Task 2.2

**Define the concept of bias in relation to your own custom dataset or the full Imsitu dataset, and a computer vision task.**

The chosen dataset consists of images of alcoholic beverages, non-alcoholic beverages, various types of fruits commonly found in cocktails, and coffee packages. Bias in the context of my dataset refers to systematic errors or inaccuracies present in the data that can skew the performance or outcomes of machine learning algorithms or computer vision models. These biases can arise from various sources and can have significant implications for the fairness and accuracy of the model’s predictions.

**List and describe the type of bias that you identified in the dataset.**

The types of biases identified are pre-existing: sampling bias, label bias, algorithmic bias, and social bias; technical: algorithmic bias; and emergent: social bias.

Sampling bias can lead to inaccuracies in classification and prediction. For instance, an overrepresentation of popular alcoholic beverages may result in models that struggle to represent lesser-known non-alcoholic alternatives.

Label bias further deepens these issues by introducing subjectivity into the labelling process. In the case of cocktail fruits, mislabelled or incomplete labels can hinder the model's ability to recognize fruits accurately in various contexts.

Additionally, social biases further compound the problem by influencing how certain objects are represented and perceived in the dataset. For example, if the dataset predominantly includes alcoholic beverages popular in Western countries, the model may struggle to accurately identify or classify beverages that are more common in other parts of the world. The same could apply to coffee brands specific to only one country.

**Discuss the possible ramifications (e.g., harm) in terms of fairness of the identified bias instance:**

(**Why, and when, is this particular instance of bias undesirable? In other words, who might be disproportionally affected by this particular instance of bias, and when does this negative effect come into play?)**

*Sampling bias*: This particular bias will mostly affect small or moderately known businesses, which in turn will also affect the bar owner’s profits since more well-known brands are more expensive while not necessarily better in quality.

*Label bias*: This bias will affect the bar’s inventory supply of fruits for cocktail preparation. These inaccuracies may lead to bar owners struggling to procure these ingredients, potentially limiting the variety and quality of cocktails offered to customers.

*Social bias*: This bias can exacerbate existing economic disparities and perpetuate inequalities in the global beverage market, favouring wealthier nations while further marginalizing economically disadvantaged countries