**Introduction/Business Problem :**

Using zip code as a key, the U.S. Department of Health and Human Services Office has maintained data for the state of California which shows "causes of death" from 1999-2016. These are broadly classified into 15 categories. The opportunity or the business problem to be solved is to minimize the number of deaths per year in California with possible early intervention by medical services as needed which could go into bringing down healthcare costs and improve quality of life and increase life expectancy of the residents of California.

**Github Link:** <https://github.com/nairsubhash/Coursera_Capstone/blob/master/California_causes_of_death.ipynb>

**Raw Data:**

### 1. California Leading Causes of Death by ZIP Code: <https://healthdata.gov/dataset/leading-causes-death-zip-code>

### 2. California zip code, latitude and longitude data: <https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/?refine.state=CA>

**Reports:** The note book (see github link above) shows the following reports:

1. Report shows deaths in 2016 in California where the cause of death was "NEP" i.e. deaths due to Nephritis, Nephrotic Syndrome and Nephrosis (Kidney diseases).
2. Top causes of death by the provided zip code during the past ~17 years (Beverly Hills zip code 90210 where heart disease and cancer are at the top).
3. Zip codes in descending order by number of deaths by suicide during the past ~17 years. The City of Carmichael (95608 zip code) in SACRAMENTO county has the highest number.
4. 15 most common causes of death per every zip code in California.
5. 5 different clusters of zip codes created via k-means clustering.

Note: Python notebook can be easily modified to accommodate more reports. My code is available as an open source on github (see link above).

**Conclusion:**

With the identification of zip code clusters, California could do the following:

1. Based on the causes of death, the state could better allocate and/or train resources such as hospitals, emergency care centers, medical, pharmaceutical, doctors, nurses etc..
2. These resources could be shared by zip codes that fall into the same cluster.
3. Overall better utilization of medical services could help better the quality and length of life for the people of California.

**Clustering on the map of California:**

