Exercise: Implementing Self-Attention in TensorFlow

Task:

Implement a basic self-attention mechanism in TensorFlow and apply it to a small sequence of words.

Instructions:

1. **Setup:**
   - Import the necessary libraries (e.g., TensorFlow).
   - Define a small vocabulary and create a sequence of words as input.
2. **Self-Attention Implementation:**
   - Implement a simple self-attention class using TensorFlow, including the following methods:
     - **__init__**: Initialize parameters (e.g., embedding dimensions, number of attention heads).
     - **call**: Perform the self-attention calculation.
3. **Apply Self-Attention:**
   - Create an instance of your self-attention class.
   - Apply self-attention to the input sequence.
4. **Visualization:**
   - Visualize the attention scores for each position in the input sequence. You can use a heatmap for clarity.
5. **Discussion:**
   - Discuss the importance of attention weights in understanding which parts of the input sequence are emphasized.
   - you are encouraged to interpret the attention scores and understand how they contribute to capturing contextual information.
6. **Experimentation:**
   - You are encouraged experiment with different hyperparameters (e.g., number of attention heads, embedding dimensions) and observe the impact on the attention patterns.
7. **Reflection:**
   - students are invited to reflect on the challenges and insights gained from implementing and experimenting with self-attention using TensorFlow.

Note:

An interesting video is : https://www.youtube.com/watch?v=mWA-PmxMBDk