

```

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 \			
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER
6			
1	536365	71053	WHITE METAL LANTERN
6			
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER
8			
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE
6			
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.
6			
...	...	...	...
.			
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL
6			
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL
4			
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE
4			
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT
3			

541909	581587	POST			POSTAGE
1					

  

	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	3.39	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	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

[541910 rows x 8 columns]

```
cleaned_data = data[data['Quantity'] >= 0]
print(cleaned_data)
```

	Invoice	StockCode	Description
Quantity \			
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER
6			
1	536365	71053	WHITE METAL LANTERN
6			
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER
8			
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE
6			
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.
6			
...	...	...	...
.			
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL
6			
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL
4			
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE
4			
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT
3			
541909	581587	POST	POSTAGE
1			

  

	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	3.39	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      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 \			
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER
6			
1	536365	71053	WHITE METAL LANTERN
6			
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER
8			
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE
6			
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.
6			
...	...	...	...
.			..
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL
6			
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL
4			
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE
4			
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT
3			
541909	581587	POST	POSTAGE
1			

  

	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	3.39	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	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

[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 \			
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER
6			
1	536365	71053	WHITE METAL LANTERN
6			
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER
8			
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE
6			
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.
6			
...	...	...	...
..			
397920	581587	22899	CHILDREN'S APRON DOLLY GIRL
6			
397921	581587	23254	CHILDRENS CUTLERY DOLLY GIRL
4			
397922	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE
4			
397923	581587	22138	BAKING SET 9 PIECE RETROSPOT
3			
397924	581587	POST	POSTAGE
1			

  

	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	3.39	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	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	489434	79323W	WHITE CHERRY LIGHTS
12			
3	489434	22041	RECORD FRAME 7" SINGLE SIZE
48			
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX
24			
...	...	...	...
.			
525456	538171	22271	FELTCRAFT DOLL ROSIE
2			
525457	538171	22750	FELTCRAFT PRINCESS LOLA DOLL
1			
525458	538171	22751	FELTCRAFT PRINCESS OLIVIA DOLL
1			
525459	538171	20970	PINK FLORAL FELTCRAFT SHOULDER BAG
2			
525460	538171	21931	JUMBO STORAGE BAG SUKI

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	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	13085.0	United Kingdom
4	2009-12-01 07:45:00	1.25	13085.0	United Kingdom
...	...	...	...	...
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
525459	2010-12-09 20:01:00	3.75	17530.0	United Kingdom
525460	2010-12-09 20:01:00	1.95	17530.0	United Kingdom

[525461 rows x 8 columns]

```
cleaned_data = df[df['Quantity'] >= 0]
```

```
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	489434	79323W	WHITE CHERRY LIGHTS
12			
3	489434	22041	RECORD FRAME 7" SINGLE SIZE
48			
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX
24			
...	...	...	...
.			..
525456	538171	22271	FELTCRAFT DOLL ROSIE
2			
525457	538171	22750	FELTCRAFT PRINCESS LOLA DOLL
1			
525458	538171	22751	FELTCRAFT PRINCESS OLIVIA DOLL
1			
525459	538171	20970	PINK FLORAL FELTCRAFT SHOULDER BAG
2			
525460	538171	21931	JUMBO STORAGE BAG SUKI
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	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      13085.0 United Kingdom
4      2009-12-01 07:45:00    1.25      13085.0 United Kingdom
...
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
525459 2010-12-09 20:01:00    3.75      17530.0 United Kingdom
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	489434	79323W	WHITE CHERRY LIGHTS
12			
3	489434	22041	RECORD FRAME 7" SINGLE SIZE
48			
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX
24			
...	...	...	...
.			
525456	538171	22271	FELTCRAFT DOLL ROSIE
2			
525457	538171	22750	FELTCRAFT PRINCESS LOLA DOLL
1			
525458	538171	22751	FELTCRAFT PRINCESS OLIVIA DOLL
1			
525459	538171	20970	PINK FLORAL FELTCRAFT SHOULDER BAG
2			
525460	538171	21931	JUMBO STORAGE BAG SUKI
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	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	13085.0	United Kingdom

4	2009-12-01 07:45:00	1.25	13085.0	United Kingdom
...	...	...	...	...
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
525459	2010-12-09 20:01:00	3.75	17530.0	United Kingdom
525460	2010-12-09 20:01:00	1.95	17530.0	United Kingdom

[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 \			
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS
12			
1	489434	79323P	PINK CHERRY LIGHTS
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	538171	22750	FELTCRAFT PRINCESS LOLA DOLL
1			
407692	538171	22751	FELTCRAFT PRINCESS OLIVIA DOLL
1			
407693	538171	20970	PINK FLORAL FELTCRAFT SHOULDER BAG
2			
407694	538171	21931	JUMBO STORAGE BAG SUKI
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	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	13085.0	United Kingdom



4	2009-12-01 07:45:00	1.25	13085.0	United Kingdom
...	...	...	...	...
407690	2010-12-09 20:01:00	2.95	17530.0	United Kingdom
407691	2010-12-09 20:01:00	3.75	17530.0	United Kingdom
407692	2010-12-09 20:01:00	3.75	17530.0	United Kingdom
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