Activity: Write prompts for Gemini

# Activity Overview

You have learned that many data professionals now use conversational AI tools like Gemini and ChatGPT to help them analyze their data and boost their productivity. Gemini and ChatGPT are both **large language models (LLMs)** that are trained on massive datasets of text and code. LLMs can generate human-like text in response to a wide range of prompts and questions. In this activity, you’ll discover the capabilities of conversational AI by writing your own prompts for Gemini.

To review the role of AI in data work, refer to the reading about [how data professionals use AI](https://www.coursera.org/learn/foundations-of-data-science/supplement/2FH0K/how-data-professionals-use-ai).

Be sure to complete this activity before moving on.

# LLM prompts and best practices

Data professionals can use LLMs to improve their data analysis, perform essential tasks, and collaborate with teammates. Here are some useful prompts for data science workflows:

* *Data cleaning*. LLMs can automate tasks such as data cleaning and coding. For example, you can ask an LLM to clean a dataset by removing missing values, outliers, and duplicate data.
* *Exploratory data analysis (EDA)*. LLMs can perform exploratory data analysis (EDA) on datasets. For example, you can ask an LLM to create data visualizations, identify patterns and trends, and calculate summary statistics.
* *Modeling*. LLMs can build and evaluate models. For example, you can ask an LLM to build a machine learning model to predict an outcome, and evaluate the performance of the model.
* *Interpreting results*. LLMs can interpret the results of models. For example, you can ask an LLM to explain the features that are most important for a model, or generate insights from the results of a model.
* *Collaboration*. LLMs can help you collaborate with teammates. For example, you can ask an LLM to create a shared document for a brainstorming session with a team of data professionals.

**Pro tip:** Be sure to structure your prompts in a way that makes it easier for the LLM to fulfill your requests and answer your questions.

The following suggestions are best practices for writing prompts for LLMs:

* *Be clear and concise in your instructions*. It is important to be clear and concise in your instructions so the LLM can understand how to help you. Details are great—just make sure they’re useful and relevant. Avoid giving the LLM unnecessary information.
* *Be precise*. When posing a question to an LLM, be precise about the input (if any) and the desired output.
* *Include a description of LLM’s role*. This reinforces the purpose of your prompt. For example, you can tell the LLM to assume the role of a data scientist by writing “Act as a data scientist” or “You are a data scientist.”
* *Provide context*. Providing context allows the LLM to understand the nuances of the relevant issue and generate more informed responses.
* *Try multiple prompts*. Trying different prompts can provide different perspectives on a problem and enable the LLM to generate a variety of useful responses.

To help get you started, consider the following specific examples of prompts that data professionals can give an LLM:

* “Act as a data scientist and write a detailed plan for a credit card fraud detection project.”
* “I have a dataset of customer purchases at an online retail store. Act as a data scientist and write Python code for data visualization and exploration.”
* “I have a dataset of customer characteristics and churn for an online video streaming service. Act as a data scientist and create a shared document for a team meeting.”
* “Act as a data generator and use Python code to generate a CSV file that contains mock employee data for a restaurant chain named Fast. The dataset has 100 rows and 5 columns. The columns are name, address, employee\_id, department\_id, email.”
* “Act as a communications expert and share best practices for explaining a data science report to a business executive with no technical background.”

**Note:** LLMsare powerful, but they are still under development – including Gemini, which is still experimental research. As a data professional, it’s important to use your own judgment when interpreting the results. LLMs can generate insights that you may not have thought of on your own; however, it’s ultimately your responsibility to verify the results and make sure they make sense.

# Step-By-Step Instructions

Review the following fictional workplace scenario. Then follow the instructions for giving prompts to Gemini at different stages of the data project.

Imagine you are a new data professional, recently hired by a healthcare company. The company sells sustainable medical devices to hospitals and clinics in urban communities. Leadership has asked the data team to develop a machine learning model to accurately predict future sales. A powerful model will help company leaders make informed decisions about inventory management, resource allocation, overall sales strategy, and more. As the newest member of the data team, you’re excited to start your first project.

1. *Project proposal*. To get started, your manager asks you to organize a kickoff meeting with the team to outline the project workflow and timeline. You want to send the invite as soon as possible, and could use some help creating a document for the meeting. **Prompt: Ask Gemini for instructions create a shared document to facilitate a brainstorming session among a team of data professionals.**
2. *Data cleaning*. Following the team meeting, you help draft a project proposal to outline key deliverables and milestones for the project. Then, the team collects the relevant data. The next step is to clean the dataset. You volunteer to perform this task for the team. The team is using the Python programming language for this project, and you’d like some coding suggestions for data cleaning. **Prompt: Ask Gemini to write Python code to clean data by removing missing values, outliers, and duplicate data.**
3. *Data visualization*. Now that the team has a clean dataset to work with, the next step is to explore and visualize the data. Your manager asks if you can help create some data visualizations to better understand the relationships between key variables. To get started, you brainstorm with B. **Prompt: Ask Gemini to suggest useful data visualizations for sales data.**
4. *Build and test machine learning models*. As a new data professional, you are not directly involved in writing code to build and test different machine learning models. However, you want to learn more about the uses of machine learning for data work as this will be an important part of your future career, and will help you better understand the current project. **Prompt: Ask Gemini about the main uses and benefits of machine learning for data work.**
5. *Executive summary*. The data team successfully builds a model that accurately predicts future sales. Now, the team is ready to share their results and insights with project stakeholders. Your manager asks you to help draft an executive summary for a meeting with company leadership. Before you begin, you want to review best practices so you can create a polished deliverable. **Prompt: Ask Gemini about best practices for creating an executive summary for business executives without a technical background.**

**Note:** Overall, Gemini is a powerful tool for data professionals. However, it’s important to remember that Gemini is not perfect. Be aware of Gemini’s (and other LLMs') limitations. These limitations include the following:

* *Gemini is not infallible*. Gemini can sometimes make mistakes, such as providing inaccurate information or generating incorrect code.
* *Gemini is not an expert in any particular field*. Gemini can learn about new topics, but it does not have the same level of understanding as an experienced human data professional.
* *Gemini cannot explain its reasoning*. Gemini can generate useful output, but it cannot explain why it’s doing what it’s doing. This can make it difficult to understand how Gemini works and to trust its results.
* *Gemini can be biased*. As an LLM, Gemini is trained on a massive dataset of text and code, and Gemini is likely to reflect the biases that are present in that dataset.

As a result of these limitations, Gemini’s responses may be inaccurate, biased, or insufficient for your purposes. As a data professional, it’s your responsibility to verify the accuracy of Gemini’s output. It’s also your job to modify or supplement Gemini’s output to fit the needs of the specific project you’re working on.