

On your team's project page, by the deadline:

- Submit a link to a pdf file that describes the following:
 - the set of questions that you want to answer (these may be somewhat vague right now),
 - the set of data sources that you have selected. Recall that you are supposed to select at least two data sources that will give you structured data, and at least 300 text documents.
 - a description of how you have extracted structured data from the two data sources.
 - what is it that you want to extract from the text documents
 - the names of open-source tools you have used in this project stage and a brief description of what they do.
- Submit links to where we can find structured data that you have extracted from the two sources, and where we can find the text documents. Do not zip this data, as that would require us to download and unzip the data. Instead, submit data in a browse-able format.

CS 838 (Spring 2017): Data Science

Project Report - Stage 1

Submitted by Group 12

Team Members:

- Deepanker Aggarwal (deepanker@cs.wisc.edu)
- Saket Saurabh (ssaurabh@cs.wisc.edu)
- Vishnu Lokhande (lokhande@cs.wisc.edu)

The Problem Statement:

In this project, we have posed an interesting data science problem that is best described through the following questions that we want to answer: