```python
import pandas as pd
import sqlite3

# Load CSVs
customers = pd.read_csv('olist_customers_dataset.csv')
geolocation = pd.read_csv('olist_geolocation_dataset.csv')
order_items = pd.read_csv('olist_order_items_dataset.csv')
order_payments = pd.read_csv('olist_order_payments_dataset.csv')
order_reviews = pd.read_csv('olist_order_reviews_dataset.csv')
orders = pd.read_csv('olist_orders_dataset.csv')
products = pd.read_csv('olist_products_dataset.csv')
sellers = pd.read_csv('olist_sellers_dataset.csv')
category_translation =
pd.read_csv('product_category_name_translation.csv')

# Connect to in-memory SQLite DB
conn = sqlite3.connect(':memory:')

# Save to SQLite
customers.to_sql('customers', conn, index=False)
geolocation.to_sql('geolocation', conn, index=False)
order_items.to_sql('order_items', conn, index=False)
order_payments.to_sql('order_payments', conn, index=False)
order_reviews.to_sql('order_reviews', conn, index=False)
orders.to_sql('orders', conn, index=False)
products.to_sql('products', conn, index=False)
sellers.to_sql('sellers', conn, index=False)
category_translation.to_sql('category_translation', conn, index=False)
```

```
71
```

```python
dfs = {
    "customers" : customers,
    "geolocation" : geolocation,
    "order_items" : order_items,
    "order_payments" : order_items,
    "order_reviews" : order_reviews,
    "orders" : orders,
    "products" : products,
    "sellers" : sellers,
    "category_translation" : category_translation

}
for name, df in dfs.items():
    print(f"\n\n  ---  for table «{name.upper()}» shape is {df.shape}
---\n")
```

```


  ---  for table «CUSTOMERS» shape is (99441, 5)  ---
  ---  for table «GEOLOCATION» shape is (1000163, 5)  ---
```

```
  ---  for table «ORDER_ITEMS» shape is (112650, 7)  ---
```

```
  ---  for table «ORDER_PAYMENTS» shape is (112650, 7)  ---
```

```
  ---  for table «ORDER_REVIEWS» shape is (99224, 7)  ---
```

```
  ---  for table «ORDERS» shape is (99441, 8)  ---
```

```
  ---  for table «PRODUCTS» shape is (32951, 9)  ---
```

```
  ---  for table «SELLERS» shape is (3095, 4)  ---
```

```
  ---  for table «CATEGORY_TRANSLATION» shape is (71, 2)  ---
```

```
for name, df in dfs.items():
    print(f"First 2 rows:\n{df.head(2)}\n\n")
```

```
First 2 rows:
                        customer_id                customer_unique_id  \
0  06b8999e2fba1a1fbc88172c00ba8bc7  861eff4711a542e4b93843c6dd7febb0   
1  18955e83d337fd6b2def6b18a428ac77  290c77bc529b7ac935b93aa66c333dc3   

   customer_zip_code_prefix          customer_city customer_state  
0                     14409                 franca             SP  
1                      9790  sao bernardo do campo             SP  


First 2 rows:
   geolocation_zip_code_prefix  geolocation_lat  geolocation_lng  \
0                         1037       -23.545621       -46.639292   
```

```
1                         1046       -23.546081       -46.644820

  geolocation_city geolocation_state
0        sao paulo                SP
1        sao paulo                SP


First 2 rows:
                           order_id  order_item_id  \
0  00010242fe8c5a6d1ba2dd792cb16214              1
1  00018f77f2f0320c557190d7a144bdd3              1

                         product_id                         seller_id
\
0  4244733e06e7ecb4970a6e2683c13e61  48436dade18ac8b2bce089ec2a041202

1  e5f2d52b802189ee658865ca93d83a8f  dd7ddc04e1b6c2c614352b383efe2d36


   shipping_limit_date  price  freight_value
0  2017-09-19 09:45:35   58.9          13.29
1  2017-05-03 11:05:13  239.9          19.93


First 2 rows:
                           order_id  order_item_id  \
0  00010242fe8c5a6d1ba2dd792cb16214              1
1  00018f77f2f0320c557190d7a144bdd3              1

                         product_id                         seller_id
\
0  4244733e06e7ecb4970a6e2683c13e61  48436dade18ac8b2bce089ec2a041202

1  e5f2d52b802189ee658865ca93d83a8f  dd7ddc04e1b6c2c614352b383efe2d36


   shipping_limit_date  price  freight_value
0  2017-09-19 09:45:35   58.9          13.29
1  2017-05-03 11:05:13  239.9          19.93


First 2 rows:
                          review_id                          order_id
\
0  7bc2406110b926393aa56f80a40eba40  73fc7af87114b39712e6da79b0a377eb

1  80e641a11e56f04c1ad469d5645fdfde  a548910a1c6147796b98fdf73dbeba33


   review_score review_comment_title review_comment_message  \
0             4                  NaN                    NaN
```

```
1             5                  NaN                    NaN

  review_creation_date review_answer_timestamp
0  2018-01-18 00:00:00     2018-01-18 21:46:59
1  2018-03-10 00:00:00     2018-03-11 03:05:13


First 2 rows:
                           order_id                       customer_id
\
0  e481f51cbdc54678b7cc49136f2d6af7  9ef432eb6251297304e76186b10a928d

1  53cdb2fc8bc7dce0b6741e2150273451  b0830fb4747a6c6d20dea0b8c802d7ef


  order_status order_purchase_timestamp    order_approved_at  \
0    delivered      2017-10-02 10:56:33  2017-10-02 11:07:15
1    delivered      2018-07-24 20:41:37  2018-07-26 03:24:27

  order_delivered_carrier_date order_delivered_customer_date  \
0          2017-10-04 19:55:00           2017-10-10 21:25:13
1          2018-07-26 14:31:00           2018-08-07 15:27:45

  order_estimated_delivery_date
0           2017-10-18 00:00:00
1           2018-08-13 00:00:00


First 2 rows:
                         product_id product_category_name  \
0  1e9e8ef04dbcff4541ed26657ea517e5            perfumaria
1  3aa071139cb16b67ca9e5dea641aaa2f                 artes

   product_name_lenght  product_description_lenght  product_photos_qty
\
0                 40.0                       287.0                 1.0

1                 44.0                       276.0                 1.0


   product_weight_g  product_length_cm  product_height_cm
product_width_cm
0             225.0               16.0               10.0
14.0
1            1000.0               30.0               18.0
20.0


First 2 rows:
                          seller_id  seller_zip_code_prefix
seller_city  \
0  3442f8959a84dea7ee197c632cb2df15                   13023
campinas
1  d1b65fc7debc3361ea86b5f14c68d2e2                   13844  mogi
guacu

  seller_state
0           SP
1           SP


First 2 rows:
    product_category_name product_category_name_english
0            beleza_saude                 health_beauty
1  informatica_acessorios         computers_accessories
```

```python
for name, df in dfs.items():
    print(f"Data Types:\n{df.dtypes}")
```

```
Data Types:
customer_id                 object
customer_unique_id          object
customer_zip_code_prefix     int64
customer_city               object
customer_state              object
dtype: object
Data Types:
geolocation_zip_code_prefix      int64
geolocation_lat                float64
geolocation_lng                float64
geolocation_city                object
geolocation_state               object
dtype: object
Data Types:
order_id                object
order_item_id            int64
product_id              object
seller_id               object
shipping_limit_date     object
price                  float64
freight_value          float64
dtype: object
Data Types:
order_id                object
order_item_id            int64
product_id              object
seller_id               object
shipping_limit_date     object
price                  float64
```

```
freight_value          float64
dtype: object
Data Types:
review_id                  object
order_id                   object
review_score                int64
review_comment_title       object
review_comment_message     object
review_creation_date       object
review_answer_timestamp    object
dtype: object
Data Types:
order_id                         object
customer_id                      object
order_status                     object
order_purchase_timestamp         object
order_approved_at                object
order_delivered_carrier_date     object
order_delivered_customer_date    object
order_estimated_delivery_date    object
dtype: object
Data Types:
product_id                     object
product_category_name          object
product_name_lenght           float64
product_description_lenght    float64
product_photos_qty            float64
product_weight_g              float64
product_length_cm             float64
product_height_cm             float64
product_width_cm              float64
dtype: object
Data Types:
seller_id                 object
seller_zip_code_prefix     int64
seller_city               object
seller_state              object
dtype: object
Data Types:
product_category_name            object
product_category_name_english    object
dtype: object

# Total payments per order using GROUP BY
query = '''
SELECT order_id, SUM(payment_value) AS total_payment
FROM order_payments
GROUP BY order_id
ORDER BY total_payment DESC
LIMIT 10
```

```
'''
pd.read_sql_query(query, conn)
```

```
                           order_id  total_payment
0  03caa2c082116e1d31e67e9ae3700499       13664.08
1  736e1922ae60d0d6a89247b851902527        7274.88
2  0812eb902a67711a1cb742b3cdaa65ae        6929.31
3  fefacc66af859508bf1a7934eab1e97f        6922.21
4  f5136e38d1a14a4dbd87dff67da82701        6726.66
5  2cc9089445046817a7539d90805e6e5a        6081.54
6  a96610ab360d42a2e5335a3998b4718a        4950.34
7  b4c4b76c642808cbe472a32b86cddc95        4809.44
8  199af31afc78c699f0dbf71fb178d4d4        4764.34
9  8dbc85d1447242f3b127dda390d56e19        4681.78
```

```
query = '''
SELECT p.product_id, ct.product_category_name_english
FROM products p
INNER JOIN category_translation ct
ON p.product_category_name = ct.product_category_name
LIMIT 10
'''
pd.read_sql_query(query, conn)
```

```
                         product_id product_category_name_english
0  1e9e8ef04dbcff4541ed26657ea517e5                     perfumery
1  3aa071139cb16b67ca9e5dea641aaa2f                           art
2  96bd76ec8810374ed1b65e291975717f                sports_leisure
3  cef67bcfe19066a932b7673e239eb23d                          baby
4  9dc1a7de274444849c219cff195d0b71                    housewares
5  41d3672d4792049fa1779bb35283ed13           musical_instruments
6  732bd381ad09e530fe0a5f457d81becb                    cool_stuff
7  2548af3e6e77a690cf3eb6368e9ab61e               furniture_decor
8  37cc742be07708b53a98702e77a21a02               home_appliances
9  8c92109888e8cdf9d66dc7e463025574                          toys
```

```
query = '''
SELECT p.product_id, p.product_category_name,
ct.product_category_name_english
FROM products p
LEFT JOIN category_translation ct
ON p.product_category_name = ct.product_category_name
LIMIT 10
'''
pd.read_sql_query(query, conn)
```

```
                         product_id  product_category_name  \
0  1e9e8ef04dbcff4541ed26657ea517e5             perfumaria
1  3aa071139cb16b67ca9e5dea641aaa2f                  artes
2  96bd76ec8810374ed1b65e291975717f          esporte_lazer
```

```
3  cef67bcfe19066a932b7673e239eb23d                  bebes
4  9dc1a7de274444849c219cff195d0b71  utilidades_domesticas
5  41d3672d4792049fa1779bb35283ed13  instrumentos_musicais
6  732bd381ad09e530fe0a5f457d81becb             cool_stuff
7  2548af3e6e77a690cf3eb6368e9ab61e        moveis_decoracao
8  37cc742be07708b53a98702e77a21a02        eletrodomesticos
9  8c92109888e8cdf9d66dc7e463025574             brinquedos

  product_category_name_english
0                     perfumery
1                           art
2                sports_leisure
3                          baby
4                    housewares
5           musical_instruments
6                    cool_stuff
7               furniture_decor
8               home_appliances
9                          toys
```

```
query = '''
SELECT product_id, COUNT(*) AS review_count
FROM order_items
WHERE order_id IN (
    SELECT order_id FROM order_reviews
)
GROUP BY product_id
ORDER BY review_count DESC
LIMIT 5
'''
pd.read_sql_query(query, conn)
```

```
                         product_id  review_count
0  aca2eb7d00ea1a7b8ebd4e68314663af           524
1  422879e10f46682990de24d770e7f83d           483
2  99a4788cb24856965c36a24e339b6058           479
3  389d119b48cf3043d311335e499d9c6b           391
4  368c6c730842d78016ad823897a372db           385
```

```
query = '''
SELECT seller_id, AVG(price) AS avg_price, SUM(freight_value) AS
total_freight
FROM order_items
GROUP BY seller_id
ORDER BY avg_price DESC
LIMIT 5
'''
pd.read_sql_query(query, conn)
```

```
                          seller_id    avg_price  total_freight
0  80ceebb4ee9b31afb6c6a916a574a1e2  6729.000000         193.21
1  ee27a8f15b1dded4d213a468ba4eb391  6499.000000         227.66
2  585175ec331ea177fa47199e39a6170a  3549.000000          53.47
3  abe021b01ba992245271b9aa422032df  3360.000000         117.24
4  a00824eb9093d40e589b940ec45c4eb0  3133.323333         379.49
```

```
conn.execute('''
CREATE VIEW IF NOT EXISTS product_revenue AS
SELECT product_id, SUM(price) AS revenue
FROM order_items
GROUP BY product_id
''')

pd.read_sql_query('SELECT * FROM product_revenue ORDER BY revenue DESC
LIMIT 5', conn)
```

```
                         product_id   revenue
0  bb50f2e236e5eea0100680137654686c  63885.00
1  6cdd53843498f92890544667809f1595  54730.20
2  d6160fb7873f184099d9bc95e30376af  48899.34
3  d1c427060a0f73f6b889a5c7c61f2ac4  47214.51
4  99a4788cb24856965c36a24e339b6058  43025.56
```

```
conn.execute('CREATE INDEX IF NOT EXISTS idx_order_id ON
order_items(order_id)')
conn.execute('CREATE INDEX IF NOT EXISTS idx_seller_id ON
order_items(seller_id)')

# Query to benefit from index
query = '''
SELECT order_id, COUNT(*) AS item_count
FROM order_items
GROUP BY order_id
ORDER BY item_count DESC
LIMIT 5
'''
pd.read_sql_query(query, conn)
```

```
                           order_id  item_count
0  8272b63d03f5f79c56e9e4120aec44ef          21
1  1b15974a0141d54e36626dca3fdc731a          20
2  ab14fdcfbe524636d65ee38360e22ce8          20
3  428a2f660dc84138d969ccd69a0ab6d5          15
4  9ef13efd6949e4573a18964dd1bbe7f5          15
```