Theme: This year's National Big Data Challenge encourages high school students across the country to apply their computational thinking efforts on UNESCO Sustainable Development Goal 4 (SDG 4) "Equity in Education". One may use federal, provincial, municipal, and NGO Open data to seek ideas for better education for all regardless of socioeconomic, cultural, and geographic background, or even pandemic-related hardships.

Research topics can include any fields of education, such as

- Family, community, and social network support in student learning and achievement. Best practices and education results in gender, race, and low income groups of students.

- The impact of the COVID-19 pandemic on lesson delivery, assessment and evaluation: personal, peer, and Big Data perspectives.

- How internet access can assist in overcoming traditional barriers to education such as availability of buildings/classrooms, teachers, learning supplies, distance to educational facilities, accommodation and provision of learning opportunities for youth with disabilities.

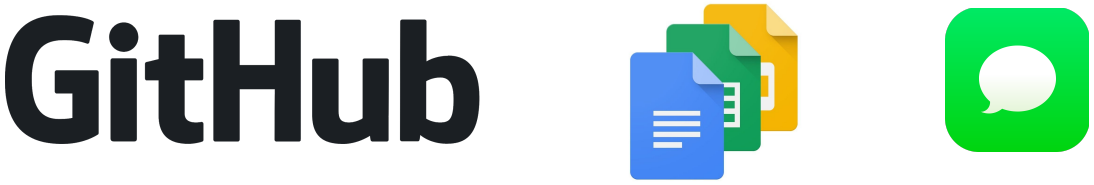
- The effectiveness of programs assisting economically disadvantaged youths on student achievement.

- Comparative analysis of in-person, blended, and online learning to achieve a more level footing for all students.

Some Powerful Statistics (topics we can base our project around):

- Today, more than 262 million children and youth are out of school.
- Six out of ten are not acquiring basic literacy and numeracy after several years in school.
- 750 million adults are illiterate, fueling poverty and marginalization.

Forms of communication:



Project Development Tools:



Important Dates:

Challenge Period (October 20, 2020 – January 22, 2021)

Crowdsource resources and investigate analytics tools (SAS, Python, Pandas, etc) choose the one you will learn and use. Workshops covering Data Science and Programming techniques will be provided to participants to help with this.

Attend mentor sessions and ask questions to learn more about anything within the realm of data science and its applications. Work on your data set for 3 months. Work together with your team, making use of your mentors, teachers, and the provided resources to analyze your data and propose solutions.

Tell the story of your data discovery through a scientific report. Use Overleaf to prepare your project report, and submit it to us.

Project Submission Deadline (January 23, 2021)

Submit your project report developed in Overleaf before the deadline (2:59 AM on January 23) for evaluation by a team of academics and industry experts.

Brainstorming:

Problem Ideas:

- How having equally funded schools is actually better (compared to the US)
    - In the US, poor neighborhoods stay poor because its the regional tax that funds schools… is that effective compared to Canada
- Has covid and working from home made education more accessible or not?
- Measuring if the hybrid method of schooling actually works (that was a school board thing for us, have different systems worked better? Even in other countries?)
- We could also look into homeschooling compared to the public school system (which is more effective) and how could we play that into a new education system that makes education more inclusive for all.
- Looking at correlation of success rate in different communities depending on how much money the education provides in a certain area
    - Does this affect class sizes

- Measuring how effective it is to provide computers and wifi (if it actually helps education)
    - Or if the computers just remain unused and it ends up being a waste of resources
    - Being in a home environment while doing school might not work for every environment

- Education in terms of income (comparing income classes and socioeconomic status's effect)

- Public vs private education
- Are different countries more effective at education than Canada and why?
- Difference in education in indigenous/ lower class communities
    - Or rural vs city

- **Canada vs. US educational**
    - See how equal distribution of wealth compared to USA's non-equal distribution of wealth for the education systems.

Public data from Canada:

https://open.canada.ca/en

How can we solve these problems?

- It requires political will, global and regional collaboration and the engagement of all governments, civil society, the private sector, youth, UN and other multilateral agencies to tackle educational challenges and build systems that are inclusive, equitable and relevant to all learners.

Our General Topic of Research: See how Canada's equal distribution of public school funding compared to the United State's non-equal distribution of public school funding affects secondary education outcomes

---

What's our objective? Our objective is to make a comparative analysis between the equal distribution of public school funding in Canada versus the non-equal distribution of public school funding in the United States at the secondary school level federally. More specifically, we are analyzing how the various education funding systems can affect the success of students. Using this data we hope to create a novel solution about funding, that will ultimately benefit the success of students using an integrative thinking method.

What are our success metrics? We have broken down the success of students in a post secondary level into 3 pillars of educational success each with sub-categories:

1. Retention Rates
    - Secondary Graduation rates
    - Total credit accumulation
    - Core courses (English, Math, and Science) drop out rates

2. Academic Performance
    - GPA
    - Reading, Mathematics, and Science Literacy Rate (https://nces.ed.gov/programs/coe/pdf/coe_cnu.pdf)
    - SAT/ACT Scores

3. Academic Engagement
    - Class absences
    - Meeting deadlines requirements??

Data Sets and Stats:
https://ourworldindata.org/measuring-education-what-data-is-available

Schedule:

Week 1 (Jan 1-3):

- Find data sets of Canadian and US school fundings….

- Take the top 5 data sets, scrape the information and data. Clean the data. Start determining and calculating various metrics…

Week 2(Jan 4-10):

Week 3(Jan 11-17):

- Start working on the report..
    - Abstract
    - Introduction
    - Materials & methods

-Keep analysing data.

- Focusing on the report:
    - Finding results…. Solutions
    - Discussion… Acknowledgement, and references..

Week 4(17-23rd):

- Polishing

ROLES:

Data Set Searcher - Deniz

Programmers - Eric, Deniz, Ana

Visuals (Scientific Report, data visualization) - Ana

Writers/Documentation: Ana and Aidan

Formatting Scientific report - Eric

Project Rubric:

https://docs.google.com/document/d/1UnzosZopa1hadZ8KAZ5SMXzKtjZYhCNRB8XQQwWG9ro/edit

1. CONTENT				
1. CONTENT				
	4- Excellent	**3- Good**	**2 - Satisfactory**	**1 – Needs Improvement**
Research Objectives	The objectives are strongly developed, explicitly stated and apparent to the reader	The objectives are developed, explicitly stated, but may sometimes digress from them.	The goals are not fully developed and are not consistently clear throughout the paper.	The goals are not developed and is generally unclear
Methodology	The methodology used is sophisticated, appropriate for the objectives of the report, and is consistent with the perspective.	The methodology used is advanced, appropriate for the purpose of the paper, and is consistent with the perspective.	The methodology used is basic, but still appropriate for the purpose of the paper, and is consistent with the perspective.	The methodology used is inadequate and not appropriate for the purpose of the paper.
Data Diversity	Diverse, representative and correlated data sets are utilized to achieve the research objectives/new interesting results.	Diverse, representative and correlated data sets are considered but not fully utilized to achieve the research objectives/new interesting results.	Diverse, representative and correlated data sets are considered but aren't fully utilized to achieve the research objectives/new interesting results.	Diverse data not considered.
Relevance	The paper makes a highly significant contribution to the field of education and equity.	The paper makes a significant contribution to the field of	The paper makes a moderate contribution to the field of	The paper makes a minimal contribution to the field of education and equity.

		education and equity.	education and equity.	
2. Results	The data sets, analysis and results are very clear, logical, and strongly support the objectives of the paper.	The data sets, analysis and results are clear, logical, and strongly support the objectives of the paper.	The data sets, analysis and results are somewhat clear, logical, and strongly support the objectives of the paper.	The data sets, analysis and results need improvement.
1. ORGANIZATION				
Language and Clarity	The paper is clear, concise, and easily understandable. The writing has little or no grammar or spelling errors.	The paper is mostly understandable, with occasional inconsistencies. Minor grammar or spelling errors are present, but do not detract from the content.	Multiple sections of the paper are difficult to read/understand. Some grammar or spelling errors are significant and detract from the content. Paper requires further editing.	The paper is difficult to read/understand due to sentence/paragraph structure, word choices, lack of explanations, etc. Significant amount of grammar and/or spelling errors which distort the sentence meanings and make it difficult to read.
Figures and Charts	Figures, tables and charts are used with clear purpose and useful to the data presentation.	Figures, tables and charts but some are extraneous.	Figures, tables and charts but many are extraneous, and the information could have been depicted in better ways.	No figures, tables, charts are used.
Use of References	Peer-reviewed conference, journal, and other approved sources	Although most of the references are professionally	Most of the references are from sources that are not	There are no sources that are professionally reliable. The

	references are included. The reader is confident that the information and ideas can be trusted.	legitimate, a few are questionable (e.g. trade books, internet sources, popular magazines). The reader is uncertain of the reliability of some of the sources.	peer- reviewed and have uncertain reliability. The reader doubts the accuracy of much of the material presented.	reader doubts the value of the material.
Conclusion	The conclusions are very well formulated and are strongly supported by the data.	The conclusions are well formulated and are supported by the data.	The conclusions are moderately effective and are only partially supported by the data.	The conclusions are minimally effective and do not appear to be supported by the data