Project 1 reference: or ADF 3

Pre-requisite:

1. ADLS container >> source(a container) >> input\_data\_file
2. ADF (all steps in week 5 daily assignment ADF file 1 and file 2)

**Workflow Recap:**

1. **Upload raw CSV data to ADLS**.
2. **Create an ADF pipeline** to ingest the data from ADLS into **Databricks**.
3. **Clean and transform** the data in **Databricks** using Spark.
4. **Save cleaned data** back to ADLS in a separate folder for easier access.
5. ~~Simulate streaming by chunking the cleaned data into time-based intervals.~~
6. ~~Create another ADF pipeline to ingest the chunks into Databricks as if they were real-time streaming data.~~
7. ~~Perform real-time data transformations and analysis in Databricks.~~
8. **Store the transformed data** or analysis results back in ADLS or a desired destination.

**If you got the error: (in second workflow)**

publishing error: table was not auto created AzureDatabricksDeltaLakeDataset1 Table is required for Copy activity.

Reason:

The error "Table was not auto created" occurs because the Databricks Delta Lake Dataset in the sink requires a pre-existing Delta table or a correctly configured file path. Here's how you can resolve it:

**1. Create a Delta Table in Databricks**

* Databricks Delta tables need to exist before writing data to them from ADF.
* Open your Databricks notebook and create the Delta table:

from pyspark.sql import SparkSession

# Create a Spark session

spark = SparkSession.builder.appName("DeltaTableSetup").getOrCreate()

# Define the path for the Delta table

delta\_table\_path = "abfss://<container>@<storage\_account>.dfs.core.windows.net/delta-tables/retail\_data/"

# Create an empty Delta table

spark.sql(f"""

CREATE TABLE IF NOT EXISTS delta.`{delta\_table\_path}` (

InvoiceNo STRING,

StockCode STRING,

Description STRING,

Quantity INT,

InvoiceDate TIMESTAMP,

UnitPrice FLOAT,

CustomerID STRING,

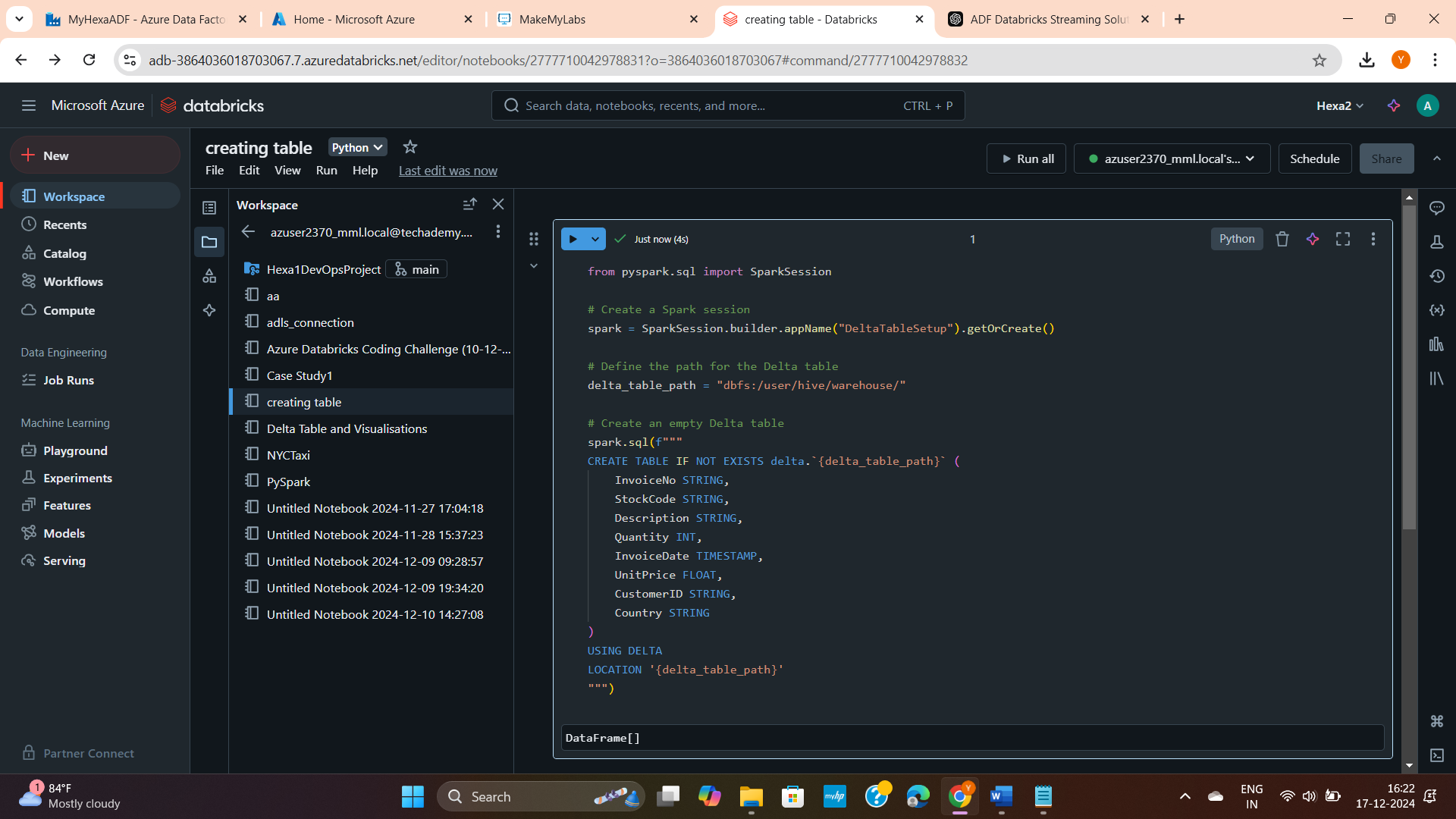
Country STRING

)

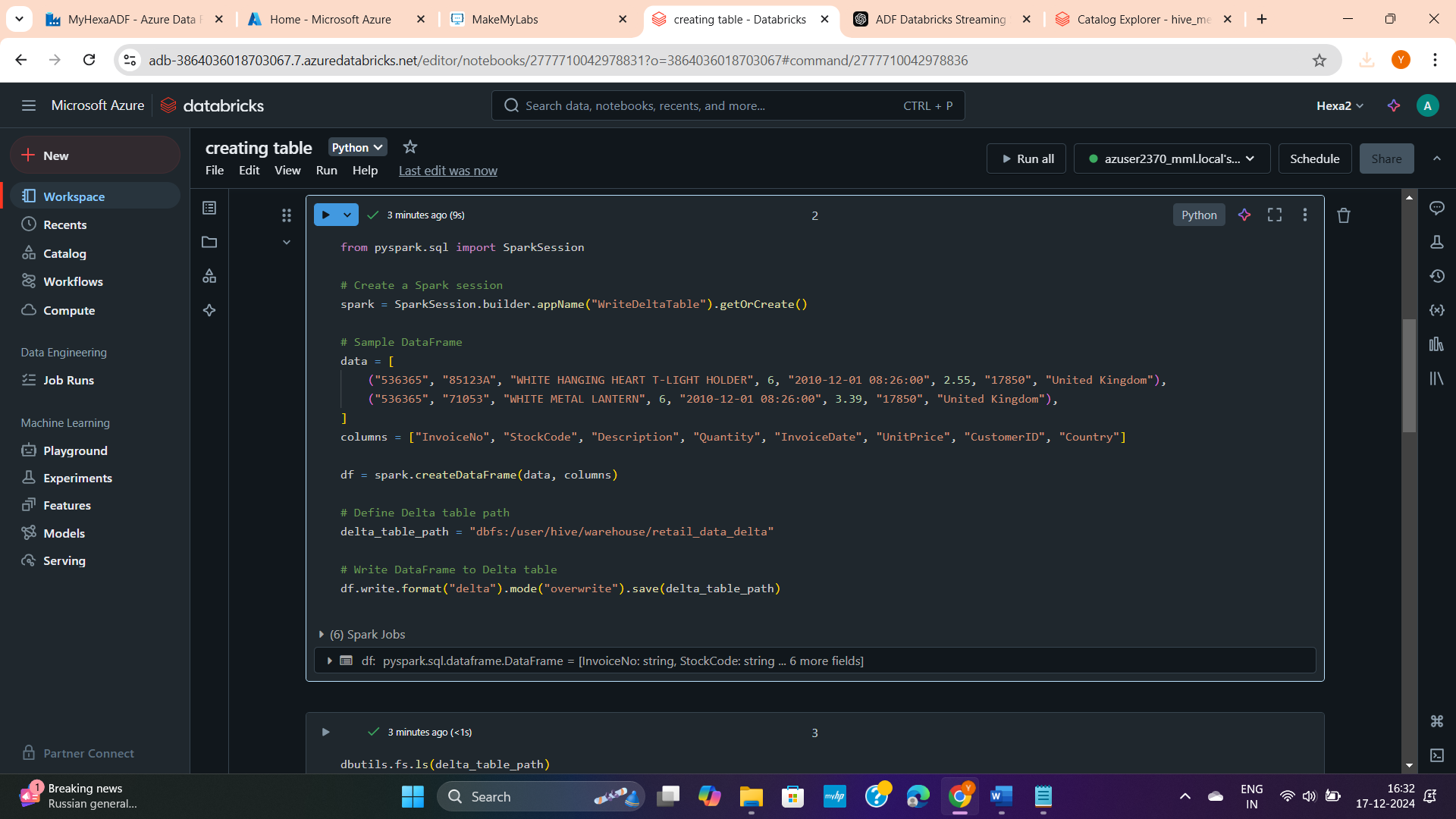
USING DELTA

LOCATION '{delta\_table\_path}'

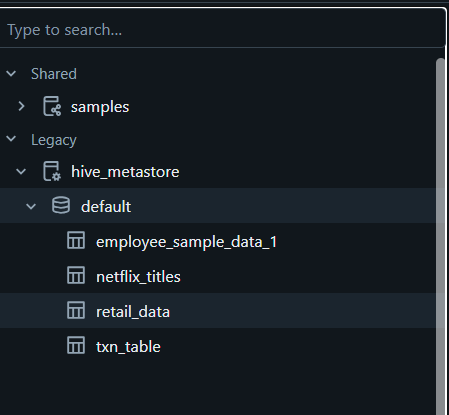
""")



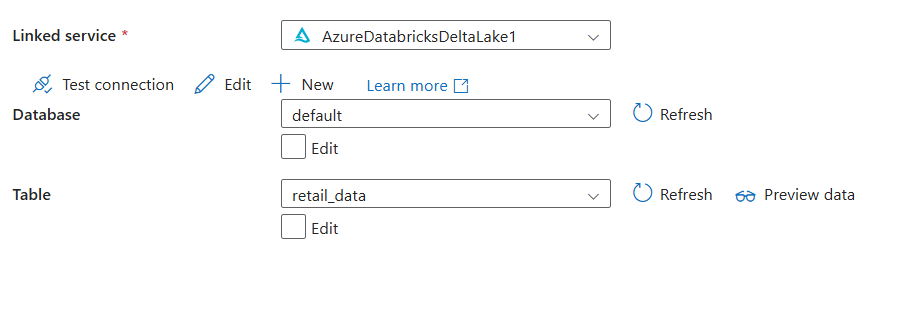
**2. You have to register your table to make it visible – for this you also want to add data as delta table cannot have empty table.**



**Refer code in:**

****

**Add sink(destination in adf):**

****

