Leverage Generative AI to unlock Data Insights

in just few minutes

Author : Rihana Msadek

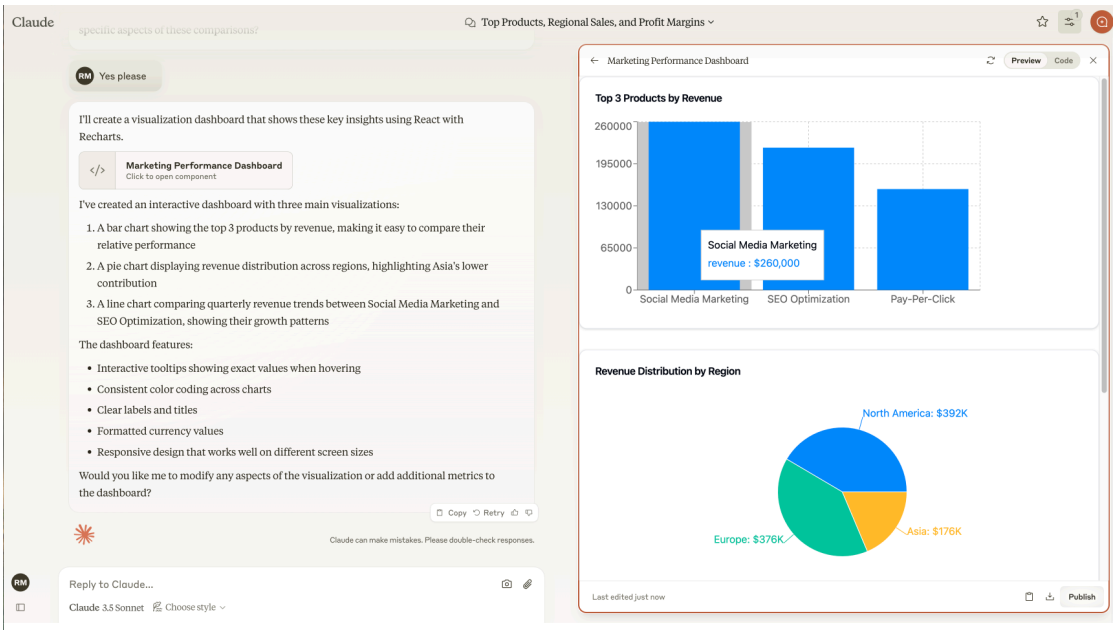
Time: 20 minutes

- **Learning Objectives:** By the end of this activity, you will be able to :
    - Understand the capabilities of Generative AI models for spreadsheet/CSV data analysis.
    - Practice formulating effective prompts to extract specific insights.
    - Analyze and interpret AI-generated results.
    - Explore different output formats for presenting data.
- **Tools:**
    - Choose any of your LLM interface :
        - Google :
            - aistudio.google.com (developer platform)
            - gemini.google.com (consumer solution)
        - DeepSeek :
            - https://chat.deepseek.com/ (consumer solution)
            - https://chat.deepseek.com/503/ (developer platform) *facing some technical issues as of Jan 31st, 2025.
        - Mistral :
            - https://console.mistral.ai/ (developer platform)
            - https://chat.mistral.ai/chat (consumer solution)
        - GPT :
            - https://platform.openai.com/docs/overview (developer platform)
            - https://chatgpt.com/ (consumer solution)
        - Anthropic / Claude :
            - https://console.anthropic.com/ (developer platform)
            - https://claude.ai/ (consumer solution)
    - The provided CSV file (Sales Data): https://github.com/rihanamsaddek/Transformative-potential-GenAI-Lab1/blob/main/Marketing%20Example%20-%20Lab%201.csv

Instructions:

1. **Introduction:**
    - Large language models like Gemini can be powerful tools for quickly analyzing data and uncovering hidden trends. This activity will guide you through the process of using LLMs to analyze a sales dataset and extract meaningful insights.
2. **Explore the Data:**
    - Download the provided CSV file and open it in a spreadsheet program or text editor.
    - Take a few minutes to familiarize yourself with the data. **What are the columns? What type of information does each column contain? What initial questions come to mind when you look at this data?**
3. First Prompts:
    - Now, let's start interacting with the LLM. Enter the following prompts one at a time and observe the model's output. Pay attention to how the LLM interprets your requests and the format of the responses.
        - Prompt: **Provide a summary of sales performance by region**.

4. Iterative Prompting (Refining Your Questions):
    - Often, you'll need to refine your initial prompts to get more specific or different types of answers. Try these:
        - **Show the top 3 products by revenue**.
        - **Which region had the lowest sales, and what was the total revenue generated there?**
        - **Compare the sales performance of the top 2 products across all regions**.
    - *Here is an example of an output that you can generate using Claude.ai*



5. **Exploring output Formats:**
    - LLMs can present data in various formats. Experiment with these prompts:
        - Show the regional sales summary in a table.
        - Present the sales data for the top product as a bar chart. (Some LLMs can generate basic charts or provide the data in a format suitable for charting)
        - Try switching platforms
            1. **Try switching models.**
            2. **Try refining the prompts.**
            3. **Try searching the web for certain prompts.**

6. **Creative Summary and Insights:**
    - Finally, let's challenge the LLM to summarize findings creatively and generate insights:
        - Write a short summary for a sales manager describing the key trends in the sales data.

Bonus Activity:
Try analyzing one of your own spreadsheets with similar prompts.
Think about how you can use LLMs to automate data analysis tasks, generate reports, or gain quick insights from your data.
Remember to always critically evaluate the LLM's output and consider the limitations of AI-generated analysis.

How to decide which model is right for you?

Here are some key considerations to help you hone in on specific models for testing and evaluation.

Governance

Industry-specific constraints will impact the type of model you need. Healthcare, finance, and government often have stringent requirements for **data privacy, security, and explainability**. This might necessitate using models that have the right type of certification, open models that allow for higher levels of transparency and customization, or even models that can be run on isolated networks and infrastructure.

Enterprise privacy at OpenAI https://openai.com/enterprise-privacy/

Gen AI Privacy at Google
https://services.google.com/fh/files/misc/genai_privacy_google_cloud_202308.pdf

Mistral Privacy policy https://mistral.ai/terms/

Anthropic privacy policy https://www.anthropic.com/legal/privacy

Use Case

What **tasks** must the model perform for your use case? How **complex** are these tasks? **Does the desired output need to be in a particular format or style**?

Performance

What factors are most important - **latency**, **cost**, or **customizability**? Different models have different strengths, and you can use your enterprise and use case priorities as a filtering mechanism.

Model Capabilities

To effectively achieve your use case and performance goals, consider the following capabilities to select the right model for your specific needs:

- **Context window:** The amount of information the model can process at once.
- **Number of parameters:** A measure of the model's complexity and learning capacity.
- **Training dataset:** The data the model has learned from, which influences its knowledge and capabilities.
- **Multimodality:** The ability to process and generate different types of data (e.g., text, images, audio).