**Big Data Technologies –Introduction to Data Engineering Project-Data science for a cause**

Anikita Mayekar, Mohanapriya, [Sridevi Emineni](https://www.facebook.com/profile.php?id=100000506632348)

**Datasource**: Change.org

**Overview**:

Founded in 2000 by a Bronx history teacher, DonorsChoose.org has raised $685 million for America’s classrooms. DonorsChoose.org has funded over 1.1 million classroom requests through the support of 3 million donors, the majority of whom were making their first-ever donation to a public school. Teachers at three-quarters of all the public schools in the U.S. have come to DonorsChoose.org to request what their students need, making DonorsChoose.org the leading platform for supporting public education. A good solution will enable DonorsChoose.org to build targeted email campaigns recommending specific classroom requests to prior donors.

**Methodology**:

The below technologies in the Spark ecosystem are used:

1. SparkSQL
2. GraphX
3. Streaming

**Language**: Scala

**Questions**:

1. Maximum,minimum,mean,median,25th 75th percentiles of Donation column
2. Donation amount for each grade level category in top 5 states.
3. What are the top words and phrases used by Teachers in Essays
4. Statewise donors chart.(number of dots on usa map) % of donor per state,city
5. % of donor per state,city?
6. What time of year do we get max donations.
7. Which State receives Highest Average Donation Amounts
8. Cities with maximum school and type of school.
9. Most common item requested all the time.
10. Most expensive item requested from all the schools.
11. what are the most frequent Categories, SubCategories, and Resource Categories for posted projects
12. How does the ratio of Funded vs NonFunded project varies by State
13. Number of project which expired vs number of project funded.