**Data Science Capstone Project**

For this week, you will be required to submit the following:

1. **A description of the problem and a discussion of the background. (15 marks)**

There are a few private companies in Oaxaca. Actually, almost all the employed people work in government. Therefore, I would like to apply analysis to data related to this issue in order to recommend to an investor, interested in Oaxaca, to open a business based on the venues that exist right now.

I am trying to recommend to an investor to open a business in Oaxaca city, Oaxaca. Therefore, I will cluster/segmentation each Postal Code in the city based on the venues that are nearby. With this information, I should be able to see areas of opportunity, to open a business based on our cluster results.

1. **A description of the data and how it will be used to solve the problem. (15 marks)**

First, I searched for a data base that contains all the Postal Codes in Oaxaca City. Luckily, I found a website with this data "https://codigo-postal.co/en-us/mexico/oaxaca/oaxaca-de-juarez/". This data set contains all postal codes in Oaxaca, Name of the neighborhood and other irrelevant data.

After, I used geocoder package to find Latitude and Longitude for each postal code, in order to find nearby venues in everyone with the function "requests.get" to download a foursquare data base. this is used to determine which are the most common venues in a radius around each neighborhood.

Then, I will group my data in order to get the top 5th common venues in each neighborhood. In order to make a Machine Learning algorithm to cluster the postal codes (Neighborhoods) based on the venues that are near each one. So, with this information, we can easily see what venues could be an investment opportunity.