

Project Proposal: FA17: CLOUD COMPUTING FOR DATA SCI: 34212

Title: An explorative analysis of H1B petitions data

Team Members:

Mani Manjusha Kottala

Yashwanth Reddy Kuruganti

Problem Description:

Data Analytics helps us to get some great insights from the huge chunks of data available. Every single application in the present era is driven by data, and being able to get the insights from this data makes our lives a lot more simpler and achieve great results. Following the recent Presidential Elections, there has been a lot of chaos in the United States about the H - 1B visas and skilled immigrants, this made us think about exploring and analyzing the H - 1B petitions so that we as International Students can get some insights which are worth the time we put in.

We will be able to find insights like, which part of the United States has the largest number of Data Scientists or which Part has the least number of Software Engineers, which industry has the most data scientist petitions for H -1B visas, we will also be finding some insights like which job role has the highest success rate in H -1B visas.

Data Description:

Size of the dataset (csv format) is around 1GB and consists of 30 lakh records for the year 2010-2016. H - 1B dataset is freely available at [kaggle.com](https://www.kaggle.com/datasets/ashwathkumar/h1b-petitions), the important columns in this Dataset are:

- Case Status
- Employer Name
- Soc Name (The name associated with the job)
- Job Title
- Full Time Position
- Prevailing Wage
- Year
- Worksite