

Advancing the Field of Data Science for Inclusive Growth

Learnings from the data.org Inclusive
Growth & Recovery Challenge

DataKind[®]

data.org



Center for
Inclusive Growth



The
ROCKEFELLER
FOUNDATION

Contents

Welcome Letter	4
Executive Summary	6
Introduction	8
Context for the Challenge	9
The data.org Challenge	10
Overview of Challenge Applicants	11
Analysis of Challenge Applications	15
Critical Topics for Further Inquiry	17
The Challenge Review Process	24
Awardees	26
Community Lattice	28
GiveDirectly and Center for Effective Global Action	28
University of Chicago, Center for Data and Computing	29
BUILD Aalborg University	29
Basel Agency for Sustainable Energy	30
Solar Sister	30
Women's World Banking	31
Fundación Capital	31
Conclusion	32
Appendix	33

Advancing the Field of Data Science for
Inclusive Growth

Learnings from the data.org Inclusive Growth
& Recovery Challenge

© DataKind® 2021

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without permission in writing from the publisher.

Welcome Letter

In early 2020, as part of its commitment to building the field of data science for social impact, **data.org** launched the \$10 Million data.org Inclusive Growth and Recovery Challenge in partnership with the **Mastercard Center for Inclusive Growth** and the **The Rockefeller Foundation**.

The goal of the Challenge was to find exceptional ideas from around the globe that use data science to encourage inclusive growth and economic recovery, especially in the wake of COVID-19. **DataKind**, a leader in the data science for social impact space that harnesses the power of data science and AI in the service of humanity, was selected as the Technical Partner for the Challenge.

We are so thankful for each of the Challenge applicants who proposed such inspiring, impactful, and data science-driven projects. The incredible array of applicants have bolstered our confidence in the landscape and future of inclusive growth and data science for social impact.

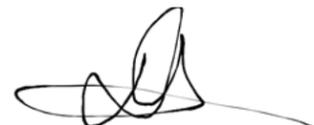
We also want to thank the hundreds of volunteer judges and our panel of Expert Judges who supported our efforts to identify breakthrough applications of data science to the field of inclusive growth and recovery. Their generosity, thoughtfulness, and commitment to this project empowered our teams with critical perspectives to identify our incredible group of awardees.

Finally, we are grateful to the Mastercard Center for Inclusive Growth and The Rockefeller Foundation, both of which are data.org's founding partners and major funders of DataKind. Without their support, expertise, and leadership, this Challenge wouldn't have been possible.

We hope that this report will empower readers with actionable insights and learnings.



Afua Bruce
Chief Program Officer
DataKind



Danil Mikhailov
Executive Director
data.org



About data.org

data.org is a platform for partnerships to build the field of data science for social impact. Launched by the Mastercard Center for Inclusive Growth and The Rockefeller Foundation, data.org works with organizations from all over the world to increase the use of data science to improve the lives of millions of people and to tackle society's most significant challenges.

About DataKind

DataKind is a global nonprofit that harnesses the power of data science and AI in the service of humanity. Named one of *Fast Company's* Top 10 Innovative Nonprofits, DataKind helps mission-driven organizations unlock their data science potential ethically and responsibly. DataKind teams talented pro bono experts with visionary social actors to collaboratively design innovative solutions to tough social challenges. Launched in 2012, DataKind is headquartered in New York City and has Chapters in Bangalore, Singapore, United Kingdom, San Francisco, and Washington DC.

About the Mastercard Center for Inclusive Growth

The Center for Inclusive Growth advances equitable and sustainable economic growth and financial inclusion around the world. The Center leverages the company's core assets and competencies, including data insights, expertise and technology, while administering the company's philanthropy, the Mastercard Impact Fund, to produce independent research, scale global programs, and empower a community of thinkers, leaders, and doers on the frontlines of inclusive growth.

About The Rockefeller Foundation

The Rockefeller Foundation advances new frontiers of science, data, and innovation to solve global challenges related to health, food, power, and economic mobility. As a science-driven philanthropy focused on building collaborative relationships with partners and grantees, The Rockefeller Foundation seeks to inspire and foster large-scale human impact that promotes the well-being of humanity throughout the world by identifying and accelerating breakthrough solutions, ideas, and conversations.

Executive Summary

This report shares learnings from the data.org Inclusive Growth and Recovery Challenge and highlights the incredible work of our inspiring group of Challenge awardees. The **data.org** \$10 Million Inclusive Growth and Recovery Challenge was launched in 2020 in partnership with the **Mastercard Center for Inclusive Growth** and **The Rockefeller Foundation**. The Challenge sought to find exceptional ideas from around the globe that apply data science to encourage inclusive growth and economic recovery. Following data.org's flagship investment in 2019, **DataKind** joined the Challenge as the Technical Partner and advisors bringing years of experience and a global network of data scientists for social good.

More than 1,260 individuals and organizations submitted applications to the Challenge across the themes of Cities and Towns, Access to Capital, and Jobs of Tomorrow, or proposed a different theme using the open track.

We received a wide range of proposals at different stages of development, from a budding idea to an established methodology seeking international replication, and from a diverse group of applicants—researchers, entrepreneurs, doctors, government representatives—from newcomers to household names. They submitted applications addressing many important topics, such as reimagining credit scoring, providing better farming knowledge to services for youth employment and gig workers, and preparing urban landscapes for extreme weather. The data.org team and partners, including the hundreds of volunteer judges, were and remain humbled by the outpouring of interest and ingenuity in finding creative ways to grow a more prosperous society and use our most cutting-edge tools and technology in the service of others.

Following a complex review process, including thousands of volunteer hours and hundreds of partner reviews, the Challenge identified eight exemplary awardees with projects that excel in the goal of applying data science for inclusive growth and recovery. Over the next two years, data.org and partners will work with awardees and many of the Challenge applicants.

Our eight Challenge awardees show the range of opportunities that exist to use data science to drive social impact for workers, entrepreneurs, and communities.

Using data mining and machine learning, Fundação Capital will arm informal workers in Mozambique with essential labor-market insights to increase income and employment opportunities. Basel Agency for Sustainable Energy will use machine learning and physics-based food modeling to enable smallholder farmers in India to access sustainable cooling facilities, thereby reducing food loss and dramatically improving livelihoods. In Nigeria and Tanzania, Solar Sister will share market insights with its network of women entrepreneurs to grow their renewable energy businesses. Addressing gender bias in lending algorithms, Women's World Banking will work with financial service providers around the world to increase credit access for low-income female entrepreneurs. Using a rich, longitudinal population dataset, Aalborg University will create interactive and actionable maps for policy-makers and urban planners that identify the geographical places in Denmark that are most vulnerable to out-migration and economic instability. Community Lattice will create a platform to predict the cost and risk of brownfield redevelopment projects in the United States to transform a community's ability to secure redevelopment funding, improve community health, and create economic opportunities. GiveDirectly and the Center for Effective Global Action will develop and implement a new model for humanitarian and development aid that enables cash transfers to be deployed effectively, accurately, and at scale to those who need them most. Lastly, to address the

digital divide, the University of Chicago will create open-source maps and toolkits that can be used to highlight inequities in broadband access and advocate for more equitable policies and investments.

Around the world, mission-driven individuals and organizations of all types, sizes, and maturity that seek to transform communities and realize lasting social change have recognized the potential of data science to achieve their goals. The Challenge clearly highlighted the high and growing degree of demand for data-driven projects in service of social missions.

Despite this global interest, there are still substantial gaps in the capacity of organizations to use data science for social impact. Social sector actors and organizations of all sizes urgently need technical support, human resources, and funding to support their capacity to reach their goals. The Challenge strengthened our conviction that there are substantial opportunities for continued growth in this field. However, there is also a clear need to support and empower organizations with the data science capacity, critical partnerships, communities of practice, and shared resources needed to realize the potential of data science for social impact.

The applications demonstrated the benefit and necessity of partnerships: some organizations are strong on the technical side but don't have social impact expertise, while others have strong social impact experience but miss the data science component. Together and translating across these different competencies, organizations can create stronger projects and have more lasting impact. Some of the strongest proposals relied exactly on this capability of using partnerships to expand and advance organizational capabilities. However, one size does not fit all, and different types of partnerships are necessary for different organizations at different stages of development. We want to support organizations holistically, not just by giving them monetary support but also including them in an environment in which they can create partnerships and receive technical support, feedback, and expert advice.

For this reason, the Challenge has sought to both support its awardees financially, as well as to provide them with the necessary technical expertise to implement their projects and achieve impact. Each award has been tailored to the specific project, organization, and partnership to ensure the maximum level of support possible.

We had an overwhelming response from private sector partners about how they could help and what they wanted to do. In the coming months, we will work with them to ensure that these collaborations will be transformative.

Applicants rightly recognized that inclusive growth and recovery touches every aspect of how individuals and communities are evolving and how data-driven and data science solutions can be employed in any of these sectors. Inclusive growth is not sector-specific, and the breadth of applications shared a diversity of solutions across areas across the social sector.

Going forward, we commit to strengthen this finding about the importance of partnerships and to work with the awardees to develop plans and grow their capacity in partnerships, training, talent, data, and shared learning. We believe these will be vital investments for any organization—funder or implementer—seeking to grow data science for social impact in a specific sector or across the field.

To transform this field, we must collaborate across organizations and sectors.

In order to fill the gap between demand and supply, the amount of investment required needs to increase exponentially—which will not happen through funding alone. Rather, we need diverse, dynamic, collaborative partnerships to meet the demand for data science for social impact.



Introduction

Pick any problem you see around the world: the raging wildfires that have devastated hectares of forests and homes; the drug epidemics that have ravaged poor communities; the world's worst refugee crisis since World War II. Confronting these problems and others—from poverty and inequality to conservation and climate change—requires the ethical application of insights drawn from timely and responsibly sourced data.

In early 2020, with the support of data.org's original founders, the Mastercard Center for Inclusive Growth and The Rockefeller Foundation, data.org—originally launched 20 years ago as a project that marshaled \$100 billion in debt forgiveness for Africa and another \$50 billion in contributions for health and development, and is now known as the ONE Campaign—relaunched to serve as a platform for partnerships to expand further the field of data science for social impact and to ensure that nonprofit and civic organizations are well-positioned to take advantage of the data revolution.

As its first major initiative, in May 2020, data.org issued an open call for breakthrough ideas that harness the power of data science to help people and communities rebound and remain resilient in the wake of COVID-19 and its economic impact. The \$10 Million Inclusive Growth and Recovery Challenge sought to foster and promote innovative, scalable solutions from people and organizations all over the world.

In the implementation of the Challenge, data.org worked closely with DataKind, a leader in the data science for social impact space that harnesses the power of data science and AI in the service of humanity. As the data.org technical partner, DataKind provided instructional webinars for potential applicants, spearheaded the recruitment and training of hundreds of expert volunteer judges from all over the world, and provided expert guidance on the quality of data science incorporated in each project. After an initial round of reviews, the best applications were reviewed by a panel of Expert Judges who helped to select the eight awardees, announced in January 2021. Throughout the two years of the Challenge funding and support, each awardee will be supported by a bevy of technical assistance partners to grow their project and organizational capacity to advance inclusive growth and recovery and in doing so build the field of data science for social impact.



Context for the Challenge

data.org was inspired to launch the Inclusive Growth and Recovery Challenge because of the transformative potential of data science to promote financial security and economic mobility for workers, help micro and small businesses access capital, and support economic development. A focus on inclusive growth was essential, given the need for economic development that benefits every individual and sector across societies. As defined by the **Mastercard Center of Inclusive Growth**, “Inclusive growth ensures the benefits of a growing economy extend to all segments of society. Unleashing people's economic potential starts with connecting them to the vital networks that power the modern economy.” Initially launched in January 2020 as the Inclusive Growth and Recovery Challenge, in recognition of the devastating effects of COVID-19 on life, livelihoods, prospects, and prosperity, the team updated the challenge to seek new and promising ways to use data science to drive inclusive growth and recovery in the wake of COVID-19.

The COVID-19 crisis has underscored the extreme vulnerability of individuals and communities to economic shocks. According to the **United Nations Sustainable Development Goals Report 2020**, global poverty is on track to rise for the first time since 1998, erasing decades of progress. As a result of the pandemic and the economic crises it has caused, the UN estimates that

71 million people were forced back into extreme poverty in 2020. In every country across the world, workers, entrepreneurs, small businesses, and the communities in which they operate are at significantly greater risk of financial instability.

The pandemic has also demonstrated the important role data science can play in solving some of the world's most pressing challenges. As private, civic, and social sector organizations tackle the health and economic implications of the pandemic, data science can inform the development of tools and insights that can be used in addressing the needs of all segments of society, especially those most vulnerable. Even outside of the COVID-19 pandemic, the need to incorporate data science to drive inclusive growth is clear, since inequality and the gap between rich and poor continues to rise and economic gains have not been distributed equitably.

Through the Challenge, data.org sought to address a systemic issue: the majority of social initiatives don't have the budget, staff, capacity, or partnerships to take full advantage of our current data revolution. With support, mission-driven organizations will be able to use data science tools and methods to make their work go further and faster, helping more people.

The data.org Challenge

The Challenge sought to tap into a broad base of expertise to catalyze innovative and scalable solutions to help individuals and communities recover and thrive, all the while building resilience to withstand future challenges.

At stake for up to ten awardees was data science talent, software, training, and funding from \$10K to \$10M USD in a mix of grant-funding, technical support and consulting, media production, marketing and promotional outreach, and software and infrastructure licenses.

Applications came from individuals and teams with a diversity of experience with data science, ranging from teams of data scientists who sought to advance their work in social impact, to those with a track record in impact seeking to integrate data science to address social issues. While data.org encouraged and accepted applications from nonprofits, for-profits, government and UN agencies, as well as individuals, all applications were required to have a charitable purpose and accrue no private benefit.

The Challenge primarily sought proposals addressing fundamental questions in the following areas:

Cities and Towns

An economic recession will hit some places harder than others. As public and private sector investments flow into cities and towns to stimulate the economy, how can leaders use data-driven insights to make the right decisions to ensure economic security in underserved communities? How can data and analytics help connect neighborhoods to the resources and networks they need to access opportunities, including quality education, affordable housing and childcare, decent jobs, and transportation?

Access to Capital

Unleashing the spirit of grassroots entrepreneurs can help communities rebound by creating much-needed jobs and growing local economies. Yet, in times of recession, the flow of credit to micro and small businesses tends to slow down. How can data science enable microentrepreneurs to gain access to capital? How can data science help us identify micro and small businesses with the potential to grow? How can data insights rethink credit-worthiness and unlock capital for high potential business owners?



Jobs of Tomorrow

Low wage workers are among the most vulnerable in today's economy. In times of economic stress, current trends toward automation and job displacement could accelerate. How can organizations use data science to help workers remain secure in precarious times? Can data insights predict trends in the labor market and connect and prepare workers for the jobs of tomorrow?

Applicants were asked to specify whether their solution was at the stage of an idea, pilot, scaling, or replication and to identify what sector they would work in, the geographic scale of the project, and which geographic region would be their primary focus.¹

During the two-month application period, the Challenge received over 1,260 applications from individuals and entities across the world. The following sections highlight trends and insights from the Challenge applicants.

¹ For the complete application form from Phase 1 and Phase 2, see Appendix.

Overview of Challenge Applicants

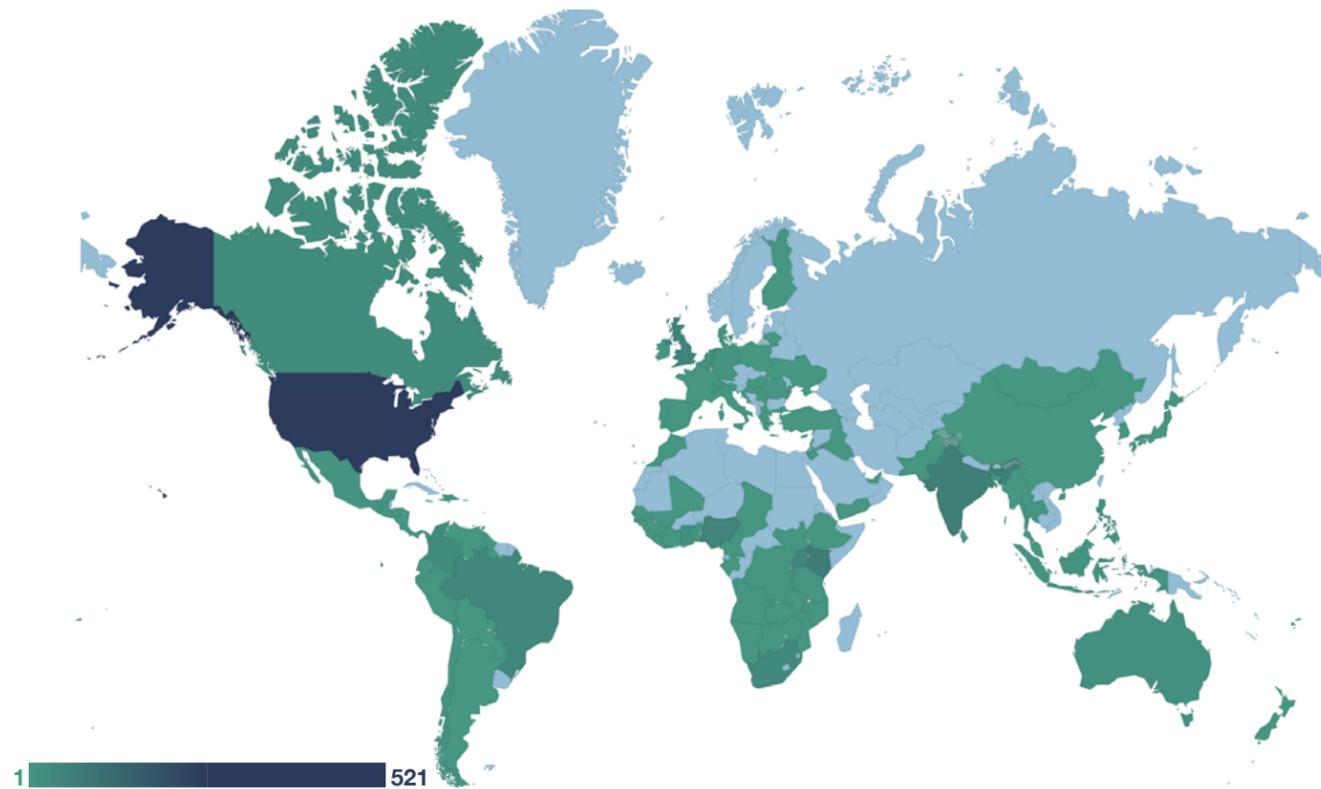
The data.org Inclusive Growth and Recovery Challenge was an international call for ideas that used data science to advance shared prosperity and help ensure an inclusive recovery. A total of 1,263 applicants submitted proposals to the data.org Challenge.

Applicant Country & Continent

The Challenge received applications from 108 countries across six continents. This geographic diversity demonstrates the many ways data science is being sought to help individuals and communities across the globe recover and thrive, while building resilience to withstand future challenges in diverse contexts.

Applicant Continent	Percentage of Total Applications
North America	>45%
Africa	>24%
Asia	>10%
Europe	>9%
South America	>8%
Oceania	>1%

Figure 1: Distribution of Applications by Country



American Samoa	3	Djibouti	1	Iraq	2	Namibia	2	Spain	7
Angola	2	Dominican Republic	1	Ireland	1	Netherlands	13	Sri Lanka	3
Argentina	4	Ecuador	3	Israel	4	New Zealand	2	Switzerland	8
Australia	17	El Salvador	2	Italy	4	Nicaragua	1	Tanzania	8
Bangladesh	7	Ethiopia	2	Jamaica	1	Nigeria	53	Thailand	3
Belgium	4	Fiji	1	Japan	1	Pakistan	5	The Democratic Republic of the Congo	7
Benin	2	Finland	1	Jordan	2	Palestine	1	Togo	1
Bolivia	1	France	5	Kenya	75	Panama	1	Trinidad and Tobago	1
Botswana	3	Gabon	1	Lebanon	1	Paraguay	2	Turkey	3
Brazil	44	Germany	11	Liberia	1	Peru	8	Uganda	44
Burundi	1	Ghana	18	Lithuania	1	Philippines	12	Ukraine	1
Cameroon	4	Greece	1	Macedonia	1	Poland	1	United Arab Emirates	1
Canada	29	Guatemala	2	Malawi	5	Portugal	1	United Kingdom	50
Chad	1	Guinea	1	Malaysia	1	Romania	1	United States Minor Outlying Islands	1
Chile	9	Guinea-Bissau	1	Mali	3	Rwanda	4	United States of America	521
China	1	Guyana	1	Mauritius	4	Senegal	4	Venezuela	1
Colombia	31	Haiti	1	Mexico	10	Serbia	1	Yemen	3
Costa Rica	1	Honduras	1	Mongolia	1	Sierra Leone	3	Zambia	4
Côte d'Ivoire	3	Hong Kong	1	Montserrat	1	Singapore	6	Zimbabwe	5
Croatia	2	Hungary	1	Morocco	1	South Africa	37		
Cyprus	1	India	68	Mozambique	3	South Korea	1		
Denmark	3	Indonesia	9	Myanmar	2	South Sudan	1		

Applicant Initial Implementation Geography

The Challenge accepted applications from individuals and teams focusing on implementing projects across the globe. The table below demonstrates the diversity of the initial implementation geographies across all applicants.

Applicant Initial Implementation Geography	Percentage of Total Applications
Asia Pacific	>5%
Europe	>2%
Global/Multi-continental	>13%
Latin America	>11%
Middle East and/or North Africa	>2%
South Asia	>6%
Sub-Saharan Africa	>28%
U.S./Canada	>29%

Applicant Type

Applicants also represented a range of teams, backgrounds, and organizations, showcasing the diversity of individuals and entities who have recognized the potential to use data science for social impact. Nearly a quarter of applications were made by individuals or teams (>19%), while others were submitted on behalf of a legally formed entity (>73%). Among applications from entities, the majority of these proposals were submitted by NGO or charitable organizations (>47%), which included both global and United States-based NGOs, academic institutions, 501(c)(3) public charities, or equivalent. Applications were also submitted by for-profit business entities (>26%), multilateral organizations including UN agencies (>2%), and government agencies (>3%).

Applications by Applicant Type

For-profit Business Entity



Government Agency



Individual



Multilateral Organization



NGO or Charitable Org (U.S.)



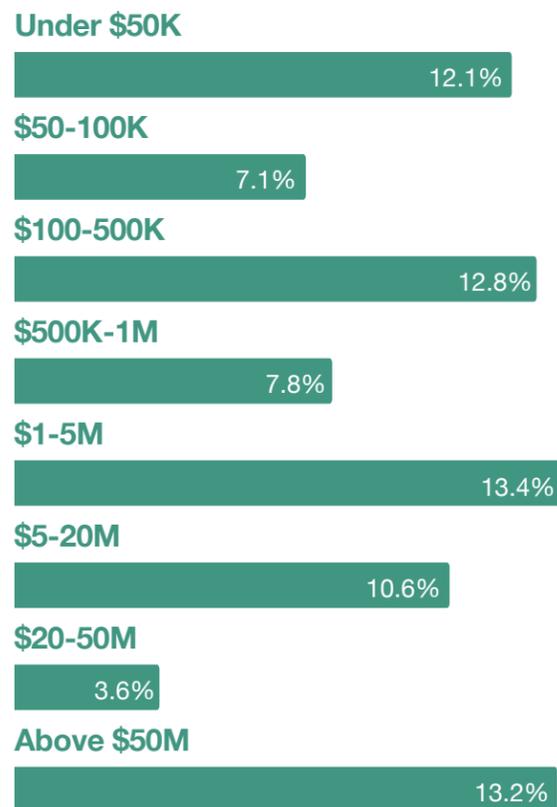
NGO or Charitable Org (Outside U.S.)



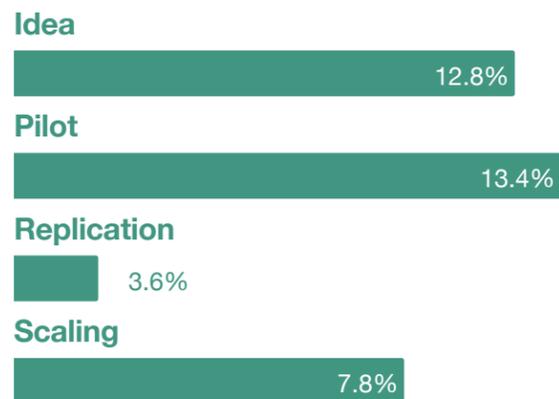
Operating Budget

Applicants also ranged in size and scope—evidence that data science capacity is a need in both large and small organizations. Many applications (>32%) were received from start-ups and small NGOs with annual budgets of under \$500K, nearly half of applications (>40%) came from well-established groups with operating budgets between \$1-50 million USD, and a small section of organizations (>13%) held budgets above \$50 million USD.

Distribution of Operating Budget Size in USD



Distribution of Project Stage



The Challenge also received projects at every stage of their lifecycle, particularly those at the idea, pilot, and scaling phases. That is, nearly one-third of applications (>32%) presented proposals at the idea stage—defined as innovations that required research, prototyping, and early testing to assess viability. Moreover, there was a large number of applications at the pilot stage (>35%), which were proposals early in their implementation that still required testing to understand feasibility. And while there were few applications in the stage of being replicated or generalized (>5%), many were at the final stage of scaling (>25%), which had moved beyond pilots and integrated innovations into existing systems to reach larger scale.

Analysis of Challenge Applications

The applications submitted to the data.org Challenge comprised a broad range of topical areas under the general umbrella of inclusive growth. The richness and variability of applications not only highlighted the individual conditions for organizational innovation but also accounted for the fundamental issues that aimed to drive sector-level impact.²

Challenge Themes

The majority of application topics (>40%) were aligned with the Challenge area Leave No Place Behind—Cities and Towns. Many applications aligned with the Challenge theme Leave No Person Behind—Jobs of Tomorrow (>27%) with data science solutions that broadly addressed vulnerable workers and the labor market. Nearly a fifth of applicants were rethinking the ways in which data science could unlock capital for high potential for small-business owners and entrepreneurs, positioning their proposals in the Leave No Entrepreneur Behind—Access to Capital area (>17%). There was also the opportunity to submit proposals based outside of these three main topic areas, in which >13% applicants chose to do.

Sectors

In addition to the three broad categories of this Challenge, applicants were asked to identify the primary sector that they sought to influence. The majority of applicants proposed to tackle issues related to economic development (>36%). These topics included areas such

as microcredit lending, support for small businesses, and agriculture, among others. Applicants also proposed to tackle challenges associated with health (>12%), education (>11%), social services (>7%), as well as topics related to energy and environment (>2%).

Applications by Sector



² Note that these applications are not fully representative of the overall distribution of inclusive growth topics in the sector, as both the open-call and applications materials were in English. The Challenge also relied on partners and government actors for dissemination, which was conducted via social media.

Applications by Primary Sustainable Development Goal

SDG 1: No Poverty

67.7%

SDG 2: Zero Hunger

36.1%

SDG 3: Good Health and Well-being

60.1%

SDG 4: Quality Education

37.1%

SDG 5: Gender Equality

34.6%

SDG 6: Clean Water and Sanitation

15.4%

SDG 7: Affordable and Clean Energy

10.5%

SDG 8: Decent Work and Economic Growth

73.6%

SDG 9: Industry Innovation and Infrastructure

56.7%

SDG 10: Reduced Inequality

76.9%

SDG 11: Sustainable Cities and Communities

65.4%

SDG 12: Responsible Consumption and Production

25.2%

SDG 13: Climate Action

14.9%

SDG 14: Life Below Water

4.4%

SDG 15: Life on Land

23.8%

SDG 16: Peace and Justice Strong Institutions

25.7%

It is important to note that these applications extended beyond a single societal topic or sector. In fact, each application touched upon one or several of the United Nations' Sustainable Development Goals (SDGs), particularly those regarding the amelioration of poverty and income inequality (i.e., SDG 1, 68%; SDG 8, 74%; SDG 10, 77%). There was also a considerable interest in topics including health (i.e., SDG 3, 60%), hunger (i.e., SDG 2, 36%), education (i.e., SDG 4, 37%), and gender equality (i.e., SDG 5, 35%). Topics on sustainable cities and environments were also frequent. Many applications touched on goals supporting sustainable cities (i.e., SDG 11, 65%), industries (i.e., SDG 9, 57%), clean water and sanitation (i.e., SDG 6, 15%), as well as responsible production (i.e., SDG 12, 25%). For a deeper analysis on SDG representation and co-occurrence of SDGs using a Market Basket Analysis, see Appendix. The chart to the left indicates the percentage of applications that touched upon each SDG.

Learnings from Natural Language Processing Analysis

As observed above, applications touched on one or multiple topics, exhibiting a range of areas under a general inclusive growth framework. Delving deeper into these topics, the Challenge applications were analyzed using a Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) algorithms, methods for unsupervised classification of documents (see Appendix for method).³ A sample of topics were uncovered, including those aligned with both the Challenge themes and other inclusive growth themes (e.g., smallholder farms, urbanization and sustainable cities, and gender equality). This probabilistic distribution of words yielded a fruitful summary of applications and the range of conditions needed for sector-level impact, exemplified using real-world examples below. The section below describes common topics and demonstrates how social sector actors are designing and implementing diverse data science solutions to address problems within the same areas. The six critical topics that we explore below will be essential for further engagement by funders and social sector actors.

³ This model conveys abstract topics across $n = 1,241$ documents (i.e., applications in English) and generates words based on their probability distribution in the text.

Critical Topics for Further Inquiry

After an expansive review of proposals and analysis of the application trends, data.org and DataKind identified six prevalent topics that are critical to inclusive growth and recovery and that have tremendously exciting applications of data science.

These topics focus on unique stakeholders and environments including individuals, groups, and communities that too often do not receive an equal share of the benefits of economic growth. Further, these topics are relevant to countries in every continent and demonstrate that, while context and solutions will vary, there are common issues within the inclusive growth arena that need attention across the world. The topics that follow will be essential for funders, organizations, and researchers to consider within the inclusive growth space.

Critical Topics for Further Inquiry



Smallholder Farmers and Agriculture

According to the **Food and Agriculture Organization of the United Nations**, about 90% of the world's 570 million farms are small, and the **International Labour Organization** reports that in Sub-Saharan Africa and Southeast Asia farming can be 50% to 78% of a country's GDP.⁴ Smallholder farmers typically operate in especially rural and impoverished communities across the world and face a myriad of challenges in enhancing their livelihoods. Too often, smallholders are unable to access loans, expand their farms, or reach their buyers. To address these and other issues faced by smallholder farmers and within the agricultural sector more broadly, applicants sought techniques that apply satellite imagery to expand access to loans or share market-insights to help farmers reach buyers, to name a few.



Affordable Housing and Neighborhoods

An economic recession will hit some places harder than others. As public and private sector investments flow into cities and towns to stimulate the economy, leaders must use data-driven insights to make the right decisions to ensure economic security in underserved communities. As we have learned from Raj Chetty's groundbreaking work with **Opportunity Zones**, data and analytics can help identify the granular resources and networks within neighborhoods that are needed to access opportunity including quality education, affordable housing and childcare, decent jobs, and transportation. Proposals on this topic sought to use mapping tools to empower policy-makers to expand affordable housing, expand access to social services through natural language processing (NLP), and identify communities without access to critical resources.



Micro, Small, and Medium Enterprises and Entrepreneurship

Unleashing the spirit of grassroots entrepreneurs can help communities rebound by creating much-needed jobs and growing local economies. Yet, in times of recession, the flow of credit to micro and small businesses tends to slow down. MSMEs represent the majority of private sector enterprises. Their survival and growth is critical to inclusive growth and recovery. Applicants in this topic sought to empower entrepreneurs and their enterprises by expanding access to credit, bringing informal businesses into the formal sector, and leveraging market-analysis to inform critical business-decisions.

⁴ "Employment in Agriculture (% of Total Employment) (Modeled ILO Estimate)." Accessed March 11, 2021. <https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS?end=2017&start=1991&view=chart>



Gender Equality

Despite significant progress over recent decades, women and girls across the world are too often left behind economically. Truly inclusive growth requires a specific focus on female entrepreneurs, workers, and youth. This is especially critical in the wake of COVID-19, as women are more likely than men to work in frontline capacities and employed in the informal sector or provide unpaid labor such as domestic work. Female entrepreneurs are often unable to access loans at the same rate as male entrepreneurs, due to ongoing bias, both explicit and implicit. Projects in this topic sought to expand access to capital for female entrepreneurs, provide critical training and support to women-owned businesses, and expand educational and vocational opportunities for young women.



Urbanization and Sustainable Development

As communities urbanize, too often inequality is exacerbated and the benefits of economic growth are unevenly distributed. Inclusive growth necessitates that growth and prosperity extend to every strata of society. Projects in this topic sought to forecast the unequal economic impacts of urbanization; understand trends related to migration and empower policymakers with the information needed to adequately support migrants; and expand opportunity for gig and informal workers.



Youth Unemployment

Empowering young people with the education and skills to start businesses or excel in the workforce is critical to inclusive growth. This is especially crucial now as, according to the **International Labour Organization**, young people aged 15-24 face an unemployment rate three times higher than adults before the crisis and have only seen their job prospects dwindle since. Projects within this topic sought to launch applications to connect youth with training opportunities, provide young people with access to the internet, and connect them with services critical to their socio-emotional development.

Capacity of Applicants

In order to build the field of data science for social impact, data.org is committed to capacity building of mission-driven organizations that seek to integrate data science into their work. In order to learn more about strengths and needs related to organizational capacity, data.org and DataKind reviewed over 100 applicants that scored highest in their data science assessment to determine their capacity for data, talent, technology, and partnerships. The capacity in each area of these 100+ applicants is displayed below on a scale of high to low.

Heat Map of the Global Ratings

	Data	Talent	Tech	Partnership
High	56	40	48	24
Medium	31	38	22	45
Low	15	24	27	36
N/A	3	3	8	0

Noting that data and partnerships described in applications were specific to the proposed projects and may not be indicative of organizations as a whole, we observed the following trends in organizational capacity.

Research institutions and universities had the highest potential for data capacity; however, on partnerships they scored quite low. Furthermore, their interventions were focused primarily on insight that could potentially be applied, rather than a full plan to reach impact.

Think tanks showed better performance on partnership commitment. However, they tended to lack experienced data science or relevant technical talent to lead the technical portion of the research.

Multilateral organizations were most equipped for diverse, multi-sector partners, especially including the public sector, though formal confirmation of partner commitments varied greatly depending on the stage of the application: idea, pilot, or scaling. The majority of the applications relied on external partnership or contractors to lead and expand the technical portion of the project.

Service providers primarily used their own internal transaction data from technology systems. Internal tech teams may lack experienced data science talent to lead, develop, and implement data science and machine learning projects. Most project proposals were self-contained solutions, so service providers also tended to have few or no partners.

Data Science Solutions

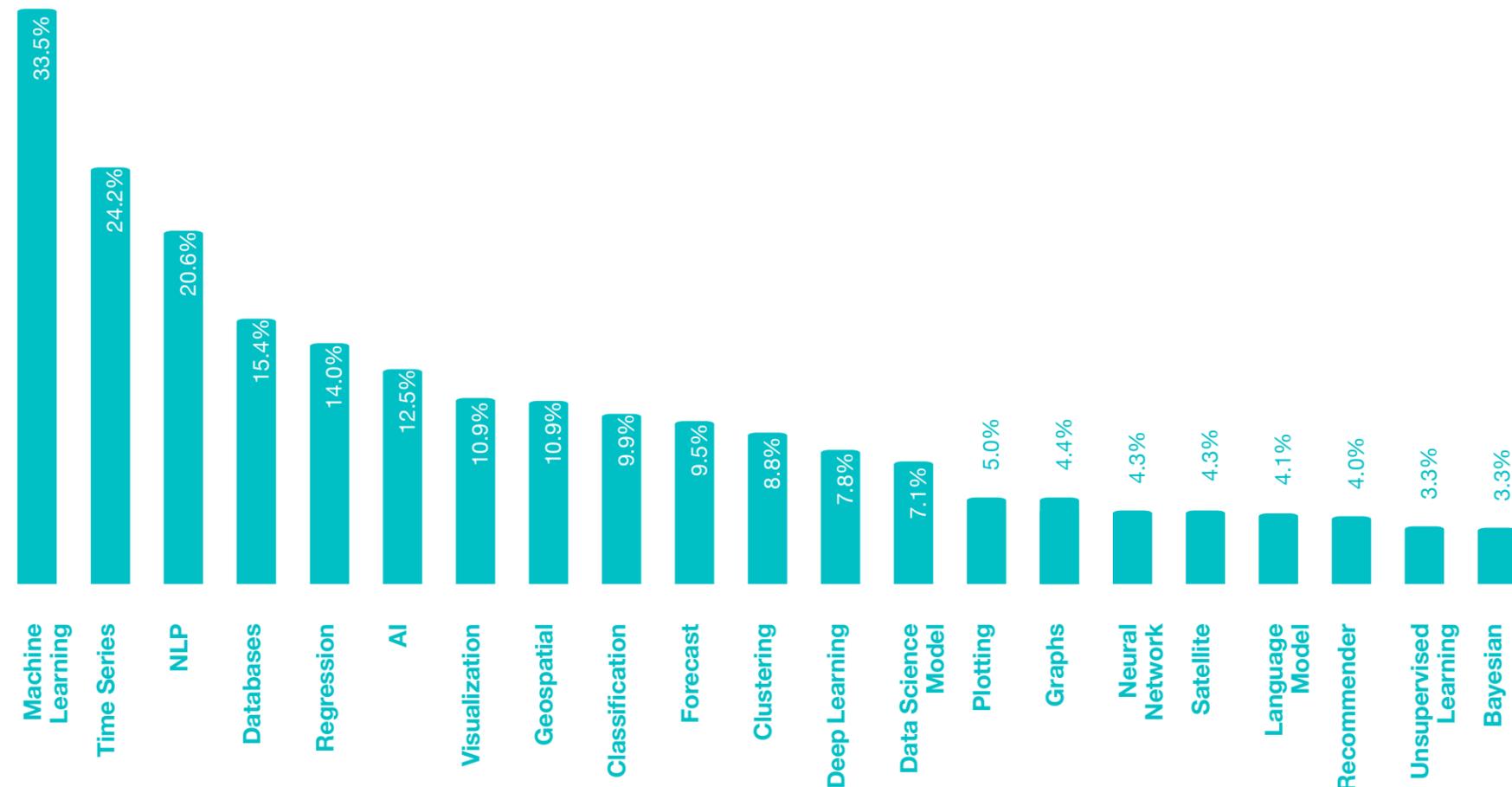
Challenge applications represented a variety of problems and data science solutions, all which addressed fundamental issues that aimed to drive sector-level impact. Both common and more advanced core analytic and data science methods were proposed across applications, which were identified using a cohort of NLP and pattern-matching techniques (see Appendix for methods). The following table highlights the common data science and AI techniques and applications proposed by applicants as assessed by DataKind.⁵

⁵ This summary is not an exhaustive list of the data science-powered approaches proposed by applicants, but rather a selection of the common techniques and applications.

Table 2: Common Techniques and Applications Proposed by Applicants

Proposed Data Science and AI Techniques	Definition and Sample Techniques
AI-Powered Techniques	
Machine Learning	A technique that learns through experience, namely through training data, to make predictions or decisions without being explicitly programmed to do so (e.g., supervised, unsupervised, and reinforcement learning). Deep Learning is a subset of machine learning based on artificial neural networks in which multiple layers of processing are used to extract progressively higher level features from data (e.g., Convolutional Neural Networks (CNNs), and Generative Adversarial Networks (GANs).
Common Applications	
Natural Language Processing	An interdisciplinary field concerned with the interactions between computers and human language, particularly how computers process and analyze large amounts of natural language data. Includes techniques such as speech-tagging, sentiment analysis, topic modeling, multilingual translation, auto-summarization, and graph models.
Geospatial Analysis	A cohort of techniques and approaches that gather, manipulate, and display geographic information system (GIS) data and imagery to create models and visualizations including satellite imagery, mapping, and spatial network analysis, among others.
Computer Vision	An interdisciplinary field broadly concerned with how computers can gain high-level understanding of images and videos including but not limited to object detection, image/video classification, and image segmentation, among others.
Predictive Modeling and Forecasting	Broadly understood as the use of statistical methods to predict outcomes including time series and forecasting and regression techniques (e.g., linear, logistic, and decision trees).
Other techniques	
Visual Analytics	Visualization of data including interactive graphics and dashboarding (e.g., Tableau and RShiny).
Data Engineering and Data Preparation	Approaches to prepare data, create ETL (Extract, Transform, Load) and data pipelines, and data warehouses.

Distribution of Data Science Terms in Applications



Delving deeper into the data science terms, the data.org Challenge applications cited using a number of AI-powered techniques among other data science applications and problem domains. The distribution of these top terms is shown here.

Across submissions, >33% of applicants cited “machine learning” in their applications, while several others (7%) proposed “deep learning” techniques and its relevant models (e.g., “neural network,” “CNN,” and “RNN”). Note that these are AI-powered techniques, which can be used in multiple problem domains and across multiple stages of project growth. For instance, the majority of applications cited using signal analyses including terms like “time series” (>24%), “regression” (>14%), “predictive modeling” (>10%), “forecast” (>9%), and “random forest” (>3%) were cited across these applications. Moreover, “NLP” was a common application that was cited in >17% of applications, in addition to other domain-specific approaches and techniques (e.g., TFIDF, text mining, and auto-summarization). Other salient applications of data science and data engineering in the Challenge included “database” (>15%), “geospatial” (>10%), and “visualization” (>10%), as well as ML methodologies like “classification” (>9%) and “clustering” (>8%). While these terms demonstrate the varying use cases and applications of data science approaches, they also highlight each project’s unique needs as they relate to their data maturity and stage. That is, early stage projects may require exploratory data analyses (EDA) and common machine learning applications, whereas scaling projects have likely already identified more advanced, domain-specific techniques.

As our analysis revealed, the Challenge applications exhibited a range of topics under the umbrella of inclusive growth, demonstrating the important role data science can play in solving some of the world’s most pressing challenges. Whether through proposals to promote financial security and economic mobility for workers, help micro and small businesses access capital, or support economic development in communities, this Challenge exemplified how individuals and teams across the world are seeking to apply data science to solve our most pressing problems.

The Challenge Review Process

Nearly
400
Judges
49
Countries
19
Time Zones
1,260
applications
3,500
Reviews
in Just
4
Weeks

To find exceptional ideas from around the world that applied data science to encourage inclusive growth and economic recovery, data.org and DataKind designed and conducted a thoughtful and rigorous evaluation of each application. DataKind led the review application process and evaluated multiple times in a holistic manner.

To begin, DataKind reached out to its global community and networks and data.org recruited diverse organizations to select and train a team of nearly 400 judges. DataKind trained and supported this team of data science experts to assess Challenge applications and identify the most promising ideas using a rubric that represented the evaluation criteria. The team represented different backgrounds, experiences, and nationalities, hailing from 49 countries across 19 time zones. The team of judges also drew from a diverse range of experience including government, private sector, NGOs, and academia and were equipped with a range of technical skills. The judges were experts in data science and machine learning, economics, finance, and other fields in the quantitative social sciences. In addition to DataKind's network of experts, volunteers from Global Data Barometer, Latin America Open Data Initiative, Mastercard, Mobile Web Ghana, the World Bank, Washington University in St. Louis, and Younglings supported the Challenge by offering expert advice in reviewing the applications.

Each of the 1,260 applications was reviewed multiple times, totaling over 3,500 reviews in just four weeks. Evaluations provided rich commentary and feedback on applications that supported data.org in identifying Challenge awardees.

Applications that passed this first round of review were then invited to submit more detailed proposals that were evaluated by senior staff at data.org, DataKind, the Mastercard Center for Inclusive Growth, and The Rockefeller Foundation. Each of these applicants also completed a 90-minute interview, including a project presentation and data readiness review. During interviews, partners acquired a deep understanding of each applicant's vision, execution plans, data readiness, and alignment with inclusive growth and recovery.

Finally, a group of prominent leaders in the fields of data science and social impact served as Expert Judges to evaluate and provide insight and recommendations on finalists. (See Expert Judges in Appendix.)

Judging Criteria and Selection

During every stage of the review process, applications were evaluated based on the Challenge's five principal criteria. Judges, Expert Judges, and staff assessed applications based on their potential impact, replicability, scalability, practicality, and breakthrough ideas.

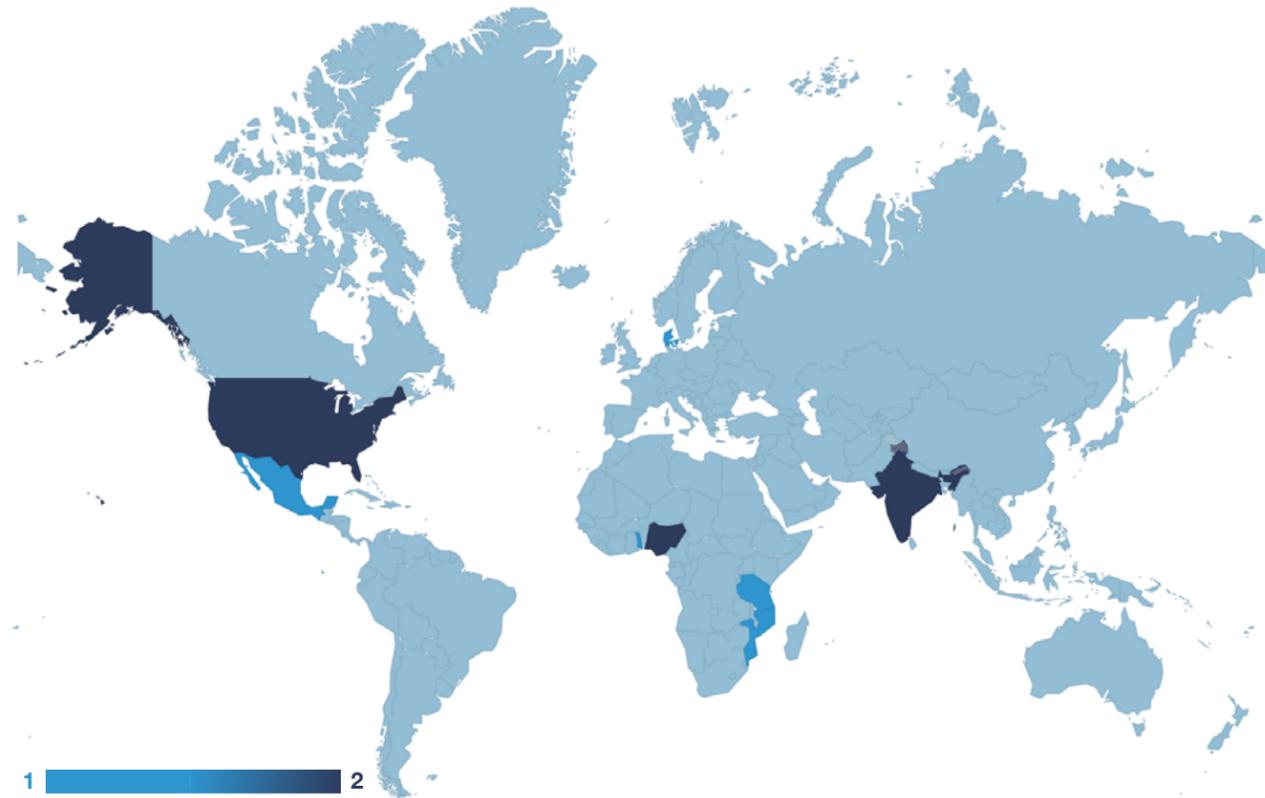
- Impactful:** Addressing an important inclusive growth and/or recovery challenge which drives toward real-world impact in areas such as Jobs of Tomorrow, Access to Capital, or Cities and Towns.
- Replicable:** Presenting a proposal that can be adapted and/or generalized to different cases, such that data science techniques and approaches could address similar inclusive growth and recovery problems elsewhere.
- Scalable:** Creating a project that can be expanded on a larger scale. Based on this application, the applicant proposes creating a project that might be expanded on a much larger scale, driving toward real-world impact in areas outside the proposed solution space.
- Practical:** Meeting reasonable resource and execution requirements with manageable risks and a realistic plan to execute on the proposal including the required data, technical expertise, and leadership needed to implement.
- Breakthrough:** Designing an insightful, thoughtful, and new application of data science that transforms how we address entrenched social challenges.

Beyond the criteria above, judges also focused on ethical considerations and potential biases during the evaluation process. Judges closely examined the risks of both failure and success of the data intervention as well as the unintended consequences that could arise from how data science is used to solve inclusive growth and recovery challenges.

Awardees

Initial Implementation Geography

Ultimately, data.org, in consultation with the Mastercard Center for Inclusive Growth, The Rockefeller Foundation, and DataKind, identified eight winning projects that demonstrate tremendous impact, replicability, scalability, practicality, and exciting applications of data science. These projects spanned eight countries across four continents.



Initial Implementation Country	Number of Projects
Denmark	1
India	2
Mexico	1
Mozambique	1
Nigeria	2
Tanzania	1
Togo	1
United States	2

We see tremendous potential for each of these projects to scale to new geographies, as well as inspire and inform similar projects across the world.

In the Cities and Towns topic area, our awardees seek to address critical inclusive growth issues facing communities in Denmark, Togo, and the United States. Aalborg University in Denmark will create an interactive and actionable map to identify the geographical places in Denmark that are most vulnerable to out-migration and economic instability. Community Lattice, a United States social enterprise, will create a platform to predict the cost and risk of brownfield redevelopment projects in the United States to transform a community's ability to secure redevelopment funding, improve community health, and create economic opportunities. GiveDirectly, a United States nonprofit, will develop and implement a new model for humanitarian support that enables cash transfers to be deployed effectively, accurately, and at scale to those who need them most. Lastly, the University of Chicago in the United States will build a model to mitigate the digital divide in Chicago.

In the Access to Capital theme, our awardees focused on diverse challenges facing entrepreneurs across the world. Basel Agency for Sustainable Energy, a Swiss nonprofit, will enable smallholder farmers in India to access sustainable cooling facilities, thereby reducing food loss and dramatically improving livelihoods for those farmers. In Nigeria and Tanzania, Solar Sister, a United States nonprofit, will share market insights with its network of women entrepreneurs to grow their renewable energy businesses. Women's World Banking, a United States nonprofit, will work with partner financial service providers across continents in India, Mexico, and Nigeria to increase credit access for low-income female entrepreneurs.

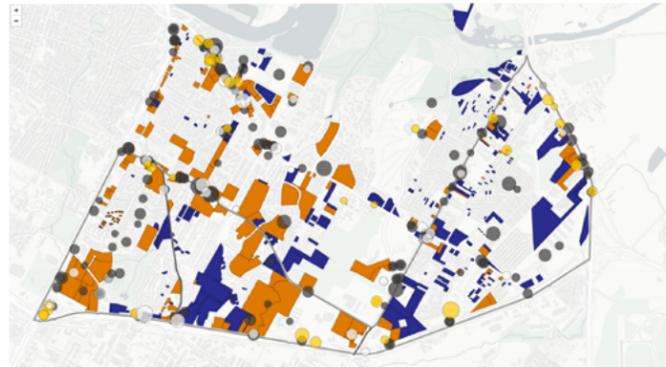
Within the Jobs of Tomorrow theme, Fundación Capital, a Panamanian nonprofit, seeks to provide informal workers in Mozambique with essential labor-market insights to increase their income and employment opportunities.

Cities and Towns: Leave No Place Behind

Community Lattice

Environmental risk model for revitalization

Country: United States
Implementation Country: United States
Partners: Kansas State University Technical Assistance to brownfields Program (KSU TAB); Fifth Ward Community Redevelopment Corporation



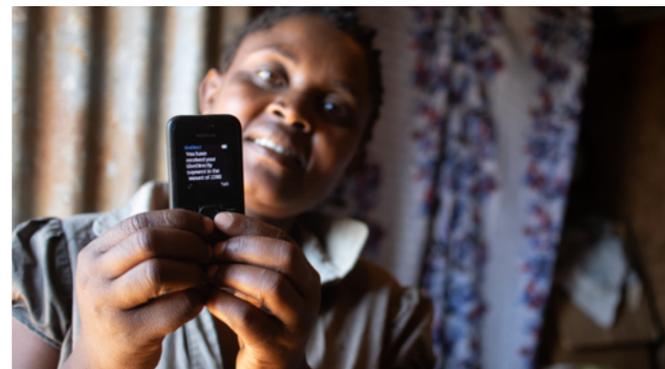
Rehabilitating underutilized properties in underserved communities—disproportionately home to low-income persons and minorities—can stimulate economic growth and quality of life improvements. However, with the potential high cost and liability for cleanup due to unknown environmental conditions, these properties (known as “brownfields”) can deter investment and exacerbate a neighborhood’s decay.

Community Lattice proposes incorporating historical brownfields clean-up data in the United States with environmental and economic data to create a site cleanup cost model to address the risks associated with brownfield redevelopment. This data and the predictive model would then be available to the public through an online platform. By accurately predicting the cost and risk of brownfield redevelopment projects using machine learning methods, the team seeks to transform a community’s ability to secure redevelopment funding, improve community health, and create economic opportunities.

GiveDirectly and Center for Effective Global Action

Using data science to target cash transfers for COVID-19 relief

Country: United States
Implementation Country: Togo
Partners: Data-Intensive Development Lab at UC Berkeley



On the road to recovery from COVID-19, many countries lack reliable and up-to-date information about economic conditions on the ground and have no way of collecting it during a pandemic. Traditional aid modalities relying on in-person enrollment and delivery are no longer safe or scalable, with governments and NGOs lacking personnel and relief taking weeks to arrive.

GiveDirectly and the Center for Effective Global Action at UC Berkeley propose addressing this challenge by developing and testing a new model for humanitarian support that enables cash transfers to be deployed effectively, accurately, and at scale to those who need them most. This project incorporates new data and computational technologies to identify people and places in economic distress and integrating data from mobile phones, satellite imagery, and traditional surveys. The plan will pilot in one or more low and middle-income countries and develop a transparent framework to scale globally.

University of Chicago, Center for Data and Computing

Mapping and mitigating the urban digital divide

Country: United States
Implementation Country: United States
Partners: KidsFirst Chicago; City Tech Collaborative



Educational and economic opportunities depend on affordable, high-speed internet access. The COVID-19 pandemic has accelerated and magnified these existing disparities, even within the same city.

The University of Chicago will pilot a project to study fiber connectivity, broadband throughput, application performance metrics, pricing data, and information about subscriptions and affordability to understand Chicago communities’ access to affordable, high-speed internet. The team will then work with technical and civic partners to design and evaluate new network architectures for improving connectivity in neighborhoods and regions that are sparsely connected. The project’s goal is to create a complete assessment of a city’s broadband infrastructure that can be used to investigate innovative infrastructure solutions, raise awareness about this critical issue, and empower policymakers, industry leaders, and the public to make better and more informed decisions. This project has the potential for tremendous impact in Chicago and to scale to cities across the U.S.

BUILD | Aalborg University

Mapping the regional quality of life

Country: Denmark
Implementation Country: Denmark



There is an increasing regional divide, witnessed in Denmark just as in much of Europe and North America. As economic inequality between regions increases, this disparity leads to enhanced migration out of underserved communities. Those leaving these communities are more often highly educated, young adults moving to large urban areas to pursue education and jobs. These persistent patterns of who leaves and who stays behind reinforce regional differences. As a result, underserved communities often struggle to ensure local public services and economic security.

BUILD will provide public authorities and decision-makers with tools to compare areas and identify those with less local economic opportunity. To spread the use of data-driven insights, BUILD is creating an interactive website containing key indicators of economic prosperity in any given area in Denmark. Indicators will include housing prices, human capital, local wage levels, transportation, migration rates, and local amenities. The website will feature interactive maps and infographics illustrating the key insights and give an overview of the regional differences.

Access to Capital: Leave No Entrepreneur Behind

Basel Agency for Sustainable Energy

Your virtual cold chain assistant

Country: Switzerland
Implementation Country: India
Partners: Empa (the Swiss Federal Laboratories for Materials Science and Technology) (Switzerland)



India is one of the world's largest food producers, yet 25% to 35% of the food produced is wasted due to a lack of proper refrigeration and other supply chain bottlenecks. Only 6% of the food produced in India currently moves through the cold chain, compared to about 60% in developed countries.

To increase this percentage and support smallholder farmers—who make up the bulk of India's hungry and poor—BASE (Basel Agency for Sustainable Energy) & Empa (the Swiss Federal Laboratories for Materials Science and Technology) will create an open access, data science-based mobile application, using machine learning and physics-based food modeling.

Solar Sister

Empowering women entrepreneurs with data science

Country: United States
Implementation Country: Nigeria & Tanzania
Partners: Fraym (United States)



Energy poverty is particularly acute in Sub-Saharan Africa, where roughly 600 million people lack electricity and 890 million cook with harmful fuels. To address this issue, Solar Sister recruits, trains, and supports local women entrepreneurs to deliver clean energy directly to homes in rural African communities.

To further support this network of women and expand their social enterprise, Solar Sister is launching Empowering Women Entrepreneurs with Data Science, a collaboration with Fraym. Through this partnership, Fraym will support Solar Sister with data science expertise—using predictive modeling, hyperlocal spatial layers, ensemble-based machine learning pipeline, and covariate matrix to train machine learning models—to provide insights on potential customers. Insights generated will help women entrepreneurs build tailored customer profiles for off-grid solar, better target their customers, and cater to their needs to help grow and sustain their renewable energy businesses. Leveraging data in this way will build healthy markets for renewable energy solutions by driving innovation and sales, in turn increasing community access to affordable and reliable modern energy, catalyzing socio-economic development, and improving community productivity and livelihood.

Women's World Banking

Innovative AI for women's financial inclusion

Country: United States
Implementation Country: Indonesia, Mexico, Nigeria
Partners: University of Zurich



Currently, female entrepreneurs are more likely to get lower premiums, higher interest rates, and increased penalties for mistakes due to out-of-date, gender-biased lending technology and practices. One billion women remain outside the formal financial system today. A solution for this is particularly important in the present climate as emerging markets look to recover from the effects of COVID-19.

To address these issues, Women's World Banking, in partnership with the University of Zurich, is exploring the implications of AI-based modeling and credit scoring on women's financial inclusion.

With a strategic focus on two of Women's World Banking's key priority markets, researchers will assess how algorithms in digital credit applications can increase lending to women borrowers, study the applications of machine learning and AI, and explore the challenges facing digital financial services as a result of COVID-19.

Jobs of Tomorrow: Leave No Worker Behind

Fundación Capital

Use of business intelligence for informal workers

Country: Panama
Implementation Country: Mozambique
Partners: UX (Mozambican tech startup); Data Elevates (U.S. social enterprise—data analytics)



In Mozambique, 96% of all workers are employed in the informal sector, facing higher poverty rates, economic insecurity, and fewer opportunities for economic advancement. To increase opportunities for informal workers, UX launched Biscate, a nonprofit digital job platform where workers can advertise their professions, experiences, and locations, and potential clients can browse services, contact workers, and leave ratings.

To increase this platform's impact, Fundación Capital and UX seek to use data mining, visualization techniques, and a machine learning-powered recommendation system to deliver real-time labor market insights directly to informal workers. This project could have a significant impact on the economic well-being of the vast majority of workers in Mozambique by increasing their job opportunities and potential income.

Conclusion

With COVID-19 and its economic impact, the data.org Challenge was an open call for breakthrough ideas that harness the power of data science and AI to help people and communities rebound and remain resilient. This Challenge has further proven that there are clear and achievable opportunities for data science and AI to solve the world’s most pressing issues.

Throughout this work, data.org and DataKind has observed the myriad ways in which organizations of all types, sizes, and maturity recognize the transformative potential of data science and are seeking to incorporate a variety of data science applications to deliver social impact. While some organizations would benefit from basic data science capacity-building, others lack the funding to implement breakthrough data science projects.

These insights are meant to contribute to the growing interest in data science for good and empower organizations through the creation of new partnerships and shared resources needed to reach social impact. Across the world, social sector organizations are ready to adopt data tools, talent, methods, and solutions that are currently concentrated in the private sector. To support a variety of techniques that generate positive impacts on people’s lives, a community of data scientists, funders, social sector organizations, private sector partners, community experts, practitioners, and policymakers must come together to support and grow current efforts.

data.org, working together with partners such as DataKind, will continue to showcase examples of the transformative potential of unlocking the power of data science for social impact and will continue to work with promising organizations and individuals to increase their capacity and talent to create lasting change. It will be thrilling to see how the Challenge awardees and applicants continue to use data science to improve the world.

Appendix

Expert Judges	34
Review Methodology	39
Challenge Application	45
Acknowledgments	51

Expert Judges

To garner critical feedback on recommended awardees for the Challenge, data.org gathered 12 prominent leaders in the fields of data science and social impact to serve as Expert Judges.



Bayo Adekanmbi

Chief Transformation Officer, MTN Nigeria

Olubayo (Bayo) Adekanmbi combines 19 years of cognate industry experience in Strategy, Marketing, Analytics, and Business Transformation from two largest economies in Africa (Nigeria and South Africa). He is a C-level executive at MTN Nigeria, Africa's largest telecommunication company, where he has also been awarded the distinguished award for using advanced analytics to drive high-impact business growth. He is the convener of Data Science Nigeria, a nonprofit that is building a world-class AI knowledge, research, and innovation ecosystem which delivers high impact and transformational research, business use applications, AI-first start-ups, employability, and social good use cases. He is the author of "The Future is Shared" and "Beginners Artificial Intelligence and Python Programming."



Raj Chetty

William A. Ackman Professor of Economics, Harvard University

Raj Chetty is the William A. Ackman Professor of Economics at Harvard University. He is also the Director of Opportunity Insights (formerly the Equality of Opportunity Project), which uses "big data" to understand how we can give children from disadvantaged backgrounds better chances of succeeding. Raj's research combines empirical evidence and economic theory to help design more effective government policies. His work on topics ranging from tax policy and unemployment insurance to education and affordable housing has been widely cited in academia, media outlets, and Congressional testimony. Raj received his Ph.D. from Harvard University in 2003 and is one of the youngest tenured professors in Harvard's history. Before joining the faculty at Harvard, he was a professor at UC Berkeley and Stanford University. Raj has received numerous awards for his research, including a MacArthur "Genius" Fellowship and the John Bates Clark medal, given to the economist under 40 whose work is judged to have made the most significant contribution to the field.



Michael Froman

Vice Chairman, Mastercard

Mike Froman serves as Vice Chairman for Mastercard. In this role, he is responsible for growing strategic partnerships, scaling new business opportunities, and advancing the company's efforts to partner with governments and other institutions to address major societal and economic issues. Mike also oversees the Mastercard Center for Inclusive Growth and is a member of the company's management committee. Prior to joining Mastercard, Mike was affiliated with the Council on Foreign Relations and continues to serve as a distinguished fellow. From 2013 to 2017, Mike served as the U.S. Trade Representative, President Barack Obama's principal advisor and negotiator on international trade and investment issues. Mike received a bachelor's degree in public and international affairs from Princeton University, a doctorate in international relations from Oxford University, and a law degree from Harvard Law School, where he was an editor of the Harvard Law Review.



Panthea Lee

Executive Director, Reboot

Panthea Lee is the Co-Founder and Executive Director of Reboot. She is a pioneer in designing and guiding multi-stakeholder processes to address complex social challenges, with experience doing so in 30+ countries, with partners including CIVICUS, Wikimedia Foundation, Open Government Partnership, MacArthur Foundation, and governments at the national, state, and local levels. The global co-design processes she's led have resulted in the launch of bold new efforts to protect human rights defenders, tackle public sector corruption, strengthen participatory democracy, advance equity in the open knowledge movement, reform leading international agencies, and drive innovation in independent media. Her work has been featured in *The Atlantic*, *The Stanford Social Innovation Review*, *Al Jazeera*, *MIT Innovations Journal*, and *Fast Company*. She serves on the boards of the RSA US, Development Gateway, and People Powered: The Global Hub for Participatory Democracy.



Laura Montoya

Founder and Executive Director, Accel.AI

Laura Montoya is the Founder and Managing Partner of Accel Impact Organizations, including Accel AI Institute, Latinx in AI, and Research Colab. She is a director with Women Who Code, advisor for Udacity's AI and Data Nanodegree, and an affiliate with the Berkman Klein Center for Internet and Society at Harvard Law. She chairs and serves on Program Committees for research workshops at AI and ML conferences including NeurIPS, ICLR, ICML, and ACM FAccT. She conducts research in reducing bias, data representations in machine learning models, the effects of AI development for developing countries, and paralleling biological and synthetic neural networks seen in mycology, entomology, and computational science. She has led sessions on social impact, tech diversity, and ethical AI development for Creative Mornings, Katapult Future Fest, Silicon Valley Future Forum, Tech Inclusion Conference, Thrival Summit, and Global Hive Summit.



Himanshu Nagpal

Deputy Director, Financial Services for the Poor, Bill & Melinda Gates Foundation

Himanshu Nagpal, Deputy Director, Financial Services for the Poor, Bill & Melinda Gates Foundation, manages the financial inclusion work in India and leads the team focused on driving usage of digital financial services by the poor and utilizing emerging technologies for driving financial inclusion. In addition, Himanshu also leads the foundation’s work on digital identity globally. Prior to joining the foundation in 2015, Himanshu spent time planning new technology products at Intel, serving the underserved markets with financial services at Capital One and HSBC Bank, working as an oil field services engineer in Nigeria with Schlumberger, and developing software solutions in India with Arthur Andersen. Himanshu holds a Bachelors of Technology in Chemical Engineering from Indian Institute of Technology Delhi and an M.B.A. from Wharton School of Business at University of Pennsylvania.



Enrica Porcari

CIO & Director of Technology, UN World Food Programme

Enrica Porcari is the UN World Food Programme’s Chief Information Officer and Director of Technology. She was also instrumental in creating WFP’s Fast IT and Telecommunications Emergency and Support Team to support global humanitarian operations. Enrica is advancing WFP’s efforts to lead a digital approach for global humanitarian operations, emergency responses, and development programs that improve life for millions of vulnerable people. Enrica is on the World Economic Forum’s Council of Drones and Aerial Mobility, and the Boards of the GSMA’s Mobile for Humanitarian Program and Big Data for Social Good Accelerator. Enrica is an Italian national and holds a Masters in Social Sciences from University of Milan and a Masters in International Procurement Management from University of Rome Tor Vergata. She is a Fellow of the Reuters Digital Vision Fellowship Programme at Stanford University.



Jake Porway

Founder, DataKind

A pioneer of the Data for Good movement, Jake Porway is an expert in the field of data and technology. As Co-Founder and Executive Director Emeritus of DataKind, he has worked alongside the nonprofit community to drive social change with the power of data science since 2011. Jake’s career spans more than a decade in the data science sector as a statistician and computer scientist. Prior to DataKind, he worked as a data scientist for *The New York Times* R&D Lab. A PopTech Social Innovation Fellow and a National Geographic Emerging Explorer, Jake was also noted as one of LinkedIn’s Next Wave Top Professionals 35 & Under and his efforts have led to DataKind being named one of *Fast Company*’s Top 10 Most Innovative Nonprofits. Jake holds a B.S. in Computer Science from Columbia University and an M.S. and Ph.D. in Statistics from UCLA.



Cassie Robinson

Senior Head of UK Portfolio, National Lottery Community Fund

Cassie Robinson is Senior Head of the UK Portfolio for the National Lottery Community Fund, a Co-founder of the Point People, and a Fellow at The Institute of Innovation and Public Purpose at UCL. Previously she was Strategic Design Director at Doteveryone and has worked in the Cabinet Office at Government Digital Service. Cassie is the Founder of the Civic Shop and Stewarding Loss and Co-founder of Tech for Good Global. She sits on the Board of Organise HQ and the Advisory Board of Participatory City Foundation. Cassie was awarded as a Nesta Creative Pioneer, a Democracy Fellow at Civic Hall, and named as a Leader in Philanthropy by the European Foundation Centre.



Bruno Sanchez

Principal Scientist, AI for Earth, Microsoft

Bruno Sanchez is the Principal Scientist at AI for Earth, Microsoft, building the “Planetary Computer.” By training, he received a Ph.D. in Astrophysics and a postdoc in rocket science. Bruno has led big data innovation at the World Bank Innovation Labs, served as VP Social Impact at the satellite company Satellogic, and was Chief Scientist at Mapbox. Bruno has published the book *Impact Science* on the role of science and research for social and environmental impact. He was awarded Mirzayan Science Policy Fellow of the U.S. National Academies of Science and a Young Global Leader of the World Economic Forum.



David Sangokoya

Project Lead, Society and Innovation, World Economic Forum

David Sangokoya is the Project Lead, Society and Innovation at the World Economic Forum in Geneva. He leads multi-stakeholder initiatives at the Forum related to driving stakeholder responsibility for social justice in technology, data, and global public goods including the Forum’s initiative on Partnering with Civil Society in the Fourth Industrial Revolution. With ten years of experience in research and nonprofit management, his background is in civil society—managing complexity, negotiating power, and responding to technological impacts. Prior to the Forum, David was the Research Manager at Data-Pop Alliance, a global coalition on big data and development created by the Harvard Humanitarian Initiative, MIT Media Lab, and Overseas Development Institute. He has previously worked as a data for good research fellow at the Governance Lab at NYU, as well as with community nonprofits in Sub-Saharan Africa and South Asia. He graduated with an M.P.A. from NYU and a B.A. with honors from Stanford University.



Rajiv J. Shah

The Rockefeller Foundation

Dr. Rajiv J. Shah serves as President of The Rockefeller Foundation. In 2009, he was appointed USAID Administrator by President Obama and unanimously confirmed by the U.S. Senate. Raj reshaped the \$20 billion agency's operations in more than 70 countries around the world by elevating the role of innovation, creating high impact public-private partnerships, and focusing U.S. investments to deliver stronger results. Raj secured bipartisan support that included the passage of two significant laws—the Global Food Security Act and the Electrify Africa Act. He led the U.S. response to the Haiti earthquake and the West African Ebola pandemic, served on the National Security Council, and elevated the role of development as part of our nation's foreign policy. Prior to his appointment at USAID, Raj served as Chief Scientist and Undersecretary for Research, Education, and Economics at the U.S. Department of Agriculture where he created the National Institute for Food and Agriculture. Raj founded Latitude Capital, a private equity firm focused on power and infrastructure projects in Africa and Asia and served as a Distinguished Fellow in Residence at Georgetown University. Raised outside of Detroit, Michigan, Raj is a graduate of the University of Michigan, the University of Pennsylvania School of Medicine, and the Wharton School of Business.



Vera Songwe

UN Under-Secretary-General and Executive Secretary, Economic Commission for Africa

Vera Songwe is the UN Under-Secretary-General and Executive Secretary of the Economic Commission for Africa, becoming the first woman to lead the institution in its 60-year history. As Executive Secretary, Vera's reforms, focusing on "ideas for a prosperous Africa," have brought to the fore critical issues of macroeconomic stability, development finance, private sector growth, poverty and inequality, the digital transformation, and trade and competitiveness. Recognized recently as one of Africa's 50 Most Powerful Women by *Forbes*, one of the 100 Most Influential Africans by *Jeune Afrique* in 2019, 100 Most Influential Africans by *New African Magazine* in 2017, and one of the 25 Africans to Watch by the *Financial Times* in 2015. Vera is acknowledged for her long-standing track record of providing policy advice and her wealth of experience in delivering development results for Africa. She has written extensively on development and economic issues including debt, infrastructure development, fiscal, and governance issues. Prior to ECA, she held a number of senior leadership roles with the International Finance Corporation and the World Bank.



Drew Zachary

Managing Director, Census Open Innovation Labs

Drew Zachary is Managing Director of Census Open Innovation Labs and Director/Co-Founder of The Opportunity Project at the U.S. Census Bureau. Previously, Drew was a Policy Advisor at the White House Domestic Policy Council and senior analyst at the U.S. Department of Commerce's Commerce Data Service. Throughout her career, Drew has worked with communities to use data, science, and technology to support economic development and leveraged the tools of human-centered design to improve how federal agencies serve the public. Drew holds an M.P.P. from Johns Hopkins University and is a Ph.D. candidate in social policy at Brandeis University. She has been recognized as one of the world's 100 Most Influential Young People in Government for 2018 by *Apolitical*, one of 20 Rising Stars by *FCW/GCN's Government Innovation Awards 2018*, and a Federal 100 winner by *FCW (Federal Computer Week)* in 2019.

Review Methodology

To analyze applications, both qualitative and quantitative methods were adopted including thematic readings on topics, NLP methods (e.g., word embeddings, pattern matching, latent Dirichlet allocation (LDA), and non-negative matrix factorization (NMF) topic models), as well as a market basket analysis to investigate the co-occurrence of UN's SDGs with each application. These methods are briefly described below, along with additional resources.

Extracting Data Science Techniques from Applications

To extract data science terms from applications, this analysis depended on both a manually curated list of common terms uncovered in the applications, as well as using common web-scraping methods. More specifically, using text from the application's "Proposed Data Science Solutions" section, this analysis first uncovered terms using Term Frequency Inverse Document Frequency (TFIDF) vectorizer, which was then manually reviewed to identify salient data-related techniques. To expand this corpus, data were also scraped from **Stackexchange.com**⁶ using the publicly available API, which gathered tags⁷ from posts between October 2018 to October 2020.

⁶ Stackexchange.com is a network of 173 question-and-answer websites, including Stackoverflow, a trusted online community for developers to learn and share their knowledge. The website provides a large corpus of available and current data science terms.

⁷ Each Stackexchange post is accompanied by tags, namely categories and data science techniques, which were used as a dictionary for the purpose of this analysis.

Forming a dictionary of 230 data science terms, this list was then compared against the 1,263 applications to determine whether each application evidenced any of the data science terms, in addition to similar terms via word embedding and fuzzy matching. Only words with similarity above 90% were counted in the analysis, which were manually reviewed to ensure quality and relevance to the application field.

Building an LDA Topic Model to Better Understand Applications

The second analysis explored the topics across application text—specifically applicants' Problem Statement—using an LDA and NMF algorithms, methods for unsupervised classification of documents. LDA conveys abstract topics across the $n = 1,241^8$ documents (i.e., applications) and generates words based on their probability distribution in the text (see Blei, Ng, Jordan, 2003 for overview). NMF, on the other hand, decomposes high-dimensional document term vectors into a lower-dimensional representation of documents by topic (see Arora, Ge, Moitra, 2012 for overview).

⁸ Twenty-two applications were submitted in languages other than English, which were excluded from this analysis.

For both algorithms, the application text was prepared using standard pre-processing and text cleaning, including the removal of punctuations and stopwords. Next, using the textacy package, a document term matrix was created and a range of values for the number of topics was analyzed. The best results were produced with NMF using 15 topics,⁹ which is displayed in Table 1.

Table 1: Application Topics Using NMF Algorithm

Problem Topic	Terms
Topic 0	people economic covid-19 need social
Topic 1	business small entrepreneur micro owner
Topic 2	job worker skill wage employer
Topic 3	farmer food smallholder farm agricultural
Topic 4	child care poor education family
Topic 5	woman gender girl man violence
Topic 6	datum science data decision collect
Topic 7	community local leader rural challenge
Topic 8	city waste urban live area
Topic 9	try problem address impact poverty
Topic 10	financial credit service access sme
Topic 11	housing neighborhood affordable wealth black
Topic 12	student school education learning teacher
Topic 13	youth unemployment young africa skill
Topic 14	health healthcare care patient mental

⁹ Topic modeling requires choosing an arbitrary number of topics. For this analysis, the selection of 15 topics yielded the best performance.

Market Basket Analysis

To understand the distribution of UN’s SDGs, this analysis used both manual tagging of applications according to relevant SDG and a Market Basket Analysis (MBA). More specifically, a team of 400 judges tagged each application according to which of the 17 SDGs were applicable. The distribution of SDGs is shown in Figure 1.

Note, however, that each application could touch upon more than one SDG. Thus, this analysis sought to understand the distribution of SDGs, namely how each co-occurred across applications as a set of “transactions.” Using an MBA, the conditional probability of each SDG co-occurring in each available instance was calculated, which is displayed in Figure 2 on the following page.

Figure 1: Applications by Primary Sustainable Development Goal



Figure 2: Conditional Probabilities According to SDG Pairings (>0.5)

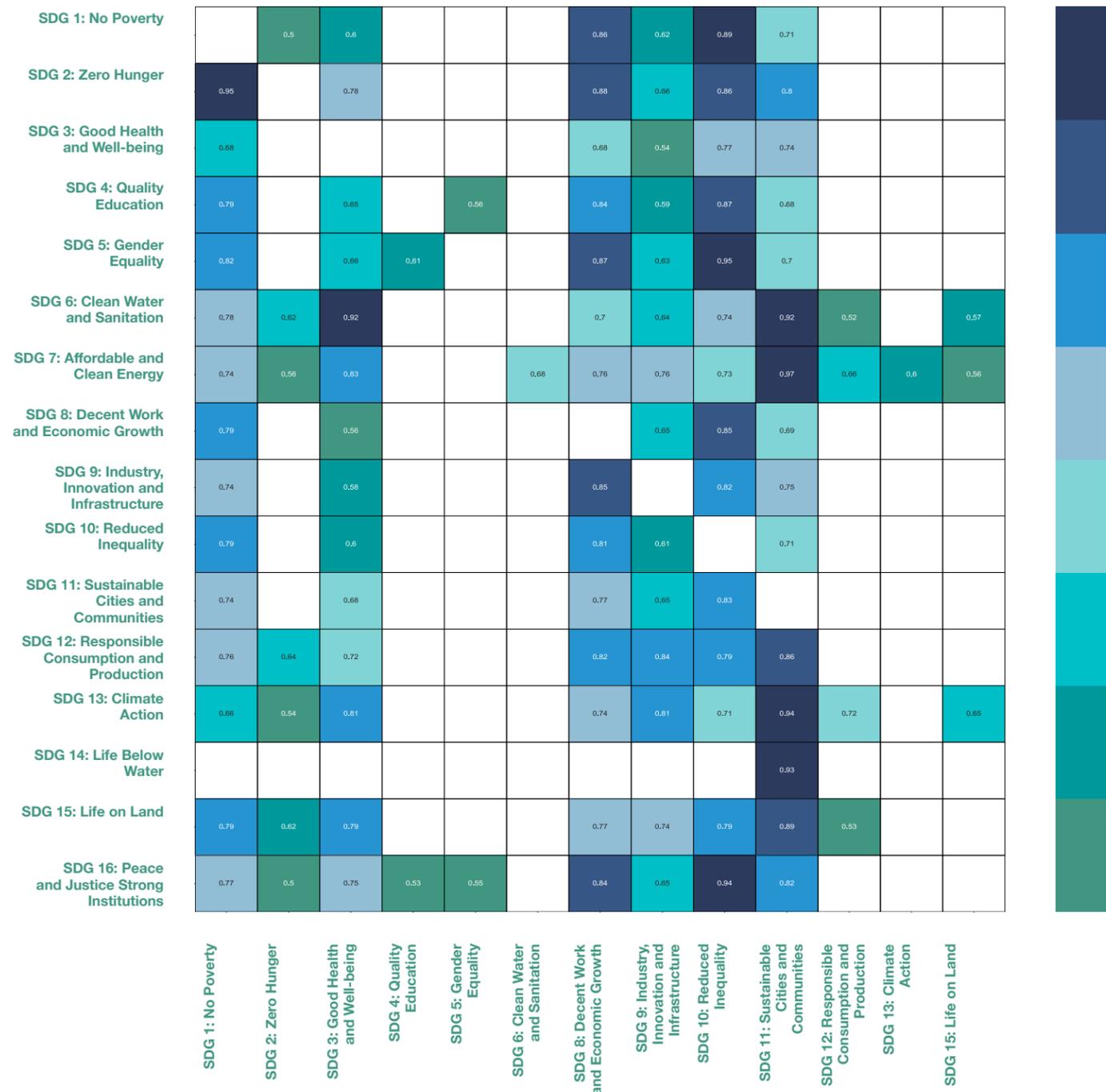
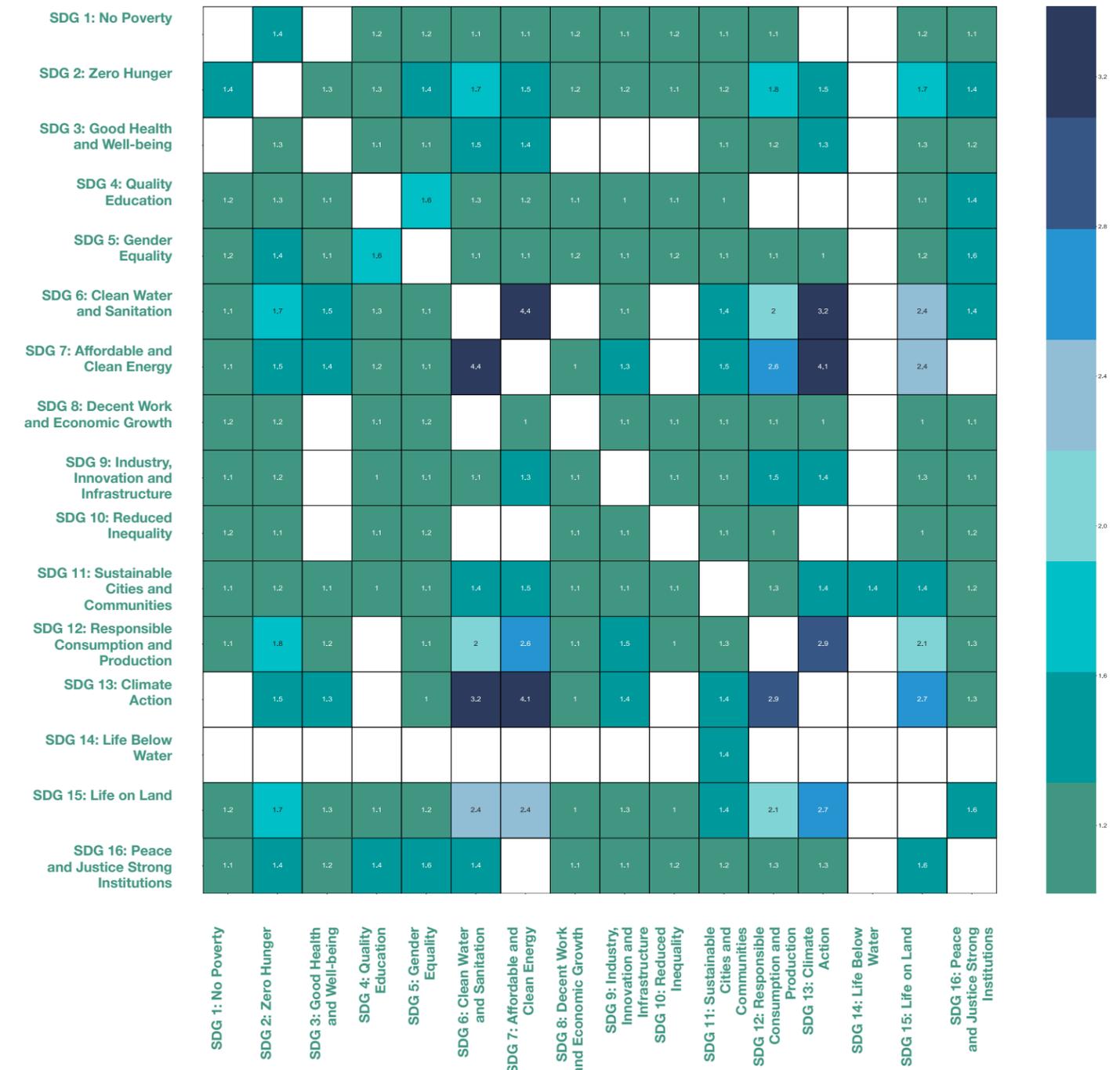


Figure 3: Lift Metric for SDG Pairings (>0.9)



The above Figure 2 reports on the conditional probability of each SDG (x-axis) co-occurring with another SDG (y-axis). For example, the probability that SDG 10, Reduced Inequality, co-occurs with SDG 5, Gender Equality, is 95%. On the contrary, SDG 10 co-occurs with Clean Water and Sanitation is less probable, just 74%. Therefore, this analysis delineates the ways in which SDGs and their common pairings are distributed across application submissions.

Challenge Application

Building off this analysis, the Lift metric was calculated to determine statistical independence.¹⁰ That is, Lift is the factor by which the co-occurrence of SDG X and SDG Y exceeds the expected probability of X and Y co-occurring.¹¹ A higher Lift score (i.e., greater than 1) suggests a higher probability of the particular SDG pairs occurring together. On the previous page, Figure 3 reports on the Lift metrics on all SDG pairings.

In Figure 3, we also observe that SDG pairings display varying dependencies, denoted by their Lift measure. For example, there is a strong association (i.e., 4.1) between SDG 7, Affordable Clean Energy, given SDG 13, Climate Action. Note, however, that more common SDGs yielded smaller Lift scores, such as SDGs 1, 5, 8, and 10. That is, more ubiquitous goals across applications implicate a higher denominator in the Lift measure. The appearance of these SDGs, which all align with the Challenge theme of inclusive growth, occurred most often to get noticeably high values. Put differently, SDG 10, Reduced Inequality, is such a common SDG, insofar that it never uniquely co-occurred with any of the other SDGs. Figure 3 reveals that 1.2 is the highest Lift score SDG 10 achieves with any of the other SDGs.

The pages that follow exhibit the Challenge application form for each phase of the online application process including instructions, questions to be answered, submission requirements, and notations for applicants.

Application: Phase One—Open to All Applicants

The data.org Inclusive Growth and Recovery Challenge seeks submissions focused on ensuring that the benefits of a rebounding economy extend to all segments of society. Please see the data.org website for detailed descriptions of primary topic areas: Jobs of Tomorrow, Access to Capital, and Cities and Towns. We welcome other proposals for using data science to advance shared prosperity, and help ensure an inclusive recovery.

1. Profile information:
 - a. Individual/lead organization name
 - b. Individual/lead organization website
 - c. Other members of the project coalition (optional)
2. Entity type/tax status (for entities only)
 - a. For-profit business entity
 - b. NGO or charitable organization located in the U.S. 501(c)(3) public charity (or equivalent)
 - c. NGO or charitable organization located outside the U.S.
 - d. Government agency
 - e. Multilateral organization (e.g., UN System Organization, the World Bank, and IMF)
3. Country: (dropdown—country list)

4. Size of annual operating budget in USD: (checkboxes—select one) (for entities only)
 - a. Under \$50K
 - b. \$50-100K
 - c. \$100-500K
 - d. \$500K-1M
 - e. \$1-5M
 - f. \$5-20M
 - g. \$20-50M
 - h. Above \$50M
5. What is the gender make-up of your executive/leadership team? (Fill in the blank—numbers only) (for entities only)
 - a. Transgender, Non-Binary, Women, Men, other
6. What is your project title:
7. Which challenge topic is your project primarily associated with? You can read more about the topic areas for the data.org Inclusive Growth and Recovery Challenge here: <https://www.data.org/challenge> (checkboxes—select one)
 - Leave No Worker Behind—Jobs of Tomorrow
 - Leave No Entrepreneur Behind—Access to Capital
 - Leave No Place Behind—Cities and Towns
 - Other (enter text)
8. Project stage:
 - a. Idea: innovations that require research, prototyping, and early testing to assess viability
 - b. Pilot: innovations at this stage are early in their implementation and require testing to understand feasibility
 - c. Scaling: innovations at this stage move beyond pilots and integrate innovations into existing systems to reach larger scale
 - d. Replication: innovations at this stage can be adapted laterally to different contexts—such as sector, geographies, and implementers
9. Project will focus on this primary sector: (select one)
 - a. Education
 - b. Health
 - c. Environment
 - d. Social Services
 - e. Transportation
 - f. Energy and Environment
 - g. International Development
 - h. Public Safety
 - i. Economic Development
 - j. Gender Equality
 - k. Other (write in)

¹⁰ Statistical independence is the $P(A \text{ and } B) = P(A) \times P(B)$. Therefore, the Lift metric is $P(A \text{ and } B) / (P(A) \times P(B)) = (P(A) \times P(B)) / (P(A) \times P(B)) = 1$.

¹¹ The Lift measure indicates the confidence that SDG X will occur given SDG Y, revealing how much better a “rule” is at predicting the result. It is calculated as the observed support—how frequently the item appears in the dataset—to that expected if SDG X and Y were independent.

10. Geographic scope of your proposed solution: (select one)

- a. Local (sub-state level)
- b. State/Province/District (subnational level)
- c. Regional (multi-state, subnational level)
- d. Country (national level)
- e. Multi-national

11. Primary initial implementation geography of your proposed solution: (select one)

- a. Asia Pacific
- b. South Asia
- c. Sub-Saharan Africa
- d. Middle East and/or North Africa
- e. Latin America
- f. Europe
- g. U.S./Canada
- h. Global/Multi-continental

Section 2: Free-Text Questions

Each question box has a maximum submission of 2,000 characters. For the following questions, please provide your best estimates below. Projects that advance to subsequent rounds will be required to prepare a comprehensive financial and in-kind support request. You will have the opportunity to fully revise support requests and not be held to these estimates.

- 12. What problem are you trying to address? Who is most impacted by this problem?
- 13. How will you solve the problem you described?
- 14. What resources do you have available in-house to complete the project, and what resources do you need to implement your proposed solution?
- 15. What is being done currently to address this problem? How and why is your solution better?
- 16. How would you define and assess the success of your proposed solution?
- 17. Why are you/your team uniquely positioned to solve this problem/implement your proposed solution?
- 18. What data science applications are you using or do you anticipate using in your proposed solution (e.g., predictive modeling, NLP techniques, and time series)?
- 19. How do you plan to ensure the technological/algorithmic accuracy of your proposed solution, and how will you maintain ongoing accuracy of your proposed data science solution?
- 20. Describe the data source(s) you have or need this project to succeed. If you do not currently have data, describe your process for obtaining data for your project and estimated timeline to procure.
- 21. How do you approach the ethical collection, management, and use of this data including data security, privacy, integrity, and transparency?
- 22. What is your high-level process and timeline to design and implement your proposed solution?
- 23. Who else is critical to your project's success? How will you go about building those partnerships?
- 24. (optional) Upload additional images, graphics, or other visual material to provide additional context. One image per upload. PNG, JPG

Licenses to Entries

By submitting an Entry to the Challenge, you automatically grant NVF, DataKind, and the Founding Partners (as defined in the data.org Challenge Terms and Conditions) (together, the “Challenge Partners”) a perpetual, worldwide, royalty-free license to use and display your Phase 1 Entry, and to share such Entry with, and sublicense such Entry to, any data.org partners for charitable and non-commercial purposes only, such as research and informing opportunities for their charitable initiatives (the “Phase 1 License”).

By submitting an Entry for Phase 2 of the Challenge, you hereby grant the Challenge Partners a perpetual, worldwide, royalty-free license to use, modify, and display any part of your Entries for Phases 1 and 2 of the Challenge, in any way and through any media whatsoever, including, but not limited to, on the data.org website, in data.org press releases or promotions, and through sharing with data.org partners, for charitable and non-commercial purposes only (the “Phase 2 License”).

If you do NOT wish to grant the Challenge Partners a license to your Phase 1 Entry, you may opt-out below, and your choices will not affect your Entry's chances for award selection. However, please note that if your Entry is selected to advance to Phase 2 of the Challenge, you must grant the Challenge Partners the Phase 2 License in order to continue participating in the Challenge.

Even with the licenses above, NVF, DataKind, and the Founding Partners will not share your Entries or any of the information or materials contained therein without first notifying you of the same and providing you with an opportunity to decline.

I/We do NOT grant the Challenge Partners a Phase 1 License.

Privacy

data.org believes in the ethical collection, management, and use of data. We also endorse data responsibility, innovation, and social impact. A commitment to data security, privacy, accountability, integrity, and transparency means we can better navigate the challenges and opportunities of the digital economy. These principles are meant to complement—and not substitute—regulatory compliance. Read our full privacy policy on our website.

During Phase 3 of the Challenge, each Finalist will undergo a data scoping exercise with DataKind. This exercise will require each Finalist to share with DataKind any and all data relevant to each Finalist's project, subject to the terms of a Data Sharing Agreement and Nondisclosure Agreement DataKind and data.org will execute with each Finalist prior to requiring any data to be shared.

Please read the complete Challenge Terms and Conditions prior to submitting your Entry.

Application: Phase Two—Open to Selected Applicants

Overview & Impact

- 1. We are interested in how entities/individuals can use data science to transform how they achieve social impact. How would this project make a unique contribution to your mission and values?
- 2. Please provide 3 examples of impact from the last 5 years of your entity's/individual history including dates and metrics.

Team

3. Who are the key members of your team, and what is their role in implementing this project? Are there gaps or unhired roles on the team? If so, how do you intend to fill those gaps or unhired roles?

Funding/Support

4. What is the amount requested to fund this project? What other (non-monetary) support do you need to make this project successful? For reference, examples of non-monetary support that may be available to awardees include media production, marketing, and promotional outreach.
 - Please also attach the project budget. A budget template is available.
5. What technical assistance would help deliver on this project and strengthen your work/entity? For reference, examples of technical assistance that may be available to awardees include technical support & consulting, software & infrastructure licenses.
6. Are you interested in pro-bono data science support?
 - Yes
 - No
7. What is the requested grant/support period for this project (up to 2 years)? (100 characters max)
8. Document your current major funding sources and the status (prospect, committed, or secured). (1000 characters max)
9. Have you received funding for this project or a similar project? If so, how much and from where? (1000 characters max)

Landscape

10. What individuals/entities do similar work in your field? What makes you/your entity best positioned within your landscape/field to be successful?
11. If this project is successful, who else might it influence within the landscape/field in which you operate?

Section 2: Project Information

Executive Summary

Write an executive summary of the proposal, including a brief description of the project, target outcomes, who the project serves, and why it matters.

Partnerships

1. Who are the key members of your team, and what is their role in implementing this project? Are there gaps or unhired roles on the team? If so, how do you intend to fill those gaps or unhired roles?

Funding/Support

2. Are you working with project partners/collaborators? If yes, please provide information about your relationship including historical collaboration and how you anticipate working with listed partners/collaborators on this proposal.
 - a. Please provide a letter of commitment for each partner indicated above. Letters should be on letterhead, dated, and signed. Note that letters of commitment will not be made public. If you are unable to obtain a letter of commitment from a partner, please provide context in your response above.

Community Engagement

3. How have you engaged, or how will you engage, the intended users or beneficiaries in this project from creation to implementation?

Practical

4. Please provide a detailed timeline including key deliverables and milestones you will use to measure progress during the project period. (option to upload attachment)

Impactful

5. If your project is successful, what social impact will be achieved? How will you measure progress toward achieving that impact? Please be as specific as possible.
6. If this project is successful, what are some potential unintended negative consequences? How would you mitigate against these negative consequences?
7. data.org is serving as a platform for partnerships to build the field of data science for social impact. We are achieving this mission through three strategic pillars: proving the **case** that data science can be the driver for social impact, building **capacity** of nonprofit organizations and governments to use data science to improve how they serve their constituents, and transforming the public **commons** with open-sourced goods that make data science for social impact achievable and possible for everyone. Please share any of the ways in which your solution would most advance **one** of these three pillars.

Sustainable

8. data.org focuses on funding sustainable solutions. If selected as an awardee, how will you ensure that the impact of your project is sustained beyond the length of support and funding? What milestones do you intend to achieve beyond the duration of data.org's support?

Breakthrough

9. The data.org Inclusive Growth and Recovery Challenge is seeking solutions that demonstrate an insightful and new application of data science. How does this project represent a breakthrough data science shift for you/your entity and the field?

Data Science

10. Please describe the data that you plan to use for this project and how it is structured. Please describe Data Type (e.g., images, text, and videos), Size (e.g., number images or rows), Attributes (e.g., images, image metadata, image labels, and how frequently data is refreshed).
11. Tell us about how you would leverage the data. What data would your project consume, and what data would it produce?
12. How will you deploy data science in the real world to address the problem? Detail any relevant partnerships or considerations.
13. Why is data science the best way to solve this problem? How would you solve it if you didn't use data science?
14. All datasets are biased in some way. How is your dataset biased, and what's your plan to mitigate the impact?
15. Will you be issuing your work under an open-source license? If not, what barriers do you see to providing your work as open-source, and what conditions would you need to make an open-source license feasible?

Section 3: Additional Information/Due Diligence

data.org is a fiscally-sponsored project of the New Venture Fund (NVF), a U.S. 501(c)(3) public charity. NVF undertakes a robust due diligence process to assess potential grant recipients' organizational and financial health to ensure that an organization has the ability to manage grant funding and successfully execute their planned project.

Please upload the documents requested below. If your organization does not have the requested documents, please contact challenge@data.org and we can discuss alternates. These documents will not be shared nor made public. If you are applying as part of a group, only the organization/individual who will receive the funding will need to submit documentation. data.org/NVF may also ask for further information as needed.

Note that if you answer "Yes" to question 1, you do not need to resubmit the following documents, unless something has changed: Proof of legal status (Question 2), Audited Financials (Question 6), or Form 990 (Question 7).

1. Has your organization applied for a grant from the New Venture Fund within the past year?
2. **Proof of legal status** IRS determination letter, 501(c)(3) equivalency determination, or an official document demonstrating charitable status in country.
3. **Names of key leaders** (Board of Directors, Trustees, and/or Senior Executives) Note: Per U.S. law, the leaders of non-U.S.-based organizations will be subject to an Office of Foreign Assets Control (OFAC) check.
4. **Organizational budget for 2020, 2021, and 2022.**
5. **Current year financial statements:** Balance Sheet (Statement of Financial Position) and Profit & Loss Statement (Statement of Activities).
6. **Audited financial statements** (last two years available). If unavailable, please provide internal unaudited financials statements for the two most recently completed fiscal years.
7. **Form 990** (last two years available) for U.S. based organizations.
8. **Confirmation of access to legal counsel.** Required for all organizations operating above \$1 Million USD annual budget.
9. As a finalist, you are entitled to receive \$2,000 USD in funding. In order to process this payment, we will need to collect a W-9 (U.S. applicants) or W-8BEN (non-U.S. applicants).
10. Additionally, to process your \$2,000 USD award, we will need you to submit the New Venture Fund Electronic Funds Transfer Form. The form is available for download in the "Quick Links" section of the platform.

Acknowledgments

Volunteer Judges

We are so grateful to the nearly 400 judges from around the world that volunteered for thousands of hours to help identify the most promising projects among Challenge applicants. We also want to recognize the following groups and their employees for providing critical volunteer support: Mastercard Center for Inclusive Growth, University of Washington, the World Bank, Global Data Barometer, Latin America Open Data Initiative, and Mobile Web Ghana.

DataKind Volunteers

This analysis would not have been possible without the valuable contributions of our volunteer data science community. Matt Harris conducted rigorous analyses using a number of NLP techniques, uncovering critical insights for the purpose of this analysis. Furthermore, Manojit Nandi leveraged his expertise and contributions to perform a Market Basket Analysis of SDG co-occurrences, informing the ways in which these applications were thematically aligned with the data.org Challenge and, broadly, the field of inclusive growth.

Report Contributors

This report and the work of the Challenge would not have been possible without the dedication and efforts of team members: Afua Bruce, Benjamin Kinsella, Caitlin Augustin, Chapin Flynn, Costanza Sciubba, Elana Stern, Evan Tachovsky, Gaia Orain, Ginger Zielinskie, Jackson Thomas, Kat Townsend, Kiwako Sakamoto, Lindsey Gottschalk, Melissa DiLoreto, and Perry Hewitt.



DataKind[®]

@DataKind datakind.org

data.org

@DataDotOrg data.org