```
!pip install kaggle

import kaggle
!kaggle datasets download ankitbansal06/retail-orders -f orders.csv

Requirement already satisfied: kaggle in c:\users\hp\anaconda3\lib\
site-packages (1.6.14)
Requirement already satisfied: urllib3 in c:\users\hp\anaconda3\lib\
site-packages (from kaggle) (1.26.14)
Requirement already satisfied: python-dateutil in c:\users\hp\
anaconda3\lib\site-packages (from kaggle) (2.8.2)
Requirement already satisfied: requests in c:\users\hp\anaconda3\lib\
site-packages (from kaggle) (2.28.1)
Requirement already satisfied: bleach in c:\users\hp\anaconda3\lib\
site-packages (from kaggle) (4.1.0)
Requirement already satisfied: six>=1.10 in c:\users\hp\anaconda3\lib\
site-packages (from kaggle) (1.16.0)
Requirement already satisfied: tqdm in c:\users\hp\anaconda3\lib\site-
packages (from kaggle) (4.64.1)
Requirement already satisfied: python-slugify in c:\users\hp\
anaconda3\lib\site-packages (from kaggle) (5.0.2)
Requirement already satisfied: certifi>=2023.7.22 in c:\users\hp\
anaconda3\lib\site-packages (from kaggle) (2024.7.4)
Requirement already satisfied: webencodings in c:\users\hp\anaconda3\
lib\site-packages (from bleach->kaggle) (0.5.1)
Requirement already satisfied: packaging in c:\users\hp\anaconda3\lib\
site-packages (from bleach->kaggle) (22.0)
Requirement already satisfied: text-unidecode>=1.3 in c:\users\hp\
anaconda3\lib\site-packages (from python-slugify->kaggle) (1.3)
Requirement already satisfied: idna<4,>=2.5 in c:\users\hp\anaconda3\
lib\site-packages (from requests->kaggle) (3.4)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\
hp\anaconda3\lib\site-packages (from requests->kaggle) (2.0.4)
Requirement already satisfied: colorama in c:\users\hp\anaconda3\lib\
site-packages (from tqdm->kaggle) (0.4.6)
Dataset URL: https://www.kaggle.com/datasets/ankitbansal06/retail-
orders
License(s): CC0-1.0
Downloading orders.csv.zip to C:\Users\Hp\python



  0%|          | 0.00/200k [00:00<?, ?B/s]
100%|##########| 200k/200k [00:00<00:00, 277kB/s]
100%|##########| 200k/200k [00:00<00:00, 276kB/s]

import zipfile
zip_ref= zipfile.ZipFile('orders.csv.zip')
zip_ref.extractall()
zip_ref.close()
```

```
import pandas as pd
df = pd.read_csv('orders.csv',na_values = ['Not Available','unknown'])
df.head(20)
df['Ship Mode'].unique()
```

```
array(['Second Class', 'Standard Class', nan, 'First Class', 'Same
Day'],
      dtype=object)
```

```
df.head(20)
```

	Order Id	Order Date	Ship Mode	Segment	Country
0	1	2023-03-01	Second Class	Consumer	United States
1	2	2023-08-15	Second Class	Consumer	United States
2	3	2023-01-10	Second Class	Corporate	United States
3	4	2022-06-18	Standard Class	Consumer	United States
4	5	2022-07-13	Standard Class	Consumer	United States
5	6	2022-03-13	NaN	Consumer	United States
6	7	2022-12-28	Standard Class	Consumer	United States
7	8	2022-01-25	Standard Class	Consumer	United States
8	9	2023-03-23	NaN	Consumer	United States
9	10	2023-05-16	Standard Class	Consumer	United States
10	11	2023-03-31	NaN	Consumer	United States
11	12	2023-12-25	NaN	Consumer	United States
12	13	2022-02-11	Standard Class	Consumer	United States
13	14	2023-07-18	Standard Class	Consumer	United States
14	15	2023-11-09	NaN	Home Office	United States
15	16	2022-06-18	Standard Class	Home Office	United States
16	17	2022-02-04	Standard Class	Consumer	United States
17	18	2023-08-04	Second Class	Consumer	United States
18	19	2022-01-23	Second Class	Consumer	United States
19	20	2022-01-11	Second Class	Consumer	United States

```
               City           State  Postal Code   Region
Category  \
0         Henderson        Kentucky        42420    South
Furniture
1         Henderson        Kentucky        42420    South
Furniture
2       Los Angeles      California        90036     West  Office
Supplies
3   Fort Lauderdale         Florida        33311    South
Furniture
4   Fort Lauderdale         Florida        33311    South  Office
Supplies
5       Los Angeles      California        90032     West
Furniture
6       Los Angeles      California        90032     West  Office
Supplies
7       Los Angeles      California        90032     West
Technology
8       Los Angeles      California        90032     West  Office
Supplies
9       Los Angeles      California        90032     West  Office
Supplies
10      Los Angeles      California        90032     West
Furniture
11      Los Angeles      California        90032     West
Technology
12          Concord  North Carolina        28027    South  Office
Supplies
13          Seattle      Washington        98103     West  Office
Supplies
14       Fort Worth           Texas        76106  Central  Office
Supplies
15       Fort Worth           Texas        76106  Central  Office
Supplies
16          Madison       Wisconsin        53711  Central  Office
Supplies
17      West Jordan            Utah        84084     West  Office
Supplies
18    San Francisco      California        94109     West  Office
Supplies
19    San Francisco      California        94109     West
Technology

   Sub Category       Product Id  cost price  List Price  Quantity  \
0     Bookcases  FUR-BO-10001798         240         260         2
1        Chairs  FUR-CH-10000454         600         730         3
2        Labels  OFF-LA-10000240          10          10         2
3        Tables  FUR-TA-10000577         780         960         5
```

```
4       Storage  OFF-ST-10000760          20          20         2
5   Furnishings  FUR-FU-10001487          50          50         7
6           Art  OFF-AR-10002833          10          10         4
7        Phones  TEC-PH-10002275         860         910         6
8       Binders  OFF-BI-10003910          20          20         3
9    Appliances  OFF-AP-10002892          90         110         5
10       Tables  FUR-TA-10001539        1470        1710         9
11       Phones  TEC-PH-10002033         750         910         4
12        Paper  OFF-PA-10002365          20          20         3
13      Binders  OFF-BI-10003656         360         410         3
14   Appliances  OFF-AP-10002311          60          70         5
15      Binders  OFF-BI-10000756           0           0         3
16      Storage  OFF-ST-10004186         610         670         6
17      Storage  OFF-ST-10000107          60          60         2
18          Art  OFF-AR-10003056          10          10         2
19       Phones  TEC-PH-10001949         170         210         3
```

```
    Discount Percent
0                  2
1                  3
2                  5
3                  2
4                  5
5                  3
6                  3
7                  5
8                  2
9                  3
10                 3
11                 3
12                 3
13                 2
14                 5
15                 5
16                 3
17                 4
18                 4
19                 3
```

```
df.columns.str.lower()
df.columns = df.columns.str.lower()

df.columns.str.replace(' ','_')
df.columns= df.columns.str.replace(' ','_')

df.head(5)
```

```
   order_id  order_date       ship_mode    segment        country  \
0         1  2023-03-01    Second Class   Consumer  United States
1         2  2023-08-15    Second Class   Consumer  United States
```

```
2         3  2023-01-10    Second Class  Corporate  United States
3         4  2022-06-18  Standard Class   Consumer  United States
4         5  2022-07-13  Standard Class   Consumer  United States

              city       state  postal_code region         category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture
4  Fort Lauderdale     Florida        33311  South  Office Supplies

  sub_category       product_id  cost_price  list_price  quantity  \
0    Bookcases  FUR-BO-10001798         240         260         2
1       Chairs  FUR-CH-10000454         600         730         3
2       Labels  OFF-LA-10000240          10          10         2
3       Tables  FUR-TA-10000577         780         960         5
4      Storage  OFF-ST-10000760          20          20         2

   discount_percent
0                 2
1                 3
2                 5
3                 2
4                 5
```

```
df['discount']=df['list_price']*df['discount_percent']*.01

df['sale_price']=df['list_price']*df['discount']
df
```

```
      order_id  order_date       ship_mode    segment
country  \
0            1  2023-03-01    Second Class   Consumer  United States

1            2  2023-08-15    Second Class   Consumer  United States

2            3  2023-01-10    Second Class  Corporate  United States

3            4  2022-06-18  Standard Class   Consumer  United States

4            5  2022-07-13  Standard Class   Consumer  United States

...        ...         ...             ...        ...            ...

9989      9990  2023-02-18    Second Class   Consumer  United States

9990      9991  2023-03-17  Standard Class   Consumer  United States

9991      9992  2022-08-07  Standard Class   Consumer  United States

9992      9993  2022-11-19  Standard Class   Consumer  United States
```

```
9993      9994  2022-07-17    Second Class   Consumer  United States


                  city       state  postal_code region         category
\
0            Henderson    Kentucky        42420  South        Furniture

1            Henderson    Kentucky        42420  South        Furniture

2          Los Angeles  California        90036   West  Office Supplies

3     Fort Lauderdale     Florida        33311  South        Furniture

4     Fort Lauderdale     Florida        33311  South  Office Supplies

...                ...         ...          ...    ...              ...

9989             Miami     Florida        33180  South        Furniture

9990       Costa Mesa  California        92627   West        Furniture

9991       Costa Mesa  California        92627   West       Technology

9992       Costa Mesa  California        92627   West  Office Supplies

9993      Westminster  California        92683   West  Office Supplies


     sub_category       product_id  cost_price  list_price
quantity  \
0       Bookcases  FUR-BO-10001798         240         260         2

1         Chairs  FUR-CH-10000454         600         730         3

2         Labels  OFF-LA-10000240          10          10         2

3         Tables  FUR-TA-10000577         780         960         5

4        Storage  OFF-ST-10000760          20          20         2

...          ...              ...         ...         ...       ...

9989  Furnishings  FUR-FU-10001889          30          30         3

9990  Furnishings  FUR-FU-10000747          70          90         2

9991       Phones  TEC-PH-10003645         220         260         2

9992        Paper  OFF-PA-10004041          30          30         4

9993   Appliances  OFF-AP-10002684         210         240         2
```

```
      discount_percent  discount  sale_price
0                    2       5.2      1352.0
1                    3      21.9     15987.0
2                    5       0.5         5.0
3                    2      19.2     18432.0
4                    5       1.0        20.0
...                ...       ...         ...
9989                 4       1.2        36.0
9990                 4       3.6       324.0
9991                 2       5.2      1352.0
9992                 3       0.9        27.0
9993                 3       7.2      1728.0

[9994 rows x 18 columns]
```

```
df['profit']=df['sale_price']-df['cost_price']

df['order_date']=pd.to_datetime(df['order_date'],format="%Y-%m-%d")

df.drop(['cost_price','list_price','discount_percent'],axis=1,inplace=
True)
df
```

```
      order_id order_date       ship_mode    segment        country  \
0            1 2023-03-01    Second Class   Consumer  United States
1            2 2023-08-15    Second Class   Consumer  United States
2            3 2023-01-10    Second Class  Corporate  United States
3            4 2022-06-18  Standard Class   Consumer  United States
4            5 2022-07-13  Standard Class   Consumer  United States
...        ...        ...             ...        ...            ...
9989      9990 2023-02-18    Second Class   Consumer  United States
9990      9991 2023-03-17  Standard Class   Consumer  United States
9991      9992 2022-08-07  Standard Class   Consumer  United States
9992      9993 2022-11-19  Standard Class   Consumer  United States
9993      9994 2022-07-17    Second Class   Consumer  United States

                  city       state  postal_code region         category
\
0           Henderson    Kentucky        42420  South        Furniture

1           Henderson    Kentucky        42420  South        Furniture

2         Los Angeles  California        90036   West  Office Supplies

3     Fort Lauderdale     Florida        33311  South        Furniture

4     Fort Lauderdale     Florida        33311  South  Office Supplies

...               ...         ...          ...    ...              ...
```

```
9989            Miami     Florida        33180  South        Furniture

9990       Costa Mesa  California        92627   West        Furniture

9991       Costa Mesa  California        92627   West       Technology

9992       Costa Mesa  California        92627   West  Office Supplies

9993      Westminster  California        92683   West  Office Supplies


     sub_category       product_id  quantity  discount  sale_price
profit
0       Bookcases  FUR-BO-10001798         2       5.2      1352.0
1112.0
1          Chairs  FUR-CH-10000454         3      21.9     15987.0
15387.0
2          Labels  OFF-LA-10000240         2       0.5         5.0
-5.0
3          Tables  FUR-TA-10000577         5      19.2     18432.0
17652.0
4         Storage  OFF-ST-10000760         2       1.0        20.0
0.0
...           ...              ...       ...       ...         ...
...
9989  Furnishings  FUR-FU-10001889         3       1.2        36.0
6.0
9990  Furnishings  FUR-FU-10000747         2       3.6       324.0
254.0
9991       Phones  TEC-PH-10003645         2       5.2      1352.0
1132.0
9992        Paper  OFF-PA-10004041         4       0.9        27.0
-3.0
9993   Appliances  OFF-AP-10002684         2       7.2      1728.0
1518.0

[9994 rows x 16 columns]

!pip install pandas mysql-connector-python sqlalchemy

Requirement already satisfied: pandas in c:\users\hp\anaconda3\lib\
site-packages (1.5.3)
Collecting mysql-connector-python
  Downloading mysql_connector_python-9.0.0-cp310-cp310-win_amd64.whl
(14.3 MB)
     ---------------------------------------- 14.3/14.3 MB 8.5 MB/s
eta 0:00:00
Requirement already satisfied: sqlalchemy in c:\users\hp\anaconda3\
lib\site-packages (1.4.39)
```

```
Requirement already satisfied: numpy>=1.21.0 in c:\users\hp\anaconda3\
lib\site-packages (from pandas) (1.23.5)
Requirement already satisfied: pytz>=2020.1 in c:\users\hp\anaconda3\
lib\site-packages (from pandas) (2022.7)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\hp\
anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: greenlet!=0.4.17 in c:\users\hp\
anaconda3\lib\site-packages (from sqlalchemy) (2.0.1)
Requirement already satisfied: six>=1.5 in c:\users\hp\anaconda3\lib\
site-packages (from python-dateutil>=2.8.1->pandas) (1.16.0)
Installing collected packages: mysql-connector-python
Successfully installed mysql-connector-python-9.0.0
```

```
from sqlalchemy import create_engine
username = "root"
password ="Ppassword#5"
host = "localhost"
database = "order_data"

data= df

engine = create_engine(f'mysql+mysqlconnector://{username}:
{password}@{host}/{database}')


data.to_sql('df_order', con=engine, if_exists='append', index=False)
```

```
9994
```

```
/*Insights drawn from the data*/
/* top 10 highest revenue generating products*/
select product_id, sum(sale_price) as sales
from df_order
group by product_id
order by sales desc
limit 10;

/*find top 5 highest selling products in each region*/

With cte As (
select region , product_id, sum(sale_price)as sales
from df_order
group by region, product_id
order by region, sales desc)
SELECT * FROM(
select * ,
row_number() over (partition by region order by sales desc)AS RN
from cte)A
WHERE RN <=5
;

/*Find month over comparison for year 2022 and 2023*/
with CTE AS(
SELECT year(ORDER_DATE) AS ORDER_YEAR, month(ORDER_DATE) AS ORDER_MONTH,
SUM(SALE_PRICE) AS SALES
FROM DF_ORDER
GROUP BY year(ORDER_DATE), month(ORDER_DATE))
SELECT order_month,
sum(case when order_year = 2022 then sales else 0 end)AS Sales_2022,
sum(case when order_year = 2023 then sales else 0 end)AS Sales_2023
from cte
group by order_month
order by order_month;
```

```
/*Insights drawn from the data*/
/*For each category which month has the highest sales*/

WITH cte as(
select category, date_format(order_date,'%Y%M'), sum(sale_price) as sales
from df_order
group by category, date_format(order_date,'%Y%M')
order by category, date_format(order_date,'%Y%M'))
select * from(
select *,
row_number() over (partition by category order by sales desc) as rn
from cte)A
WHERE rn <=1;


/*Which sub category has highest growth by profit in 2023 compared to 2022*/

with CTE AS(
SELECT sub_category, year(ORDER_DATE) AS ORDER_YEAR,  SUM(SALE_PRICE) AS SALES
FROM DF_ORDER
GROUP BY sub_category,year(ORDER_DATE)),
cte2 as(
SELECT sub_category,
sum(case when order_year = 2022 then sales else 0 end)AS Sales_2022,
sum(case when order_year = 2023 then sales else 0 end)AS Sales_2023
from cte
group by sub_category)
select  *,
(sales_2023- sales_2022)*100/sales_2022
from cte2
order by (sales_2023- sales_2022)*100/sales_2022 desc limit 1
;
```