INFO – 7250

Suryanarayana Malladi

001052531

Project Proposal

The project will use the dataset from US government’s Department of Education. The dataset is called College Scorecard. College Scorecard contains data about institutions and their field of study.

The institute data is from 1996 to 2019. Each institute has almost 2000 rows worth of data. Each institute also has information about their field of study in another csv file.

The kinds of analysis I intend to perform are

1. Find number of schools per zip code
2. Find schools with highest enrollment by state.
3. Find schools which give the highest scholarships.
4. Sort and display schools by graduation rate.
5. Sort by number of students who receive scholarships in a zip code.
6. Sort states by number of incoming students.
7. Highest decrease/increase percentage in enrollment in 2019.
8. Colleges in which students graduate with lowest debt.
9. List of colleges which offer Masters degree in Computer Engineering
10. List of College name and its url for a particular state offering Doctorate degree

The above analysis covers all kinds of Map Reduce Analysis covered in classes till data.