Installation Instructions

Install important libraries

producer.py

pip install confluent\_kafka

pip install faker

cassandra\_consumer.py

pip install pyspark

pip install cassandra-driver

mysql\_consumer.py

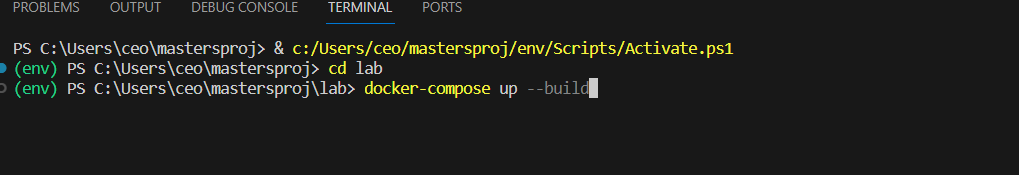
pip install pyspark

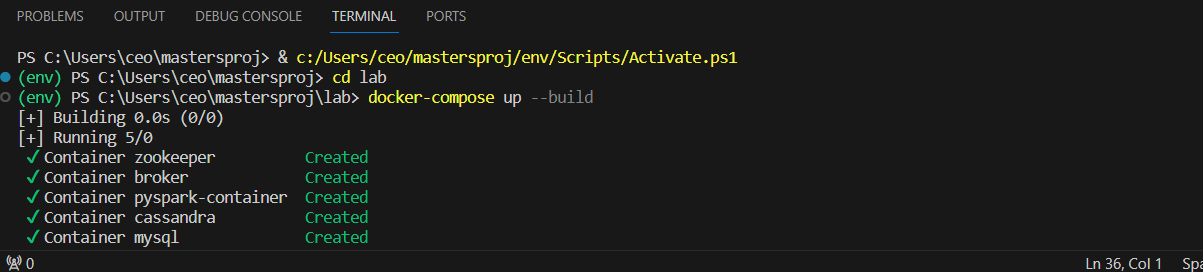
pip install mysql-connector-java==8.0.26

Running the project

PS C:\Users\ceo\mastersproj>

enter into the python environment using: env/Scripts/Activate.ps1





Open a new tab and run the producer.py



pyspark container: cassandra\_consumer.py

(env) PS C:\Users\ceo\mastersproj\lab>

Enter: docker exec -it pyspark-container /bin/bash

(env) PS C:\Users\ceo\mastersproj\lab> docker exec -it pyspark-container /bin/bash

(base) jovyan@bd8ef966c0dc:~$

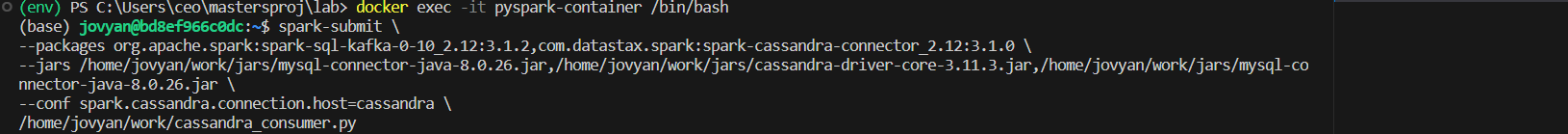
Enter: spark-submit \

--packages org.apache.spark:spark-sql-kafka-0-10\_2.12:3.1.2,com.datastax.spark:spark-cassandra-connector\_2.12:3.1.0 \

--jars /home/jovyan/work/jars/mysql-connector-java-8.0.26.jar,/home/jovyan/work/jars/cassandra-driver-core-3.11.3.jar,/home/jovyan/work/jars/mysql-connector-java-8.0.26.jar \

--conf spark.cassandra.connection.host=cassandra \

/home/jovyan/work/cassandra\_consumer.py



pyspark container: mysql\_consumer.py

(env) PS C:\Users\ceo\mastersproj\lab>

Enter: docker exec -it pyspark-container /bin/bash

(env) PS C:\Users\ceo\mastersproj\lab> docker exec -it pyspark-container /bin/bash

(base) jovyan@bd8ef966c0dc:~$

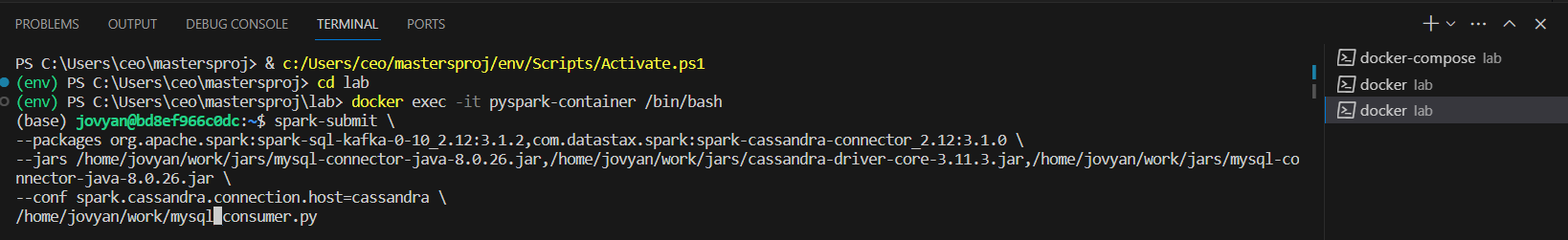
Enter: spark-submit \

--packages org.apache.spark:spark-sql-kafka-0-10\_2.12:3.1.2,com.datastax.spark:spark-cassandra-connector\_2.12:3.1.0 \

--jars /home/jovyan/work/jars/mysql-connector-java-8.0.26.jar,/home/jovyan/work/jars/cassandra-driver-core-3.11.3.jar,/home/jovyan/work/jars/mysql-connector-java-8.0.26.jar \

--conf spark.cassandra.connection.host=cassandra \

/home/jovyan/work/mysql\_consumer.py



cassandra container:

(env) PS C:\Users\ceo\mastersproj\lab>

Enter: docker exec -it cassandra cqlsh -u cassandra -p cassandra

cassandra@cqlsh>

Enter: describe keyspaces;

cassandra@cqlsh>

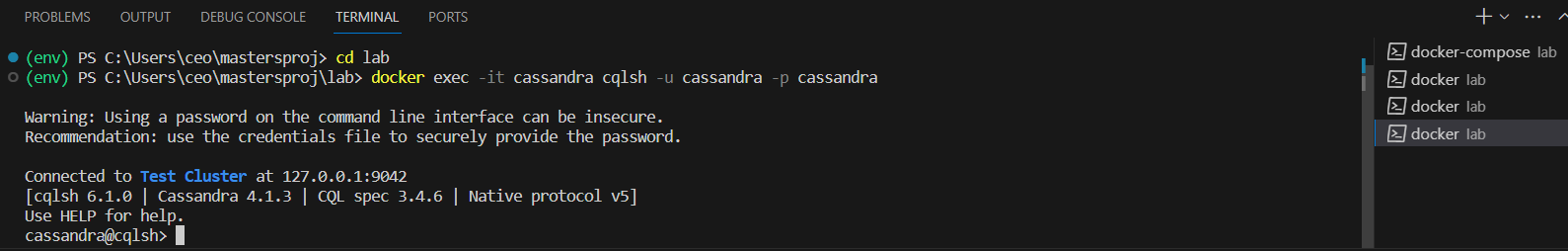
Enter: use car\_parts;

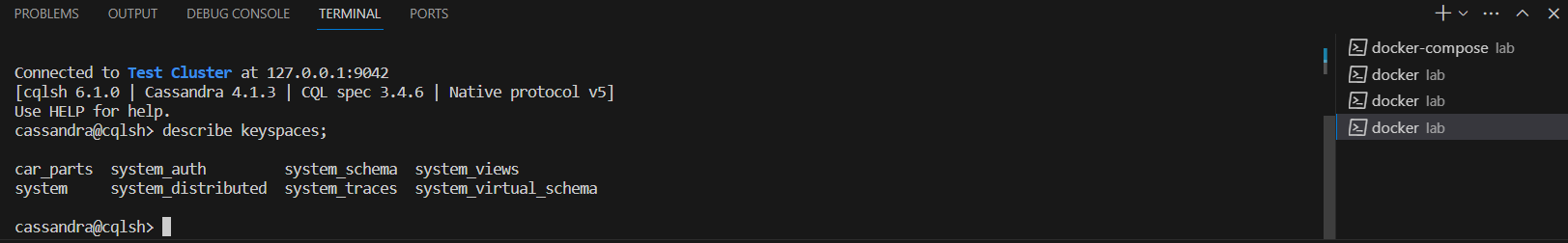
cassandra@cqlsh> use car\_parts;

Enter: describe tables;

cassandra@cqlsh:car\_parts>

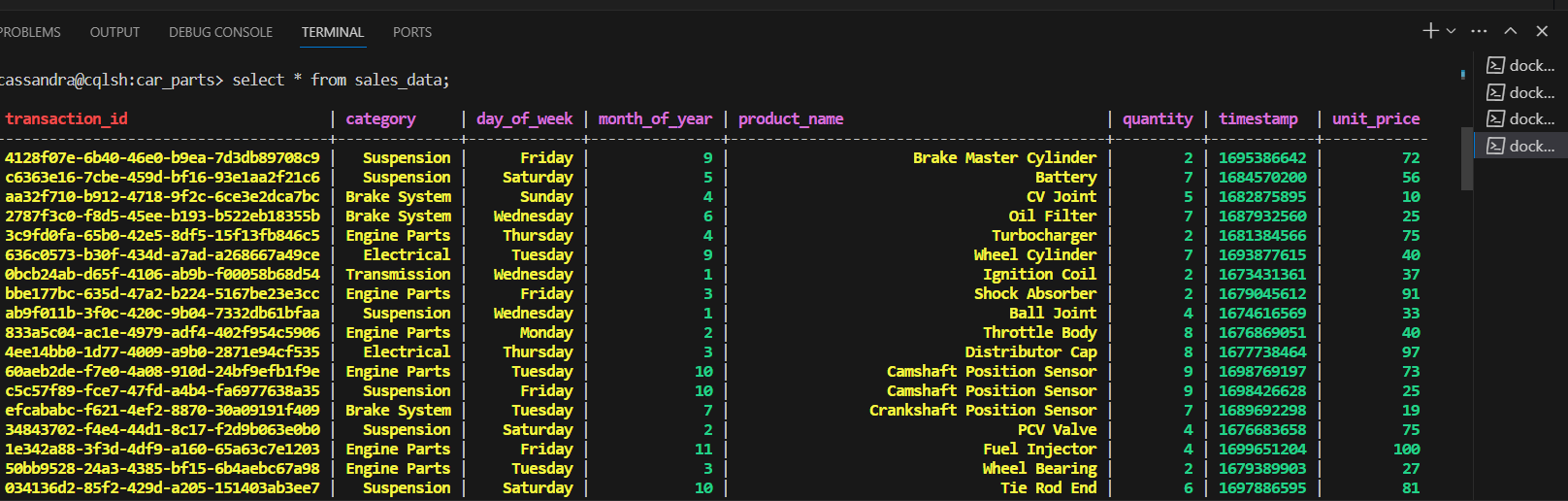
Enter: select \* from sales\_data;











mysql container:

(env) PS C:\Users\ceo\mastersproj\lab>

Enter docker exec -it mysql bash

(env) PS C:\Users\ceo\mastersproj\lab> docker exec -it mysql bash

bash-4.4#

Enter: mysql -u myuser -p

bash-4.4# mysql -u myuser -p

Enter password: mypassword

mysql>

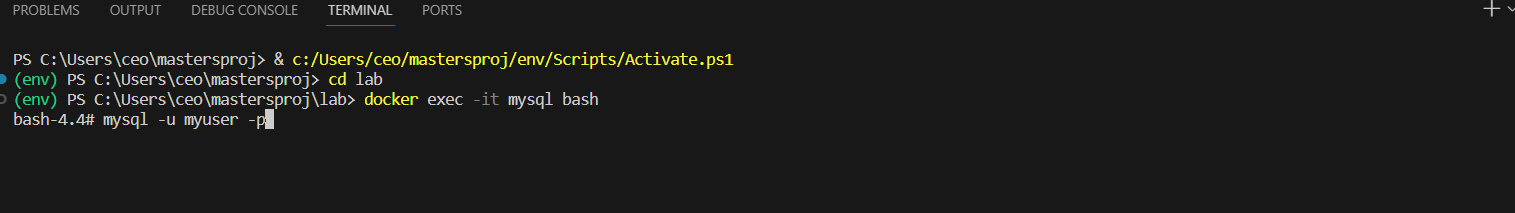
Enter: show databases;

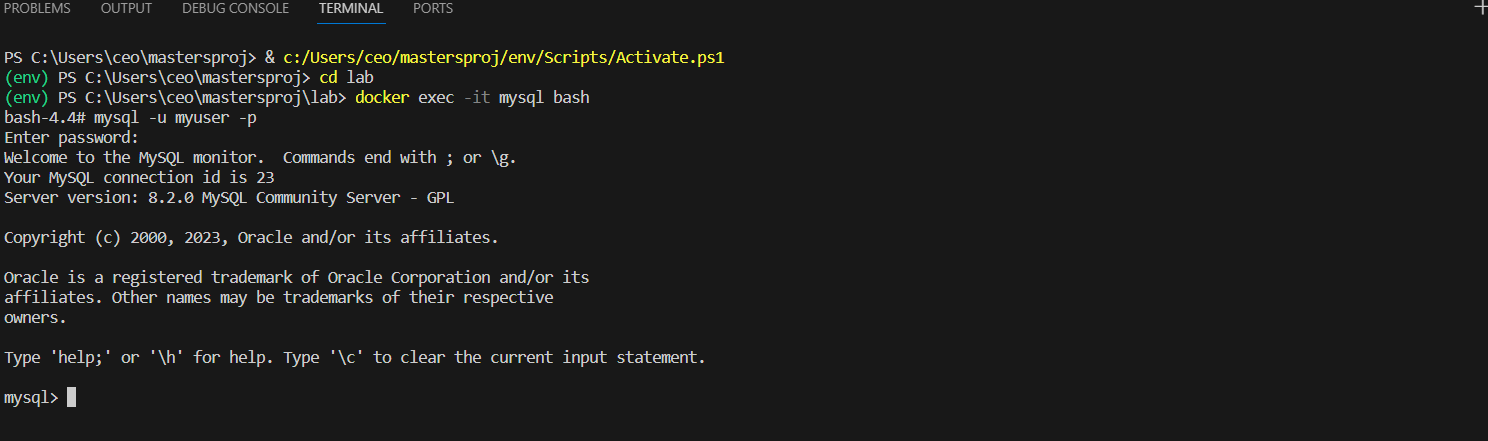
mysql> Enter: use mydatabase;

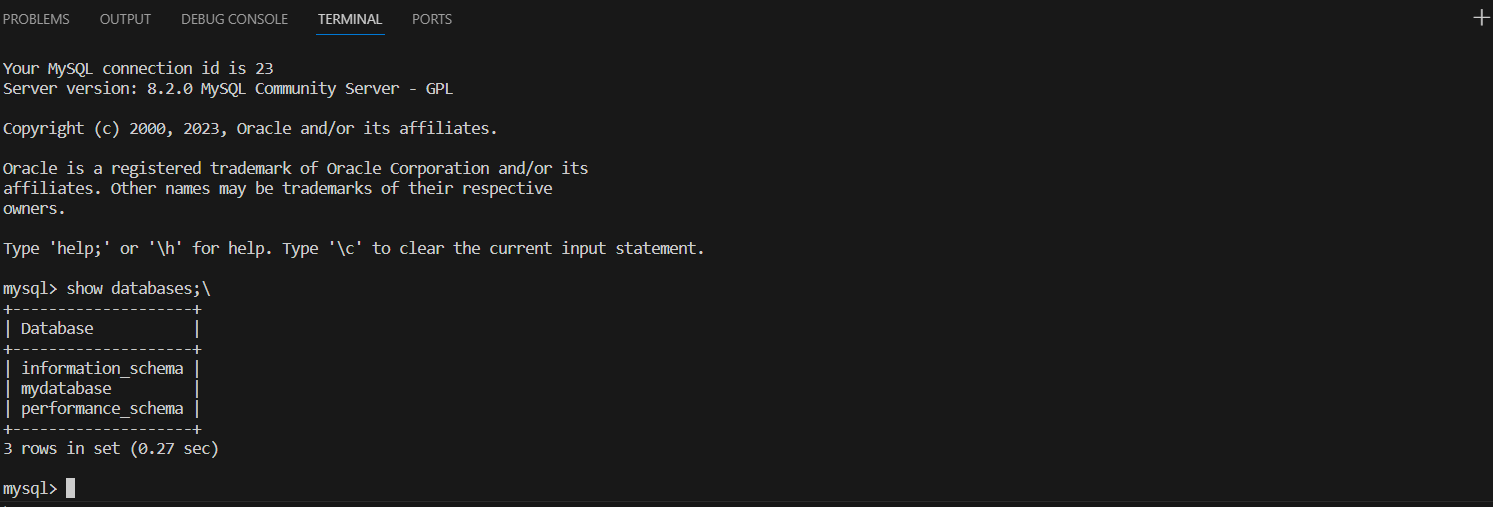
mysql> Enter show tables;

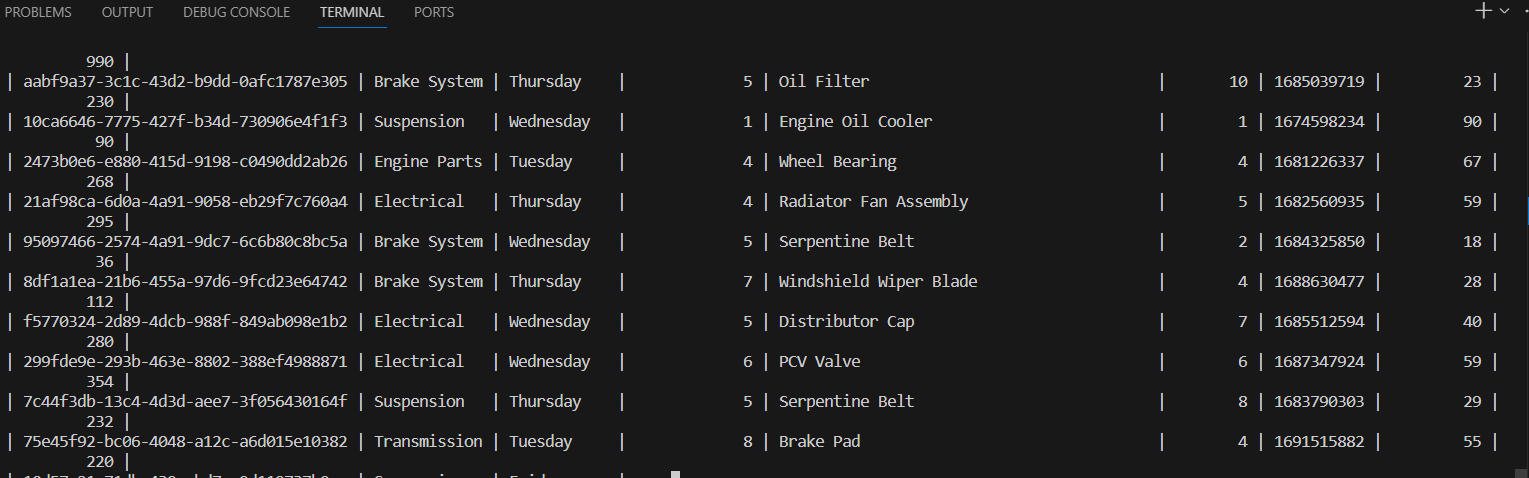
Enter select command

mysql> select \* from sales\_data;





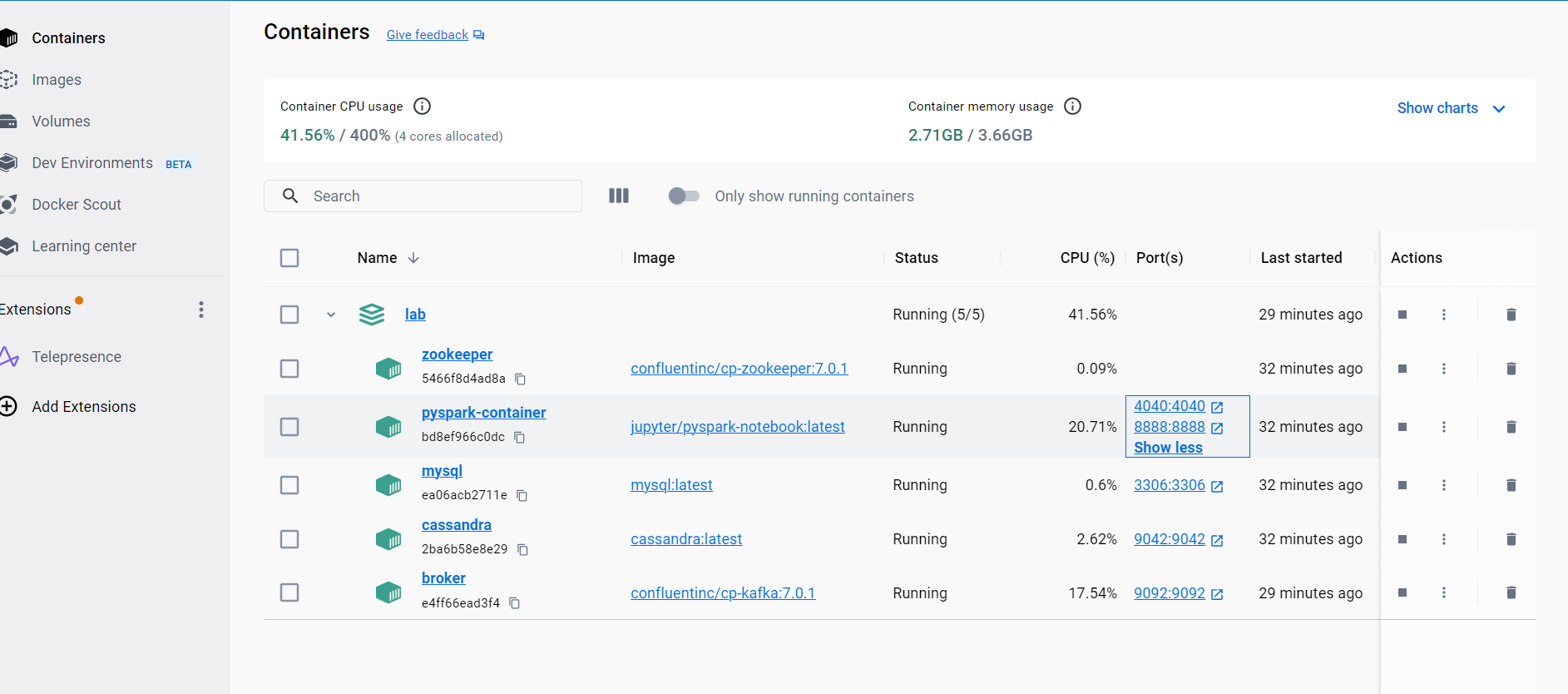


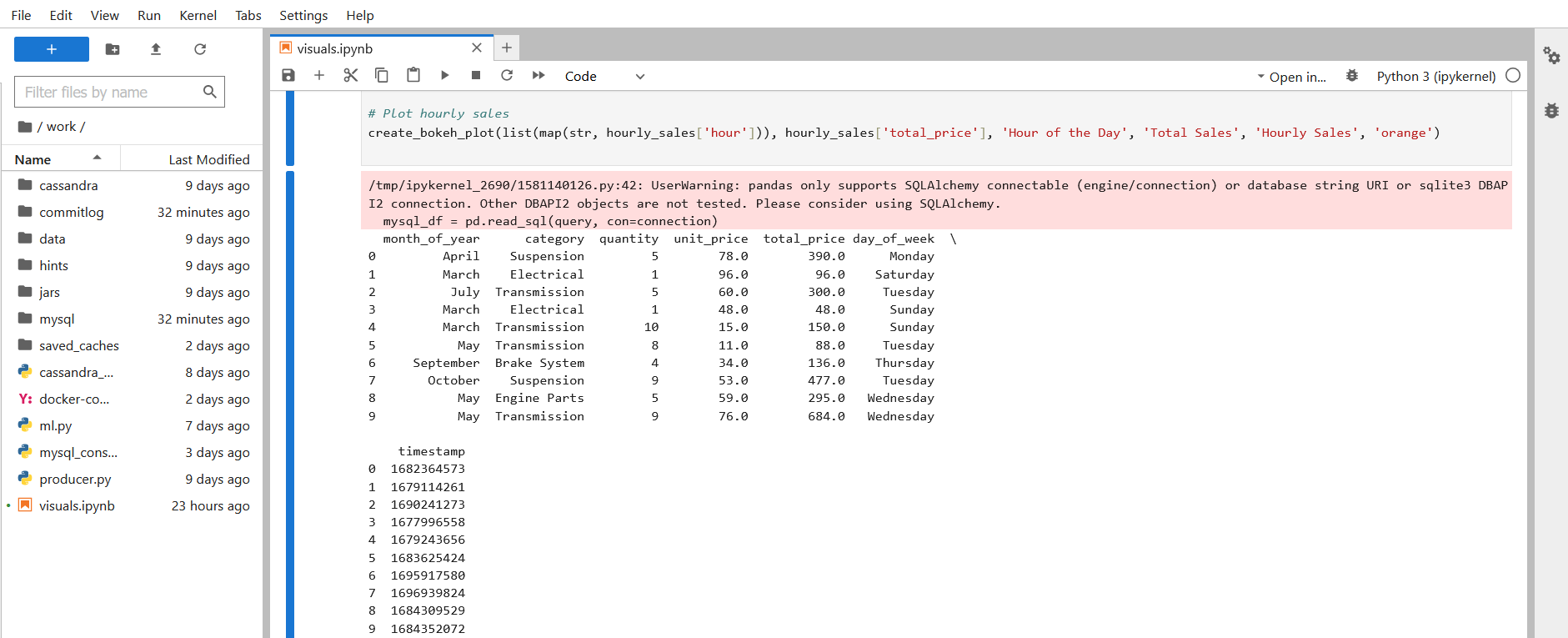


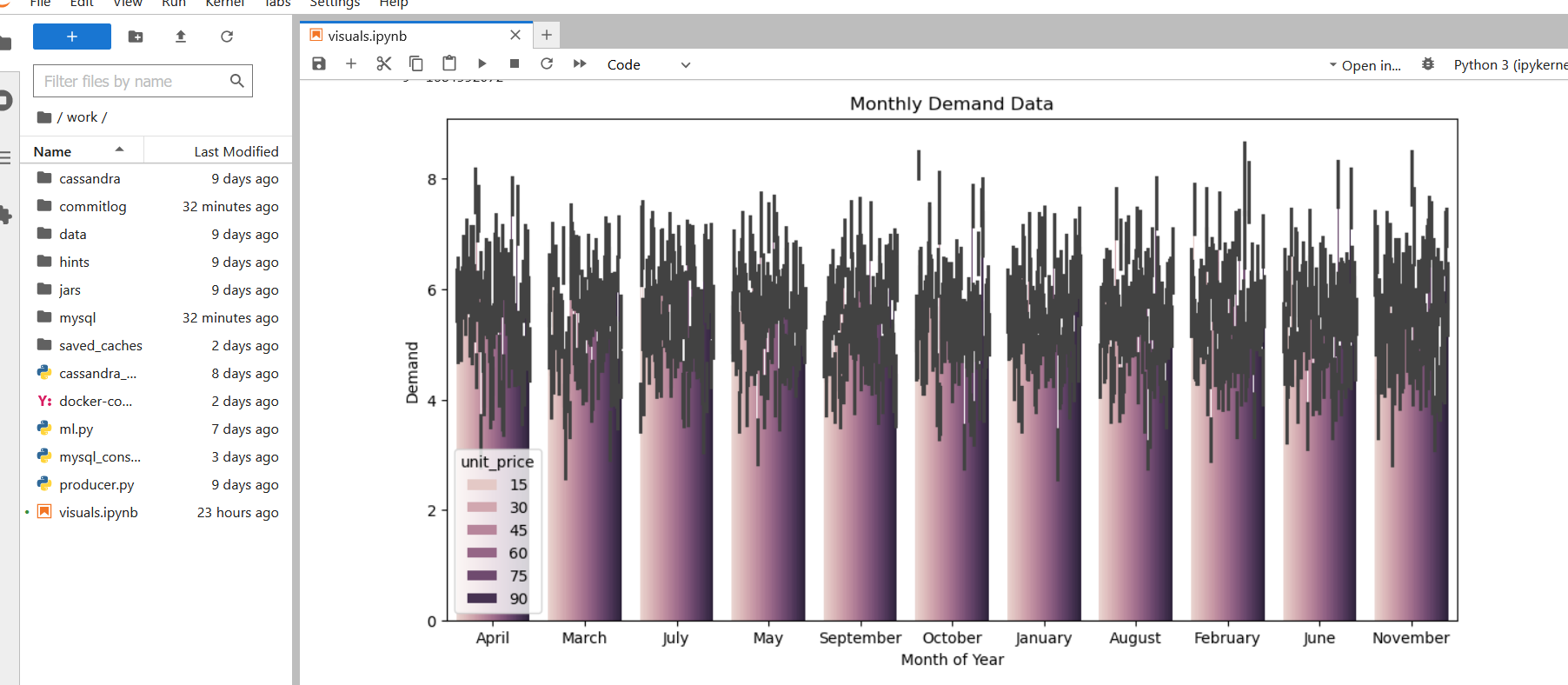
notebook: visuals.py

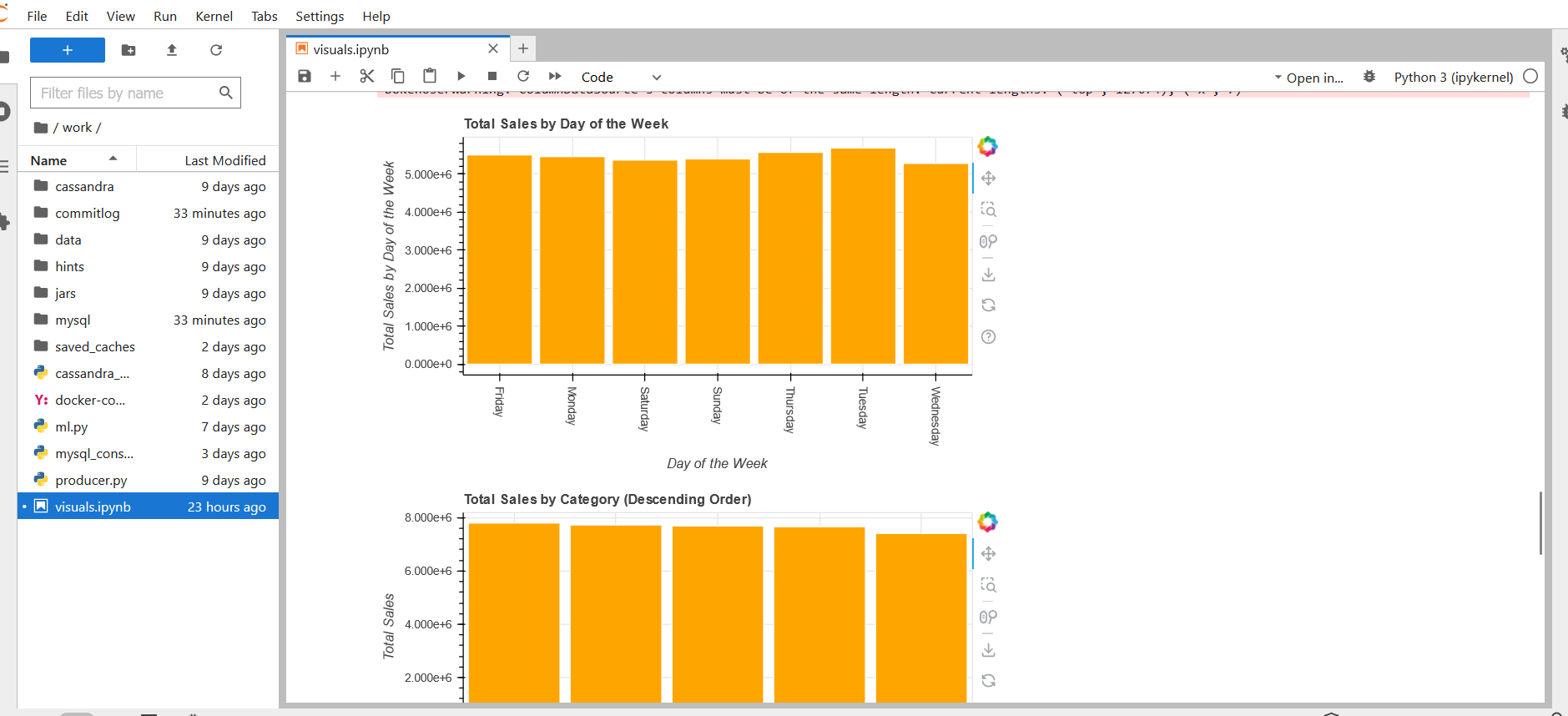
Click on pyspark container port : 8888:8888

click on visuals.ipynb

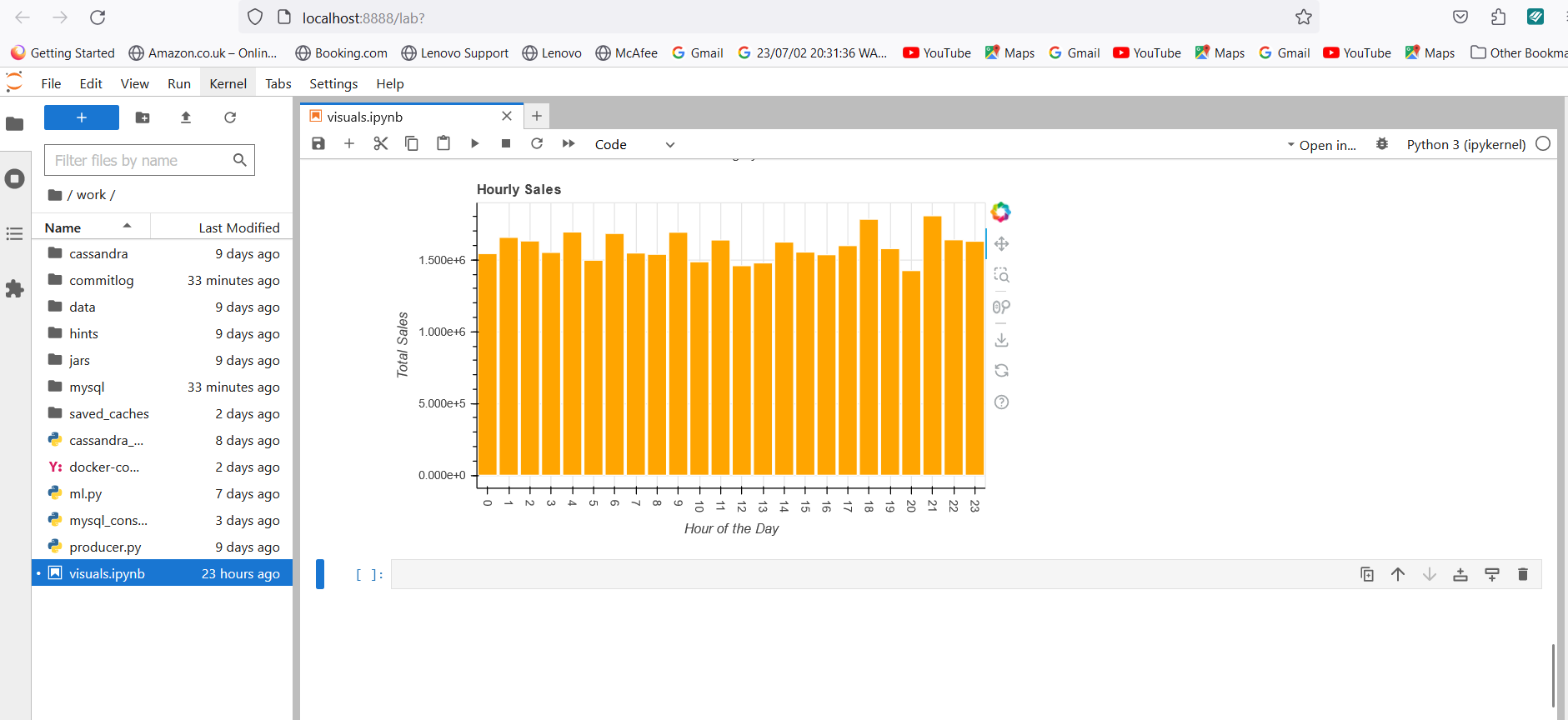












Project github link: https://github.com/omag72/mastersproj