

Estudiante:

Heiner Romero Leiva

Máster en Big Data & Data Science

Noviembre, 2021



Descripción de la tarea:

Has sido contratado por una multinacional como Data Engineer. Se trata de una empresa global que vende hardware informático, incluidos almacenamiento, placas base, RAM, tarjetas de video y CPUs. Te dan acceso a su base de datos relacional, la cual está montada en un servidor Oracle. El modelo de datos (tablas, campos y relaciones) es el siguiente:

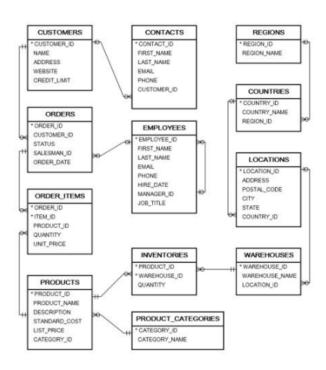
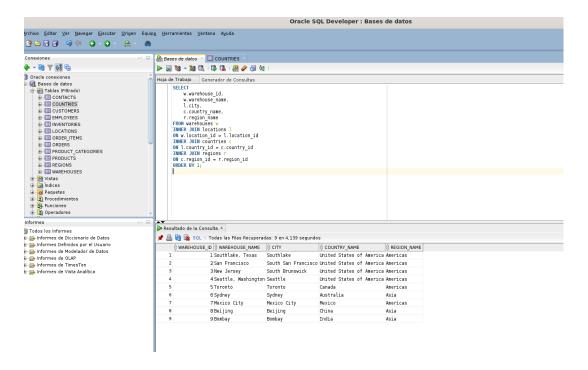


Tabla	Descripción	Registros
CONTACTS	Tabla con información de la persona de contacto de los clientes	319 registros
COUNTRIES	Tabla con información de países	25 registros
CUSTOMERS	Tabla con información de clientes	319 registros
EMPLOYEES	Tabla con información de empleados	107 registros
INVENTORIES	Tabla con información del inventario de productos	1112 registros
LOCATIONS	Tabla con ubicación de los almacenes	23 registros
ORDERS	Tabla de pedidos	105 registros
ORDER_ITEMS	Tabla con las líneas de los pedidos	665 registros
PRODUCT_CATEGORIES	Tabla con las categorías de productos	5 registros
PRODUCTS	Tabla con la información de los productos	288 registros
REGIONS	Tabla con las regiones donde la compañía opera	4 registros
WAREHOUSES	Tabla con información de los almacenes	9 registros

Debes crear las sentencias SQL necesarias para poder responder a las siguientes tareas:

- 1. ¿Cuáles son los datos de los almacenes que tiene la compañía? Se necesita:
 - Identificador del almacén.
 - Nombre del almacén.
 - Nombre de la ciudad, país y región donde está ubicado.



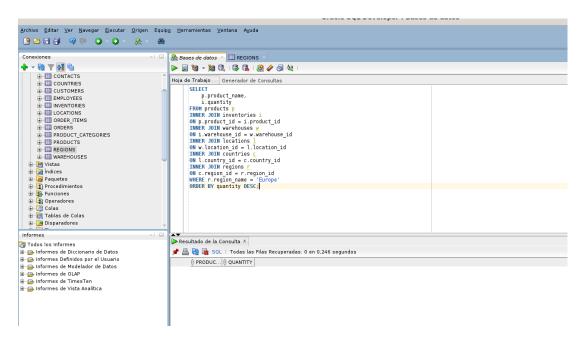
Query:

SELECT

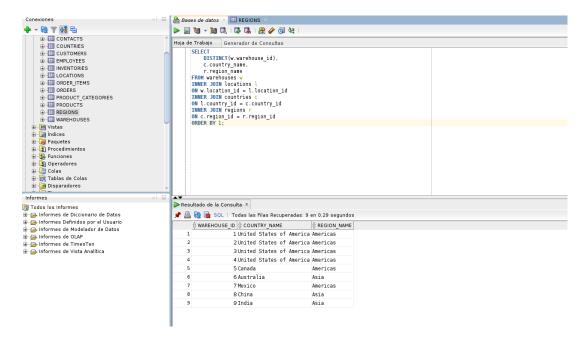
w.warehouse_id,
w.warehouse_name,
l.city,
c.country_name,
r.region_name
FROM warehouses w
INNER JOIN locations l
ON w.location_id = l.location_id
INNER JOIN countries c
ON l.country_id = c.country_id
INNER JOIN regions r
ON c.region_id = r.region_id
ORDER BY 1;

2. ¿Cuál es el nombre del producto que tiene más stock en Europa?

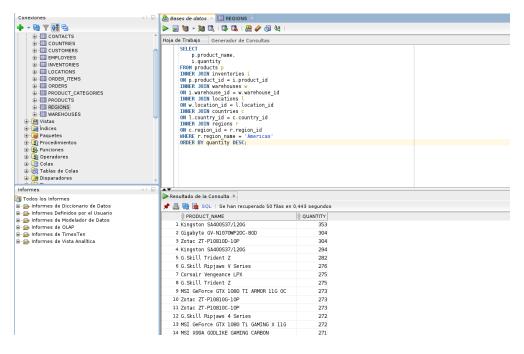
Para este caso en particular me percaté de que en Europa no hay almacén solo existen en América y en Asia, entonces lo que hice fue crear el query con "Europe" pero como vi que no arrojaba resultados, hice otro query para ver si efectivamente no habían almacenes en Europa.

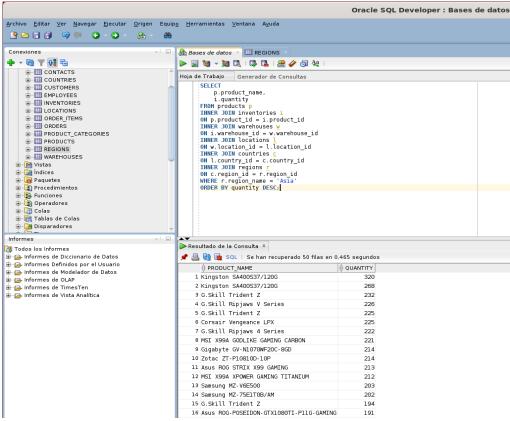


Con este query me di cuenta de que en efecto en Europa no hay almacenes, solo en América y Asia.

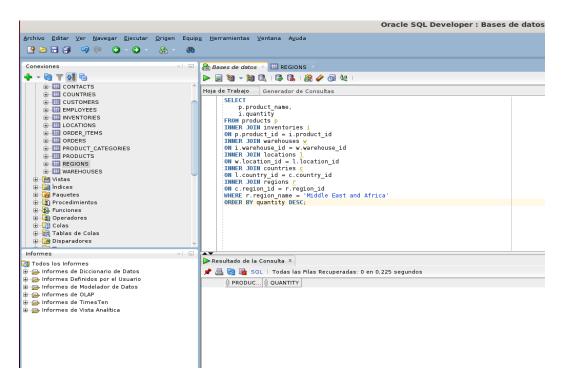


Luego lo que hice fue comprobar si el query estaba bien hecho y arrojaba resultados, así que lo probe para "Americas" y "Asia".





Finalmente también comprobé con "Middle East and Africa" para corroborrar que no arrojara ningún dato.



Queries:

/* Query para Europa*/

```
SELECT
```

p.product_name,

i.quantity

FROM products p

INNER JOIN inventories i

ON p.product_id = i.product_id

INNER JOIN warehouses w

ON i.warehouse_id = w.warehouse_id

INNER JOIN locations 1

ON w.location_id = 1.location_id

INNER JOIN countries c

ON 1.country id = c.country id

INNER JOIN regions r

ON c.region id = r.region id

WHERE r.region_name = 'Europe'

ORDER BY quantity DESC;

/* Query para comprobar ubicación de los almacenes*/

SELECT

DISTINCT(w.warehouse_id)

c.country_name

r.region_name

FROM warehouses w

INNER JOIN locations 1

ON w.location_id = l.location_id

INNER JOIN countries c

ON l.country_id = c.country_id

INNER JOIN regions r

ON c.region_id = r.region_id

/* Query para América*/

SELECT

p.product_name,

i.quantity

FROM products p

INNER JOIN inventories i

ON p.product_id = i.product_id

INNER JOIN warehouses w

ON i.warehouse_id = w.warehouse_id

INNER JOIN locations 1

ON w.location_id = l.location_id

INNER JOIN countries c

ON l.country_id = c.country_id

INNER JOIN regions r

ON c.region_id = r.region_id

WHERE r.region_name = 'Americas'

ORDER BY quantity DESC;

/* Query para Asia*/

```
SELECT
```

p.product_name,

i.quantity

FROM products p

INNER JOIN inventories i

ON p.product_id = i.product_id

INNER JOIN warehouses w

ON i.warehouse_id = w.warehouse_id

INNER JOIN locations 1

ON w.location_id = l.location_id

INNER JOIN countries c

ON l.country_id = c.country_id

INNER JOIN regions r

ON c.region_id = r.region_id

WHERE r.region_name = 'Asia'

ORDER BY quantity DESC;

/* Query para Medio Este y África */

SELECT

p.product_name,

i.quantity

FROM products p

INNER JOIN inventories i

ON p.product_id = i.product_id

INNER JOIN warehouses w

ON i.warehouse_id = w.warehouse_id

INNER JOIN locations 1

ON w.location_id = 1.location_id

INNER JOIN countries c

ON l.country_id = c.country_id

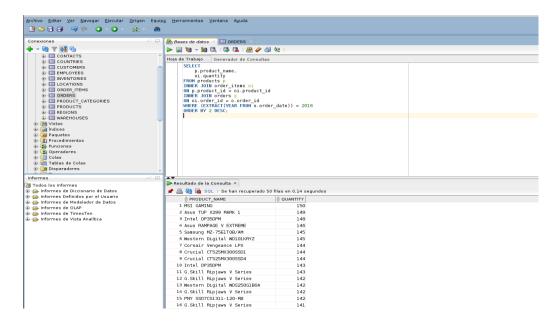
INNER JOIN regions r

ON c.region_id = r.region_id

WHERE r.region_name = 'Middle East and Africa'

ORDER BY quantity DESC;

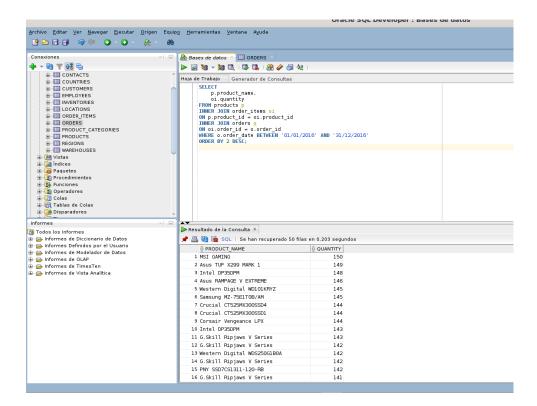
3. ¿Cuál es el producto que ha vendido más unidades durante 2016?



El producto con más venta de unidades para 2016 es MSI GAMING con 150 unidades.

```
p.product_name,
oi.quantity,
FROM products p
INNER JOIN order_items oi
ON p.product_id = oi.product_id
INNER JOIN orders o
ON oi.order_id = o.order_id
WHERE (EXTRACT(YEAR FROM o.order_date)) = 2016
ORDER BY 2 DESC;
```

También lo comprobé hacienda otro tipo de sentencia con la fecha y fue el mismo:



Query:

```
SELECT
```

p.product_name,

oi.quantity,

FROM products p

INNER JOIN order_items oi

ON p.product_id = oi.product_id

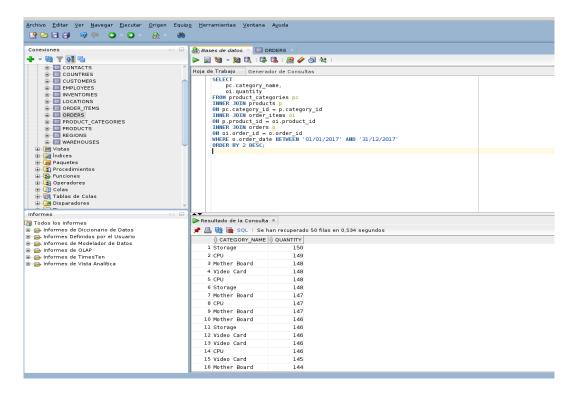
INNER JOIN orders o

ON oi.order_id = o.order_id

WHERE o.order_date BETWEEN '01/01/2016' AND '31/12/2016'

ORDER BY 2 DESC;





La categoría es Storage con 150 unidades para 2017.

```
SELECT

pc.category_name,
oi.quantity

FROM product_categories pc

INNER JOIN products p

ON pc.category_id = p.category_id

INNER JOIN order_items oi

ON p.product_id = oi.product_id

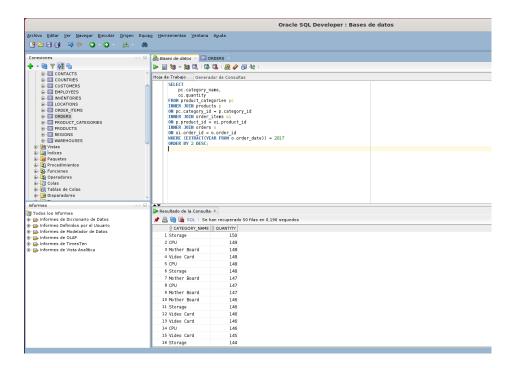
INNER JOIN orders o

ON oi.order_id = o.order_id

WHERE o.order_date BETWEEN '01/01/2017' AND '31/12/2017'

ORDER BY 2 DESC;
```

De igual forma lo comprobé usando otro tipo de sentencia para extraer la variable de año y dio el mismo resultado.



Query:

```
SELECT
```

pc.category_name,
oi.quantity

FROM product_categories pc

INNER JOIN products p

ON pc.category_id = p.category_id

INNER JOIN order_items oi

ON p.product_id = oi.product_id

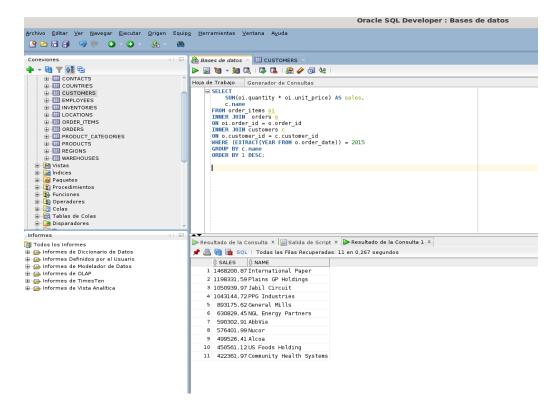
INNER JOIN orders o

ON oi.order_id = o.order_id

WHERE (EXTRACT(YEAR FROM o.order_date)) = 2017

ORDER BY 2 DESC;

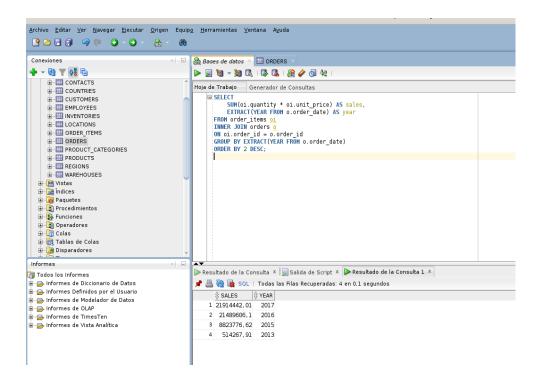
5. ¿Cuál es el nombre del cliente cuyo gasto ha sido más alto en 2015?



Para el año 2015 el cliente: International Paper es el que registra el gasto más alto con \$1.4M. Cabe aclarar algo, y es que aunque la base de datos contiene 319 registros de clientes, hay unos que aunque registrados, no hacen compras.

```
SELECT
SUM(oi.quantity * oi.unit_price) AS sales,
c.name
FROM order_items oi
INNER JOIN orders o
ON oi.order_id = o.order_id
INNER JOIN customers c
ON o.customer_id = c.customer_id
WHERE (EXTRACT(YEAR FROM o.order_date)) = 2015
GROUP BY c.name
ORDER BY 1 DESC;
```

6. ¿Cuánto ha facturado la compañía en cada uno de los años de los que tiene datos?



Query:

SELECT

SUM(oi.quantity * oi.unit_price) AS sales, EXTRACT(YEAR FROM o.order_date) AS year

FROM order_items oi

INNER JOIN orders o

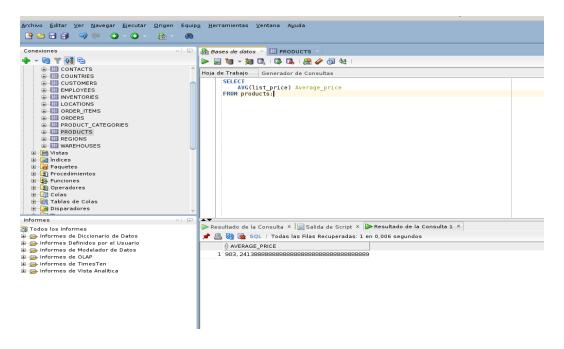
ON oi.order_id = o.order_id

GROUP BY EXTRACT(YEAR FROM o.order_date)

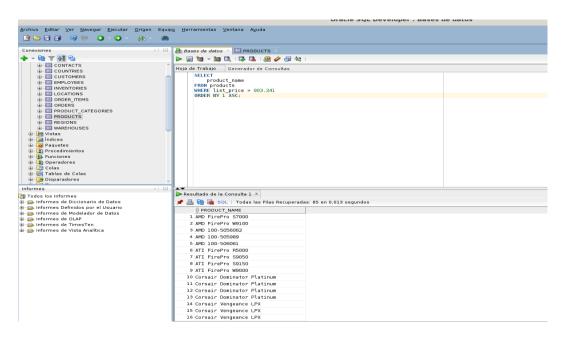
ORDER BY 2 DESC;

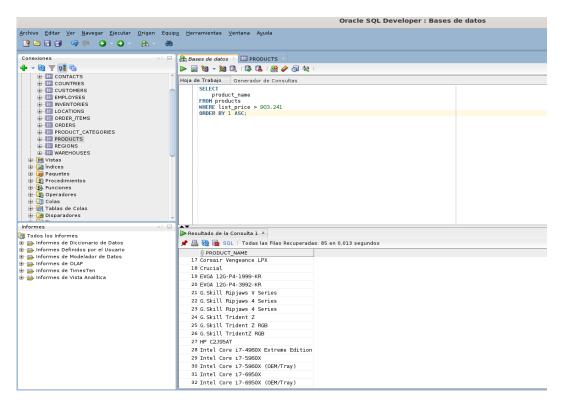
7. ¿Cuáles son los nombres de los productos cuyo precio es superior la media?

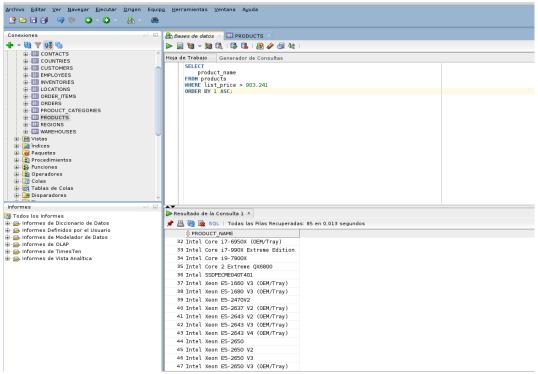
Aquí primero calculé la media de la lista de productos.

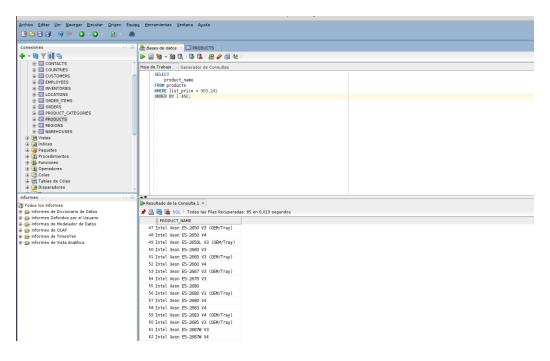


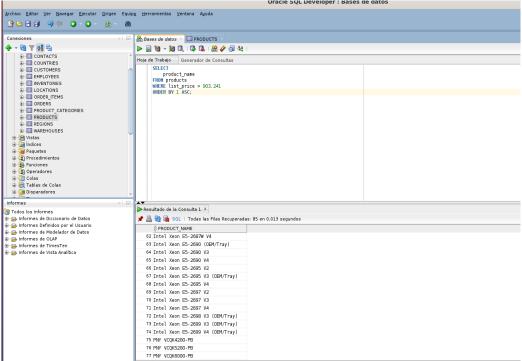
Luego con esa media, me di a la tarea de hacer un filtro con el list_price, con todos aquellos productos que tuviesen un precio superior a \$903.24 y el resultado me arrojó que hay 85 productos cuyo precio es mayor a la media y esos productos los ordené en forma ascendente.

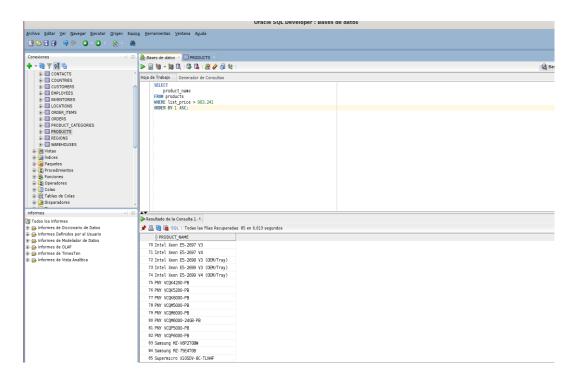












Queries:

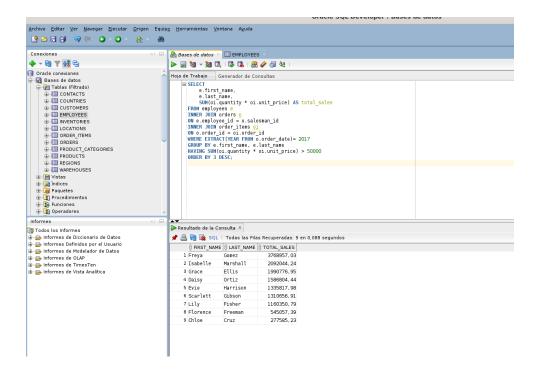
SELECT

AVG(list_price) Average_price FROM products;

SELECT

product_name
FROM products
WHERE list_price > 903,241
ORDER BY 1 ASC;

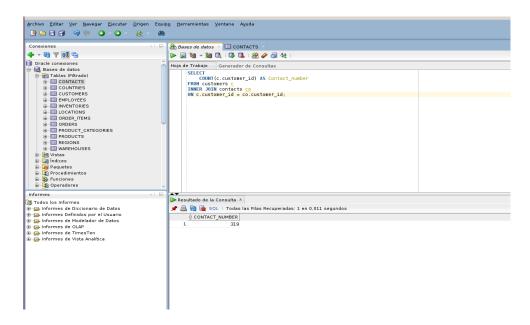
8. ¿Cuáles son los empleados (nombre y apellido) que han vendido más de 50K durante 2017?



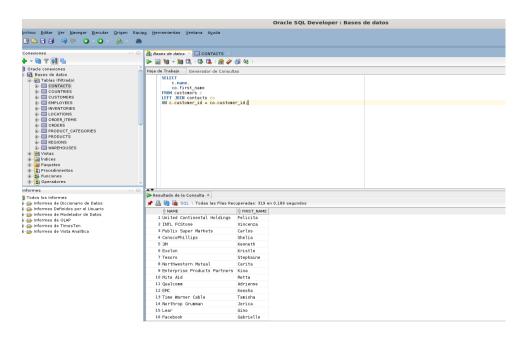
```
select
e.first_name,
e.last_name,
sum(oi.quantity * oi.unit_price) AS total_sales
FROM employees e
INNER JOIN orders o
ON e.employee_id = o.salesman_id
INNER JOIN order_items oi
ON o.order_id = oi.order_id
WHERE EXTRACT(YEAR FROM o.order_date)= 2017
GROUP BY e.first_name, e.last_name
HAVING SUM(oi.quantity * oi.unit_price) > 50000
ORDER BY 3 DESC;
```

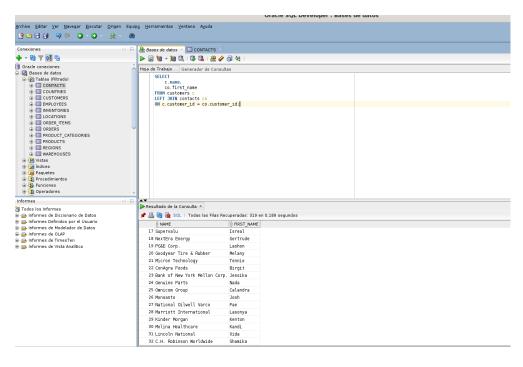
9. ¿Cuánto clientes no tienen persona de contacto?

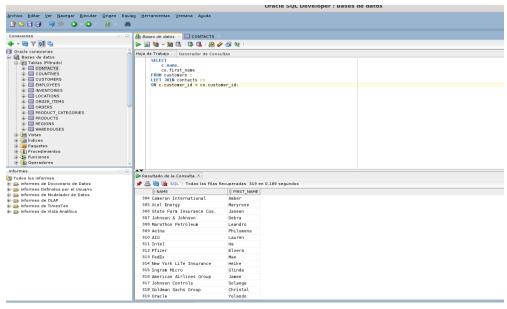
Aquí me di cuenta que necesitaba hacer un join para ver si alguno de esos clientes no tenía persona de contacto, entonces el primero que hice fue un INNER JOIN e hice un conteo para ver si tanto en la tabla de "customers" como en la de "contacs" había correspondencia y sí había, 319 registros arrojó (misma cantidad que se dio en la table inicial) es decir, todos los clientes tienen persona de contacto.



Sin embargo quise comprobar con un left join a ver si había correspondencia de la table "customers" con la tabla "contacts".







Al final comprobé que efectivamente en los 319 registros de la tabla "customers" todos tenían su correspondencia con la tabla de "contacts" y no había ningún registro nulo o huérfano.

```
SELECT

COUNT(c.CUSTOMER_ID) AS Contact_number
FROM customers c
INNER JOIN contacts co
ON c.customer_id = co.customer_id

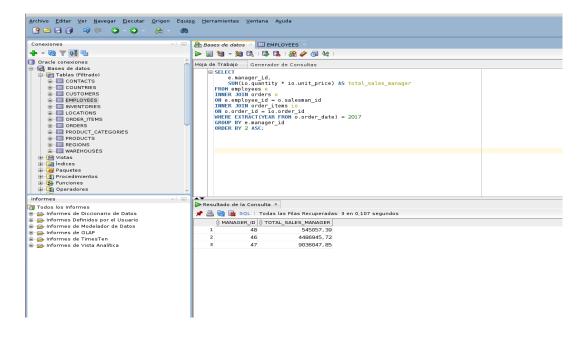
SELECT

c.name,
co.first_name
FROM customers c
LEFT JOIN contacts co
ON c.customer_id = co.customer_id
```

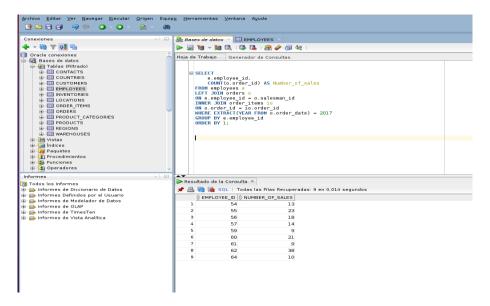
10. ¿Cuál es el Manager (nombre y apellido identificado por el campo manager_id) que menos ha vendido durante 2017?

El manager que menos vendió durante el 2017 fue el manager identificado con el número o registro "48" (en la tabla de employees, no tenía ni el nombre ni el apellido, este correspondía al nombre y apellido del vendedor).

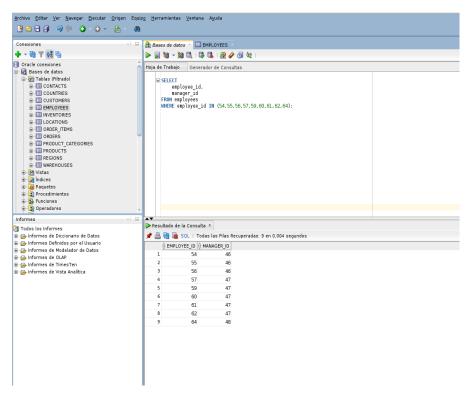
Sin embargo, luego de encontrar el resultado me generó curiosidad porqué para el 2017 solo tengo 3 managers, 46, 47 y 48, y en la tabla habían más registros de managers, entonces hice otras consultas.



Con esta consulta me quise cerciorar si todos los empleados vendieron en 2017, ya que vi muchos vendedores en la tabla con diferentes managers que no eran efectivamente los 46, 47 y 48 y me di cuenta de que en realidad solo unos cuantos empleados vendieron en 2017, no todos los que aparecen en la tabla.



Luego lo que hice fue comprobar si esos empleados le reportaban a alguno de los tres managers (46, 47, 48) y efectivamente ellos le reportaban a alguno de los mencionados.



Queries:

```
SELECT
      e.manager_id,
      SUM(io.quantity * io.unit_price) AS total_sales_manager
FROM employees e
INNER JOIN orders o
ON e.employee_id = o.salesman_id
INNER JOIN order_items io
ON o.order_id = io.order_id
WHERE EXTRACT(YEAR FROM o.order_date)= 2017
GROUP BY e.manager_id
ORDER BY 2 ASC;
SELECT
      e.employee_id,
      COUNT(o.order_id) AS Number_of_sales
FROM employees e
LEFT JOIN orders o
ON e.employee_id = o.salesman_id
INNER JOIN order_items io
ON o.order_id = io.order_id
WHERE EXTRACT(YEAR FROM o.order_date) = 2017
GROUP BY e.employee_id
ORDER BY 1;
SELECT
      employee_id,
      manager_id
FROM employees
WHERE employee_id IN (54,55,56,57,59,60,61,62,64);
                                   FIN -----
```