Final Report

Columbia-Lehman Public Interest Technology Data Science Corps (PIT-DSC)

PIT-UN Network Challenge 2021 Cohort

PI: Tian Zheng (Columbia University) and Jennifer Laird (Lehman College)

Results on Goals and Objectives

Project GOAL: Columbia University and Lehman College propose to implement a first step towards a Public Interest Technology-Data Science Corps (PIT-DSC).

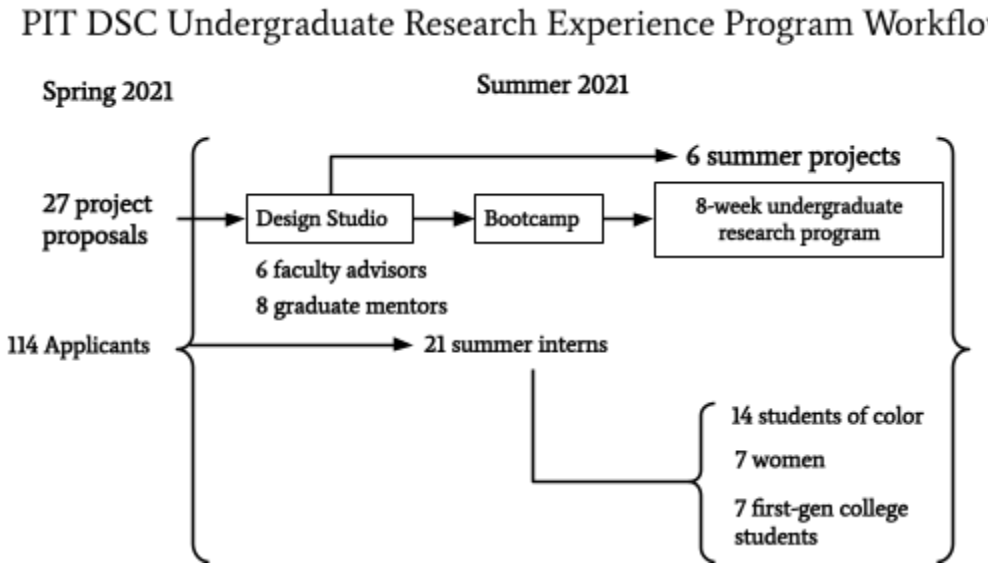
Project Results: In 2021, we successfully launched the proposed PIT-DSC program as a partnership between Columbia University and Lehman College, established partnerships with local organizations and agencies, and engaged faculty in mentoring students in working on public interest technology data science projects. We provide more detailed project results under each of the following originally proposed project objectives.

Project objective 1) pilot a summer PIT-DSC Scholars Undergraduate Research Experience (URE) program in partnership with Town+Gown:NYC (a city-wide university-community partnership program).

Project Results: In Spring 2021, 27 proposals were sourced from community organizations via Town+Gown:NYC (a city-wide university-community partnership program). Based on the evaluation of project impact, data science needs, and project readiness, six proposals were selected to proceed. Unselected projects were invited to work with graduate students in Statistics at Columbia to either further develop their project proposal or explore potential data-driven solutions.

Out of 50+ applicants, eight graduate student mentors were selected and participated in a 6-week “design studio.” The graduate students worked with NYC public interest project owners, turning the selected proposals into project designs for the 8-week Undergraduate Research Experience (URE) program.

A total of 114 undergraduate students applied for the URE program at Columbia and Lehman. These applicants were evaluated for their potential contributions to the PIT DSC projects. 21 undergraduate PIT interns (10 Columbia, 11 Lehman) were selected.



January - March	Program initiation; Project sourcing
April - May	Project selection; Summer intern recruitment and selection; Mentor recruitment and selection; Faculty engagement
June	Design-studio project development; Project owner engagement; Mentor training
July - August	PIT-DSC summer program: Bootcamp; Project check-ins; Mentor training; Proposal development for expansion; Partner engagement; Evaluation surveys
September - December	~~Fall poster share; Focus group discussions;~~ Project owner and partner engagement; Curriculum development discussion

A bootcamp was offered on ethical data science, team building, and PIT project management.

Project objective 2) Deepen collaboration between Columbia and Lehman to improve diversity and inclusion in data science pipelines.

Project Results: There are several ways the Columbia-Lehman PIT-DSC collaboration is deepening and diversifying the data science pipeline. After the completion of the 2021 summer program, 13 of the undergraduates expressed interest in pursuing PIT careers; 5 expressed interest in pursuing PIT graduate education. Two of the 2021 Lehman PIT-DSC interns, both first-generation students of color, have since been accepted into the Columbia Statistics MA program with a full scholarship. After completing the

PIT-DSC program, another Lehman student was later hired as a full-time data scientist (before graduating from Lehman). Lehman students have since been recruited into Columbia's Summer Training for Aspiring Researchers (STAR) program, a new summer educational program for students who aspire to do research in statistics and data science. After participating in the 2021 PIT-DSC program, a PIT-DSC alum organized the Lehman Data Science Association (DSA), a student-led club at Lehman. Another example of the deepening collaboration is that PIT-DSC alums from Lehman are giving back to the next cohort of Columbia and Lehman students. Two Lehman students from the 2021 cohort have been hired as bootcamp instructors for the 2022 PIT-DSC program.

Project objective 3) Connect with stakeholders and partners for scaling up the PIT-DSC community: data scientists as mentors, non-profit organizations as project sponsors, and institutions and faculty as collaborators.

Project Results: The connections and infrastructure we established are best illustrated by the PIT-DSC Summer 2021 projects with local community organizations and agencies, their participating students, and mentors.

Brownsville Partnership - Using existing data to analyze home problems and their possible correlation with health problems in Brownsville, Brooklyn.

Graduate Mentor: Victoria Cui (Columbia); Interns: Alix Leon (Lehman) and Amanda Lopez (Lehman)

Bronx Council for Environmental Quality - Using data science techniques to analyze and graphically represent the connections between the Bronx's demographic and environmental conditions in order to identify environmental justice policies and help stakeholders and decision-makers make informed policy decisions for the Bronx.

Graduate Mentor: Zi Fang (Columbia); Interns: Kayla Bernard (Lehman), Gabriel Fernandez (Columbia), Ashe Lewis (Columbia), and Eusebia Vazquez (Lehman)

Park Spending - Department of Citywide Administrative Services (DCAS)- Create an interactive map of NYC park spending across neighborhoods by zip code and neighborhood

Graduate Mentor: Ellen Chen (Columbia); Interns: Matias Alvarez (Columbia), Eric Aragundi (Lehman), Nicki Camberg (Columbia), and Henry Ovalle (Lehman)

Energy Management – Department of Citywide Administrative Services (DCAS) - The housing project focuses on unsubsidized multi-family housing available on the private market in medium-density neighborhoods. Participants will create a new algorithm that can be updated periodically to adjust what is considered affordable.

Graduate Mentor: Amber Zhang (Columbia); Interns: Davar Archibald (Lehman), Anthony Illescas (Lehman), Kyle Neary (Columbia), Catherine O'Brien (Columbia)

STEM programs - Science for New York (Sci4NY) - A look into inequalities in STEM programs in NYC public schools. Supplemental analyses will focus on teacher satisfaction with resources and testing scores.

Graduate Mentor: Peter Kwuak (Columbia); Interns: Madeline Angelides (Columbia), Bryan De Los Santos (Lehman), Nathaniel Lowe (Columbia), Goodness Martins (Lehman)

Parsing PDF reports - Science for New York (Sci4NY) - Generating a systematic method of parsing and reorganizing NYC community district reports pdf documents and district profile websites with an ultimate goal of visualizing science-related policies in NYC.

Graduate Mentor: Jen Wang (Columbia); Interns: Juan Alvarez (Lehman), Simon Aytes (Lehman), Blake Mills (Columbia), Nixon McKenzie (Columbia)

Project objective 4) Evaluate values of PIT Data Science URE in terms of learning outcomes, ethics, pipeline diversity & inclusion; identify stakeholders' interests and curricular elements for Public Interest Technology education.

Project Results: We identified the following learning outcomes when designing and reviewing the PIT DSC design studio and the summer URE programs.

Learning Outcomes of the PIT DSC "Design Studio"

- ☐ Students assist project owners to articulate their vision by formalizing their ideas with mathematical rigor and data.
- ☐ Students learn to set achievable and relevant milestones with a reasonable deadline for time-sensitive projects.
- ☐ Students learn to manage expectations around data science projects and learn to balance the perfection vs speed trade-off.
- ☐ Students learn to create reproducible models and documentation meant for a wide audience with different backgrounds.

Learning Outcomes of PIT DSC

- ☐ **PIT DSC interns** develop practical data science skills by solving real-world PIT problems and a deeper appreciation of the intended and unintended impacts of data-driven solutions.
- ☐ **Project owners** receive data-driven products for their projects, gain experience working with data scientists, and develop a deeper understanding of data science.
- ☐ **Graduate mentors** support the project to ensure learning experiences for students and the delivery of project outcomes. They will also develop skills in project management and communication

Pipeline Diversity and Inclusion

- ☐ The first cohort of PIT-DSC interns included 14 students who identify as Black, Hispanic, or Latina/o/xof color, 7 women, and 7 first-generation college students.
- ☐ Within Columbia, the PIT-DSC program has become a model, and as a result, the Columbia School of Engineering is now partnering with Lehman to fund a wide array of undergraduate research opportunities for under-represented students interested in using their STEM skills to build inclusive communities.
- ☐ The PIT-DSC program has also served as a launching pad for a *Wall Street Journal* data science internship that is targeted at under-represented students (two of the 2021 PIT-DSC interns secured the WSJ internship in the fall of 2021).

Expanding connections to stakeholders

- ☐ The PIT DSC design studio continued to work with local organizations and expanded beyond the organizations mentioned above. PI Zheng has also been working with Columbia's Data Science Institute on a new initiative called "Executive in Residence" that will explore training and pipeline opportunities that deepen connections between academic education programs and data science workforce in the public sector.

Curricular Elements for PIT Education

- ☐ PIT bootcamp curriculum - Following the 2021 summer program, we developed a 2022 bootcamp module on "work preferences." Drawing from academic research on cultural differences in management and organizational behavior, the module allowed teams to share and discuss their expectations related to communication and leadership.
- ☐ After the completion of the 2021 PIT-DSC program, CUNY funded a study on "Revealing the Hidden Curriculum in Data Science Research Programs." Led by PIT-DSC PIs Zheng, Laird, and Lehman sociologist Brittany Fox-Williams, the project is currently gathering diary, survey, and focus group data from the 2022 program participants. The data will be used to identify and disseminate the unwritten expectations, values, and norms that are the basis for success in the PIT-DSC program.

[redacted]

- ☐ [redacted]
- ☐ [redacted]

Lessons learned

- ☐ PIT research projects require ongoing development and strong ties between faculty, students, and project owners. The projects require constant reassessment. Learning goals and outcomes often have to be revised depending on data availability. There is

extensive variation across projects and project owners in their capacity to identify objectives and communicate with students.

- ☐ For all students to contribute, the work environment has to be inclusive. Achieving and maintaining an inclusive environment is particularly challenging during a pandemic and over Zoom.
- ☐ Students have to become self-driven collaborators (not interns for hire). Training in project management and “managing up” can help undergraduates assume the roles of collaborator and administrator.
- ☐ Technical expertise is not enough; faculty and graduate students have to learn how to motivate and manage interdisciplinary teams from different backgrounds. Discussions about best practices for engaging first-generation students can generate concrete strategies for building inclusive environments.

Media Generated

PIT DSC Website
Press Release of BCEQ
Columbia Neighbors Article (See attached)

Certification

All The Trustees of Columbia University in the City of New York activities conducted with the Grant funds were and are consistent with charitable purposes as set forth in Section 501(c)(3) of the Internal Revenue Code, and The Trustees of Columbia University in the City of New York complied with all provisions and restrictions contained in this Agreement, including, for example and without limitation, those provisions relating to lobbying and political activity.

Columbia Neighbors Article

Learn How Students Used Data Science to Explore Health, Wealth, and Education in the Bronx

By Victoria Benitez

September 09, 2021

A team of undergraduate students from Columbia University and Lehman College, CUNY, used sophisticated data visualization techniques to explore the correlation between environmental factors and statistical measures of school dropout rates, income, and health disparities as part of the Bronx Council for Environmental Quality's (BCEQ) research project, *Environmental Determinants of Health, Wealth, and Education*.

Over the summer, the undergraduates applied their data science skills to projects that focus on the need to address quality of life issues for New York City residents in underserved communities. Graduate students from the Columbia Department of Statistics served as mentors during the program.

The students presented their report via Zoom at a public BCEQ meeting on September 8. The report findings include:

- The dropout rate among public school students who live near highways in the Bronx is substantially higher than among those who don't.

- Bronx residents who live farther from parks suffer from more health problems than those who live near green spaces. Additionally, positive effects of living near a park are less significant if the park is dirty or in disrepair, as those parks go unused.

The completed report is a call to action for Bronx residents about the effects of bad environmental policy. “The series of maps produced by this research raise troubling policy implications for the maintenance and operations of parks, schools, and highway corridors,” said Dr. Robert Fanuzzi, president of BCEQ, an all-volunteer organization that is celebrating 50 years of environmental advocacy.

“Columbia is working to improve the public good by expanding and diversifying the Public Interest Technology data science workforce,” said Columbia Professor Tian Zheng, principal investigator for the Columbia-Lehman Public Interest Technology Data Science Corps (PIT-DSC). “Our partnership with Lehman and BCEQ advances our mission to create data science solutions that have a societal impact.”

Earlier this year, BCEQ submitted a proposal to the Columbia-Lehman PIT-DSC, which was facilitated by the Town + Gown program, a New York City-wide partnership between universities and local communities that bring academics and practitioners together to create plans for workable solutions to public issues. The proposal was selected for PIT-DSC’s inaugural summer undergraduate research program. The research was also supported by a grant from New America’s Public Interest Technology University Network.

As part of the project, BCEQ board members met weekly with the research team of students from various disciplines, including sociology, math, statistics, and computer science. Led by Zi Fang, a graduate student in statistics at Columbia, the students used datasets that incorporated Census data, New Yorkers for Parks reports, and city and state agency databases to come up with their findings.

“We are excited with the work of the students. We asked them to help us find out why the Bronx is still 62 out of 62 in the state for negative health outcomes,” said BCEQ Board member Karen Argenti. “The program helped BCEQ find a road to environmental justice in the Bronx. The numbers and their correlations tell a story of less funding, staffing, maintenance, and planning for open spaces, education workforce, and development in the borough. In other words, systemic racism. I think the students learned a lot, as well.”

The project was a success for everyone involved. “We are thrilled to be partnering with BCEQ and Columbia on a project that uses our students’ expertise to understand environmental conditions in the Bronx,” said Lehman Assistant Professor Jennifer Laird, co-principal investigator for PIT-DSC. “This project provides invaluable data science training for our students, most of whom are from the Bronx, and it provides stakeholders with the evidence they need to make informed policy decisions.”

Tags Public Health Local

List of intellectual property and assets created with the Grant

Design Studio Assets

- ☐ Templates for developing project proposals
- ☐ Starter codes and research designs
- ☐ Training materials for communication, collaboration, conflict resolution, expectation management, and identification of action items and deliverables

PIT DSC summer URE program assets

- ☐ Project submission template
- ☐ Project review criteria and process
- ☐ Student application form and review criteria
- ☐ Student recruitment info session materials
- ☐ Five-day bootcamp on ethical data science and responsible PIT collaboration, including slides, discussion, and team activities focused on solving social problems
- ☐ Mentor training through the design studio and weekly check-ins
- ☐ Project milestone report templates
- ☐ Evaluation surveys for students, mentors, and project owners

Project repositories

1. https://github.com/alixlm19/Brownsville-project
2. https://drive.google.com/drive/folders/145OJVgaNFk5SBph7YjGHv4Qxm20i4t_Q?usp=sharing
3. https://github.com/EllenChen5/NYC-Park-Budget-Analysis
4. https://github.com/Zhuohan-Amber/Affordable-Housing-Estimation
5. https://github.com/sci4ny/PIT-DSC-STEM-Inequality
6. https://github.com/jenn2325/PIT-DSC_Program