

# Cultural Relevance for Epistemic Practice in High School Computational Data Mining

Alex Acquah  
*College of Education*  
*The University of Texas at El Paso*  
El Paso, USA  
aacquah@miners.utep.edu

Amanda Barany  
*Graduate School of Education*  
*University of Pennsylvania*  
Philadelphia, USA  
amanda.barany@gmail.com

Andi Scarola  
*Department of Computer Science*  
*The University of Texas at El Paso*  
El Paso, USA  
adscarola@miners.utep.edu

Michael A. Johnson  
*Department of Learning Technologies*  
*University of North Texas*  
Denton, USA  
michaeljohnson.unt@gmail.com

Sayed Mohsin Reza  
*Computer Science*  
*Pennsylvania State University*  
Harrisburg, USA  
skr6024@psu.edu

Christopher Rivera  
*Computer Science*  
*The University of Texas at El Paso*  
El Paso, USA  
carivera7@miners.utep.edu

Justice T. Walker  
*College of Education*  
*The University of Texas at El Paso*  
El Paso, USA  
jtwalker@utep.edu

**Abstract**—This research-to-practice full paper examines the integration of culturally responsive computing (CRC) within high school data science education. It presents a theoretical and practical framework that leverages cultural backgrounds to enhance the teaching and learning processes in computational data mining. The main objective of this research is to demonstrate how integrating CRC can improve educational experiences by making them more inclusive and engaging for students from diverse cultural backgrounds. By applying CRC, educators can tailor learning experiences that transmit technical skills and conceptual knowledge and resonate more deeply with students’ personal and social identities. This approach has the potential to foster deeper engagement and understanding of the material, thereby enhancing learning outcomes. Our research operates within the broader context of ongoing efforts to integrate data science across K-12 education, advocating for educational practices that acknowledge and utilize students’ cultural identities as assets rather than deficits. The theoretical foundations for these practices are rooted in existing literature that emphasizes the importance of culturally relevant pedagogies in computing and data science education. Implementing these frameworks in classroom settings represents a vital intersection of educational research and practice, aiming to address disparities in engagement and achievement among minority students.

This work leverages qualitative analysis of how high school students utilize their cultural and social orientations within a structured “coding like a data miner” workshop. The dataset includes students’ final project presentations, which showcase their ability to integrate cultural relevance into their epistemic data science practices. We highlight how these integrations influence learners’ understanding of and engagement with data science concepts. The results are anticipated to contribute to the discourse on how data science education can be made more accessible and effective through culturally responsive methodologies. This contribution is valuable for preparing a diverse student

population to thrive in a data-driven future, ensuring they are equipped with technical skills and can apply them within their cultural contexts.

**Key Words**—Culturally Relevant Computing, Epistemic Practices, and Computational Data Mining, Data Science Education

## I. INTRODUCTION

The integration of computational data science within STEM education is increasingly recognized for its potential to drive significant innovation and societal progress. Studies have demonstrated that engaging students in this dynamic field enhances their technical skills and deepens their understanding of complex data systems, which are crucial in the modern digital landscape. [1], [2], [3] Educators are striving to introduce high school students to data science by designing activities that engage them in what we define as “epistemic practices.” Epistemic practices involve the processes used to generate, assess, and disseminate knowledge in data science [4], [5]. These practices have demonstrated utility for equipping students with the skills necessary to effectively navigate and manipulate data, enabling them to create insightful and contextually relevant knowledge. Despite these advancements, challenges remain around supporting inclusivity in current educational strategies. Existing data science programs tend to emphasize what students lack rather than capitalizing on the diverse skills and perspectives they bring to the classroom. For instance, [6] focuses on computational thinking, and Next Generation Science Standards [7] present the Bootstrap: Data Science curriculum, designed to connect data science

and algebra pedagogy. This often leads to a disconnect for marginalized learners, who might feel that these programs do not reflect their experiences or value their contributions. Culturally responsive computing (CRC) has emerged as a powerful framework to address these concerns. CRC promotes an educational approach that values and integrates students' varied cultural backgrounds into the learning process, making STEM fields more inclusive and relatable [8], [9]. This approach acknowledges and actively incorporates students' cultural identities, enhancing their engagement and participation.

Building upon these foundational ideas, recent initiatives have tailored learning activities to teach computational data mining and integrate it with open-ended inquiries and culturally relevant computing [10], [11]. These practices encourage students to explore topics of personal and cultural significance through data analysis, which has the potential to enhance engagement and understanding of the material. This approach addresses the critique that much of traditional data science education may disconnect from social issues or fail to convey the inherent messiness of epistemic practices, which is crucial for authentic learning experiences [12], [13]. This study examines how high school students utilize their social and cultural orientations in a data mining workshop to explore how cultural relevance intertwines with their epistemic practices. We can gain insights into student learning outcomes through a qualitative case study of workshop artifacts, such as participants' final projects. This analysis helps us understand how various elements contribute to more effective and inclusive educational practices in computational data mining and broader STEM education. Ultimately, this prepares a diverse student population to navigate the complexities of the future in data science education. We presented students' final project presentations, where they engaged in metacognitive processes throughout the data mining stages, to address the following research questions: (1) How do students integrate their cultural backgrounds into their data mining projects, and how does this integration support their understanding of data mining concepts? (2) Do culturally relevant epistemic practices influence students' ability to generate, evaluate, and communicate knowledge in their final data mining projects?

The first research question investigates how students integrate their unique cultural perspectives into their data mining projects. By examining students' final presentations, we can discern how they embed cultural elements into their data analysis, enhancing their understanding and application of data science concepts. This analysis will evaluate the role of Culturally Responsive Computing (CRC) in deepening students' connections to data science education. The second research question concentrates on the epistemic practices employed in students' projects, specifically how integrating cultural relevance influences their methods for generating, evaluating, and sharing knowledge. This approach will provide insights into the impact of cultural considerations on students' epistemic strategies and overall learning outcomes.

By analyzing the students' presentations for evidence of these practices, the study seeks to determine how CRC influ-

ences the epistemic quality of their work and their proficiency in presenting complex data-driven insights.

## II. BACKGROUND

The expanding discussion on incorporating data science into K–12 education highlights the need to adapt our teaching strategies to the realities of the digital era. Recent literature, including studies by [14] and [15], illustrates the increasing relevance of data science not only in higher education but also at the elementary and secondary levels, as emphasized by [16]. This push for inclusion in the K–12 curriculum is gaining momentum, backed by robust support from global policymakers who advocate for educational frameworks that equip young learners with critical data fluency skills necessary for their future.

However, amidst these progressive movements, there remains a significant need for more clarity about what effective data science education should look like for younger learners and how it can be integrated into the already dense school curricula without causing undue strain [17]. This unclear definition presents a critical opportunity for pedagogical innovation but raises concerns about whether current educational approaches adequately prepare students for a data-driven world. This situation calls for a detailed exploration of potential educational models and strategies to effectively bridge this gap, ensuring that data science education is comprehensive and accessible for all students.

While data science education is perceived as a natural extension of mathematics [7] or computer science, notable programs and curricula have begun exploring this integration with significant outcomes [18]. Yet, the focus remains predominantly on standalone courses. This narrow focus could limit data science education's broader applicability and relevance across various disciplines.

In addressing these concerns, it is essential to consider a shift away from data science as a standalone subject toward its application and integration across various disciplines. The literature suggests that data science education should not be siloed within mathematics or computer science but should enrich learning across all subject areas, including social studies, science, and the arts [17], [19]. This interdisciplinary approach can make learning more relevant and engaging by connecting it with real-world issues and students' personal interests. However, the current focus predominantly remains on standalone courses, which may not effectively harness the full potential of data science education in fostering a comprehensive understanding and applicability across different areas of knowledge.

To bridge this gap, this paper proposes adopting a Culturally Relevant Computing (CRC) framework that recognizes and leverages students' diverse cultural backgrounds as assets in the learning process. This approach aligns with the perspectives and methodologies advocated by [8], [9], [20], who emphasize the importance of culturally responsive pedagogies in computing. This solution is also informed by the works of [21], [22], [23] and other scholars advocating for culturally relevant

computing. By integrating CRC, educational practices in data science can become more inclusive, engaging, and effective, particularly for marginalized groups. Furthermore, focusing on epistemic practices within data science—processes involved in creating, evaluating, and disseminating knowledge—can give students a deeper understanding of how to interact with and interpret data. [24], [25] discuss the importance of nurturing these epistemic practices, which are essential for developing critical thinking and problem-solving skills in a data-driven world. By conceptualizing data science, education through the lens of CRC, and robust epistemic practices. This paper aims to outline a comprehensive framework that addresses the identified educational gaps and enhances the relevance and applicability of data science education in preparing students for future challenges. Informed by these theoretical foundations, the 'Coding Like a Data Miner' project aims to apply these concepts within a pedagogical framework practically. The following section details the methodologies employed to implement and assess the effectiveness of culturally relevant data science education, as proposed in the background.

### III. METHOD

#### A. Context and Participants

The "Coding Like a Data Miner" initiative, supported by the National Science Foundation, represents a novel approach in educational settings by integrating data science into curricula through a sandbox-style format specifically designed for social media platforms like Twitter/X. This project aims to enable learners to extract and analyze real-world data that resonates with their interests and cultural backgrounds, fostering a connection between data science practices and their everyday lives [10], [12]. Such an approach ensures equitable teaching of epistemic practices by embedding them within culturally relevant topics.

The curriculum was implemented over two weeks through an online workshop, featuring four hours of daily instruction from Monday to Friday. Platforms like Zoom and Google Colaboratory enabled this digital learning environment, fostering interactive and engaging educational experiences. The initial week introduced the basics of data mining, covering topics such as data pre-processing, statistical analysis, coding, data collection, and visualization. This stage involved guided learning, where participants adapted and experimented with pre-existing code samples and data sets [10], [26]—the subsequent week encouraged a shift towards independent inquiry, allowing students to formulate their research questions and develop comprehensive data collection, analysis, and presentation strategies. The program concluded with each participant delivering a detailed presentation of their projects, showcasing data science methods' application to issues pertinent to their interests.

The workshop was attended by eight adolescents aged 13 to 17, representing a diverse mix of ethnic backgrounds, including three of Asian descent, three African American/Black, one Latinx, and one Caucasian/White, with equal gender representation. Due to connectivity issues, analyses were conducted

on presentations from seven participants. The study focuses on the experiences of four high school participants, John, Samira, Joiy, and Joseph (pseudonyms), aged between 14 and 17. Their contributions were particularly insightful for understanding how culturally responsive data science education can impact minority students. We purposefully sampled these projects because their work offered valuable insights into how data science education interfaces with culturally diverse backgrounds. Each of these projects represents a unique perspective and engages with significant socio-political themes through their projects, making their contributions valuable for understanding the broader impact of culturally responsive data science education, especially among minority students. John's project, focusing on the online gaming community, delves into how digital players express their interests. Samira's investigation into addressing women's rights/women's evolution represents a sociopolitical concern. Joiy's exploration of diabetes prevalence in relation to age and gender taps into public health discourse influenced by cultural and lifestyle factors. Joseph's work on the Israel-Palestine conflict addresses the complexities of geopolitical discourse shaped by media narratives. These projects exemplify how students from varied backgrounds use data science tools to engage with and reflect upon issues that resonate with their personal experiences and cultural histories. This is particularly relevant in culturally responsive education, which seeks to validate students' personal identities and experiences as integral elements of their learning process. Moreover, analyzing these projects allows us to understand the effectiveness of data science education in fostering critical thinking and problem-solving skills across diverse contexts. By focusing on these four participants, the study seeks to offer a better understanding of how students leverage their cultural and socio-political awareness to formulate research questions and use data science methodologies to explore answers.

#### B. Data Sources and Analysis

The analysis utilized video recordings of the students' final presentations and their corresponding slides, either in Google Slides or PowerPoint format, focusing on the different phases of the data mining process. This included forming research questions, collecting and processing data, analyzing it, and finally communicating the findings. These artifacts provided a rich source of information for understanding how participants applied data mining techniques within culturally relevant contexts. This study employs the qualitative case study methodology for data analysis [27], [28], [29] and defines a case study as an empirical inquiry that explores a contemporary phenomenon within its real-life context, mainly when the boundaries between the phenomenon and context are blurred. This method is ideal for exploring the nuanced ways students engage with data mining, emphasizing the search for deep understanding and meaning, the researcher's active role in data collection and analysis, and using a deductive approach to develop a rich, descriptive end product. This study's use of culturally relevant themes and qualitative case study methodology addresses existing educational gaps and

pushes the conversation on making data science education more inclusive and impactful for diverse student populations.

We present three cases of participants’ final project presentations that illustrate the key themes in the data mining workshop to understand. (1) How do students integrate their cultural backgrounds into their data mining projects, and how does this integration support their understanding of data mining concepts? (2) Do culturally relevant epistemic practices influence students’ ability to generate, evaluate, and communicate knowledge in their final data mining projects?

Seven participants engaged in the final projects following the completion of the brainstorming phase of the implementation. During this initial phase of the participants’ project development, they were guided through a structured brainstorming process designed to help them clearly define and articulate the framework for their projects. This guidance was designed to help them identify and refine the core components of their projects, which are the problem, the need, and the solution. Here’s a breakdown of each component: **Need:** This part of the framing involved students articulating why it is important to address the problem. This typically involves demonstrating the relevance of the problem to their community, field of study, or a broader societal context. The ‘need’ helps justify the research and often drives the motivation behind the project. **Problem:** Students were prompted to identify a specific issue or challenge that their project would address. This could be a gap in existing knowledge, a societal issue, or a technical challenge. By defining a clear problem, students could focus their research and project development efforts more effectively. **Solution:** Students were then asked to propose a potential solution or set of solutions to the problem identified. This could involve designing a new product, proposing a change in existing processes, or suggesting new areas for further research. The solution phase encourages students to think critically and creatively about how to address the need and solve the problem effectively. By providing prompts for these three key elements—problem, need, and solution—we aimed to ensure that students developed well-thought-out, feasible, and impactful projects. This structured approach helped participants clearly articulate their research questions and align their projects, with practical outcomes and real-world applications. Such a framework supports students in moving beyond mere idea generation to detailed project planning and execution, as shown in Figure 1 below.

#### IV. RESULT

We present three themes identified in the participants’ final project from the “Coding like a Data Miner” workshop. These projects were selected based on their ability to illustrate the integration of participants’ cultural history, personal interests, sociopolitical concerns, and the strategies and rationales employed. Each project showcases the participants’ unique cultural or sociopolitical interests, providing insights into their epistemic practices within data science.

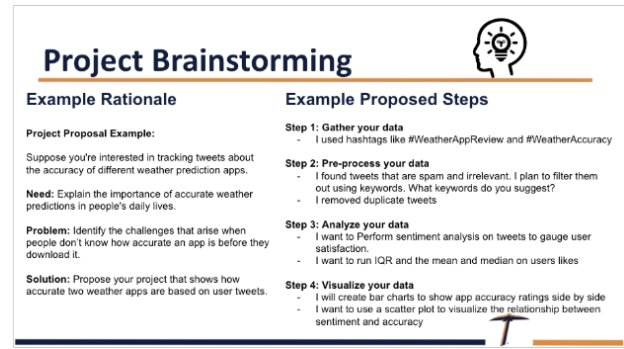


Fig. 1. Project Brainstorming Guide

TABLE I  
THEMES FROM PARTICIPANTS’ PROJECTS

Theme	Explanation
Cultural History	Involves aspects of a participant's culture that influence their perspective or involvement in a project. This includes ethnicity, language, customs, and traditions.
Personal Interest	Involves an artifact or project initiative that directly relates to the personal passions or hobbies of the participant. This might include projects that are used for personal enjoyment or professional gain.
Sociopolitical Concerns	Concerns that are associated with both social and political factors. This theme relates to how the project addresses issues like equality, justice, or social policies.
Strategies and Rationale Employed	Refers to the particular methods and reasons employed by participants to conduct the project. This includes the pedagogical approaches, technologies used, and the rationale behind these choices, emphasizing how they align with

#### A. Project 1: Analyzing League of Legends Community Trends Which Demonstrate Personal Interest and Formation of Research Questions

John (pseudonym), a 17-year-old student, embarked on a project titled “Analyzing League of Legends Community Trends.” His research was driven by his self-described personal interest in understanding the current popular trends within the League of Legends video game community. By focusing on this topic, John aimed to identify what aspects of the game and its community content engage players the most. He utilized the social media platform “X” (formerly Twitter) to analyze user-generated content, providing a perspective on the community’s priorities and preferences. The core of John’s project was forming a research question that sought to highlight the prevailing interests and discussions within the League of Legends community. This methodical approach involved several key steps: (1) conceptualizing the idea based on his personal interests in the gaming community,

(2) formulating a focused research question, (3) collecting data from social media posts, (4) analyzing this data to detect patterns and trends, and finally, (5) employing data visualization tools such as charts and word clouds to present his findings effectively. A section of the process used in his project is shown in Figure 2 below. John initiated this project to merge his personal passion for gaming with his academic pursuits in data science, making the learning process more engaging and directly relevant to real-world applications. This project exemplifies how educational endeavors can effectively incorporate students' interests, applying modern technological tools to analyze and interpret complex datasets. By showcasing how to harness social media as a powerful analytical tool, his project is a model for other researchers interested in similar studies. Furthermore, it highlights the importance of aligning educational projects with students' interests, which can be an initial step to enhancing engagement and the practical relevance of academic research. This project demonstrates the integration of cultural and personal interests with research question formation.

#### *B. Project 2: Addressing Women's Rights/Women's Evolution, which Represents a Sociopolitical Concern*

Samira, a 17-year-old high school student, embarked on a research project to explore the evolution of women's rights over time. She described her investigation as deeply personal, motivated by her identity as a young black female and her desire to understand and articulate the challenges and advancements concerning women's rights. Samira utilized social media analytics, specifically analyzing tweets tagged with #WomenRights, to gain insights into current discussions about women's rights and their representation in online communities. Figure 3 shows a section of her project. Samira identified a significant theme that resonated with her socio-political concerns. She formulated a research question to understand whether women's rights have progressed and if women today face fewer obstacles than in the past. To tackle this question, she chose Twitter as her data source, recognizing its utility in capturing real-time public opinion and discourse. Samira collected data through the social media platform, focusing on tweets that discussed various aspects of women's rights. She then analyzed this data using quantitative methods such as frequency analysis and qualitative approaches like content analysis to identify prevalent themes and sentiments. The results were visualized using pie charts, word clouds, and histograms. The creation of this project was driven by Samira's aim to provide a voice to young women who may feel marginalized or underrepresented. By highlighting the evolution of women's rights, she sought to offer a platform for discussion and empowerment. This project provided a space for advocacy, linking broader socio-political concerns with the academic study of evolving gender norms. By analyzing the discourse around women's rights on social media, this project transcends personal interest, illustrating the impact of social research on public awareness and policy dialogue. It serves as a model for how academic research can address vital societal

## Rationale: Need

### Project Topic:

League of Legends has an extensive community with different aspects, I want to see what parts of the league content and its community are popular at the moment.

### Importance:

It is important to know what is going on in the community and see what is popular to stay in the loop. Also see what the makers of league are doing with the game.



## Rationale: Solution

### Solution:

One way to accurately know what is occurring within League of Legends is by looking at a social media platform like Twitter, now called "X".

This mini-project plans by collecting data from looking at the contents of League related tweets to see what the netizens of the internet are interested in.



#LeagueOfLegends

## Visualizations

### Step 4: Visualize your data

I created a word cloud from the collected Tweet data.

Pros of this visualization is the most frequently said words are easily shown which can be analyzed to show what is popular in the game currently.

The cons of this visualization is that you are unable to see the context of the words.



Fig. 2. Project on Analyzing League of Legends

issues, reflecting the ways students can harness their studies to contribute to important socio-political conversations.

This approach has the potential to enrich the educational experience and enhance student engagement by connecting learning outcomes with real-world applications and social change. Her findings offer insights into women's rights for the community, encouraging informed discussions and actions. Moreover, her methodology showcases how social media data can be leveraged to gain real-time insights into public opinion, offering a template for other researchers interested in similar topics. Samira's project highlights the enriching experience of pursuing research driven by her socio-political concern. It highlights the importance of allowing young researchers to explore topics that resonate with their identities, which can lead to profound learning experiences and contributions to societal discourse.



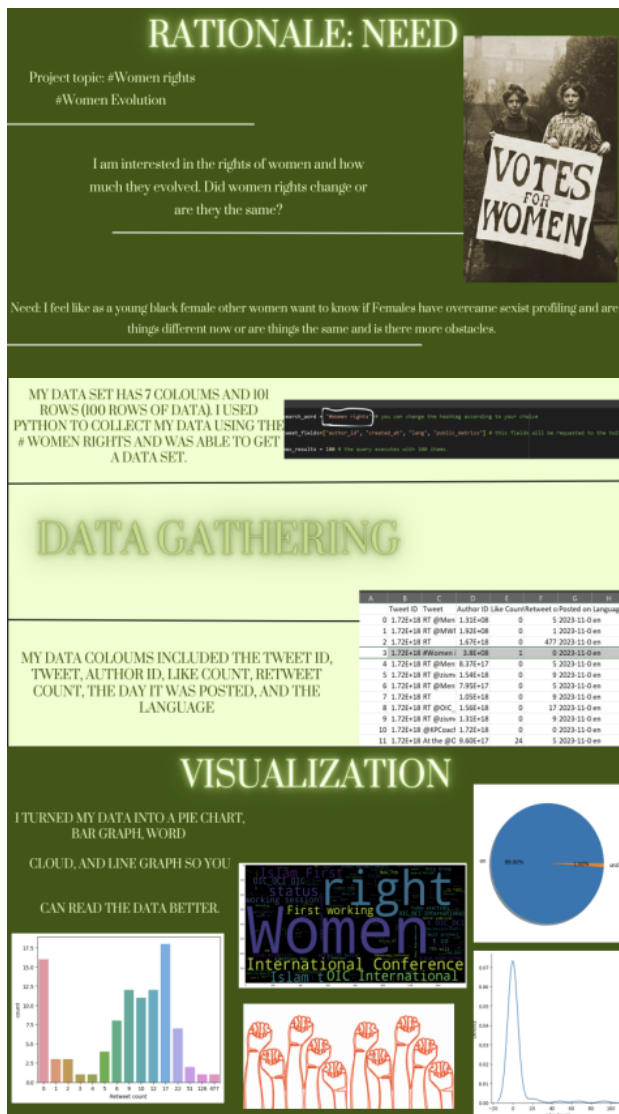


Fig. 3. Project on Women's Rights/Women's Evolution

### C. Project 3: Israel-Palestine Conflict Which Demonstrates Strategies and Rationales Employed

Jackson, a 17-year-old student, embarked on a data science project aimed at dissecting the complexities of the Israel-Palestine conflict, mainly focusing on the misinformation surrounding it. He chose to investigate how different narratives are framed and propagated through social media, specifically Twitter, now called "X." This project focuses on the current dynamics of public opinion and the spread of misinformation online. He employed a strategic approach to identify, analyze, and clarify the misinformation in discussions about the Israel-Palestine conflict. His process began with formulating a straightforward research question: How is the Israel-Palestine conflict represented on social media, and what misinformation exists within these discussions? To tackle this question, Jackson methodically collected tweets that mentioned the conflict, employing data mining techniques to

analyze the content and sentiment of these tweets. He then used data visualization tools, such as charts, to present his findings in a manner that demonstrated the prevalence of specific themes and misinformation, as presented in Figure 4. Jackson's project aimed to highlight the online discourse surrounding the Israel-Palestine conflict, distinguishing between factual and fabricated content that often muddies understanding of such complex issues. His project aimed to contribute to the broader discourse by providing insights that could lead to more informed discussions and a better understanding of How Social Media Can Influence Public Perceptions of Geopolitical Issues. Jackson's work highlights the vital role of critical thinking and media literacy in the digital age. By showcasing how data science can be applied to real-world problems, the project serves as an educational model that encourages other students to engage critically with information they encounter online. It also highlights the importance of data mining skills in navigating the modern information landscape, where misinformation can easily be spread on global platforms. This approach is crucial for cultivating a generation of informed individuals who can thoughtfully engage with complex global issues.

### D. Project 4: Addressing How Gender and Age Affect your Likelihood of Getting Diabetes, Demonstrating how People's Cultural background can Affect Their Chances of Getting Diabetes

Joiy, a 14-year-old student, embarked on a research project titled "Addressing How Gender and Age Affect Your Likelihood of Getting Diabetes," where she investigated the intricate connections between age, gender, and the prevalence of diabetes. This project resonates deeply with the theme of cultural history, as Joiy aimed to uncover how entrenched cultural eating habits and lifestyle choices influenced by one's background contribute to health outcomes related to diabetes. Joiy began her project by posing a critical question: Are certain age groups and genders more susceptible to diabetes, and how do their cultural practices around diet and health influence this susceptibility? This question led her to utilize Twitter, a platform where diverse global voices discuss and share their lifestyles and health practices. She focused on hashtags like # diabetes, # diabetesstatistics, and # cdc\_diabetes to gather data, aiming to capture a broad spectrum of individual and community health narratives that reflect both personal choices and culturally ingrained habits. Through her methodical approach, Joiy collected and analyzed tweets to identify prevalent discussions around diabetes, which often included mentions of dietary habits, age-specific health concerns, and gender-specific health discussions. She applied quantitative analysis to determine the frequency of discussions linked to age and gender, and qualitative analysis to understand the context and nuances behind these discussions. The findings were meticulously represented through visual aids like charts and graphs, which helped illustrate the complex interplay between cultural habits and diabetes risk. The motivation behind Joiy's project was not just academic curiosity but a driven purpose to educate and inform. She recognized

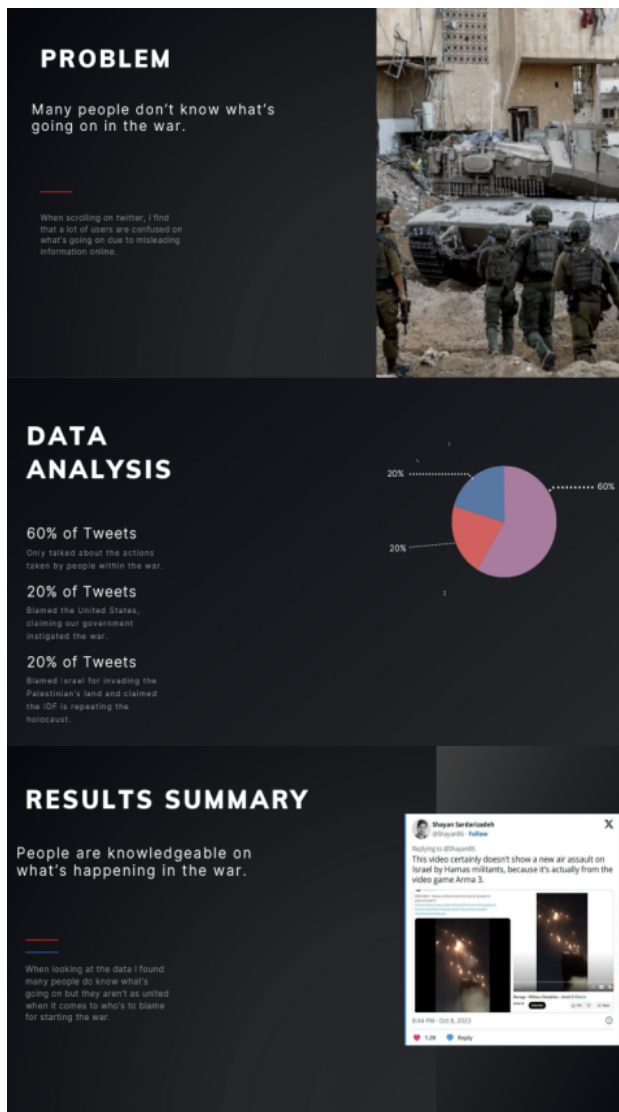


Fig. 4. Project on Israel-Palestine Conflict

that many people are unaware of how their age, gender, and deeply rooted cultural practices around food and health could affect their diabetes risk. Highlighting these factors, her project aimed to foster greater awareness and encourage preventative measures within communities. The implications of Joiy's research extend beyond the academic sphere. Sections of her project are presented in Figure 5.

By linking diabetes risk to cultural practices and demographic factors, she provides valuable insights that can influence public health policies and individual behaviors. Her work underscores the need for culturally sensitive health education that respects and incorporates traditional practices while promoting healthy choices. In essence, Joiy's project serves as a valuable example of how young researchers can use data science to explore and address health-related issues. It shows that academic projects, especially those conducted by younger students, can have a meaningful impact on public

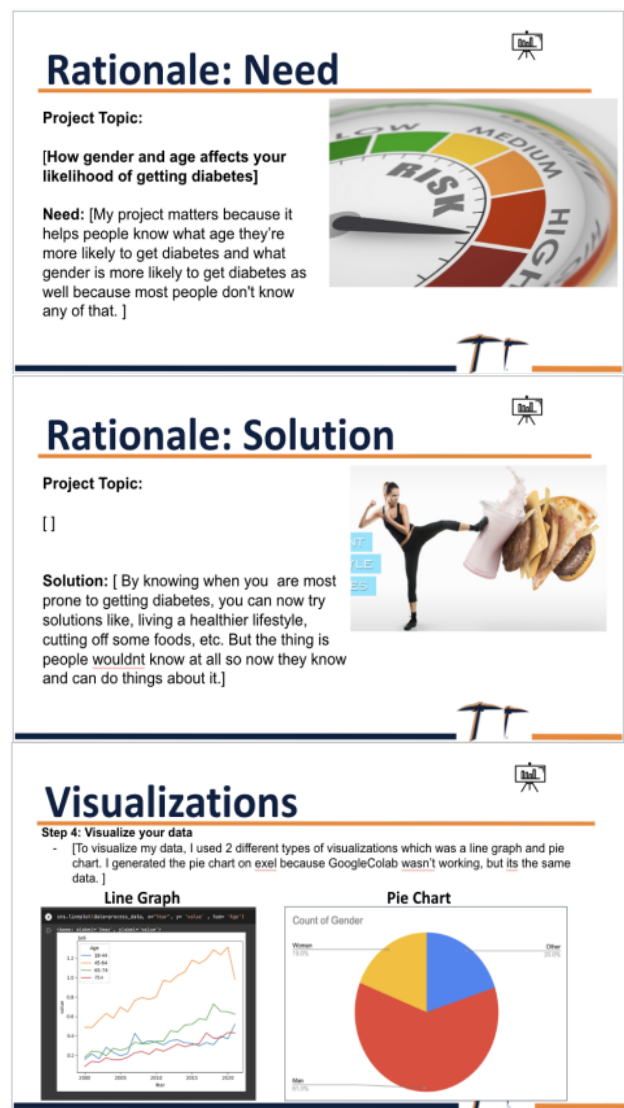


Fig. 5. Addressing How Gender and Age Affect Your Likelihood of Getting Diabetes

understanding and health outcomes, particularly when they illuminate the connections between culture, lifestyle, and disease.

## V. DISCUSSION AND IMPLICATIONS

With an emphasis on cultural relevance and epistemic practices, this research highlights the importance of aligning educational content and methods with students' diverse backgrounds and experiences. Here are some of the primary implications derived from this study:

### A. Enhanced Student Engagement and Learning Outcomes:

The study suggests that students might be engaged and achieve better learning outcomes when their cultural backgrounds are considered in the design and implementation of data science curricula. This engagement is facilitated by incorporating culturally relevant content that resonates with their

personal and community experiences, making the learning process more meaningful and relatable.

#### *B. Promotion of Inclusive Educational Practices:*

The results underscore the need for inclusive educational practices for all students, particularly those from marginalized communities. Culturally responsive computing (CRC) is an effective framework to ensure that all students feel valued and included in the learning process, which is essential for fostering a supportive and productive educational environment.

#### *C. Development of Critical Thinking and Problem-Solving Skills:*

The emphasis on epistemic practices within the curriculum supports the development of essential skills such as critical thinking and problem-solving. These skills are crucial for students as they navigate and make sense of complex data sets, enabling them to draw meaningful conclusions and make informed decisions based on the data.

#### *D. Curriculum Development:*

For curriculum developers, this study might support the development of educational policies and curricula that better cater to the needs of a diverse student population. Integrating CRC and epistemic practices into data science education can serve as a model for how other disciplines might approach curriculum development to enhance equity and access in education.

#### *E. Future Research Directions:*

This research opens several avenues for further investigation, particularly in exploring how different aspects of culturally relevant pedagogies can be effectively implemented in various educational contexts. Additional studies could also examine the long-term impacts of these educational strategies on students' career choices and success in STEM fields. Some limitations must be acknowledged in order to interpret the findings and thoroughly guide future research. One notable limitation is that the study's reliance on qualitative data from final presentations may introduce subjective biases in interpretation. Although qualitative methods provide deep insights into participants' experiences and perceptions, they depend heavily on the researcher's interpretation, which can be influenced by personal biases or theoretical inclinations. Moreover, the data captured represents a snapshot in time and does not account for longitudinal changes in students' attitudes or performances. This could provide a more comprehensive understanding of the impacts of culturally responsive computing over time. Moreover, the study's focus on culturally responsive computing narrows the scope of the research to specific pedagogical approaches, possibly overlooking other significant factors that influence learning outcomes in data science education, such as prior knowledge, personal interest in the subject matter, or other socioeconomic factors that might affect student engagement and success. These limitations suggest a need for further research involving larger, more diverse populations,

possibly incorporating mixed methods to triangulate data and validate findings. Longitudinal studies could provide insights into the lasting impacts of culturally responsive pedagogies in computational data science education. Moreover, future research could explore the interplay of culturally responsive computing with other pedagogical elements to paint a fuller picture of what effectively supports diverse learners in data science education.

## VI. ACKNOWLEDGMENT

This work was supported by a grant from the National Science Foundation(#2137708). Any opinions, findings, conclusions, and/or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation or the University of Texas at El Paso

## VII. REFERENCES

### REFERENCES

- [1] U.S. Bureau of Labor Statistics, "Employment projections," 2021, accessed: 2024-08-08. [Online]. Available: <https://www.bls.gov/>
- [2] National Science Foundation, "National center for science and engineering statistics (ncses)," 2022, accessed: 2024-08-08. [Online]. Available: <https://www.nsf.gov/statistics/>
- [3] E. Galasso, "Big data, causal inference, and "good data mining"?" World Bank Blogs, 2014, accessed: 2024-08-08. [Online]. Available: <https://blogs.worldbank.org/impactevaluations/big-data-causal-inference-and-good-data-mining>
- [4] G. Kelly, "Inquiry, activity and epistemic practice," in *Teaching scientific inquiry*. Brill, 2008, pp. 99–117.
- [5] S. Barzilai and C. A. Chinn, "On the goals of epistemic education: Promoting apt epistemic performance," *Journal of the Learning Sciences*, 2017.
- [6] D. Weintrop, E. Beheshti, M. Horn, K. Orton, K. Jona, L. Trouille, and U. Wilensky, "Defining computational thinking for mathematics and science classrooms," *Journal of science education and technology*, vol. 25, pp. 127–147, 2016.
- [7] S. Krishnamurthi, E. Schanzer, J. G. Politz, B. S. Lerner, K. Fisler, and S. Dooman, "Data science as a route to ai for middle-and high-school students," *arXiv preprint arXiv:2005.01794*, 2020.
- [8] K. A. Scott, K. M. Sheridan, and K. Clark, "Culturally responsive computing: A theory revisited," *Learning, media and technology*, vol. 40, no. 4, pp. 412–436, 2015.
- [9] A. Acquah, A. Barany, M. A. Johnson, A. S. Scarola, C. Rivera, and J. T. Walker, "Sandboxing in data science: An exploration of youth learning using open-inquiry approaches for computing-based data mining," in *Proceedings of the 18th International Conference of the Learning Sciences - ICLS 2024*, R. Lindgren, T. I. Asino, E. A. Kyza, C.-K. Looi, D. T. Keifert, and E. Suárez, Eds. International Society of the Learning Sciences, 2024, pp. 2433–2434.
- [10] A. Barany, S. Reza, M. Johnson, A. Barrera, O. Badreddin, C. Fuentes, and J. T. Walker, "Towards the design of a culturally relevant curriculum for equitable, data mining-based cs education," in *Proceedings of the 17th International Conference of the Learning Sciences-ICLS 2023*, pp. 1498–1501. International Society of the Learning Sciences, 2023.
- [11] J. T. Walker, A. Barany, A. Acquah, S. M. Reza, K. Guzman, M. Johnson, O. Badreddin, and A. Barrera, "Sandbox data science: Culturally relevant k-12 computing," 2024.
- [12] J. T. Walker, A. Barany, A. Acquah, S. M. Reza, A. Barrera, K. D. R. Guzman, and M. A. Johnson, "Coding like a data miner: A sandbox approach to computing-based data science for high school student learning," in *2023 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2023, pp. 1–5.



- [13] A. Barany, A. D. Scarola, A. Acquah, S. Reza, M. Johnson, and J. T. Walker, "Learning designs that empower: Navigating sandbox data science at the intersection of computing, big data, and social media," in *Learning with Social Media in an Algorithmic Age: Opportunities and Challenges for Education*, C. Greenhow, B. Staudt Willet, and J. Carpenter, Eds. Information and Learning Sciences, in press.
- [14] R. A. Estrellado, E. Freer, J. M. Rosenberg, and I. C. Velásquez, *Data science in education using R*. Routledge, 2020.
- [15] D. A. McFarland, S. Khanna, B. W. Domingue, and Z. A. Pardos, "Education data science: Past, present, future," *AERA Open*, vol. 7, p. 23328584211052055, 2021.
- [16] National Academies of Sciences, Engineering, and Medicine, *Data Science for Undergraduates: Opportunities and Options*. The National Academies Press, 2018. [Online]. Available: <https://doi.org/10.17226/25104>
- [17] S. Jiang, A. Nocera, C. Tatar, M. M. Yoder, J. Chao, K. Wiedemann, W. Finzer, and C. P. Rosé, "An empirical analysis of high school students' practices of modelling with unstructured data," *British Journal of Educational Technology*, vol. 53, no. 5, pp. 1114–1133, 2022.
- [18] R. Gould, S. Machado, T. Johnson, and J. Molynoux, "Introduction to data science v 5.0," 2018.
- [19] E. Gebre, "Conceptions and perspectives of data literacy in secondary education," *British Journal of Educational Technology*, vol. 53, no. 5, pp. 1080–1095, 2022.
- [20] T. Monroe-White, "Emancipatory data science: a liberatory framework for mitigating data harms and fostering social transformation," in *Proceedings of the 2021 on Computers and people research conference*, 2021, pp. 23–30.
- [21] V. R. Lee and V. Delaney, "Identifying the content, lesson structure, and data use within pre-collegiate data science curricula," *Journal of Science Education and Technology*, vol. 31, no. 1, pp. 81–98, 2022.
- [22] G. Gay, *Culturally responsive teaching: Theory, research, and practice*. teachers college press, 2018.
- [23] G. Ladson-Billings, "Toward a theory of culturally relevant pedagogy," *American educational research journal*, vol. 32, no. 3, pp. 465–491, 1995.
- [24] S. Hongladarom, "Cross-cultural epistemic practices," *Social epistemology*, vol. 16, no. 1, pp. 83–92, 2002.
- [25] Y. Mor and R. Abdu, "Responsive learning design: Epistemic fluency and generative pedagogical practices," *British journal of educational technology*, vol. 49, no. 6, pp. 1162–1173, 2018.
- [26] M. Johnson, A. Barany, A. Barrera, A. Acquah, S. M. Reza, and J. T. Walker, "Coding like a data miner: Reflections on epistemic plurality in sandbox data science with youth," 2024.
- [27] R. Stake, *Case study methods in educational research*. American Educational Research Association, 1980.
- [28] S. Merriam, "Case study research in education: A qualitative approach," 1988.
- [29] R. K. Yin, *Case study research: Design and methods*. sage, 2009, vol. 5.