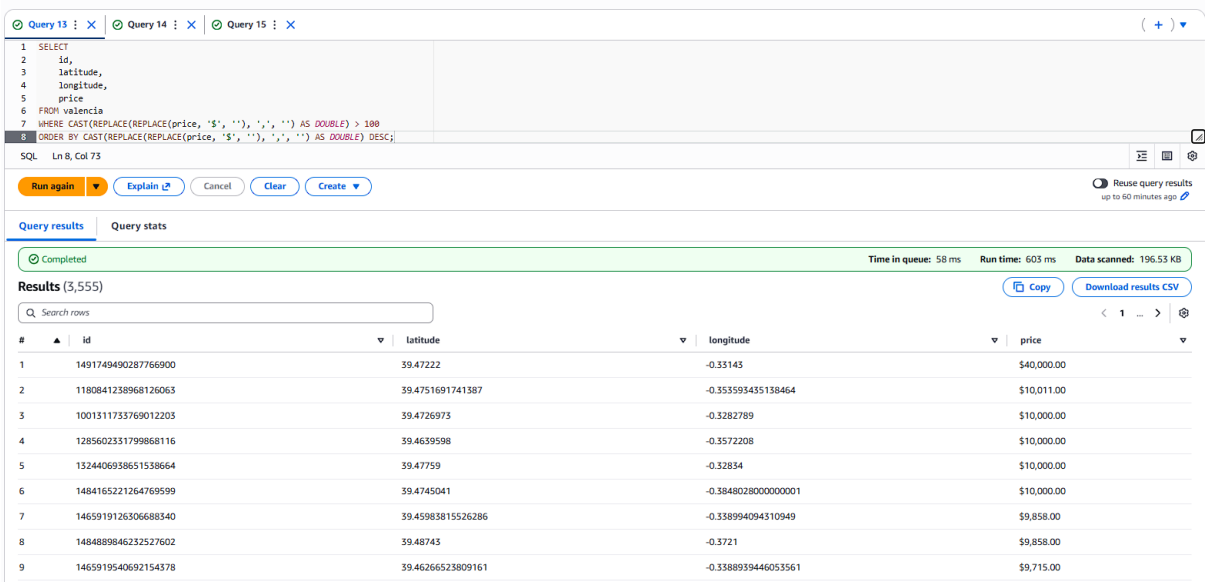
Imágenes de referencia en pasos del examen

Paso 2, inciso 1: Filtrado de los precios superiores a 100 dolares en Valencia. Query tanto resultado de la query en athena.



```
SELECT
    id,
    latitude,
    longitude,
    price
FROM valencia
WHERE CAST(REPLACE(REPLACE(price, '$', ''), ',', '') AS DOUBLE) > 100
ORDER BY CAST(REPLACE(REPLACE(price, '$', ''), ',', '') AS DOUBLE) DESC;
```

Query 13 | Query 14 | Query 15

SQL Ln 8, Col 73

Run again | Explain | Cancel | Clear | Create

Reuse query results up to 60 minutes ago

Query results | Query stats

Completed — Time in queue: 58 ms — Run time: 603 ms — Data scanned: 196.53 KB

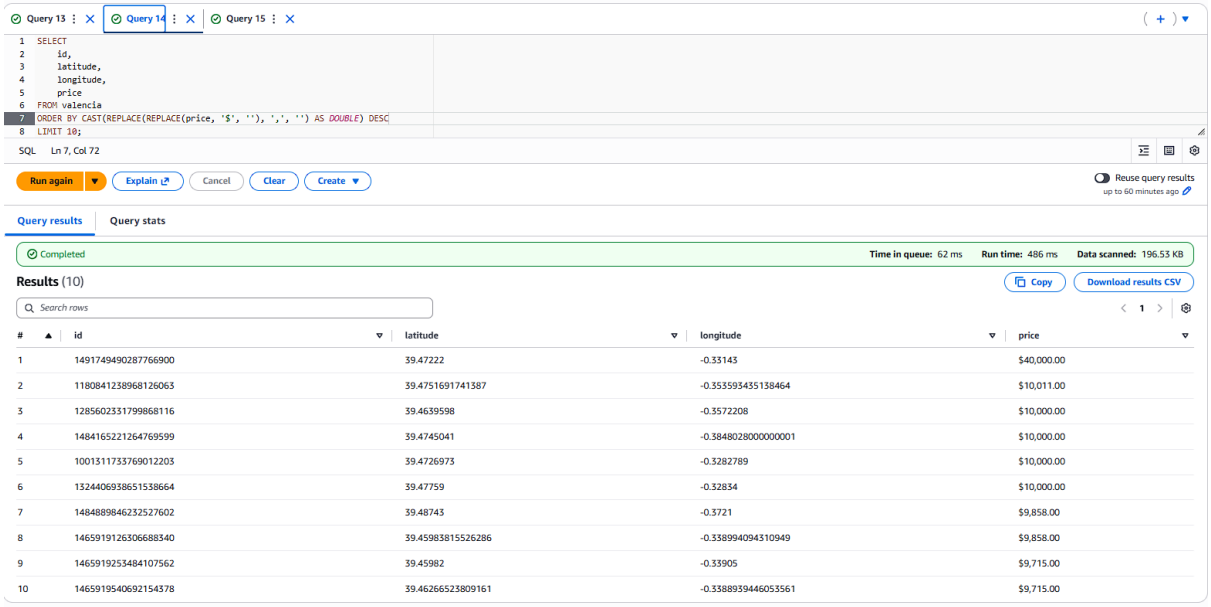
Results (3,555)

Copy | Download results CSV

Search rows

#	id	latitude	longitude	price
1	14917494902877669 00	39.47222	-0.33143	$40,000.00
2	11808412389681260 63	39.4751691741387	-0.353593435138464	$10,011.00
3	10013117337690122 03	39.4726973	-0.3282789	$10,000.00
4	12856023317998681 16	39.4639598	-0.3572208	$10,000.00
5	13244069386515386 64	39.47759	-0.32834	$10,000.00
6	14841652212647695 99	39.4745041	-0.3848028000000001	$10,000.00
7	14659191263066883 40	39.45983815526286	-0.338994094310949	$9,858.00
8	14848898462325276 02	39.48743	-0.3721	$9,858.00
9	14659195406921543 78	39.46266523809161	-0.3388939446053561	$9,715.00

Paso 2, inciso 2: Selección de los 10 primeros id con mayor precio en Valencia. Query tanto resultado de la query en athena.



```
SELECT
    id,
    latitude,
    longitude,
    price
FROM valencia
ORDER BY CAST(REPLACE(REPLACE(price, '$', ''), ',', '') AS DOUBLE) DESC
LIMIT 10;
```

Query 13 | Query 14 | Query 15

SQL Ln 7, Col 72

Run again | Explain | Cancel | Clear | Create

Reuse query results up to 60 minutes ago

Query results | Query stats

Completed — Time in queue: 62 ms — Run time: 486 ms — Data scanned: 196.53 KB

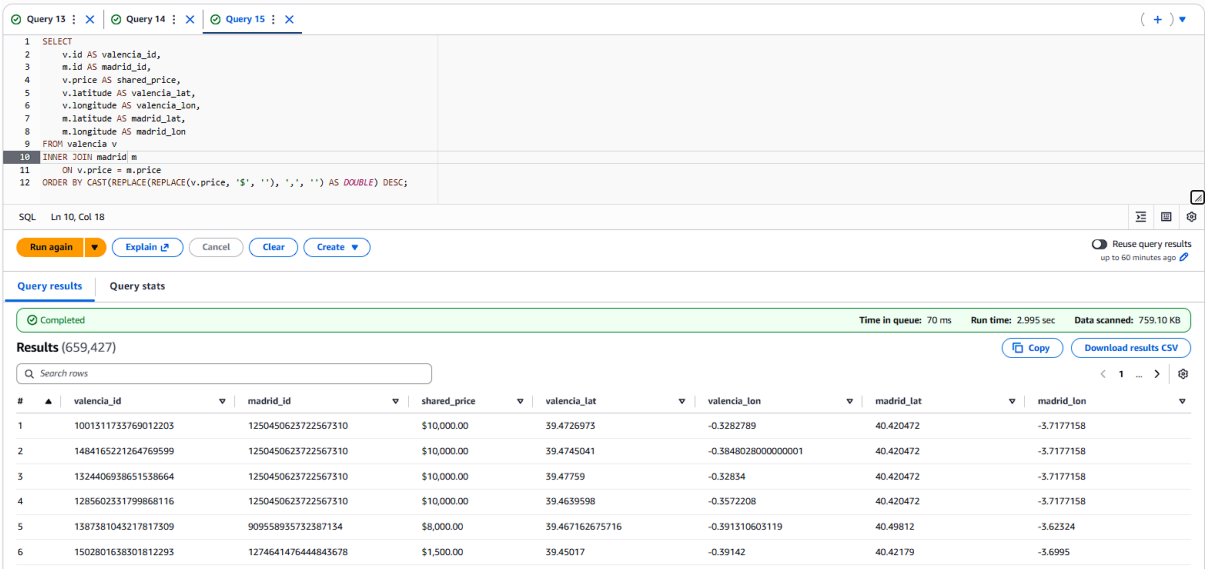
Results (10)

Copy | Download results CSV

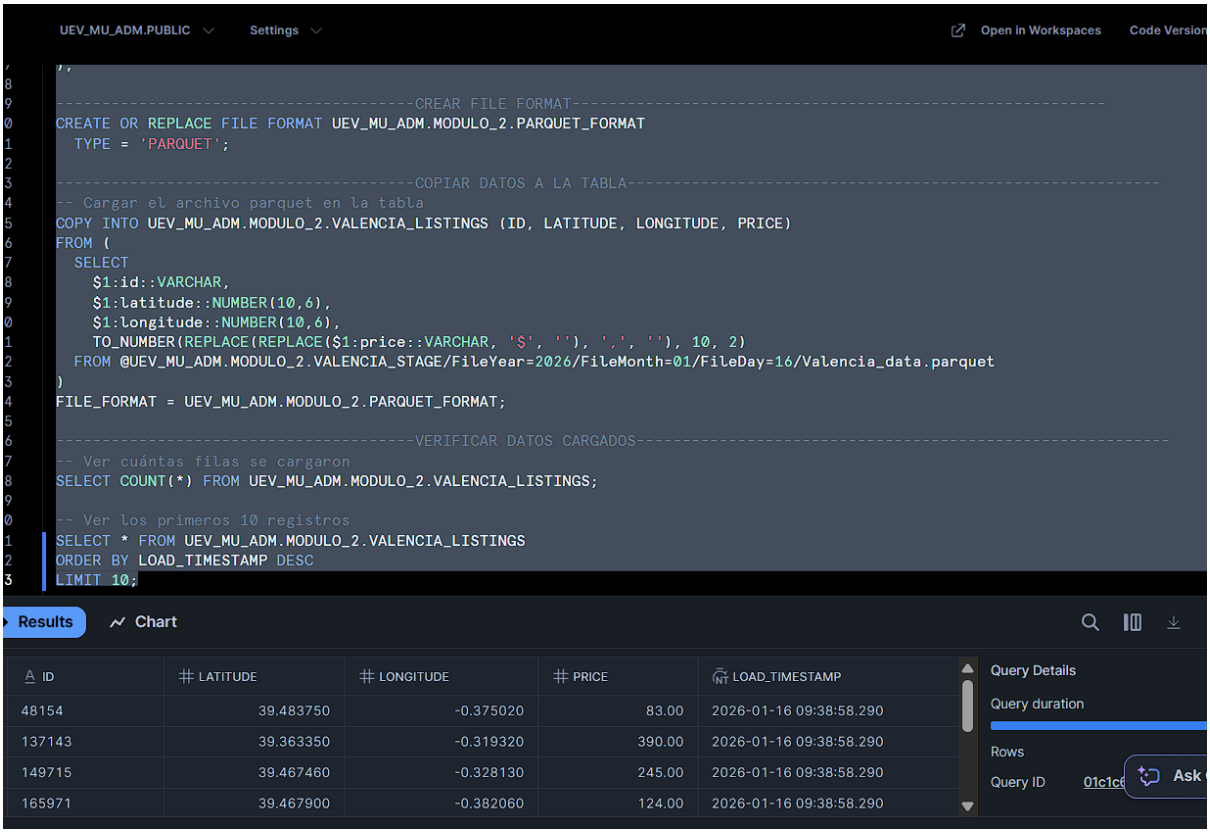
Search rows

#	id	latitude	longitude	price
1	14917494902877669 00	39.47222	-0.33143	$40,000.00
2	11808412389681260 63	39.4751691741387	-0.353593435138464	$10,011.00
3	12856023317998681 16	39.4639598	-0.3572208	$10,000.00
4	14841652212647695 99	39.4745041	-0.3848028000000001	$10,000.00
5	10013117337690122 03	39.4726973	-0.3282789	$10,000.00
6	13244069386515386 64	39.47759	-0.32834	$10,000.00
7	14848898462325276 02	39.48743	-0.3721	$9,858.00
8	14659191263066883 40	39.45983815526286	-0.338994094310949	$9,858.00
9	14659192534841075 62	39.45982	-0.33905	$9,715.00
10	14659195406921543 78	39.46266523809161	-0.3388939446053561	$9,715.00

Paso 2, inciso 3: Join de Valencia y Madrid para ver que ids tienen el mismo precio. Query tanto resultado de la query en athena.



Paso 3: ingesta de datos del archivo de valencia de S3 a snowflake:



Paso 4: transient table del 90% de los apartamentos mas caros de valencia, mediante una tarea:
Se hizo un select para ver los registros cargados:

```sql
-- Ver cuántos registros se cargaron (debe ser ~90% del total)
SELECT COUNT(*) as total_top_90 FROM UEV_MU_ADM.MODULO_2.VALENCIA_TOP_90_PRICES;

-- Comparar con el total original
SELECT
    (SELECT COUNT(*) FROM UEV_MU_ADM.MODULO_2.VALENCIA_LISTINGS) as total_original,
    (SELECT COUNT(*) FROM UEV_MU_ADM.MODULO_2.VALENCIA_TOP_90_PRICES) as total_top_90,
    ROUND((SELECT COUNT(*) FROM UEV_MU_ADM.MODULO_2.VALENCIA_TOP_90_PRICES) * 100.0 /
```

Results | Chart

TOTAL_TOP_90
7072

Se hizo otro para comparar si si es el 90%:

```sql
-- Comparar con el total original
SELECT
    (SELECT COUNT(*) FROM UEV_MU_ADM.MODULO_2.VALENCIA_LISTINGS) as total_original,
    (SELECT COUNT(*) FROM UEV_MU_ADM.MODULO_2.VALENCIA_TOP_90_PRICES) as total_top_90,
    ROUND((SELECT COUNT(*) FROM UEV_MU_ADM.MODULO_2.VALENCIA_TOP_90_PRICES) * 100.0 /
          (SELECT COUNT(*) FROM UEV_MU_ADM.MODULO_2.VALENCIA_LISTINGS), 2) as porcentaje;

-- Ver los precios mínimo y máximo del top 90%
SELECT
    MIN(PRICE) as precio_minimo_top90,
    MAX(PRICE) as precio_maximo_top90,
    AVG(PRICE) as precio_promedio_top90
FROM UEV_MU_ADM.MODULO_2.VALENCIA_TOP_90_PRICES;

-- Ver los primeros 10 registros
SELECT * FROM UEV_MU_ADM.MODULO_2.VALENCIA_TOP_90_PRICES
ORDER BY PRICE DESC
```

Results | Chart

	TOTAL_ORIGINAL	TOTAL_TOP_90	PORCENTAJE
1	7844	7072	90.16

Otro para los minimos y máximos dentro de ese 90%

```sql
-- Ver los precios mínimo y máximo del top 90%
SELECT
    MIN(PRICE) as precio_minimo_top90,
    MAX(PRICE) as precio_maximo_top90,
    AVG(PRICE) as precio_promedio_top90
FROM UEV_MU_ADM.MODULO_2.VALENCIA_TOP_90_PRICES;

-- Ver los primeros 10 registros
SELECT * FROM UEV_MU_ADM.MODULO_2.VALENCIA_TOP_90_PRICES
ORDER BY PRICE DESC
LIMIT 10;

-----------------------------------MONITOREAR LA TAREA-----------------------------------
-- Ver el historial de ejecución de la tarea
SELECT *
FROM TABLE(INFORMATION_SCHEMA.TASK_HISTORY(
    TASK_NAME => 'TASK_LOAD_TOP_90_PRICES'
```

Results | Chart

PRECIO_MINIMO_TOP90	PRECIO_MAXIMO_TOP90	PRECIO_PROMEDIO_TOP90
42.00	40000.00	180.68567746

Paso 7. Creacion de parquet particionado en base a los datos logrados en los pasos anteriores: