

RETRIEVE SQL DATA DAN EKSPLORASI		
Praktikan	Aslab	
Nama: xxxx	Annur Hangga Prihadi	065001800028
Nim: xxxx	Faiz Kumara	065001800003

## PRAKTIKUM 1

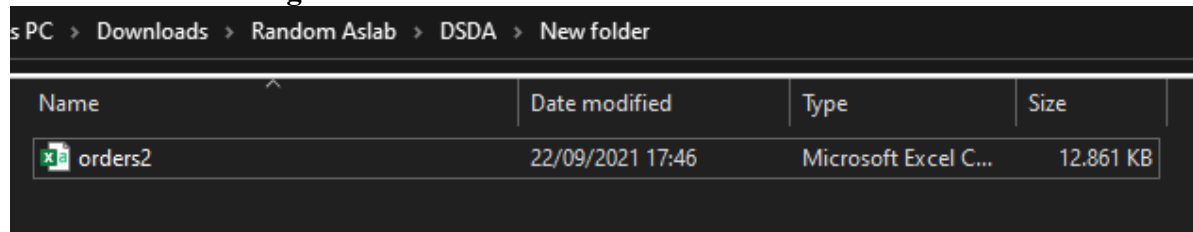
### DATA SAINS DAN ANALITIK


Pertemuan praktikum pertama kali ini adalah memasang dan menyesuaikan perangkat lunak maupun library yang dibutuhkan saat praktikum, mengetahui cara untuk mengambil data dari RDBMS, dan eksplorasi data sederhana dari data yang telah diambil oleh praktikan.

## Latihan 1

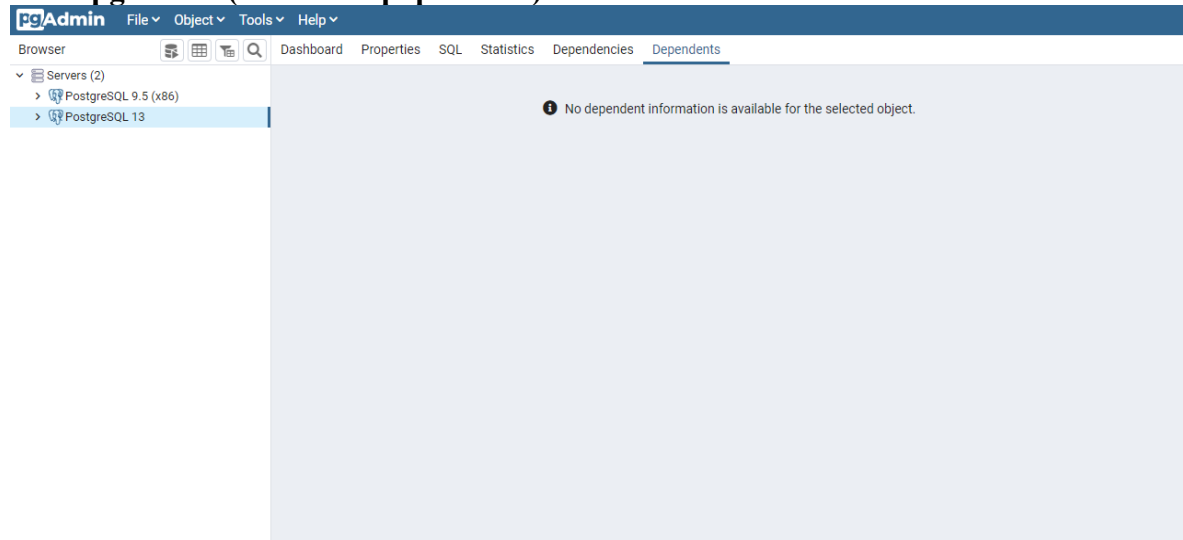
### Pre-praktikum

Pastikan sudah mengunduh file orders.csv

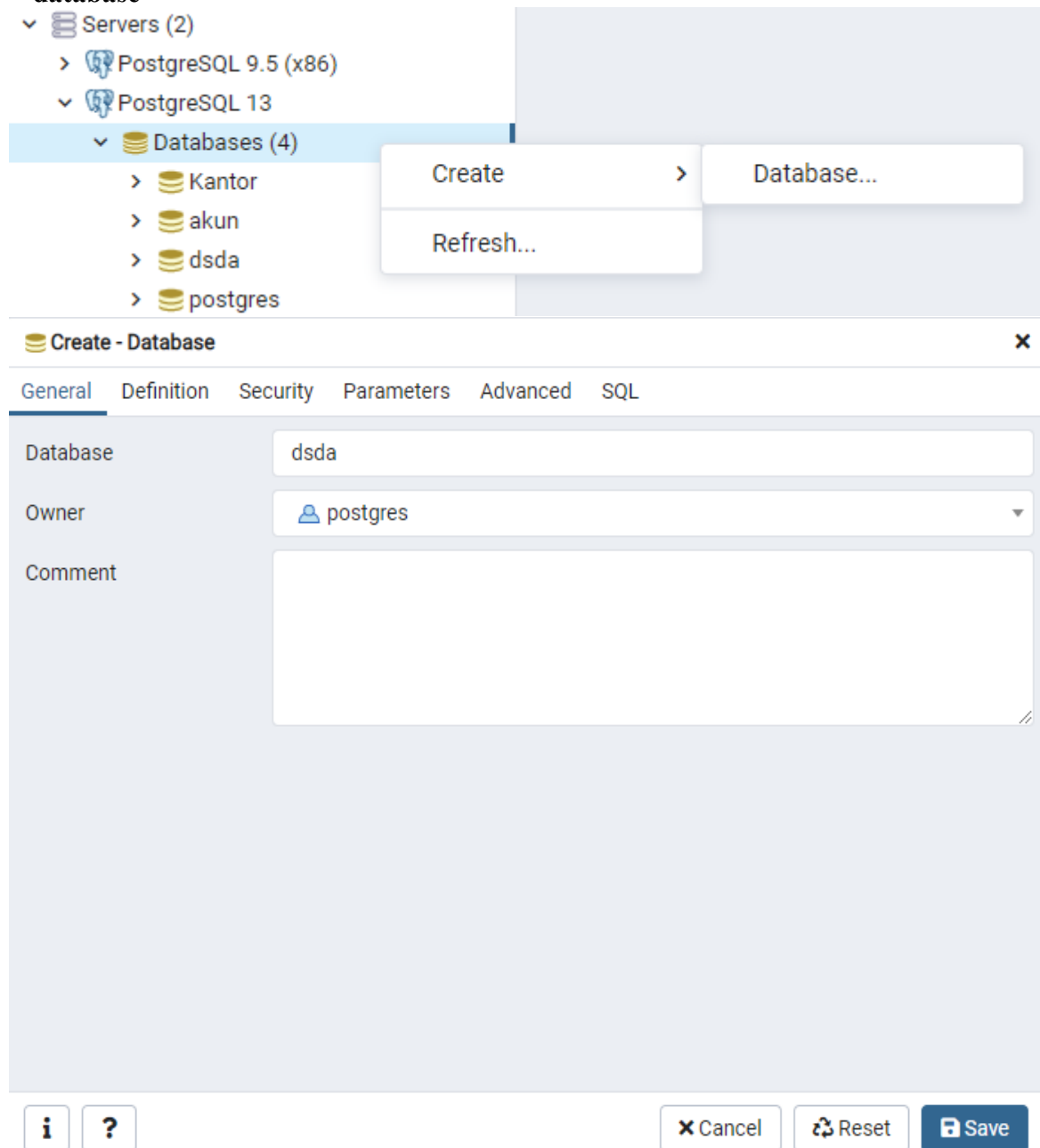


s PC > Downloads > Random Aslab > DSDA > New folder			
Name	Date modified	Type	Size
 orders2	22/09/2021 17:46	Microsoft Excel C...	12.861 KB

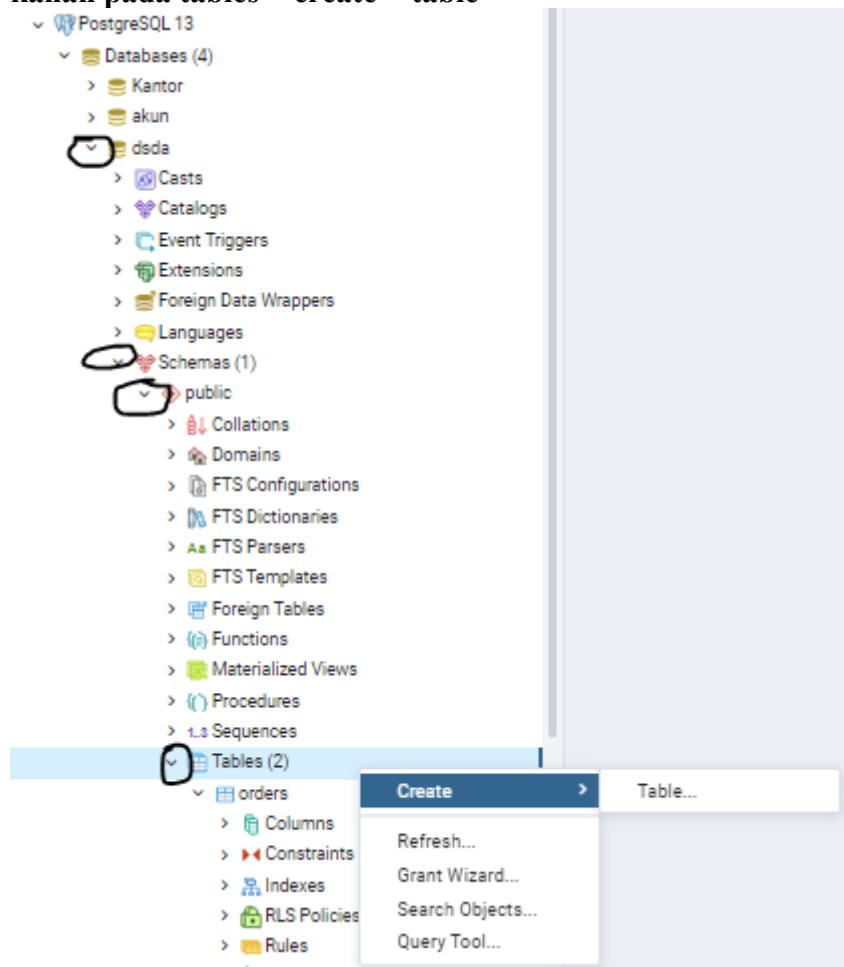
Buka pgAdmin (Versi berapapun bisa)



**Buat database dengan nama dsda dengan cara klik kanan pada icon database > create > database**



**Buat tabel dengan cara memperluas database dsda > schemas > public > tables > klik kanan pada tables > create > table**



**Beri nama table bebas terserah kalian, lalu klik columns**

A screenshot of the 'Create - Table' dialog box. The 'General' tab is selected. The 'Name' field contains 'order2'. The 'Owner' dropdown shows 'postgres'. The 'Schema' dropdown shows 'public'. The 'Tablespace' dropdown shows 'Select an item...'. The 'Partitioned table?' checkbox is unchecked, with 'No' displayed. There is a large text area for 'Comment'. At the bottom, there are buttons for 'Cancel', 'Reset', and 'Save', along with information and help icons.

## Tambah kolom dan tipe data sesuai dengan data order

A	B	C	D	E	F	G	H	I	J	K
orderid	customerid	campaignid	orderdate	city	state	zipcode	paymenttype	totalprice	numorderlines	numunits

Create - Table

General

Columns

Advanced

Constraints

Partitions

Parameters

Security

SQL

Inherited from table(s)

Columns

+

	Name	Data type	Length/Precision	Scale	Not NULL?	Primary key?
<input checked="" type="checkbox"/>	orderid	character varying			<input type="checkbox"/> No	<input type="checkbox"/> No
<input checked="" type="checkbox"/>	customerid	character varying			<input type="checkbox"/> No	<input type="checkbox"/> No
<input checked="" type="checkbox"/>	campaignid	character varying			<input type="checkbox"/> No	<input type="checkbox"/> No
<input checked="" type="checkbox"/>	orderdate	timestamp without time zone			<input type="checkbox"/> No	<input type="checkbox"/> No
<input checked="" type="checkbox"/>	city	character varying			<input type="checkbox"/> No	<input type="checkbox"/> No
<input checked="" type="checkbox"/>	state	character varying			<input type="checkbox"/> No	<input type="checkbox"/> No
<input checked="" type="checkbox"/>	zipcode	character varying			<input type="checkbox"/> No	<input type="checkbox"/> No
<input checked="" type="checkbox"/>	paymenttype	character varying			<input type="checkbox"/> No	<input type="checkbox"/> No
<input checked="" type="checkbox"/>	totalprice	double precision			<input type="checkbox"/> No	<input type="checkbox"/> No
<input checked="" type="checkbox"/>	numorderlines	character varying			<input type="checkbox"/> No	<input type="checkbox"/> No
<input checked="" type="checkbox"/>	numunits	character varying			<input type="checkbox"/> No	<input type="checkbox"/> No

i

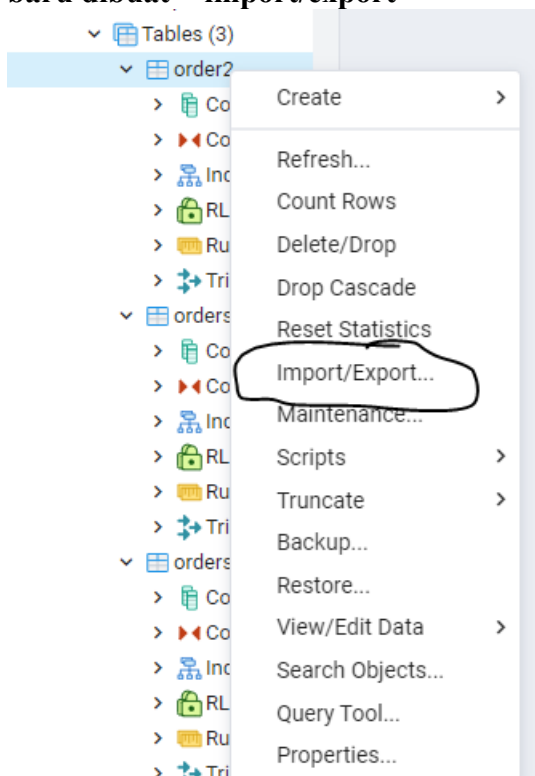
?

Cancel

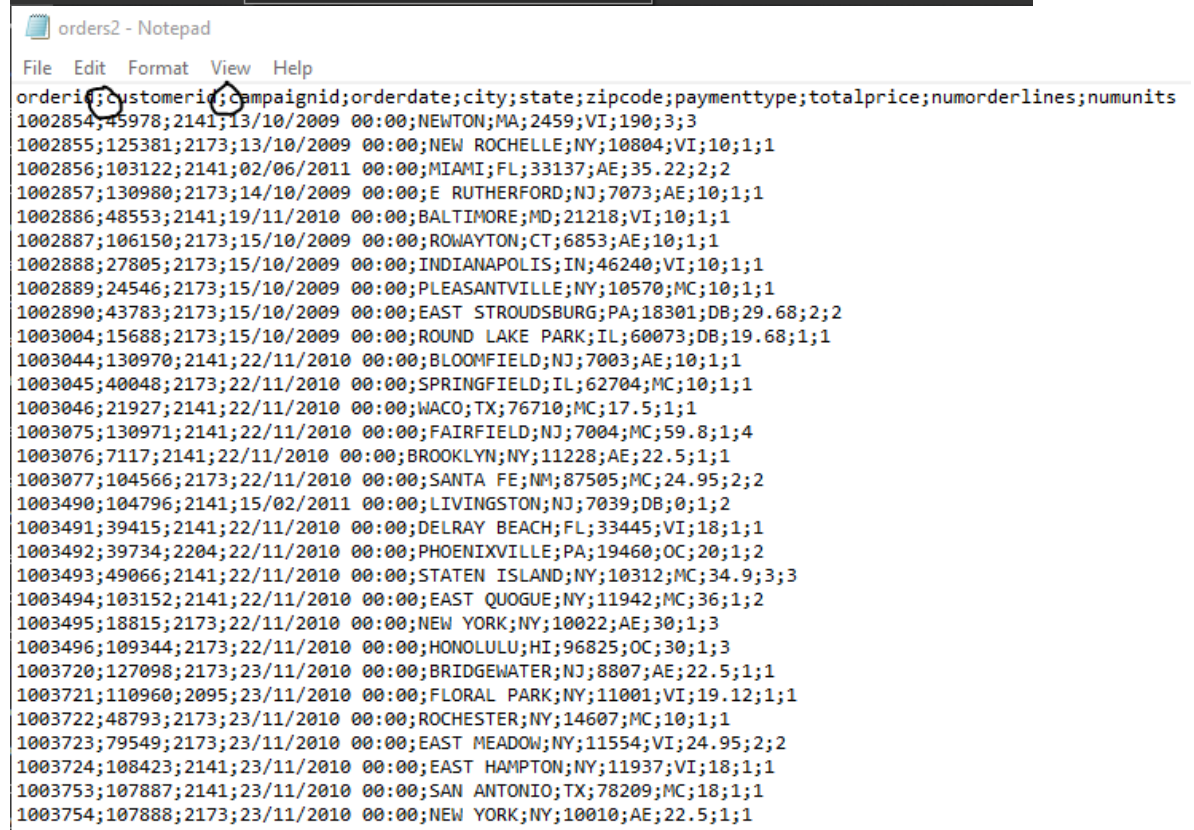
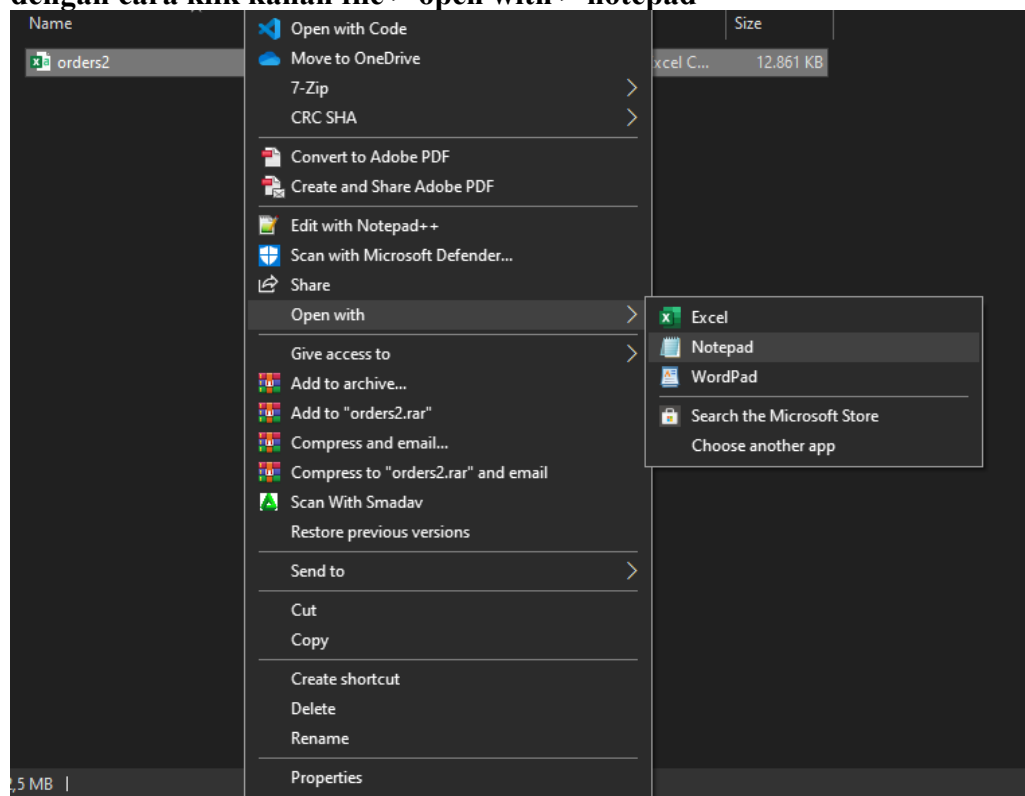
Reset

Save

Import data csv yang telah diunduh di-step pertama dengan cara klik kanan tabel yang baru dibuat > import/export



Sebelum import pastikan dahulu delimiter/pembatas dari data csv yang diunduh dengan cara klik kanan file > open with > notepad



Terlihat di dalam file orders milik aslab terdapat delimiter berupa tanda ; (Masing-masing praktikan kemungkinan mendapatkan delimiter yang berbeda, jadi tolong lebih teliti)

Setelah mengetahui delimiter atau pembatas lanjut ke tahap import file, lakukan langkah berikut

1. Klik export untuk berubah ke import
2. Pilih directory file penyimpanan file csv yang kalian unduh di-step awal
3. Pilih yes pada header
4. Sesuaikan delimiter dengan yang kalian dapat di-step sebelumnya
5. Jika mengikuti step akan muncul notifikasi sukses

Import/Export data - table 'order2'

Options Columns

Import/Export 1 **Import**

**File Info**

Filename 2 C:\Users\hangg\Downloads\Random Aslab\DSDA\New folder\orders2.csv

Format csv

Encoding Select an item...

**Miscellaneous**

OID No

Header 3 **Yes**

Delimiter 4 ;

Quote "

Escape '.

Cancel OK

Import - Copying table data

Copying table data 'public.order2' on database 'dsda' and server (localhost:5432)

Wed Sep 22 2021 18:10:42 GMT+0700 (Western Indonesia Time)

1.37 seconds

More details... Stop Process

Successfully completed.

## Memasang library yang dibutuhkan

```
import sys
!{sys.executable} -m pip install psycopg2
!{sys.executable} -m pip install pandas
import psycopg2
import pandas as pd
```

```
In [1]: import sys
!{sys.executable} -m pip install psycopg2
!{sys.executable} -m pip install pandas
import psycopg2
import pandas as pd

Requirement already satisfied: psycopg2 in c:\users\hangg\anaconda3\lib\site-packages (2.9.1)
Requirement already satisfied: pandas in c:\users\hangg\anaconda3\lib\site-packages (1.2.4)
Requirement already satisfied: pytz>=2017.3 in c:\users\hangg\anaconda3\lib\site-packages (from pandas) (2021.1)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\hangg\anaconda3\lib\site-packages (from pandas) (2.8.1)
Requirement already satisfied: numpy>=1.16.5 in c:\users\hangg\anaconda3\lib\site-packages (from pandas) (1.20.1)
Requirement already satisfied: six>=1.5 in c:\users\hangg\anaconda3\lib\site-packages (from python-dateutil>=2.7.3->pandas) (1.15.0)
```

## Menyambungkan koneksi worksheet jupyter dengan PostgreSQL

```
conn = psycopg2.connect(host="localhost", port = xxxx, database="dsda", user="postgres", password="xxxx")
conn.set_session(autocommit=True)
cur = conn.cursor()
sql = "SELECT * FROM public.nama_tabel"
cur.execute(sql)
baris = cur.fetchall()
```

--Sesuaikan port, password, dan nama\_tabel dengan milik kalian--

```
In [2]: conn = psycopg2.connect(host="localhost", port = 5432, database="dsda", user="postgres", password="hanggaa")
conn.set_session(autocommit=True)
```

```
In [3]: cur = conn.cursor()
```

```
In [5]: sql = "SELECT * FROM public.order2"
cur.execute(sql)
baris = cur.fetchall()
```

## Menampilkan n data pertama

```
order = pd.DataFrame([[ij for ij in i]for i in baris])
order.rename(columns={0:'order_id',1:'customer_id',2:'campaign_id',3:'order_date',4:'city',5:'state',6:'zipcode',
7:'payment_type',8:'total_price',9:'num_order',10:'num_units'},inplace=True)
order.head(4)
```

```
In [5]: order = pd.DataFrame([[ij for ij in i]for i in baris])
order.rename(columns={0:'order_id',1:'customer_id',2:'campaign_id',3:'order_date',4:'city',5:'state',6:'zipcode',7:'payment_type'
order.head(4)
```

Out[5]:

	order_id	customer_id	campaign_id	order_date	city	state	zipcode	payment_type	total_price	num_order	num_units
0	1002854	45978	2141	2009-10-13	NEWTON	MA	2459	VI	190.00	3	3
1	1002855	125381	2173	2009-10-13	NEW ROCHELLE	NY	10804	VI	10.00	1	1
2	1002856	103122	2141	2011-06-02	MIAMI	FL	33137	AE	35.22	2	2
3	1002857	130980	2173	2009-10-14	E RUTHERFORD	NJ	7073	AE	10.00	1	1

## Menampilkan rangkuman statistik dari data integer dan dimensi tabel

```
order.describe()
order.shape
```

```
In [6]: order.describe()
```

Out[6]:

	total_price
count	192983.000000
mean	71.036808
std	183.510368
min	0.000000
25%	16.000000
50%	25.950000
75%	46.900000
max	9848.960000

```
In [7]: order.shape
```

Out[7]: (192983, 11)

**Menampilkan hasil dari rata-rata dan simpangan baku 1 kolom maupun berdasarkan kelompok**

```
order['total_price'].mean()
order.groupby('order_date')['total_price'].mean()
```

```
In [8]: order['total_price'].mean()
Out[8]: 71.03680790531962

In [9]: order.groupby('order_date')['total_price'].mean()
Out[9]: order_date
2009-10-04    18.489282
2009-10-05    16.907114
2009-10-06    31.735349
2009-10-07    22.062759
2009-10-08    33.134118
...
2016-09-16    54.217333
2016-09-17    34.912500
2016-09-18    47.739091
2016-09-19    46.532273
2016-09-20    47.200000
Name: total_price, Length: 2541, dtype: float64
```

```
order['total_price'].std()
order.groupby('order_date')['total_price'].std()
```

```
In [11]: order['total_price'].std()
Out[11]: 183.5103681036432

In [13]: order.groupby('order_date')['total_price'].std()
Out[13]: order_date
2009-10-04    20.436475
2009-10-05    10.895262
2009-10-06    71.229757
2009-10-07    30.588732
2009-10-08    74.703053
...
2016-09-16    50.183828
2016-09-17    41.822191
2016-09-18    41.695064
2016-09-19    53.153625
2016-09-20    3.181981
Name: total_price, Length: 2541, dtype: float64
```



## Latihan 2

1. Gunakan data orders untuk mengetahui pengaruh data kolom lain terhadap total harga dari pesanan.
  - a. Berapakah nilai rata-rata dari total harga untuk tiap kategori campaign\_id, city, payment\_tipe dan Jelaskan makna dari keterkaitan tiap kategori tersebut terhadap variabel total harga!
  - b. Berapakah nilai standar deviasi dari total harga untuk tiap kategori campaign\_id, city, payment\_tipe dan Jelaskan makna dari keterkaitan tiap kategori tersebut terhadap variabel total harga!

### Lampiran Screenshot hasil a dan b

[Input screenshot disini](#)

### Makna dari hasil nilai rata-rata dan standar deviasi

[Ketik makna disini](#)