

Tarea evaluable

CE_5075 8.1

Apartado 2

Big data aplicado



Índice

APARTADO 1	2
1 - CONFIGURACIÓN DEL CLÚSTER	2
2 - QUERIES	7

APARTADO 2

Como ya sabemos, en <http://insideairbnb.com/get-the-data/> podemos encontrar los datos de los alojamientos de Airbnb. En concreto, volvemos a centrarnos en los de la isla de Menorca. Ahora vamos a trabajar con el archivo con datos detallados, dentro de `listings.csv.gz`

A partir del archivo CSV que contiene, debes desarrollar un **cuadro de mando en Databricks**, usando **Databricks SQL**, que muestre estos datos de la mejor manera posible.

Debes utilizar **al menos 5 elementos visuales distintos** (no tienen por qué ser los mismos que usamos en el ejemplo de los apuntes). También debes implementar **al menos 4 filtros (afinadores)** (los filtros no cuentan entre los 5 elementos visuales diferentes requeridos). El cuadro de mando también debe tener un **título**.

Para que el cuadro de mando sea efectivo, debes hacer una buena elección de los elementos visuales que presentes, cuidando la disposición y la estética general.

1 - CONFIGURACIÓN DEL CLÚSTER

Antes de realizar la práctica es necesario configurar el entorno así como nos muestran los apuntes. Primero se requiere actualizar el plan actual al Premium. Una vez hecho esto, accedemos al área de trabajo

En la barra lateral podremos ver la nueva sección SQL, pulsaremos en “Nuevo -> Más -> Almacén de SQL” para crear el warehouse.

Configuramos los parámetros y ajustes así como se nos muestra en los apuntes y estará listo para usarse.

Nuevo almacén de SQL

Nombre:

Tamaño del clúster: 4 DBU/h

Parada automática: ☒ Tras minutos de inactividad

Expansión: Mín. Máx. clústeres (4 DBU)

Tipo: ☒ Serverless ☐ Pro ☐ Clásico

Opciones avanzadas

Microsoft Azure | databricks

Buscar datos, cuadernos, recientes y más...

+ Nuevo

- Workspace
- Recientes
- Catálogo
- Flujos de trabajo
- Cómputo
- Marketplace

SQL

- Editor de SQL
- Consultas
- Tableros
- Genie
- Alertas
- Historial
- Almacenes de SQL**

Ingeniería de datos

Ejecuciones

Almacenes SQL >

apartado_2_sql_wh

Información general Detalles de conexión Monitoreo

Estado	● En ejecución
Nombre	apartado_2_sql_wh (ID: 383559e4fe26f48a)
Tipo	Serverless
Tamaño del clúster	2X-Pequeño
Terminación automática	Tras 10 minutos de inactividad
Expansión	Conteo de clústeres: activo 1 mín. 1 máx. 1
Canal	Actual (v 2025.15)
Creador	David Ramírez

Cargamos los datos de la misma manera que se ha explicado en el apartado 1.

Microsoft Azure | databricks

Buscar datos, cuadernos, recientes y más... % + P

+ Nuevo

- Workspace
- Recientes
- Catálogo**
- Flujos de trabajo
- Cómputo
- Marketplace

SQL

- Editor de SQL
- Consultas
- Tableros
- Genie
- Alertas
- Historial
- Almacenes de SQL

Ingeniería de datos

Ejecuciones

Explorador de catálogos

Enviar comentarios

Escribir para buscar...

- Comparticiones de Delta recibidas
 - samples
- Heredado
 - hive_metastore
 - default
 - goalscorers
 - listings**
 - results

default >

listings

Información general Datos de muestra Detalles Permisos

Acerca de esta tabla

Propietario	David Ramírez
Fuente de datos	Delta
Última actualización	hace 6 segundos
Tamaño	2.2MiB, 1 archivo

Descripción

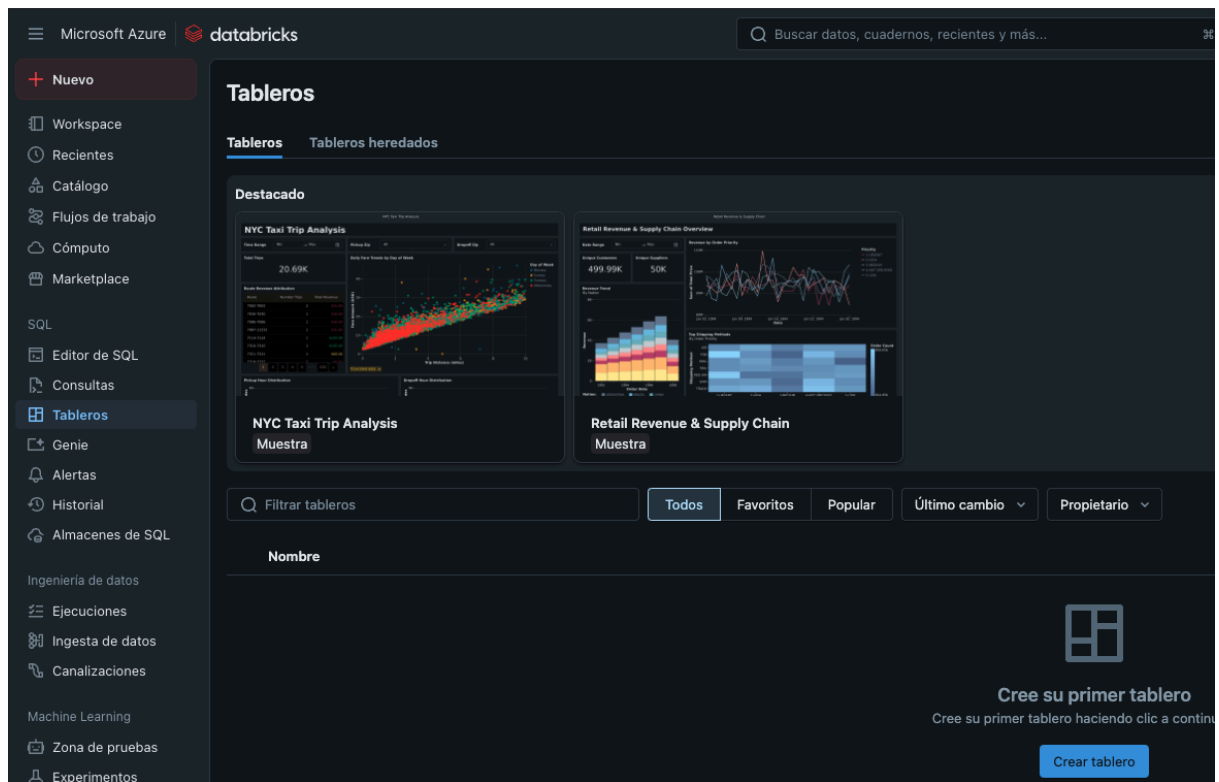
Created by the file upload UI

Filtrar columnas...

Columna
id

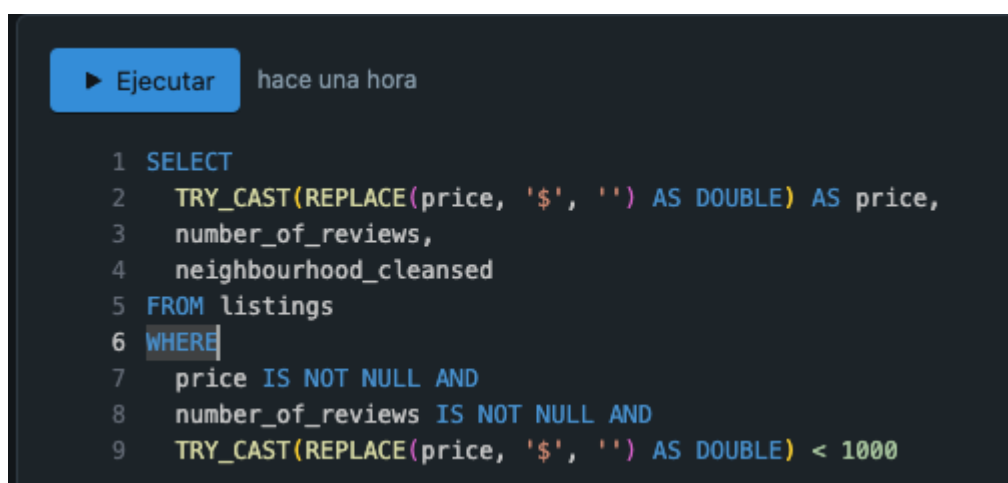
2 - CREAR DASHBOARD

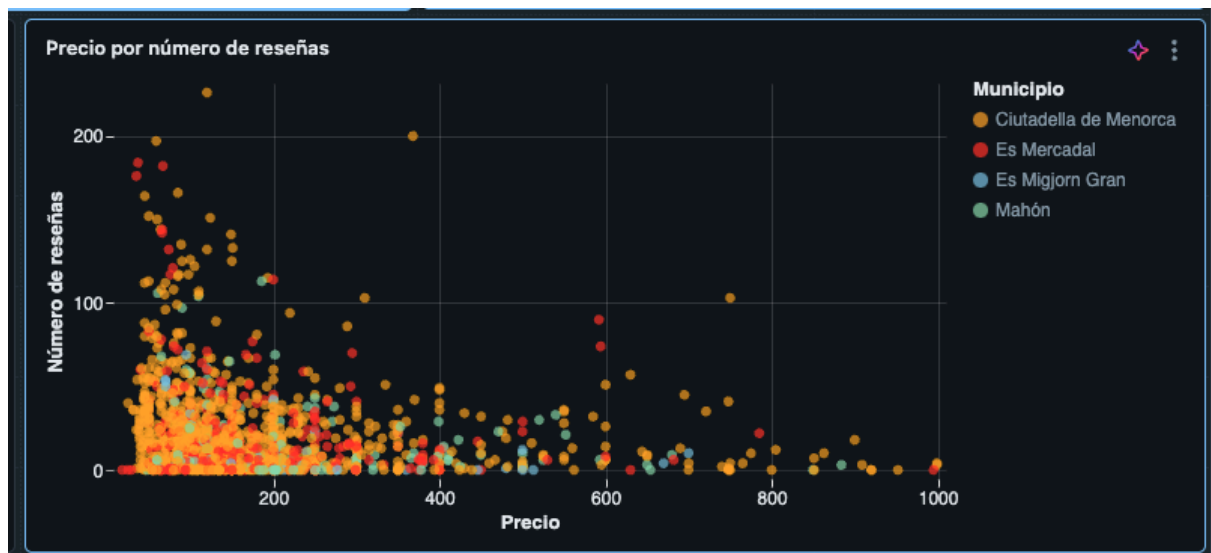
Vamos a proceder a crear el Dashboard, en mi caso llamado “Tablero” ya que está en español. Y pulsamos “Crear tablero”.



Creamos un dashboard y añadimos la consulta SQL para filtrar los datos. Después asignamos estos datos a un gráfico. Este proceso se repite para cada gráfico, a continuación mostraré imágenes de las consultas y los gráficos.

1. Precio por número de reseñas - DIAGRAMA DE DISPERSIÓN





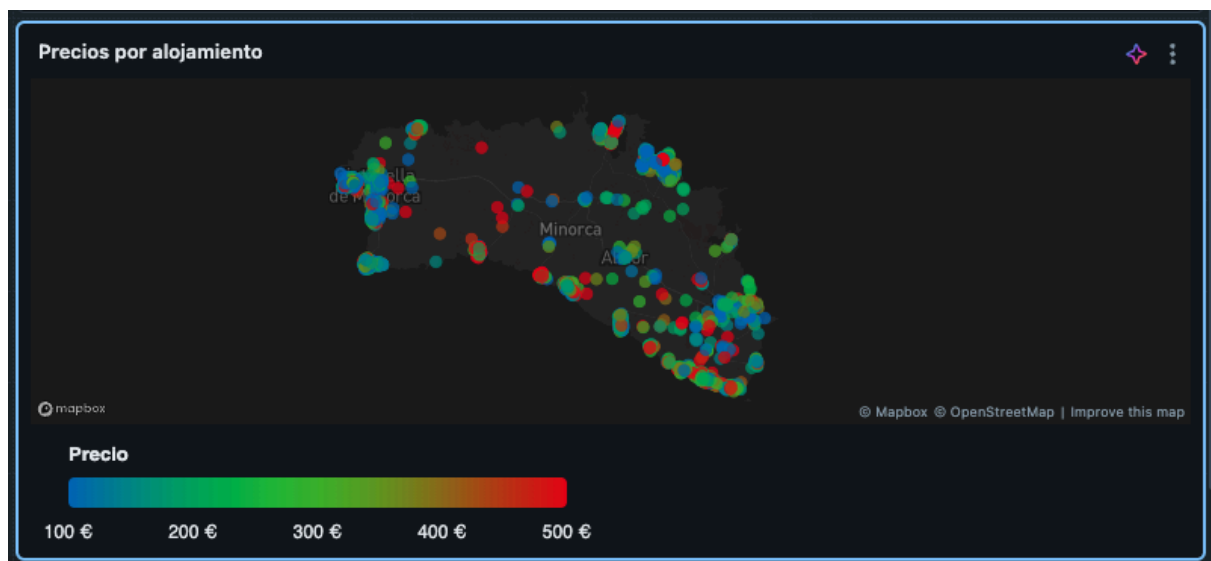
2. Mapa de alojamientos por precio - MAPA DE PUNTOS

Ejecutar hace 8 horas

```

1 SELECT
2   id,
3   latitude,
4   longitude,
5   TRY_CAST(REPLACE(REPLACE(price, '$', ''), ',', '')) AS FLOAT) AS price
6 FROM listings
7 WHERE latitude IS NOT NULL AND longitude IS NOT NULL AND price IS NOT NULL;

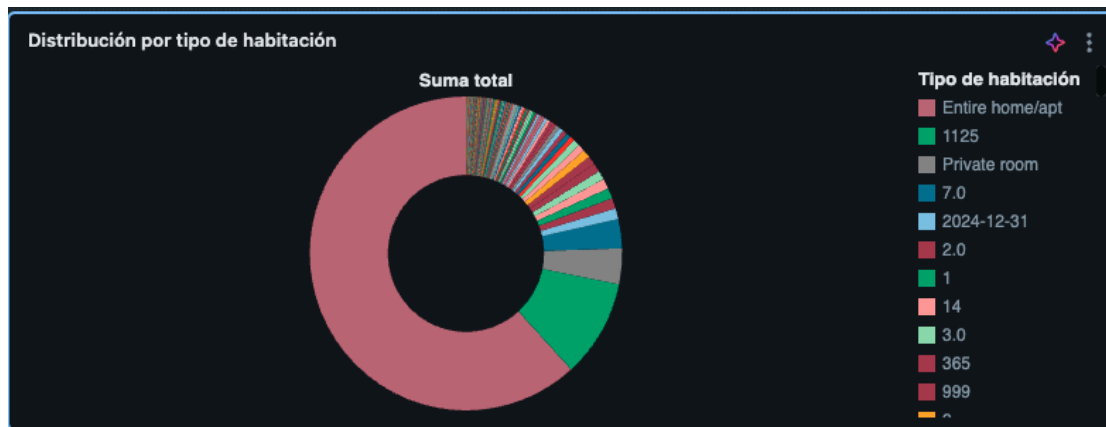
```



3. Distribución de tipos de habitación - GRÁFICO DE ANILLO

```
▶ Ejecutar hace 7 horas

1 SELECT
2   room_type,
3   COUNT(*) AS total,
4   ROUND(100.0 * COUNT(*) / SUM(COUNT(*) OVER ()), 2) AS percentage
5 FROM listings
6 WHERE room_type IS NOT NULL
7 GROUP BY room_type
8 ORDER BY total DESC
```



4. Alojamientos por host - TABLA

```
▶ Ejecutar hace 30 minutos

1 SELECT host_name, host_id, COUNT(*) AS total_listings
2 FROM listings
3 WHERE host_id IS NOT NULL AND host_name IS NOT NULL
4 GROUP BY host_id, host_name
5 ORDER BY total_listings DESC
```

Alojamientos por host

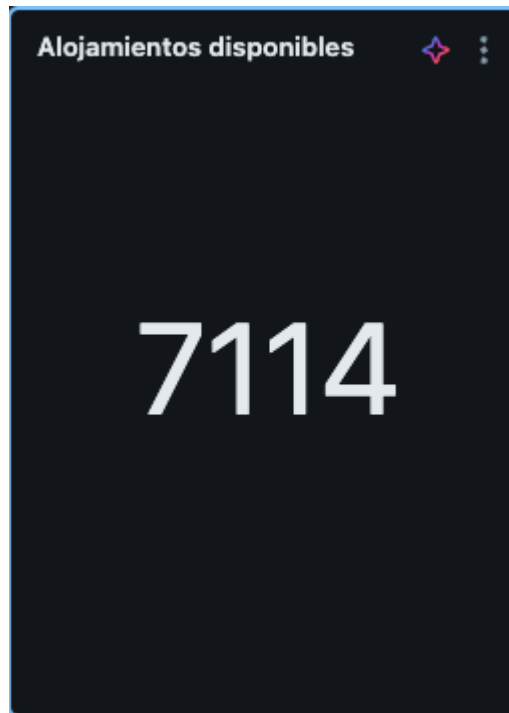
Nombre del host	Alojamientos totales
World2Meet	62
OrfilaSternal	58
t	51
Som	49
Holidays	45
MauterVillas	44

< 1 2 3 4 5 ... 65 >

5. Total de alojamientos disponibles - CONTADOR

```
▶ Ejecutar hace 26 minutos

1 SELECT COUNT(*) AS total_alojamientos
2 FROM listings
```



RESULTADO FINAL

