#) (click)
2. File > Save a copy in drive
3. Silahkan untuk edit dan ketik pada file colab masing-masing
4. Kamu juga bisa mererefresh tsb untuk [google colab](#) melihat codingan terbaru



<https://colab.research.google.com/drive/1g5ICJHsmBhNCNEpj69TndS-uDDIdDwX2#scrollTo=pS0RDBljahpn>



[https://github.com/harisariefkamis/Final-Project-DAB14/blob/main/Final\\_Project\\_Python.ipynb](https://github.com/harisariefkamis/Final-Project-DAB14/blob/main/Final_Project_Python.ipynb)



# Outline



- DEFENISI GOOGLLE COLABORATORY
- Apa itu Google Colab ?
- Fitur Google Colab
- Bahasa Python For Google Colab
- Definisi Bahasa Python
- Keuntungan Menggunakan Python
- Final Project Python with Google Colab

# Apa itu Google Colab ?



## what is google collaboration?

Google Colab, or Google Colaboratory, is a cloud-based platform that allows you to write, run, and share **Python** code via a web browser. In short, Google Colab is a valuable tool for anyone who wants to work with Python code, especially in the fields of data science, machine learning , and programming.

# Fitur Google Colab



## **Fitur utama Google Colab meliputi:**

1. Anda dapat menulis kode dalam sel yang dapat dijalankan secara individual dan melihat outputnya langsung di browser.
2. Tidak perlu instalasi software atau pengaturan lingkungan lokal. Gunakan GPU dan TPU gratis untuk tugas komputasi berat.
3. Bagikan notebook Colab Anda dengan orang lain untuk bekerja bersama secara real-time.
4. Gunakan dengan akun Google Anda secara GRATIS !!!

# Python for Google Colab



## Python programming language and definition?

### Definition:

Programming language: A language used to give instructions to a computer. High level: Easier for humans to read and write than machine languages. Versatile: Can be used for a wide variety of tasks. Easy to learn: Python syntax is easy to understand, even for people who are just learning programming. Widely used: In various fields and industries.

Python is a high-level programming language that is versatile, easy to learn, and widely used. This language is popular in various fields, including:

### Keuntungan:

1. Syntax Python mudah dipahami, bahkan bagi orang yang baru belajar pemrograman.
2. Kode Python dapat dijalankan di berbagai platform, seperti Windows, Mac, dan Linux.
3. Python tersedia secara gratis dan dapat diubah dan didistribusikan oleh siapa saja.
4. Ada komunitas Python yang besar dan aktif yang dapat membantu Anda belajar dan memecahkan masalah.

# Python for Google Colab



**Sebelum dan sesudah memulai pengerjaan Berdo'a dulu ya guys!!:**  
**> Save a copy in Drive anda' sebelum memulai workspace baru**

IMPORT DATA DI GOOGLE COLAB

instal pandas,numpy,matplotlib,scikit-learn,dan tensorflow

```
!pip install pandas
```

```
!pip install numpy
```

```
!pip install matplotlib
```

```
!pip install seaborn
```

```
# !pip install openpyxl
```

```
# !pip install scikit-learn
```

```
# !pip install tensorflow
```

```
import pandas as pd
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
from pandas.tseries.offsets import BDay
```

```
# import openpyxl as xlsx
```

```
# import scikit-learn. as scln
```

```
# import pandas  
# import numpy  
# import openpyxl  
# import matplotlib  
# import scikit-learn  
# import tensorflow  
# import seaborn  
# from pandas.tseries.offsets import BDay
```

Membuat dan  
menampilkan  
komentar di code  
kamu





# Python for Google Colab

variable	class	description
<b>order_detail:</b>		
id	object	angka unik dari order / id_order
customer_id	object	angka unik dari pelanggan
order_date	object	tanggal saat dilakukan transaksi
sku_id	object	angka unik dari produk (sku adalah stock keeping unit)
price	int64	harga yang tertera pada tagging harga
qty_ordered	int64	jumlah barang yang dibeli oleh pelanggan
before_discount	float64	nilai harga total dari produk ( $price * qty\_ordered$ )
discount_amount	float64	nilai diskon produk total
after_discount	float64	nilai harga total produk ketika sudah dikurangi dengan diskon
is_gross	int64	menunjukkan pelanggan belum membayar pesanan
is_void	int64	menunjukkan pelanggan sudah melakukan pembayaran
is_ref	int64	menunjukkan transaksi sudah selesai
payment_id	int64	angka unik dari metode pembayaran
<b>sku_detail:</b>		
id	object	angka unik dari produk (dapat digunakan untuk key saat join)
sku_name	object	nama dari produk
base_price	float64	harga barang yang tertera pada tagging harga / price
costs	int64	cost of goods sold / total biaya untuk menjual 1 produk
category	object	kategori produk
<b>customer_detail:</b>		
id	object	angka unik dari pelanggan
registered_date	object	tanggal pelanggan mulai mendaftarkan diri sebagai anggota
<b>payment_detail:</b>		
id	int64	angka unik dari metode pembayaran
payment_method	object	metode pembayaran yang digunakan

Dataset:

Data yang digunakan adalah data yang berasal dari Tokopedia (***bukan data sesungguhnya***). Mengenai penjelasan dataset adalah sebagai berikut:



# Python for Google Colab



# @title SOURCE DATA

```
path_od = "https://raw.githubusercontent.com/dataskillsboost/FinalProjectDA11/main/order_detail.csv" #data order_detail
path_pd = "https://raw.githubusercontent.com/dataskillsboost/FinalProjectDA11/main/payment_detail.csv" #data
payment_detail
path_cd = "https://raw.githubusercontent.com/dataskillsboost/FinalProjectDA11/main/customer_detail.csv" #data
customer_detail
path_sd = "https://raw.githubusercontent.com/dataskillsboost/FinalProjectDA11/main/sku_detail.csv" #data sku_detail
```

```
df_project_od
=pd.read_csv('https://raw.githubusercontent.com/dataskillsboost/FinalProjectDA11/main/o
rder_detail.csv')
df_project_cd =
pd.read_csv('https://raw.githubusercontent.com/dataskillsboost/FinalProjectDA11/main/pa
yment_detail.csv')
df_project_pd =
pd.read_csv('https://raw.githubusercontent.com/dataskillsboost/FinalProjectDA11/main/cu
stomer_detail.csv')
df_project_sd =
pd.read_csv('https://raw.githubusercontent.com/dataskillsboost/FinalProjectDA11/main/sk
u_detail.csv')

print ('df_project_od')
print ('df_project_pd')
print ('df_project_cd')
print ('df_project_sd')
```

Mengimport dan  
membaca data  
file csv.excel



# Python for Google Colab



```
# menampilkan 10 baris dataframe(df)
df_project.head(10)
```

```
# menampilkan informasi dataframe(df)
df_project.info()
```

```
# menampilkan beberapa baris terakhir dari sebuah
dataframe.(df)
df_project.tail(5)
```

```
# menampilkan informasi statistik deskriptif
dataframe(df)
df_project.describe()
```

```
# menampilkan filter dataframe(df) berdasarkan
column
df_project.filter('customer_')
```

```
# mengelompokan dataframe(df) berdasarkan column
df_project.groupby('customer_id')
```

## #Menjalankan SQL di Colab

```
from sqlite3 import connect
conn = connect(':memory:')
df_project_od.to_sql('order_detail',conn, index=False,
if_exists='replace')
df_project_pd.to_sql('payment_detail', conn, index=False,
if_exists='replace')
df_project_cd.to_sql('sku_detail', conn, index=False,
if_exists='replace')
df_project_sd.to_sql('customer_detail', conn, index=False,
if_exists='replace')
```

Membuat dan menjalankan  
SQLite di lingkungan python



# Python for Google Colab



```
# menampilkan 5 baris dataframe(df)  
df_project.head(5)
```

```
df_project.head()
```

	id	customer_id	order_date	sku_id	price	qty_ordered	before_discount	discount_amount	after_discount	is_gros
0	ODR9939707760w	C713589L	2021-11-19	P858068	26100	200	5220000	2610000	2610000	
1	ODR7448356649d	C551551L	2021-11-19	P886455	1971942	5	9859710	2464927	7394782	
2	ODR4011281866z	C685596L	2021-11-25	P678648	7482000	1	7482000	2065344	5416655	
3	ODR3378927994s	C830683L	2021-11-22	P540013	3593680	1	3593680	1455440	2138239	
4	ODR4904430099k	C191766L	2021-11-21	P491032	4413220	1	4413220	1059172	3354047	

5 rows × 22 columns





### df\_project.dtypes

```
id            object
customer_id   object
order_date    object
sku_id        object
price         int64
qty_ordered   int64
before_discount  int64
discount_amount int64
after_discount  int64
is_gross      int64
is_valid      int64
is_net        int64
payment_id    int64
payment_method object
sku_name      object
base_price    int64
cogs          int64
category      object
registered_date object
day           object
month         object
month_num     int64
dtype: object
```

### #Mengubah tipe kolom Date menjadi Datetime

```
df_project['order_date']= pd.to_datetime(df_project['order_date'])
df_project['registered_date']=
pd.to_datetime(df_project['registered_date'])
df_project.dtypes
```

```
id            object
customer_id   object
order_date    datetime64[ns]
sku_id        object
price         int64
qty_ordered   int64
before_discount  int64
discount_amount int64
after_discount  int64
is_gross      int64
is_valid      int64
is_net        int64
payment_id    int64
payment_method object
sku_name      object
base_price    int64
cogs          int64
category      object
registered_date datetime64[ns]
day           object
month         object
month_num     int64
dtype: object
```



# Python for Google Colab



## No 1.

Dear Data Analyst,

Akhir tahun ini, perusahaan akan memberikan hadiah bagi pelanggan yang memenangkan kompetisi Festival Akhir Tahun. Tim Marketing membutuhkan bantuan untuk menentukan perkiraan hadiah yang akan diberikan pada pemenang kompetisi nantinya. Hadiah tersebut akan diambil dari TOP 5 Produk dari Kategori Mobiles & Tablets selama tahun 2022, dengan jumlah kuantitas penjualan (valid = 1) paling tinggi.

Mohon bantuan, untuk mengirimkan data tersebut sebelum akhir bulan ini ke Tim Marketing. Atas bantuan yang diberikan, kami mengucapkan terima kasih.

Regards

Tim Marketing

Solusi no. 1

Memfilter data dengan valid = 1

Memfilter data dengan Category = Mobiles & Tablets

Memfilter data pada transaksi selama 2022

Gunakan groupby berdasarkan sku\_name

Gunakan sort\_values untuk mengurutkan data

Gunakan head untuk menampilkan top 5 produk



# Python for Google Colab



```
df_project.info()
In [ ]:
# @title `Jawaban no.1` # Menambahkan filter pada data frame df_filter = df_project[(df_project['is_valid']==1)&\ (df_project.category == "Mobiles & Tablets")&\
(df_project.order_date>= "2022-01-01")&\ (df_project.order_date<= "2022-12-31")]
In [ ]:
# Mengelompokan dataframe su_name & qty_ordered pd.DataFrame(df_filter.groupby('sku_name')['qty_ordered'].sum())\
.reset_index().sort_values('qty_ordered',ascending=False)\ .reset_index(drop=True)\ .head()
Out[ ]:
```

	sku_name	qty_ordered
0	IDROID_BALRX7-Gold	1000
1	IDROID_BALRX7-Jet black	31
2	Infinix Hot 4-Gold	15
3	samsung_Grand Prime Plus-Black	11
4	infinix_Zero 4-Grey	10



# Python for Google Colab



```
# Menelompokkan dan menjumlahkan qty_ordered + customer_id
df_filter.groupby('sku_name')\ .agg({"qty_ordered": "sum", "customer_id": "nunique"})\
.head()
Out[ ]:
```

sku_name	qty_ordered	customer_id
Apple iPhone 6S Plus 16GB Silver	3	3
IDROID_BALRX7-Gold	1000	1
IDROID_BALRX7-Jet black	31	2
Infinix Hot 4-Gold	15	1
MotoZPlay -Black	1	1





# Python for Google Colab



No 2

Dear Data Analyst,

Menindaklanjuti meeting gabungan Tim Warehouse dan Tim Marketing, kami menemukan bahwa ketersediaan stock produk dengan Kategori Others pada akhir 2022 kemarin masih banyak.

Kami mohon bantuan untuk melakukan pengecekan data penjualan kategori tersebut dengan tahun 2021 secara kuantitas penjualan. Dugaan sementara kami, telah terjadi penurunan kuantitas penjualan pada 2022 dibandingkan 2021. (Mohon juga menampilkan data ke-15 kategori) Apabila memang terjadi penurunan kuantitas penjualan pada kategori Others, kami mohon bantuan untuk menyediakan data TOP 20 nama produk yang mengalami penurunan paling tinggi pada 2022 jika dibanding dengan 2021. Hal ini kami gunakan sebagai bahan diskusi pada meeting selanjutnya.

Mohon bantuan untuk mengirimkan data tersebut paling lambat 4 hari dari hari ini. Atas bantuan yang diberikan, kami mengucapkan terima kasih.  
Regards

Tim Warehouse

## SOLUSI NO 2

1. Memfilter data dengan valid = 1
2. Memfilter data pada transaksi selama 2021
3. Memfilter data pada transaksi selama 2022
4. Gunakan groupby berdasarkan sku\_name masing-masing tahun
5. Menggabungkan kedua data dengan merge
6. Lakukan pengurangan kolom qty 2022 dengan qty 2021



# Python for Google Colab



# @title `Jawaban no. 2.1`

# Menambahkan filter pada data frame

```
df_filter = df_project[(df_project['is_valid']==1)&\
    (df_project.category=="Mobiles & Tablets")&\
    (df_project.order_date>= "2021-01-01")&\
    (df_project.order_date<= "2021-12-31")]
```

# Menambahkan dataframe tahun 2021 & 2022

```
data_2021 = pd.DataFrame(df_project[(df_project['is_valid']==1)&\
    (df_project.order_date.dt.year==2021)]\
    .groupby('category').agg({"qty_ordered":"sum"}))
```

```
data_2022 = pd.DataFrame(df_project[(df_project['is_valid']==1)&\
    (df_project.order_date.dt.year==2022)]\
    .groupby('category').agg({"qty_ordered":"sum"}))
```

```
data_2021.head(10)
```



# Python for Google Colab



```
data_2021 = pd.DataFrame(df_project[(df_project['is_valid']==1)&(df_project['category']=="Others")&\
(df_project.order_date.dt.year==2021)].groupby('sku_name').agg({"qty_ordered":"sum"}))
```

```
data_2022 = pd.DataFrame(df_project[(df_project['is_valid']==1)&(df_project['category']=="Others")&\
(df_project.order_date.dt.year==2022)].groupby('sku_name').agg({"qty_ordered":"sum"}))
```

```
data_all_sku = data_2021.merge(data_2022, on="sku_name", how="left", suffixes=('_2021', '_2022')).reset_index()
data_all_sku
```

```
data_all_sku['selisih'] = data_all_sku['qty_ordered_2022']\
-data_all_sku['qty_ordered_2021']\
.head()
```

```
top20 = data_all_sku.sort_values('selisih').sort_values('selisih').head(20)
top20
```



# Python for Google Colab



## No 3

Dear Data Analyst,

Terkait ulang tahun perusahaan pada 2 bulan mendatang, Tim Digital Marketing akan memberikan informasi promo bagi pelanggan pada akhir bulan ini. Kriteria pelanggan yang akan kami butuhkan adalah mereka yang sudah melakukan check-out namun belum melakukan pembayaran (`is_gross = 1`) selama tahun 2022. Data yang kami butuhkan adalah ID Customer dan Registered Date.

Mohon bantuan, untuk mengirimkan data tersebut sebelum akhir bulan ini ke Tim Digital Marketing. Atas bantuan yang diberikan, kami mengucapkan terima kasih.

Regards

Tim Digital Marketing

Klik di sini untuk tips

1. Memfilter data dengan `gross = 1`
2. Memfilter data dengan `valid = 0`
3. Memfilter data dengan `net = 0`
4. Memfilter data pada transaksi selama 2022



# Python for Google Colab



```
# @title `Jawaban No 3.1`  
# df.info()  
df_project.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 5884 entries, 0 to 5883  
Data columns (total 22 columns):  
#   Column          Non-Null Count  Dtype  
---  -  
0   id               5884 non-null   object  
1   customer_id      5884 non-null   object  
2   order_date       5884 non-null   datetime64[ns]  
3   sku_id           5884 non-null   object  
4   price            5884 non-null   int64  
5   qty_ordered      5884 non-null   int64  
6   before_discount  5884 non-null   int64  
7   discount_amount  5884 non-null   int64  
8   after_discount   5884 non-null   int64  
9   is_gross          5884 non-null   int64  
10  is_valid          5884 non-null   int64  
11  is_net            5884 non-null   int64  
12  payment_id        5884 non-null   int64  
13  payment_method    5884 non-null   object  
14  sku_name          5884 non-null   object  
15  base_price        5884 non-null   int64  
16  cogs              5884 non-null   int64  
17  category          5884 non-null   object  
18  registered_date   5884 non-null   datetime64[ns]  
19  day               5884 non-null   object  
20  month             5884 non-null   object  
21  month_num         5884 non-null   int64  
dtypes: datetime64[ns](2), int64(12), object(8)  
memory usage: 1011.4+ KB
```

```
df_project =  
df_project[(df_project.is_gross==1)&(df_project.is_valid==0)&(df_proje  
ct.is_net==0)&(df_project.order_date.dt.year==2022)][['customer_id','re  
gistered_date']]
```

```
df_project.to_csv('https://raw.githubusercontent.com/dataskillsboost/Fi  
nalProjectDA11/main/order_detail.csv', index=False)
```

#Jalankan kode ini untuk mendownload file

```
from google.colab import files  
df_project.to_csv('audience_list.csv', encoding = 'utf-8-  
sig',index=False) #yang akan di download column pada  
tabel[customer_id dan registered_date]  
files.download('audience_list.csv')
```



# Python for Google Colab



## No 4

Dear Data Analyst,

Pada bulan October hingga Desember 2022, kami melakukan campaign setiap hari Sabtu dan Minggu. Kami hendak menilai, apakah campaign tersebut cukup berdampak pada kenaikan penjualan (before\_discount). Mohon bantuan untuk menampilkan data:

Rata-rata harian penjualan weekends (Sabtu dan Minggu) vs rata-rata harian penjualan weekdays (Senin-Jumat) per bulan tersebut.

Apakah ada peningkatan penjualan pada masing-masing bulan tersebut.

Rata-rata harian penjualan weekends (Sabtu dan Minggu) vs rata-rata harian penjualan weekdays (Senin-Jumat) keseluruhan 3 bulan tersebut.

Mohon bantuan untuk mengirimkan data tersebut paling lambat minggu depan. Atas bantuan yang diberikan, kami mengucapkan terima kasih.

Regards

Tim Campaign

Klik di sini untuk tips

1. Buatlah kolom baru untuk day, month, dan month number
2. Memfilter data dengan valid = 1
3. Memfilter data dengan day termasuk 'Saturday', 'Sunday' dengan menggunakan fungsi isin dan simpan dalam variable 'weekends'
4. Memfilter data dengan day termasuk 'Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday' dengan menggunakan fungsi isin dan simpan dengan variabel 'weekday'
5. Memfilter data pada transaksi selama 2022 bulan Oktober sampai dengan Desember
6. Gunakan groupby berdasarkan pada masing-masing variabel
7. Menggabungkan kedua data dengan merge
8. Tampilkan pada grafik batang untuk melihat perbedaannya



# Python for Google Colab



df\_project.head()

```
df_project.head()
```

	id	customer_id	order_date	sku_id	price	qty_ordered	before_discount	discount_amount	after_discount	is_gros
0	ODR9939707760w	C713589L	2021-11-19	P858068	26100	200	5220000	2610000	2610000	
1	ODR7448356649d	C551551L	2021-11-19	P886455	1971942	5	9859710	2464927	7394782	
2	ODR4011281866z	C685596L	2021-11-25	P678648	7482000	1	7482000	2065344	5416655	
3	ODR3378927994s	C830683L	2021-11-22	P540013	3593680	1	3593680	1455440	2138239	
4	ODR4904430099k	C191766L	2021-11-21	P491032	4413220	1	4413220	1059172	3354047	

5 rows × 24 columns



# Python for Google Colab



```
# @title Menambahkan Filter Tahun
```

```
# Menambahkan filter pada data frame
```

```
df_filter = df_project[(df_project['is_valid']==1)&\
    (df_project.category=="Mobiles & Tablets")&\
    (df_project.order_date>= "2021-01-01")&\
    (df_project.order_date<= "2021-12-31")]
```

```
# @title `Jawaban 4.1`
```

```
df_project['day_name'] = df_project['order_date'].dt.day_name()
```

```
df_project['is_weekend'] = ["weekend" if x in ["Saturday", "Sunday"] else "weekdays" for x in df_project['day_name']]
```

```
df_project["is_weekend"]
```

```
0    weekdays
```

```
1    weekdays
```

```
2    weekdays
```

```
3    weekdays
```

```
4     weekend
```

```
...
```

```
5879  weekdays
```

```
5880  weekdays
```

```
5881  weekdays
```

```
5882  weekdays
```

```
5883   weekend
```

```
Name: is_weekend, Length: 5884, dtype: object
```





# Python for Google Colab



# Tulis kode Anda di bawah ini. Dapat menggunakan lebih dari 1 blok kode

```
df_project[(df_project.order_date.dt.month>=10)&\
(df_project.order_date.dt.month<=12)&\
(df_project.order_date.dt.year==2022)].groupby('is_weekend').agg({"before_discount":"sum"})
```

before\_discount

is\_weekend

weekdays 275986968

weekend 103536728

This is just learning material for all of us

Thank you for studying here...



# Follow me!

Instagram :harisariefkamis

LinkedIn :harisariefkamis

Bootcamp Data Analysis  
by @myskill.id

