Pros and Cons of Under-Sampling

Pros:

1. Definitely avoid the model to be overfitting on the original dataset, which tends to skew towards majority
2. Save time and storage by having a relatively small dataset. Less data means businesses need less storage and time to gain valuable insights.

Cons:

1. Lead to information loss, including potentially important and useful information
2. Samples of the majority class chosen could be biased. Inaccurate analysis originates from inaccurate representation of the majority population

Pros and Cons of Over-Sampling

Pros:

1. No information loss compared to Under-Sampling

Cons:

1. Increase the likelihood of overfitting.
2. Greatly increase the size of the training set, leading to the increase of training time and required memory.

There is no clear conclusion as to which modeling techniques may outperform the other one. So if we take the business context into consideration, we may think about the training time and training memory as top priorities. Under-sampling requires less time and allocated memory with a risk of information loss. Over-sampling needs more time and memory with a probability of overfitting. It is a trade-off.