**Project Report**

**Project Title**

**Diagnostics for Detecting Cardiovascular**

**Team Leader:**

Shivanoor Vignesh 21R11A0597

**Team Members:**

G.Sandhya Rani 21R11A0569

Dhathri 21R11A0545

Hasini 21R11A0518

Pooja 21R11A0501

PSBN Sriya 21R11A0595

N Sai Charan 21R11A0588

MD.Akhbar Ali 21R11A0586

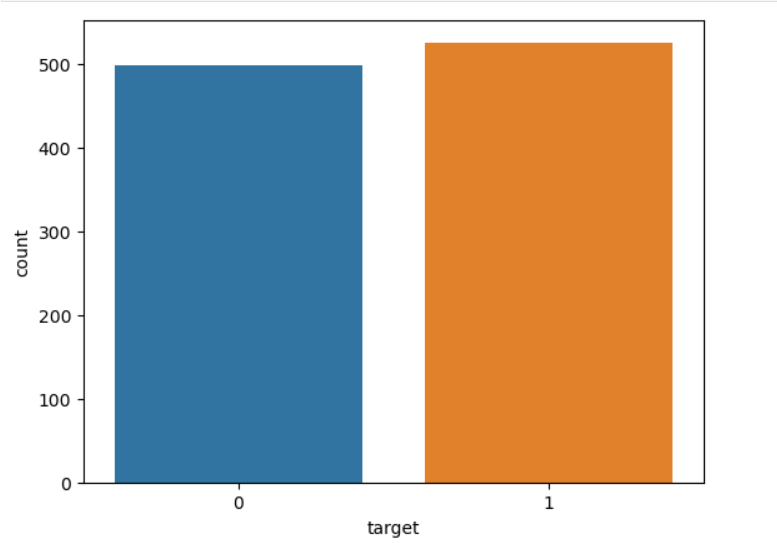
**Exploratory Data Analysis**

**Import data set**

For importing the dataset and to perform Exploratory Data Analysis we have to import some packages or library which are essential.

* import pandas
* import NumPy
* import seaborn
* import matplotlib

**COUNT OF TARGET(total no of heart disease patients):**



**Observation:**

The above bar graph illustrates the count of total number of heart disease patients.That is,’0’ indicating possibility of not having any .heart disease ,where as ‘1’ indicating the possibility of having heart disease.

**Category “0”:**

The count for this category is almost reaching 500 units.

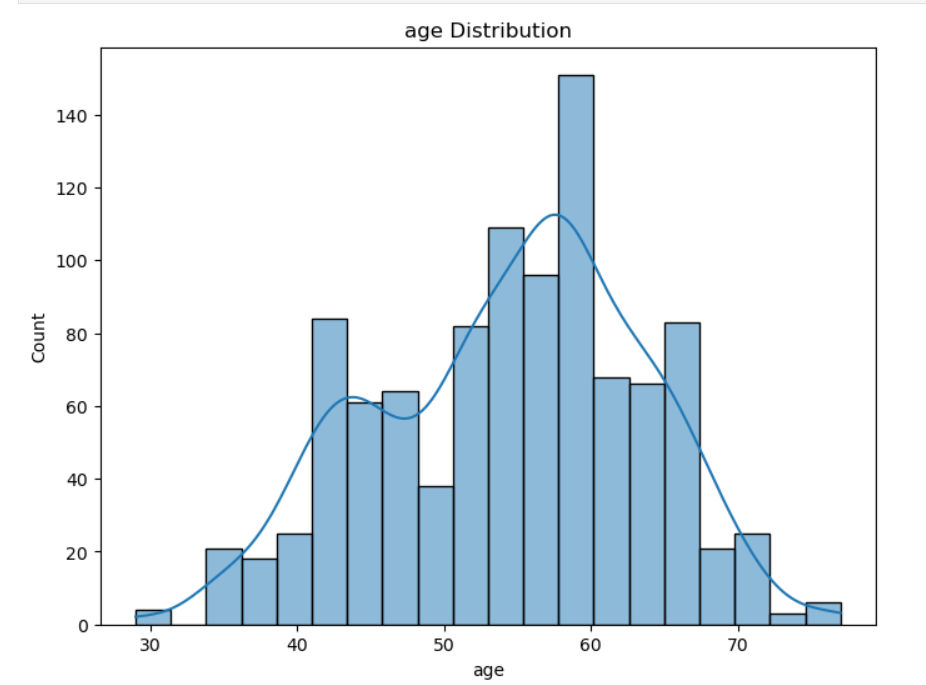
This suggests that most of the people are not having any heart disease.

**Category “1”:**

The count for this category is slightly higher than category “0”.

This suggests that even though there are more people falling under category “0”,there are even more people who are suffering from heart diseases.

**Age Distribution:**

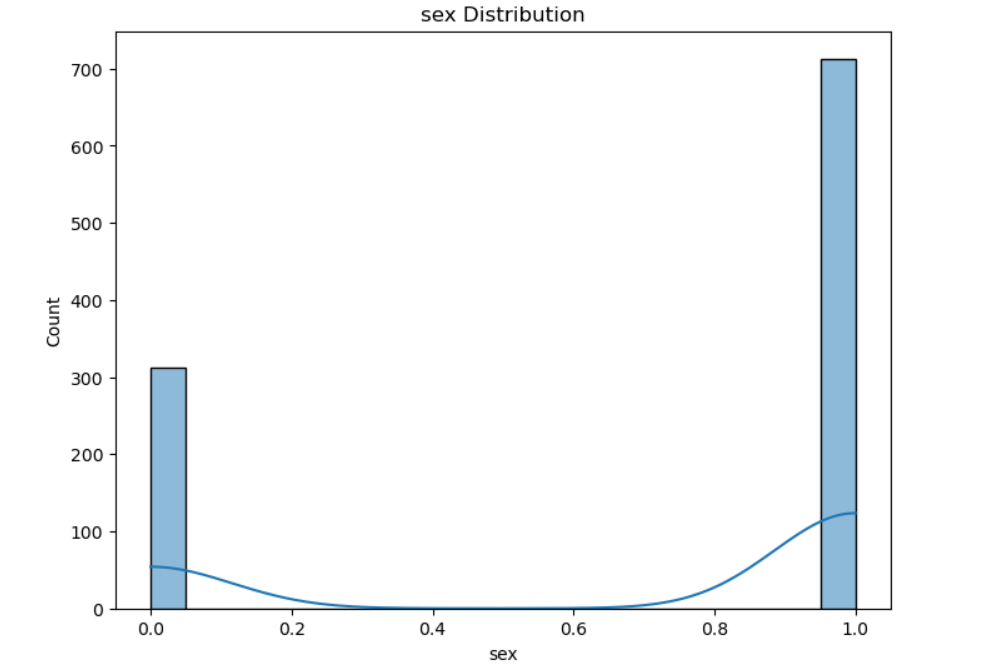


**Observation:**

The histogram visually represents age distribution and prevalence among individuals. Notably, it reveals that the age group between 55 and 60 has a higher number of people compared to other age groups.

This correlation strongly suggests an increased risk or susceptibility to heart-related issues for individuals within the 55-60 age range..

**Gender Distribution:**



**Observation:**

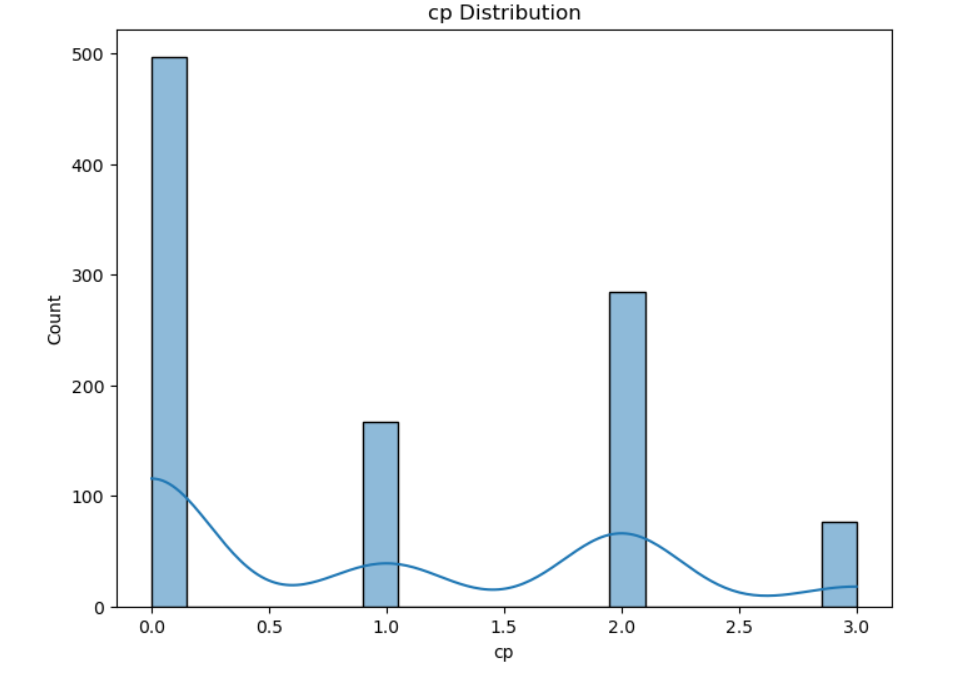
The histogram visually represents the distribution of sex (0 for females, 1 for males) concerning heart diseases.

The large difference in counts between males and females suggests a significant gender-based gap in the occurrence of heart diseases.

An examination of the data reveals a substantial count, with males registering a count exceeding 700 units, while females account for around 330 units

The discernible difference in counts leads to the conclusion that males exhibit a higher admission rate for heart diseases compared to females

**Chest Pain Distribution graph:**



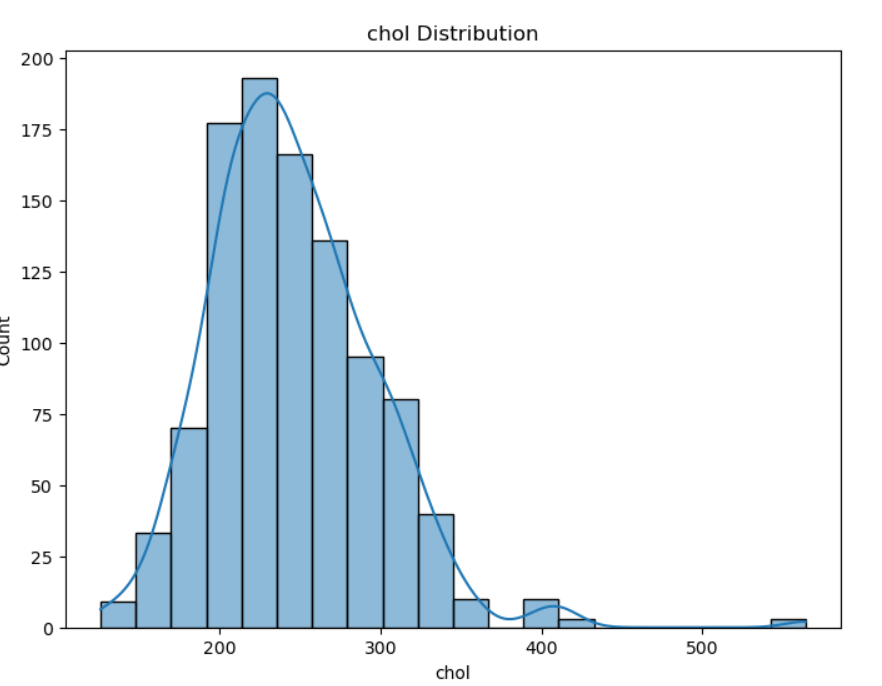
**Observation:**

The above histogram visually represents the distribution of chest pain experienced (Value 0: typical angina, Value 1: atypical angina, Value 2: non-anginal pain, Value 3: asymptomatic)

As we can observe that typical angina has the highest count, almost 500 units, than other values.And followed by non-anginal pain with around 300 units than other 2 values.

Atypical angina has approximately 270 units, while asymptomatic is notably lower, ranging from 80 to 90 units.

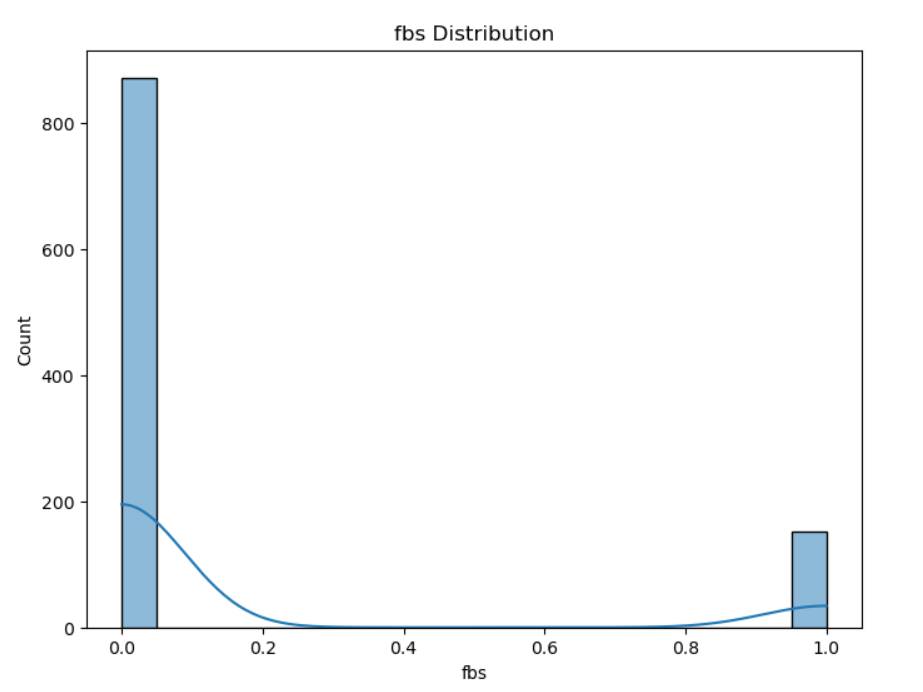
**Cholesterol Distribution graph:**



**Observation:**

The histogram visually depicts serum cholesterol levels, emphasizing that a notable concentration of individuals falls within the 240-260 range, indicating higher levels compared to other ranges.

**Fbs Distribution graph:**



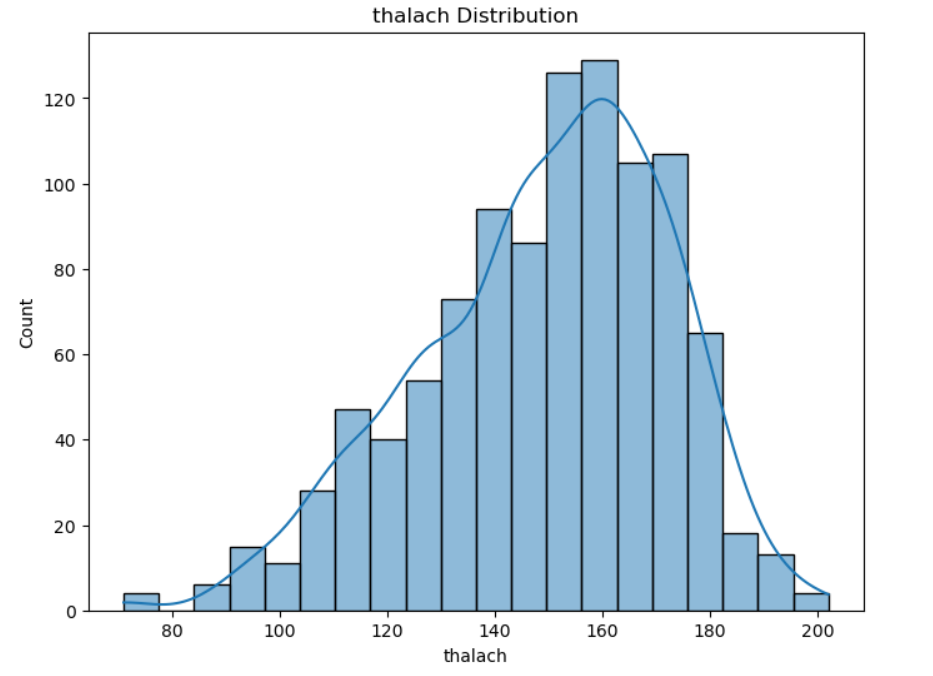
**Observation:**

The above histogram provides the visual representation of fbs distribution (fasting blood sugar(> 120 mg/dl ,1 = true; 0 = false).

"1" (true) indicates that an individual's fasting blood sugar level is greater than 120 mg/dL, suggesting an elevated blood sugar level that might pose a risk, as per the mentioned health context.

"0" (false) indicates that the individual's fasting blood sugar level is not above 120 mg/dL, suggesting a normal or lower blood sugar level based on the specified threshold.

**Thalach Distribution:**



**Observation:**

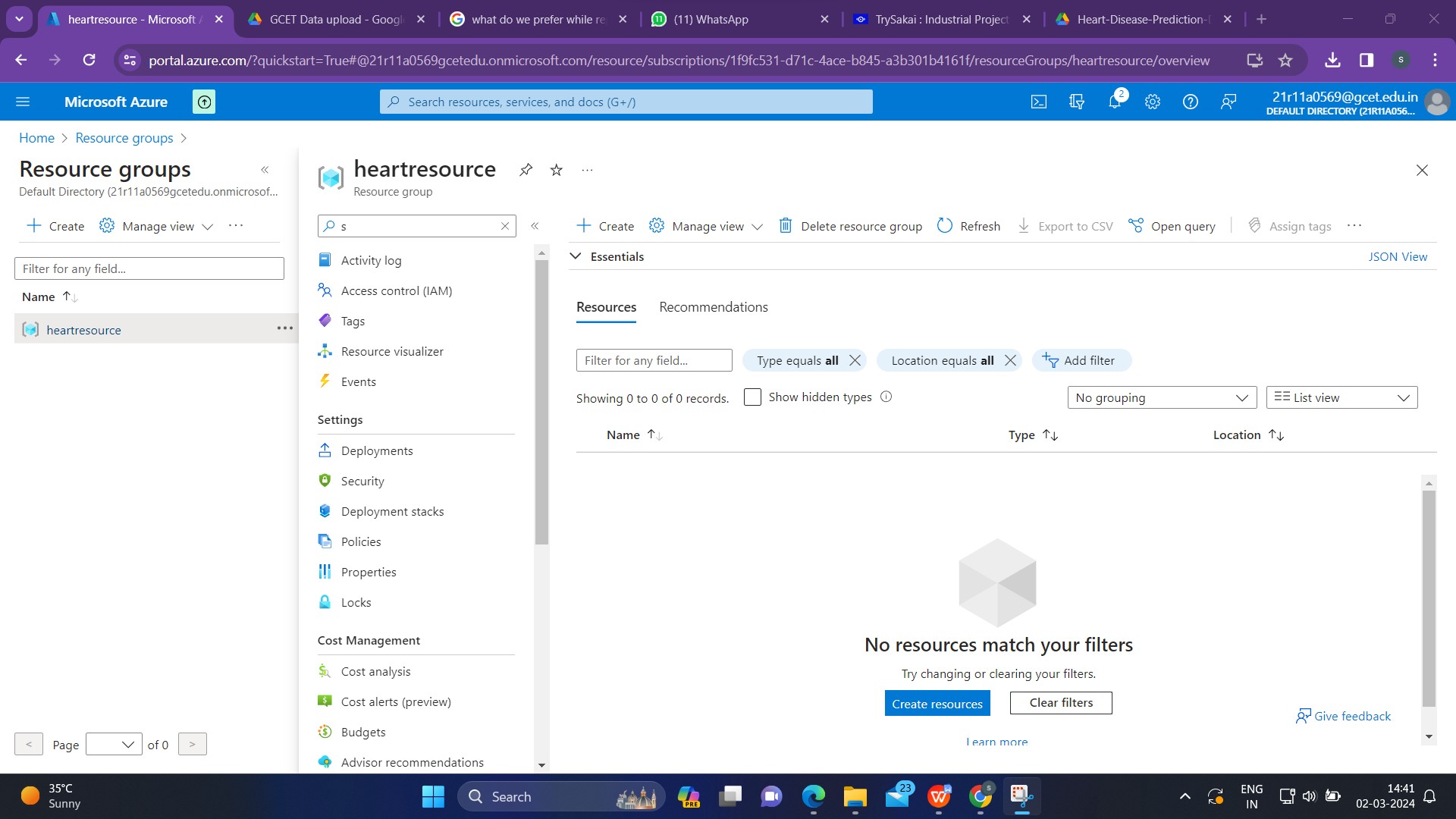
The histogram visually illustrates the distribution of thalach (maximum heart rate)

Notably, the data shows a peak around 160 units, indicating that the maximum heart rate observed in the dataset is most commonly recorded at this value.

This suggests that a significant proportion of individuals in the dataset have a maximum heart rate of approximately 160 units.

**Modelling**

**Resource group creation:**



To create a resource group in a cloud platform like Microsoft Azure, one has to generally follow these steps:

1.Sign in to the Cloud Platform: Go to the Azure portal or the relevant cloud provider's portal.

2.Search for Resource Group:In the Azure portal, we'll typically find a search bar at the top. Enter "Resource Group" to locate the service.

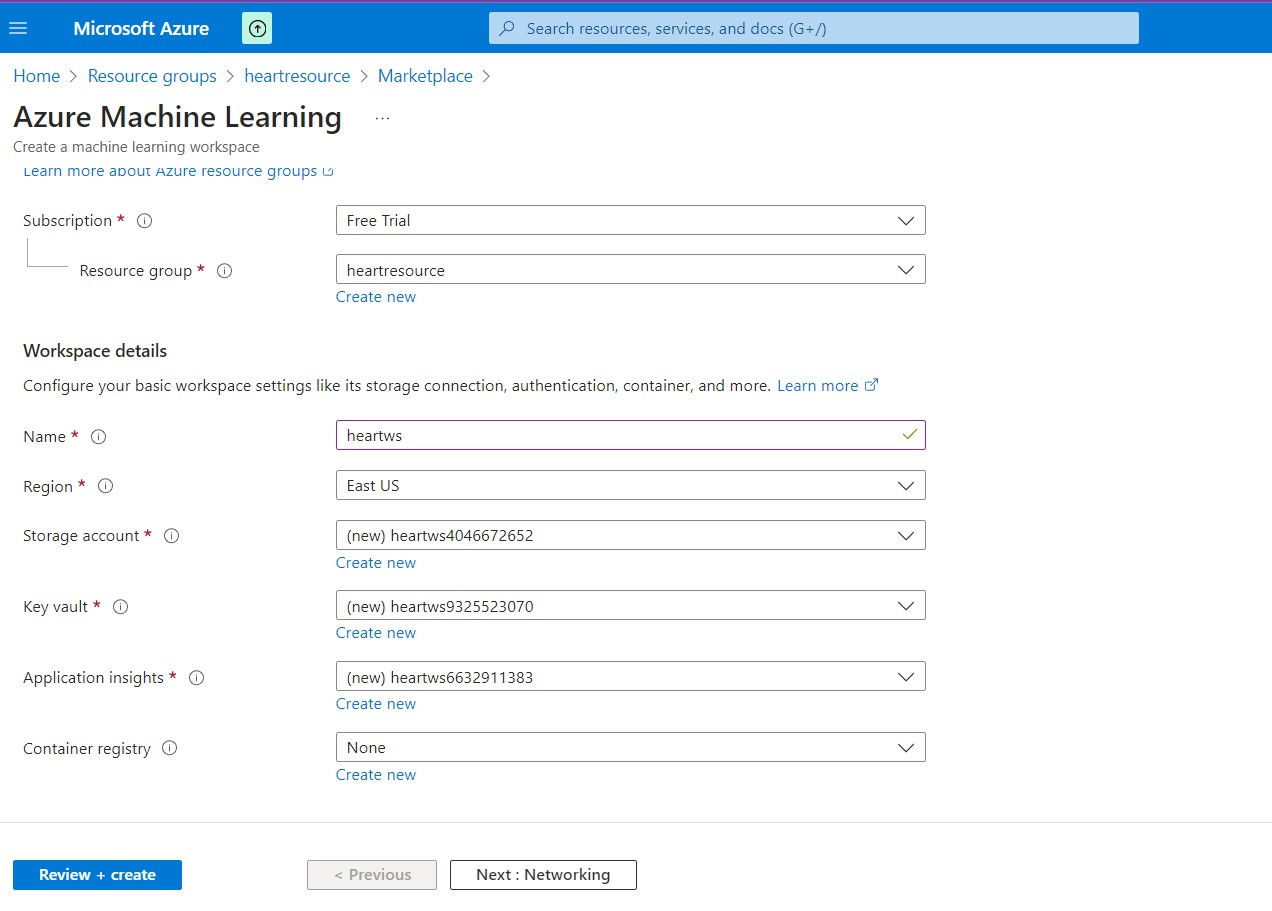
3.Click on Create:Once we find the Resource Group service, click on it.

4.Providing Information:we'll usually see a "Create" button. Click on it.

5.Enter a name for your resource group.Choose the location (in this case, East US).

6.Complete the Creation.

**Creating a Workspace:**



1.After creating resource group, go back to the created resource group and click on create button

2.Look for the search bar in marketplace and enter "Azure Machine Learning" to find the service.

3.Click on the Azure Machine Learning service in the search results.

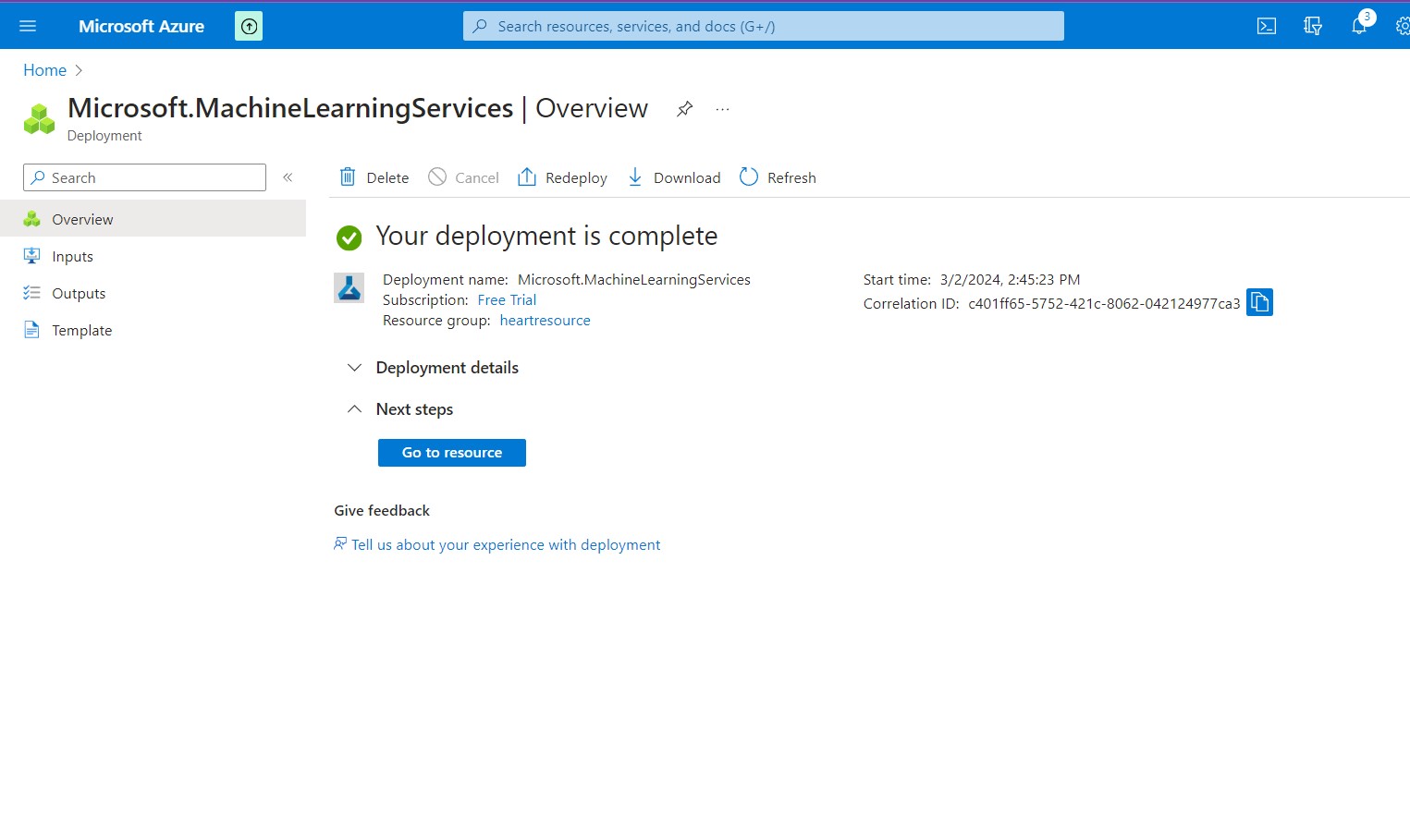
4.We'll typically find a "Create" or "Add" button. Click on it to start the process of creating an Azure Machine Learning workspace.

5.Provide Workspace Details:

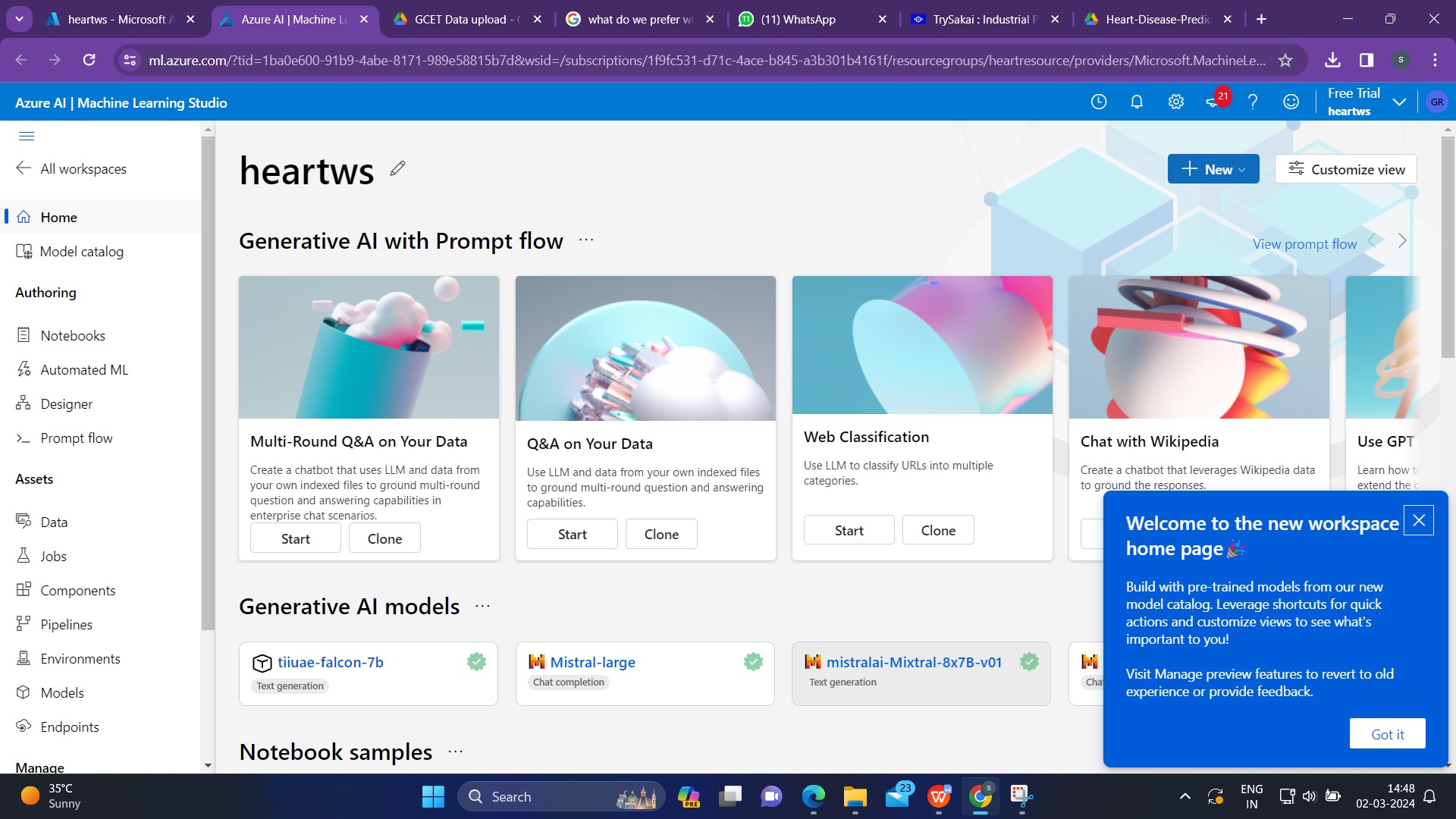
* During the creation process, you'll be prompted to provide details for your Azure Machine Learning workspace.
* Enter a unique name for your workspace.
* Select the subscription you want to use.
* Choose the resource group you created earlier.

6.Specify the region or location (East US).

7.Complete the Creation



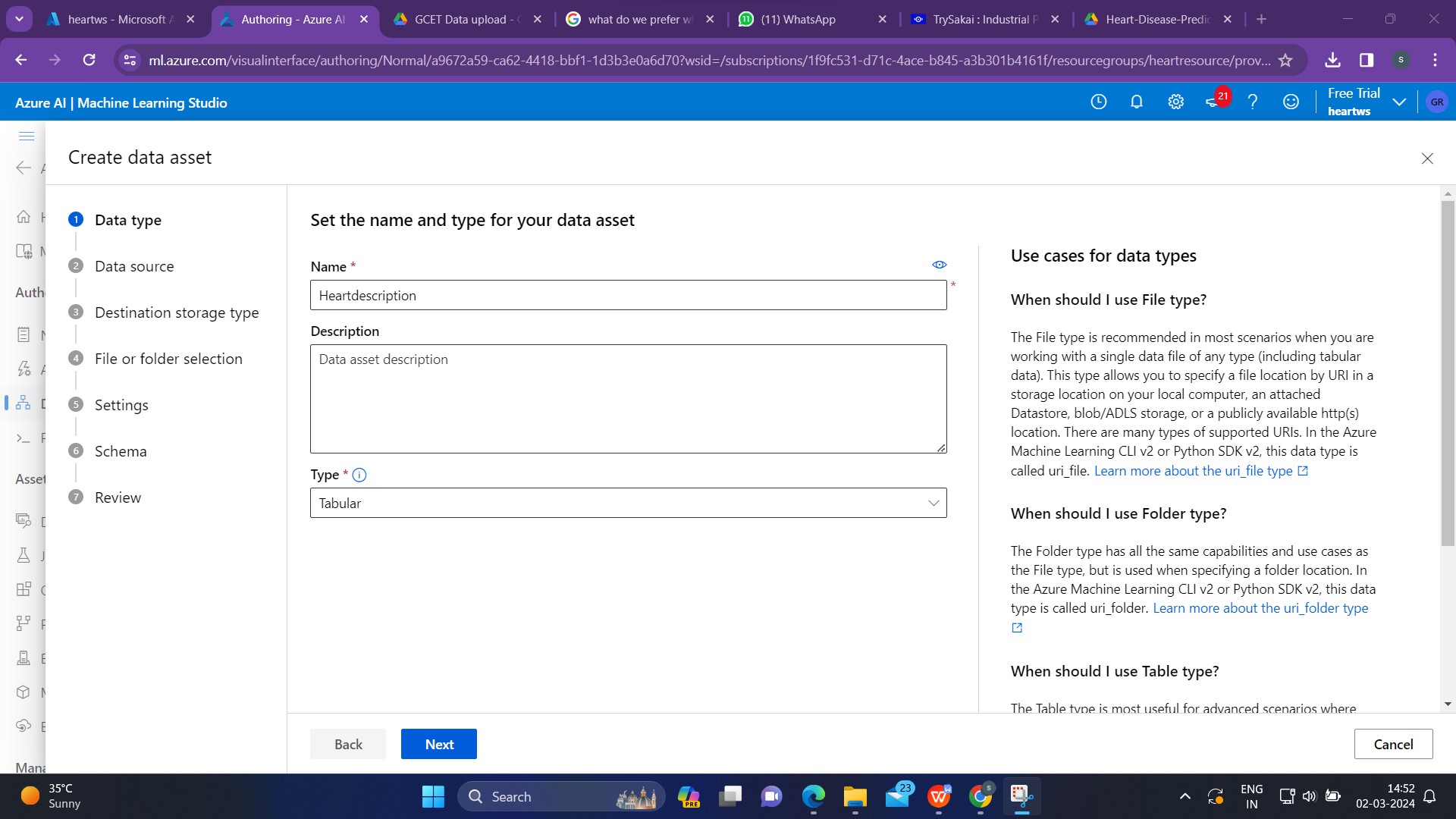
**Data Uploading**



After successfully completing the deployment,go to the workspace which we have created and launch it.

Within the Azure Machine Learning studio, look for a section or tab labeled "Design" or "Designer." Click on it then next click on ‘+” for creating pipline .

**Data type:**

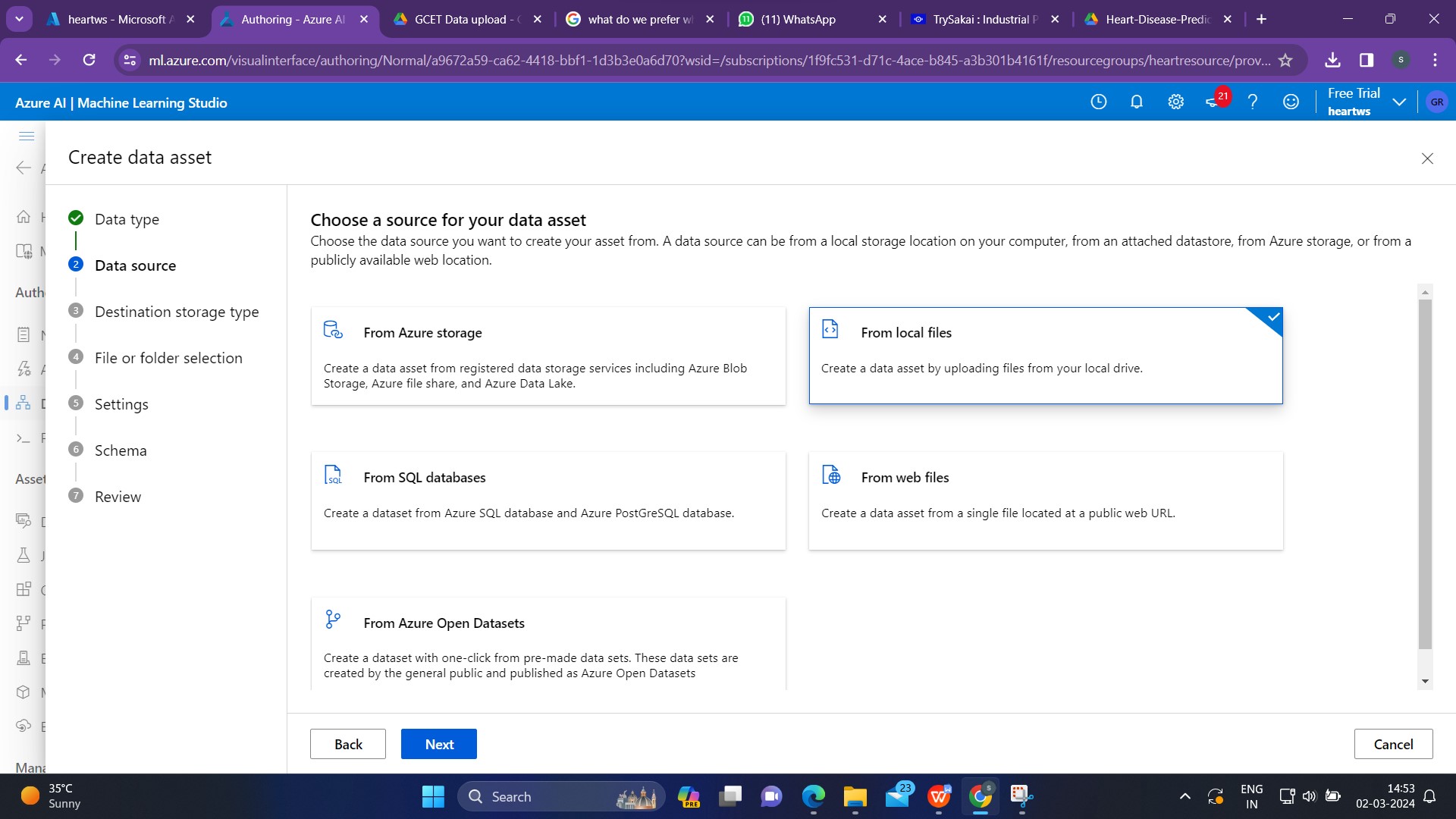


Within the Designers section, locate the option for working with data.

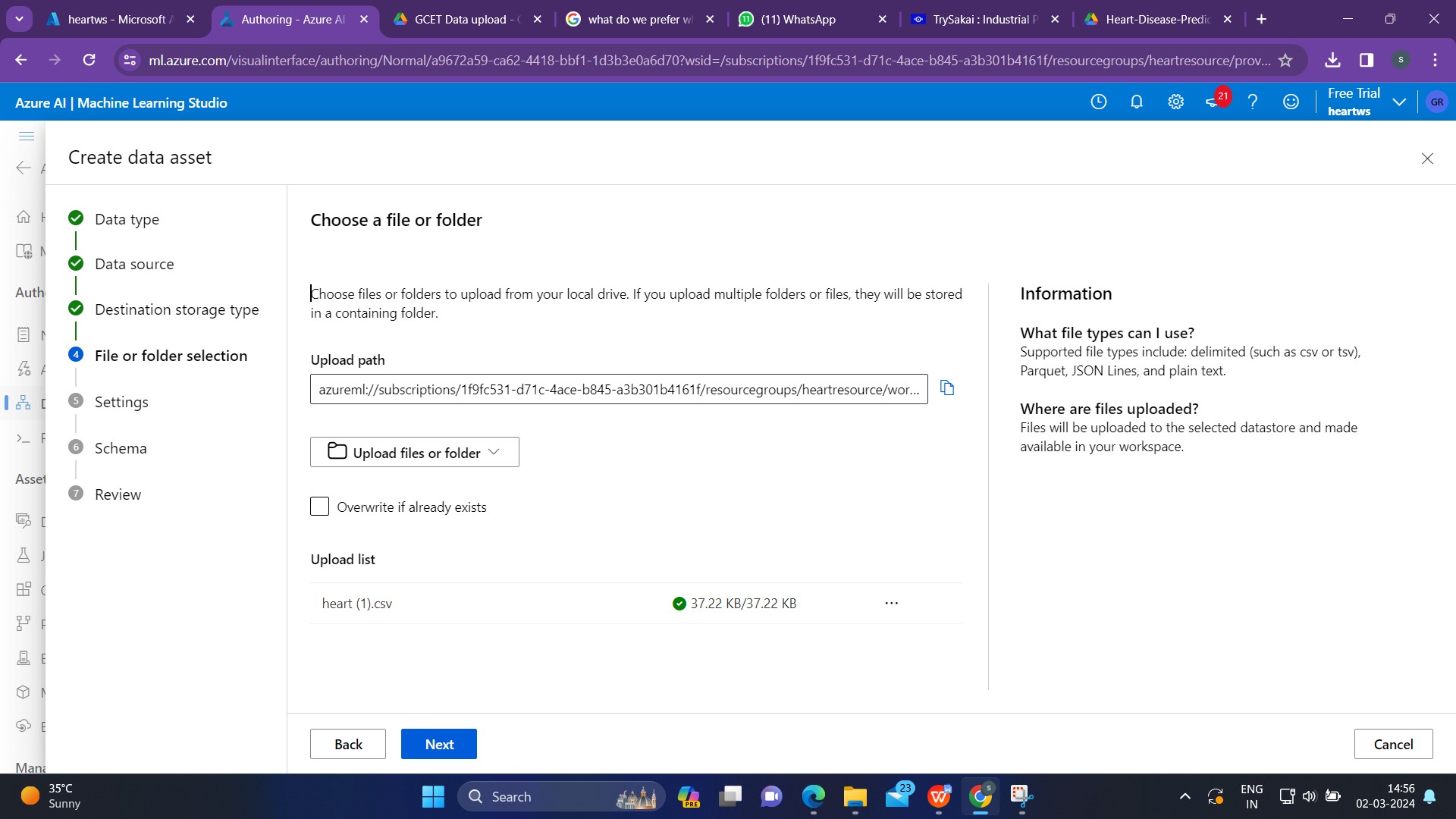
Once we're in the data section,create a new data asset by clicking on “Create” button.

Provide the necessary details for data asset, including the name, description(optional).

**Data source:**

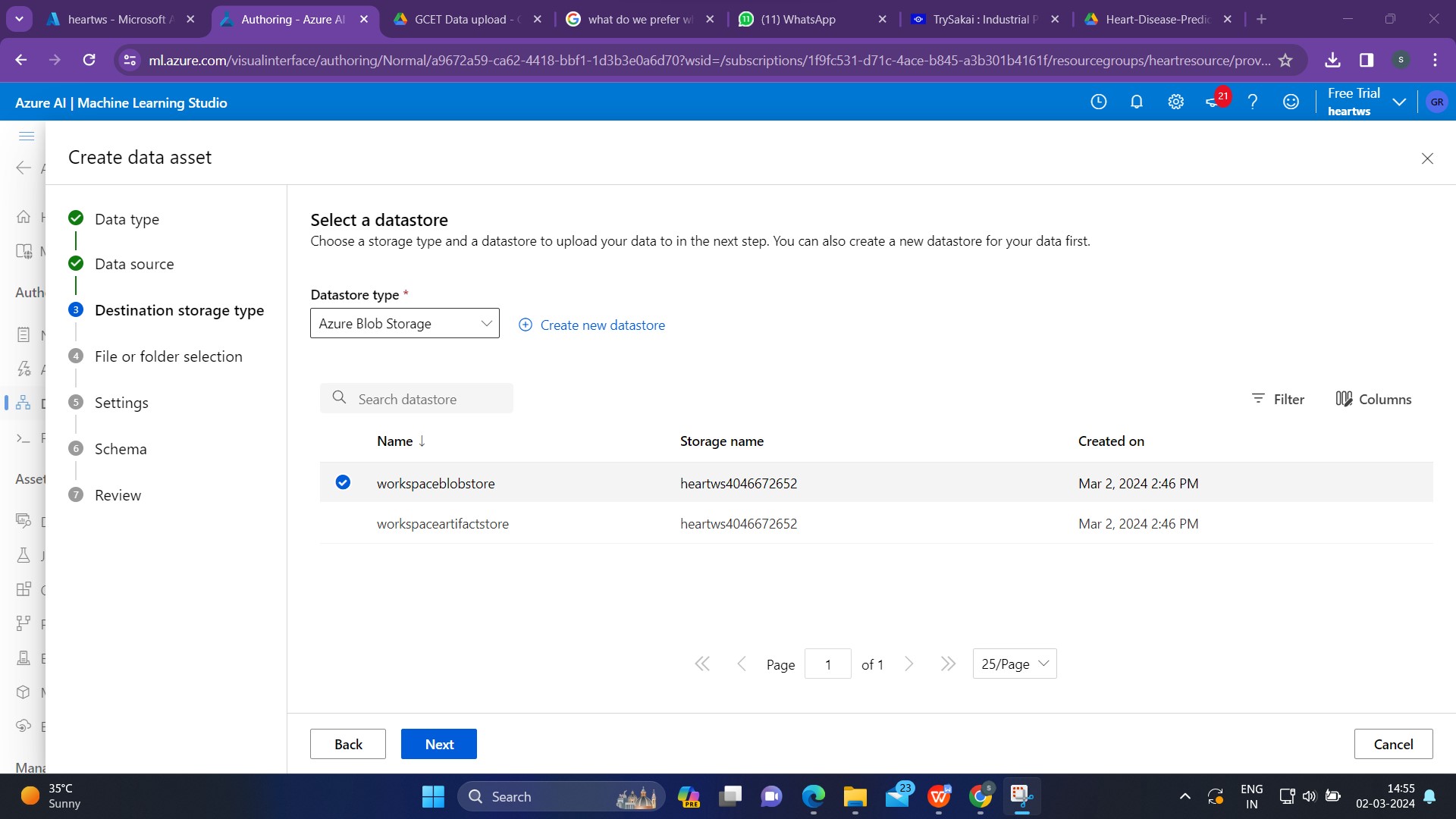


For choosing a data source,select the option “From local files”.



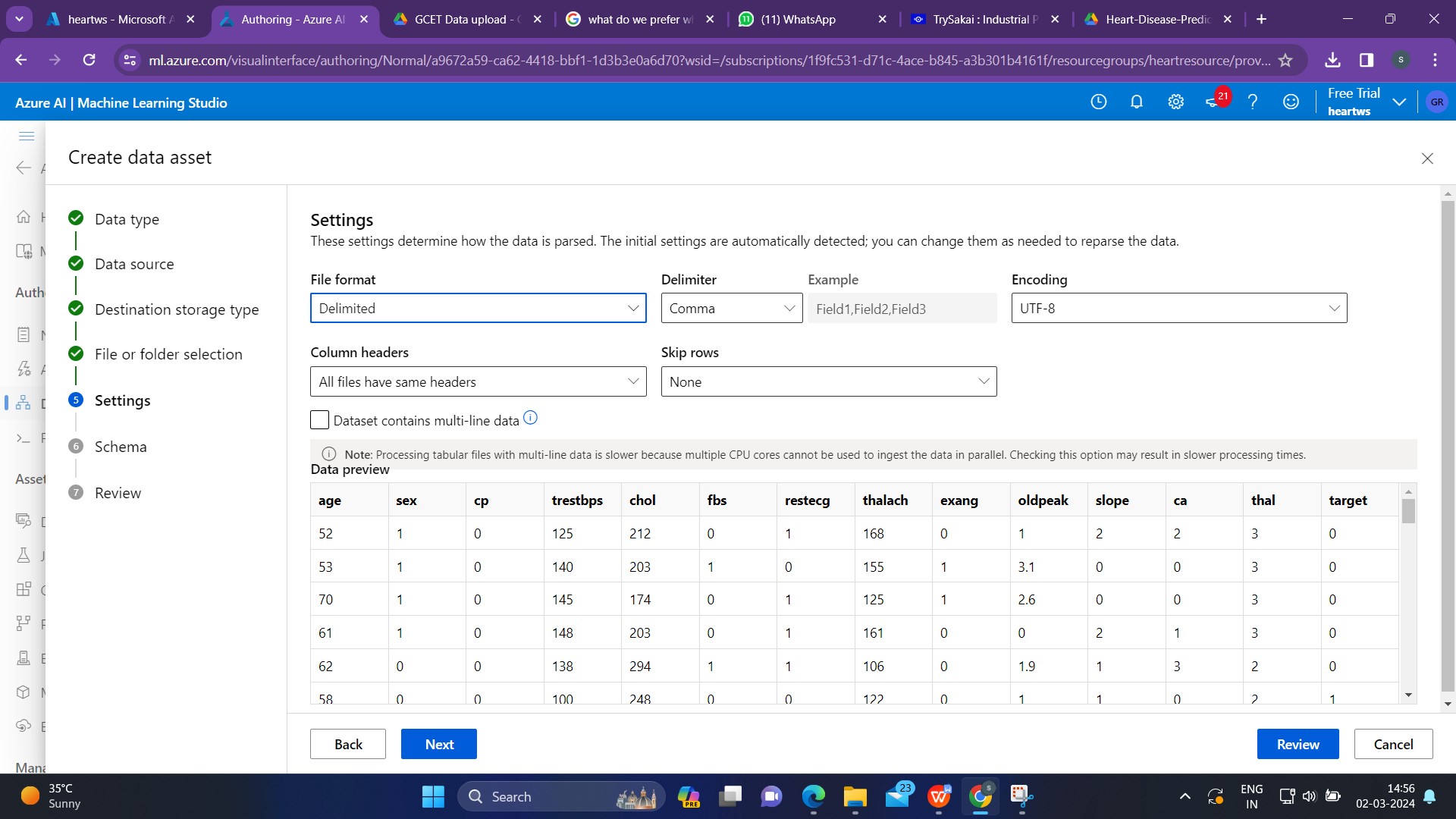
we'll likely see a button that enables us to upload files. Click on it to browse our local computer and select the CSV file we want to upload.

Destination Storage type:



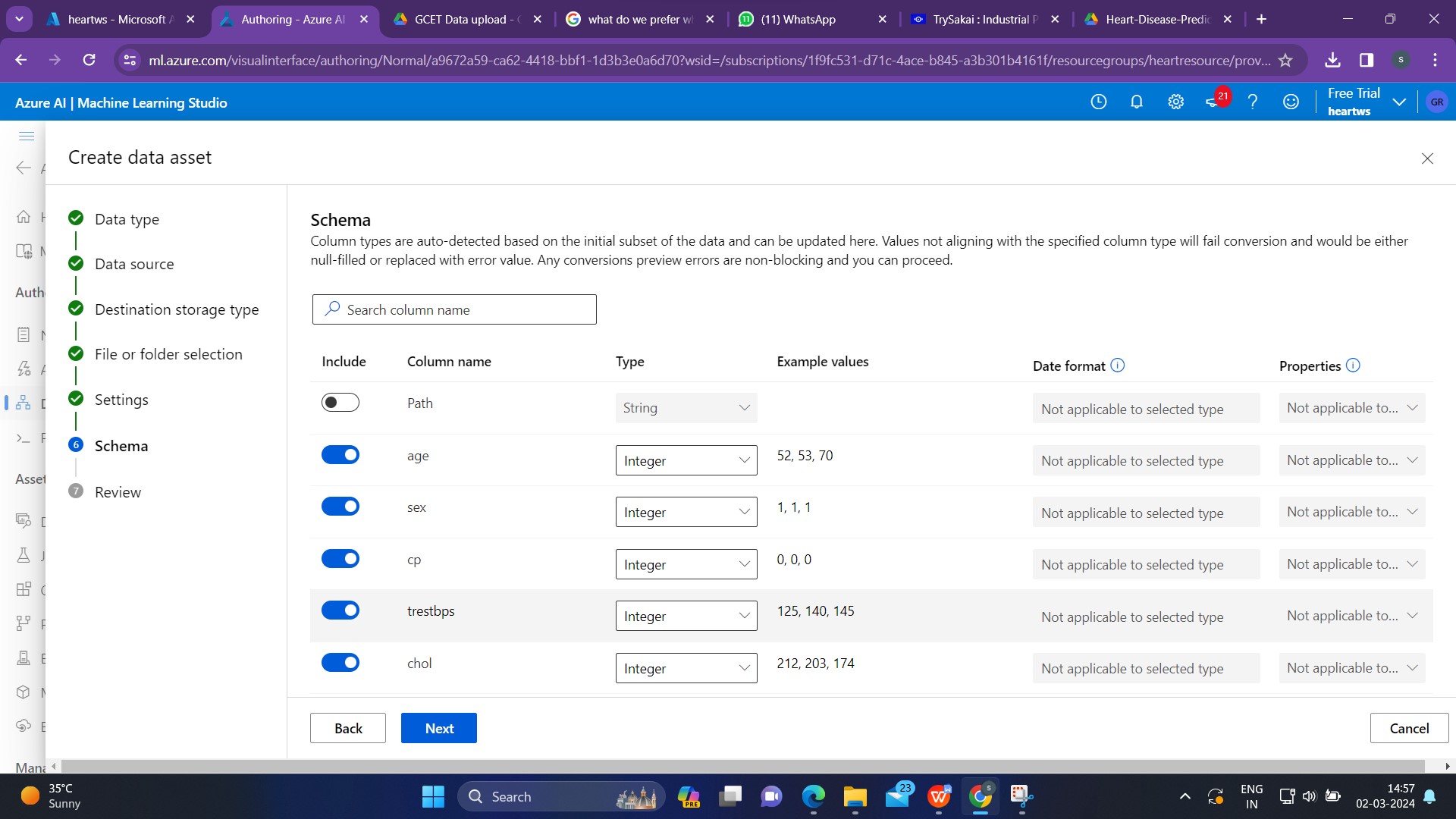
Always select Azure blob store.

Settings:



Here it will show the columns present in our data set and all the details of our data set.If we want to change anything we can change it here.

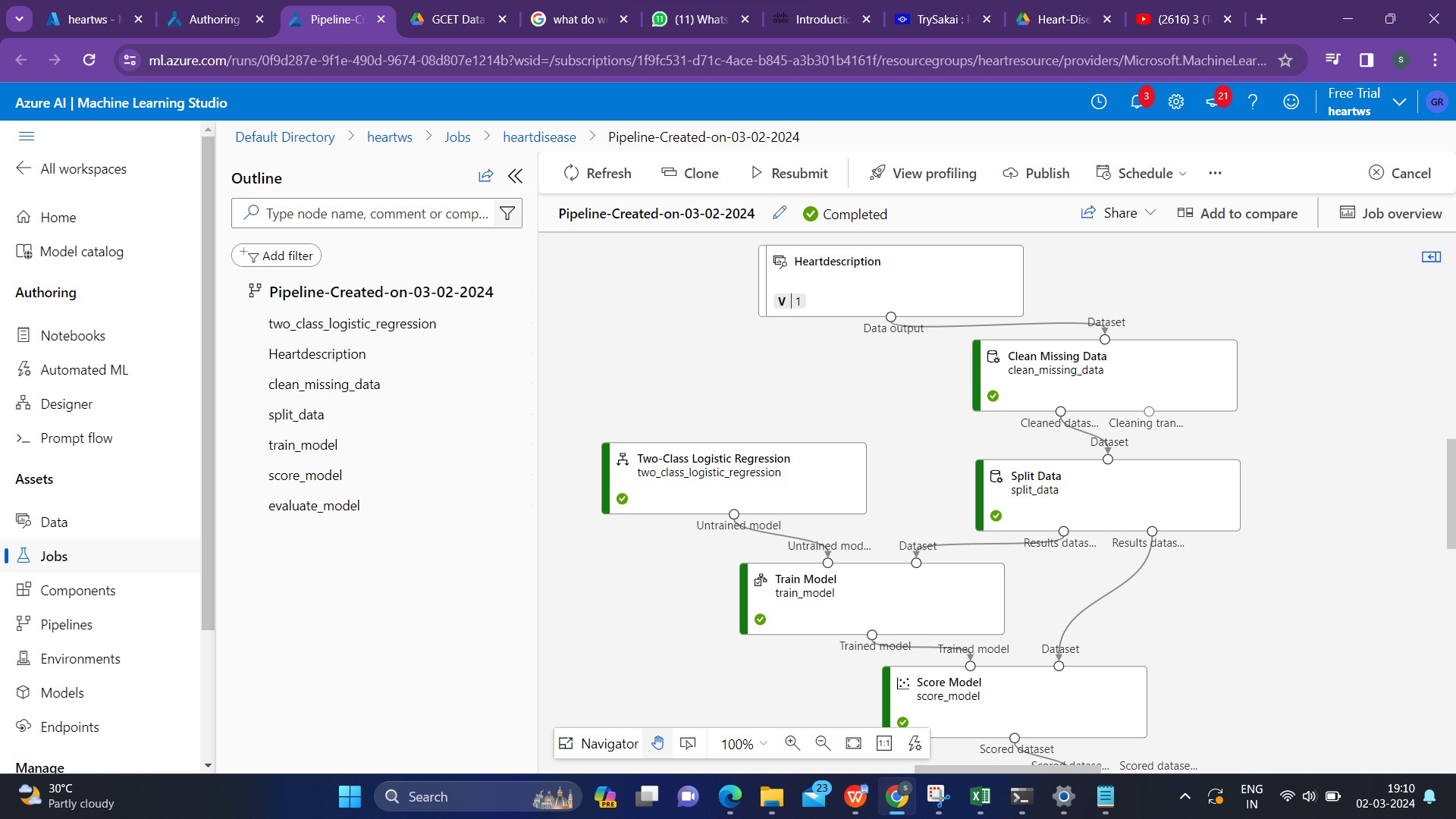
Schema:

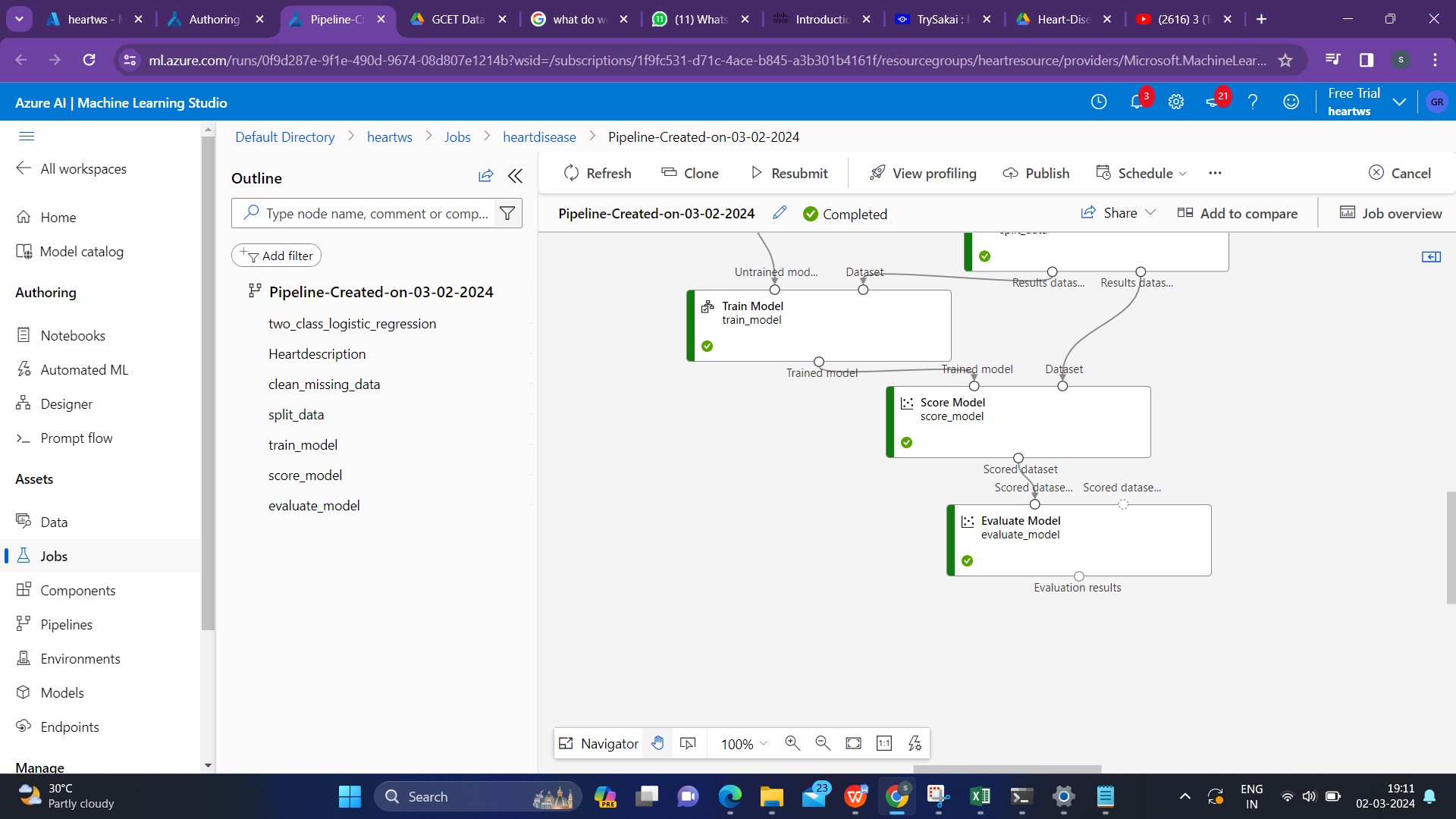


Schema step allows us to disable a column which has uniq values .

And last step is to review and click on create and data asset is created.

**Ml Pipeline Flow**





Here's a simplified flow of how a typical machine learning pipeline might look:

Data Ingestion:

The pipeline starts with data ingestion, where raw data is collected from various sources. This could include Azure Blob Storage, Azure SQL Database, or other data repositories.

Data Preparation:

Once the data is ingested, it undergoes preprocessing and cleaning.

For cleaning double click the **Clean Missing Data** module, and in the pane on the right, select **Edit column**. Then in the **Columns to be cleaned** window, select **With rules**, in the **Include** list select **Column names**.

This step includes handling missing values, transforming features, and preparing the data for model training.

Model Training:

The preprocessed data is used to train machine learning models. This step involves selecting an algorithm, splitting the data into training and validation sets, and training the model using historical data.

For splitting the data double click on Split data and in th fraction column write the ratio in which we want to split our dataset.

Model Evaluation:

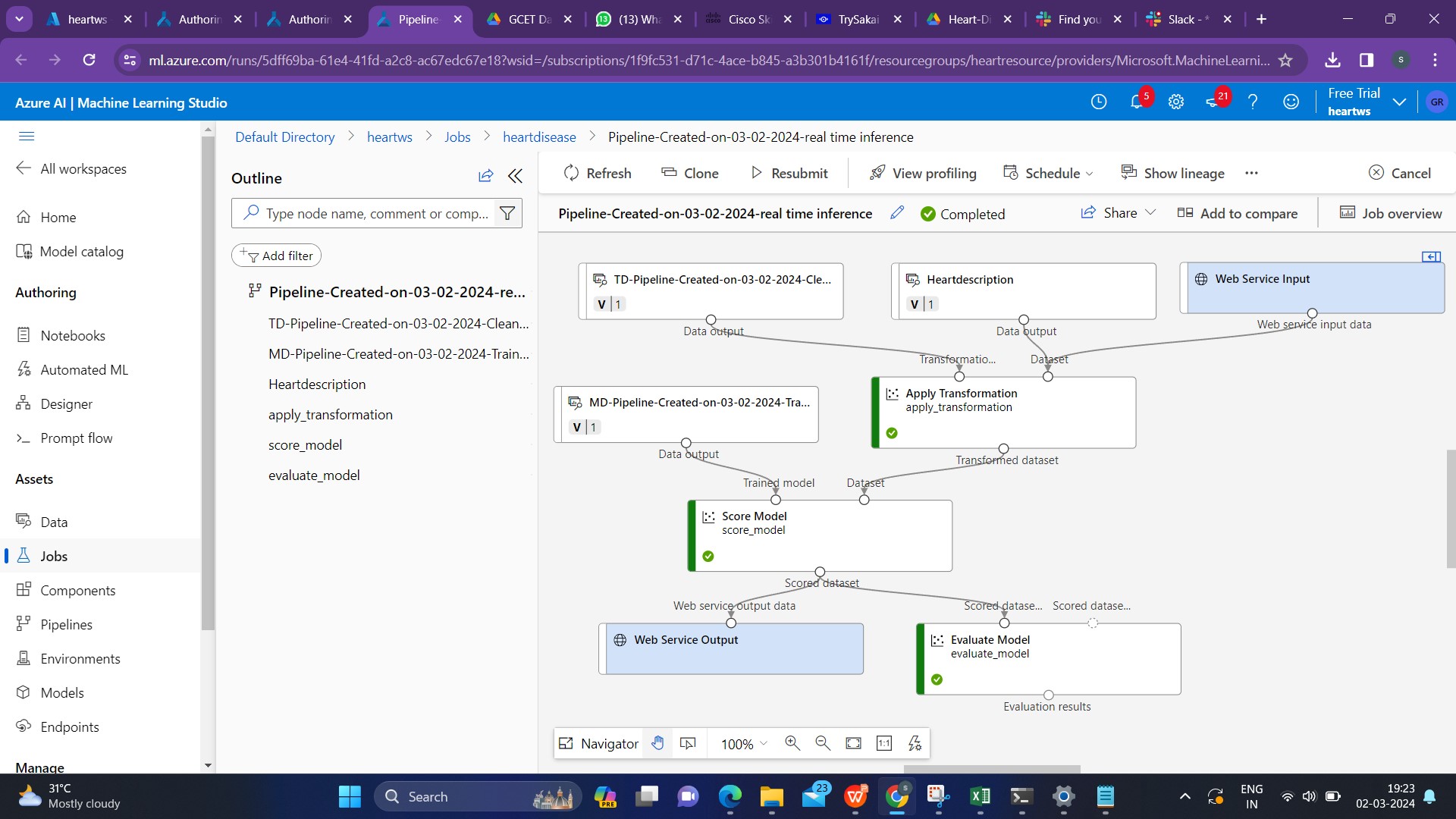
The trained model is then evaluated using a separate set of data not used during training. This step helps assess the model's performance and ensures it generalizes well to new, unseen data.

Model Deployment:

SelectConfigure & Submit at the top of the page to open the Set up pipeline job dialogue.

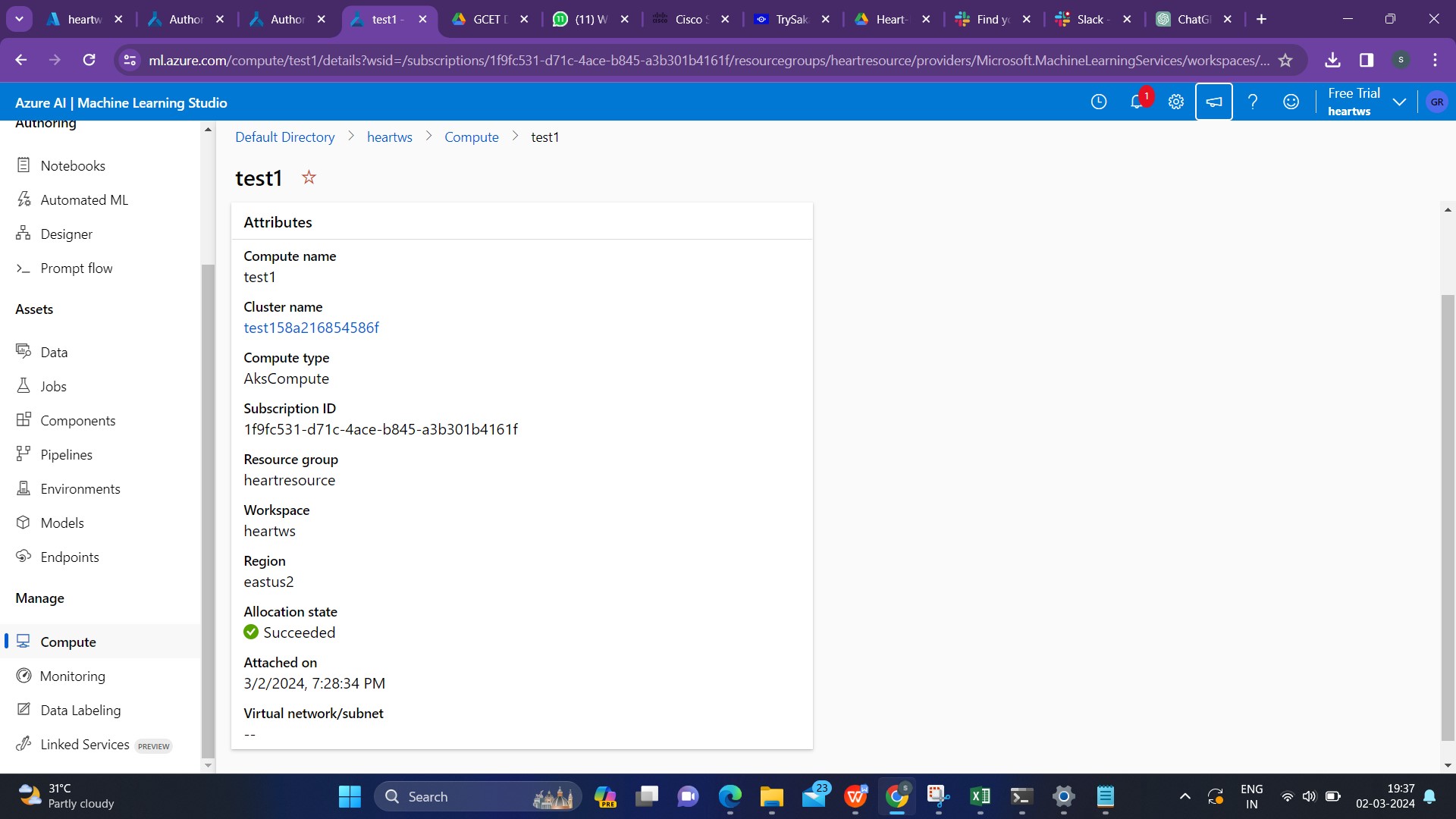
Once a satisfactory model is obtained, it can be deployed for inference. This involves making predictions on new data using the trained model.

**Real Time Inference Pipeline:**



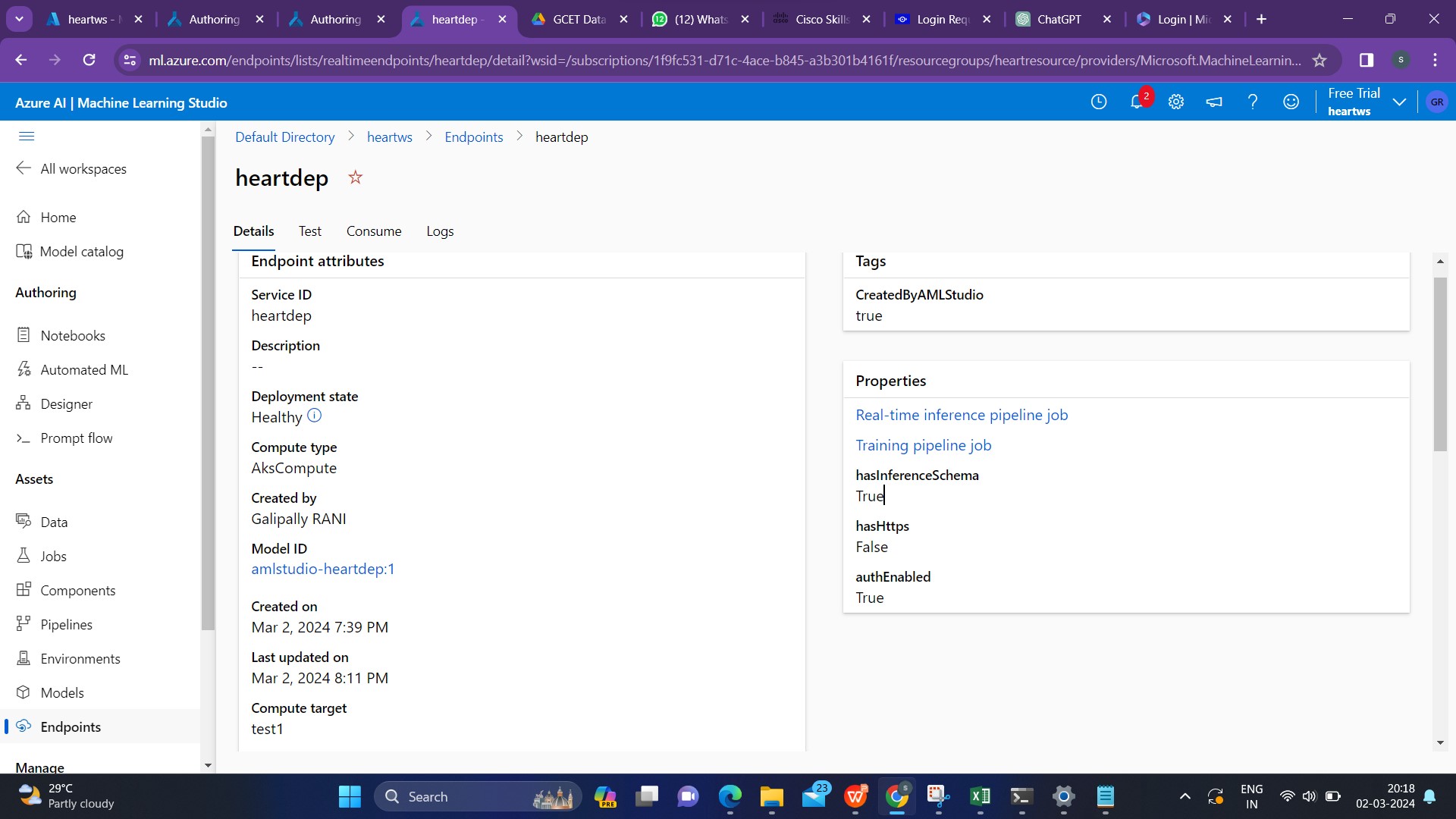
1. In the Create inference pipeline drop-down list, select **Real-time inference pipeline**. After few seconds, a new version of pipeline  will be opened.
2. Go to components and search for web service input and drag it to our model which is right side of the canvas.
3. Web service input is basically for the user to give input and Web Server Output is for getting the output.

**Compute Cluster**



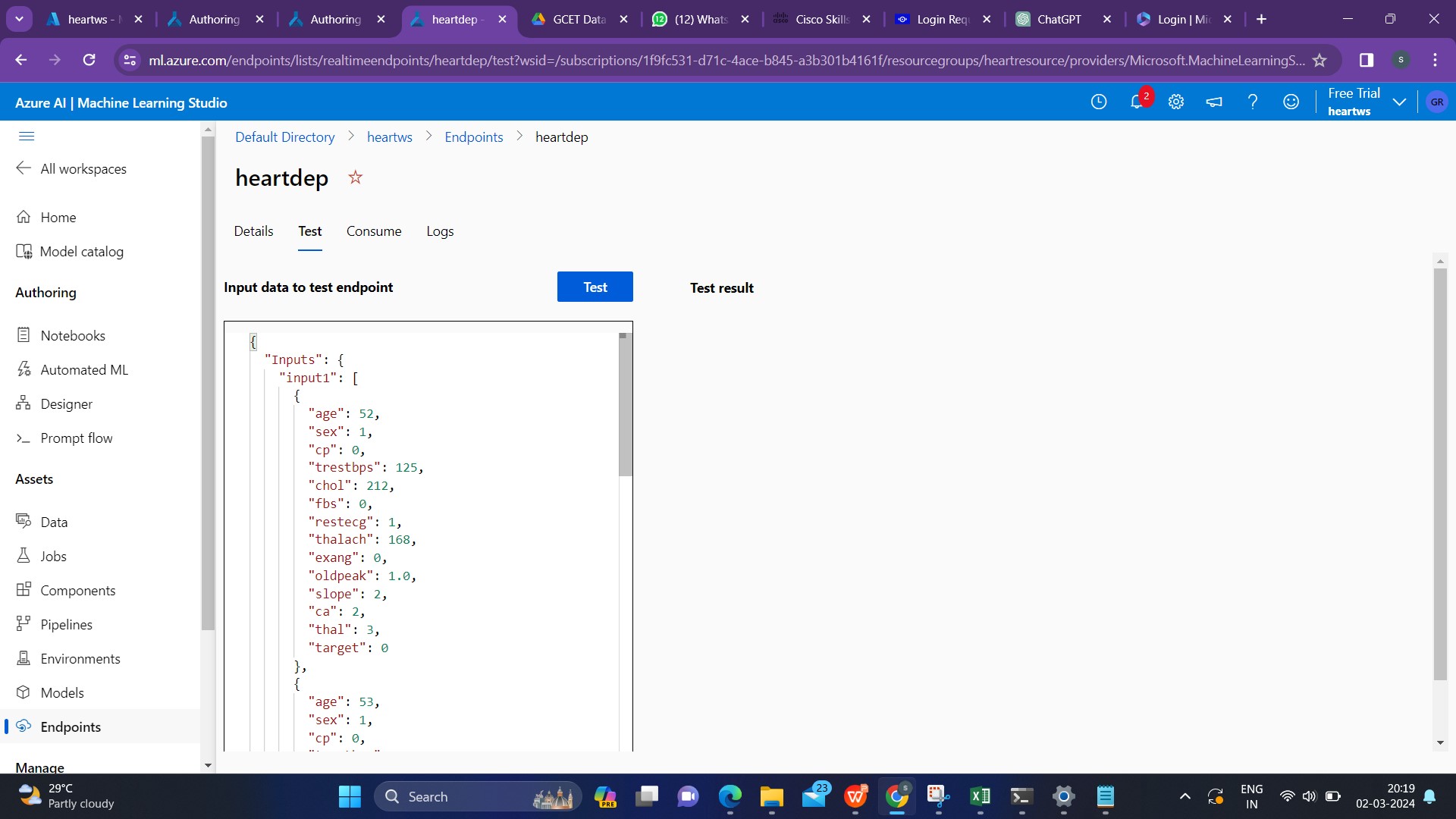
After successfully deploying the real time inference,go to compute and create a  **cluster (Ask compute) In Kubemetes clusters. As we can see the above fig,it shows a Ask compute cluster called test1.**

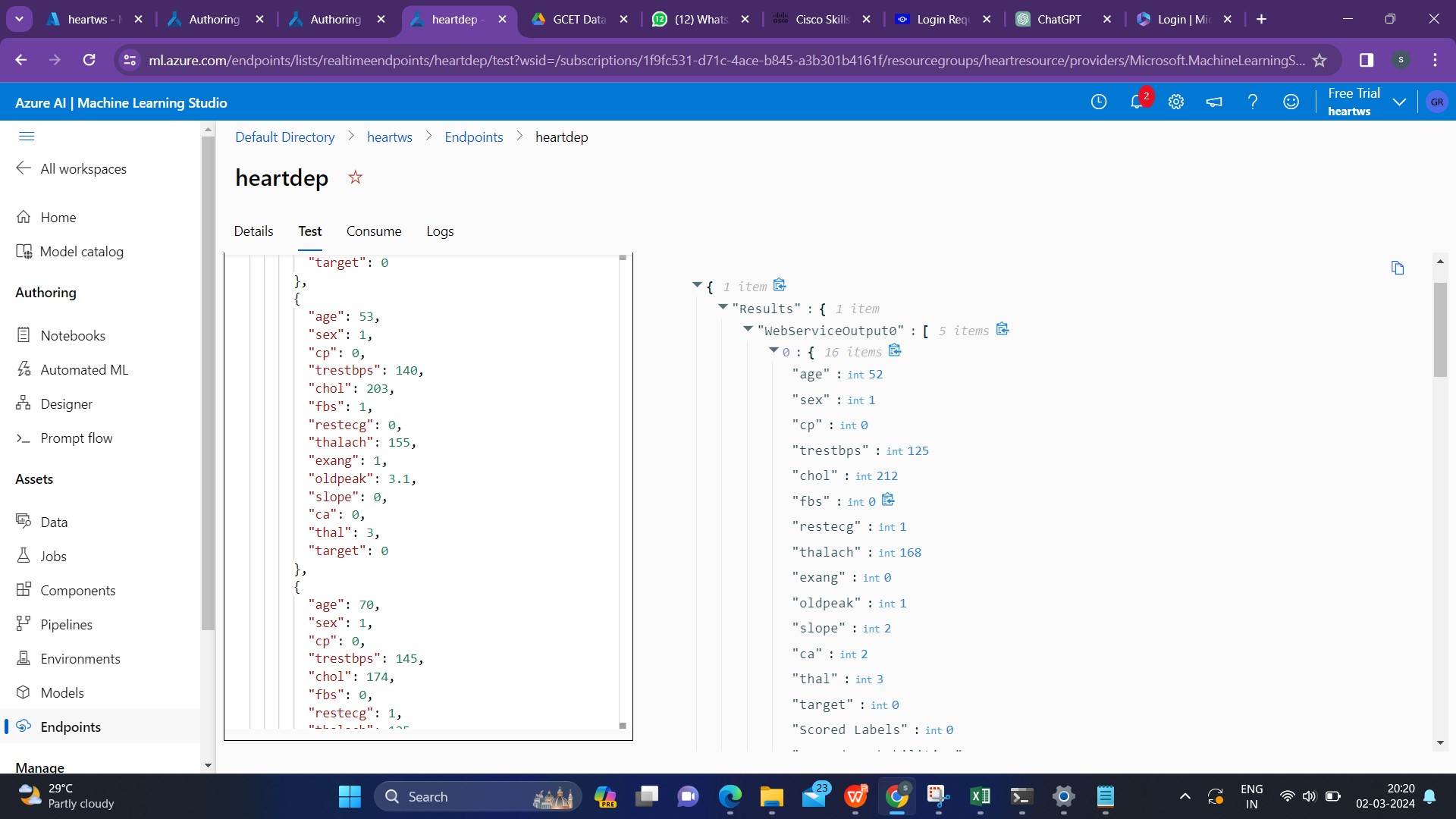
**Endpoints**



Next go to Endpoints and click on deploy and wait until the Deployment state changes to healthy state.

So after successful deployment go to the **Test** and click on test to get the output/Test result for the data in the data set.





INPUT:

{

"Inputs": {

"input1": [

{

"age": 52,

"sex": 1,

"cp": 0,

"trestbps": 125,

"chol": 212,

"fbs": 0,

"restecg": 1,

"thalach": 168,

"exang": 0,

"oldpeak": 1.0,

"slope": 2,

"ca": 2,

"thal": 3,

"target": 0

},

{

"age": 53,

"sex": 1,

"cp": 0,

"trestbps": 140,

"chol": 203,

"fbs": 1,

"restecg": 0,

"thalach": 155,

"exang": 1,

"oldpeak": 3.1,

"slope": 0,

"ca": 0,

"thal": 3,

"target": 0

},

{

"age": 70,

"sex": 1,

"cp": 0,

"trestbps": 145,

"chol": 174,

"fbs": 0,

"restecg": 1,

"thalach": 125,

"exang": 1,

"oldpeak": 2.6,

"slope": 0,

"ca": 0,

"thal": 3,

"target": 0

},

{

"age": 61,

"sex": 1,

"cp": 0,

"trestbps": 148,

"chol": 203,

"fbs": 0,

"restecg": 1,

"thalach": 161,

"exang": 0,

"oldpeak": 1.0,

"slope": 2,

"ca": 1,

"thal": 3,

"target": 0

},

{

"age": 62,

"sex": 0,

"cp": 0,

"trestbps": 138,

"chol": 294,

"fbs": 1,

"restecg": 1,

"thalach": 106,

"exang": 0,

"oldpeak": 1.9,

"slope": 1,

"ca": 3,

"thal": 2,

"target": 0

}

]

},

"GlobalParameters": {}

}

OUTPUT:

{

1 item

"Results":{1 item

"WebServiceOutput0":[5 items

0:{16 items

"age":int52

"sex":int1

"cp":int0

"trestbps":int125

"chol":int212

"fbs":int0

"restecg":int1

"thalach":int168

"exang":int0

"oldpeak":int1

"slope":int2

"ca":int2

"thal":int3

"target":int0

"Scored Labels":int0

"Scored Probabilities":float0.2828159049000958

}

1:{16 items

"age":int53

"sex":int1

"cp":int0

"trestbps":int140

"chol":int203

"fbs":int1

"restecg":int0

"thalach":int155

"exang":int1

"oldpeak":float3.1

"slope":int0

"ca":int0

"thal":int3

"target":int0

"Scored Labels":int0

"Scored Probabilities":float0.027217881321284845

}

2:{16 items

"age":int70

"sex":int1

"cp":int0

"trestbps":int145

"chol":int174

"fbs":int0

"restecg":int1

"thalach":int125

"exang":int1

"oldpeak":float2.6

"slope":int0

"ca":int0

"thal":int3

"target":int0

"Scored Labels":int0

"Scored Probabilities":float0.03423137634944725

}

3:{16 items

"age":int61

"sex":int1

"cp":int0

"trestbps":int148

"chol":int203

"fbs":int0

"restecg":int1

"thalach":int161

"exang":int0

"oldpeak":int1

"slope":int2

"ca":int1

"thal":int3

"target":int0

"Scored Labels":int0

"Scored Probabilities":float0.331548835643165

}

4:{16 items

"age":int62

"sex":int0

"cp":int0

"trestbps":int138

"chol":int294

"fbs":int1

"restecg":int1

"thalach":int106

"exang":int0

"oldpeak":float1.9

"slope":int1

"ca":int3

"thal":int2

"target":int0

"Scored Labels":int0

"Scored Probabilities":float0.10305861811683052

}

]

}

}

**Observation**

* The model's predictions suggest that the fourth individual is more likely to have heart disease compared to the others.
* The second individual has the lowest probability, indicating a lower risk of heart disease in this case.
* The first, third, and fifth individuals fall in between, with varying probabilities.
* The model considers demographic data, clinical symptoms, and test results for predicting heart disease.
* Traditional risk factors are present in all cases, yet the model predicts a low probability of heart disease.
* This highlights the complexity of cardiovascular disease prediction and the importance of comprehensive risk assessment.
* The differences in predicted probabilities show the personalized nature of health risk assessments.