

Task 3: SQL for Data Analysis

1) Show first 10 customers.

```
1 SELECT * FROM olist_customers_dataset
2 LIMIT 10;
```

	customer_id	customer_unique_id	customer_zip_code_prefix	customer_city	customer_state
1	06b8999e2fba1a1fbc88172c00ba8bc7	861eff4711a542e4b93843c6dd7febb0	14409	franca	SP
2	18955e83d337fd6b2def6b18a428ac77	290c77bc529b7ac935b93aa66c333dc3	9790	sao bernardo do campo	SP
3	4e7b3e00288586ebd08712fdd0374a03	060e732b5b29e8181a18229c7b0b2b5e	1151	sao paulo	SP
4	b2b6027bc5c5109e529d4dc6358b12c3	259dac757896d24d7702b9acbbff3f3c	8775	mogi das cruzeiras	SP
5	4f2d8ab171c80ec8364f7c12e35b23ad	345ecd01c38d18a9036ed96c73b8d066	13056	campinas	SP
6	879864dab9bc3047522c92c82e1212b8	4c93744516667ad3b8f1fb645a3116a4	89254	jaragua do sul	SC
7	fd826e7cf63160e536e0908c76c3f441	addec96d2e059c80c30fe6871d30d177	4534	sao paulo	SP
8	5e274e7a0c3809e14aba7ad5aae0d407	57b2a98a409812fe9618067b6b8ebe4f	35182	timoteo	MG
9	5adf08e34b2e993982a47070956c5c65	1175e95fb47ddff9de6b2b06188f7e0d	81560	curitiba	PR
10	4b7139f34592b3a31687243a302fa75b	9afe194fb833f79e300e37e580171f22	30575	belo horizonte	MG

Execution finished without errors.
Result: 10 rows returned in 30ms
At line 1:
SELECT * FROM olist_customers_dataset
LIMIT 10;

2) Show customer_id and customer_city (first 10 rows).

SQL 1*

```
1 SELECT customer_id, customer_city FROM olist_customers_dataset LIMIT 10;
```

	customer_id	customer_city
1	06b8999e2fba1a1fbc88172c00ba8bc7	franca
2	18955e83d337fd6b2def6b18a428ac77	sao bernardo do campo
3	4e7b3e00288586ebd08712fdd0374a03	sao paulo
4	b2b6027bc5c5109e529d4dc6358b12c3	mogi das cruzeiras
5	4f2d8ab171c80ec8364f7c12e35b23ad	campinas
6	879864dab9bc3047522c92c82e1212b8	jaragua do sul
7	fd826e7cf63160e536e0908c76c3f441	sao paulo
8	5e274e7a0c3809e14aba7ad5aae0d407	timoteo
9	5adf08e34b2e993982a47070956c5c65	curitiba
10	4b7139f34592b3a31687243a302fa75b	belo horizonte

Execution finished without errors.
Result: 10 rows returned in 28ms
At line 1:
--SELECT * FROM olist_customers_dataset
--LIMIT 10;
SELECT customer_id, customer_city FROM olist_customers_dataset LIMIT 10;

3) Customers in Sao Paulo.

SQL 1*				
1	SELECT customer_id, customer_unique_id, customer_city, customer_state			
2	FROM olist_customers_dataset			
3	WHERE customer_city = 'sao paulo';			

	customer_id	customer_unique_id	customer_city	customer_state
1	4e7b3e00288586ebd08712fdd0374a03	060e732b5b29e8181a18229c7b0b2b5e	sao paulo	SP
2	fd826e7cf63160e536e0908c76c3f441	addec96d2e059c80c30fe6871d30d177	sao paulo	SP
3	eabebad39a88bb6f5b52376faec28612	295c05e81917928d76245e842748184d	sao paulo	SP
4	9b8ce803689b3562defaad4613ef426f	7f3a72e8f988c6e735ba118d54f47458	sao paulo	SP
5	2938121a40a20953c43caa8c98787fcb	482441ea6a06b1f72fe9784756c0ea75	sao paulo	SP
6	cb721d7b4f271fd87011c4c83462c076	a5844ba4bfc8d0cc61d13027c7e63bcc	sao paulo	SP
7	6e359a57a91f84095cc64e1b351aef8c	2e6a42a9b5cbb0da62988694f18ee295	sao paulo	SP
8	e3109970a3fe8021d5ff82c577ce5606	a8654e2af5da6bb72f52c22b164855e1	sao paulo	SP
9	a3b0fda37bae14cf754877bed475e80c	c9158d089637ab443c78984d20da7fc0	sao paulo	SP
10	d3b6830d18c7de943d1e707d1f061d40	27cf4b153010911a0957150255a6c6db	sao paulo	SP
11	a9b0d1c26105279e1b8edc63d06bd668	3d49f4455a3947c8dd7e972b3ad8cb76	sao paulo	SP
12	03f846ad03437d864a8d2a22976dcafe	7677c213007e9a6ec9267ea50b5ce5bc	sao paulo	SP
13	0aae2862f8eac77f10a34f44860720ac	cd076285a12f40041b32f5ad8c98699f	sao paulo	SP
14	6c9a5923526346cbc0bd7bbd92269c01	cf6d4152d758efc43910e0141ae5b912	sao paulo	SP
15	8264e3518163dd09211870b24a5d741d	67d21c8bea9d6017d1b124d3879dd815	sao paulo	SP
16	38d1cd89306128348ffdf4cc23f3a50a	d491a65a6ef3c04e145d37395996bad7	sao paulo	SP

Execution finished without errors.
 Result: 15540 rows returned in 70ms
 At line 1:
 SELECT customer_id, customer_unique_id, customer_city, customer_state
 FROM olist_customers_dataset
 WHERE customer_city = 'sao paulo';

4) Customers ordered by city (ascending).

SQL 1*				
1	SELECT customer_id, customer_unique_id, customer_city, customer_state			
2	FROM olist_customers_dataset			
3	ORDER BY customer_city ASC;			

	customer_id	customer_unique_id	customer_city	customer_state
1	f11eb8f0b8...	64ee476500...	abadia dos...	MG
2	a23e3f9a2b...	afddf43a03...	abadia dos...	MG
3	9e01f714a2...	e1feae9083...	abadia dos...	MG
4	576d71ddb2...	8d76b55918...	abadiania	GO
5	5e9e1ae42e...	28af9604f7...	abaete	MG
6	ff0d62f8be...	ab26557b28...	abaete	MG
7	962b41ee52...	8b8545eb71...	abaete	MG
8	d47c8bb51d...	189af2582a...	abaete	MG
9	2bf14bf7a1...	1605eaf574...	abaete	MG
10	24d97ea7bb...	5bd046eccd...	abaete	MG
11	a3d3c38e58...	5a4fa4919c...	abaete	MG
12	e67618ef41...	a0e363eb99...	abaete	MG
13	0852882426...	8f7f547488...	abaete	MG
14	225634426c...	cala0307fe...	abaete	MG
15	3c46a6771b...	2a19a29632...	abaete	MG
16	5371894984...	2d69de392f...	abaete	MG

Execution finished without errors.
 Result: 99441 rows returned in 360ms
 At line 1:
 SELECT customer_id, customer_unique_id, customer_city, customer_state
 FROM olist_customers_dataset
 ORDER BY customer_city ASC;

5) Total customers per state.

SQL 1*		
1	SELECT customer_state,COUNT(customer_id) AS total_customers	
2	FROM olist_customers_dataset	
3	GROUP BY customer_state;	
	customer_state	total_customers
1	AC	81
2	AL	413
3	AM	148
4	AP	68
5	BA	3380
6	CE	1336
7	DF	2140
8	ES	2033
9	GO	2020
10	MA	747
11	MG	11635
12	MS	715
13	MT	907
14	PA	975
15	PB	536
16	PE	1652
Execution finished without errors. Result: 27 rows returned in 80ms At line 1: SELECT customer_state,COUNT(customer_id) AS total_customers FROM olist_customers_dataset GROUP BY customer_state;		

6) Total customers per state.

SQL 1*	
1	SELECT customer_unique_id
2	FROM olist_customers_dataset
3	WHERE customer_id IN (
4	SELECT customer_id
5	FROM olist_customers_dataset
6	GROUP BY customer_id
7	HAVING COUNT(*) > 1);
customer_unique_id	
Execution finished without errors. Result: 0 rows returned in 106ms At line 1: SELECT customer_unique_id FROM olist_customers_dataset WHERE customer_id IN (SELECT customer_id FROM olist_customers_dataset GROUP BY customer_id	

7) Total number of customers.

```
1 SELECT COUNT(customer_id) AS total_customers
2 FROM olist_customers_dataset;
3
```

	total_customers
1	99441

Execution finished without errors.
Result: 1 rows returned in 28ms
At line 1:
SELECT COUNT(customer_id) AS total_customers
FROM olist_customers_dataset;

8) Total unique customers.

```
SQL 1*
1 SELECT COUNT(DISTINCT customer_unique_id) AS unique_customers
2 FROM olist_customers_dataset;
```

	unique_customers
1	96096

Execution finished without errors.
Result: 1 rows returned in 242ms
At line 1:
SELECT COUNT(DISTINCT customer_unique_id) AS unique_customers
FROM olist_customers_dataset;

9) Total customers per state (descending order).

SQL 1*		
1	SELECT customer_state, COUNT(customer_id) AS total_customers	
2	FROM olist_customers_dataset	
3	GROUP BY customer_state	
4	ORDER BY total_customers DESC;	

	customer_state	total_customers
1	SP	41746
2	RJ	12852
3	MG	11635
4	RS	5466
5	PR	5045
6	SC	3637
7	BA	3380
8	DF	2140
9	ES	2033
10	GO	2020
11	PE	1652
12	CE	1336
13	PA	975
14	MT	907

Execution finished without errors.
Result: 27 rows returned in 89ms
At line 1:
SELECT customer_state, COUNT(customer_id) AS total_customers
FROM olist_customers_dataset
GROUP BY customer_state
ORDER BY total_customers DESC;

10) Total customers per city (descending order).

SQL 1*		
1	SELECT customer_city, COUNT(customer_id) AS total_customers	
2	FROM olist_customers_dataset	
3	GROUP BY customer_city	
4	ORDER BY total_customers DESC;	

	customer_city	total_customers
1	sao paulo	15540
2	rio de janeiro	6882
3	belo horizonte	2773
4	brasilia	2131
5	curitiba	1521
6	campinas	1444
7	porto alegre	1379
8	salvador	1245
9	guarulhos	1189
10	sao bernardo do campo	938
11	niteroi	849
12	santo andre	797
13	osasco	746
14	santos	713

Execution finished without errors.
Result: 4119 rows returned in 104ms
At line 1:
SELECT customer_city, COUNT(customer_id) AS total_customers
FROM olist_customers_dataset
GROUP BY customer_city
ORDER BY total_customers DESC;

11) Self join: Customers living in the same city.

SQL 1*			
1	SELECT a.customer_id AS customer1, b.customer_id AS customer2, a.customer_city		
2	FROM olist_customers_dataset AS a		
3	INNER JOIN olist_customers_dataset AS b		
4	ON a.customer_city = b.customer_city		
5	WHERE a.customer_id <> b.customer_id;		
	customer1	customer2	customer_city
1	06b8999e2f...	01077b7579...	franca
2	06b8999e2f...	019b7076ac...	franca
3	06b8999e2f...	01edba0c13...	franca
4	06b8999e2f...	02611d7a9e...	franca
5	06b8999e2f...	0273ad500c...	franca
6	06b8999e2f...	0445283ab2...	franca
7	06b8999e2f...	0489975a32...	franca
8	06b8999e2f...	059571e029...	franca
9	06b8999e2f...	078e990360...	franca
10	06b8999e2f...	08134f392c...	franca
11	06b8999e2f...	08f33b15b6...	franca
12	06b8999e2f...	0c51809f15...	franca
13	06b8999e2f...	0dcee6eaf8...	franca
14	06b8999e2f...	139490b244...	franca
15	06b8999e2f...	13c8337f68...	franca
16	06b8999e2f...	13d2f11be3...	franca
17	06b8999e2f...	13f048312e...	franca
18	06b8999e2f...	155716e7d7...	franca
19	06b8999e2f...	176f7fae9b...	franca

12) Query from the view: state-wise customer count.

SQL 1*		
1	SELECT * FROM view_state_customers;	
	customer_state	total_customers
1	AC	81
2	AL	413
3	AM	148
4	AP	68
5	BA	3380
6	CE	1336
7	DF	2140
8	ES	2033
9	GO	2020
10	MA	747
11	MG	11635

Execution finished without errors.
Result: 27 rows returned in 91ms
At line 2:
SELECT * FROM view_state_customers;