USER-content identification objectives

### executive summary

In the website hosting and provider business, one of the most difficult challenges in building a successful business according to [Convertica](https://www.converticacommerce.com/), a leading expert in multichannel marketing and business development, is building valuable content. The key to having valuable content is having end-user participation in the content generation process; which not only reduces the amount of overhead for your company by reducing the workload of content generation, but also increases the overall volume as well.

Collecting large volumes of user-generated content has many positives for a start-up website, however, with such large volumes of content, curation becomes a dauting task that is too unwieldly for manual intervention on every piece of new information. To this end, we would ideally have a systematic process in place to automatically identify the user posted content and tag it appropriately.

Having such a system of automatic content recognition for image based would have enormous benefits, such as being able to automatically tag and categorize new user uploaded content so that the site maintains a canonical structure, relevant content can be automatically placed under relevant sub-sections on the site, and the information flow is not bottlenecked by a manual process that keeps the site moving organically around the clock. An additional benefit of such a system would include flagging inappropriate materials automatically, so that we don’t accidently alienate a large portion of our user base by a single bad actor uploading inappropriate content.

### Research design

For this problem we used a sample user-generated dataset that was used for a machine learning competition on a well-known data science hub, [Kaggle.com](https://www.kaggle.com/). This dataset contains an image database of 25,000 high-resolution images. These images are pre-labeledf

### Technical Overview

### Conclusion