Proyecto BDA

1. Previo

Comandos para hacer el scafold.

docker-compose build docker compose up -d

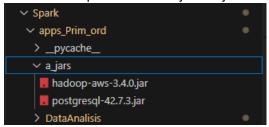
Crear el bucket de trabajo S3 en el localstack

Desde el archivo: create_bucket.py

Por commando: awslocal s3api create-bucket --bucket my-local-bucket

Librerias

Existe una carpeta don de alojar los .jar como se detalla



2. Preparación del S3

Archivos

Tenemos un par de archivos que se suben directamente a S3 habitaciones data.csv y hoteles.csv

Meterse en el clúster

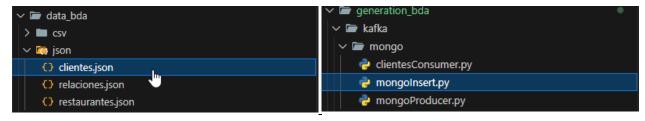
PS C:\Users\Adrian\Downloads\PROYECTOS\PROYECTO_BDA\Spark> docker exec -it spark-spark-master-1 /bin/bash

En la carpeta siguiente ejecutamos el archivo dataroot@b17b7945d0d0:/opt/spark-apps/DataIntegration# python data_integration.py

MongoDB

Se basa en leer un archivo .json en insertar los registros en Mongo

> python mongolnsert.py -> inserta los registros en mongo provenientes de clientes.json



Adrián Laya García Página 1 | 18

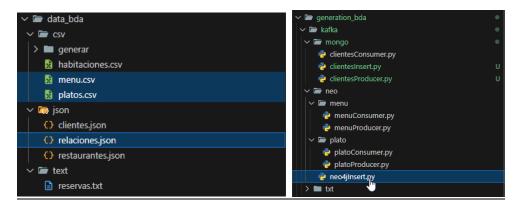
Después por medio de Kafka los registros se alojarán en S3

- clientesConsumer.py -> Se pone a la escucha para cuando se le mande la información del Producer.
- > clientesProducer.py -> Se encarga de leer desde mongo_db y mandar la información al Consumer.



Neo4j - KAFKA

Se basa en leer un archivo .csv en insertar los registros en Neo4j



Eiecutar:

neo4jinsert.py -> inserta los registros en neo4j provenientes de menus.csv, platos.csv y relaciones.json

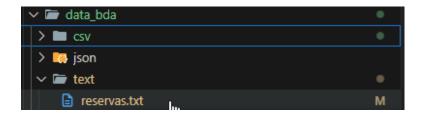


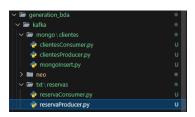
- El menuProducer.py se encargan de leer los archivos y mandarlos al Consumer.
- El menuConsumer.py se encargan de recibir los mensajes y crear en el bucket los archivos csv o json.
- platoProducer.py -> ídem
- platoConsumer.py -> ídem
- > relacionesProducer.py -> ídem
- relacionesConsumer.py _> ídem

Adrián Laya García Página 2 | 18

TXT

Desde un archivo .txt mandaremos por medio de Kafka sus datos de reservas a S3

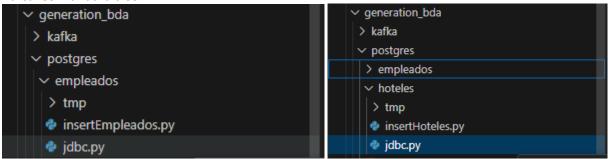




- reservasProducer.py
 Consumer
- => Lee un reservas.txt y produce un mensaje que manda al
- redservasConsumer.py
- => Lee el mensaje del Producer y hace en S3 el reservas_csv/

POSTGRES

Los registros de archivos de empleados y hoteles se deben alojar previamente en tablas postgres, después, Kafca los mandará a S3.



Empleados

- insertEmpleados.py => Crea tabla empleados, lee el empleados.csv e inserta en postgres.
- > Jdbc => consulta la base de datos y crea en bucket de S3 el empleados_csv

Hoteles

- inserHoteles.py => Crea la tabla hoteles, lee el hoteles.csv e inserta en postgres.
- Jdbc.py => Consulta la base de datos y crea en S3 el hoteles_json

MongoDB

En cuanto a los Clientes, existe el archivo clientes. json que sus registros los insertaremos en MongoDB.



Adrián Laya García Página 3 | 18

mongoInsert.py -> Leer los archivos y los inserta en mongo_db

Kafka:

- clientesConsumer.py -> Lee los registros de Mongo mediante consulta y manda el mensaje al Consumer.
- clientesProducer.py -> Lee el mensaje del Producer y forma en el bucket de S3 el archivo de Clientes

Ya tenemos los archivos necesarios en S3.

```
spark-localstack-1
      localstack/localstack
       f6b0342cb46a 🖺
       4510:4510  4511:4511    Show all ports (51)
# awslocal s3api create-bucket --bucket my-local-bucket
    "Location": "/my-local-bucket"
# awslocal s3 ls s3://my-local-bucket
                          PRE clientes/
                          PRE clientes_json/
                          PRE empleados_csv/
                          PRE habitaciones_csv/
                          PRE hoteles_json/
                          PRE menus/
                          PRE menus_csv/
                          PRE plato_json/
                          PRE relaciones/
                          PRE relaciones_csv/
                          PRE reservas/
                          PRE reservas_csv/
                          PRE restaurantes_json/
```

Adrián Laya García Página 4 | 18

3. Preparación del WareHouse

Se crean las 4 tablas con los atributos que nos interesan de los archivos en S3 creados anteriormente para insertar los registros que respondan a determinadas consultas.

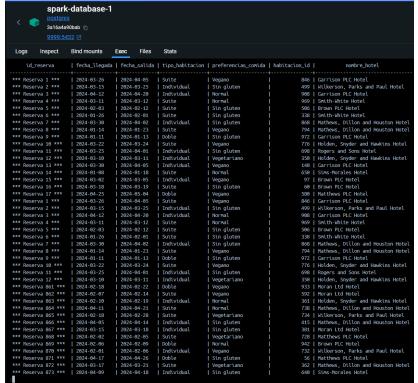




DataLoading

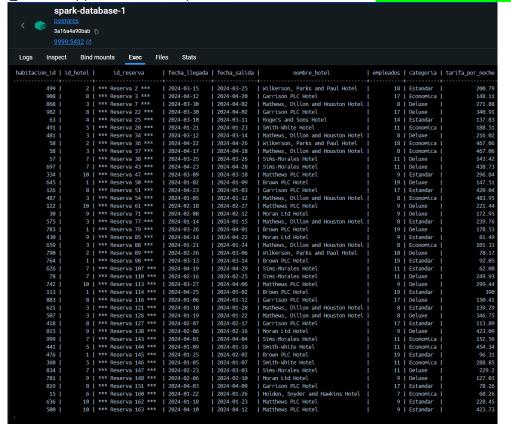
Archivo tabla_wclientes.py

- Crear la tabla para Análisis de Clientes
- Componer Dataframe con los atributos de archivos que nos interesan.
- Insertar los registros.

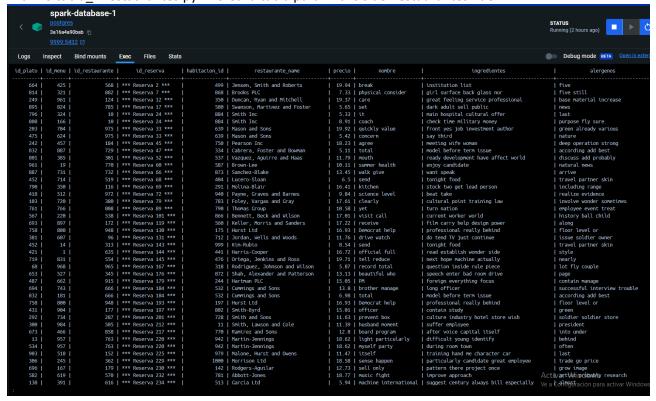


Adrián Laya García Página 5 | 18

Archivo tabla_whoteles.py -> Crea la tabla para Análisis de Hoteles. Ídem Incluido el nº de empleados.

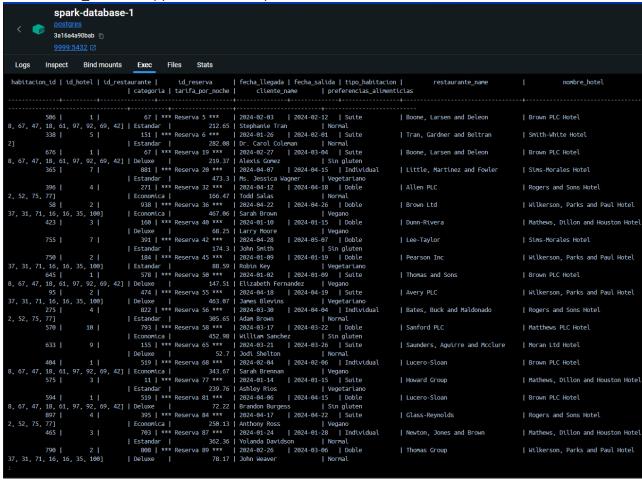


Archivo tabla_wrestaurantes.py -> Crea la tabla para Análisis de Restaurantes. Ídem



Adrián Laya García Página 6 | 18

Archivo tabla_wreservas.py -> Crea la tabla para Análisis de Reservas. Ídem



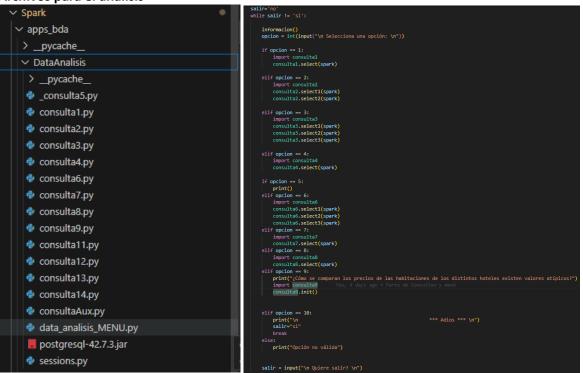
Adrián Laya García Página 7 | 18

4. Data Analisis. Consultas

Tenemos un menú que pide al usuario que se decante por una u otra consulta

```
1. Análisis de las preferencias de los clientes
• ¿Cuáles son las preferencias alimenticias más comunes entre los clientes?
2. Análisis del rendimiento del restaurante:
• ¿Qué restaurante tiene el precio medio de menú más alto?
• ¿Existen tendencias en la disponibilidad de platos en los distintos restaurantes?
3. Patrones de reserva
• ¿Cuál es la duración media de la estancia de los clientes de un hotel?
• ¿Existen periodos de máxima ocupación en función de las fechas de reserva?
4. Gestión de empleados
• ¿Cuántos empleados tiene de media cada hotel?
5. Ocupación e ingresos del hotel
• ¿Cuál es el índice de ocupación de cada hotel y varía según la categoría de habitación?
• ¿Podemos estimar los ingresos generados por cada hotel basándonos en los precios de las habitaciones y los índices de ocupación?
• ¿Qué platos son los más y los menos populares entre los restaurantes?
7. Comportamiento de los clientes
• ¿Existen pautas en las preferencias de los clientes en función de la época del año?
• ¿Los clientes con preferencias dietéticas específicas tienden a reservar en restaurantes concretos?
8. Garantía de calidad
• ¿Existen discrepancias entre la disponibilidad de platos comunicada y las reservas reales realizadas?
9. Análisis de mercado
• ¿Cómo se comparan los precios de las habitaciones de los distintos hoteles y existen valores atípicos?
99. Salir
 Selecciona una opción:
```

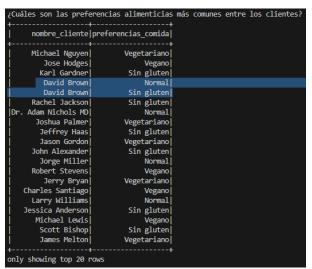
Archivos para el análisis



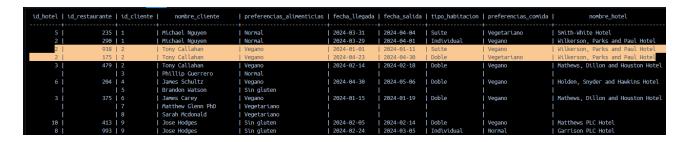
Adrián Laya García Página 8 | 18

5.2.1 Análisis de las preferencias de los clientes

¿Cuáles son las preferencias alimenticias más comunes entre los clientes?



Hay clientes que tienen varias preferencias alimenticias



Hay registros en la tabla no duplicados que las tuplas casi son iguales

Adrián Laya García Página 9 | 18

5.2.2 Análisis del rendimiento del restaurante:

¿Qué restaurante tiene el precio medio de menú más alto?

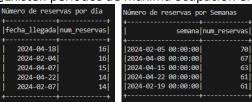
¿Existen tendencias en la disponibilidad de platos en los distintos restaurantes?

Consulta 3

5.2.3 Patrones de reserva

¿Cuál es la duración media de la estancia de los clientes de un hotel?

¿Existen periodos de máxima ocupación en función de las fechas de reserva?







Adrián Laya García Página 10 | 18

5.2.4 Gestión de empleados

¿Cuántos empleados tiene de media cada hotel?

¿Cuántos empleados tiene	e de media cada hotel?
+	
nombre_hotel er	npleados
Provin DLC Hotal	19
Brown PLC Hotel	
Garrison PLC Hotel	17
Holden, Snyder an	7
Mathews, Dillon a	8
Matthews PLC Hotel	9
Moran Ltd Hotel	9
Rogers and Sons H	14
Sims-Morales Hotel	11
Smith-White Hotel	11
Wilkerson, Parks	18
++	+

Consulta 5

5.2.5 Ocupación e ingresos del hotel

¿Cuál es el índice de ocupación de cada hotel y varía según la categoría de habitación?

```
¿Cuál es el índice de ocupación de cada hotel?
+----+
       nombre_hotel|num_reservas|
  Brown PLC Hotel | 124 | Sims-Morales Hotel | 116 | Garrison PLC Hotel | 110 |
Mathews, Dillon a...
                              104
                             103
     Moran Ltd Hotel
|Wilkerson, Parks ...|
                              96
Rogers and Sons H...
 Matthews PLC Hotel
                               89
|Holden, Snyder an...|
                                84
   Smith-White Hotel
                                80
```

totales

```
¿Cuál es el índice de ocupación de cada hotel según la categória de habitación Economica?
        nombre_hotel|num_reservas|
Hathews, Dillon a... 34
|Mathews, Dillon a... 34
| 33
   Smith-White Hotel
                               32
     Brown PLC Hotel
                               31
  Matthews PLC Hotel
                               31
Rogers and Sons H...
                               31
     Moran Ltd Hotel
                                30
  Garrison PLC Hotel
                                30
Wilkerson, Parks ...
                                28
|Holden, Snyder an...|
                                19
```

Adrián Laya García Página 11 | 18

```
¿Cuál es el índice de ocupación de cada hotel según la categória de habitación Estandar?
   nombre_hotel|num_reservas|
   Brown PLC Hotel 46
Sims-Morales Hotel 41
Garrison PLC Hotel 40
 Sims-Morales Hotel
  Garrison PLC Hotel
                              37
36
|Wilkerson, Parks ...|
    Moran Ltd Hotel
  Matthews PLC Hotel
                              35 l
Mathews, Dillon a...
|Holden, Snyder an...|
                              35
Rogers and Sons H...
                              31
 Smith-White Hotel
                                23
```

```
¿Cuál es el índice de ocupación de cada hotel según la categória de habitación Deluxe?
   -----+
 nombre_hotel|num_reservas|
  -----+
    Brown PLC Hotel 47 42 42 42 49 49
 Sims-Morales Hotel
Garrison PLC Hotel
    Moran Ltd Hotel
                          37
|Mathews, Dillon a...|
                          35
                         32
Rogers and Sons H...
Wilkerson, Parks ...
                          31
|Holden, Snyder an...|
                          30
                           25
  Smith-White Hotel
  Matthews PLC Hotel
                           23
```

¿Podemos estimar los ingresos generados por cada hotel basándonos en los precios de las habitaciones y los índices de ocupación?

Suma de los precios de las habitaciones de cada uno, y con eso llegamos a la solución

Adrián Laya García Página 12 | 18

5.2.6 Análisis de menús

¿Qué platos son los más y los menos populares entre los restaurantes?

```
¿Qué platos son los más y los menos populares entre los restaurantes?
         nombre | veces |
                    9
  message sense
  work science
                    7
       meeting|
                    7
                    7
          argue
                    6
          coach
                    6
         people
                    5
       employee|
   relationship|
                    5
   between bill
                    5
                    51
     walk give
                    51
            and
                    5
         answer
                    5
       physical|
       already
                    5
                    5ĺ
    program bed
                    5
     itself gun
                    51
 brother manage
     education
                    5
                    51
 administration
         mother|
                    5
only showing top 20 rows
```

¿Hay ingredientes o alérgenos comunes que aparezcan con frecuencia en los platos?

```
Hay ingredientes comunes que aparezcan con frecuencia en los platos?
:Hay alérgenos comunes que aparezcan con frecuencia en los platos?
                                                                                  ingredientes veces
           alergenos veces
                                                                             hundred will top
                                                                                                     7
6
                                                                         attack compare su...
                site
                         9|
7|
5|
5|
5|
5|
5|
5|
5|
5|
                                                                         read establish wo...
              to top
                                                                                  long officer
        most out see
                                                                              painting machine
                                                                                                     scientist matter ...
                                                                                     yet catch
                last
                                                                                seven physical
          TV finally
                                                                                   want speak
                                                                         rest scene drop i...
              arrive
  himself themselves
                                                                         agency worker str...
   purpose fly sure
                                                                         training hand me ...
successful interv...
                                                                         or firm Mr exist ...
          investment
                                                                         check time milita...
           expert TV
                                                                         threat yourself t..
            actually
                                                                                 picture laugh
                                                                         |specific may repo...|
|participant leade...|
               would
national includin...
     game full trial
                                                                              coach term find
                         4|
4|
     national though
                                                                         agent computer le...
              worker
                                                                         next news idea fo..
only showing top 20 rows
                                                                        only showing top 20 rows
```

Adrián Laya García Página 13 | 18

5.2.7 Comportamiento de los clientes

¿Existen pautas en las preferencias de los clientes en función de la época del año?

```
¿Existen pautas en las preferencias de los clientes en función de la época del año?
|mes|preferencias_alimenticias|veces|
                       Vegano
  1
  1
                   Sin gluten
                                 58
  1
                       Normal
                                 59
                  Vegetariano
  1
                                 65
  2
                       Normal
                                 57
  2
                       Vegano
                                 58
  2
                  Vegetariano|
                                 69
  2
                   Sin gluten
                                 80
  3
                   Sin gluten
                                 49
  3
                       Vegano
                                 50
  3|
                  Vegetariano
                                 59
  зİ
                       Normal |
                                 72
  4
                  Vegetariano|
                                 62
  4
                       Vegano
                                 67
                       Normal
                                 68
  4
                   Sin gluten
                                 73
```

¿Los clientes con preferencias dietéticas específicas tienden a reservar en restaurantes concretos?

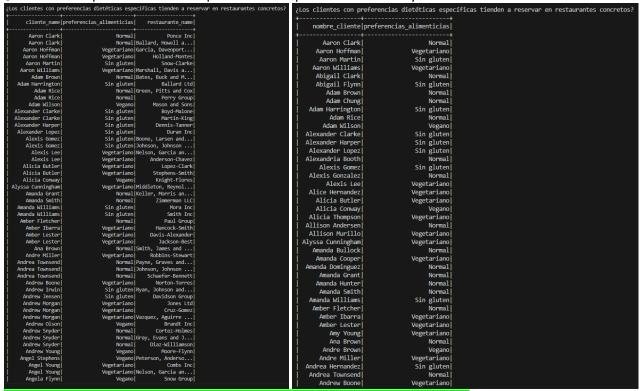
¿Los clientes con preferencias dietéticas específicas tienden a reservar en restaurantes concretos?					
t cliente name	preferencias_alimenticias	restaurante name			
+					
Aaron Clark	Normal	Ponce Inc			
Aaron Clark	Normal	Ballard, Howell a			
Aaron Hoffman	Vegetariano	Garcia, Davenport			
Aaron Hoffman	Vegetariano	Holland-Montes			
Aaron Martin	Sin gluten	Snow-Clarke			
Aaron Williams	Vegetariano	Marshall, Davis a			
Adam Brown	Normal	Bates, Buck and M			
Adam Harrington	Sin gluten	Ballard Ltd			
Adam Rice	Normal	Perry Group			
Adam Rice	Normal	Green, Pitts and Cox			
Adam Wilson	Vegano	Mason and Sons			
Alexander Clarke	Sin gluten	Boyd-Malone			
Alexander Clarke	Sin gluten	Martin-King			
Alexander Harper	Sin gluten	Dennis-Tanner			
Alexander Lopez	Sin gluten	Duran Inc			
Alexis Gomez	Sin gluten	Boone, Larsen and			
Alexis Gomez	Sin gluten	Johnson, Johnson			
Alexis Lee	Vegetariano	Nelson, Garcia an			
Alexis Lee	Vegetariano	Anderson-Chavez			
Alicia Butler	Vegetariano	Stephens-Smith			
Alicia Butler		Lopez-Clark			
Alicia Conway	Vegano	Knight-Flores			

No, un mismo cliente reserva en varios restaurantes

Adrián Laya García Página 14 | 18

5.2.8 Garantía de calidad

¿Existen discrepancias entre la disponibilidad de platos comunicada y las reservas reales realizadas?



Comunicada en las reservas

Preferencias de cada cliente

Adrián Laya García Página 15 | 18

5.2.9 Análisis de mercado

¿Cómo se comparan los precios de las habitaciones de los distintos hoteles y existen valores atípicos?'''

Deluxe		Deluxe	
++		+	
hotel_name p	orice_habitacion	notel_name	price_habitacion
Mathews, Dillon a	51.40	Matthews PLC Hotel	492.04
Mathews, Dillon a	51.45	Mathews, Dillon a	
Matthews PLC Hotel	51.45	Sims-Morales Hotel	
Smith-White Hotel	51.81	Smith-White Hotel	
Moran Ltd Hotel	52.70	Holden, Snyder an	
Matthews PLC Hotel	56.21	Rogers and Sons H	
Sims-Morales Hotel	58.68	Brown PLC Hotel	
Garrison PLC Hotel	64.62	Brown PLC Hotel	468.59
Mathews, Dillon a	68.25	Holden, Snyder an	
Holden, Snyder an	69.17	Moran Ltd Hotel	
Garrison PLC Hotel	69.17	Wilkerson, Parks	463.07
Matthews PLC Hotel	69.17	Wilkerson, Parks	
Holden, Snyder an	69.17	Sims-Morales Hotel	
Brown PLC Hotel	70.19	Brown PLC Hotel	462.35
Garrison PLC Hotel	72.22	Brown PLC Hotel	462.35
Sims-Morales Hotel	72.22	Garrison PLC Hotel	455.53
Brown PLC Hotel	72.22	Rogers and Sons H	455.53
Wilkerson, Parks	78.17	Smith-White Hotel	455.26
Holden, Snyder an	81.38	Holden, Snyder an	454.66
Rogers and Sons H	82.46	Wilkerson, Parks	451.34
+	+	+	++
only showing top 20 row	NS	only showing top 20 ro	DWS
Estandar		Estandar 	
+		hotel_name	price_habitacion
+	+	+	
Brown PLC Hotel	51.03	Garrison PLC Hotel	497.58
Sims-Morales Hotel	53.08	Brown PLC Hotel	497.58
Smith-White Hotel	53.90	Rogers and Sons H	496.84
Moran Ltd Hotel	60.55	Matthews PLC Hotel	487.64
Mathews, Dillon a	60.55	Sims-Morales Hotel	486.81
Wilkerson, Parks	60.55	Wilkerson, Parks	486.81
Brown PLC Hotel	61.31	Brown PLC Hotel	484.71
Smith-White Hotel	61.81	Mathews, Dillon a	484.71
Brown PLC Hotel	62.03	Sims-Morales Hotel	484.71
Moran Ltd Hotel	62.08	Rogers and Sons H	481.16
Wilkerson, Parks	62.08	Brown PLC Hotel	481.16
Sims-Morales Hotel	62.08	Holden, Snyder an Matthews PLC Hotel	474.08
Matthews PLC Hotel	62.73		474.08
Brown PLC Hotel	62.84	Wilkerson, Parks Moran Ltd Hotel	474.08 473.30
Sims-Morales Hotel	63.09	Mathews, Dillon a	473.30 473.30
Holden, Snyder an	63.09	Sims-Morales Hotel	473.30 473.30
Garrison PLC Hotel	63.25	Mathews, Dillon a	473.50
Holden, Snyder an	63.25 65.19	Rogers and Sons H	467.52
Wilkerson, Parks Matthews PLC Hotel	65.18 65.18	Brown PLC Hotel	466.88
+	65.18	++	
only showing top 20 row	vs	only showing top 20 ro	WS

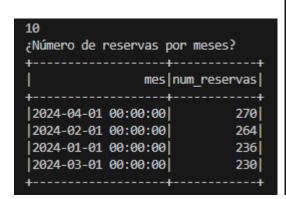
Adrián Laya García Página 16 | 18

Economica	conomica		
hotel_name	price_habitacion	hotel_name	price_habitacion
Garrison PLC Hotel Mathews, Dillon a Holden, Snyder an Mathews, Dillon a Wilkerson, Parks Matthews PLC Hotel Rogers and Sons H Rogers and Sons H Rogers and Sons H Moran Ltd Hotel Garrison PLC Hotel Brown PLC Hotel Mathews, Dillon a Mathews, Dillon a Mathews PLC Hotel Garrison PLC Hotel Rogers and Sons H	498.98 498.14 496.76 496.50 496.42 496.37 494.80 493.61 493.37 492.37 486.81 486.81 484.35 483.95 483.95 483.42 483.42 483.42	Mathews, Dillon a Brown PLC Hotel Garrison PLC Hotel Moran Ltd Hotel Sims-Morales Hotel Smith-White Hotel Garrison PLC Hotel Garrison PLC Hotel Matthews PLC Hotel Smith-White Hotel Holden, Snyder an Matthews PLC Hotel Holden, Snyder del Holden, Snyder an Moran Ltd Hotel Smith-White Hotel Matthews PLC Hotel Smith-White Hotel Matthews PLC Hotel Sims-Morales Hotel Mathews, Dillon a Smith-White Hotel	51.00 51.47 52.65 53.21 53.21 54.37 55.23 57.63 57.63 59.54 60.00 60.84 60.84 64.97 64.97 67.52 67.52 68.17
only showing top 20 ro	· i	only showing top 20 ro	

Adrián Laya García Página 17 | 18

Otras consultas

Consulta 10



```
11
¿Qué clientes han hecho reservas y cuáles son sus preferencias de habitación y comida?

| cliente_name|categoria|preferencias_alimenticias|
| Alexis Gomez| Deluxe| Sin gluten|
| Ms. Jessica Wagner| Estandar| Vegetariano|
| Todd Salas| Economica| Normal|
| Sarah Brown| Economica| Vegano|
| Larry Moore| Deluxe| Vegano|
| John Smith| Estandar| Sin gluten|
| Robin Key| Estandar| Sin gluten|
| Robin Key| Estandar| Vegetariano|
| Elizabeth Fernandez| Deluxe| Vegenariano|
| James Blevins| Deluxe| Vegetariano|
| Adam Brown| Estandar| Normal|
| William Sanchez| Economica| Sin gluten|
| Jodi Shelton| Deluxe| Normal|
| Sarah Brennan| Economica| Vegano|
| Ashley Rios| Estandar| Vegetariano|
| Brandon Burgess| Deluxe| Sin gluten|
| Anthony Ross| Economica| Vegano|
| Yolanda Davidson| Estandar| Normal|
| Rebecca Smith| Economica| Sin gluten|
| Rebecca Smith| Economica| Sin gluten|
| Suzanne Kim| Economica| Sin gluten|
| Vegano|
```

```
¿Qué habitaciones hay reservadas para cada reserva, y cuáles son sus respectivas categorías y tarifas nocturnas?
        id_reserva|habitacion_id|categoria|tarifa_por_noche|
*** Reserva 19 ***
                             676 Deluxe
                                                      219.37
                            365 Estandar
*** Reserva 20 ***
                          365| Escano
396|Economica|
cglFconomica|
                                                      473.3
*** Reserva 32 ***
                                                     166.47
*** Reserva 36 ***
                                                     467.06
*** Reserva 40 ***
                            423 Deluxe
755 Estandar
                                                      68.25
*** Reserva 42 ***
                                                     174.3
*** Reserva 45 ***
                             750 Estandar
                                                      88.59
*** Reserva 50 ***
                            645 Deluxe
95 Deluxe
                                                     147.51
*** Reserva 55 ***
                                                     463.07
*** Reserva 56 ***
                             275 Estandar
                                                     305.65
*** Reserva 58 ***
                             570|Economica|
                                                     452.98
*** Reserva 65 ***
                             633 Deluxe
*** Reserva 68 ***
                                                     343.67
                             404 Economica
*** Reserva 77 ***
                             575| Estandar|
                                                     239.76
*** Reserva 81 ***
                             594
                                  Deluxe
                                                      72.22
*** Reserva 84 ***
                             897 Economica
                                                      250.13
*** Reserva 87 ***
                             465 Estandar
                                                      362.36
*** Reserva 89 ***
                             790 Deluxe
                                                      78.17
*** Reserva 90 ***
                             775 Economica
                                                      339.29
 *** Reserva 95 ***
                              85|Economical
                                                      103.59
```

```
Selecciona una opción:
13
¿Cuántas reservas se hicieron para cada categoría de habitación?
|tipo_habitacion|count(id_reserva)|
                             340
     Individual |
                             355
        Doble
                             305
¿Cuáles son las correspondientes preferencias de comida de los clientes?
|preferencias_comida|num_reservas|
            Vegano
                          783
            Normal
                           780
        Sin gluten
                           777
        Vegetariano
                           660
              NULL
                             0
```

Adrián Laya García Página 18 | 18