

## Cleaning a PostgreSQL Database

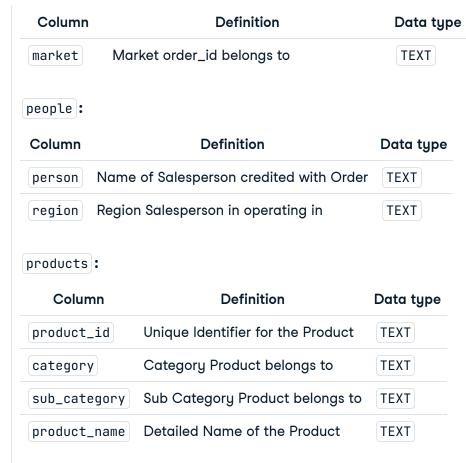
In this project, you will work with data from a hypothetical Super Store to challenge and enhance your SQL skills in data cleaning. This project will engage you in identifying top categories based on the highest profit margins and detecting missing values, utilizing your comprehensive knowledge of SQL concepts.

## **Data Dictionary:**

O	n	a	$\sim$	n	_	•
u	Ι.	u	<b>—</b>	1 . 3	<b>7</b>	

Column	Definition	Data type	Comments
row_id	Unique Record ID	INTEGER	
order_id	ldentifier for each order in table	TEXT	Connects to order_id in returned_orders table
order_date	Date when order was placed	TEXT	
market	Market order_id belongs to	TEXT	
region	Region Customer belongs to	TEXT	Connects to region in people table
product_id	Identifier of Product bought	TEXT	Connects to product_id in products table
sales	Total Sales Amount for the Line Item	DOUBLE PRECISION	
quantity	Total Quantity for the Line Item	DOUBLE PRECISION	
discount	Discount applied for the Line Item	DOUBLE PRECISION	
profit	Total Profit earned on the Line Item	DOUBLE PRECISION	

Column	Definition	Data type	
returned	Yes values for Order / Line Item Returned	TEXT	
order_id	ldentifier for each order in table	TEXT	



As you can see in the Data Dictionary above, date fields have been written to the orders table as TEXT and numeric fields like sales, profit, etc. have been written to the orders table as Double Precision. You will need to take care of these types in some of the queries. This project is an excellent opportunity to apply your SQL skills in a practical setting and gain valuable experience in data cleaning and analysis. Good luck, and happy querying!

~	category	product_name	product_total_sales
0	Furniture	Hon Executive Leather Armchair, Adjustable	58
1	Furniture	Office Star Executive Leather Armchair, Adjustable	
2	Furniture	Harbour Creations Executive Leather Armchair, Adjustable	50
3	Furniture	SAFCO Executive Leather Armchair, Black	4:
4	Furniture	Novimex Executive Leather Armchair, Adjustable	40
5	Office Supplies	Eldon File Cart, Single Width	39
6	Office Supplies	Hoover Stove, White	3
7	Office Supplies	Hoover Stove, Red	32
8	Office Supplies	Rogers File Cart, Single Width	29
9	Office Supplies	Smead Lockers, Industrial	28
10	Technology	Apple Smart Phone, Full Size	86
11	Technology	Cisco Smart Phone, Full Size	76
12	Technology	Motorola Smart Phone, Full Size	7
13	Technology	Nokia Smart Phone, Full Size	7:
11	Technology	Canon imageCLASS 2200 Advanced Copier	6:

```
Projects Data
                DataFrame as i
-- impute_missing_values
WITH missing AS (
    SELECT product_id,
        discount,
        market,
        region,
        sales,
        quantity
    FROM orders
    WHERE quantity IS NULL
),
unit_prices AS (SELECT o.product_id,
    CAST(o.sales / o.quantity AS NUMERIC) AS unit_price
FROM orders o
RIGHT JOIN missing AS m
    ON o.product_id = m.product_id
    AND o.discount = m.discount
WHERE o.quantity IS NOT NULL
)
SELECT DISTINCT m.*,
    ROUND(CAST(m.sales AS NUMERIC) / up.unit_price,0) AS calculated_quantity
FROM missing AS m
INNER JOIN unit_prices AS up
```