

golden_100_text_to_sql

id	category	natural_language	sql_query
1	Aggregation	How many customers are there in each state?	SELECT customer_state, COUNT(customer_id) AS customer_count FROM customers GROUP BY customer_state ORDER BY customer_count DESC;
2	Aggregation	What are the top 10 cities with the most customers?	SELECT customer_city, COUNT(customer_id) AS customer_count FROM customers GROUP BY customer_city ORDER BY customer_count DESC LIMIT 10;
3	Basic Selection	Show me the details of a specific customer with the unique ID '861eff4711a542e4b93843c6dd7febb0'.	SELECT * FROM customers WHERE customer_unique_id = '861eff4711a542e4b93843c6dd7febb0';
4	Basic Selection	List all orders placed by the customer with ID '06b8999e2fba1a1fbc88172c00ba8bc7'.	SELECT * FROM orders WHERE customer_id = '06b8999e2fba1a1fbc88172c00ba8bc7';
5	Aggregation	How many orders have been placed each month?	SELECT STRFTIME('%Y-%m', order_purchase_timestamp) AS month, COUNT(order_id) AS order_count FROM orders GROUP BY month ORDER BY month;
6	Aggregation	What is the current status of all orders?	SELECT order_status, COUNT(order_id) AS order_count FROM orders GROUP BY order_status;
7	Filtering	Find all orders that were delivered later than the estimated delivery date.	SELECT order_id, order_estimated_delivery_date, order_delivered_customer_date FROM orders WHERE order_delivered_customer_date > order_estimated_delivery_date;
8	Aggregation	What is the average time between order purchase and approval?	SELECT AVG(JULIANDAY(order_approved_at) - JULIANDAY(order_purchase_timestamp)) AS avg_approval_time_days FROM orders;
9	Joins	Show me the full details of all orders, including customer information.	SELECT o.*, c.customer_unique_id, c.customer_city, c.customer_state FROM orders o JOIN customers c ON o.customer_id = c.customer_id;
10	Joins	Which customers have placed more than 5 orders?	SELECT c.customer_unique_id, COUNT(o.order_id) AS order_count FROM customers c JOIN orders o ON c.customer_id = o.customer_id GROUP BY c.customer_unique_id HAVING order_count > 5;
11	Aggregation	How many products are there in each category?	SELECT product_category_name, COUNT(product_id) AS product_count FROM products GROUP BY product_category_name ORDER BY product_count DESC;
12	Basic Selection	What is the English translation for all product category names?	SELECT * FROM product_category_name_translation;
13	Filtering	List the top 10 heaviest products.	SELECT product_id, product_category_name, product_weight_g FROM products ORDER BY product_weight_g DESC LIMIT 10;
14	Basic Selection	What are the dimensions (length, height, width) of the product with ID '1e9e8ef04dbcf4541ed26657ea517e5'?	SELECT product_length_cm, product_height_cm, product_width_cm FROM products WHERE product_id = '1e9e8ef04dbcf4541ed26657ea517e5';
15	Filtering	Find products with a description length greater than 1000 characters.	SELECT product_id, product_description_lenght FROM products WHERE product_description_lenght > 1000;
16	Aggregation	How many photos does each product have on average, grouped by category?	SELECT product_category_name, AVG(product_photos_qty) AS avg_photo_count FROM products GROUP BY product_category_name;
17	Basic Selection	Show me all products in the 'beleza_saude' category.	SELECT * FROM products WHERE product_category_name = 'beleza_saude';
18	Aggregation	What is the total number of products?	SELECT COUNT(product_id) FROM products;
19	Filtering	List all product categories that have the word 'livros' in their name.	SELECT DISTINCT product_category_name FROM products WHERE product_category_name LIKE '%livros%';
20	Joins	Get the product category name in English for each product.	SELECT p.product_id, t.product_category_name_english FROM products p JOIN product_category_name_translation t ON p.product_category_name = t.product_category_name;
21	Aggregation	How many sellers are there in each state?	SELECT seller_state, COUNT(seller_id) AS seller_count FROM sellers GROUP BY seller_state ORDER BY seller_count DESC;
22	Aggregation	What are the top 10 cities with the most sellers?	SELECT seller_city, COUNT(seller_id) AS seller_count FROM sellers GROUP BY seller_city ORDER BY seller_count DESC LIMIT 10;
23	Basic Selection	Show me the details of the seller with ID '3442f8959a84dea7ee197c632cb2df15'.	SELECT * FROM sellers WHERE seller_id = '3442f8959a84dea7ee197c632cb2df15';
24	Joins	Which sellers have sold the most items?	SELECT s.seller_id, s.seller_city, COUNT(o.order_item_id) AS items_sold FROM sellers s JOIN order_items oi ON s.seller_id = oi.seller_id GROUP BY s.seller_id, s.seller_city ORDER BY items_sold DESC LIMIT 10;
25	Joins	What is the total revenue generated by each seller?	SELECT s.seller_id, SUM(oi.price) AS total_revenue FROM sellers s JOIN order_items oi ON s.seller_id = oi.seller_id GROUP BY s.seller_id ORDER BY total_revenue DESC;
26	Filtering	Find sellers who have not sold any products yet.	SELECT s.seller_id FROM sellers s LEFT JOIN order_items oi ON s.seller_id = oi.seller_id WHERE oi.order_id IS NULL;
27	Aggregation	What is the average freight value for each seller?	SELECT seller_id, AVG(freight_value) AS avg_freight FROM order_items GROUP BY seller_id;
28	Basic Selection	List all sellers from the city of 'sao paulo'.	SELECT * FROM sellers WHERE seller_city = 'sao paulo';
29	Aggregation	How many sellers are there for each zip code prefix?	SELECT seller_zip_code_prefix, COUNT(seller_id) AS seller_count FROM sellers GROUP BY seller_zip_code_prefix;
30	Joins	Show the seller ID and the date of their last sale.	SELECT oi.seller_id, MAX(oi.order_purchase_timestamp) AS last_sale_date FROM order_items oi JOIN orders o ON oi.order_id = o.order_id GROUP BY oi.seller_id;
31	Aggregation	What is the total number of items sold for each order?	SELECT order_id, COUNT(order_item_id) AS item_count FROM order_items GROUP BY order_id;
32	Aggregation	What is the average price of items sold?	SELECT AVG(price) AS average_price FROM order_items;
33	Filtering	Show me the most expensive item sold.	SELECT * FROM order_items ORDER BY price DESC LIMIT 1;
34	Aggregation	How many items of each product have been sold?	SELECT product_id, COUNT(order_item_id) AS items_sold FROM order_items GROUP BY product_id ORDER BY items_sold DESC;
35	Aggregation	What is the total freight value for all orders?	SELECT SUM(freight_value) AS total_freight FROM order_items;
36	Basic Selection	List all order items for a specific order with ID 'e481f51cbdc54678b7cc49136f2d6af7'.	SELECT * FROM order_items WHERE order_id = 'e481f51cbdc54678b7cc49136f2d6af7';
37	Aggregation	What are the different payment types used and how frequent is each?	SELECT payment_type, COUNT(order_id) AS frequency FROM order_payments GROUP BY payment_type;
38	Aggregation	What is the average number of payment installments?	SELECT AVG(payment_installments) AS avg_installments FROM order_payments;
39	Aggregation	Show me the total payment value for each order.	SELECT order_id, SUM(payment_value) AS total_payment FROM order_payments GROUP BY order_id;
40	Filtering	Find orders paid with more than one payment method.	SELECT order_id, COUNT(payment_sequential) AS payment_methods FROM order_payments GROUP BY order_id HAVING payment_methods > 1;
41	Aggregation	What is the average review score for all orders?	SELECT AVG(review_score) AS average_review_score FROM order_reviews;
42	Aggregation	How many reviews are there for each score?	SELECT review_score, COUNT(review_id) AS review_count FROM order_reviews GROUP BY review_score;
43	Basic Selection	Show me all reviews with a score of 1.	SELECT review_comment_title, review_comment_message FROM order_reviews WHERE review_score = 1;
44	Aggregation	What is the average time taken to answer a review?	SELECT AVG(JULIANDAY(review_answer_timestamp) - JULIANDAY(review_creation_date)) AS avg_answer_time_days FROM order_reviews;
45	Basic Selection	List all reviews for a specific order ID 'e481f51cbdc54678b7cc49136f2d6af7'.	SELECT * FROM order_reviews WHERE order_id = 'e481f51cbdc54678b7cc49136f2d6af7';
46	Filtering	Find orders that have a review but no comment message.	SELECT order_id FROM order_reviews WHERE review_comment_message IS NULL OR review_comment_message = '';
47	Aggregation	How many reviews were created each month?	SELECT STRFTIME('%Y-%m', review_creation_date) AS month, COUNT(review_id) AS review_count FROM order_reviews GROUP BY month ORDER BY month;
48	Filtering	Show me the review comments that contain the word 'great'.	SELECT review_comment_message FROM order_reviews WHERE review_comment_message LIKE '%great%';
49	Aggregation	What is the highest review score given?	SELECT MAX(review_score) FROM order_reviews;
50	Joins	Which orders have not been reviewed yet?	SELECT o.order_id FROM orders o LEFT JOIN order_reviews r ON o.order_id = r.order_id WHERE r.review_id IS NULL;
51	Basic Selection	What are the latitude and longitude for the zip code prefix 1001?	SELECT geolocation_lat, geolocation_lng FROM geolocation WHERE geolocation_zip_code_prefix = 1001;
52	Aggregation	How many unique zip code prefixes are there in the geolocation data?	SELECT COUNT(DISTINCT geolocation_zip_code_prefix) FROM geolocation;
53	Basic Selection	List all cities in the state of 'SP'.	SELECT DISTINCT geolocation_city FROM geolocation WHERE geolocation_state = 'SP';
54	Aggregation	Find the geographical boundaries (min/max lat/lng) for the state of 'RJ'.	SELECT MIN(geolocation_lat), MAX(geolocation_lat), MIN(geolocation_lng), MAX(geolocation_lng) FROM geolocation WHERE geolocation_state = 'RJ';
55	Aggregation	How many geolocation entries are there for each city?	SELECT geolocation_city, COUNT(*) AS entry_count FROM geolocation GROUP BY geolocation_city ORDER BY entry_count DESC;
56	Joins	Show the customer's city and state along with their geolocation coordinates.	SELECT c.customer_id, c.customer_city, c.customer_state, g.geolocation_lat, g.geolocation_lng FROM customers c JOIN geolocation g ON c.customer_zip_code_prefix = g.geolocation_zip_code_prefix;
57	Joins	Find the seller's city and state along with their geolocation coordinates.	SELECT s.seller_id, s.seller_city, s.seller_state, g.geolocation_lat, g.geolocation_lng FROM sellers s JOIN geolocation g ON s.seller_zip_code_prefix = g.geolocation_zip_code_prefix;
58	Aggregation	What is the average latitude and longitude for all locations?	SELECT AVG(geolocation_lat), AVG(geolocation_lng) FROM geolocation;
59	Filtering	List zip code prefixes that appear in more than one city.	SELECT geolocation_zip_code_prefix, COUNT(DISTINCT geolocation_city) AS city_count FROM geolocation GROUP BY geolocation_zip_code_prefix HAVING city_count > 1;
60	Filtering	Get the state for a given latitude and longitude range.	SELECT DISTINCT geolocation_state FROM geolocation WHERE geolocation_lat BETWEEN -23.6 AND -23.5 AND geolocation_lng BETWEEN -46.7 AND -46.6;
61	Aggregation	How many qualified leads were there in total?	SELECT COUNT(mql_id) FROM leads_qualified;
62	Aggregation	When was the first contact made with any lead?	SELECT MIN(first_contact_date) FROM leads_qualified;
63	Aggregation	How many leads came from each origin?	SELECT origin, COUNT(mql_id) AS lead_count FROM leads_qualified GROUP BY origin;
64	Basic Selection	List all leads that came from the 'social' origin.	SELECT * FROM leads_qualified WHERE origin = 'social';
65	Aggregation	How many leads were converted into sellers (closed leads)?	SELECT COUNT(mql_id) FROM leads_closed;
66	Aggregation	What is the conversion rate from qualified lead to a closed deal?	SELECT (SELECT COUNT(*) FROM leads_closed) * 100.0 / (SELECT COUNT(*) FROM leads_qualified) AS conversion_rate;
67	Aggregation	Who are the top 5 Sales Development Representatives (SDRs) by the number of closed deals?	SELECT sdr_id, COUNT(mql_id) AS deals_closed FROM leads_closed GROUP BY sdr_id ORDER BY deals_closed DESC LIMIT 5;
68	Aggregation	What is the most common business segment for new sellers?	SELECT business_segment, COUNT(mql_id) AS segment_count FROM leads_closed GROUP BY business_segment ORDER BY segment_count DESC LIMIT 1;
69	Basic Selection	How many closed leads have a company?	SELECT COUNT(mql_id) FROM leads_closed WHERE has_company = 1;
70	Aggregation	What is the average declared monthly revenue of new sellers?	SELECT AVG(declared_monthly_revenue) FROM leads_closed;
71	Aggregation	Show me the lead type distribution for closed leads.	SELECT lead_type, COUNT(mql_id) AS lead_type_count FROM leads_closed GROUP BY lead_type;
72	Aggregation	What is the average declared product catalog size for each business type?	SELECT business_type, AVG(declared_product_catalog_size) AS avg_catalog_size FROM leads_closed GROUP BY business_type;
73	Basic Selection	List all closed leads with a 'cat' lead behavior profile.	SELECT * FROM leads_closed WHERE lead_behaviour_profile = 'cat';
74	Aggregation	How many leads were won each month?	SELECT STRFTIME('%Y-%m', won_date) AS month, COUNT(mql_id) AS won_leads_count FROM leads_closed GROUP BY month ORDER BY month;
75	Joins	Show the journey of a lead from qualified to closed.	SELECT q.mql_id, q.first_contact_date, q.landing_page_id, q.origin, c.won_date, c.business_segment FROM leads_qualified q JOIN leads_closed c ON q.mql_id = c.mql_id;
76	Joins	What is the average review score for each product category?	SELECT p.product_category_name, AVG(r.review_score) AS avg_review_score FROM products p JOIN order_items oi ON p.product_id = oi.product_id JOIN order_reviews r ON oi.order_id = r.order_id GROUP BY p.product_category_name;
77	Joins	List the top 5 best-selling products by revenue.	SELECT p.product_id, t.product_category_name_english, SUM(oi.price) AS total_revenue FROM products p JOIN order_items oi ON p.product_id = oi.product_id JOIN product_category_name_translation t ON p.product_id = t.product_id ORDER BY total_revenue DESC LIMIT 5;
78	Joins	What is the average delivery time for each seller's state?	SELECT s.seller_state, AVG(JULIANDAY(o.order_delivered_customer_date) - JULIANDAY(o.order_purchase_timestamp)) AS avg_delivery_days FROM orders o JOIN order_items oi ON o.order_id = oi.order_id JOIN sellers s ON o.customer_id = s.customer_id GROUP BY s.seller_state;
79	Joins	Which customers have bought products from sellers in their own state?	SELECT DISTINCT c.customer_unique_id FROM customers c JOIN orders o ON c.customer_id = o.customer_id JOIN order_items oi ON o.order_id = oi.order_id JOIN sellers s ON oi.seller_id = s.seller_id WHERE c.customer_state = s.seller_state;
80	Joins	What is the total payment value for each product category?	SELECT p.product_category_name, SUM(op.payment_value) AS total_payment FROM products p JOIN order_items oi ON p.product_id = oi.product_id JOIN order_payments op ON oi.order_id = op.order_id GROUP BY p.product_category_name;
81	Aggregation	Find the percentage of orders that have a review score of 5.	SELECT (SELECT COUNT(*) FROM order_reviews WHERE review_score = 5) * 100.0 / (SELECT COUNT(*) FROM orders) AS percentage_5_star_reviews;
82	Analytics	What is the most common payment type for each product category?	WITH CategoryPayment AS (SELECT p.product_category_name, op.payment_type, COUNT(*) as payment_count, ROW_NUMBER() OVER(PARTITION BY p.product_category_name ORDER BY payment_count) DESC) as cp SELECT cp.product_category_name, cp.payment_type FROM CategoryPayment cp WHERE cp.payment_count = 1;
83	Aggregation	List customers who have only ever made one order.	SELECT c.customer_unique_id FROM customers c JOIN orders o ON c.customer_id = o.customer_id GROUP BY c.customer_unique_id HAVING COUNT(o.order_id) = 1;
84	Joins	What is the average product weight for orders with a freight value over 50?	SELECT AVG(p.product_weight_g) FROM products p JOIN order_items oi ON p.product_id = oi.product_id WHERE oi.freight_value > 50;
85	Filtering	Show the order status for orders with a total payment value greater than 1000.	SELECT o.order_id, o.order_status, SUM(op.payment_value) AS total_payment FROM orders o JOIN order_payments op ON o.order_id = op.order_id GROUP BY o.order_id, o.order_status HAVING total_payment > 1000;
86	Joins	Which sellers have the highest average review scores?	SELECT oi.seller_id, AVG(r.review_score) AS avg_review_score FROM order_items oi JOIN order_reviews r ON oi.order_id = r.order_id GROUP BY oi.seller_id ORDER BY avg_review_score DESC LIMIT 10;
87	Aggregation	What is the time difference between the first and last order?	SELECT JULIANDAY(MAX(order_purchase_timestamp)) - JULIANDAY(MIN(order_purchase_timestamp)) AS days_between_first_last_order FROM orders;
88	Joins	Find the total revenue per state, based on customer location.	SELECT c.customer_state, SUM(op.payment_value) AS total_revenue FROM customers c JOIN orders o ON c.customer_id = o.customer_id JOIN order_payments op ON o.order_id = op.order_id GROUP BY c.customer_state;
89	Analytics	What is the most frequently purchased product category in each state?	WITH StateCategory AS (SELECT c.customer_state, p.product_category_name, COUNT(*) as purchase_count, ROW_NUMBER() OVER(PARTITION BY c.customer_state ORDER BY purchase_count) DESC) as sc SELECT sc.customer_state, sc.product_category_name FROM StateCategory sc WHERE sc.purchase_count = 1;
90	Aggregation	Show the total number of orders and the total revenue for each month.	SELECT STRFTIME('%Y-%m', o.order_purchase_timestamp) AS month, COUNT(DISTINCT o.order_id) AS number_of_orders, SUM(op.payment_value) AS total_revenue FROM orders o JOIN order_payments op ON o.order_id = op.order_id GROUP BY month;
91	Filtering	List all products that have never been sold.	SELECT p.product_id, p.product_category_name FROM products p LEFT JOIN order_items oi ON p.product_id = oi.product_id WHERE oi.order_id IS NULL;
92	Aggregation	What is the average number of products per order?	SELECT AVG(item_count) FROM (SELECT order_id, COUNT(order_item_id) AS item_count FROM order_items GROUP BY order_id) AS order_counts;
93	Analytics	Find the top 3 sellers in terms of revenue for each state.	WITH SellerStateRevenue AS (SELECT s.seller_state, s.seller_id, SUM(oi.price) as total_revenue, ROW_NUMBER() OVER(PARTITION BY s.seller_state ORDER BY SUM(oi.price) DESC) as rn FROM sellers s JOIN order_items oi ON s.customer_id = o.customer_id GROUP BY s.seller_state, s.seller_id) as ssr SELECT ssr.seller_state, ssr.seller_id FROM ssr WHERE rn <= 3;
94	Joins	What is the distribution of business types for sellers in each state?	SELECT s.seller_state, lc.business_type, COUNT(lc.seller_id) AS seller_count FROM sellers s JOIN leads_closed lc ON s.seller_id = lc.seller_id GROUP BY s.seller_state, lc.business_type ORDER BY s.seller_state, lc.business_type;
95	Joins	Show the average declared monthly revenue for sellers by their lead origin.	SELECT lc.origin, AVG(lc.declared_monthly_revenue) AS avg_revenue FROM leads_qualified lq JOIN leads_closed lc ON lq.mql_id = lc.mql_id GROUP BY lc.origin;
96	Joins	Which landing pages generate the most valuable leads (in terms of declared monthly revenue)?	SELECT lq.landing_page_id, SUM(lc.declared_monthly_revenue) AS total_declared_revenue FROM leads_qualified lq JOIN leads_closed lc ON lq.mql_id = lc.mql_id GROUP BY lq.landing_page_id ORDER BY total_declared_revenue DESC;
97	Joins	What is the average time from first contact to a won deal for each SDR?	SELECT lc.sdr_id, AVG(JULIANDAY(lc.won_date) - JULIANDAY(lq.first_contact_date)) AS avg_days_to_win FROM leads_qualified lq JOIN leads_closed lc ON lq.mql_id = lc.mql_id GROUP BY lc.sdr_id;
98	Aggregation	Compare the average product catalog size of sellers who have a company versus those who do not.	SELECT has_company, AVG(declared_product_catalog_size) AS avg_catalog_size FROM leads_closed GROUP BY has_company;
99	Joins	What is the total freight value per customer state?	SELECT c.customer_state, SUM(oi.freight_value) AS total_freight FROM customers c JOIN orders o ON c.customer_id = o.customer_id JOIN order_items oi ON o.order_id = oi.order_id GROUP BY c.customer_state;
100	Analytics	Find the correlation between the length of the review comment and the review score.	SELECT review_score, AVG(LENGTH(review_comment_message)) AS avg_comment_length FROM order_reviews WHERE review_comment_message IS NOT NULL GROUP BY review_score ORDER BY review_score;