

```
In [1]: import sys
```

```
In [2]: !{sys.executable} -m pip install pymysql
```

Requirement already satisfied: pymysql in c:\programdata\anaconda3\lib\site-packages (1.1.0)

```
In [4]: import pymysql as pm
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [6]: connection=pm.connect(host='localhost',user='root',password='',database='supermarket')
cursor=connection.cursor()
desc1=pd.read_sql('desc customer_details;',con=connection)
print(desc1)
```

	Field	Type	Null	Key	Default	Extra
0	ROW_ID	int(10)	NO	PRI	None	
1	CUSTOMER-ID	varchar(20)	NO		None	
2	CUSTOMER-NAME	varchar(20)	NO		None	
3	SEGMENT	char(10)	NO		None	
4	MARKET	char(10)	NO		None	
5	REGION	char(10)	NO		None	
6	SALES	varchar(10)	NO		None	
7	QUANTITY	int(10)	NO		None	
8	DISCOUNT	varchar(10)	NO		None	
9	PROFIT	varchar(10)	NO		None	
10	ADDRESS	varchar(50)	NO		None	

C:\Windows\Temp\ipykernel_10660\1410343410.py:3: UserWarning: pandas only supports SQLAlchemy connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connection. Other DBAPI2 objects are not tested. Please consider using SQLAlchemy.

```
desc1=pd.read_sql('desc customer_details;',con=connection)
```

```
In [7]: connection=pm.connect(host='localhost',user='root',password='',database='super:
cursor=connection.cursor()
desc2=pd.read_sql('desc order_details;',con=connection)
print(desc2)
```

	Field	Type	Null	Key	Default	Extra
0	ROW_ID	int(20)	NO	PRI	None	
1	ORDER_ID	varchar(20)	NO		None	
2	ORDER_DATE	varchar(20)	NO		None	
3	SHIP-DATE	varchar(15)	NO		None	
4	SHIP_MODE	varchar(22)	NO		None	
5	PRODUCT_ID	varchar(30)	NO		None	
6	CATEGORY	varchar(20)	NO		None	
7	SUB-CATEGORY	varchar(20)	NO		None	
8	PRODUCT-NAME	varchar(50)	NO		None	
9	SHIPPING-COST	varchar(10)	NO		None	
10	ORDER-PRIORITY	varchar(10)	NO		None	

C:\Windows\Temp\ipykernel_10660\2400787648.py:3: UserWarning: pandas only supports SQLAlchemy connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connection. Other DBAPI2 objects are not tested. Please consider using SQLAlchemy.

```
desc2=pd.read_sql('desc order_details;',con=connection)
```

```
In [10]: connection=pm.connect(host='localhost',user='root',password='',database='super:
cursor=connection.cursor()
show1=pd.read_sql('select * from order_details;',con=connection)
print(show1)
```

C:\Windows\Temp\ipykernel_10660\4030484799.py:3: UserWarning: pandas only supports SQLAlchemy connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connection. Other DBAPI2 objects are not tested. Please consider using SQLAlchemy.

```
show1=pd.read_sql('select * from order_details;',con=connection)
```

	ROW_ID	ORDER_ID	ORDER_DATE	SHIP-DATE	SHIP_MODE \
0	1	CA-2012-124891	31-07-2012	31-07-2012	Same Day
1	2	IN-2013-77878	5/2/2013	7/2/2013	Second Class
2	3	IN-2013-71249	17-10-2013	18-10-2013	First Class
3	4	ES-2013-1579342	28-01-2013	30-01-2013	First Class
4	5	SG-2013-4320	5/11/2013	6/11/2013	Same Day
...
51285	51286	IN-2014-62366	19-06-2014	19-06-2014	Same Day
51286	51287	US-2014-102288	20-06-2014	24-06-2014	Standard Class
51287	51288	US-2013-155768	2/12/2013	2/12/2013	Same Day
51288	51289	MX-2012-140767	18-02-2012	22-02-2012	Standard Class
51289	51290	MX-2012-134460	22-05-2012	26-05-2012	Second Class

	PRODUCT_ID	CATEGORY	SUB-CATEGORY \
0	TEC-AC-10003033	Technology	Accessories
1	FUR-CH-10003950	Furniture	Chairs
2	TEC-PH-10004664	Technology	Phones
3	TEC-PH-10004583	Technology	Phones
4	TEC-SHA-10000501	Technology	Copiers
...
51285	OFF-FA-10000746	Office Supplies	Fasteners
51286	OFF-AP-10002906	Office Supplies	Appliances
51287	OFF-EN-10001219	Office Supplies	Envelopes
51288	OFF-BI-10000806	Office Supplies	Binders
51289	OFF-PA-10004155	Office Supplies	Paper

	PRODUCT-NAME	SHIPPING-COST \
0	Plantronics CS510 - Over-the-Head monaural Wir...	933.57
1	Novimex Executive Leather Armchair, Black	923.63
2	Nokia Smart Phone, with Caller ID	915.49
3	Motorola Smart Phone, Cordless	910.16
4	Sharp Wireless Fax, High-Speed	903.04
...
51285	Advantus Thumb Tacks, 12 Pack	0.01
51286	Hoover Replacement Belt for Commercial Guardsm...	0.01
51287	#10- 4 1/8" x 9 1/2" Security-Tint Envelopes	0.01
51288	Acco Index Tab, Economy	0
51289	Eaton Computer Printout Paper, 8.5 x 11	0

	ORDER-PRIORITY
0	Critical
1	Critical
2	Medium
3	Medium
4	Critical
...	...
51285	Medium
51286	Medium
51287	High
51288	Medium
51289	High

[51290 rows x 11 columns]

```
In [11]: connection=pm.connect(host='localhost',user='root',password='',database='supermarket')
cursor=connection.cursor()
show2=pd.read_sql('select * from customer_details;',con=connection)
print(show2)
```

C:\Windows\Temp\ipykernel_10660\1679967303.py:3: UserWarning: pandas only supports SQLAlchemy connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connection. Other DBAPI2 objects are not tested. Please consider using SQLAlchemy.

```
show2=pd.read_sql('select * from customer_details;',con=connection)
```

	ROW_ID	CUSTOMER-ID	CUSTOMER-NAME	SEGMENT	MARKET	REGION
\						
0	1	RH-19495	Rick Hansen	Consumer	US	East
1	2	JR-16210	Justin Ritter	Corporate	APAC	Oceania
2	3	CR-12730	Craig Reiter	Consumer	APAC	Oceania
3	4	KM-16375	Katherine Murray	Home Offic	EU	Central
4	5	RH-9495	Rick Hansen	Consumer	Africa	Africa
...
51285	51286	KE-16420	Katrina Edelman	Corporate	APAC	North Asia
51286	51287	ZC-21910	Zuschuss Carroll	Consumer	US	Central
51287	51288	LB-16795	Laurel Beltran	Home Offic	US	West
51288	51289	RB-19795	Ross Baird	Home Offic	LATAM	South
51289	51290	MC-18100	Mick Crebagga	Consumer	LATAM	Central

	SALES	QUANTITY	DISCOUNT	PROFIT	\
0	2309.65	7	0	762.1845	
1	3709.395	9	0.1	-288.765	
2	5175.171	9	0.1	919.971	
3	2892.51	5	0.1	-96.54	
4	2832.96	8	0	311.52	
...	
51285	65.1	5	0	4.5	
51286	0.444	1	0.8	-1.11	
51287	22.92	3	0	11.2308	
51288	13.44	2	0	2.4	
51289	61.38	3	0	1.8	

	ADDRESS
0	New York City New York United States -10024
1	Wollongong New South Wales Australia -
2	Brisbane Queensland Australia -
3	Berlin Berlin Germany -
4	Dakar Dakar Senegal -
...	...
51285	Kure Hiroshima Japan -
51286	Houston Texas United States -77095
51287	Oxnard California United States -93030
51288	Valinhos São Paulo Brazil -
51289	Tipitapa Managua Nicaragua -

[51290 rows x 11 columns]

In [12]: show1

Out[12]:

	ROW_ID	ORDER_ID	ORDER_DATE	SHIP- DATE	SHIP_MODE	PRODUCT_ID	CATEGORY	C/
0	1	CA-2012-124891	31-07-2012	31-07-2012	Same Day	TEC-AC-10003033	Technology	Ac
1	2	IN-2013-77878	5/2/2013	7/2/2013	Second Class	FUR-CH-10003950	Furniture	
2	3	IN-2013-71249	17-10-2013	18-10-2013	First Class	TEC-PH-10004664	Technology	
3	4	ES-2013-1579342	28-01-2013	30-01-2013	First Class	TEC-PH-10004583	Technology	
4	5	SG-2013-4320	5/11/2013	6/11/2013	Same Day	TEC-SHA-10000501	Technology	
...
51285	51286	IN-2014-62366	19-06-2014	19-06-2014	Same Day	OFF-FA-10000746	Office Supplies	
51286	51287	US-2014-102288	20-06-2014	24-06-2014	Standard Class	OFF-AP-10002906	Office Supplies	A
51287	51288	US-2013-155768	2/12/2013	2/12/2013	Same Day	OFF-EN-10001219	Office Supplies	I
51288	51289	MX-2012-140767	18-02-2012	22-02-2012	Standard Class	OFF-BI-10000806	Office Supplies	
51289	51290	MX-2012-134460	22-05-2012	26-05-2012	Second Class	OFF-PA-10004155	Office Supplies	

51290 rows × 11 columns



In [13]: show2

Out[13]:

	ROW_ID	CUSTOMER-ID	CUSTOMER-NAME	SEGMENT	MARKET	REGION	SALES	QUANTITY
0	1	RH-19495	Rick Hansen	Consumer	US	East	2309.65	7
1	2	JR-16210	Justin Ritter	Corporate	APAC	Oceania	3709.395	9
2	3	CR-12730	Craig Reiter	Consumer	APAC	Oceania	5175.171	9
3	4	KM-16375	Katherine Murray	Home Offic	EU	Central	2892.51	5
4	5	RH-9495	Rick Hansen	Consumer	Africa	Africa	2832.96	8
...
51285	51286	KE-16420	Katrina Edelman	Corporate	APAC	North Asia	65.1	5
51286	51287	ZC-21910	Zuschuss Carroll	Consumer	US	Central	0.444	1
51287	51288	LB-16795	Laurel Beltran	Home Offic	US	West	22.92	3
51288	51289	RB-19795	Ross Baird	Home Offic	LATAM	South	13.44	2
51289	51290	MC-18100	Mick Crebagga	Consumer	LATAM	Central	61.38	3

51290 rows × 11 columns



In [18]: x=show1['ORDER_DATE']
y=show1['SHIPPING-COST']

In [16]: x

```
Out[16]: 0      31-07-2012
          1      5/2/2013
          2     17-10-2013
          3     28-01-2013
          4      5/11/2013
          ...
          51285   19-06-2014
          51286   20-06-2014
          51287    2/12/2013
          51288   18-02-2012
          51289   22-05-2012
          Name: ORDER_DATE, Length: 51290, dtype: object
```

In [19]: y

```
Out[19]: 0      933.57
          1      923.63
          2      915.49
          3      910.16
          4      903.04
          ...
          51285    0.01
          51286    0.01
          51287    0.01
          51288     0
          51289     0
          Name: SHIPPING-COST, Length: 51290, dtype: object
```

In [31]: show1.count()

```
Out[31]: ROW_ID      51290
          ORDER_ID    51290
          ORDER_DATE  51290
          SHIP-DATE   51290
          SHIP_MODE    51290
          PRODUCT_ID  51290
          CATEORY      51290
          SUB-CATEGORY 51290
          PRODUCT-NAME 51290
          SHIPPING-COST 51290
          ORDER-PRIORITY 51290
          dtype: int64
```



```
In [32]: show2.count()
```

```
Out[32]: ROW_ID          51290  
CUSTOMER-ID          51290  
CUSTOMER-NAME        51290  
SEGMENT              51290  
MARKET               51290  
REGION              51290  
SALES                51290  
QUANTITY             51290  
DISCOUNT            51290  
PROFIT               51290  
ADDRESS              51290  
dtype: int64
```

```
In [36]: show1.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 51290 entries, 0 to 51289  
Data columns (total 11 columns):  
#   Column                Non-Null Count  Dtype  
---  ---  
0   ROW_ID                51290 non-null  int64  
1   ORDER_ID              51290 non-null  object  
2   ORDER_DATE            51290 non-null  object  
3   SHIP-DATE             51290 non-null  object  
4   SHIP_MODE              51290 non-null  object  
5   PRODUCT_ID            51290 non-null  object  
6   CATEGORY              51290 non-null  object  
7   SUB-CATEGORY          51290 non-null  object  
8   PRODUCT-NAME          51290 non-null  object  
9   SHIPPING-COST         51290 non-null  object  
10  ORDER-PRIORITY        51290 non-null  object  
dtypes: int64(1), object(10)  
memory usage: 4.3+ MB
```

In [37]: show2.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51290 entries, 0 to 51289
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  -
0   ROW_ID          51290 non-null  int64
1   CUSTOMER-ID     51290 non-null  object
2   CUSTOMER-NAME   51290 non-null  object
3   SEGMENT         51290 non-null  object
4   MARKET          51290 non-null  object
5   REGION          51290 non-null  object
6   SALES           51290 non-null  object
7   QUANTITY        51290 non-null  int64
8   DISCOUNT       51290 non-null  object
9   PROFIT          51290 non-null  object
10  ADDRESS         51290 non-null  object
dtypes: int64(2), object(9)
memory usage: 4.3+ MB
```

In [39]: show1.shape

Out[39]: (51290, 11)

In [40]: show2.shape

Out[40]: (51290, 11)

In [41]: show1.describe()

Out[41]:

	ROW_ID
count	51290.00000
mean	25645.50000
std	14806.29199
min	1.00000
25%	12823.25000
50%	25645.50000
75%	38467.75000
max	51290.00000

In [42]: show2.describe()

Out[42]:

	ROW_ID	QUANTITY
count	51290.00000	51290.000000
mean	25645.50000	3.476545
std	14806.29199	2.278766
min	1.00000	1.000000
25%	12823.25000	2.000000
50%	25645.50000	3.000000
75%	38467.75000	5.000000
max	51290.00000	14.000000

In [43]: show1.isnull()

Out[43]:

	ROW_ID	ORDER_ID	ORDER_DATE	SHIP- DATE	SHIP_MODE	PRODUCT_ID	CATEGORY	CATEG
0	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	
...	
51285	False	False	False	False	False	False	False	
51286	False	False	False	False	False	False	False	
51287	False	False	False	False	False	False	False	
51288	False	False	False	False	False	False	False	
51289	False	False	False	False	False	False	False	

51290 rows × 11 columns



```
In [44]: show2.isnull()
```

```
Out[44]:
```

	ROW_ID	CUSTOMER-ID	CUSTOMER-NAME	SEGMENT	MARKET	REGION	SALES	QUANTITY	I
0	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	
...	
51285	False	False	False	False	False	False	False	False	
51286	False	False	False	False	False	False	False	False	
51287	False	False	False	False	False	False	False	False	
51288	False	False	False	False	False	False	False	False	
51289	False	False	False	False	False	False	False	False	

51290 rows × 11 columns



```
In [45]: show1.isnull().sum()
```

```
Out[45]: ROW_ID          0
ORDER_ID          0
ORDER_DATE        0
SHIP-DATE         0
SHIP_MODE         0
PRODUCT_ID        0
CATEGORY          0
SUB-CATEGORY      0
PRODUCT-NAME      0
SHIPPING-COST     0
ORDER-PRIORITY    0
dtype: int64
```

```
In [46]: show2.isnull().sum()
```

```
Out[46]: ROW_ID          0
CUSTOMER-ID        0
CUSTOMER-NAME      0
SEGMENT            0
MARKET             0
REGION             0
SALES              0
QUANTITY           0
DISCOUNT          0
PROFIT             0
ADDRESS            0
dtype: int64
```

In [51]: show1.tail()

Out[51]:


	ROW_ID	ORDER_ID	ORDER_DATE	SHIP- DATE	SHIP_MODE	PRODUCT_ID	CATEGORY	CA
51285	51286	IN-2014- 62366	19-06-2014	19-06- 2014	Same Day	OFF-FA- 10000746	Office Supplies	f
51286	51287	US-2014- 102288	20-06-2014	24-06- 2014	Standard Class	OFF-AP- 10002906	Office Supplies	A
51287	51288	US-2013- 155768	2/12/2013	2/12/2013	Same Day	OFF-EN- 10001219	Office Supplies	E
51288	51289	MX-2012- 140767	18-02-2012	22-02- 2012	Standard Class	OFF-BI- 10000806	Office Supplies	
51289	51290	MX-2012- 134460	22-05-2012	26-05- 2012	Second Class	OFF-PA- 10004155	Office Supplies	



In [52]: show1.head()

Out[52]:

	ROW_ID	ORDER_ID	ORDER_DATE	SHIP- DATE	SHIP_MODE	PRODUCT_ID	CATEGORY	CATEG
0	1	CA-2012-124891	31-07-2012	31-07-2012	Same Day	TEC-AC-10003033	Technology	Access
1	2	IN-2013-77878	5/2/2013	7/2/2013	Second Class	FUR-CH-10003950	Furniture	C
2	3	IN-2013-71249	17-10-2013	18-10-2013	First Class	TEC-PH-10004664	Technology	Ph
3	4	ES-2013-1579342	28-01-2013	30-01-2013	First Class	TEC-PH-10004583	Technology	Ph
4	5	SG-2013-4320	5/11/2013	6/11/2013	Same Day	TEC-SHA-10000501	Technology	Co



In [53]: show2.head()

Out[53]:

	ROW_ID	CUSTOMER-ID	CUSTOMER-NAME	SEGMENT	MARKET	REGION	SALES	QUANTITY	DIS
0	1	RH-19495	Rick Hansen	Consumer	US	East	2309.65	7	
1	2	JR-16210	Justin Ritter	Corporate	APAC	Oceania	3709.395	9	
2	3	CR-12730	Craig Reiter	Consumer	APAC	Oceania	5175.171	9	
3	4	KM-16375	Katherine Murray	Home Offic	EU	Central	2892.51	5	
4	5	RH-9495	Rick Hansen	Consumer	Africa	Africa	2832.96	8	



In [54]: show2.tail()

Out[54]:

	ROW_ID	CUSTOMER-ID	CUSTOMER-NAME	SEGMENT	MARKET	REGION	SALES	QUANTITY	I
51285	51286	KE-16420	Katrina Edelman	Corporate	APAC	North Asia	65.1	5	
51286	51287	ZC-21910	Zuschuss Carroll	Consumer	US	Central	0.444	1	
51287	51288	LB-16795	Laurel Beltran	Home Office	US	West	22.92	3	
51288	51289	RB-19795	Ross Baird	Home Office	LATAM	South	13.44	2	
51289	51290	MC-18100	Mick Crebagger	Consumer	LATAM	Central	61.38	3	

In [55]: len(show1)

Out[55]: 51290

In [56]: len(show2)

Out[56]: 51290

In [58]: show1.loc(1)

Out[58]: <pandas.core.indexing._LocIndexer at 0x115858fe800>

In [59]: show1.sample()


Out[59]:

	ROW_ID	ORDER_ID	ORDER_DATE	SHIP-DATE	SHIP_MODE	PRODUCT_ID	CATEGORY	CATEG
49669	49670	US-2014-163251	12/5/2014	16-05-2014	Standard Class	OFF-PA-10000032	Office Supplies	f

```
In [60]: show2.sample()
```

Out[60]:

	ROW_ID	CUSTOMER-ID	CUSTOMER-NAME	SEGMENT	MARKET	REGION	SALES	QUANTITY	I
33829	33830	DP-13165	David Philippe	Consumer	EU	Central	48.63	1	



```
In [ ]:
```