watsonx.ai Challenge Guide

Analyze customer review comments to gain insights and determine product improvements with watsonx.ai

Introduction

Foundation models provided by watsonx.ai can be used to read and summarize written, text-based input to help businesses understand customer information, such as customer feedback about the products that business offers. Knowledge about customer feedback would help the business understand what improvements or changes are required in their products.

In this challenge, participants will use customers' product review comments for a fictious company and produce an effective summary of those reviews. This is achieved through the careful use of prompt engineering with the IBM Granite model (a decoder-only model known for its exceptional performance).

Glossary of terms

notebook: An interactive document that contains executable code, descriptive text for that code, and the results of any code that is run.

prompt: Data, such as text or an image, that prepares, instructs, or conditions a foundation model's output.

prompting: The process of providing input to a foundation model to induce it to produce output.

prompt engineering: The process of designing natural language prompts for a language model to perform a specific task.

data set: A collection of data, usually in the form of rows (records) and columns (fields) and contained in a file or database table.

token: A discrete unit of meaning or analysis in a text, such as a word or sub-word (part of a word).

inferencing: The process of using a trained machine learning model to make predictions on new data.

application programming interface (API): An interface that allows an application program that is written in a high-level language to use specific data or functions of the operating system or another program.

ROUGE score: ROUGE, or Recall-Oriented Understudy for Gisting Evaluation, is a set of metrics and a software package used for evaluating automatic summarization and machine translation software in natural language processing. The metrics compare an automatically produced summary or translation against a reference or a set of references (human-produced) summary or translation.

Generative AI

Generative AI has been proven to be effective at tasks like answering questions and summarization.

Sentiment analysis aims to determine if an input has certain characteristics, such as being *positive* or *negative*.
Category classification is determining what offer to provide based the on given customer comment.

Before you begin

Make sure you have set up your IBM Cloud environment by following the instructions to set up your account ahead of the challenge.

Data sets

To get the required data sets for this challenge, download the watsonx.ai-summarization-track-files.zip and extract it.

The extracted data folder contains the following files:

- train.csv
- test.csv

Use the data present in the **train.csv** to build the prompts and also to understand the expected summary format. Then test those prompts using the data present in the **test.csv** file. The Jupyter notebook (explained below) has examples on how to complete these steps.

The .zip file also contains the notebook file "**watsonx.ai-summarize-product-reviews-notebook.ipynb**".

Note: All the data and notebooks provided in the zip file have been cleared by IBM Legal for use in the watsonx Challenge.

Understanding this challenge

This challenge focuses on assessing customer reviews for a fictitious product and doing the following task:

- Summarization

During this challenge, you will carefully design prompts with the **IBM Granite model** (IBM Granite – granite-13b-chat-v2) to generate a summary of the product reviews highlighting the features liked by the customer and the ones for which they complained.

Task: Summarize product reviews

This task consists of summarizing product reviews by listing features liked by the customers and features customers have complained about. This will help the product company decide on improvements to the product and customer experience.

The sample prompt for this task might look like the following image:

```
summarize_instruction = """

You are a retail expert. You will be given a list of product reviews. Your task is to generate a short summary, only from given product reviews from an ecommerce site to give feedback to another customer. When generating responses, prioritize correctness, i.e., ensure that your response is correct given the context and user query, and that it is grounded in the context.

[Document]
{reviews}
[End]


<|assistant|>

"""
```

Your goal is to accurately generate a summary of product reviews which will help the company understand customer feedback, not only to improve the product features, but also to understand positive feedback shared by customers.

For reference, in the training data, the summary column gives the expected summary by a human that is to be generated by the prompt and model so you can compare your results.

Once you are satisfied with the prompt, you can copy it back to the Python notebook and run the notebook to calculate the accuracy score (rouge score – in cell number 13).

Submitting your accuracy score

Using the right prompting for the summarization task, your challenge is to improve the Accuracy score (Rouge score) in the Jupyter notebook (explained below). Cell number 13 is calculating product reviews summarisation task. You will be required to submit a Rouge score as a challenge submission.

Get started with the watsonx Challenge

In this watsonx Challenge you must complete the following two lab exercises:

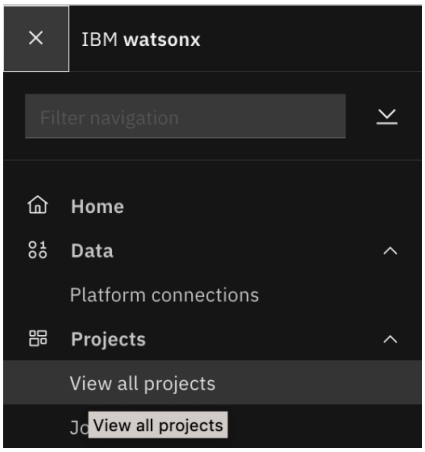
- Using the watsonx.ai Prompt Lab
- Using Jupyter Notebook

Use the watsonx.ai Prompt Lab

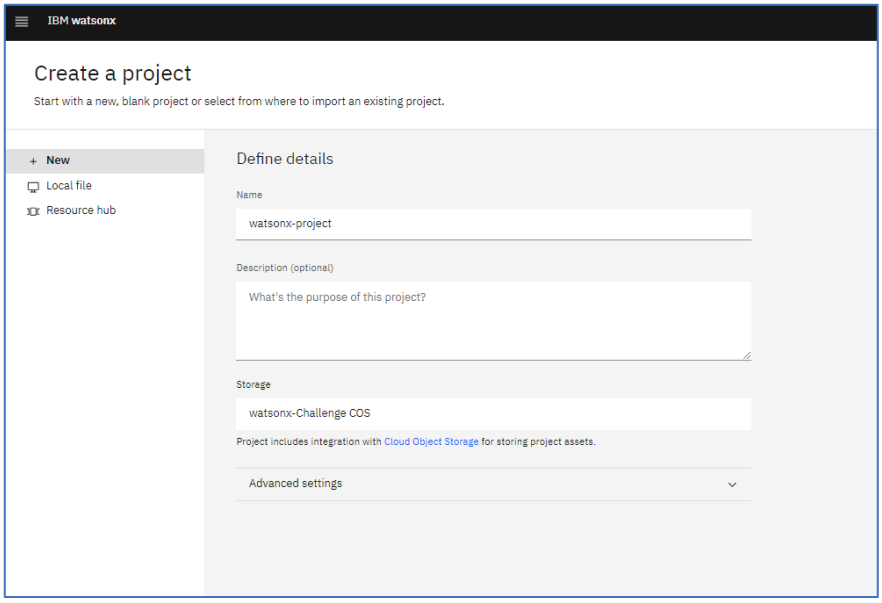
The watsonx.ai Prompt Lab is an easy-to-use interface that helps you experiment with foundation models and build prompts.

In the Prompt Lab in IBM watsonx.ai, you can experiment with prompting different foundation models, explore sample prompts, and save and share your best prompts. Before you start using Prompt Lab, you will first need to Create a Project.

1. Log in to watsonx (https://dataplatform.cloud.ibm.com/wx/home?context=wx) using your IBMid.
2. From the navigation menu at the top-left, select **View all projects.**

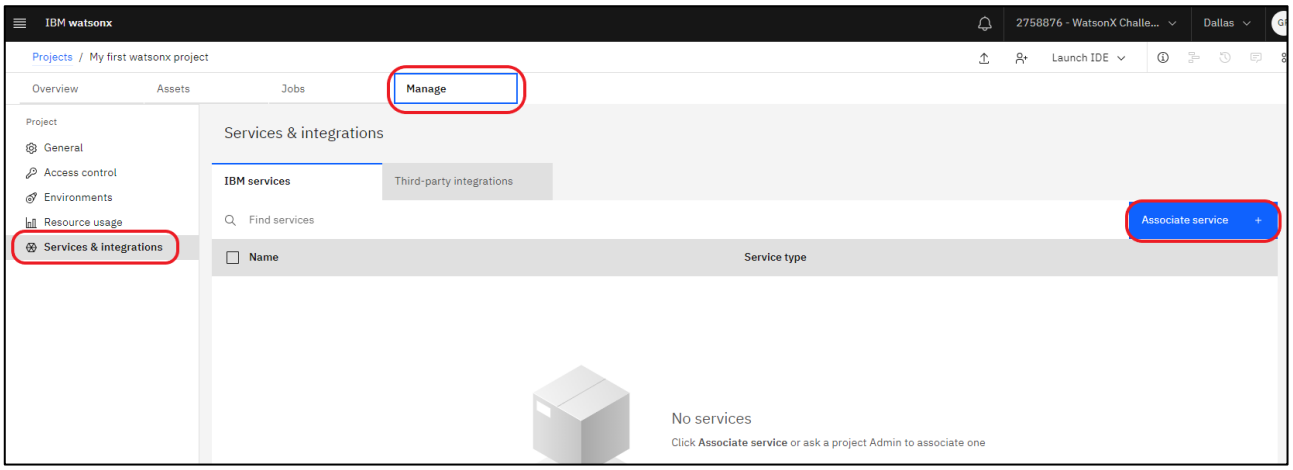


3. Select **New Project > Create an empty project**.
4. Enter a name for the project. In the **Storage** panel, “watsonx – Challenge COS” will be selected for you.

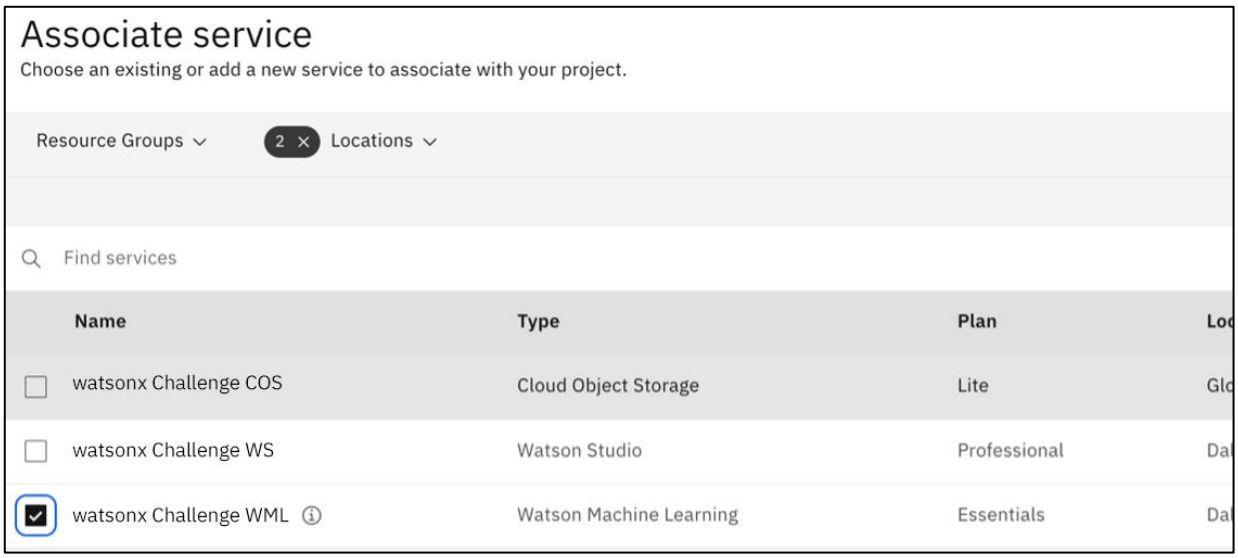


5. Click **Create**. You will be taken to the project’s overview dashboard.

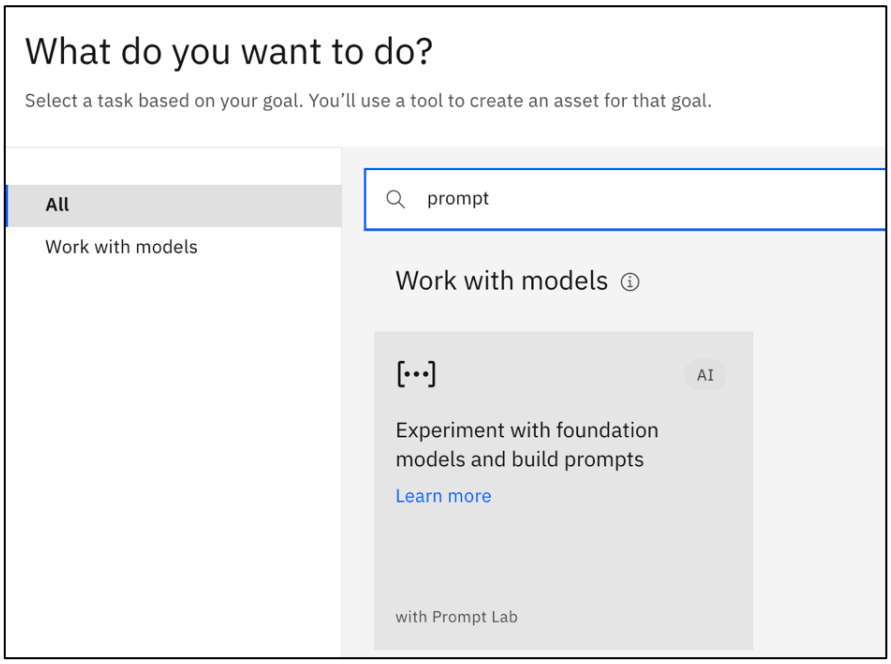
6. Select the **Manage** tab on the dashboard, select the **Services & integrations** on the left pane, and click **Associate service +**.



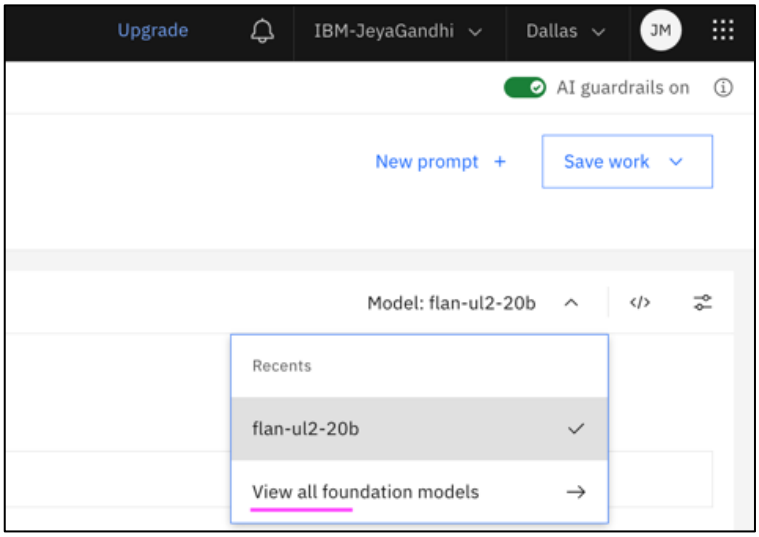
7. From the list of services, select the "**watsonx-Challenge WML**" Watson Machine Learning instance and then click **Associate**.



8. Select the **Assets** tab, click **New asset +**, and search for "prompt".
9. Select the "**Experiment with foundation models and build prompts**" tile.



10. The Prompt Editor opens. Select the **Structured** view in the prompt lab.
11. The selected **Model** is flan-ul2-20b by default. Change this by opening the Model drop-down menu and select **View all foundation models > granite-13b-chat-v2**.



12. Under the **Set up** section, enter the following in the **Instruction** field: "Analyze and summarize customer product reviews extracting the features liked and disliked by the customers."
13. Specify one or more pairs of example inputs and the corresponding desired outputs under **Examples** (You can click **Add example +** to add more examples):

- **Input**:

  1: Very nice freezer works perfect, I'm so happy this product
  2: Nice size, quiet, keeps food cool, freezer too small, yet good for making ice cubes in a tray.
  3: Needed a mini fridge for my office, kept on searching on many sites but most were above my budget until I found this fridge on a good price. It's got decent space and shelves to organize food and drinks. The little mini freezer is quite which is a nice addition too.
  4: I bought this for my wife's home office. It's a nice little fridge and I love the storage area in the door which can hold cans. The freezer section is on the small side, but you don't get a bigger freezer section on a mini fridge without going up significantly in price it seems.
  5: Does well keeping the beers cold. Did pretty good In The 120 degrees last summer. Freezer doesn't freeze but that's not what we bought it for.

- **Output**:

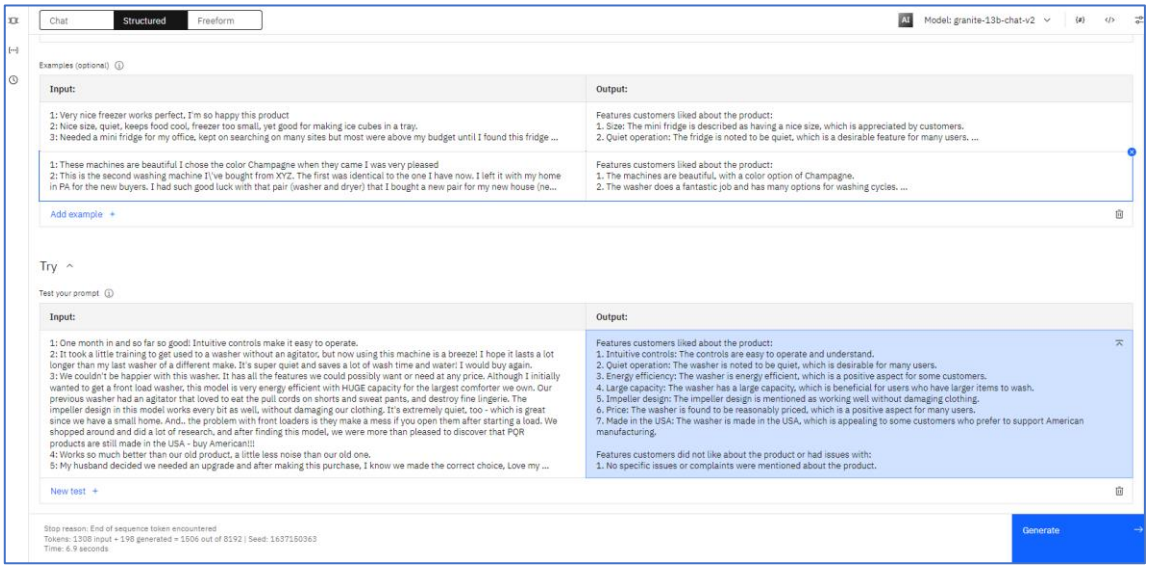
  Features customers liked about the product:
  1. Size: The mini fridge is described as having a nice size, which is appreciated by customers.
  2. Quiet operation: The fridge is noted to be quiet, which is a desirable feature for many users.
  3. Temperature control: The fridge is effective at keeping food cool, indicating good temperature control.
  4. Storage space: The storage area in the door can hold cans, which is a convenient feature for some customers.
  5. Price: The fridge is found to be reasonably priced, which is a positive aspect for many users.

  Features customers did not like about the product:
  1. Freezer size: The freezer section is on the small side, which some customers found to be a drawback.
  2. Price: Although the price is reasonable for some customers, others felt that it could be lower.
  3. Limited freezing capability: The freezer does not freeze items which may not meet the expectations of some customers.

14. Under the **Try** section, enter the following input so you can test your prompt and generate an output:
    - **Input**:

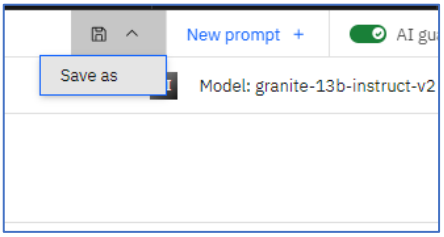
1: One month in and so far, so good! Intuitive controls make it easy to operate.
2: It took a little training to get used to a washer without an agitator, but now using this machine is a breeze! I hope it lasts a lot longer than my last washer of a different make. It's super quiet and saves a lot of wash time and water! I would buy again.
3: We couldn't be happier with this washer. It has all the features we could possibly want or need at any price. Although I initially wanted to get a front load washer, this model is very energy efficient with HUGE capacity for the largest comforter we own. Our previous washer had an agitator that loved to eat the pull cords on shorts and sweat pants, and destroy fine lingerie. The impeller design in this model works every bit as well, without damaging our clothing. It's extremely quiet, too - which is great since we have a small home. And.. the problem with front loaders is they make a mess if you open them after starting a load. We shopped around and did a lot of research, and after finding this model, we were more than pleased to discover that PQR products are still made in the USA - buy American!!!
4: Works so much better than our old product, a little less noise than our old one.
5: My husband decided we needed an upgrade and after making this purchase, I know we made the correct choice, Love my PQR!

15. Click the **Generate** button. The values you entered are sent to the model and an output is generated.

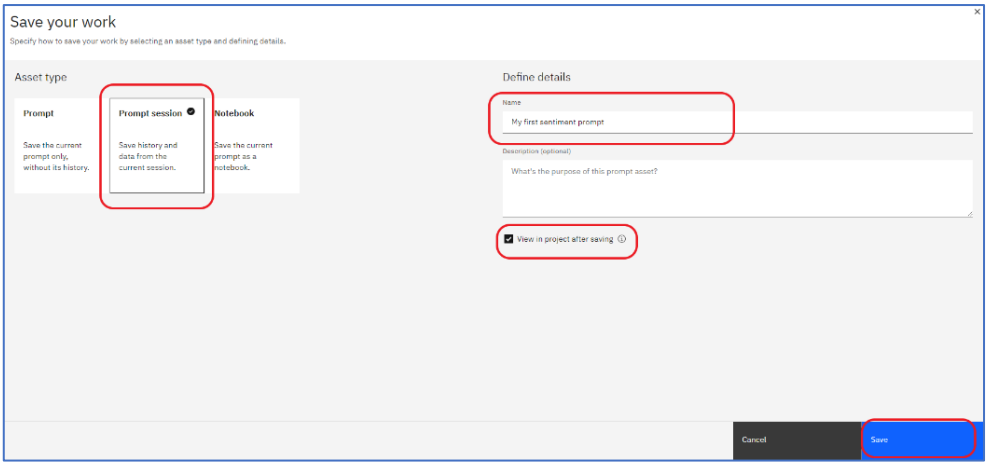


Note: Make use of the model parameters and make necessary changes to fine tune your prompt output.

16. Save your prompt lab session for later use. On the top of the Prompt lab, select **Save work** dropdown button and then select the **Save as** option.



17. A **Save your work** widget will appear. Select **Prompt session** under the **Asset type** option. Enter a **name** and check the **View in project after saving** option under the **Define details** section. Finally, click the **Save** button.



18. Once you save, you will see the saved work under the **Assets** tab.

Use Jupyter notebooks

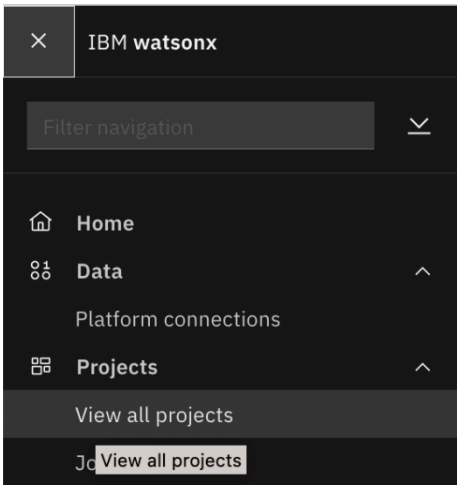
Before you can run the notebook, complete the following steps:

- Generate an IBM Cloud API key (See the setup instructions for details). You will be prompted for the IBM Cloud API key while running the notebook. Ensure that the key is easily accessible after you generate it.
- Access the Jupyter Notebook from the watsonx-ai-summarization-track-files.zip that you downloaded, or you can upload via URL (refer the instructions for uploading the notebook).

Notebook exercise steps:

1. Log in to watsonx (https://dataplatform.cloud.ibm.com/wx/home?context=wx) using your IBMid.

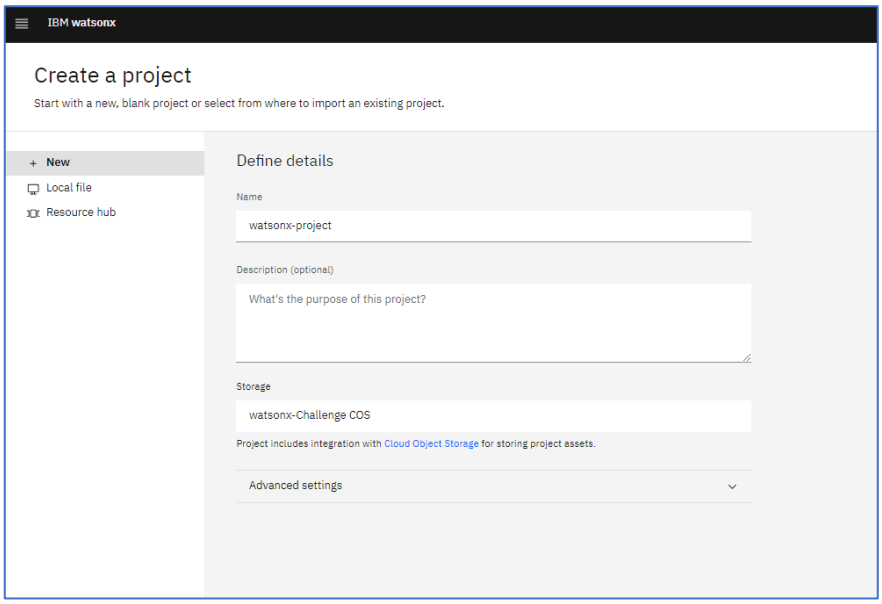
2. From the menu, select **View all projects**.



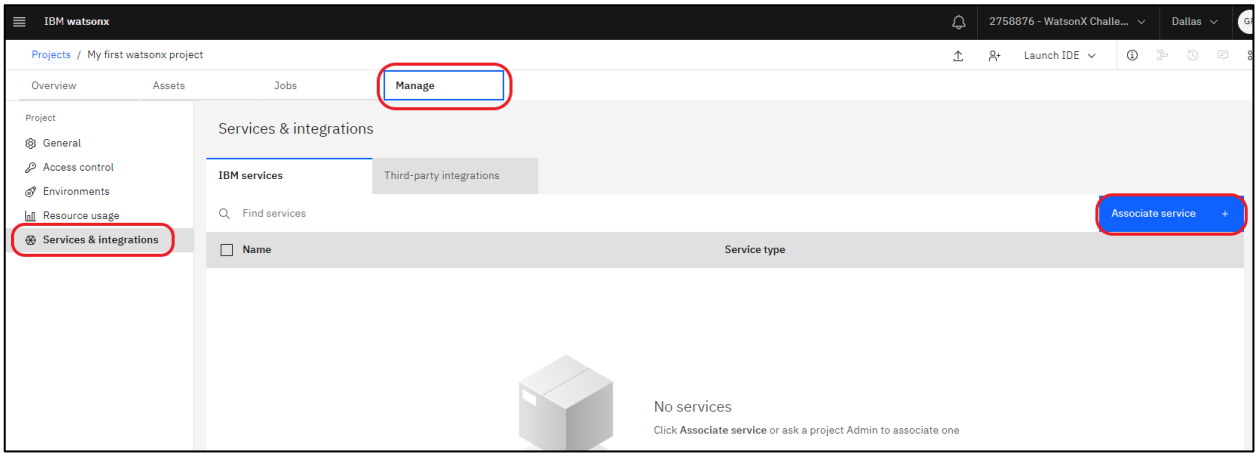
1. Select **New Project > Create an empty project**.

   NOTE: If you have already created a project for Prompt Lab exercise, you can select the same project and skip to step 8.

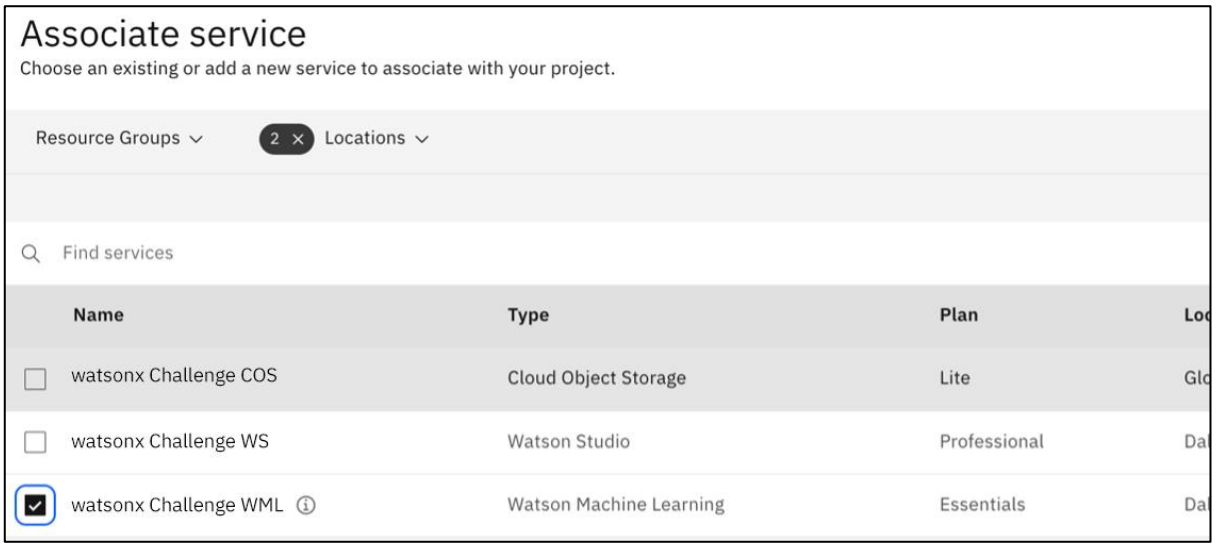
2. Give any name to the project. In the **Storage** panel, “watsonx – Challenge COS” will be selected for you.



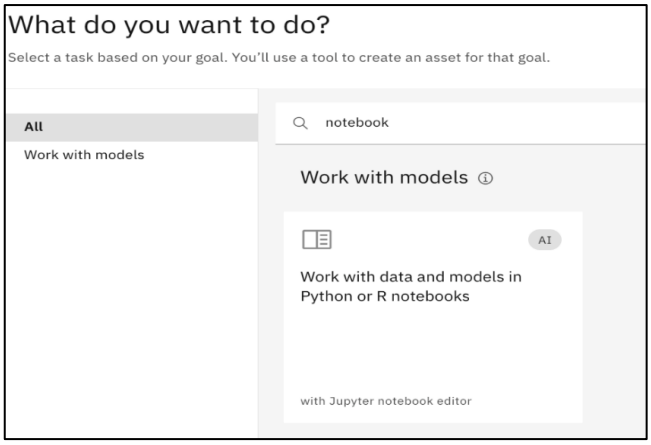
3. Click **Create**. You will be taken to the project's overview dashboard.
4. Select the **Manage** tab on the dashboard, select the **Services & integrations** on the left pane, and click **Associate service +**.



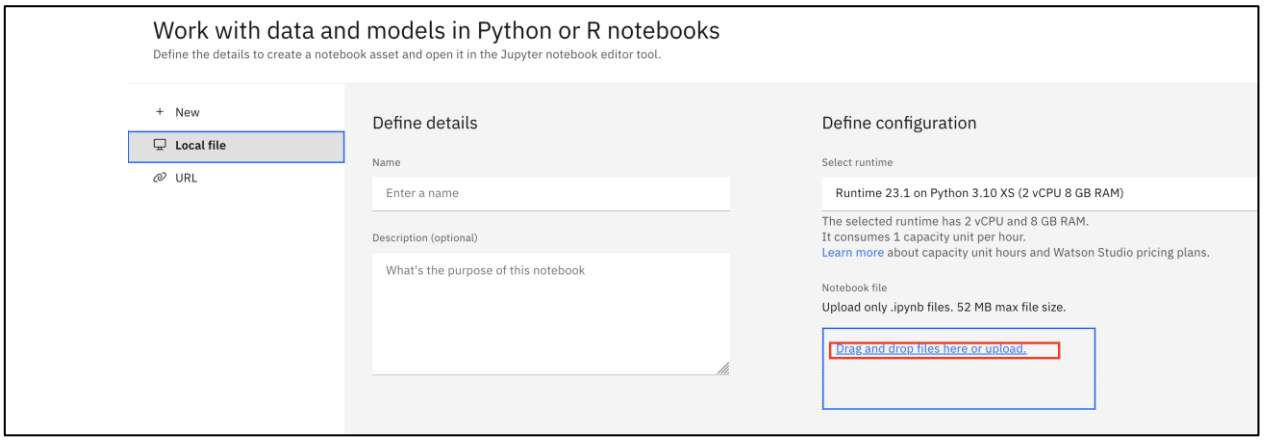
5. From the list of services, select the "**watsonx-Challenge WML**" Watson Machine Learning instance and then click **Associate**.



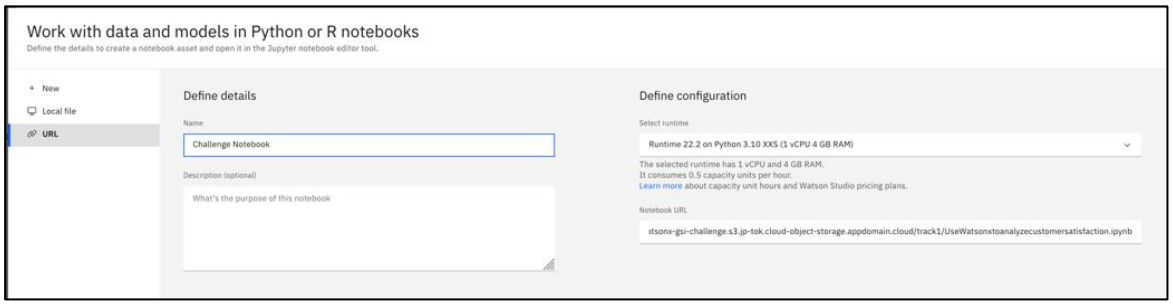
6. On the **Assets** tab, click **New asset +** and search for ‘notebook’.
7. Select **Work with data and models in Python or R notebooks** card option.



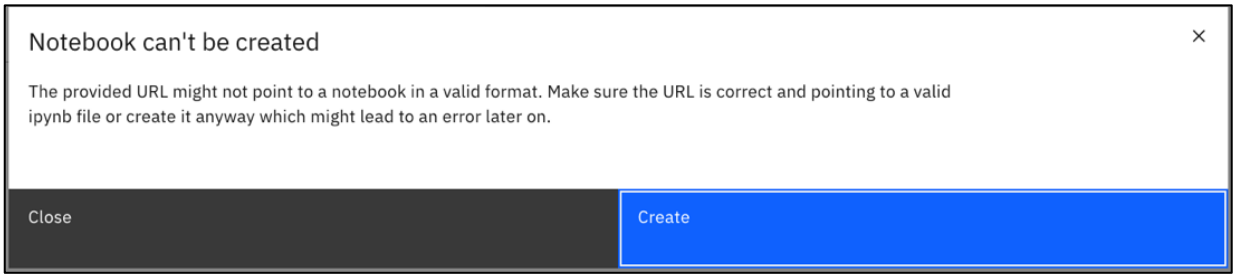
8. Select **Local file** from the left side menu.
9. Click on “**Drag and drop files here or upload**” and upload the “**watsonx.ai-summarize-product-reviews-notebook.ipynb**” notebook file provided with.zip file. Click the **Create** button.



a. If a notebook fails to create with previous step, create a notebook from a URL. Select URL from the left side menu, enter a Name for the notebook and in the Notebook URL field, add the URL: https://ibm-watsonx-challenge-2024.s3.us.cloud-object-storage.appdomain.cloud/watsonx.ai-summarize-product-reviews-notebook.ipynb

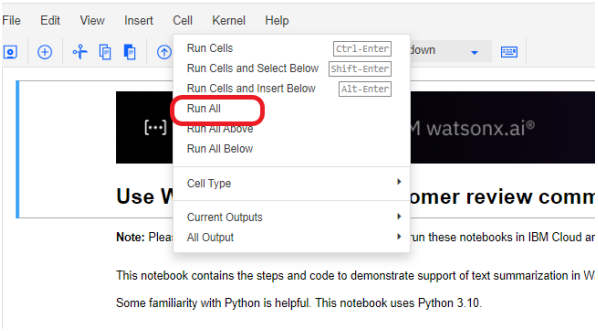


10. You might get an error as below. You can ignore it and click **Create**.

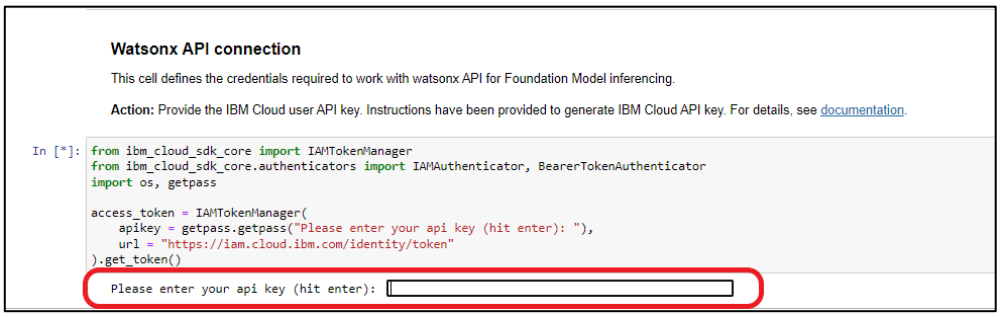


11. Now, you can edit the notebook and follow the instructions in the notebook to proceed with the challenge.
12. If you are a beginner working with Jupyter notebook, follow these instructions to run the notebook:

a. Select the **Cell** tab at the top of the notebook and select the **Run All** option. This will execute all the cells in the notebook.

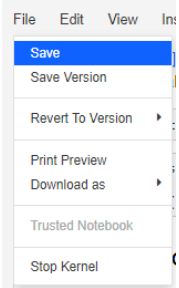


b. Next, make sure to enter the **IBM Cloud API key** that you had created as part of the [setup instructions] for the notebook to continue execution of all the cells.



c. Once the notebook completes execution of all the cells in the notebook, a **rouge1 score** will be generated.

d. You can save the notebook environment by selecting **File** on top left of the notebook and click the **Save** option. The saved notebook will be accessible under the project's **Assets** tab.



13. Either utilize the Prompt Lab or Jupyter notebook's *Evaluate the model, prompt and parameters section* to improve model accuracy

score (rouge1) and submit your highest rouge1 score at end of the challenge.

Final deliverables

Take a screen capture of ***rouge1 score*** that you have tried to improve for the product review summarization in the Jupyter Notebook, and save this image. At the end of this watsonx Challenge, you will submit this image to demonstrate your ***rouge1 score***.