Citizen Friendly Report of DiversityDataKids.org

Team 6:

Ali Hussain Ladiwala
Anil Gubbala
Urja Naik
Priyanka Cornelius

Diversitydatakids.org ?

❖ It was established in 2014 with the support of W.K. Kellogg Foundation and the Robert Wood Johnson Foundation to carry out rigorous research to improve child well-being and racial ethnic equity in opportunities for children.

❖ It has a thriving community of users of their research and data.

❖ Basically, diversitydatakids.org is a research project that examines who our children are, whether they have what they need to grow up healthy and achieve their full potential, whether social policies are well designed to improve children's lives and how to make them better to improve equity.
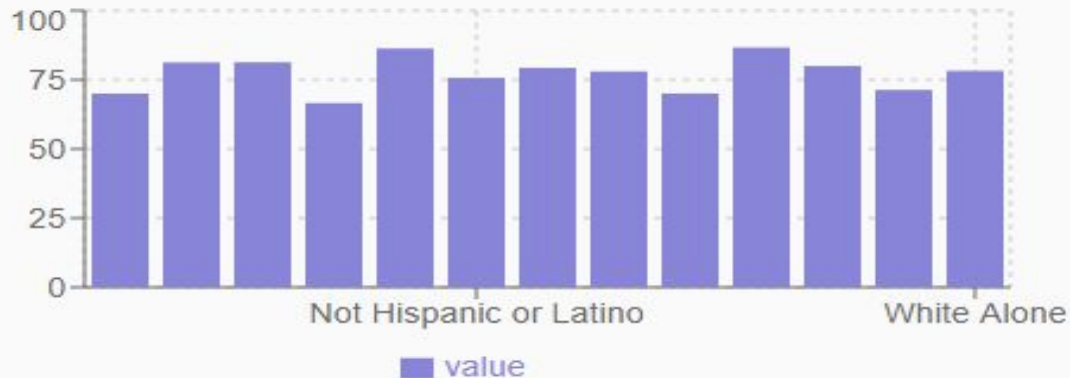
Comparison

Current Website (how it presents data?)

geoid	name	year	total_est	aian_est	api_est	asian_est	black_est	hisp_est	nhisp_est	nhopi_est	nhwhite_es
02000US1	Northeast Region	2011-2015	4.2412066	6.7760377	2.0987678	2.0667834	4.9462686	5.2160101	4.0168548	8.4929228	3.9530537
02000US1	Northeast Region	2008-2012	4.1410174	6.5575967	1.9354409	1.910761	4.6729469	5.1005025	3.9399326	6.1335983	3.8960171
02000US1	Northeast Region	2012-2016	4.2846689	6.9207163	2.0521421	2.0221786	4.8971171	5.2190385	4.0635853	7.7042804	4.0355515

Our Website

Selected Record data:

In between the years 2005-2009 the total Percent of Male labor force participation rate, ages 16-64 is 76. 041122 % with a mean value of 77 for various ethnicities. This Percent is lower than all the Census Divisions average by a value of 2. 79 %. In the selected, Some Other Race Alone has registered the highest value of 86. 662567 %. Whereas Black or African American Alone has recorded the lowest value of 66. 601349 %.

Project Demo

Persona and Use Case

❖ Our website is basically targeted on U.S children who doesn't get the proper and healthy environment in the family for their development

❖ Our website is citizen friendly so it would reach out to maximum people and they would also help to contribute in the research and help in the development of the children and change the policy to provide equal opportunities to all children.
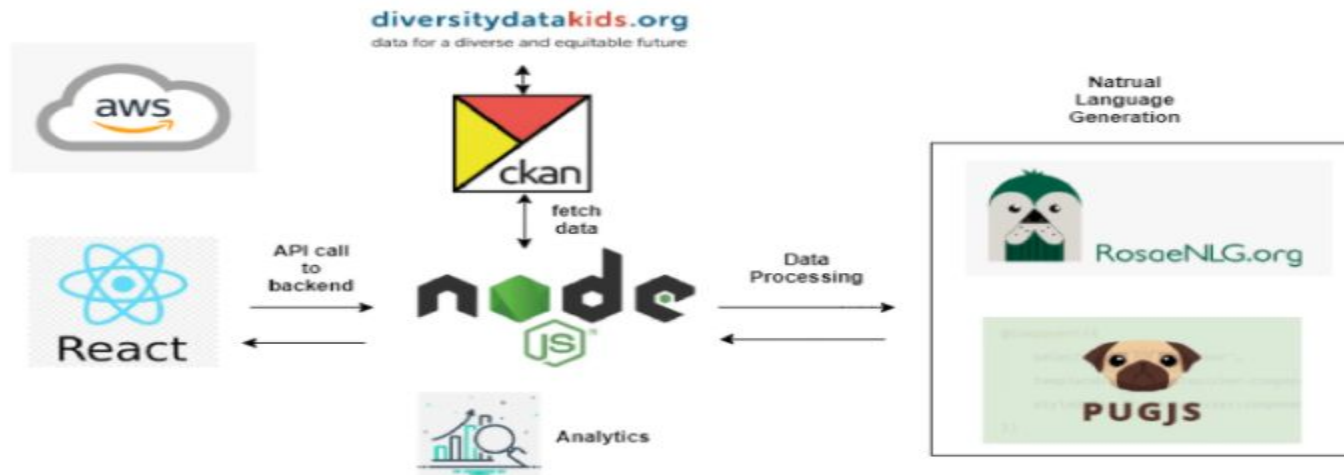
What is Ckan(Comprehensive Knowledge Archive Network) ?

❖ CKAN is an open-source DMS (data management system) for storage and distribution of open data.

❖ Mainly used by public institutions seeking to share their data with the general public. (used by various government data catalogues like data.gov)

❖ CKAN maintains information in PostgreSQL databases.

❖ Searches are implemented in Apache Solr.

❖ CKAN's API to interact with their data

❖ CKAN makes it easy to publish, share and use data.

PostgreSQL & Apache Solr

❖ Send requests in SQL format.

❖ https://data.diversitydatakids.org/api/3/action/datastore_search_sql?sql=SELECT * from "c5dd6311-2069-4338-bc8c-271a5a25e056" WHERE name LIKE 'California'

❖ For searching datasets - Solr queries.

❖ https://data.diversitydatakids.org/api/3/action/package_search?fq=&facet.field=["vocab_Subtopic"]&rows=0

Project Architecture

Natural language Generation

❖ Software process that generates natural language text
❖ A subfield of artificial intelligence (AI)
❖ Use Cases : automates the production of repetitive reports and texts based on structured data
❖ Summarize data to readable format.

Stages

- ❖ Content Determination : what data to process
- ❖ Document Structuring: organization of the information
- ❖ Aggregation: Sentence creation
- ❖ Lexical choice: Selecting of words (synonyms)
- ❖ Referring expression generation: identifying the objects
- ❖ Realization: Actual text creation

RosaeNLG & Pug js

❖ RosaeNLG is an open source Natural Language Generation Library written in Javascript based on pug template

❖ RosaeNLG works both in node.js and in the browser, for client side NLG

❖ Pug is a template engine for Node and for the browser

❖ Write reusable HTML content

NLG in our project

- ❖ Fetch data from CKAN API
- ❖ Generate statistics on the input data based on the format of each dataset
- ❖ Send dataset values & statistics to RosaeNLG
- ❖ Generate human readable Text
- ❖ Display information to the user.

Future Scope

❖ Adding more graphical Data to the website for better visualization of the data

❖ Add more analytics in NLG like what's the trend in past few years or trend of a particular ethnic group.

❖ Data visualization for the overview of whole dataset.