**Project Overview**

This project offers a comprehensive glossary of terms and definitions, organized by modules, to aid in the exploration and analysis of key concepts. The glossary is available as a CSV file, which can be utilized in both spreadsheet software and Python notebooks, providing flexibility and advanced functionalities.

**Key Features**

* **Spreadsheet Usage:** The CSV file can be seamlessly integrated with spreadsheet software like Google Sheets or Microsoft Excel. This enables users to effortlessly view, sort, and filter the glossary by terms, modules, and definitions, facilitating quick access to specific information.
* **Python Notebook Integration:** For users seeking more advanced functionalities, the project provides a Python notebook that offers:  
  + **Sorting:** Sort the glossary by terms or modules, enabling customized organization of the data.
  + **Similarity Search:** Leverage pre-calculated vector embeddings to identify semantically similar terms. This feature allows users to explore related concepts and discover connections within the glossary.

Visit Modules/Glossary.ipynb for examples

**Additional Resources**

For data science enthusiasts seeking further learning and development opportunities, the project also includes a curated list of resources:

**General Data Science Topics:**

* **freeCodeCamp:**
  + Python:<https://www.freecodecamp.org/news/search/?query=python%20basics>
  + R:<https://www.freecodecamp.org/news/r-programming-course/>
  + Calculus:<https://www.freecodecamp.org/news/learn-college-calculus-in-free-course/>
  + Linear Algebra:<https://www.freecodecamp.org/news/search?query=linear%20algebra>
  + Statistics:<https://www.freecodecamp.org/news/search?query=statistics>
  + Computer Science:<https://www.freecodecamp.org/news/search?query=computer%20science%20basics>
  + SQL:<https://www.freecodecamp.org/news/search?query=sql>
  + Database Management:<https://www.freecodecamp.org/news/search?query=database%20management>
  + Version Control:<https://www.freecodecamp.org/news/search?query=version%20control>
  + Cloud Computing:<https://www.freecodecamp.org/news/search?query=cloud>
* **YouTube:**
  + Calculus: [https://www.youtube.com/watch?v=HfACrKJ\\_Y2w](https://www.youtube.com/watch?v=HfACrKJ_Y2w)
  + Linear Algebra:<https://www.youtube.com/watch?v=JnTa9XtvmfI>
  + Statistics:<https://www.youtube.com/watch?v=xxpc-HPKN28>

**Machine Learning:**

* **YouTube:** [https://www.youtube.com/watch?v=vStJoetOxJg&amp;list=PLkDaE6sCZn6FNC6YRfRQc\\_FbeQrF8BwGI](https://www.youtube.com/watch?v=vStJoetOxJg&list=PLkDaE6sCZn6FNC6YRfRQc_FbeQrF8BwGI)
* **Coursera:**<https://www.coursera.org/specializations/machine-learning-introduction>

**Deep Learning:**

* **YouTube:** [https://www.youtube.com/watch?v=CS4cs9xVecg&amp;list=PLkDaE6sCZn6Ec-XTbcX1uRg2\\_u4xOEky0](https://www.youtube.com/watch?v=CS4cs9xVecg&list=PLkDaE6sCZn6Ec-XTbcX1uRg2_u4xOEky0)
* **Coursera:**<https://www.coursera.org/specializations/machine-learning-introduction>

**Additional Notes**

* **Installation:** Ensure you have the necessary Python libraries installed, including pandas, scikit-learn, and transformers.
* **Model:** The Python notebook might utilize a specific pre-trained sentence transformer model. Feel free to experiment with different models from the Hugging Face Model Hub to observe their impact on the similarity search results.
* **Customization:** Tailor the code and parameters, such as the number of similar terms to return, to align with your specific requirements and preferences.

This project offers a versatile and robust approach to exploring and analyzing glossary data. By combining the user-friendliness of a spreadsheet with the advanced capabilities of Python and sentence embeddings, users can gain a deeper understanding of terms, their definitions, and their relationships within the context of data science.