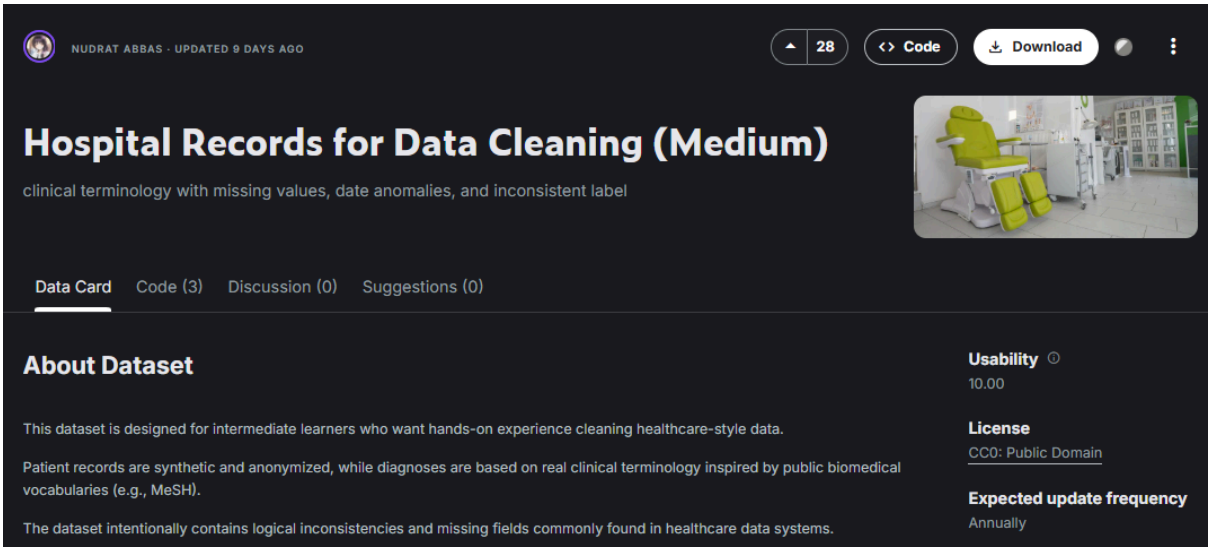
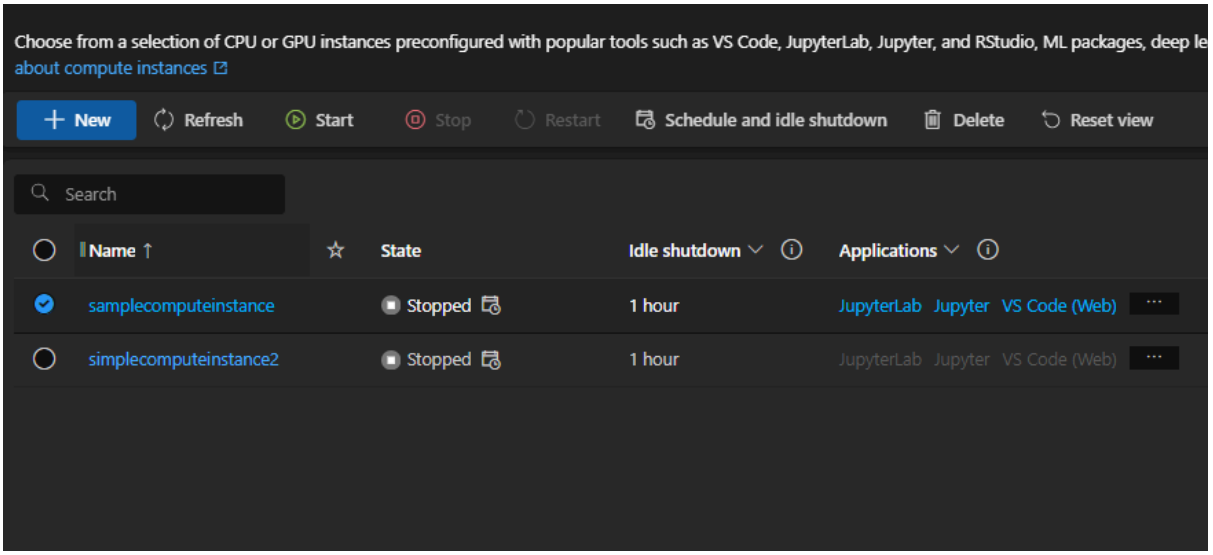
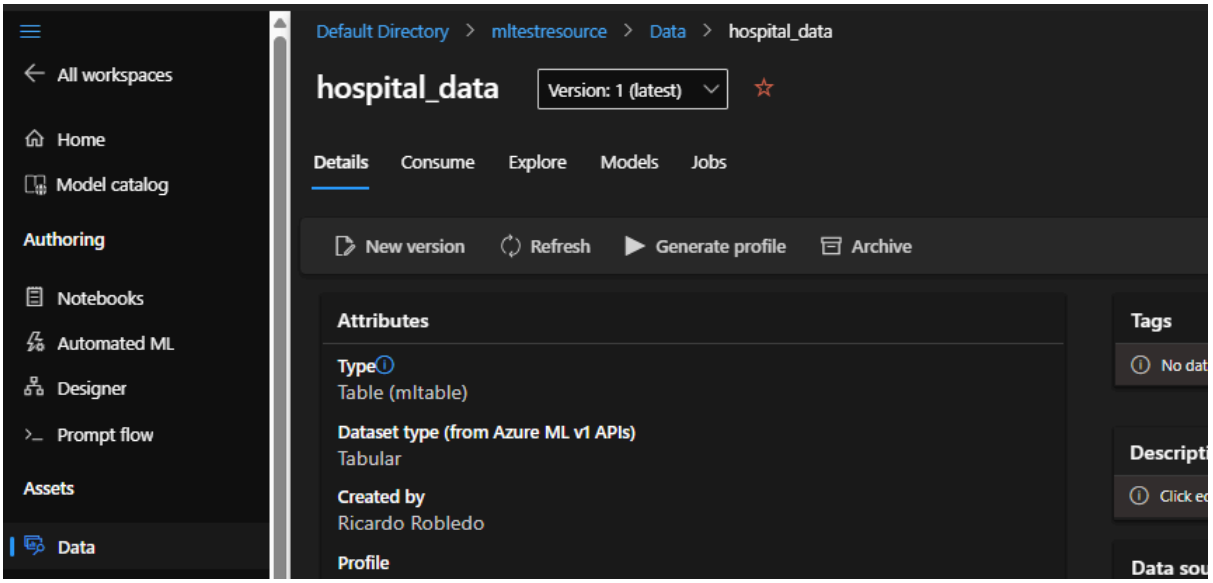


We will use Azure ML Designer to develop a model that will help us with this dataset for hospitals. In this case we will develop a model to predict the cases in a month given the diagnosis, day and month.

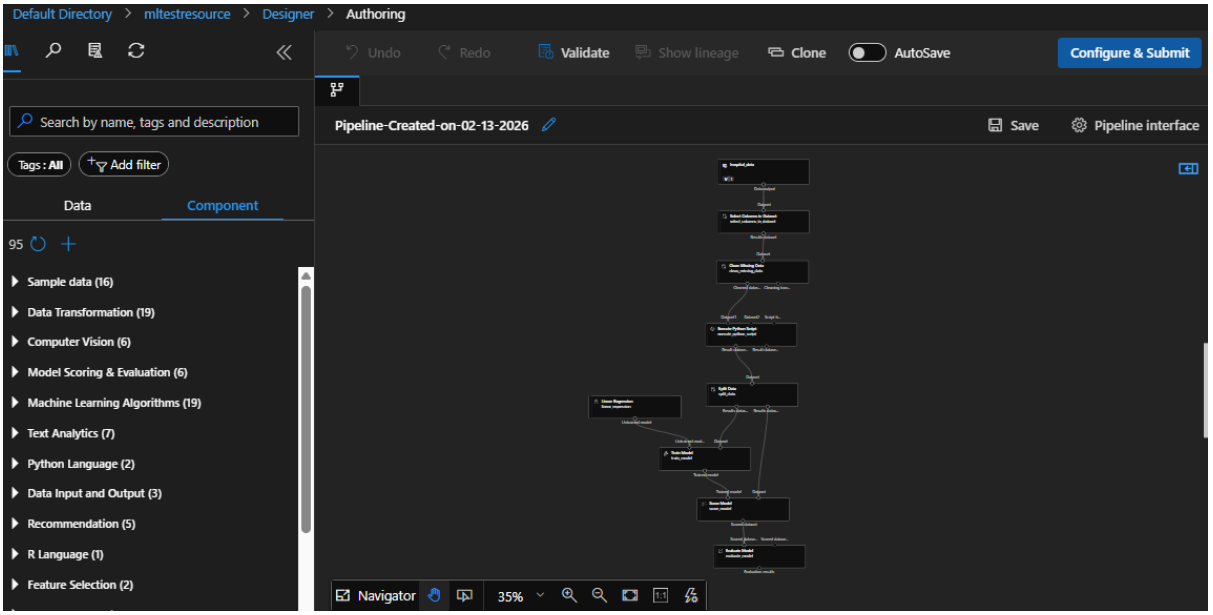
We need to start or create a Compute instance. You need to go to Azure ML Workspace and the select Compute option, then turn on or create your instance.



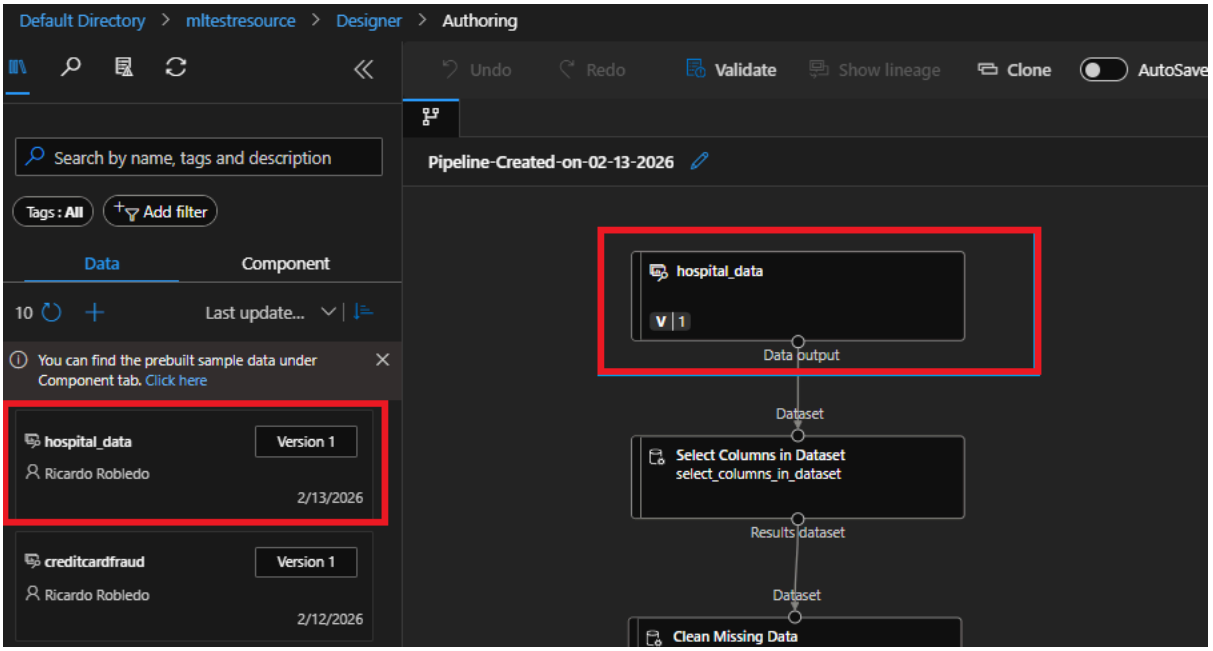
Then you need to have a dataset, for that you need to have a storage account and upload the dataset in a container instance, after that create a data asset with that file.



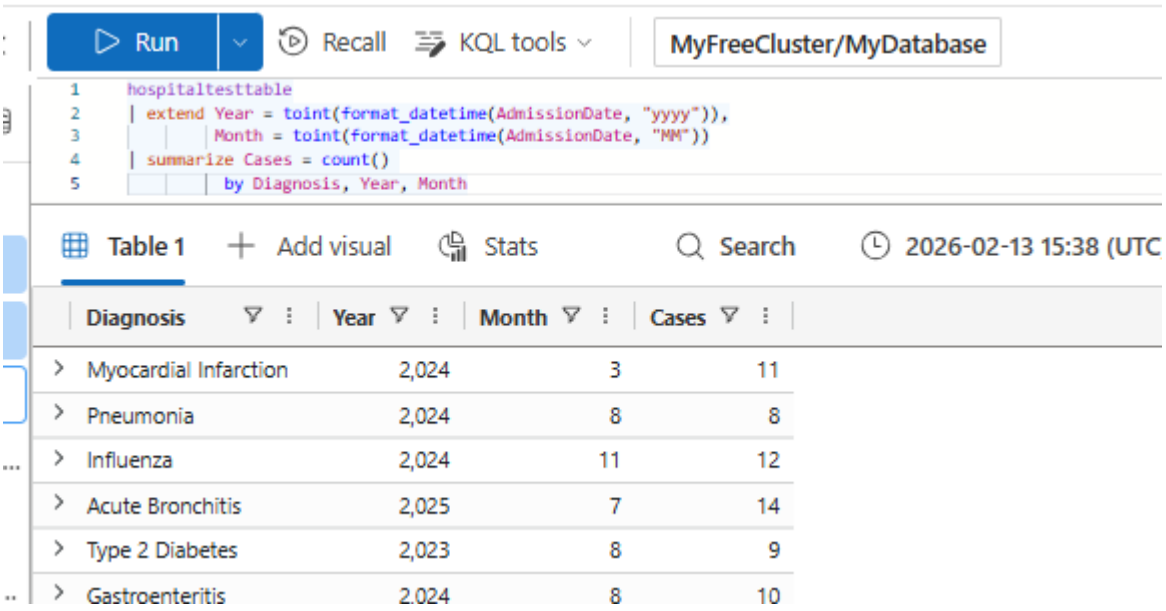
Then we go to Designer and create a template to create our model. In this case I have a pipeline that I will explain.



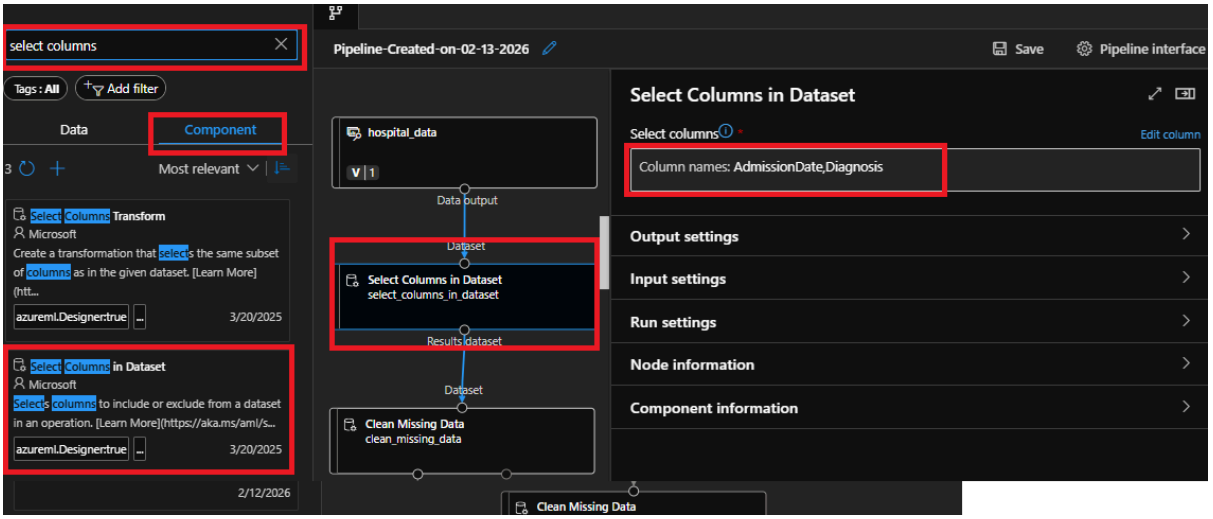
Go to data and select the dataset.



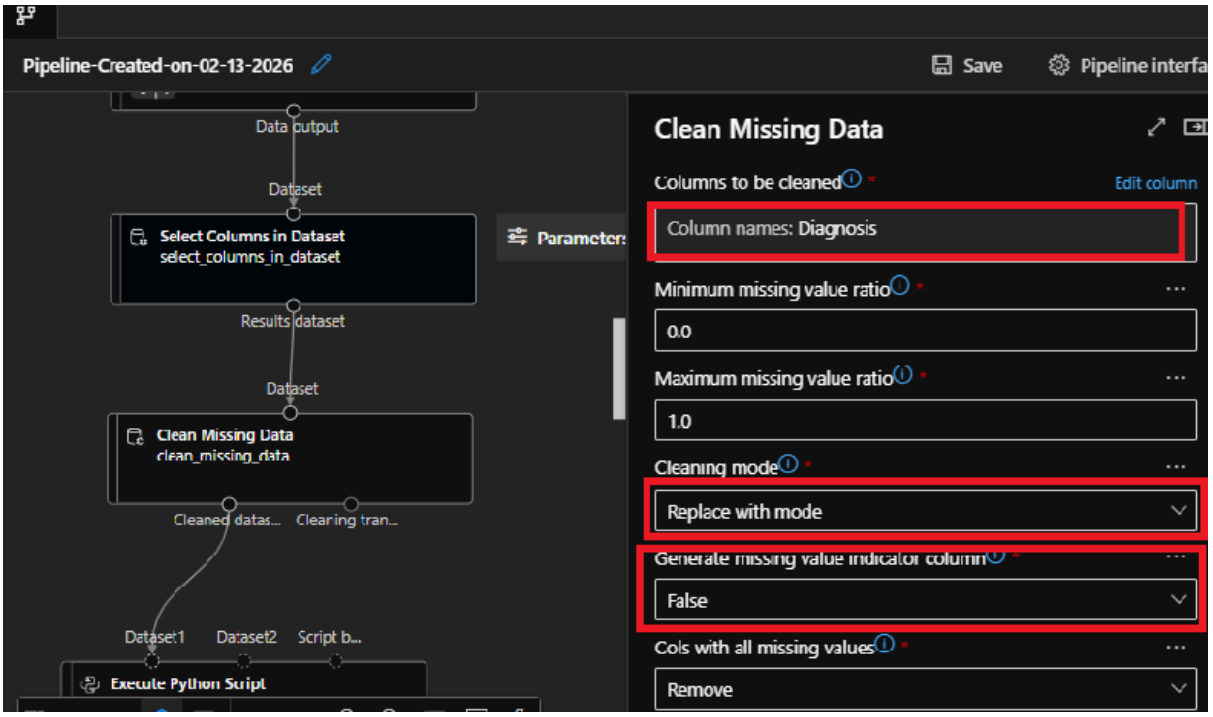
Ok, it is important to know which columns we will use, in my case I used Azure Data Explorer and Kusko to explore data, we will make a model that predicts the total count of cases depending on the diagnosis.



Now we must go to select components and drag Select Columns in Dataset, after that select the next columns.



Now we need to use the component Clean Missing Data. Select column Diagnosis, replace with mode and not generate another column.



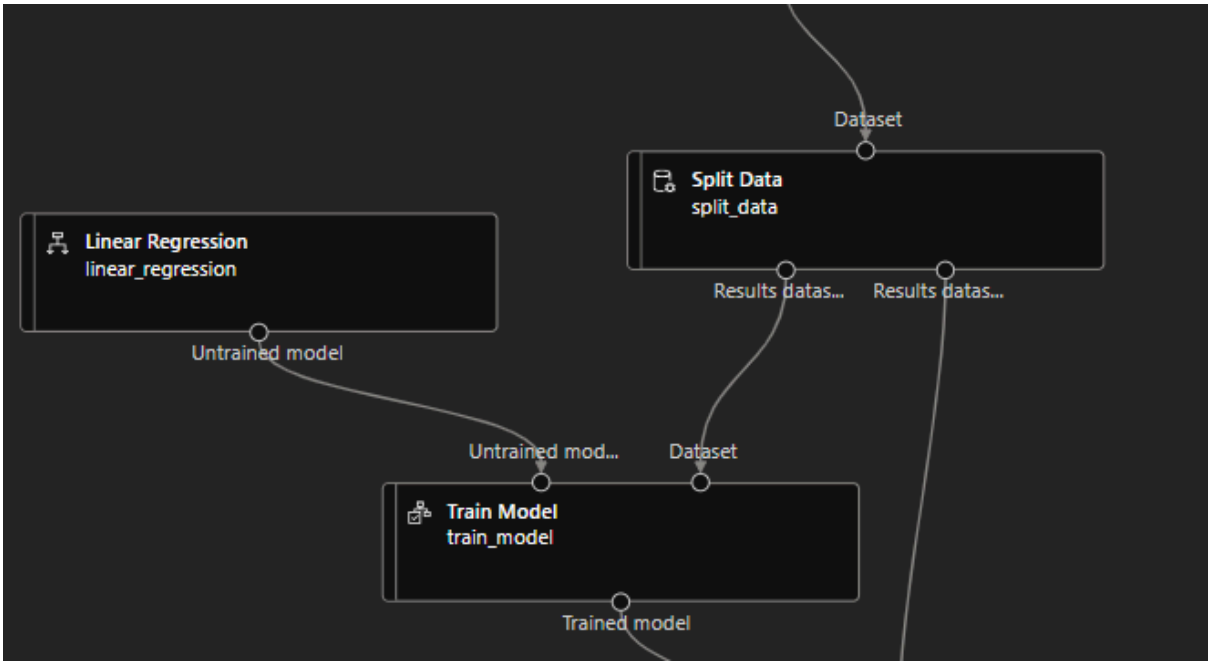
The next is to put a Execute Python Script module, the code is the next:

```
import pandas as pd

def azureml_main(dataframe1=None, dataframe2=None):

    # Copiar solo columnas necesarias
    df = dataframe1[["Diagnosis", "AdmissionDate"]].copy()
```

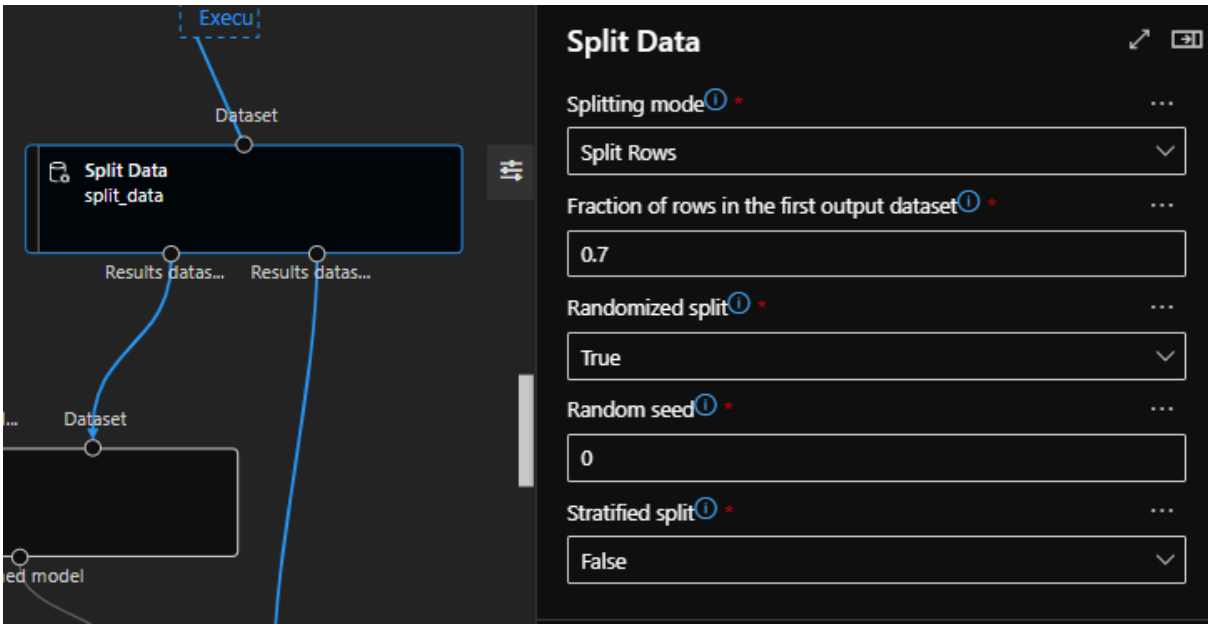
```python
    # Convertir a datetime
    df["AdmissionDate"] = pd.to_datetime(df["AdmissionDate"])

    # Crear Year y Month
    df["Year"] = df["AdmissionDate"].dt.year
    df["Month"] = df["AdmissionDate"].dt.month

    # Agrupar y contar
    result = (
        df.groupby(["Diagnosis", "Year", "Month"])
          .size()
          .reset_index(name="Cases")
          .sort_values(["Year", "Month", "Cases"],
                       ascending=[True, True, False])
    )

    # Devolver SOLO las columnas deseadas explícitamente
    result = result[["Diagnosis", "Year", "Month", "Cases"]]

    return result,
```

In a brief summary it only group the diagnosis by month and year, then we only return the columns that we will use to predict the Cases.

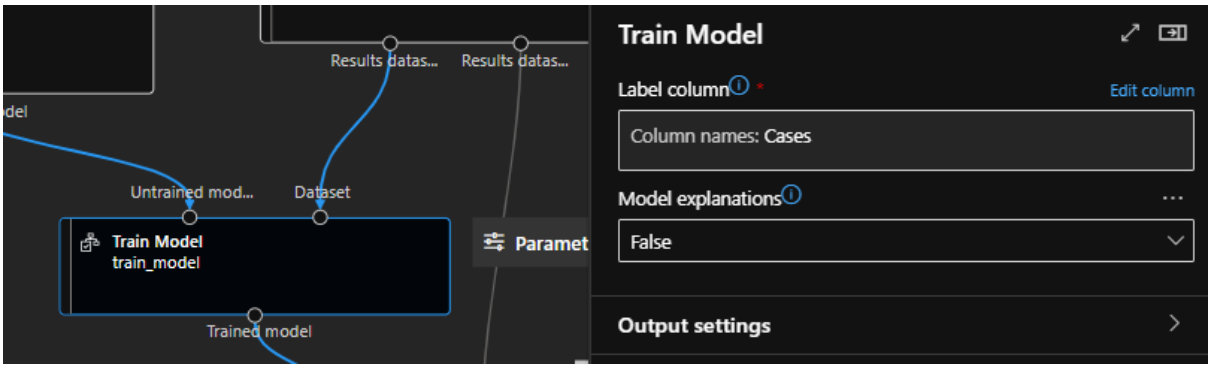
Ok, we will use 3 modules, they are: Split Data, Liner Regression and Train Model.



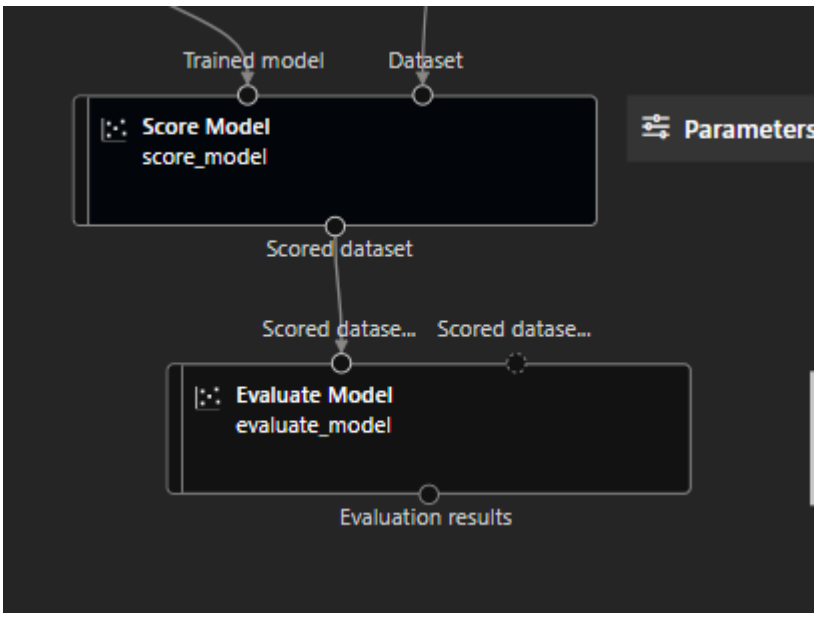
These are the settings for Split Data. Like we have temporal data, we don´t need randomized split neither stratified split.



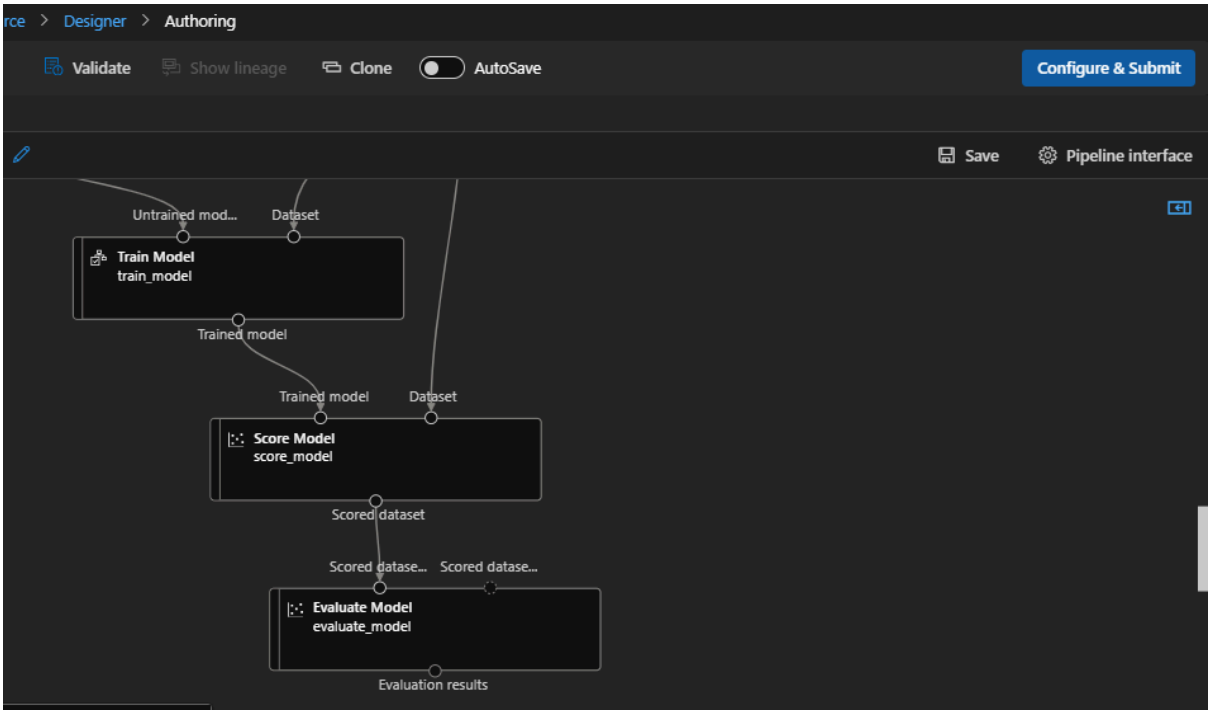
Train Model is very simple, only select the target that is Cases.



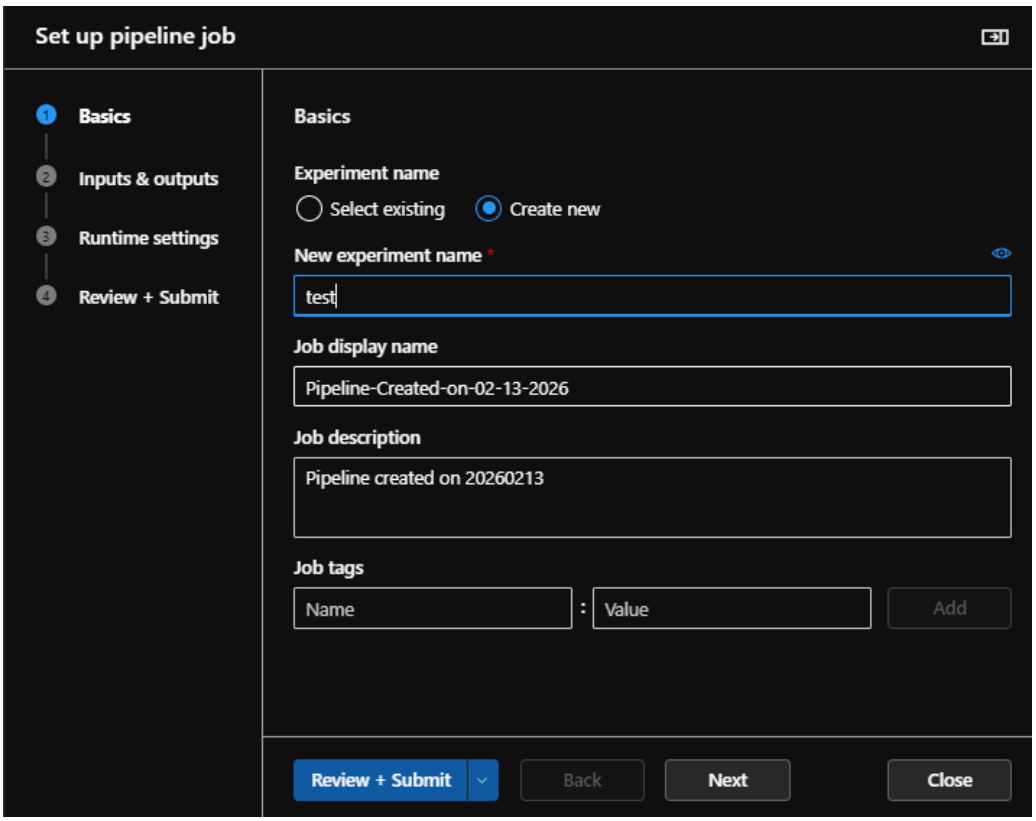
After that we need only 2 modules that are Score Model and Evaluate Model.



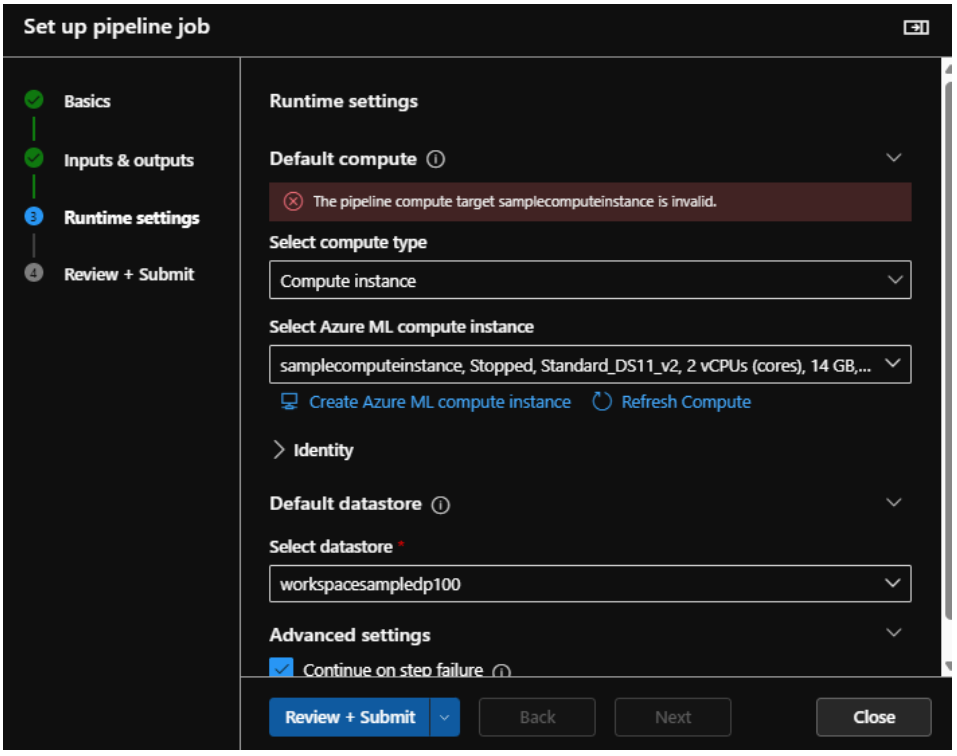
Very well, we only need to execute our pipeline, to that push Configure & Submit.



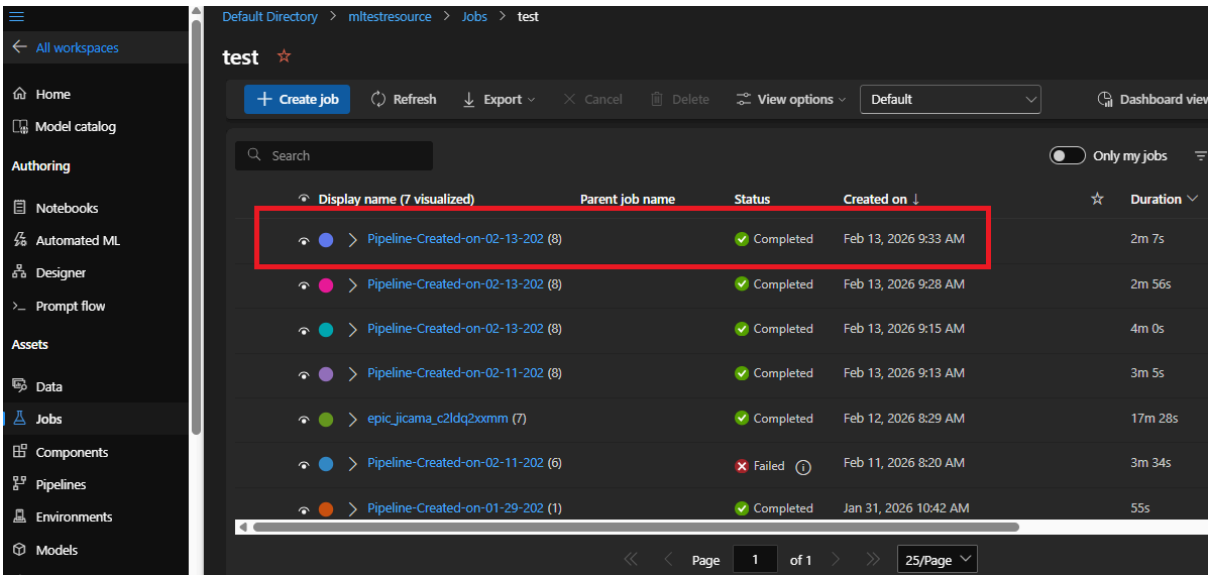
Give it a name to the experiment.



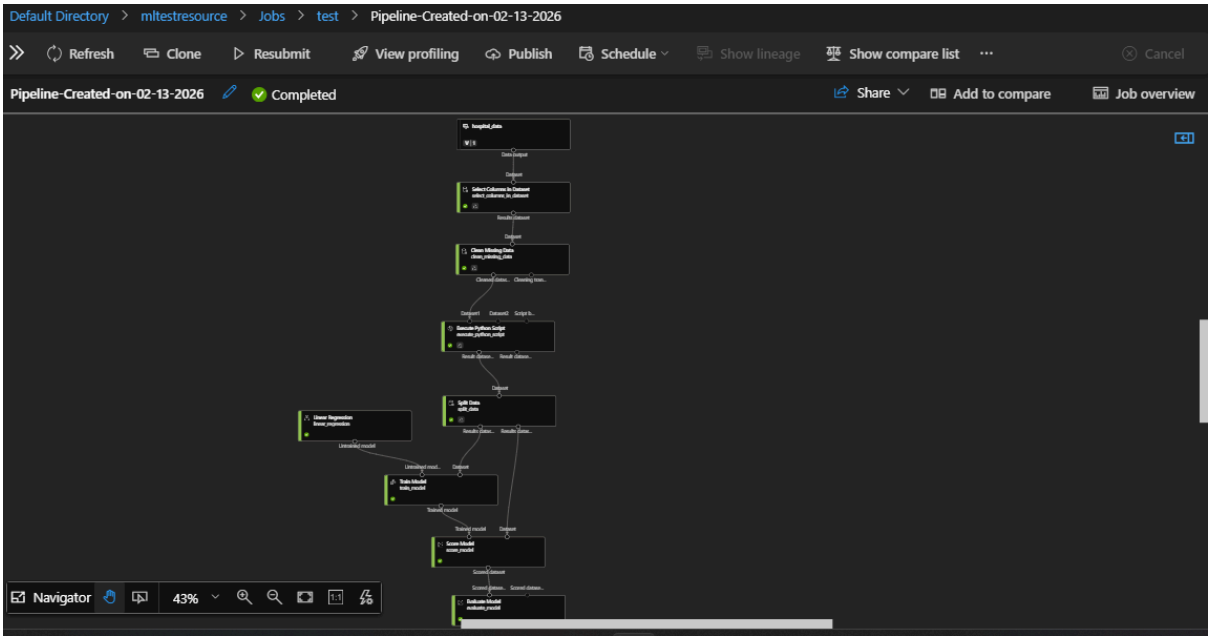
Select Compute instance, and Review + Submit.



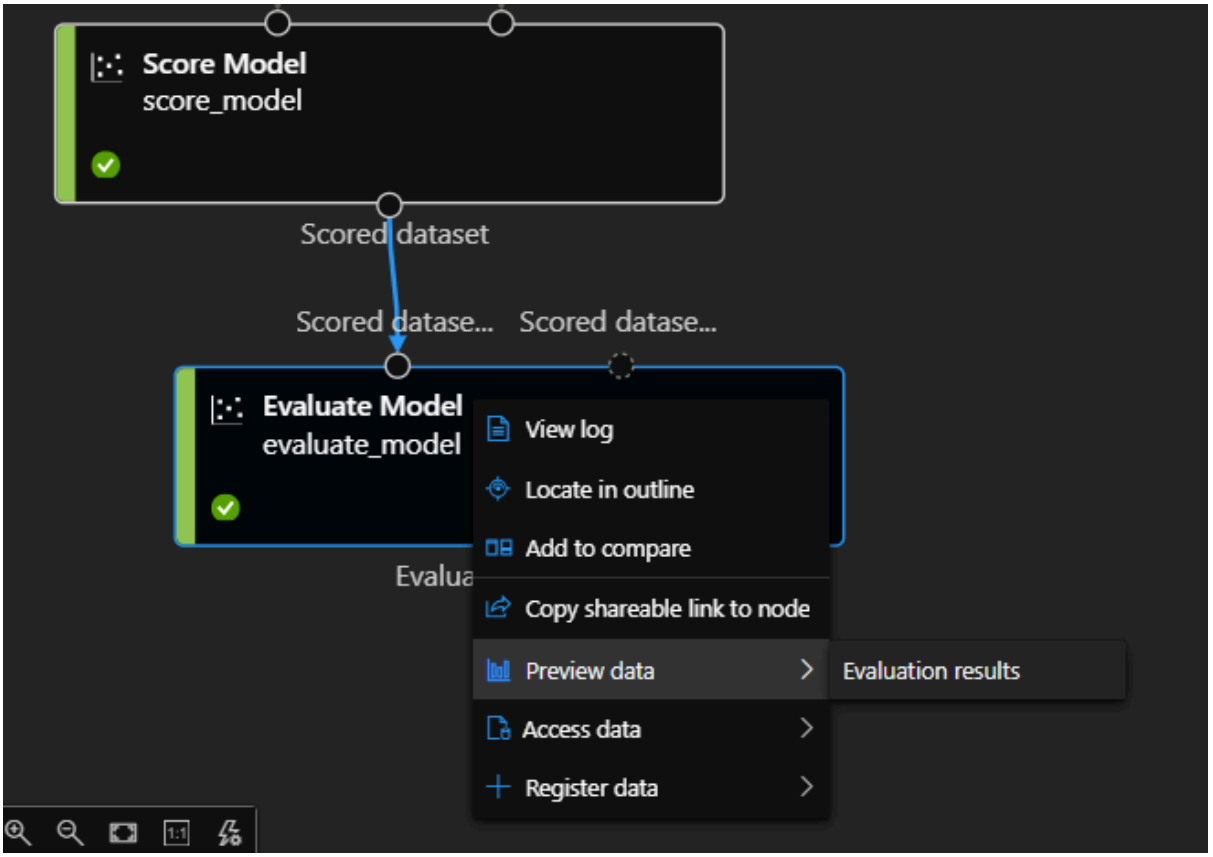
after that the pipeline will be created, so after some minutes we can go to Jobs in the menu and Select pipeline.



OK!, this is good, all the modules have executed successfully.



Find the Evaluate Model module and do right click and select the next option..



These are our results.