

Comandos para hacer el scafold.

docker-compose build

docker compose up -d

localstack

create\_bucket.py

# -> Crea el bucket de trabajo en S3

1.1 Archivos

Meterse en el cluster



docker exec -it b17b7945d0d028dfc3b28a0892fb5ebbfd57e36d63ce2077ccf0b8f2e52dce65 /bin/bash

dataroot@b17b7945d0d0:/opt/spark-apps/DataIntegration#

Ejecutar: python data\_integration.py -> Sube al bucket de S3 de

* habitaciones\_data.csv y
* hoteles.csv

2 MongoDB

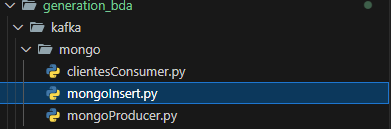
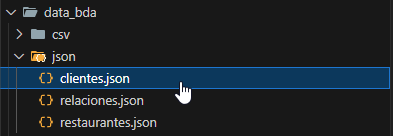
Se basa en leer un archivo json en insertar los registros en Mongo

Hacer uno que lea csv

Hacer los 2 archivos que unifiquen todo y simplifiquen las ejecuciones

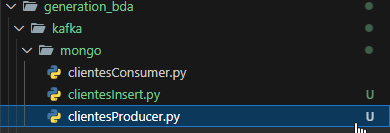
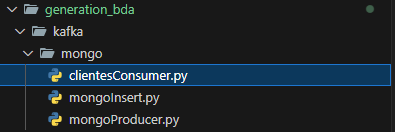
Ejecutar: python mongoInsert.py -> inserta los registros en mongo provenientes de

* clientes.json

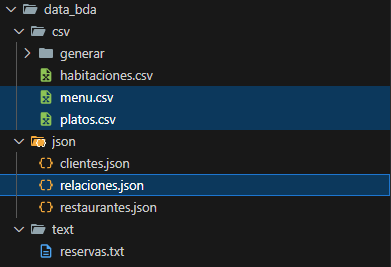
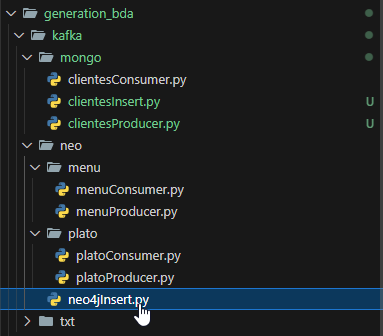
 

clientesConsumer.py -> Se pone a la escucha para cuando se le mande la información del Producer.

clientesProducer.py -> Se encarga de leer desde mongo\_db y mandarla (mensaje estructurado) al Consumer.

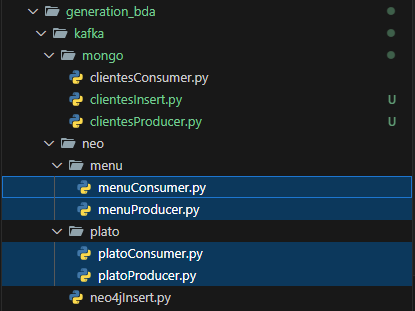
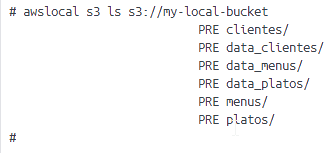
 

3 Neo4j



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 a los consumers.

El menuConsumer.py se encargan de recibir los mensajes y crear en el bucket los archivos csv menus y platos.

platoProducer.py

platoConsumer.py

relacionesProducer.py

relacionesConsumer.py

.TXT

reservasProducer.py => Lee un reservas.txt y produce un mensaje

redservasConsumer.py => Lee el mensaje del Producer y hace en S3 el reservas\_csv/

POSTGRES

Empleados

insertEmpleados.py => Crea tabla empleados, lee el empleados.csv e inserta en postgres.

Jdbc => consulta la base de datos y crea en S3 el empleados\_csv

Hoteles

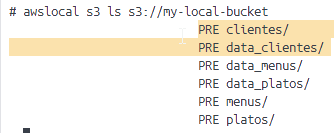
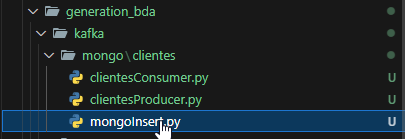
inserHoteles.py => Crea la tabla hoteles, lee el hoteles.csv (que lea un json tambien) e inserta en postgres.

Jdbc.py => Consulta la base de datos y crea en S3 el hoteles\_json

Ya tenemos los archivos.

3.1 Postgres

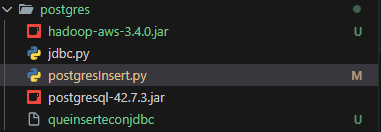
mongoInsert.py



# -> Leer los archivos y los inserta en mongo\_db

clientesConsumer.py

clientesProducer.py

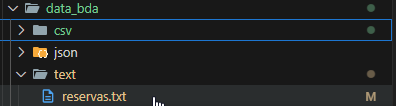
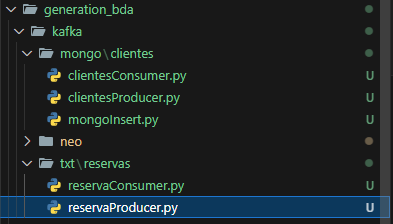
3.1 Postgres empleados. csv y hoteles.csv

postgresinsert.py

# -> inserta en postgres empleados y hoteles

# -> Existe otro archivo que inserta con spark mediante jdbc a postgres desde un csv

3.2 Kafka



Producer

# -> Lee el .txt y manda el mensaje al Consumer

Consumer

# \_> Lee el mensaje e inserta los registros en postgres.

Preparación del WareHouse

Se crean las 4 tablas con los atributos que nos interesan y se juntan dataframes para insertar los registros.

Creación de las tablas:

DataLoading -> tabla\_wclientes.py -> Crea la tabla para Análisis de Clientes.

* Crear la tabla.
* Componer Dataframe con los atributos.
* Insertar los registros

DataLoading -> tabla\_whoteles.py -> Crea la tabla para Análisis de Hoteles.

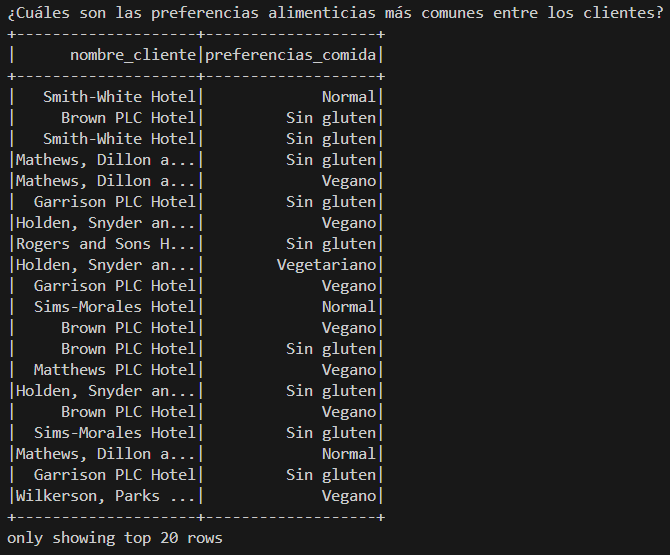
DataLoading -> tabla\_wrestaurantes.py -> Crea la tabla para Análisis de Restaurantes.

DataLoading -> tabla\_wreservas.py -> Crea la tabla para Análisis de Reservas.

Consulta1 OK

5.2.1 Análisis de las preferencias de los clientes

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

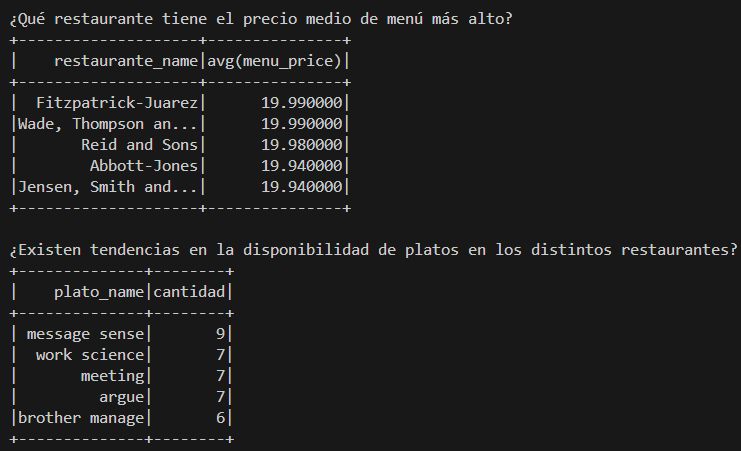


Consulta2 OK

5.2.2 Análisis del rendimiento del restaurante:

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

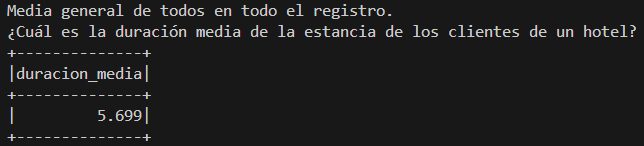
¿Existen tendencias en la disponibilidad de platos en los distintos restaurantes?



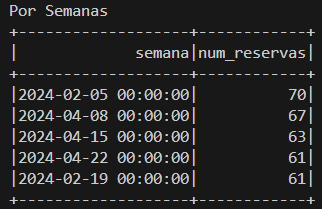
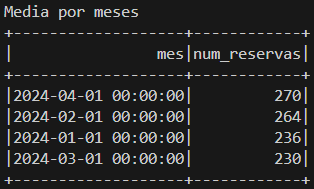
Consulta3

5.2.3 Patrones de reserva

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



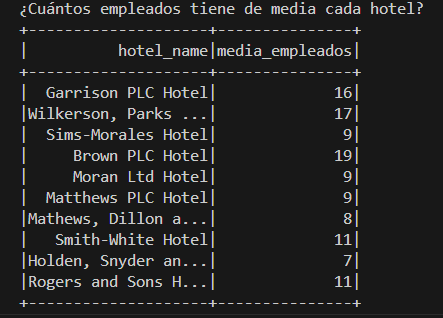
¿Existen periodos de máxima ocupación en función de las fechas de reserva?

 no es media

Consulta4

5.2.4 Gestión de empleados

¿Cuántos empleados tiene de media cada hotel?



Consulta5

5.2.5 Ocupación e ingresos del hotel

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

habitación?

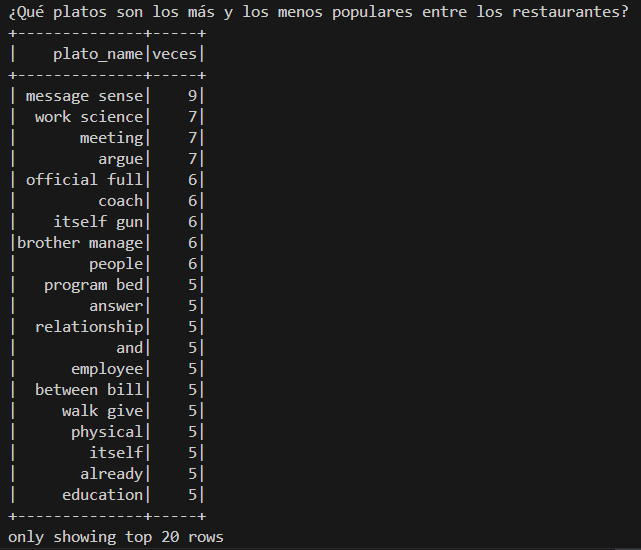
¿Podemos estimar los ingresos generados por cada hotel basándonos en los

precios de las habitaciones y los índices de ocupación?

Consulta6

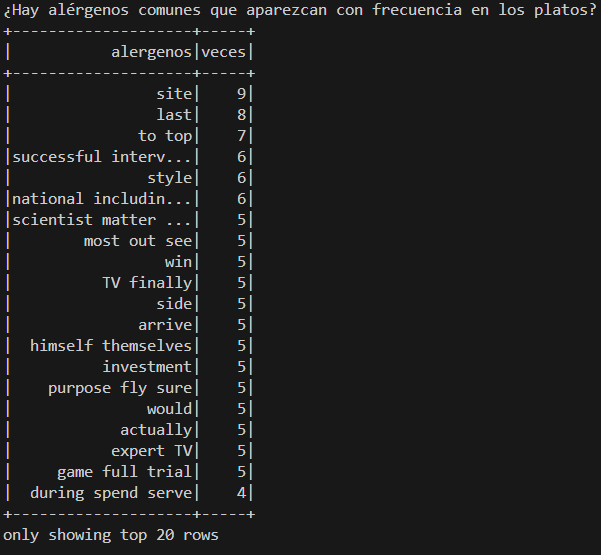
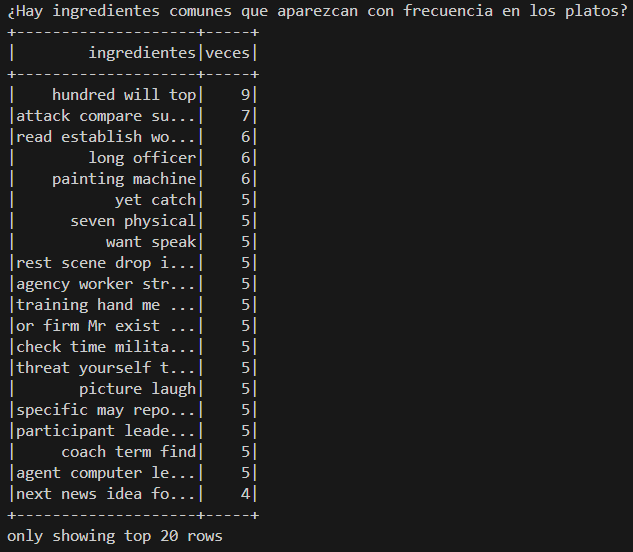
5.2.6 Análisis de menús

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



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

platos?

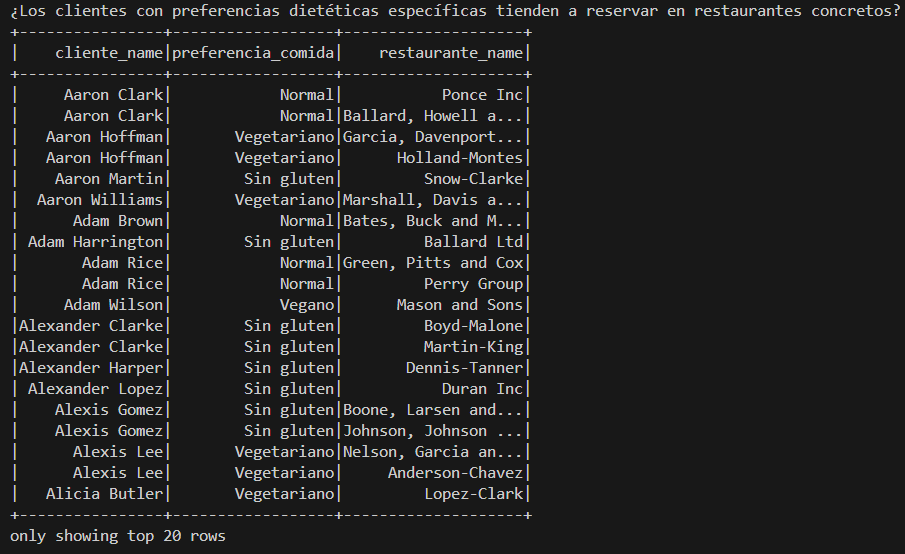
Consulta7

5.2.7 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?



No, un mismo cliente reserva en varios restaurantes

Falta la Consulta8

5.2.8 Garantía de calidad

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

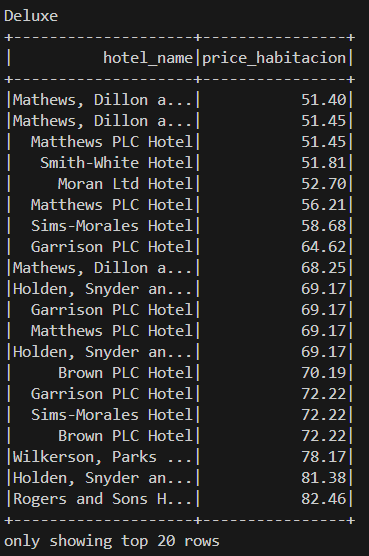
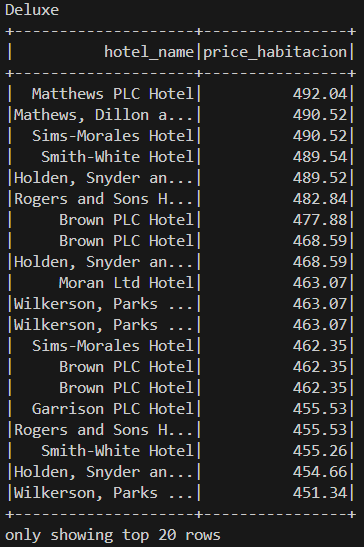
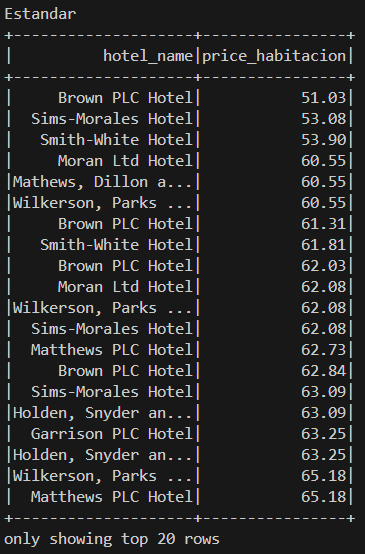
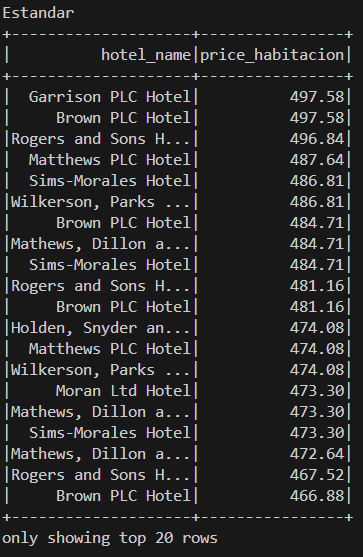
reales realizadas?

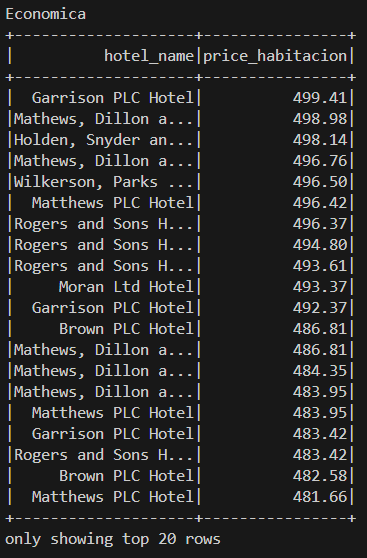
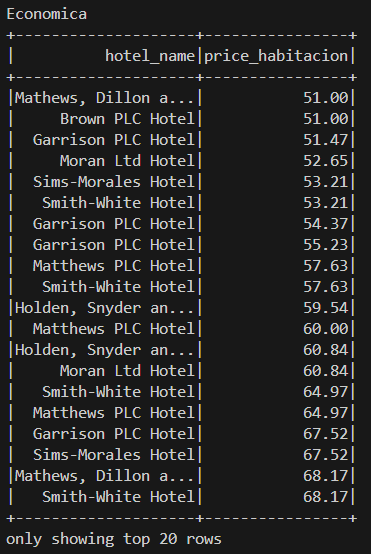
Consulta 9

5.2.9 Análisis de mercado

¿Cómo se comparan los precios de las habitaciones de los distintos hoteles y

existen valores atípicos?'''

 **** **** ****

**** ****

