**Deliverable 1)** Describe, using an abstract, your approach and methodology. Include a visual representation of your analytic process flow.

**Deliverable 2)** Return the uncategorized content dataset with the categories identified.

**Deliverable 3)** Document your code and reference the analytic process flow-diagram from deliverable one (1).

**Abstract**

Upon first thought, the approach was to see if there was any sort of feature reduction that could be done.

An idea we had was to query the location data and find the exact address of stores, which could potentially be added to the data as a feature. However, upon running the logistic regressor across all features in the dataset, there was no significant correlation found between location, or for that matter, any of the features, and transaction class.

Because we had a labeled train dataset, we knew we were going to use some kind of supervised learning technique. We ended up rolling with a convolutional neural network, which [Soni et. al](https://arxiv.org/abs/2203.05173) in a 2022 paper find to do remarkably well with text data in various natural language processing contexts.

Soni et al.: https://arxiv.org/abs/2203.05173