**Lab 17: Calculating embeddings from unsupervised clusters**

**Introduction**

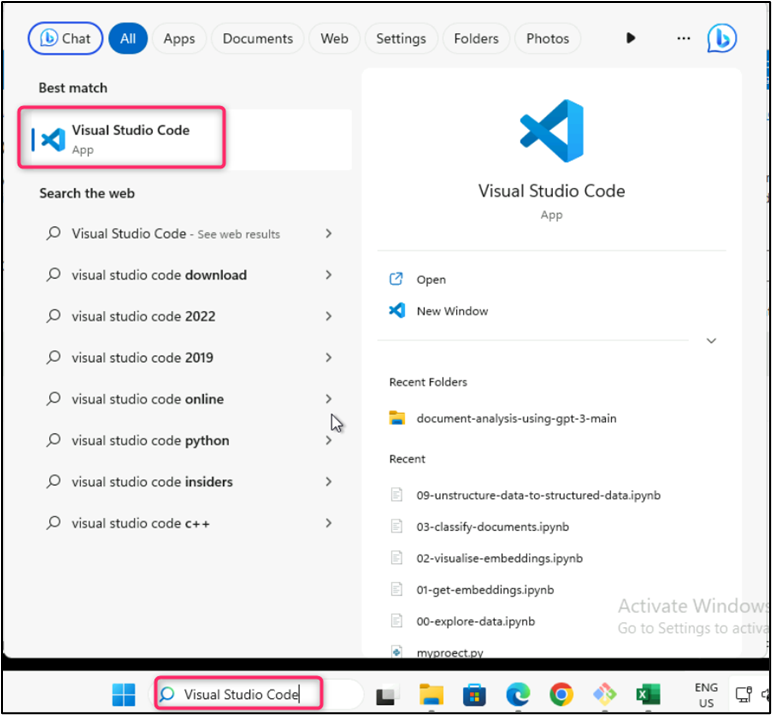
In this lab, you'll be using a subset of the Top 10,000 Popular Movies Dataset to calculate embeddings on movie descriptions and then you'll be applying k-means to find similar clusters. Once you've these clusters, you'll be using a prompt to extract the topics from each cluster.

**Objectives**

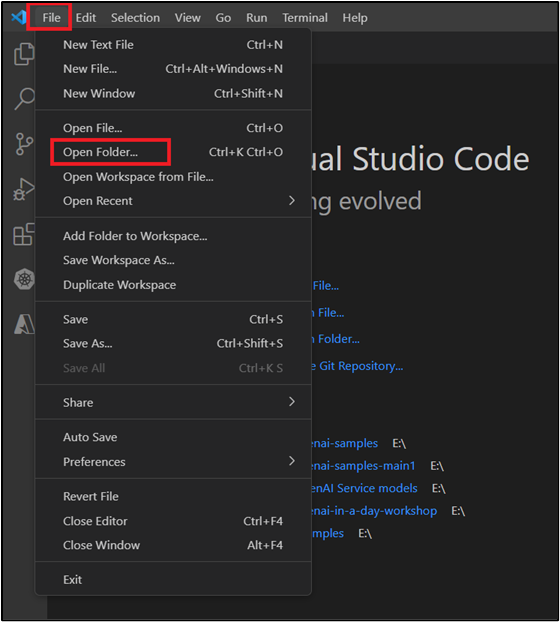
* To configure Azure OpenAI Service, define embedding model and encoding, and load the movies data in Visual Studio Code.
* To perform unsupervised clustering, visualize clusters, and extract topics.
* To build a movie recommendation system using embeddings.

**Task 1: Using Embeddings for unsupervised clustering with named clusters**

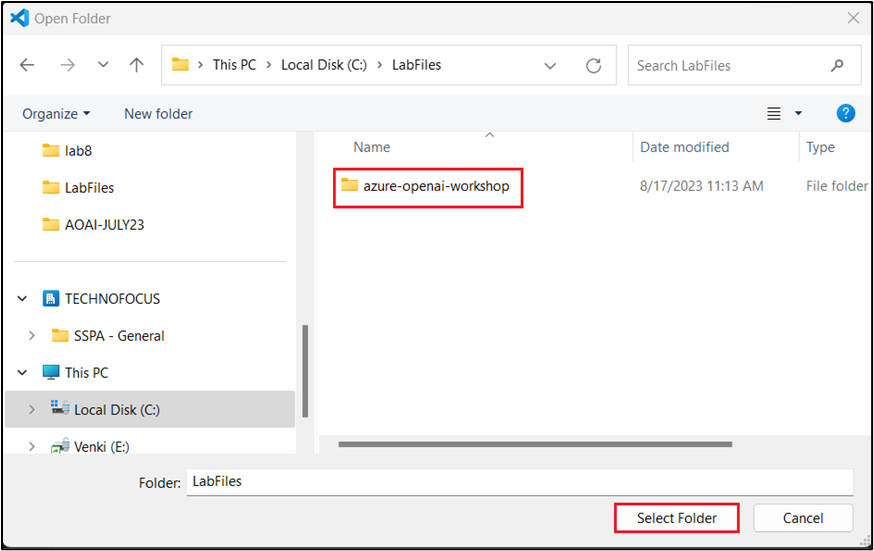
1. In your Windows search box, type Visual Studio Code, then click on **Visual Studio Code**.



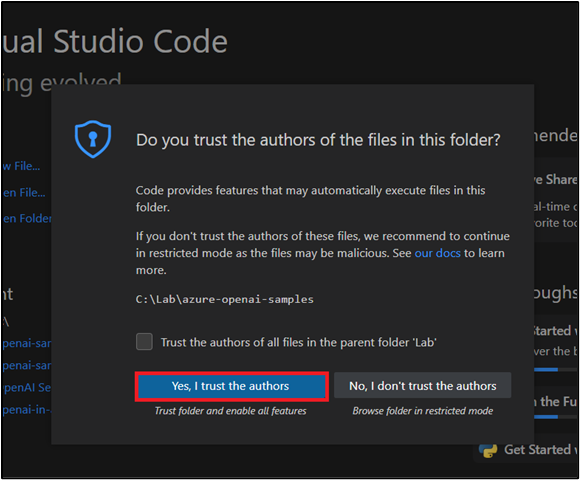
1. In the **Visual Studio Code** editor, click on **File**, then navigate and click on **Open Folder.**



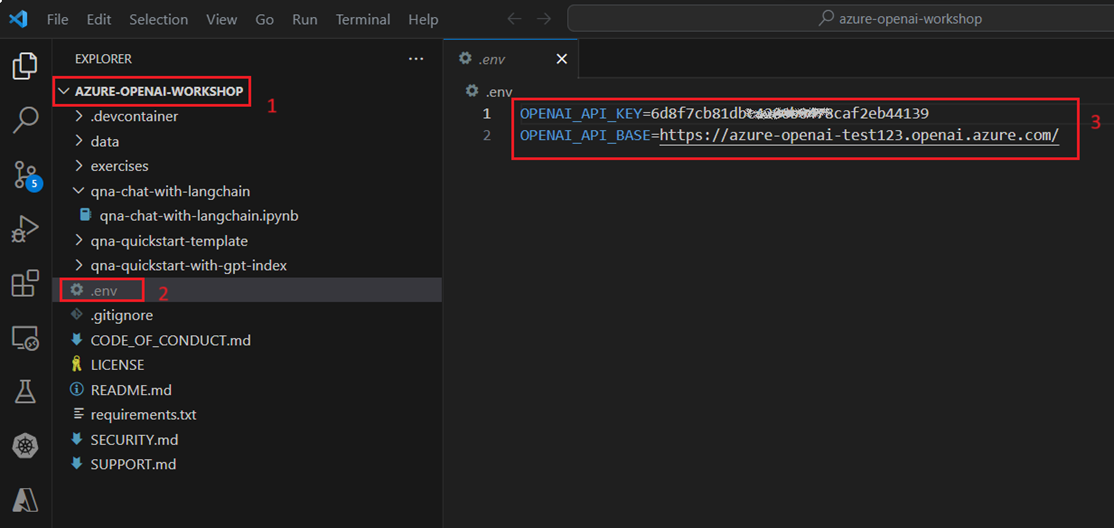
1. Navigate to **C:\LabFiles**, select **azure-openai-workshop** folder.



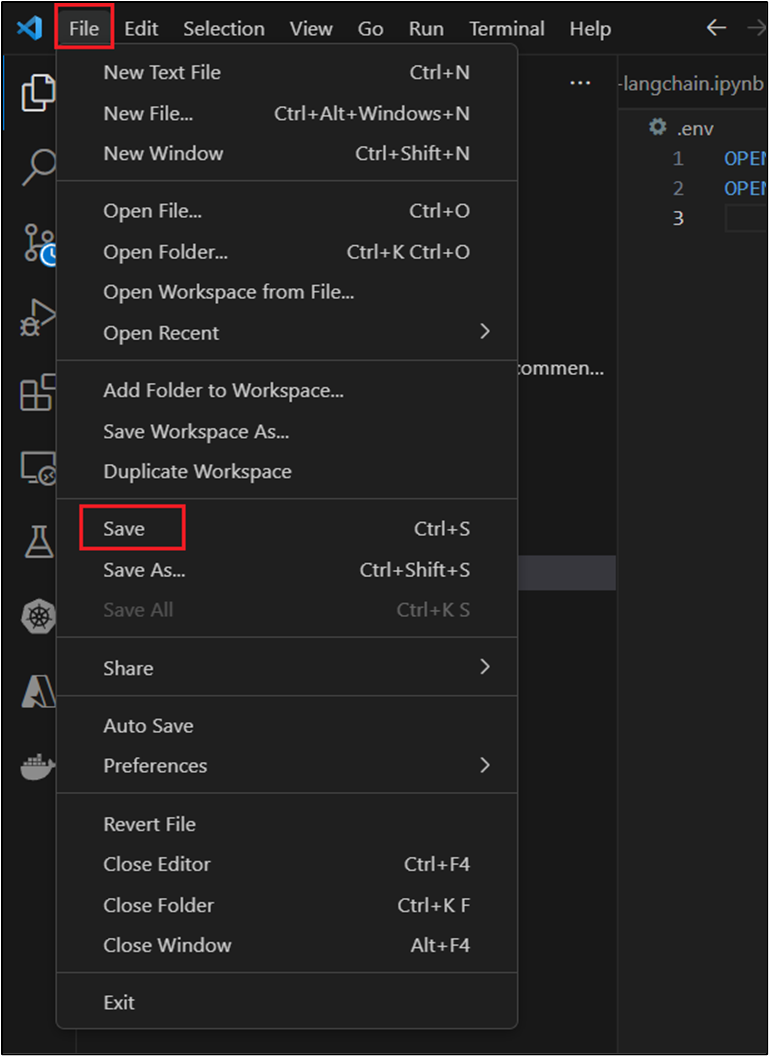
1. If you see a dialog box - **Do you trust the authors of the files in this folder?** then click on **Yes, I trust the author**.



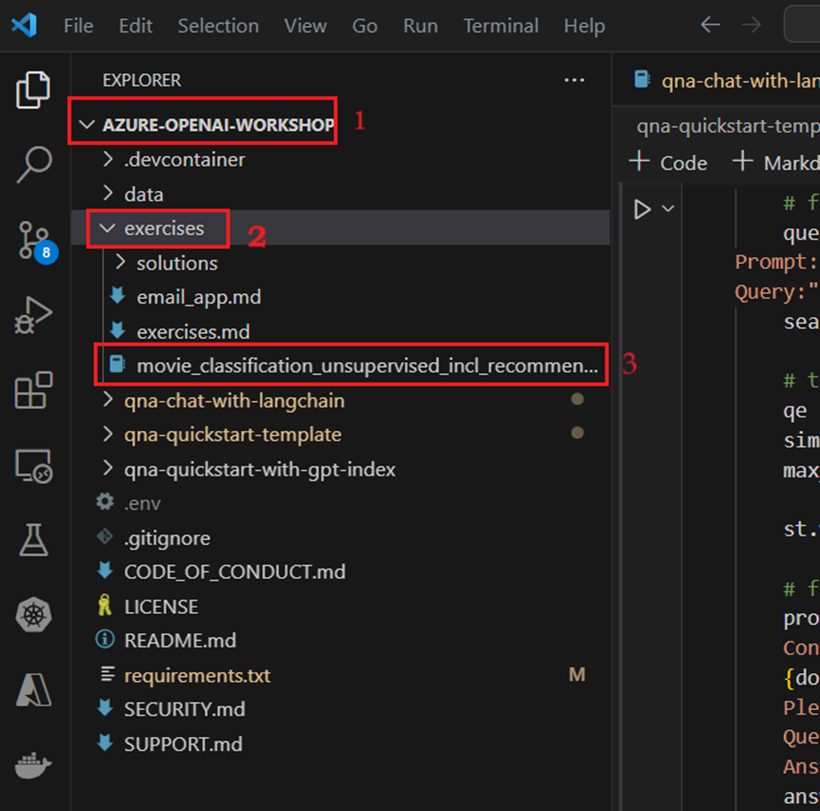
1. In Visual Studio Code, navigate and click on **.env** file.
2. In the **.env** file, replace **OPENAI\_API\_KEY** and **OPEN\_API\_BASE** values with the Azure Openai Service Key and Endpoint values that you have saved in your notepad in the **Lab 1 > Task 2**.



1. Click on **File** and then click on **Save**.



1. In Visual Studio Code, under **AZURE-OPENAI-WORKSHOP**, navigate and click on **exercises** and then click on **movie\_classification\_unsupervised\_incl\_recommendations\_exercise.ipynb** notebook.

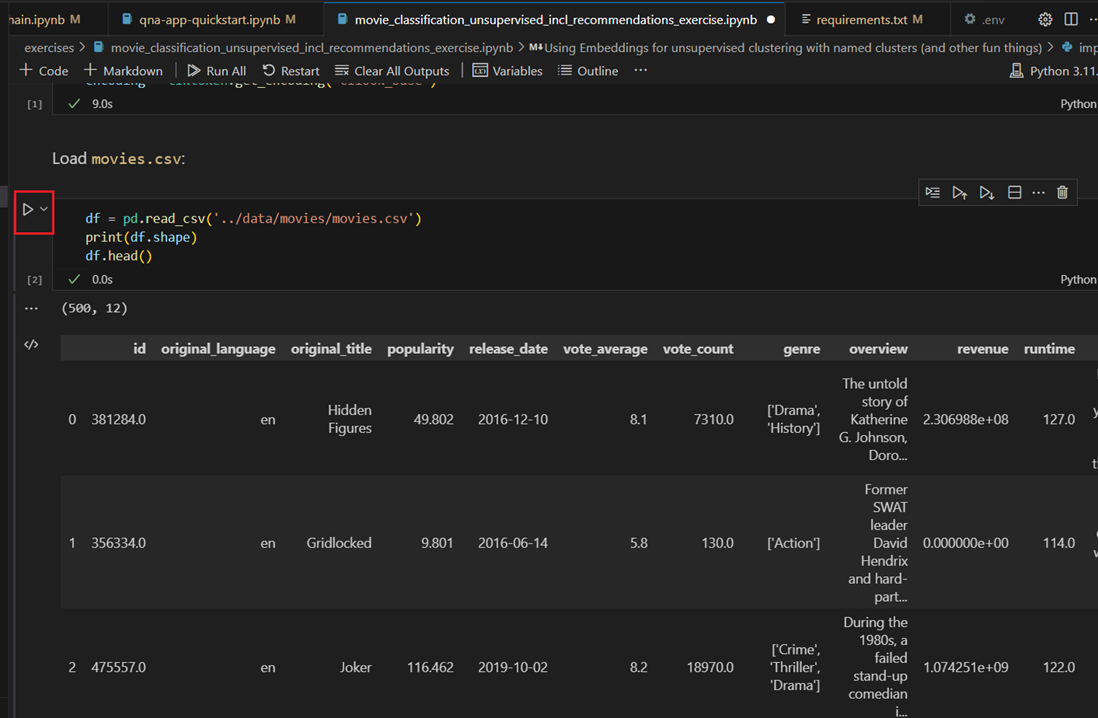


1. In the main page of Visual Studio Code editor, execute the 1st cell by clicking on the **play** button. If prompted to select the environment, then select **Python version 3.11.5(or later version)** path as shown in the below image.

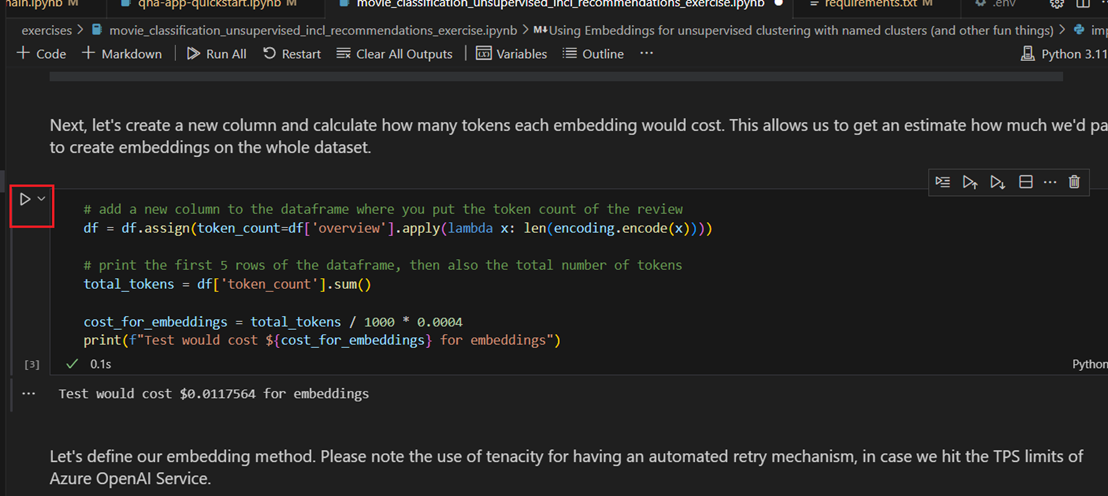
A screenshot of a computer

Description automatically generated

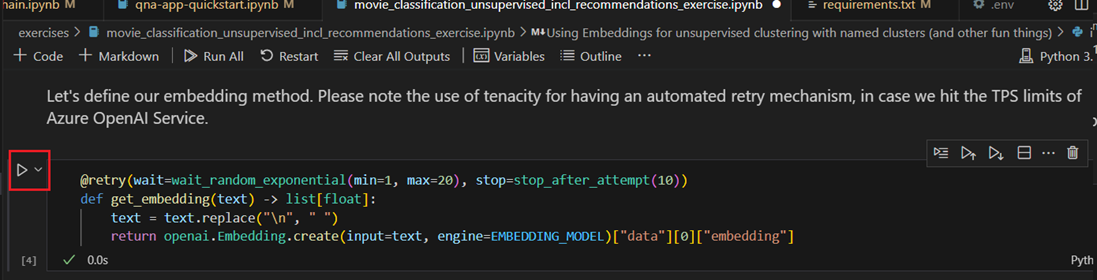
1. To load the movies file, select and execute the 2nd cell by clicking on the **play** button.



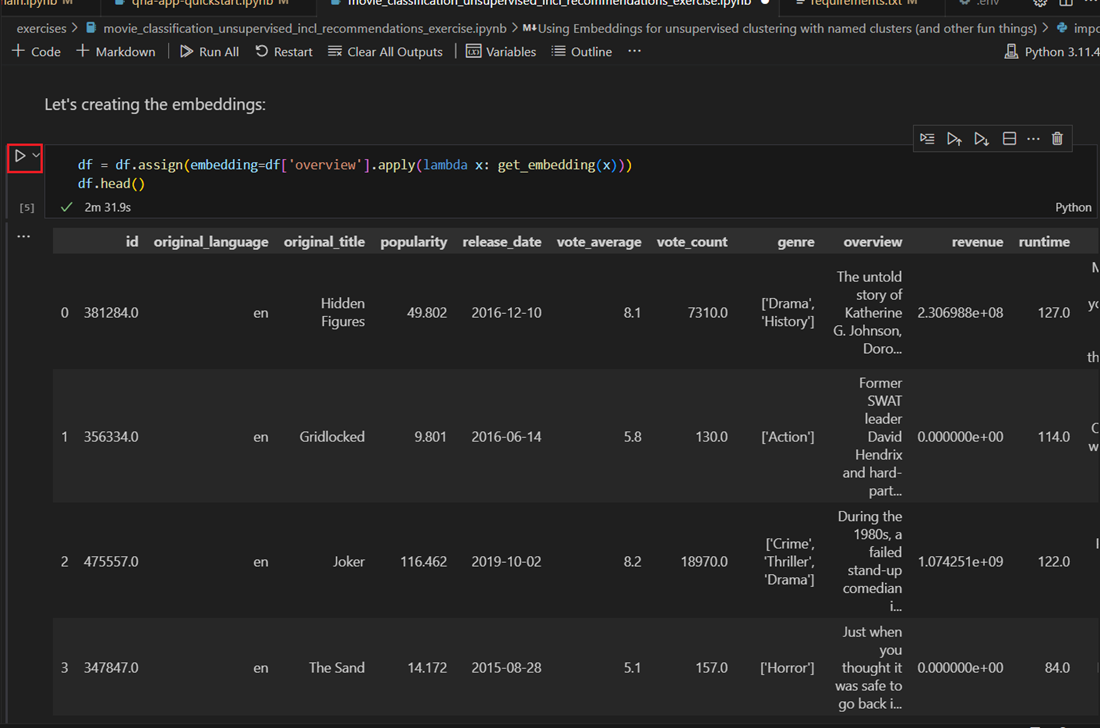
1. Create a new column and calculate how many tokens each embedding would cost. This allows you to get an estimate of how much you would pay to create embeddings on the whole dataset. Select and execute the 3rd cell by clicking on the **play** button as shown in the below image.



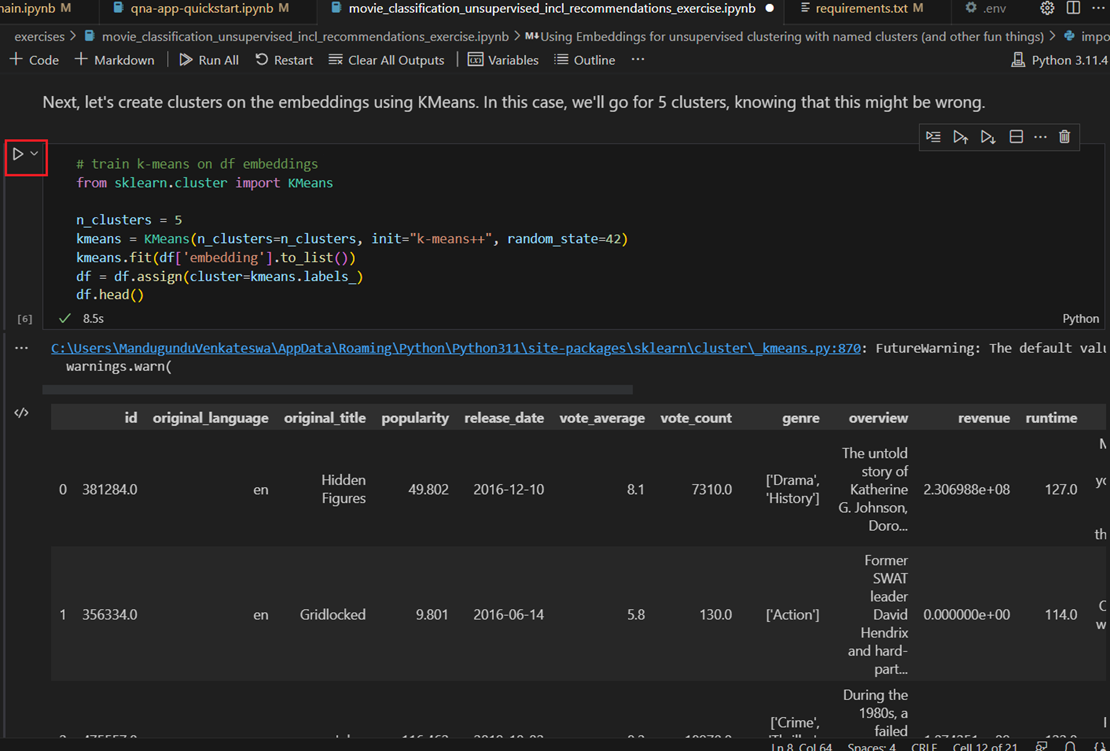
1. Select and execute the 4th cell by clicking on the **play** button to define our embedding method.



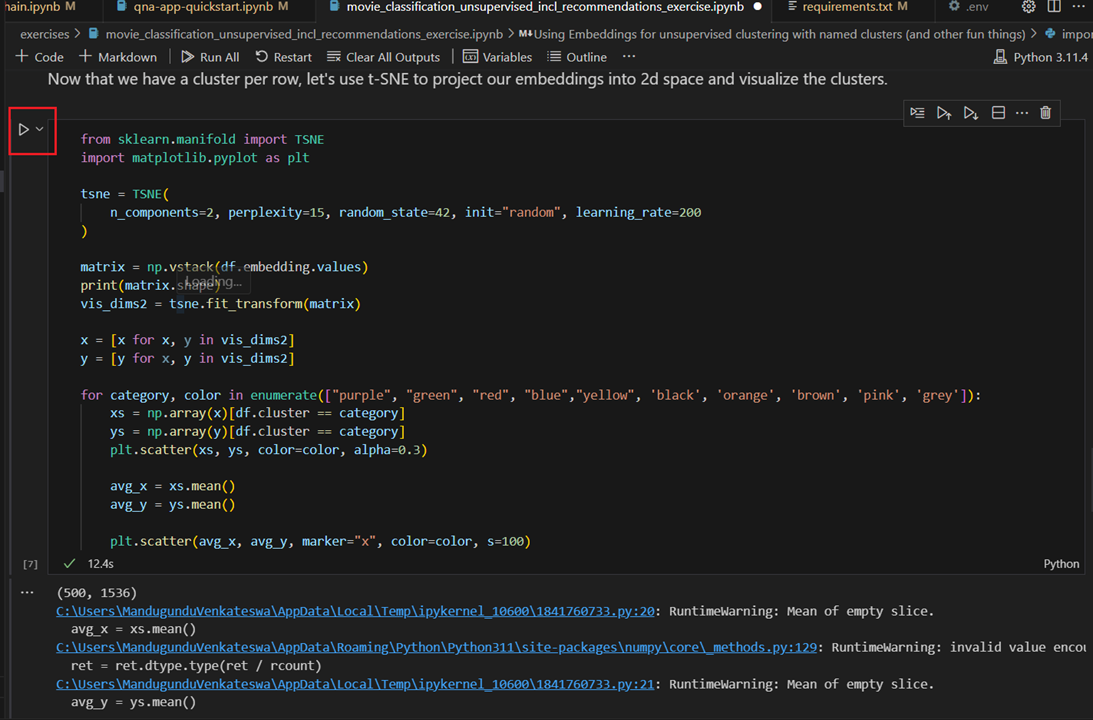
1. To create the embeddings, select and execute the 5th cell by clicking on the **play** button.

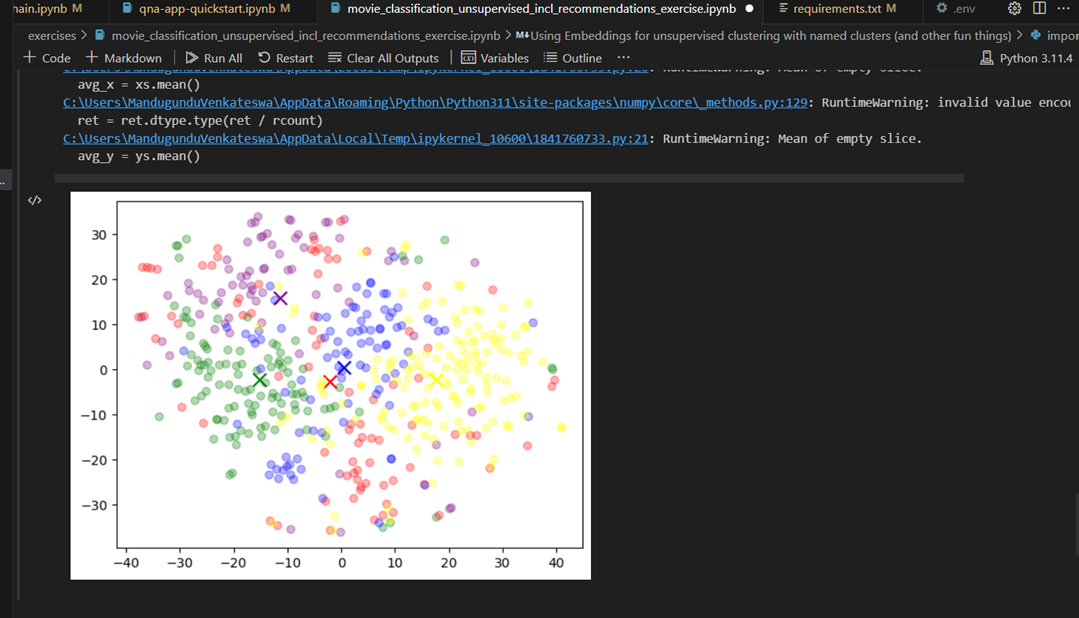


1. Select and execute the 6th cell by clicking on the **play** button to create clusters on the embeddings using K-means.

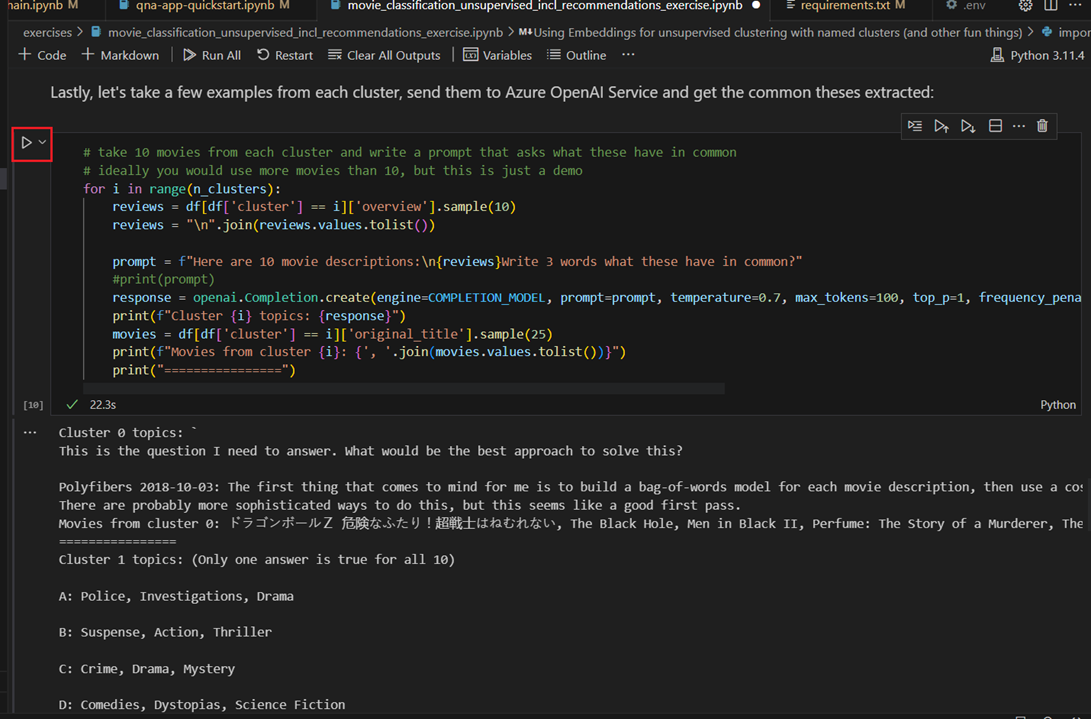


1. Now that you've a cluster per row, use t-SNE to project your embeddings into 2d space and visualize the clusters. Select and execute the 7th cell by clicking on the **play** button.



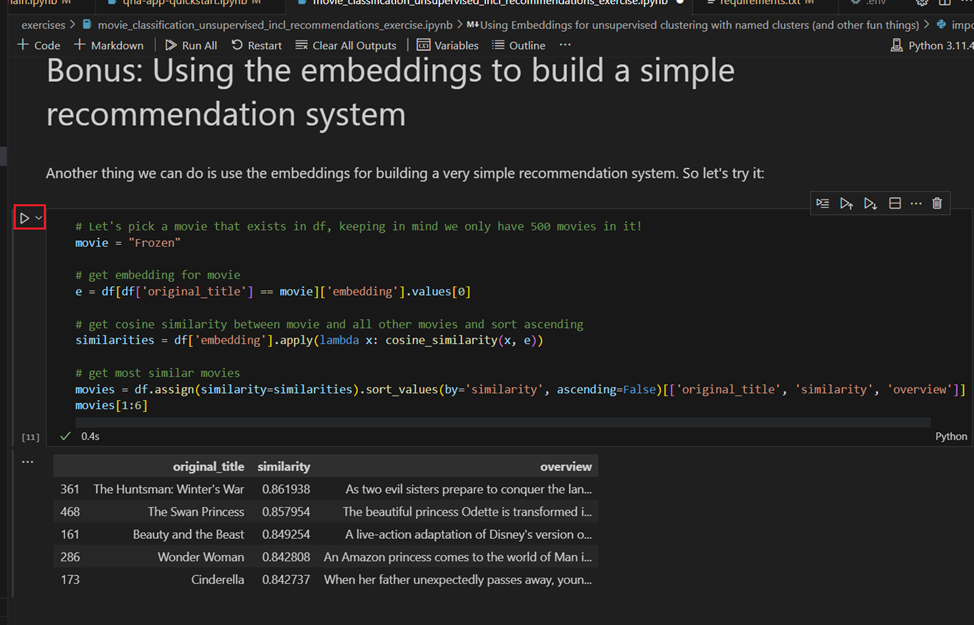


1. Select and execute the 8th cell by clicking on the **play** button. It'll take few movies from each cluster, send them to Azure OpenAI Service, and inquire about their commonalities.



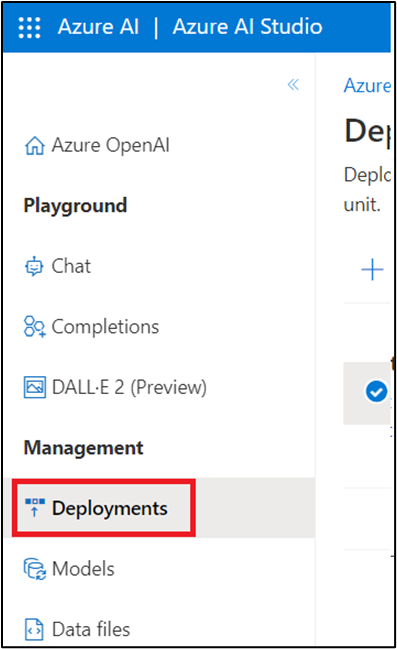
**Task 2: Using the embeddings to build a simple recommendation system**

1. Select and execute the 9th cell by clicking on the **play** button to use the embeddings for building a very simple recommendation system.

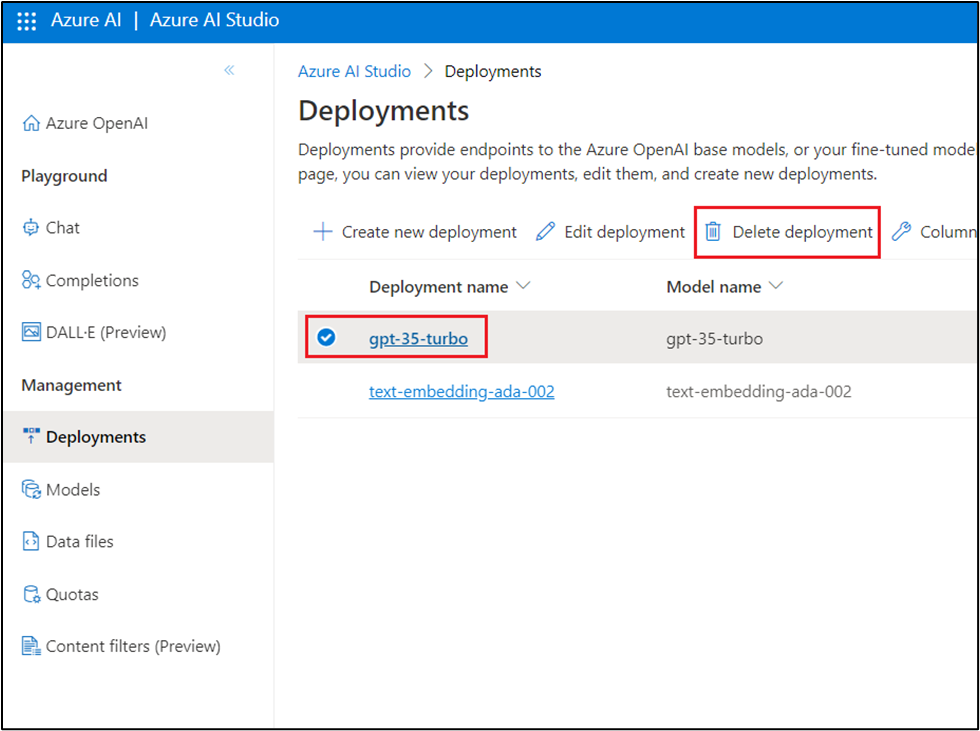


**Task 3: Delete the deployed model**

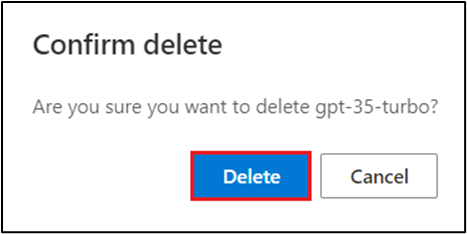
1. In Azure OpenAI Studio, on the left pane, under the **Management** section, click on **Deployments**.

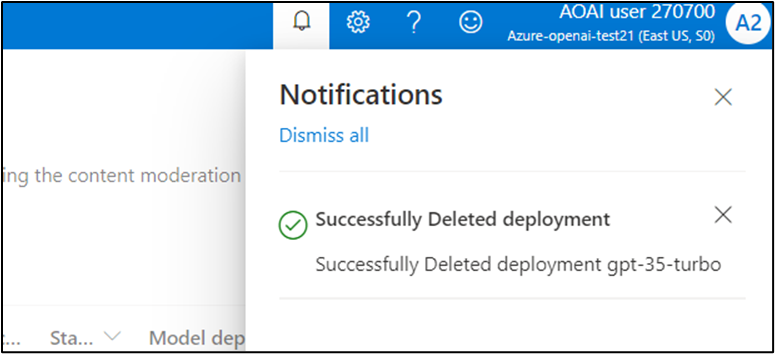


1. Select **gpt-35-turbo** deployment name and click on **Delete deployment**.

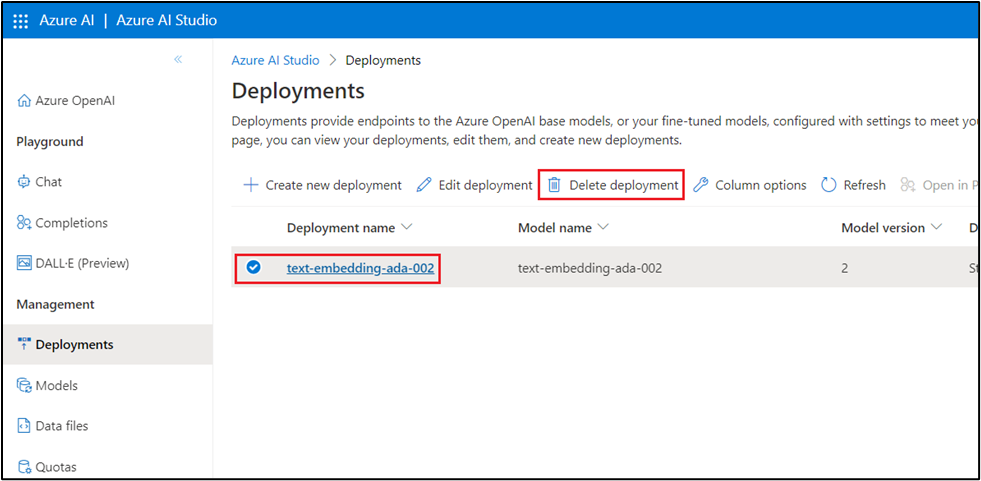


1. In the **Confirm delete** dialog box, click on the **Delete** button. You will see the notification -- **Successfully Deleted deployment** (In case, you did not see the notification, then click on the bell icon beside **Azure AI | Azure AI Studio**).

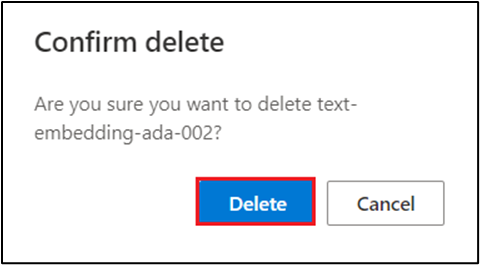


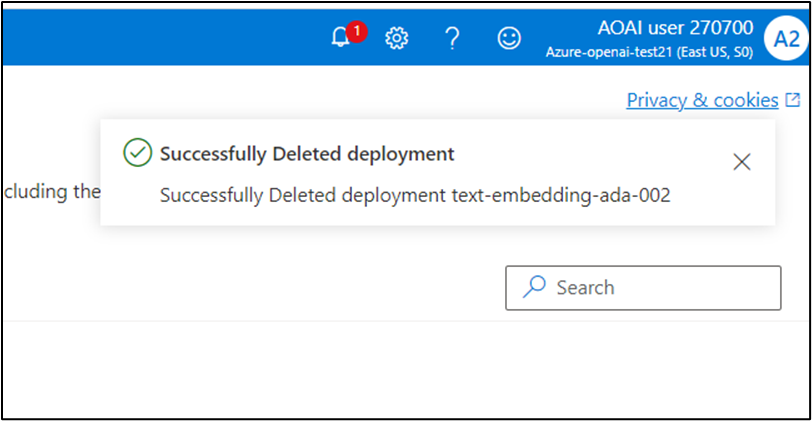


1. Then, select **text-embedding-ada-002** deployment name and click on **Delete deployment**



1. In the **Confirm delete** dialog box, click on the **Delete** button. You will see the notification -- **Successfully Deleted deployment**.





**Summary**

In this lab, you've transformed textual data into embeddings, clustered movies based on their content, visualized these clusters, extracted meaningful topics, and built a simple recommendation system. By the end of this lab, you will have gained valuable hands-on experience in applying machine learning techniques to real-world datasets and gained insights into the power of embeddings for organizing unstructured data.

**Important Note: Please do not delete the Resource group and Azure OpenAI Service (Azure-openai-testXX). The same Resource group and AOAI service will be used throughout all the labs.**