**Capstone ideas**

1. **India Water Quality Data**

Context:

There has been increased interest among the public about the Environment and living conditions in India. Especially, after since many manufacturing units are being planned, people are worried about how it will affect the underground water quality and the environment. Government of India, under the Ministry of Drinking Water and Sanitation has released the Water Quality Affected Data for 2009, 2010, 2011 and 2012. The objective here is to analyze this data alongside with Forest, Industries, Habitation, and development projects data in the same area (panchayat) to figure out whether there is any connection between the development effort and the quality of water getting affected. This effort will identify such associations and create awareness such that people and the Govt. can act in time to avoid further deterioration of the water resources.

Data:

I will get Data for this project from Kaggle - India Water Quality Data dataset.

1. **Top 500 Indian Cities**

Context:

This data set merging **the census 2011 of Indian Cities with Population more than 1 Lac** and **City wise number of Graduates from the Census 2011**, to create a visualization of where the future cities of India stands today.

Data:

I will get Data for this project from Kaggle - Top 500 Indian Cities dataset.

1. **IMDB** **5000 Movie Data**

Context:

How can we tell the greatness of a movie before it is released in cinema?

This question puzzled me for a long time since there is no universal way to claim the goodness of movies. Many people rely on critics to gauge the quality of a film, while others use their instincts. But it takes the time to obtain a reasonable amount of critics review after a movie is released. And human instinct sometimes is unreliable.

Data:

I will get Data for this project from Kaggle - IMDB 5000 Movie dataset.