

Figure 14. Randomly selected examples from **PixMo-Points**. Even when text has been cut off, all points still appear in the image. Our templated prompts can be ungrammatical for some of these options, but we find they are still sufficient to let the model respond correctly to natural language instructions.

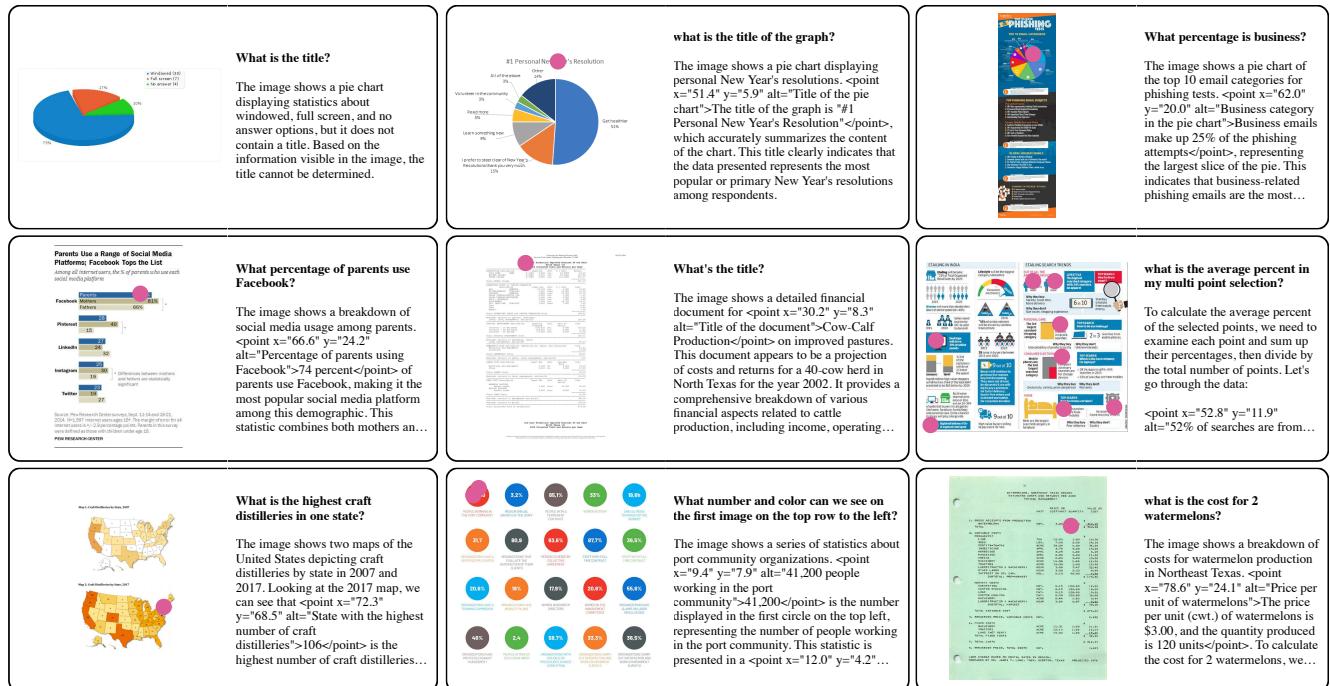


Figure 15. Randomly selected examples from the experimental **PixMo-Points** data that includes points with explanations.