**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. Which states has more donations?
3. Top 10 states with more number of projects
4. Total number of requests for donation in each project type
5. Donation amount for each grade level category in top 5 states.
6. Top donation amount for schools in the past 3 years.
7. Which schools has more number of free lunch requests?
8. Which Vendor has done maximum donations across?
9. What are the top words and phrases used by Teachers in Essays
10. Which category and grade are with most project completed fully funded?
11. What percent of donors are from outside of USA?
12. How many donors are have donated more than once?
13. What product is been donated the most.?
14. Satatewise donors chart.(number of dots on usa map) % of donor per state,city
15. % of donor per state,city?
16. States with highest Donors to Population Ratio(importing data from usa census)  
    https://www.enchantedlearning.com/usa/states/population.shtml
17. %of donor who are teachers.
18. Top donors and their top contributions
19. Numer of Donations per Year
20. What time of year do we get max donations.
21. What days per month do we get max donation(everypaycheck)
22. Which State receives Highest Average Donation Amounts
23. Statewise donation graph of in last 5 years
24. %of gender of teachers posting project on the site. Pie chart of teacher prefix
25. Growth in number of teachers per year.
26. Which metro type school post more project.(numbers of projects per metro school type)
27. Cities with maximum school and type of school.
28. Most common item requested all the time.
29. Top 10 items requested in bulk
30. Most expensive item requested from all the schools.
31. Famous vendors one with most fulfilled projects.
32. what are the most frequent Categories, SubCategories, and Resource Categories for posted projects
33. what kind of project are most fulfilled
34. in which quarter the projects are most posted
35. Grade Categories and Resource Categories hainvg the highest project costs
36. How does the ratio of Funded vs NonFunded project varies by State
37. Number of project which expired vs number of project funded.
38. What are the top expired projects.?