```
import sqlalchemy as sqa
import pandas as pd
import pyodbc
import pymysql
username = 'root'
password = 'jobaer126'
host = 'localhost'
port = '3306'
database = 'kaggle_data'
url = 'mysql+pymysql://{0}:{1}@{2}:{3}/{4}'.format(username, password,
host, port, database)
mysql engine = sqa.create engine(url)
mysql engine
Engine(mysql+pymysql://root:***@localhost:3306/kaggle data)
import pandas as pd
file path = 'D:\\excel\\online retail II.xlsx' # Replace with your
actual CSV file path
sheet name = 'Year 2010-2011'
# Load data from CSV file
data = pd.read excel(file path, sheet name)
print(data)
       Invoice StockCode
                                                 Description
Quantity \
                 85123A WHITE HANGING HEART T-LIGHT HOLDER
       536365
6
1
                  71053
                                         WHITE METAL LANTERN
       536365
6
2
                              CREAM CUPID HEARTS COAT HANGER
       536365
                 84406B
8
3
                 84029G KNITTED UNION FLAG HOT WATER BOTTLE
       536365
6
4
       536365
                 84029E
                              RED WOOLLY HOTTIE WHITE HEART.
6
                                CHILDREN'S APRON DOLLY GIRL
541905 581587
                  22899
541906 581587
                               CHILDRENS CUTLERY DOLLY GIRL
                  23254
                             CHILDRENS CUTLERY CIRCUS PARADE
541907 581587
                  23255
541908 581587
                  22138
                               BAKING SET 9 PIECE RETROSPOT
```

```
541909 581587
                   P0ST
                                                    POSTAGE
1
                          Price Customer ID
                                                     Country
              InvoiceDate
      2010-12-01 08:26:00
                           2.55
                                     17850.0
                                              United Kingdom
0
                            3.39
1
      2010-12-01 08:26:00
                                     17850.0
                                              United Kingdom
2
      2010-12-01 08:26:00
                            2.75
                                     17850.0
                                              United Kingdom
3
      2010-12-01 08:26:00
                            3.39
                                     17850.0
                                              United Kingdom
4
      2010-12-01 08:26:00
                            3.39
                                              United Kingdom
                                     17850.0
                            2.10
541905 2011-12-09 12:50:00
                                     12680.0
                                                      France
541906 2011-12-09 12:50:00
                           4.15
                                     12680.0
                                                      France
541907 2011-12-09 12:50:00
                           4.15
                                     12680.0
                                                      France
541908 2011-12-09 12:50:00
                          4.95
                                     12680.0
                                                      France
541909 2011-12-09 12:50:00 18.00
                                     12680.0
                                                      France
[541910 rows x 8 columns]
cleaned data = data[data['Quantity'] >= 0]
print(cleaned data)
      Invoice StockCode
                                                Description
Ouantity \
                         WHITE HANGING HEART T-LIGHT HOLDER
       536365 85123A
6
1
                                        WHITE METAL LANTERN
       536365
                  71053
6
2
       536365 84406B
                              CREAM CUPID HEARTS COAT HANGER
8
3
       536365 84029G KNITTED UNION FLAG HOT WATER BOTTLE
6
4
                             RED WOOLLY HOTTIE WHITE HEART.
       536365
                 84029E
6
541905 581587
                  22899
                               CHILDREN'S APRON DOLLY GIRL
541906 581587
                              CHILDRENS CUTLERY DOLLY GIRL
                  23254
                             CHILDRENS CUTLERY CIRCUS PARADE
541907 581587
                  23255
                               BAKING SET 9 PIECE RETROSPOT
541908 581587
                  22138
541909 581587
                   P0ST
                                                    POSTAGE
              InvoiceDate Price Customer ID
                                                     Country
0
      2010-12-01 08:26:00
                            2.55
                                     17850.0
                                              United Kingdom
1
      2010-12-01 08:26:00
                                              United Kingdom
                            3.39
                                     17850.0
```

```
2
       2010-12-01 08:26:00
                            2.75
                                               United Kingdom
                                       17850.0
3
       2010-12-01 08:26:00
                             3.39
                                       17850.0
                                               United Kingdom
4
       2010-12-01 08:26:00
                            3.39
                                       17850.0
                                               United Kingdom
541905 2011-12-09 12:50:00
                            2.10
                                       12680.0
                                                        France
541906 2011-12-09 12:50:00
                            4.15
                                       12680.0
                                                       France
541907 2011-12-09 12:50:00
                            4.15
                                       12680.0
                                                       France
541908 2011-12-09 12:50:00
                            4.95
                                       12680.0
                                                       France
541909 2011-12-09 12:50:00 18.00
                                       12680.0
                                                       France
[531286 rows x 8 columns]
# Create a copy first
cleaned data = data[data['Quantity'] >= 0].copy()
# Then drop missing values
cleaned data.dropna(subset=['Customer ID', 'Description'],
inplace=True)
print(cleaned data)
       Invoice StockCode
                                                 Description
Quantity \
        536365
                 85123A
                          WHITE HANGING HEART T-LIGHT HOLDER
6
1
        536365 71053
                                         WHITE METAL LANTERN
6
2
                              CREAM CUPID HEARTS COAT HANGER
        536365
                 84406B
8
3
        536365
                 84029G KNITTED UNION FLAG HOT WATER BOTTLE
6
4
                              RED WOOLLY HOTTIE WHITE HEART.
        536365
                 84029E
6
                                CHILDREN'S APRON DOLLY GIRL
541905
       581587
                  22899
6
                               CHILDRENS CUTLERY DOLLY GIRL
541906 581587
                  23254
541907 581587
                             CHILDRENS CUTLERY CIRCUS PARADE
                  23255
541908 581587
                  22138
                               BAKING SET 9 PIECE RETROSPOT
541909 581587
                    P0ST
                                                     POSTAGE
               InvoiceDate
                           Price Customer ID
                                                      Country
                                               United Kingdom
       2010-12-01 08:26:00
                            2.55
                                       17850.0
       2010-12-01 08:26:00
                            3.39
1
                                       17850.0
                                               United Kingdom
2
       2010-12-01 08:26:00
                            2.75
                                      17850.0
                                               United Kingdom
```

```
3
       2010-12-01 08:26:00
                              3.39
                                                  United Kingdom
                                         17850.0
4
       2010-12-01 08:26:00
                                                  United Kingdom
                              3.39
                                         17850.0
                               . . .
541905 2011-12-09 12:50:00
                              2.10
                                         12680.0
                                                          France
541906 2011-12-09 12:50:00
                              4.15
                                         12680.0
                                                          France
541907 2011-12-09 12:50:00
                              4.15
                                         12680.0
                                                          France
541908 2011-12-09 12:50:00
                              4.95
                                         12680.0
                                                          France
541909 2011-12-09 12:50:00 18.00
                                         12680.0
                                                          France
[397925 rows x 8 columns]
import os
print(os.getcwd())
cleaned data.to csv('C:\\Users\\USER\\cleaned sales data.csv',
index=False)
C:\Users\USER
file path ='C:\\Users\\USER\\Documents\\cleaned sales data.csv'
data1 = pd.read_csv(file_path)
print(data1)
        Invoice StockCode
                                                     Description
Quantity \
                             WHITE HANGING HEART T-LIGHT HOLDER
         536365
                    85123A
6
1
                    71053
                                             WHITE METAL LANTERN
         536365
6
2
         536365
                    84406B
                                 CREAM CUPID HEARTS COAT HANGER
8
3
         536365
                    84029G
                            KNITTED UNION FLAG HOT WATER BOTTLE
6
4
                                 RED WOOLLY HOTTIE WHITE HEART.
         536365
                    84029E
6
. . .
397920
         581587
                     22899
                                   CHILDREN'S APRON DOLLY GIRL
6
                                  CHILDRENS CUTLERY DOLLY GIRL
         581587
                     23254
397921
397922
         581587
                     23255
                                CHILDRENS CUTLERY CIRCUS PARADE
397923
         581587
                     22138
                                  BAKING SET 9 PIECE RETROSPOT
397924
         581587
                      P<sub>0</sub>ST
                                                         POSTAGE
                InvoiceDate
                              Price
                                     Customer ID
                                                          Country
        2010-12-01 08:26:00
                               2.55
                                                   United Kingdom
0
                                          17850.0
        2010-12-01 08:26:00
1
                               3.39
                                                   United Kingdom
                                          17850.0
```

```
2
        2010-12-01 08:26:00
                              2.75
                                        17850.0
                                                 United Kingdom
3
                              3.39
                                                 United Kingdom
        2010-12-01 08:26:00
                                        17850.0
4
        2010-12-01 08:26:00
                              3.39
                                        17850.0
                                                 United Kingdom
397920
       2011-12-09 12:50:00
                              2.10
                                        12680.0
                                                          France
397921
       2011-12-09 12:50:00
                              4.15
                                        12680.0
                                                          France
397922 2011-12-09 12:50:00
                              4.15
                                        12680.0
                                                          France
397923
        2011-12-09 12:50:00
                              4.95
                                        12680.0
                                                          France
397924 2011-12-09 12:50:00
                             18.00
                                        12680.0
                                                          France
[397925 rows x 8 columns]
data1.to sql('year 2010-2011', con=mysql engine, if exists='append',
index=False)
397925
import pandas as pd
file path = 'D:\\excel\\online_retail_II.xlsx' # Replace with your
actual CSV file path
sheet name = 'Year 2009-2010'
# Load data from CSV file
df = pd.read excel(file path, sheet name)
print(df)
       Invoice StockCode
                                                   Description
Quantity \
0
        489434
                   85048 15CM CHRISTMAS GLASS BALL 20 LIGHTS
12
1
        489434
                  79323P
                                            PINK CHERRY LIGHTS
12
2
                                          WHITE CHERRY LIGHTS
        489434
                  79323W
12
                                 RECORD FRAME 7" SINGLE SIZE
3
        489434
                   22041
48
        489434
                   21232
                               STRAWBERRY CERAMIC TRINKET BOX
4
24
525456 538171
                   22271
                                         FELTCRAFT DOLL ROSIE
                                 FELTCRAFT PRINCESS LOLA DOLL
525457 538171
                   22750
525458 538171
                               FELTCRAFT PRINCESS OLIVIA DOLL
                   22751
                           PINK FLORAL FELTCRAFT SHOULDER BAG
525459 538171
                   20970
                                       JUMBO STORAGE BAG SUKI
525460 538171
                   21931
```

```
2
               InvoiceDate
                           Price Customer ID
                                                      Country
0
       2009-12-01 07:45:00
                            6.95
                                       13085.0
                                               United Kingdom
1
       2009-12-01 07:45:00
                             6.75
                                               United Kingdom
                                       13085.0
                            6.75
2
       2009-12-01 07:45:00
                                       13085.0
                                               United Kingdom
3
       2009-12-01 07:45:00
                            2.10
                                       13085.0
                                               United Kingdom
                                               United Kingdom
4
       2009-12-01 07:45:00
                             1.25
                                       13085.0
                             . . .
525456 2010-12-09 20:01:00
                            2.95
                                       17530.0
                                               United Kingdom
525457 2010-12-09 20:01:00
                            3.75
                                       17530.0
                                               United Kingdom
525458 2010-12-09 20:01:00
                            3.75
                                       17530.0
                                               United Kingdom
                                       17530.0
                                               United Kingdom
525459 2010-12-09 20:01:00
                            3.75
525460 2010-12-09 20:01:00
                                               United Kingdom
                            1.95
                                      17530.0
[525461 rows x 8 columns]
cleaned data = df[df['Quantity'] >= 0]
print(cleaned data)
                                                 Description
       Invoice StockCode
Quantity \
        489434
                  85048 15CM CHRISTMAS GLASS BALL 20 LIGHTS
12
1
                 79323P
                                           PINK CHERRY LIGHTS
        489434
12
                                         WHITE CHERRY LIGHTS
2
        489434
                 79323W
12
3
        489434
                   22041
                                RECORD FRAME 7" SINGLE SIZE
48
4
                               STRAWBERRY CERAMIC TRINKET BOX
        489434
                   21232
24
525456
       538171
                   22271
                                        FELTCRAFT DOLL ROSIE
525457 538171
                  22750
                                 FELTCRAFT PRINCESS LOLA DOLL
                               FELTCRAFT PRINCESS OLIVIA DOLL
525458 538171
                   22751
                   20970
                           PINK FLORAL FELTCRAFT SHOULDER BAG
525459 538171
525460 538171
                  21931
                                       JUMBO STORAGE BAG SUKI
               InvoiceDate
                           Price Customer ID
                                                      Country
       2009-12-01 07:45:00
                            6.95
                                               United Kingdom
0
                                       13085.0
       2009-12-01 07:45:00
1
                            6.75
                                       13085.0
                                               United Kingdom
2
       2009-12-01 07:45:00
                            6.75
                                      13085.0
                                               United Kingdom
```

```
3
       2009-12-01 07:45:00
                             2.10
                                                United Kingdom
                                       13085.0
4
       2009-12-01 07:45:00
                                                United Kingdom
                             1.25
                                       13085.0
                              . . .
525456 2010-12-09 20:01:00
                             2.95
                                       17530.0
                                                United Kingdom
                                       17530.0
525457 2010-12-09 20:01:00
                             3.75
                                                United Kingdom
525458 2010-12-09 20:01:00
                                                United Kingdom
                             3.75
                                       17530.0
525459 2010-12-09 20:01:00
                                                United Kingdom
                             3.75
                                       17530.0
525460 2010-12-09 20:01:00
                             1.95
                                       17530.0
                                                United Kingdom
[513135 rows x 8 columns]
# Create a copy first
cleaned data = df[df['Quantity'] >= 0].copy()
# Then drop missing values
cleaned data.dropna(subset=['Customer ID', 'Description'],
inplace=True)
print(cleaned_data)
       Invoice StockCode
                                                  Description
Quantity \
0
        489434
                  85048 15CM CHRISTMAS GLASS BALL 20 LIGHTS
12
1
        489434
                 79323P
                                           PINK CHERRY LIGHTS
12
2
                                          WHITE CHERRY LIGHTS
        489434
                  79323W
12
3
        489434
                   22041
                                 RECORD FRAME 7" SINGLE SIZE
48
                               STRAWBERRY CERAMIC TRINKET BOX
        489434
                   21232
24
                                         FELTCRAFT DOLL ROSIE
525456 538171
                   22271
                   22750
                                 FELTCRAFT PRINCESS LOLA DOLL
525457 538171
525458 538171
                   22751
                               FELTCRAFT PRINCESS OLIVIA DOLL
1
                           PINK FLORAL FELTCRAFT SHOULDER BAG
525459 538171
                   20970
525460 538171
                   21931
                                       JUMBO STORAGE BAG SUKI
                            Price Customer ID
               InvoiceDate
                                                       Country
0
       2009-12-01 07:45:00
                             6.95
                                       13085.0
                                                United Kingdom
1
       2009-12-01 07:45:00
                             6.75
                                                United Kingdom
                                       13085.0
2
       2009-12-01 07:45:00
                             6.75
                                       13085.0
                                                United Kingdom
3
       2009-12-01 07:45:00
                                                United Kingdom
                             2.10
                                       13085.0
```

```
4
       2009-12-01 07:45:00
                              1.25
                                                 United Kingdom
                                        13085.0
525456 2010-12-09 20:01:00
                              2.95
                                        17530.0
                                                 United Kingdom
525457 2010-12-09 20:01:00
                                                 United Kingdom
                              3.75
                                        17530.0
525458 2010-12-09 20:01:00
                              3.75
                                        17530.0
                                                 United Kingdom
525459 2010-12-09 20:01:00
                              3.75
                                        17530.0
                                                 United Kingdom
                                                 United Kingdom
525460 2010-12-09 20:01:00
                              1.95
                                        17530.0
[407695 rows x 8 columns]
import os
print(os.getcwd())
cleaned data.to csv('C:\\Users\\USER\\sales 2009 10.csv', index=False)
C:\Users\USER
file path = 'C:\\Users\\USER\\Documents\\sales 2009 10.csv'
df= pd.read csv(file path)
print(df)
        Invoice StockCode
                                                     Description
Quantity
                            15CM CHRISTMAS GLASS BALL 20 LIGHTS
0
         489434
                    85048
12
                   79323P
                                             PINK CHERRY LIGHTS
1
         489434
12
2
         489434
                   79323W
                                            WHITE CHERRY LIGHTS
12
3
         489434
                    22041
                                   RECORD FRAME 7" SINGLE SIZE
48
4
         489434
                    21232
                                 STRAWBERRY CERAMIC TRINKET BOX
24
. . .
407690
         538171
                    22271
                                           FELTCRAFT DOLL ROSIE
2
407691
                    22750
                                   FELTCRAFT PRINCESS LOLA DOLL
         538171
407692
                                 FELTCRAFT PRINCESS OLIVIA DOLL
         538171
                    22751
407693
         538171
                    20970
                             PINK FLORAL FELTCRAFT SHOULDER BAG
407694
         538171
                    21931
                                         JUMBO STORAGE BAG SUKI
                InvoiceDate
                              Price
                                     Customer ID
                                                          Country
        2009-12-01 07:45:00
                                                  United Kingdom
0
                               6.95
                                         13085.0
1
        2009-12-01 07:45:00
                               6.75
                                         13085.0
                                                  United Kingdom
2
        2009-12-01 07:45:00
                               6.75
                                                  United Kingdom
                                         13085.0
3
        2009-12-01 07:45:00
                                         13085.0
                                                  United Kingdom
                               2.10
```

```
4
       2009-12-01 07:45:00
                             1.25
                                       13085.0 United Kingdom
                                       17530.0 United Kingdom
407690 2010-12-09 20:01:00
                             2.95
407691 2010-12-09 20:01:00
                             3.75
                                       17530.0 United Kingdom
                                       17530.0 United Kingdom
407692 2010-12-09 20:01:00
                             3.75
407693 2010-12-09 20:01:00
                             3.75
                                       17530.0 United Kingdom
407694 2010-12-09 20:01:00
                             1.95
                                       17530.0 United Kingdom
[407695 rows x 8 columns]
df.to_sql('year_2009_2010', con=mysql_engine, if_exists='append',
index=False)
407695
```