Driving Social Impact harnessing Data: Addressing Social Problems through Public Interest Technology

1st Ridhi Jhamb
Lead Data Scientist
DataLab
Philadelphia, USA
ridhij@seas.upenn.edu

2nd Elena Eneva
Sr. Principal AI Research Scientist
Accenture
Bay Area, USA
elenaeneva@gmail.com

***Abstract*—As the adoption of Data Science and AI surges across various industries, the gap between the social sector and advanced data technology remains evident. To bridge this chasm, initiatives that leverage technological advancements for societal benefits are imperative. This paper delves deep into the transformative power of "tech for social good" illuminating the innovative data-driven solutions embraced by Sewanee DataLab, to address pressing social problems. The paper particularly explores the application of data-driven techniques across five pivotal projects, aiming to generate actionable insights and policy recommendations. These tools and techniques not only address immediate societal issues but also highlight the potential of technology as a force in amplifying social welfare and impact. In the healthcare domain, our focus centered on the uninsured, strategically analyzing high-need areas to amplify the reach and impact of medical care. Concurrently, in addressing food security, advanced predictive modeling enabled us to precisely forecast meal consumption, resulting in reduced waste and ensuring food availability in economically challenged regions. Meanwhile, our work in the housing sector involved a rigorous examination of allocation algorithms to eradicate biases, advocating for equitable housing distribution. The mental health landscape called for an in-depth exploration of college student data, unearthing patterns pivotal for crafting targeted interventions. Lastly, our propensity to serve the globally displaced - refugees, manifested in a comprehensive dashboard, streamlining resource allocation and bolstering advocacy efforts for global refugee funding. Each of these projects exemplifies the potent synergy of data science and domain-specific knowledge, underscoring the profound societal impact when technology is wielded judiciously and ethically.**

I. Introduction

Data Science for Social Good [1] is an emerging field that leverages cutting-edge technologies to devise innovative solutions that positively impact society. Through its focus on the greater good, the field aims to revolutionize how we approach social challenges, potentially leading to a more equitable and sustainable future. Using data science for social good diverges from its industrial counterpart in its primary objective and purpose. The concept revolves around identifying and solving social problems, such as healthcare access, climate change, food security, and public policy challenges. Unlike the industrial use, collaborations involve non-profit organizations, government agencies, humanitarian groups, and community partners, all with the shared goal of serving the public interest.

It is crucial to acknowledge the concept of Human-in-the-Loop (HITL) and Society-in-the-loop (SITL) systems when applying data science to social good initiatives. Social challenges are deeply complex, often rooted in nuanced ethical and cultural contexts that algorithms might overlook or misinterpret. By integrating human judgment with automated processes, HITL ensures that decisions are not only statistically accurate but also ethically and morally sound. This approach addresses potential biases, acknowledges data limitations, and fosters greater trust and acceptance among stakeholders, serving as a vital checkpoint to ensure that technological solutions genuinely benefit society without perpetuating existing disparities. Moreover, Data Science solutions for social problems should be designed with inclusivity and ethics, ensuring that they cater to diverse populations, including marginalized and underrepresented groups while being relevant and beneficial to the stakeholders.

Similar to data ethics, data privacy and security play a pivotal role when dealing with sensitive data, which is often the case while working on social problems. it's essential to ensure that the data is protected and used responsibly through the pipeline. Incorporating transparency, fairness and explainability in algorithm design not only builds trust but also allows those unfamiliar with the underlying technology (non-technical stakeholders) to scrutinize and suggest improvements. Hence, recognizing the importance of openness about how the technology works, and its decision-making processes to create a tool that is user-friendly and easily interpretable should be prioritized.

In a practical application of these principles, Sewanee DataLab emerges as a vibrant data science for social good program where we partner with government and nonprofits organizations, and leverage their data, to devise feasible solutions for their distinct problems. In 2023, we worked on five different problems, helping partners tackle challenges from the bottom up. With a diverse set of partner organizations, ranging from regional nonprofits to international public sector partners, the methods deployed to help them also vary. Further, identifying the sub-sector each organization belongs to becomes pivotal to allocate resources

and experts who guide these projects by providing domain knowledge and strategies.

The goals we defined for Sewanee DataLab are:

- Nurturing Future Data Scientists: We are dedicated to providing comprehensive training and guidance to aspiring data scientists hailing from diverse backgrounds. By offering a platform for skill development, hands-on experience, and exposure to real-world projects, we aim to equip these individuals with the expertise required to excel in the dynamic field of data science.
- Empowering Collaborative Partnerships: Another core facet of our mission is to empower our partners with the knowledge and tools needed to navigate the realm of data science successfully. Through collaborative efforts, we work side by side with our partners, offering insights, strategies, and best practices to approach and execute data science projects effectively.
- Driving Social Impact: Central to our mission is the aim to tackle significant social challenges. We are committed to leveraging data science as a catalyst for positive change, addressing pressing societal issues head-on. By employing data-driven approaches, we aspire to contribute meaningfully to the advancement of social good, making a meaningful difference in the lives of individuals and communities facing critical challenges.

As we continue to step towards achieving these goals, we ensure that we focus on the bull's eye, which is to use technology in the best interest of the people in need, directly or indirectly. This paper showcases five distinct social problems and their targeted data driven solutions designed by data scientists at DataLab. While these problems are unique and exhibit individuality in their essence, what proves intriguing is the web of connections and synergies that weave through them, rendering them remarkably relatable. It is within these overlaps and integrations that the true complexity of societal issues unfolds, presenting an opportunity for innovative resolutions that transcend isolated solutions and embrace a holistic perspective.

Our intention is to thoroughly explore and present these open-source projects, fostering a vibrant exchange of ideas within the field. By sharing insights into our methodologies, challenges, and outcomes, we aspire to facilitate a collective learning experience. This open discourse is integral to fostering innovation and enhancing the impact of our work. Through transparent discussions, we aim to inspire the development of new solutions that build upon the foundation laid by these projects. By engaging in collaboration, we can collectively propel the integration of data-driven approaches into various sectors, amplifying their positive influence on social challenges.

II. PROJECTS

A. Bridging the healthcare gap

In the US, 33 million people are medically uninsured and Tennessee contributes about a million of those people. This is about 12% of TN's population, which ranks the state in the bottom 18% of the entire country. Many areas are currently suffering from the lack of accessible health care due to the unavailability of health care coverage [2]. Partners for Healing (PH), a non-profit organization, has taken on the mission of aiding the uninsured by furnishing free healthcare services to local communities that experience medical under-service. Despite their clinic's capacity to accommodate more patients, its usage remains incomplete. The intent of this project was to thoroughly assess the counties they serve (Franklin, Moore, and Coffee), along with adjacent counties, to identify communities in high need of PH's services. This analytical approach will serve as the foundation for devising a coordinated action plan aimed at channeling expansion efforts, enabling the organization to efficiently target new patients.

- Data

The primary data sources utilized encompass data from the Tennessee Department of Health, the PH Clinic, US Census, and various external sources. The Tennessee Department of Health Data incorporates one year of both inpatient and outpatient data. Access to this sensitive information was secured through explicit permissions, and the team underwent training to ensure compliance. The PH Clinic Data encompasses patient encounter and referral details spanning the period from 2009 to 2023. For privacy and anonymity, patient history and social history datasets were omitted from the analysis. In the assessment of demographic trends within Tennessee counties, US Census Data estimations spanning 2018 to 2021 were employed. This analysis specifically focused on variables such as rates of uninsurance, employment, disability, federal poverty levels, and age distribution. Furthermore, external data sources such as maps and classification tables were integrated to augment the depth of the analysis.

- Methodology

The healthcare data being extremely sensitive, data security was a pressing concern with this project. The team went through multiple trainings to ensure safe and sensitive access to the data. The first phase involved cleaning, pre-processing and merging TN Health Data, US Census Data, and PH Clinic Data. This was followed by a series of aggregations to convert the data into a usable format for analysis. The US Census Data specifically required restructuring due to merging issues, and it was then grouped by specific categories, including Disability Status, Employment Status, Federal Poverty Level, Age Ranges, and Uninsured Status [3].

In the subsequent analysis stage, focus was directed towards extracting insights from each dataset. Concerning the US

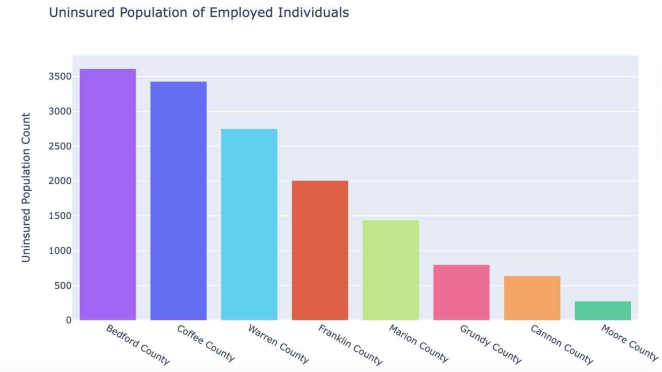

Fig. 1. Uninsured population in counties of operation

Census Data, interactive visualizations were generated across various categories, analyzing the uninsured population within each. Furthermore, the utilization of zip codes facilitated the creation of interactive geospatial heatmaps using python libraries like GeoPandas and Plotly, effectively evaluating the uninsured population and healthcare resources available to them. These heatmaps underscored specific locations needing targeted interventions. Within the analysis of PH Clinic Data, interactive graphs were formulated to uncover pivotal trends around clinic demographic statistics, patient conditions, ER overuse, and the clinic's temporal usage trends.

- Results & Recommendations

The end result materialized as an interactive dashboard comprising three primary tabs, each replete with multiple graphs and insights drawn from the data. One of the primary aspirations for the dashboard was for it to be a tool that the clinic would utilize to grow its patient population. This dashboard proved to be an invaluable asset to our partner, offering streamlined access to data-derived insights for grant applications and the presentation of evidence to the board of directors. The dashboard's maintenance demands remain relatively modest, and its capacity for effortless updates with fresh data ensures its sustained utility for future analyses. Ultimately, it was desired that entities such as PH would leverage the open-source code provided and harness this tool to address their distinct requirements. Given the uniformity of rural healthcare challenges spanning the nation, PH stands as a potential exemplar for other modest clinics in embracing data science as a vehicle for expansion.

B. Transforming the Path to Housing

Currently, there are over half a million individuals in the United States facing homelessness. In the year 2023, about 1500 people in the Southeastern Tennessee area alone were found to be facing homelessness by the Chattanooga Regional Homeless Coalition (CRHC). CRHC partnered with DataLab to optimize their housing assistance intake process so that their resources are allocated equitably amongst the unhoused individuals [4]. The objective of the project was to identify potential biases and their origin within the existing resource allocation tools and processes used by CRHC. The multifold solution involved employing Machine Learning models and statistical tests to identify and address these biases. Additionally, an in-depth analysis of the survey responses was conducted to gain insights into their distributions.

- Data

The data has been sourced from four primary channels: the Point-in-Time Survey, VI-SPDAT (Vulnerability Index - Service Prioritization Decision Assistance Tool), PVA (Place Value Assessment), and HMIS (Homeless Management Information System). The Point-in-Time Survey, conducted annually in January, involves manual collection of data concerning the homeless population within a Continuum of Care's geographic area, aimed at securing funding from the Department of Housing and Urban Development. The VI-SPDAT serves as a triage tool for evaluating and prioritizing the needs and vulnerabilities of individuals experiencing homelessness, spanning datasets from 2018 to 2023. PVA, another triage tool, assesses homelessness across five categories represented by a five-digit score, with a dataset encompassing three months of records from November 2022 onward. HMIS functions as a local IT system, compiling client-level and housing data pertaining to homeless individuals and families.

- Methodology

In order to enable CRHC to identify biases in their most recent housing assistance resource allocation process (PVA), a dashboard inspired by Conley's 2020 study [5] that revealed biases in VI-SPDAT was provided to them. Given PVA dataset's limited size and the absence of outcome data, automation code was employed to dynamically update the dataset as new data is added, generating essential visualizations.

Conley's methods included: A bivariate t-test: The independent t-test is used to test for a significant difference in means between two independent groups or conditions. It compares the means of a continuous outcome variable for two groups. Chi-square test: The chi-square test determines if there is a significant association between two categorical variables. It compares the observed frequencies in a contingency table with the expected frequencies under the assumption of independence between the variables.

Following a meticulous pipeline involving data cleaning, pre-processing, and de-duplication, the data underwent thorough exploration and analysis. Variables constituted questions from the homeless assistance survey that served as information for CRHC to calculate a score for each individual, which eventually helps in prioritizing them. Upon scrutinizing patterns and interdependencies among features, statistical tests were conducted to assess correlations and dependencies amongst bias related features such as race, gender, ethnicity in relation to the final scores assigned to individuals. Additionally, a logic to evaluate variable relationships and their impact on the VI-SPDAT score was developed using feature importance and

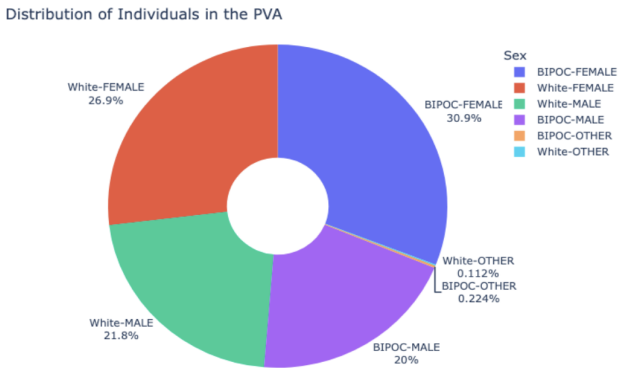

Fig. 2. Race-Gender breakdown for population catered by PVA

feature correlation information given by multiple predictive models.

The evaluative report delved deeper into correlations involving subscores, scores, race, and gender using the methodologies mentioned earlier. In this comprehensive report, detailed explanations, well-researched recommendations, and tailored insights were provided to address CRHC's concerns. Dedicated sections were assigned to each statistically significant correlation, outlining findings, methods, and research-based recommendations. The structured approach aimed to effectively address CRHC's inquiries and furnish them with valuable insights into resource allocation.

- Results & Recommendations

Among various discoveries, one particularly enlightening revelation was that within demographic groups exhibiting statistically significant population sizes, the central tendency of scores was observed to be the lowest for BIPOC Females. This observation suggested that they generally received lower scores compared to other demographic segments, resulting in a lower standing on the priority list for accessing housing and resources. All these insights concerning the origin of bias in the PVA were enclosed in the evaluative report and the dashboard provided to CRHC. The project played a pivotal role in aiding CRHC's recognition of biases and vulnerabilities related to equity within the PVA. Consequently, this realization significantly influenced their decision to transition to an alternative algorithm.

C. Improving Student Well-being

The global increase in personal and interpersonal challenges faced by students eventually leading to mental health related struggles is concerning. The University of the South embarked on this project in collaboration with DataLab to address the growing concerns regarding the escalation of personal and interpersonal challenges experienced by college students, often culminating in mental health-related adversities. The objective was to comprehend these challenges through a study driven by data, enabling the implementation of well-informed interventions [6]. An analysis of the annual student survey data was undertaken to uncover patterns and avenues for impactful outreach, service utilization, and the augmentation of student well-being. This analysis particularly focused on areas such as social isolation, loneliness, suicidality, substance abuse, and sexual abuse.

- Data

The dataset was derived from the Healthy Minds Survey, an online survey conducted on post-secondary students. It provides insights into mental health outcomes, knowledge, attitudes, and service utilization of its takers. The focus was directed towards data spanning the years 2017 to 2022, with the omission of 2021 to mitigate anomalies arising from the pandemic. The dataset encompassed details concerning student demographics, their health, behaviors, and lifestyle mapped to mental health outcomes in terms of service utilization [7]. A major challenge was ensuring a trauma informed methodology and data security at all points, considering the sensitivity of the data.

- Methodology

The approach aimed to provide valuable insights on student well-being, with a focus on specific challenges such as loneliness, sexual assault, and substance abuse. To achieve this, a data-driven methodology was adopted, with a particular focus on analysis of trends, interpretable data visualizations, and creation of two dashboards [7][8]. To the raw survey data, data pre-processing techniques such as data cleaning, imputation, and outlier detection were applied to ensure data's quality and reliability. This step also involved encoding categorical variables and normalizing numerical data. This was followed by exploratory data analysis, which involved generating summary statistics, identifying patterns, and exploring potential relationships between variables. Further, visualizations played a crucial role in making the data more interpretable by representing trends, patterns, and key insights. Visualizations helped in identifying trends related to loneliness, sexual assault, and substance abuse, among other factors affecting student well-being.

The user-friendly dashboards were thoughtfully designed to effectively communicate the gathered insights, ensuring accessibility for non-experts in data analysis, including our partners. The two dashboards created as part of this methodology serve distinct purposes. The public dashboard is designed to be accessible and comprehensible to a wide audience, including university administrators and stakeholders. On the other hand, the private dashboard is accessible only to our partners, enhancing their capacity to advocate for resources and support from within the institution. Given the limitations of small sample sizes and the voluntary nature of the survey, the insights gleaned from the data proved to be relevant, enabling our partners to identify important aspects of student well-being that warrant attention and intervention.

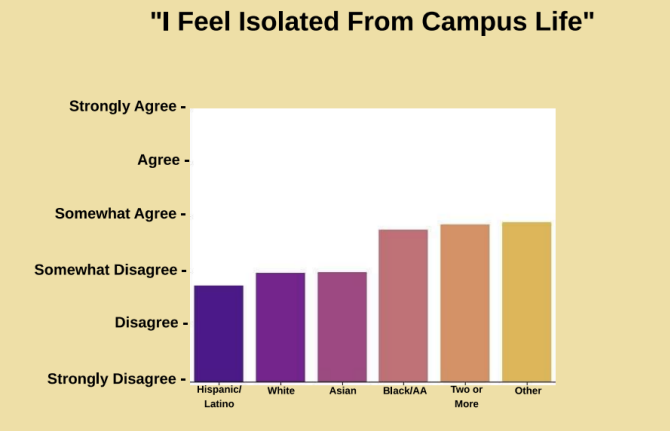

Fig. 3. Student mental health survey outcomes

- Results & Recommendations

The insights gained from the provided tools can be employed by the University's wellness center to enhance their pre-existing outreach and support resources framework for students. This, in turn, will enable them to refine their outreach and preventive initiatives more efficiently. Furthermore, the dashboards will equip the partner to advocate for supplementary financial backing, thereby amplifying the resources already accessible to Sewanee students.

D. Decoding Displacement

In recent years, the scale of the global refugee crisis has expanded rapidly [9][10]. Millions have been displaced due to violence, economic instability, and the ripple effects of climate change. To advocate for much-needed refugee funding, we partnered with the USA for UNHCR to compare data on regional refugee movement in recent years with the funding needs of nations in crisis outlined in UNHCR's annual Global Appeals Report. Using this data, we built the Displacement Crisis Dashboard, an interactive tool displaying visualizations and insights, enabling the USA for UNHCR to right-size supply and optimize resource allocation based on refugee demographics and global appeals.

- Data

The data was obtained from the United Nations High Commissioner for Refugees (UNHCR) and includes both static and dynamic data on forcibly displaced people. The Refugee Data Finder tool on UNHCR's website was used to sort data based on refugee demographics and status, but since it cannot display age/gender and refugee status simultaneously, separate datasets were downloademerged. The dataset is primarily disaggregated by age, gender, and country of origin/country of asylum. The second important data source was the budgeting data from UNHCR's annual Global Appeal Report, which provides financial information for refugee aid worldwide, including requested budget by countries and the expenditure [11]. The Global Appeal data is organized by geographic categories and further broken down into impact and outcome areas that range from short-term emergency aid to long-term structural solutions.

- Methodology

As the initial phase, two datasets retrieved from the UNHCR Refugee Data Finder tool were merged. The first dataset encompassed geographical details regarding classifications of displaced individuals, while the second furnished demographics of these displaced individuals. Alongside this, budgetary data for 2022 was aggregated, outlining the needs-based budget specified in the Global Appeal, stitched with the real expenditure for each country. Following the processing, cleaning, and normalization of the datasets, preliminary visualizations were generated utilizing Plotly, a Python library known for its user-friendly visualization capabilities. For interactive maps, Folium was employed, linking geospatial data to color displays based on distinct dataset features. The analysis and visualizations addressed crucial inquiries concerning refugee movement and the funding requirements of various countries.

The ultimate dashboard offered valuable insights into refugee categorization, demographics, and budget allocation. It showcased interactive global maps, highlighting the total displaced population and budget allocation details. Furthermore, it provided an in-depth demographic analysis based on factors like country of asylum, age, gender, population, and location type. The funding section visualized budget allocation, actual expenditure, and the gaps in funding for each country operating under the UNHCR's mandate.

Through a series of extensive engagements with UNHCR executives, the project established a vital dialogue aimed at addressing the persistent funding gaps within the organization. These interactions involved in-depth discussions, collaborative brainstorming sessions, and meticulous research into the existing systems and strategies employed by the UNHCR to bridge the funding gap. Drawing upon the insights from these discussions, a set of well-defined target metrics and statistics were defined to be displayed on the dashboard. These metrics were curated to not only address the immediate funding challenges but also to provide a comprehensive overview of the organization's financial landscape at a visual level. The dashboard successfully met the goal of promoting right-sizing and efficient resource allocation for USA for UNHCR.

- Results & Recommendations

As the count of forcibly displaced individuals surged at a rapid pace, UNHCR found itself needing more resources to cope with the escalating numbers seeking refuge. The interactive maps and graphs within our dashboards proved highly effective for UNHCR to gain a contextual understanding of the funding prerequisites for each region. These tools provided enhanced intelligence in resource allocation to address the needs of the displaced. The dashboard included interactive global maps, which highlighted the total displaced population and provided details on budget allocation. It also featured demographic

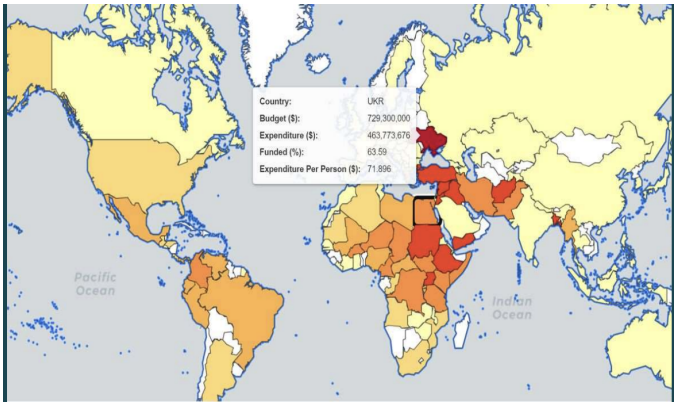

Fig. 4. Underfunding by region

analysis based on factors such as the country of asylum, age, gender, population, and location type. The funding section of the dashboard visualized budget allocation, actual expenditure, and funding gaps for each country operating under the UNHCR's mandate.

This project stood as a pioneering effort, unraveling trends through the harmonization of Refugee demographics and Global Appeal datasets. Our work yielded invaluable insights into the funding dynamics of various regions under UNHCR's purview. Our partners can extend the capabilities of these dashboards by integrating additional data, further augmenting their potential for comprehensive analysis.

E. Feeding Futures

As per USDA, almost 33 percent of the population of the IFSA countries was considered food insecure in 2022 [12], highlighting the significance of initiatives like the South Cumberland Summer Meal Program (SCSMP), which aims to combat hunger in Middle Tennessee. Unavailability of a consistently reliable meal prediction system posed challenges for the program resulting in either under-service or over-service, with a substantial 20-50% of meals being wasted. These wasted meals do not receive reimbursement, adversely affecting the program's budget for future operations. To address this problem, we applied machine learning techniques on historical meals data to develop an adaptable forecasting tool to reduce food waste and program costs by accurately predicting meal consumption in economically distressed areas.

- Data

Daily Meal Count Form Data (Summer Meal Program Data): 8 years of daily meal count records collected through a Google Form, providing information on meal types, delivery time, number of meals prepared, meals served to eligible and non-eligible participants, damaged meals, and leftovers. Schedule Data: Dataset included schedules of meal sites' operating days and times over the 8-year period. It is crucial for understanding the timing of meal services and integrating it with other datasets. Menu Data: Dataset included menus used by the program during the 8-year period, with variations over time. Integrating it with other datasets helped assess the impact of menu changes on meal consumption. ncluded addresses of meal sites operating within the program during the 8 years, essential for clustering and understanding the geographic distribution of meal services. Weather Data: Real-time weather information around the South Cumberland Plateau was used to assess its impact on meal consumption and participant attendance. Event Data: Dates and names of national and local events occurring during the program's operation months (June and July) helped us understand the influence of community events' on meal service planning.

- Methodology

Utilizing seven years of historical data on meal site operations provided by our partners, trends were analyzed, new features were synthesized and machine learning methods were applied to develop a model capable of forecasting meal demand in various regions with a two-week anticipation. As the first step, data acquisition was completed after integrating data from multiple data sources listed above. The underlying methodology involved data cleaning, with two main phases, manual and automated. In the manual phase, data sanity checks were done using pivot tables and visual validation. Subsequent data cleaning was automated using Python, involving data validation, handling missing data, and removal of redundancies and duplication. Data cleaning pipelines were developed to function for two different kinds of meal count datasets, congregate and non-congregate in fashion. Further, transformation of data to align with the forecasting frequency was done. After the completion of data cleaning, an exploration of visualizations was undertaken by the team to comprehend the relationships and patterns present among variables for feature engineering and to provide insights to the partners. Subsequently, new features were derived from the existing dataset, prioritizing those that demonstrated increased utility for enhancing the accuracy of predictions.

For the purpose of predictive modeling, the XGBoost Regressor (XGBR) was utilized to forecast the range of people expected to arrive at meal site regions two weeks in advance. Separate predictions were made for congregate and non-congregate data, with error calculations adjusting for underserved predictions, which were finally ensemble together. Fetching dynamic variables for the unseen data, for instance weather information two weeks in advance was automated using API calls. The explainable model factored in variables that influence predictions and weighed the important variables highly while calculating the cost function. Predictive outcomes were also generated via the Random Forest bagging technique, which failed to cater to high bias in the data as efficiently as boosting. The model was integrated with the dashboard, enabling personalized predictions through user input, revealing the significance of features, percentage distribution, weather particulars, and the range of predictions for the next two weeks. The aspect of model retraining was introduced, allowing partners to update data and enhance accuracy by

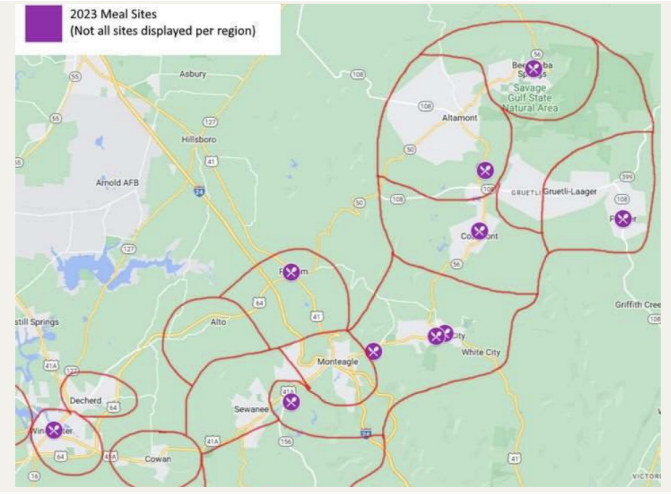

Fig. 5. Meal site clusters

integrating new information as and when it is available.

Employing machine learning methods on raw meals data received from each meal site along with external data sources like weather, schedules and events made it a state-of-the-art approach to reduce food wastage and ensure food security. Moreover, South Cumberland Plateau AmeriCorps VISTA Project being a small scale project, posed significant hurdles, particularly in cleaning and validating the data. The raw data available with our partners were physical copies of forms filled out at meal sites daily. The process of digitizing the data introduced potential human errors, missing and duplicate data, making data cleaning and validation arduous and time-consuming. Further, the lack of standardized working schedules and menus for each operation site, which were based on human hunches and estimations, became a challenge for our model's accuracy due to the absence of contextual information. Nevertheless, overcoming these challenges yielded substantial value when this became one of our most impactful works where we not only helped our partner by developing a valuable tool to solve their specified problem but also pioneered the digital operations for them by establishing their very first clean dataset and technology pipelines. Additionally, we remained dedicated to accommodating our partner's limited technical expertise, ensuring that the tool we created was user-friendly and easily interpretable, even for individuals unfamiliar with the underlying technology.

- Results & Recommendations

The findings led to the creation of a dashboard designed for the partner, showcasing essential visual insights and offering a platform for inputting data and in turn receiving predictions from the model. In addition to this private dashboard, a public dashboard was created intended for the community, supplying resources and information highlighting the locations of meal and food bank sites.

The tools provided empowered the SCSMP to enhance and modernize its prevailing meal distribution strategy through data-informed decision-making. By leveraging the precision of these predictions, SCSMP could effectively expand its services, thus achieving enhanced operational efficiency. This, in turn, enables them to increase the range of their impact to a broader spectrum of individuals, consequently contributing significantly to the reduction of food insecurity. Beyond the direct influence on SCSMP's operations, the objective is to instigate action against food insecurity by illustrating data-driven approaches that can be implemented. The overarching aim is to impact communities by empowering other organizations like SCSMP to make data-informed choices based on our discoveries.

III. CONCLUSION

The fusion of data technology and social impact has emerged as a potent force for addressing critical societal issues. Within the rapidly evolving realm of Data Science, there is a need for a distinctive shift towards utilizing these tools for addressing pressing societal challenges as opposed to mere industrial applications. Innovative technologies, when combined with collaborative efforts between nonprofits, governments, and communities, have the potential to drive groundbreaking and equitable solutions. From challenges related to well-being, housing assistance, refugee aid, public healthcare, to food security, data technology reveals the interconnectedness of these issues and paves the way for comprehensive solutions. Interactive visualizations serve as powerful tools, providing real-time insights and representations of vast amounts of data. Predictive modeling, on the other hand, allows stakeholders to anticipate future trends and challenges, enabling proactive measures rather than reactive responses. Central to the effectiveness of these technological solutions in the social sector is their emphasis on user-centric designs. Recognizing that stakeholders come from varied backgrounds with differing levels of technical expertise, these tools need to be both accessible and intuitive. This inclusivity ensures that a wider audience can benefit from and contribute to the data-driven discourse. But the impact of these initiatives extends beyond the immediate, tangible results they produce. They play a pivotal role in nurturing a culture that values innovation and collaboration. As industries and communities harness the power of data, they are not only addressing the challenges of today but are also laying the groundwork for a more collaborative, innovative and sustainable future.

REFERENCES

[1] Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., ... Clopath, C. (2020). AI for social good: unlocking the opportunity for positive impact. Nature Communications, 11(1), 2468.
[2] Ayanian, J. Z., Weissman, J. S., Schneider, E. C., Ginsburg, J. A., Zaslavsky, A. M. (2000). Unmet health needs of uninsured adults in the United States. Jama, 284(16), 2061-2069.
[3] Walker, K. (2023). Analyzing us census data: Methods, maps, and models in R. CRC Press.
[4] Rita, N., Garboden, P. M., Darrah-Okike, J. (2023). "You Have to Prove that You're Homeless": Vulnerability and Gatekeeping in Public Housing Prioritization Policies. City Community, 22(2), 83-104.
[5] Cronley, C. (2022). Invisible intersectionality in measuring vulnerability among individuals experiencing homelessness–critically appraising the VI-SPDAT. Journal of Social Distress and Homelessness, 31(1), 23-33.

[6] Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H. C., Jeste, D. V. (2019). Artificial intelligence for mental health and mental illnesses: an overview. Current psychiatry reports, 21, 1-18.

[7] Lipson, S. K., Eisenberg, D. (2018). Mental health and academic attitudes and expectations in university populations: results from the healthy minds study. Journal of Mental Health, 27(3), 205-213.

[8] Narita, Z., Banawa, R., Zhou, S., DeVylder, J., Koyanagi, A., Oh, H. (2022). Loneliness and psychotic experiences among US university students: Findings from the Healthy Minds Study 2020. Psychiatry research, 308, 114362.

[9] Kang, Y. D. (2021). Refugee crisis in Europe: determinants of asylum seeking in European countries from 2008–2014. Journal of European Integration, 43(1), 33-48.

[10] Ramasubramanian, S., Miles, C. (2018). Framing the Syrian refugee crisis: A comparative analysis of Arabic and English news sources. International Journal of Communication, 12, 19.

[11] Smrkolj, M. (2009). International institutions and individualized decision-making: An example of UNHCR's refugee status determination. In The Exercise of Public Authority by International Institutions: Advancing International Institutional Law (pp. 165-193). Berlin, Heidelberg: Springer Berlin Heidelberg.

[12] Baquedano, F. G., Zereyesus, Y. A., Valdes, C., Ajewole, K. (2021). International Food Security Assessment 2021-31 (No. 1962-2021-2203).