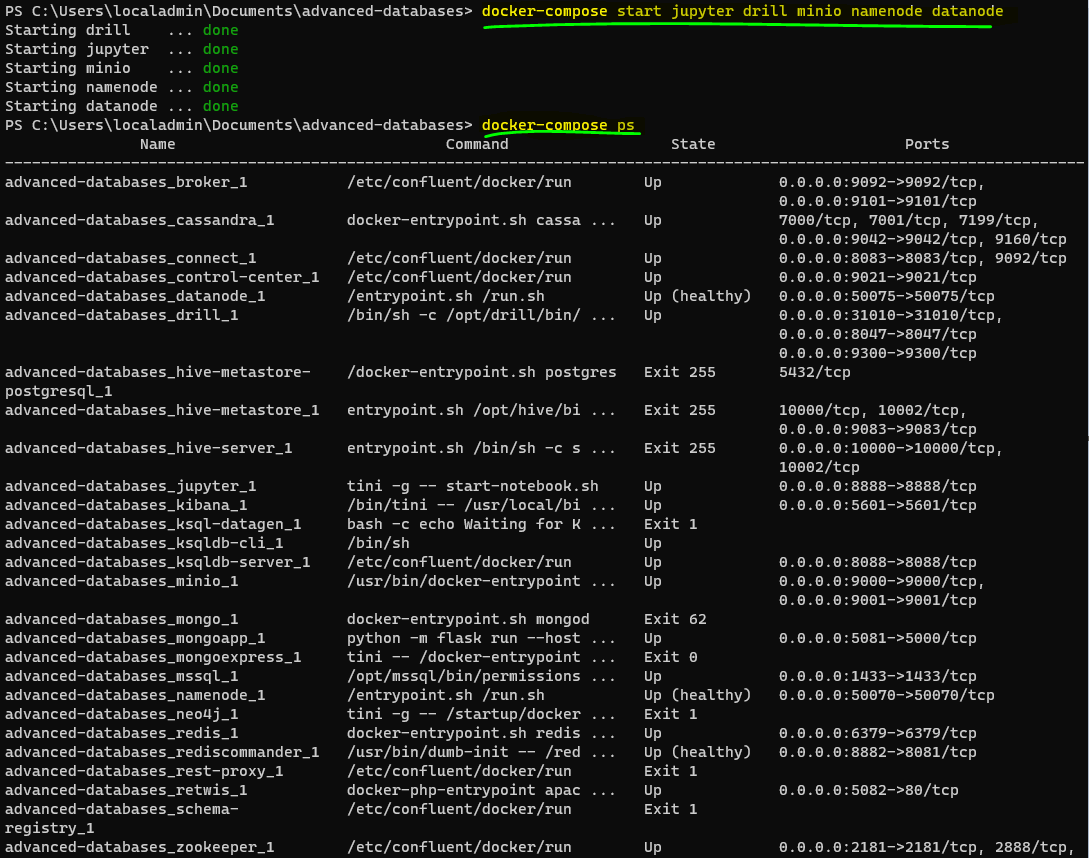
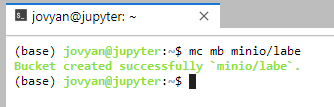
Schemeless SQL: Drill

Before the question, just to upload files:

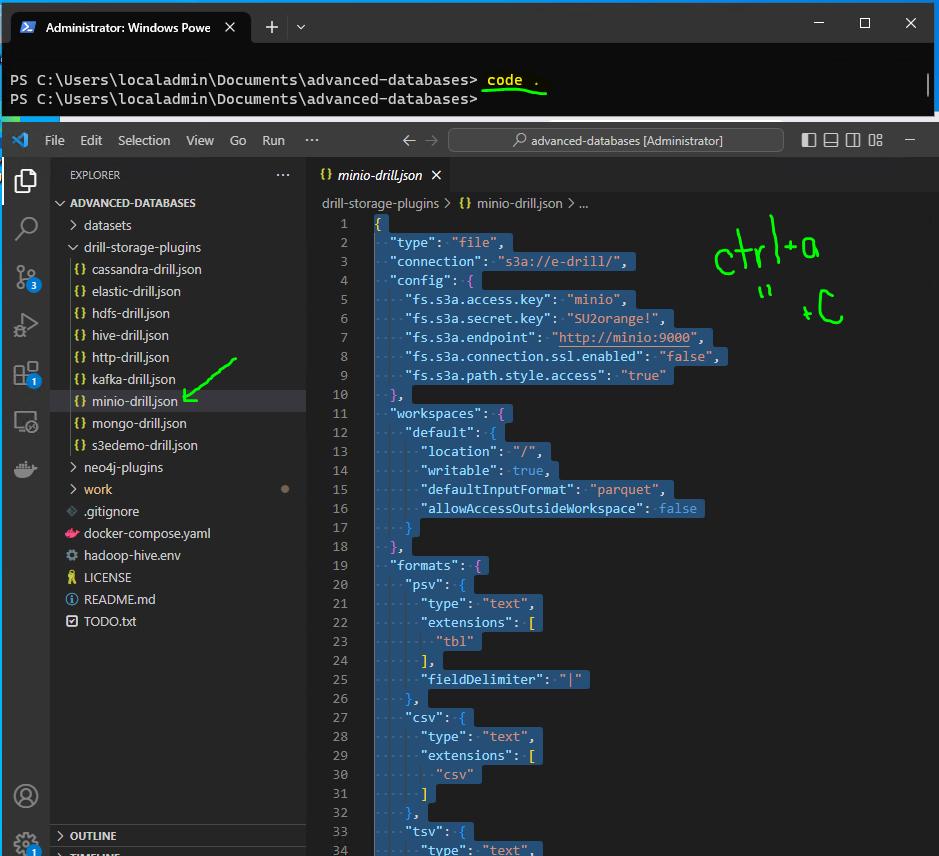
* In Command Line…



* In Jupyter…

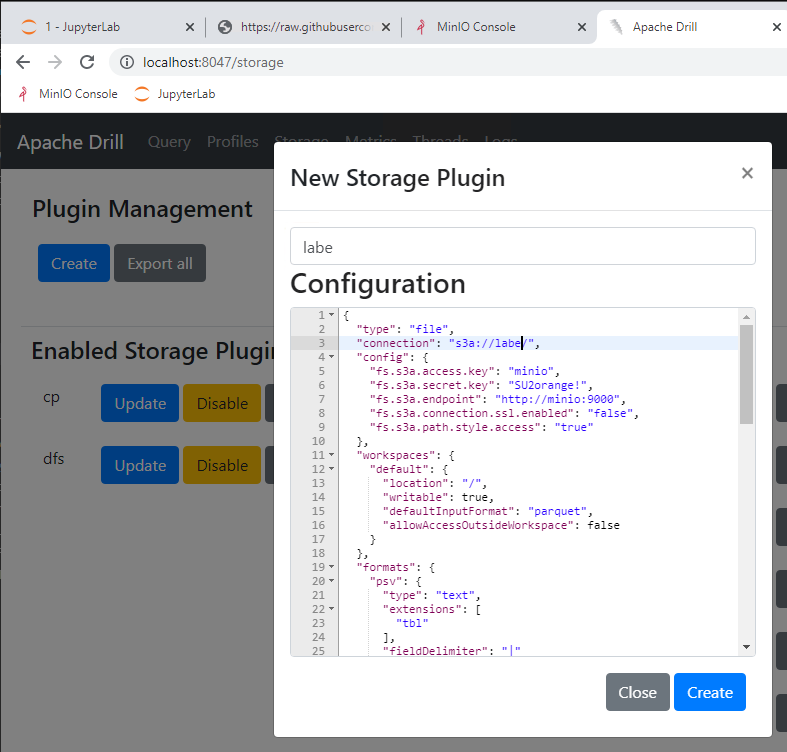


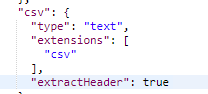
* In Minio > Lab E > Upload file

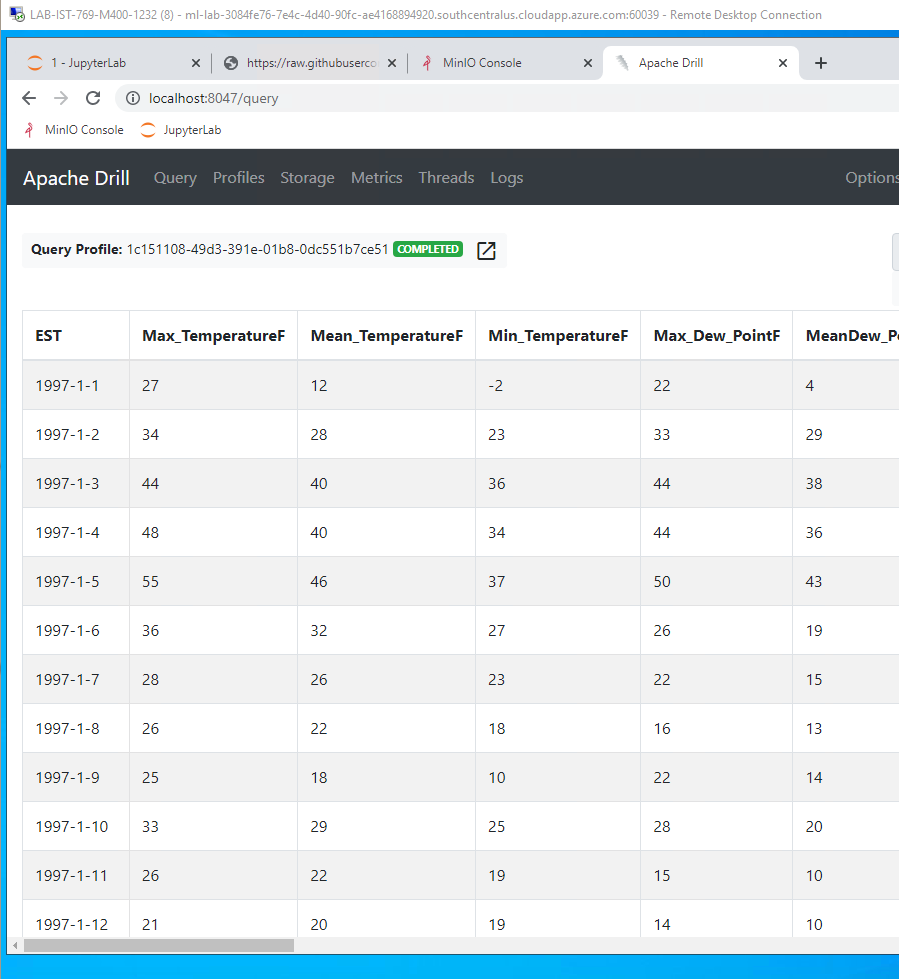


**Paste into Drill (on local host) > Storage > Create**

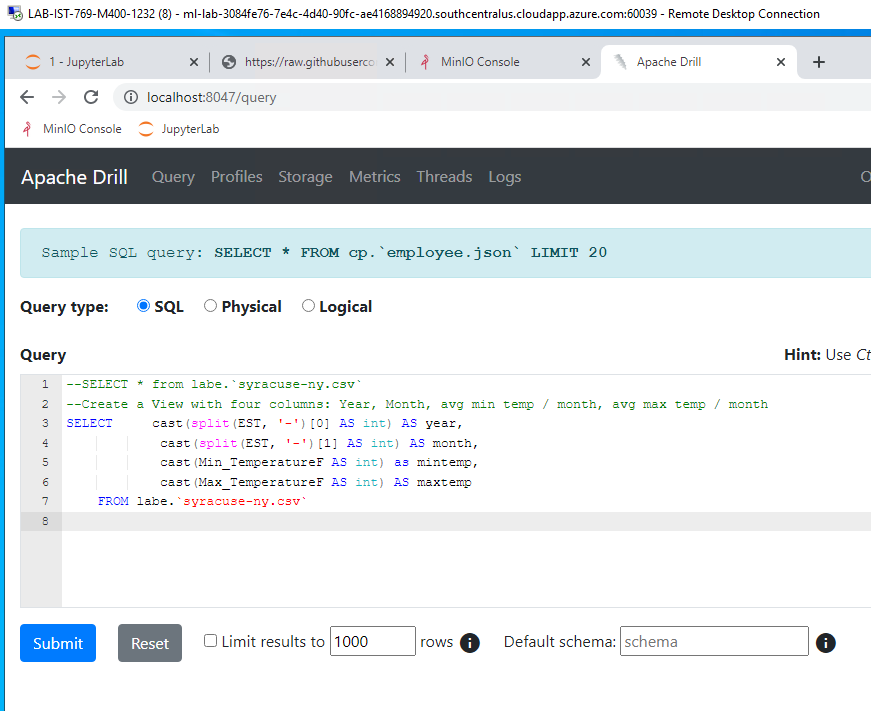
1. Configure a Drill storage plugin for the Minio **labe** bucket. Then write a Drill query for **syracuse-ny.csv** to demonstrate you can read the file with headers.

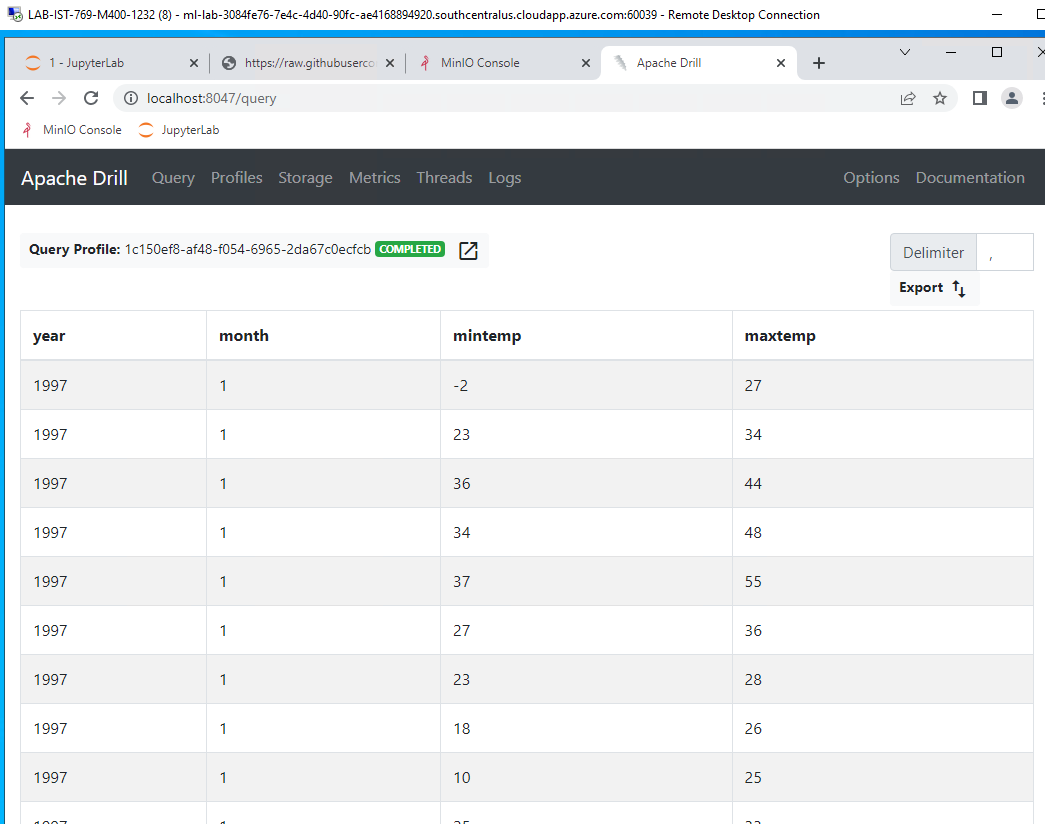




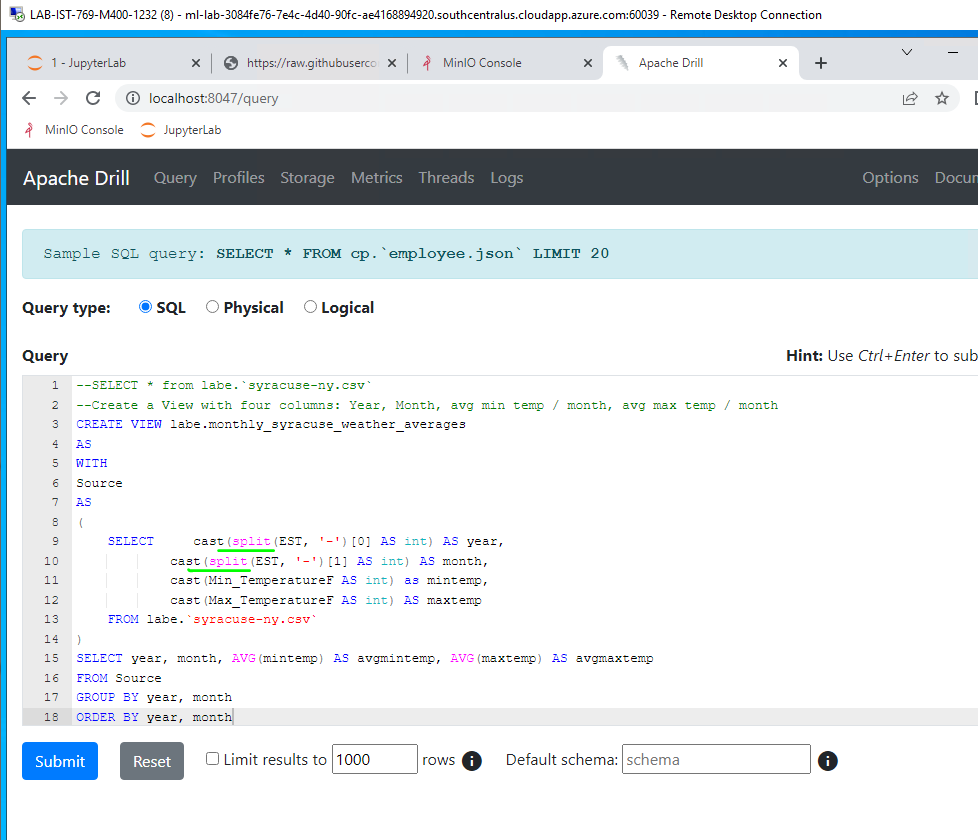


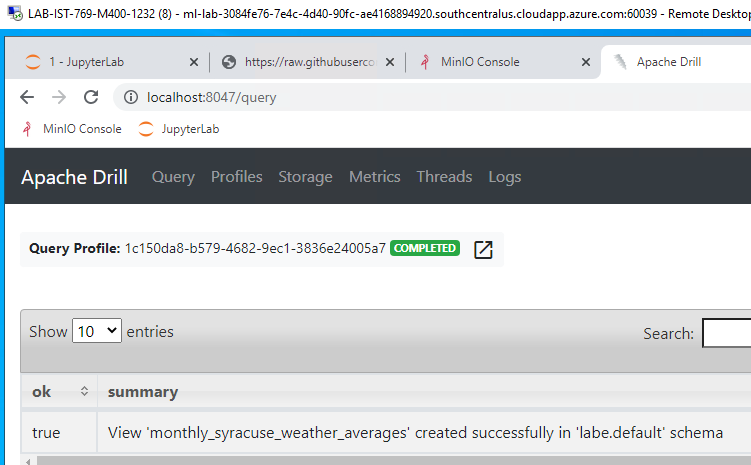
1. Write a Drill SQL query to get the overall average min and max temperatures by year and month. Use Drill’s SPLIT() function to separate Year, Month. You might need to use cast() to ensure the min and max temperatures are numeric types. You output should include four columns: Year, Month, the average minimum temperature for that month, and the average maximum temperature for that month.



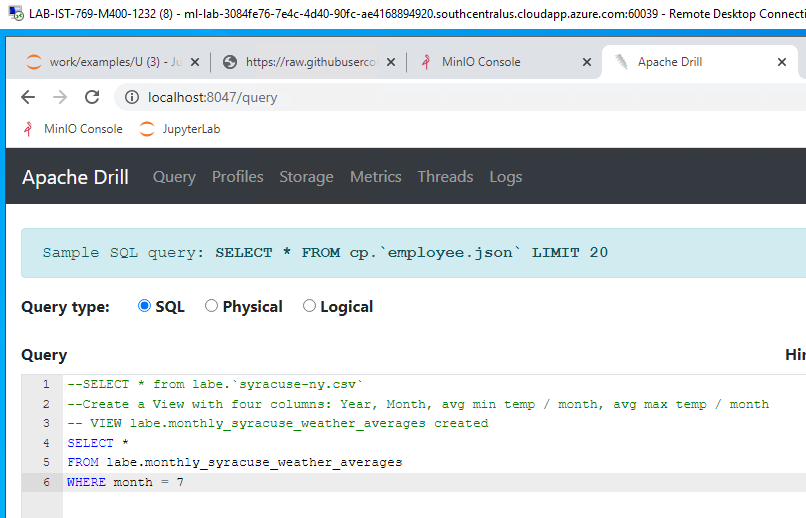


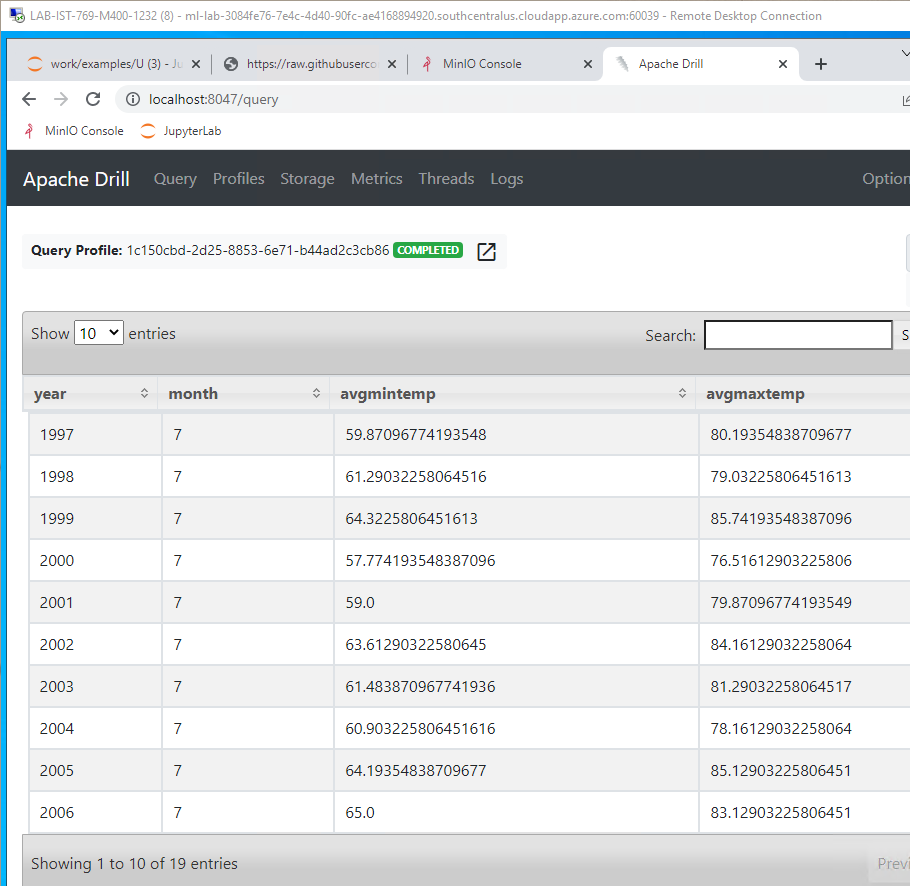
1. Create a view called **monthly\_syracuse\_weather\_averages** from the query you wrote in Question 2 and store it back on the **labe** bucket. (If you cannot get Question 2 working, use a similar query.) Provide your Drill SQL code and a screenshot showing the view file is on the Minio bucket.  
   NOTE: If you get an error about an immutable object, you need to change your storage config so you can write to the storage location.



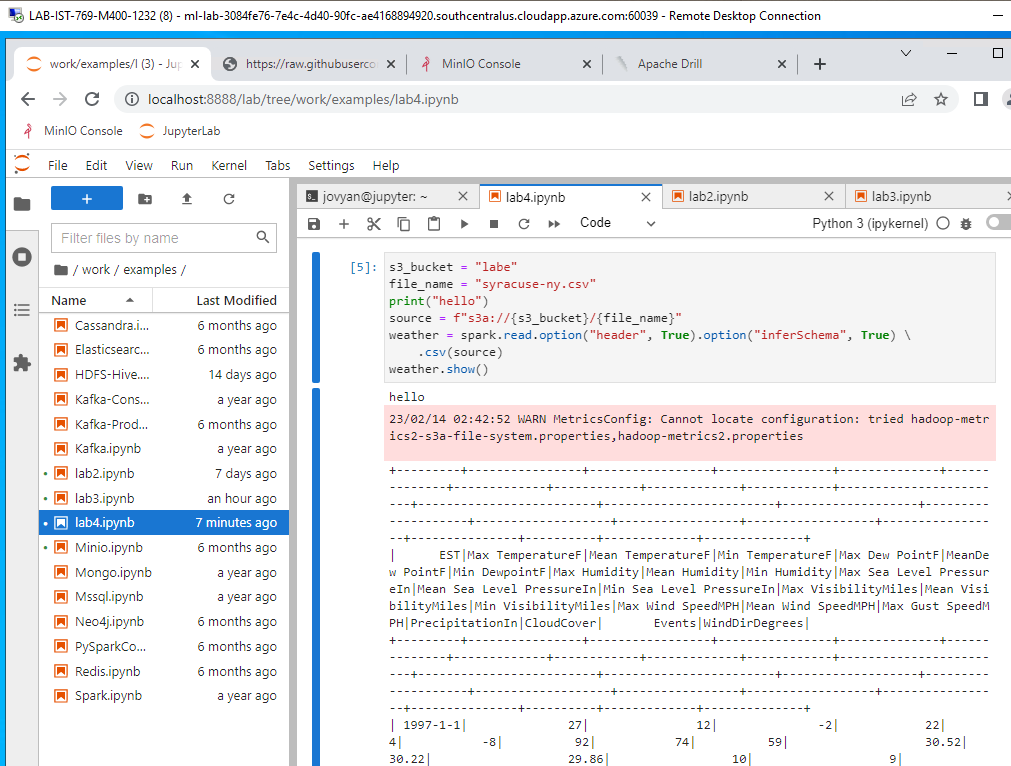


1. Use the view you created in Question 3 to show the weather data for only the month of July.

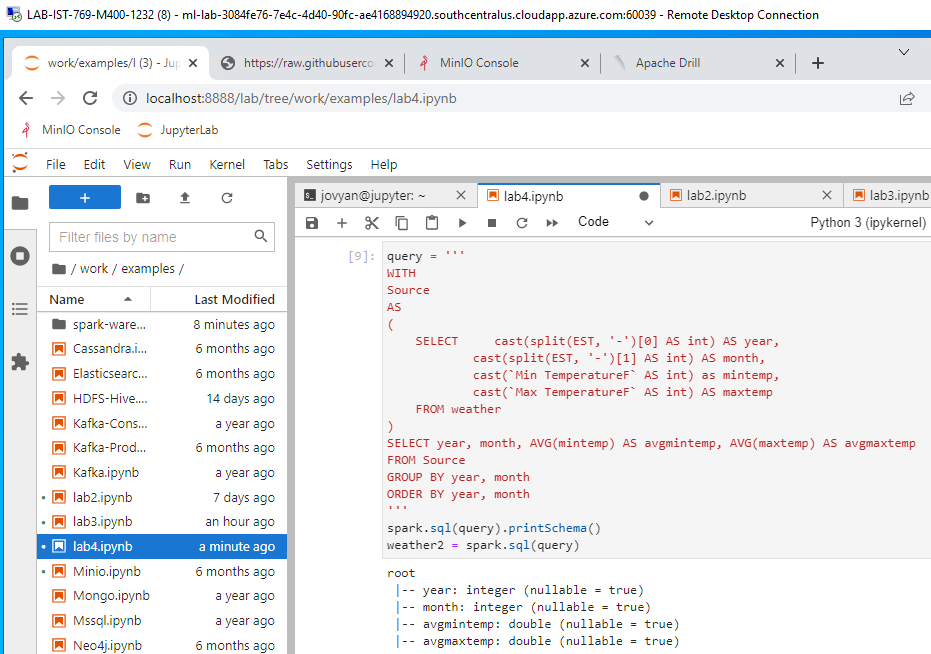




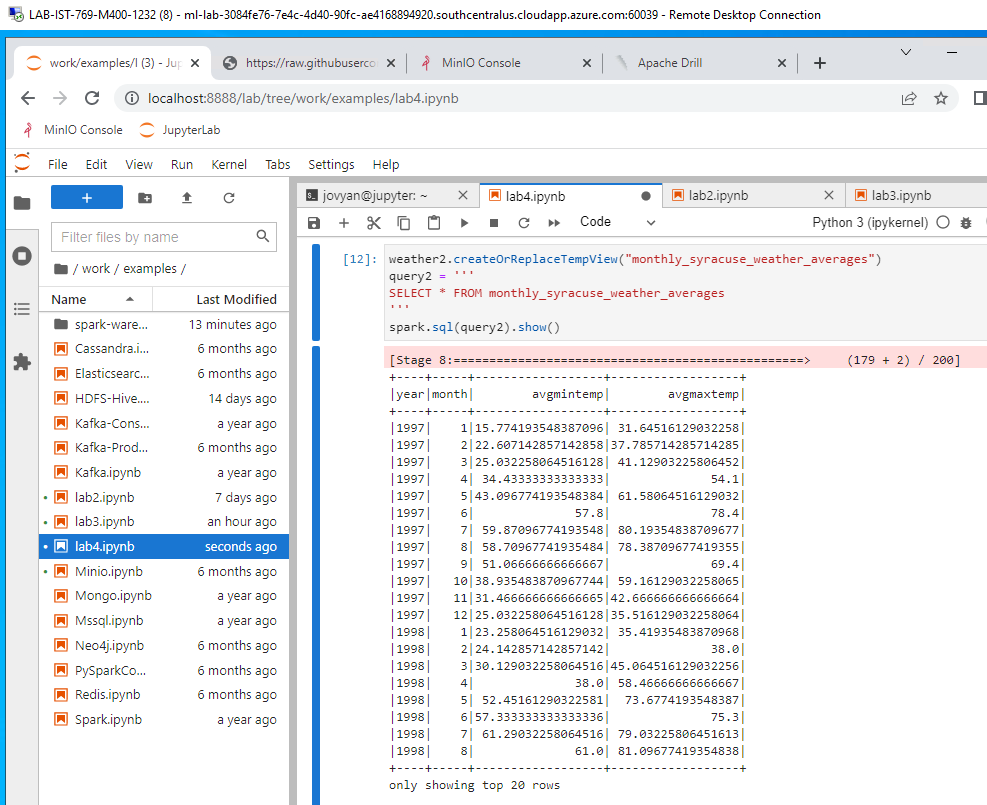
1. Configure Spark to read from Minio **labe** bucket, then load **syracuse-ny.csv** into a DataFrame and register it as the table **weather**.



1. Rewrite Question 2 using pure Spark SQL and the **weather** temp view. NOTE: There will be some subtle differences with how you must write the code, so be sure to **printSchema()** so you can see what the columns are.



1. Save the output from the DataFrame in Question 6 to the temp view **monthly\_syracuse\_weather\_averages**. Prove the view is there by querying it.



1. CHALLENGE YOURSELF! At the bottom of the **work/content/E-Drill-Spark.ipynb** file there is a section called Big Data to Small Data. Try to write a complete program that:
   1. Inputs a month 1–12 at run-time
   2. Displays a scatter plot of min/max average monthly temperatures, where year is on the X-axis

