

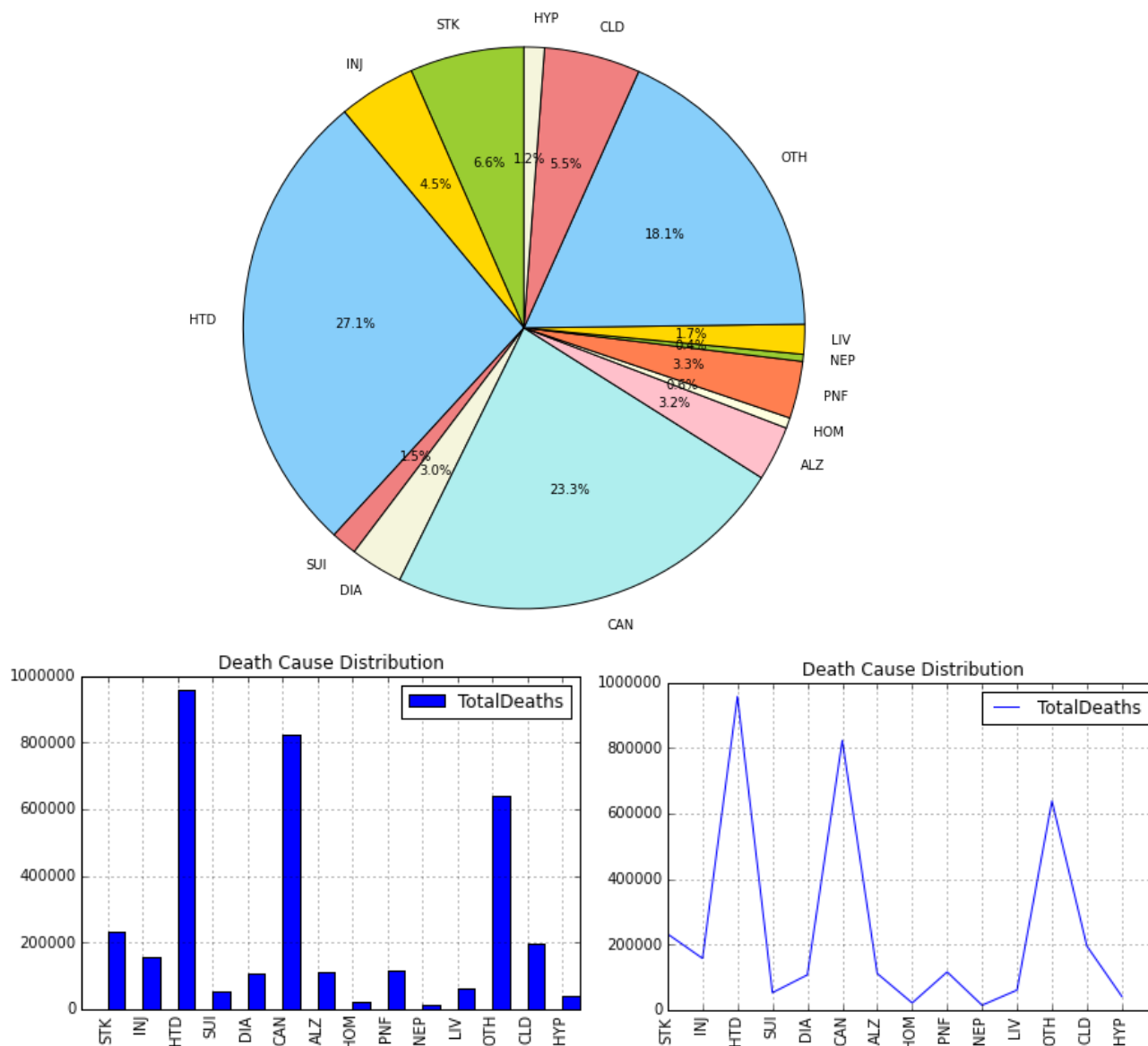
Health Data Analysis:

Name: Pooja Yelure

git-hub link: <https://github.com/devpoo/DeathCauseAnalysis>

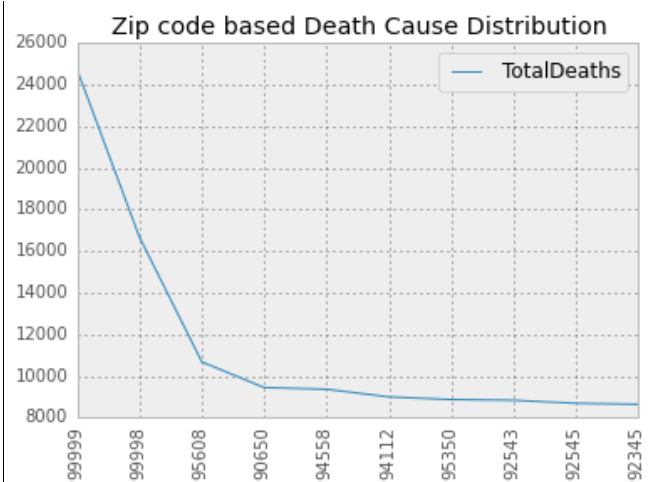
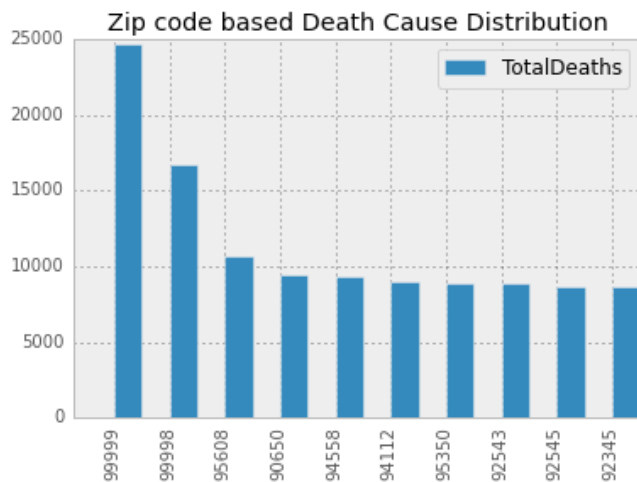
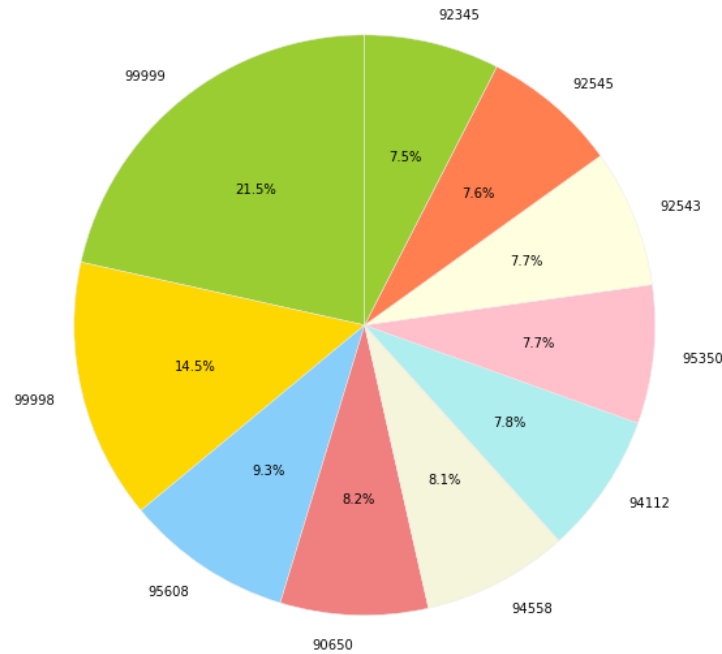
Data used for the assignment shows distribution of leading death causes in California from the years 1999 to 2013. I have made an attempt to find insights from the data to understand the distribution based various factors like causes, number of deaths its distribution over year and zip code. I have also tried to see how the trend of death causes has been over years and for different years. IBM Bluemix Apache spark service was used to answer these questions. I have listed the questions and answers answered in the report. Data Sources: <http://catalog.data.gov/dataset/leading-causes-of-death-by-zip-code-1999-2013>

Distribution of Death Causes from 1999-2013



The above figures show that the HTD, OTH and CAN have been the major causes for death.

Zip based distribution of death causes from 1999-2013



Zip Codes with highest mortality rate

10 Zip codes with highest mortality rate were used to plot the data. Zip code 99999 has the highest mortality rate over given years.

Scatter plot for zip code based death cause distribution:

The scatter plot of zip based distribution of deaths due to different causes is plotted below.

