### Why Visual Genome Dataset

#### 1. Multi-Modal Dataset

a. multi modal means information from multiple, distinct data types (or modalities) to describe a single concept

b. Visual Genome Dataset comprised of images that have more than one type of data to be understood

c. key characteristics of Visual Genome Dataset being Multi-Modal is Visual Modality (Images), Numerical Modality (Bounding Boxes in coordinates), and Textual Modality (Object Names & Descriptions).

#### 2. Ease of processing

a. The Hugging Face datasets library of python provided access to the data, very few lines of code. Adding simplicity to the process.

b. The raw dataset has a nested structure, where each row represents an image containing a list of objects. Thus, there is a need to iterates through each image and creates a separate row for every single object, transforming the data into a clean, tabular format.

#### 3. High quality information

a. The values provided by the dataset are rich. For instance, from just the limited samples we processed, we concluded that the most frequent objects were identified as desk, wall, and chair. Which suggests the dataset subset are mostly on indoor.

b. All object bounding boxes were confirmed to have valid, positive dimensions, ensuring the integrity of the localization data.