

Task-4

1. Create Database

```
CREATE DATABASE olist;  
USE olist;
```

2. Create Tables

-- Customers

```
CREATE TABLE customers (  
    customer_id VARCHAR(50) PRIMARY KEY,  
    customer_unique_id VARCHAR(50),  
    customer_zip_code_prefix INT,  
    customer_city VARCHAR(100),  
    customer_state CHAR(2)  
);
```

-- Orders

```
CREATE TABLE orders (  
    order_id VARCHAR(50) PRIMARY KEY,  
    customer_id VARCHAR(50),  
    order_status VARCHAR(20),  
    order_purchase_timestamp DATETIME,  
    order_approved_at DATETIME,  
    order_delivered_carrier_date DATETIME,  
    order_delivered_customer_date DATETIME,  
    order_estimated_delivery_date DATETIME,  
    FOREIGN KEY (customer_id) REFERENCES customers(customer_id)  
);
```

-- Products

```
CREATE TABLE products (  
    product_id VARCHAR(50) PRIMARY KEY,  
    product_category_name VARCHAR(100),  
    product_name_length INT,  
    product_description_length INT,  
    product_photos_qty INT,  
    product_weight_g INT,  
    product_length_cm INT,  
    product_height_cm INT,  
    product_width_cm INT  
);
```

-- Sellers

```
CREATE TABLE sellers (  
    seller_id VARCHAR(50) PRIMARY KEY,  
    seller_zip_code_prefix INT,  
    seller_city VARCHAR(100),  
    seller_state CHAR(2)  
);
```

-- Order Items

```
CREATE TABLE order_items (  
  order_id VARCHAR(50),  
  order_item_id INT,  
  product_id VARCHAR(50),  
  seller_id VARCHAR(50),  
  shipping_limit_date DATETIME,  
  price DECIMAL(10,2),  
  freight_value DECIMAL(10,2),  
  PRIMARY KEY (order_id, order_item_id),  
  FOREIGN KEY (order_id) REFERENCES orders(order_id),  
  FOREIGN KEY (product_id) REFERENCES products(product_id),  
  FOREIGN KEY (seller_id) REFERENCES sellers(seller_id)  
);
```

-- Order Payments

```
CREATE TABLE order_payments (  
  order_id VARCHAR(50),  
  payment_sequential INT,  
  payment_type VARCHAR(20),  
  payment_installments INT,  
  payment_value DECIMAL(10,2),  
  PRIMARY KEY (order_id, payment_sequential),  
  FOREIGN KEY (order_id) REFERENCES orders(order_id)  
);
```

-- Order Reviews

```
CREATE TABLE order_reviews (  
  review_id VARCHAR(50) PRIMARY KEY,  
  order_id VARCHAR(50),  
  review_score INT,  
  review_comment_title TEXT,  
  review_comment_message TEXT,  
  review_creation_date DATETIME,  
  review_answer_timestamp DATETIME,  
  FOREIGN KEY (order_id) REFERENCES orders(order_id)  
);
```

-- Geolocation

```
CREATE TABLE geolocation (  
  geolocation_zip_code_prefix INT,  
  geolocation_lat DECIMAL(10,8),  
  geolocation_lng DECIMAL(10,8),  
  geolocation_city VARCHAR(100),  
  geolocation_state CHAR(2)  
);
```

-- Product Category Translation

```
CREATE TABLE product_category_name_translation (  
    product_category_name VARCHAR(100),  
    product_category_name_english VARCHAR(100)  
);
```

3. Load Data from CSV (put CSV files inside secure_file_priv path)

-- Customers

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server  
8.0/Uploads/olist_customers_dataset.csv'  
INTO TABLE customers  
FIELDS TERMINATED BY ','  
ENCLOSED BY ''''  
LINES TERMINATED BY '\n'  
IGNORE 1 ROWS  
(customer_id, customer_unique_id, customer_zip_code_prefix, customer_city, customer_state);
```

-- Orders

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server  
8.0/Uploads/olist_orders_dataset.csv'  
INTO TABLE orders  
FIELDS TERMINATED BY ','  
ENCLOSED BY ''''  
LINES TERMINATED BY '\n'  
IGNORE 1 ROWS  
(order_id, customer_id, order_status, @order_purchase_timestamp, @order_approved_at,  
@order_delivered_carrier_date, @order_delivered_customer_date,  
@order_estimated_delivery_date)  
SET  
order_purchase_timestamp = NULLIF(@order_purchase_timestamp,""),  
order_approved_at = NULLIF(@order_approved_at,""),  
order_delivered_carrier_date = NULLIF(@order_delivered_carrier_date,""),  
order_delivered_customer_date = NULLIF(@order_delivered_customer_date,""),  
order_estimated_delivery_date = NULLIF(@order_estimated_delivery_date,"");
```

-- Products

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server  
8.0/Uploads/olist_products_dataset.csv'  
INTO TABLE products  
FIELDS TERMINATED BY ','  
ENCLOSED BY ''''  
LINES TERMINATED BY '\n'  
IGNORE 1 ROWS  
(product_id, product_category_name, @product_name_length, @product_description_length,  
@product_photos_qty, @product_weight_g, @product_length_cm, @product_height_cm,  
@product_width_cm)  
SET  
product_name_length = NULLIF(@product_name_length,""),  
product_description_length = NULLIF(@product_description_length,""),  
product_photos_qty = NULLIF(@product_photos_qty,""),
```

```
product_weight_g = NULLIF(@product_weight_g,"),
product_length_cm = NULLIF(@product_length_cm,"),
product_height_cm = NULLIF(@product_height_cm,"),
product_width_cm = NULLIF(@product_width_cm,");
```

-- Sellers

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server
8.0/Uploads/olist_sellers_dataset.csv'
INTO TABLE sellers
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS
(seller_id, seller_zip_code_prefix, seller_city, seller_state);
```

-- Order Items

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server
8.0/Uploads/olist_order_items_dataset.csv'
INTO TABLE order_items
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS
(order_id, order_item_id, product_id, seller_id, @shipping_limit_date, @price, @freight_value)
SET
shipping_limit_date = NULLIF(@shipping_limit_date,"),
price = NULLIF(@price,"),
freight_value = NULLIF(@freight_value,");
```

-- Order Payments

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server
8.0/Uploads/olist_order_payments_dataset.csv'
INTO TABLE order_payments
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS
(order_id, payment_sequential, payment_type, payment_installments, payment_value);
```

-- Order Reviews

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server
8.0/Uploads/olist_order_reviews_dataset.csv'
IGNORE INTO TABLE order_reviews
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS
(review_id, order_id, review_score, review_comment_title, review_comment_message,
@review_creation_date, @review_answer_timestamp)
SET
```

```
review_creation_date = NULLIF(@review_creation_date,""),  
review_answer_timestamp = NULLIF(@review_answer_timestamp,"");
```

-- Geolocation

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server  
8.0/Uploads/olist_geolocation_dataset.csv'  
INTO TABLE geolocation  
FIELDS TERMINATED BY ','  
ENCLOSED BY ''''  
LINES TERMINATED BY '\n'  
IGNORE 1 ROWS  
(geolocation_zip_code_prefix, geolocation_lat, geolocation_lng, geolocation_city,  
geolocation_state);
```

-- Product Category Translation

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server  
8.0/Uploads/product_category_name_translation.csv'  
INTO TABLE product_category_name_translation  
FIELDS TERMINATED BY ','  
ENCLOSED BY ''''  
LINES TERMINATED BY '\n'  
IGNORE 1 ROWS  
(product_category_name, product_category_name_english);
```

List all customers from São Paulo:

```
SELECT * FROM customers WHERE  
customer_city = 'sao paulo';
```

	customer_id	customer_unique_id	customer_zip_code_prefix	customer_city	customer_state
▶	000379cdec625522490c315e70c7a9fb	0b83f73b19c2019e182fd552c048a22c	4841	sao paulo	SP
	0005aefbb696d34b3424dccc0a0e9fd0	616309b2eeb7bd9c05b0fdbab28e6c6	3052	sao paulo	SP
	00062b33cb9f6fe976afdcff967ea74d	f90f55ee274a4ae21510b386134b09cd	2306	sao paulo	SP
	001028b78fd413e19704b3867c369d3a	e57a935f49ffd73b93c18eaaa04efa84	5387	sao paulo	SP
	001051abfcfbed9f87b4266213a5df1	4ea5df5937187bd3176aee08d4782104	2251	sao paulo	SP

Find top 10 orders by purchase date:

```
SELECT order_id, order_purchase_timestamp  
FROM orders  
ORDER BY order_purchase_timestamp DESC  
LIMIT 10;
```

	order_id	order_purchase_timestamp
▶	10a045cdf6a5650c21e9cfefb60384c16	2018-10-17 17:30:18
	b059ee4de278302d550a3035c4cdb740	2018-10-16 20:16:02
	a2ac6dad85cf8af5b0afb510a240fe8c	2018-10-03 18:55:29
	616fa7d4871b87832197b2a137a115d2	2018-10-01 15:30:09
	392ed9afd714e3c74767d0c4d3e3f477	2018-09-29 09:13:03
	869997f9e01f39d184956b5c6bccfdbe	2018-09-26 08:40:15
	5aac76cf7b07dd06fa4d50bf461d2f40	2018-09-26 08:40:15
	ed3efbd3a87bea76c2812c66a0b32219	2018-09-20 13:54:16
	bd35b677fd239386e9861d11ae98ab56	2018-09-17 17:21:16
	ea844c92cf978ea23321fa7fe5871761	2018-09-13 09:56:12
*	NULL	NULL

Count orders by status:

```
SELECT order_status, COUNT(*) AS total_orders  
FROM orders  
GROUP BY order_status;
```

	order_status	total_orders
▶	delivered	96478
	unavailable	609
	shipped	1107
	canceled	625
	invoiced	314
	processing	301
	approved	2
	created	5

Average delivery time (days):

```
SELECT
AVG(DATEDIFF(order_delivered_customer_date,
order_purchase_timestamp)) AS avg_delivery_days
FROM orders
WHERE order_delivered_customer_date IS NOT
NULL;
```

	avg_delivery_days
▶	12.4973

Aggregation & GROUP BY

Top 5 cities by number of customers:

```
SELECT customer_city, COUNT(*) AS
total_customers
FROM customers
GROUP BY customer_city
ORDER BY total_customers DESC
LIMIT 5;
```

	customer_city	total_customers
▶	sao paulo	15540
	rio de janeiro	6882
	belo horizonte	2773
	brasilgia	2131
	curitiba	1521

Advanced Aggregation

Total revenue by product category:

```
SELECT p.product_category_name, SUM(oi.price)
AS revenue
FROM products p
JOIN order_items oi ON p.product_id =
oi.product_id
GROUP BY p.product_category_name
ORDER BY revenue DESC;
```

	product_category_name	revenue
▶	beleza_saude	1258681.34
	relogios_presentes	1205005.68
	cama_mesa_banho	1036988.68
	esporte_lazer	988048.97
	informatica_acessorios	911954.32
	moveis_decoracao	729762.49
	cool_stuff	635290.85
	utilidades_domesticas	632248.66
	automotivo	592720.11
	ferramentas_jardim	485256.46
	brinquedos	483946.60
	bebes	411764.89
	perfumaria	300124.87

Top 5 orders with highest freight value:

```
SELECT order_id, SUM(freight_value) AS
total_freight
FROM order_items
GROUP BY order_id
ORDER BY total_freight DESC
LIMIT 5;
```

Result Grid	Filter Rows:	Expe
order_id	total_freight	
▶ cf4659487be50c0c317cff3564c4a840	1794.96	
2455cbeb73fd04b170ca2504662f95ce	1002.29	
cfed507ac357129f750f05a0d7d71b15	711.33	
71dab1155600756af6de79de92e712e3	626.64	
17784b9fbb37fb0bdc230d8ed6f6b355	502.98	

Indexes (Optimization)

Create index on orders date:

```
CREATE INDEX idx_orders_purchase_date ON
orders(order_purchase_timestamp);
```


More Queries

Count of products per seller:

```
SELECT seller_id, COUNT(DISTINCT product_id)
AS total_products
FROM order_items
GROUP BY seller_id
ORDER BY total_products DESC
LIMIT 10;
```

	seller_id	total_products
▶	4a3ca9315b744ce9f8e9374361493884	399
	cca3071e3e9bb7d12640c9fbe2301306	322
	d91fb3b7d041e83b64a00a3edfb37e4f	315
	fa1c13f2614d7b5c4749cbc52fecda94	289
	7142540dd4c91e2237acb7e911c4eba2	266
	6560211a19b47992c3666cc44a7e94c0	256
	da8622b14eb17ae2831f4ac5b9dab84a	222
	ea8482cd71df3c1969d7b9473ff13abc	220
	3d871de0142ce09b7081e2b9d1733cb1	202
	7c67e1448b00f6e969d365cea6b010ab	198

Orders delayed beyond estimated delivery:

```
SELECT order_id, order_estimated_delivery_date,
order_delivered_customer_date
FROM orders
WHERE order_delivered_customer_date >
order_estimated_delivery_date;
```

	order_id	order_estimated_delivery_date	order_delivered_customer_date
▶	0005a1a1728c9d785b8e2b08b904576c	2018-03-29 00:00:00	2018-03-29 18:17:31
	00063b381e2406b52ad429470734ebd5	2018-08-07 00:00:00	2018-08-07 13:56:52
	000e906b789b55f64edcb1f84030f90d	2017-12-07 00:00:00	2017-12-09 17:27:23
	0017afd5076e074a48f1f1a4c7bac9c5	2017-05-19 00:00:00	2017-05-23 08:32:07
	001c85b5f68d2be0cb0797afc9e8ce9a	2017-12-14 00:00:00	2017-12-22 18:37:40
	001d8f0e34a38c37f7dba2a37d4eba8b	2017-05-24 00:00:00	2017-05-26 13:14:50
	0030d783f979fbc5981e75613b057344	2017-12-15 00:00:00	2018-01-06 00:25:43
	00324b3eda39ba5ecce3945823e3594c	2018-04-30 00:00:00	2018-05-03 20:29:24
	0032d07457ae9c806c79368d7d9ce96b	2018-04-17 00:00:00	2018-04-29 21:08:59
	00335b686d693c7d72deeb12f8e89227	2017-08-11 00:00:00	2017-09-12 20:58:45
	00378c6c981f234634c0b9d6128df6dd	2018-02-26 00:00:00	2018-02-26 21:36:00
	003a94f778ef8cfd50247c8c1b582257	2018-08-15 00:00:00	2018-08-24 00:18:38
	003d804ee0e1b856881cd18e0cc0d4c	2017-12-18 00:00:00	2018-01-10 23:19:43