We chose to use clustering as a way to gain understanding of company profiles. Before running the clustering algorithm, we decided to separate our data by macroregions and US regions since the US made up the majority of the data we had. We used k-modes clustering because our data was all categorical. Using two separate elbow plots for the regions we defined, we found the optimal number of clusters to use to be 8.

While there was a variation between funding round types and category codes shown between clusters, we saw that the region is predominantly North America, which suggested that most startups, at least within this dataset, are based in and operated there. From this we also drew some insight as to how market segments can vary across regions. For example, software appeared twice as the most common category in that cluster (clusters 1 & 2). However, they had different regions and funding type that suggested that the software industry in North America seemed to be more mature than Europes, since the most frequent funding round type was Series-b compared to Venture (which varies more and doesn't have defined series). We can say that perhaps the opposite is true when comparing startups in the biotech industries between Europe and North America (clusters 3 & 5). However, since "other" doesn't provide much information, we can't say for sure this is the case.

We saw that from cluster 4, North America seems to appreciate riskier ventures particularly in the web category since "angel" was the most common type of funding. From cluster 7, we saw that startups in the enterprise category are usually more mature, due to the series-c+ dominant funding.

It seemed to be the standard for most startups in clusters 2, 3, and 4 to sell their company and become acquired as opposed to clusters 0 and 7 where there was a more even ratio between startups who decide to go public and startups that get acquired. This could be because of the predominant region or category present in each cluster.

In the US, compared to worldwide cluster membership, it seemed that again, 2 and 4 prefer to be acquired rather than going public. However, we saw that they both tend to follow the same pattern, suggesting that North America accounted for most of the trends being shown, since it is a large portion of our dataset.