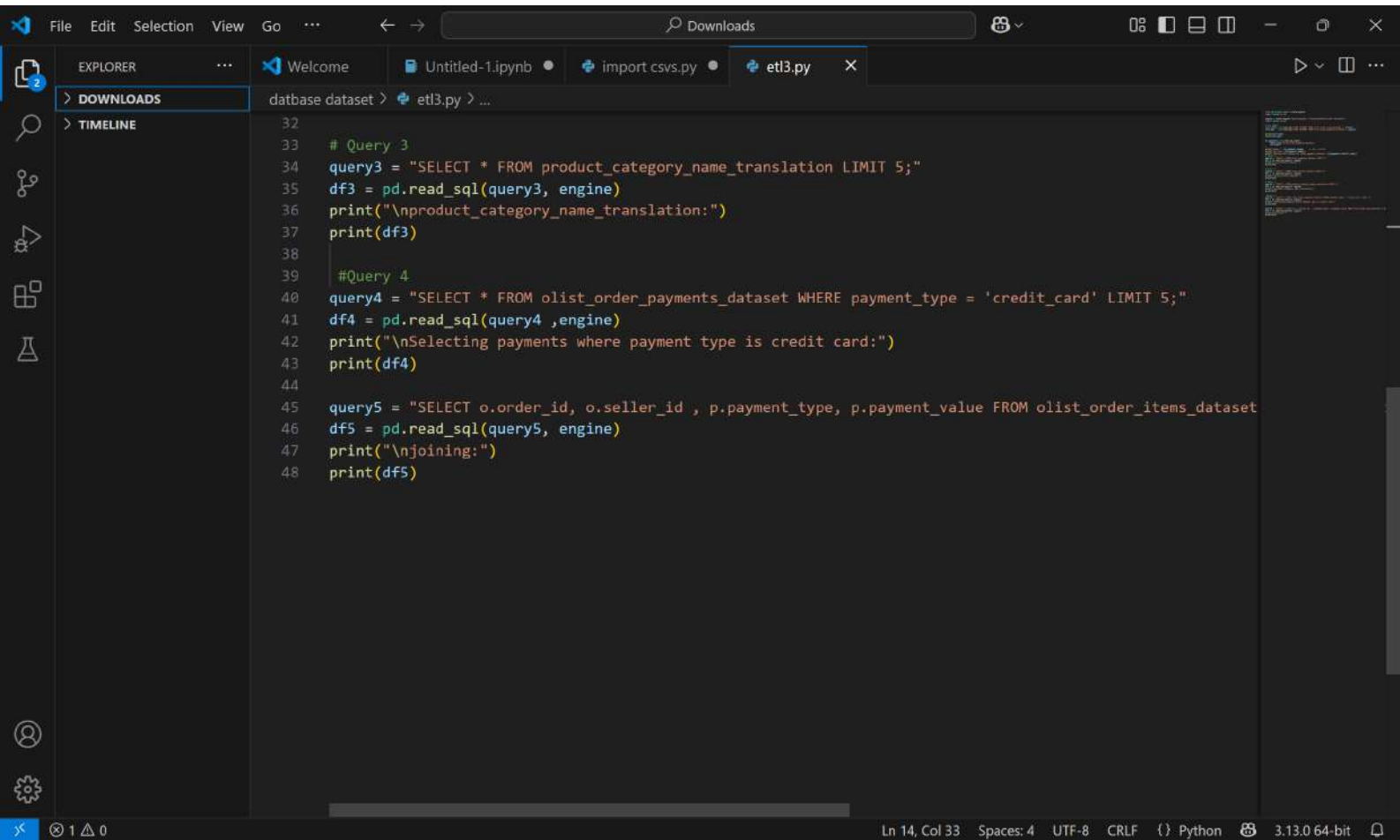


```
database dataset > etl3.py > ...
1  from sqlalchemy import create_engine
2  import pandas as pd
3
4  engine = create_engine('mysql+pymysql://root:@localhost/retail business')
5  import pandas as pd
6
7  # For MySQL
8  cols_items = pd.read_sql("SHOW COLUMNS FROM olist_order_items_dataset;", engine)
9  cols_pay = pd.read_sql("SHOW COLUMNS FROM olist_order_payments_dataset;", engine)
10
11 print(cols_items)
12 print(cols_pay)
13
14 df_payments = pd.read_sql_table(
15     table_name='olist_order_payments_dataset',
16     con=engine
17 )
18 print("shape :", df_payments.shape)    # rows, columns
19 print("top rows:",df_payments.head())
20 print("checking null values for order_payments_dataset:",df_payments.isnull().sum())
21 # Query 1
22 query1 = "SELECT * FROM olist_products_dataset LIMIT 5;"
23 df1 = pd.read_sql(query1, engine)
24 print("olist_products_dataset:")
25 print(df1)
26
27 # Query 2
28 query2 = "SELECT * FROM olist_sellers_dataset LIMIT 5;"
29 df2 = pd.read_sql(query2, engine)
30 print("\nolist_sellers_dataset:")
31 print(df2)
32
```

Ln 14, Col 33 Spaces: 4 UTF-8 CRLF {} Python 3.13.0 64-bit



```
32
33 # Query 3
34 query3 = "SELECT * FROM product_category_name_translation LIMIT 5;"
35 df3 = pd.read_sql(query3, engine)
36 print("\nproduct_category_name_translation:")
37 print(df3)
38
39 #Query 4
40 query4 = "SELECT * FROM olist_order_payments_dataset WHERE payment_type = 'credit_card' LIMIT 5;"
41 df4 = pd.read_sql(query4, engine)
42 print("\nSelecting payments where payment type is credit card:")
43 print(df4)
44
45 query5 = "SELECT o.order_id, o.seller_id, p.payment_type, p.payment_value FROM olist_order_items_dataset
46 df5 = pd.read_sql(query5, engine)
47 print("\njoining:")
48 print(df5)
```

Ln 14, Col 33 Spaces: 4 UTF-8 CRLF {} Python 3.13.0 64-bit

FileEditSelectionViewGo...Downloads

EXPLORER

> DOWNLOADS  
> TIMELINE

OUTPUT

DEBUG CONSOLE

TERMINAL

PROBLEMS 1

PORTS

PS C:\Users\varsh\Downloads> & C:/Users/varsh/AppData/Local/Programs/Python/Python313/python.exe "c:/Users/varsh/Downloads/database\_dataset/etl3.py"

	Field	Type	Null	Key	Default	Extra
0	order_id	text	YES		None	
1	order_item_id	bigint(20)	YES		None	
2	product_id	text	YES		None	
3	seller_id	text	YES		None	
4	shipping_limit_date	text	YES		None	
5	price	double	YES		None	
6	freight_value	double	YES		None	

	Field	Type	Null	Key	Default	Extra
0	order_id	text	YES		None	
1	payment_sequential	bigint(20)	YES		None	
2	payment_type	text	YES		None	
3	payment_installments	bigint(20)	YES		None	
4	payment_value	double	YES		None	

shape : (103886, 5)  
top rows:

	order_id	payment_sequential	payment_type	payment_installments	payment_v
0	b81ef226f3fe1789b1e8b2acac839d17	1	credit_card	8	99.33
1	a9810da82917af2d9aefd1278f1dcfa0	1	credit_card	1	24.39
2	25e8ea4e93396b6fa0d3dd708e76c1bd	1	credit_card	1	65.71
3	ba78997921bbcdc1373bb41e913ab953	1	credit_card	8	107.78
4	42fdf880ba16b47b59251dd489d4441a	1	credit_card	2	128.45

checking null values for order\_payments\_dataset:

order_id	0
payment_sequential	0
payment_type	0
payment_installments	0
payment_value	0

dtype: int64  
olist\_products\_dataset:

	product_id	product_category_name	...	product_height_cm	product_width_cm
0	1e9e8ef04dbcff4541ed26657ea517e5	perfumaria	...	10.0	14.0
1	3aa071139cb16b67ca9e5dea641aaa2f	artes	...	18.0	20.0
2	96bd76ec8810374ed1b65e291975717f	esporte_lazer	...	9.0	15.0
3	cef67bcfe19066a932b7673e239eb23d	bebes	...	4.0	26.0
4	9dc1a7de274444849c219cff195d0b71	utilidades_domesticas	...	17.0	13.0

Python 3.13.0 64-bit

Ln 14, Col 33 Spaces: 4 UTF-8 CRLF {} Python 3.13.0 64-bit

