Team DC20023 Abstract

In this case, we were given a dataset listing all the training events held by Maryland SBDC during the last five years, including the topics, format, as well as other information on the events themselves, and the demographics of the entrepreneurs that attended those events. Our goal is to find out the historical market of the training attendees, the topics of interest, to examine whether there is a significant difference in the demographics of the attendees that attended workshops in comparison to those who attended face-to-face training. And hence we are making different strategies using several models for finding out how we can recommend the people about the events conducted and mainly on what basis. The end goal is that maximum people should attend an event and hence proper recommendations regarding what type of workshop and which format need to be done. We are planning to utilize data visualization tools(ArcGIS, SAS, Tableau, Python, R) to identify geographical clusters of training attendees for different training formats.

We used the Binary Logit Model to identify the relevant demographic factors that influenced attendee's choice of attending a certain format of the training events. By looking at the p-value and coefficient estimate we are able to sort out the factors with significant influence and generate the exact increase in odds of attending a certain form of event. And so the factors with extreme influence are the ones we use to provide recommendations to the people for attending the workshops. In addition, we have reviewed a second dataset, US clustering mapping which helps us in determining the strategy for the organization according to the region, employment, the business environment, demographics, innovation, economic prosperity. This can be used to compare with our dataset and hence both can be used together.