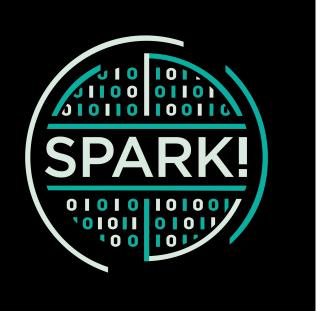# PITCH DAY - CS506 DATA SCIENCE

SPARK!

SPARK! MISSION

INNOVATIVE +
INTERDISCIPLINARY

INTRODUCING THE BU FACULTY OF COMPUTING
& DATA SCIENCES.

[EXPERIENTIAL LEARNING]

Powered by **SPARK!**

To help students discover their passion(s), find community, prepare for their future, and make **real-world impact** through **innovation and experiential learning.**

SPARK! TEAM

Ziba Cranmer
Director

James Grady
Creative Director (CFA)

Langdon White
Technical Director (SE)

Savannah Majarwitz
Associate Program Manager

Lydia Holck
Operations Director

Isabel Torres
Community & Outreach Mgr
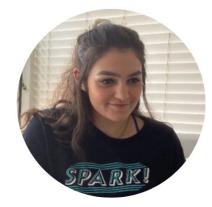
Thao Nguyen
UXD Program Manager

Ian Saucy
Assistant Solutions Engineer

Michelle Voong
Data Solutions Engineer

Other Spark! Support assigned to teams:

Expert In Residence*

Project Managers

Technical Engineers

*EIRs only for *some* teams

WORK FROM HOME - BU SARGENT COLLEGE

The mission of BU's College of Health and Rehabilitation Sciences: Sargent College is to advance, preserve, disseminate, and apply knowledge in the health and rehabilitation sciences. Client: Karen Jacobs, Associate Dean, Digital Learning and Innovation Sargent College

Project Description: The researchers used the ecological momentary assessment approach to conduct a 6-month study to assess the physical, cognitive, and mental effects of a remote work environment, on workers' well-being. 70 participants were recruited for the study. Participants were given a Garmin watch to wear which "pinged" them three times a day to share information such as their current location, musculoskeletal discomfort, and number of breaks. On Fridays, participants completed the E-Work and the Flourishing scale surveys. The computer workstation survey was completed once a month on Wednesday. Three months of data have been collected and researchers want to understand what insights can be gleaned so far as a benchmark to compare to six months of data when the study is completed.

Data & Sources: Three Month Data; Working From the Home Environment - DATA

Some questions to explore: Does the data prove or disprove the following predictions?
- Hypothesis: Participants' age will negatively correlate with financial and material stability
- Hypothesis: Participants working in healthcare will have lower mental health scores on the Flourishing Scale than those working in other industries.
- Hypothesis: Participants' stress algorithm will be inversely correlated to their number of breaks.

Expected Final Deliverable:
Create a report with key visualizations and answers to your client's questions. This report must include tables with the data and what statistics your team used to analyze the data. This research and data analysis will likely be used in preparation for presenting at a conference or manuscript. The research is funded by a grant from the Office Ergonomics Research Committee (OERC) to provide them with data and publish it widely.

WEATHER IMPACT ON WASTE - BU SUSTAINABILITY

BU Sustainability supports the transformation of Boston University's planning, operations, and culture toward a sustainable and equitable future.

Client: Sarah Healey, BU Sustainability, Zero Waste Manager

Project Description:

BU Sustainability has pulled multiple spreadsheets with data from their 3rd party vendor Contelligent and our waste vendor Casella. Contelligent provides PSI fullness monitoring of our compactors. The client has provided detailed data sets from each monitor, overall data, temperature data, and waste generation spreadsheet by data from Casella. The analysis of these datasets will inform BU Sustainability how they can potentially improve where to store waste if there are adverse weather effects.

Questions:

- Does temperature impact waste generation?
- If so, in what ways (i.e. more recycling, more of all materials, less recycling, etc.).
- Can we use temperature as a predictor of waste generation and service level requirements?

Expected Final Deliverable: Report with a data set for FY21 that overlays the data from the 2 vendors into a single file. Provide a comprehensive data sheet with a data analysis section that works to solve the questions provided by the client

Boston University Sustainability

BOSTON UNIVERSITY

BU Sustainability.

Understanding How Weather Impacts Waste

Zero Waste

Reimagining waste as a resource.

The conservation of all resources by means of responsible production, consumption, reuse, and recovery of products, packaging, and materials without burning and with no discharges to land, water, or air that threaten the environment or human health.[1]

- Rethink/Redesign
- Reduce
- Reuse
- Recycle/Compost
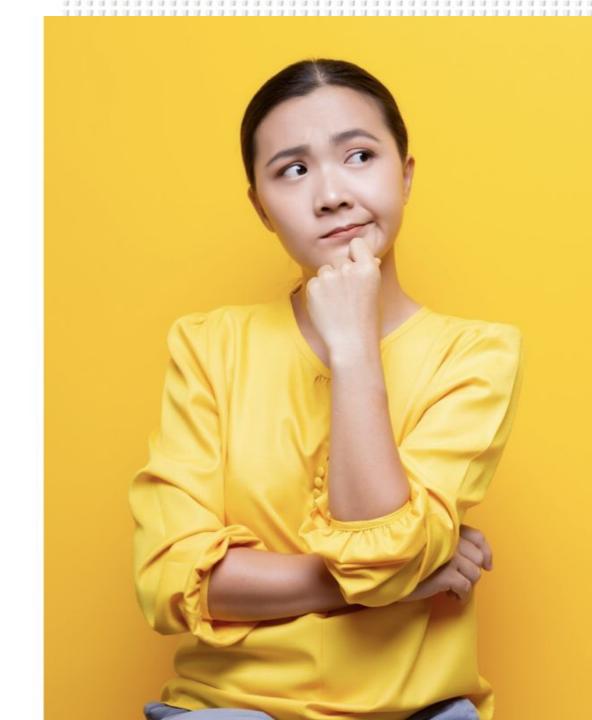- Culture Change
- Market Development

Internationally peer-reviewed definition of Zero Waste: zwia.org/zero-waste-definition/

Rethink

Using Data to Rethink how we manage waste

We collect data from monitors on our compactors. This data calculates how often the compactor gets used and the temperature when it gets used.

- Does temperature impact waste generations?
- If so, in what ways (i.e., more recycling, more total waste, etc.)
- Can we use temperature as a predictor of waste generation and service level requirement?

Sarah Healey, Zero Waste Manager
Srhealey@bu.edu
BU Sustainability
Facilities Management

Thank You.

BOSTON
UNIVERSITY

FAIR HOUSING DISCRIMINATION - BOSTON CITY COUNCILOR: RUTHZEE LOUIJEUNE

Client: As City Councilor At-Large, Ruthzee Louijeune is fighting for affordable housing and homeownership, support for our immigrant communities, improved funding and teacher diversity at our schools, and using all the tools at the city's disposal to address racial justice and equity.

Project Description: Councilor Ruthzee wants to understand if there is discrimination in lending/ banking for individuals seeking home ownership at the neighborhood level by examining who is participating in first time home ownership programs and who is receiving loans and purchasing homes.

    Extension: This is an opportunity to get creative with additional data that you think would tell an interesting story, one your team finds compelling. You will create a mid-semester proposal for this extension project that the client will approve.

Data & Sources: Parcel Database (MAPC), Redlining Data, Census/ Namsor

Some questions to explore:
- What are the patterns of home ownership over the past 5-10 years in Boston based on race/ethnicity and home type?
- Are there indicators of bias based on these ownership patterns?

Expected Final Deliverable: The final deliverable should be in the format of a written report (word) with conclusions and data visualizations, as well as details about the process of analysis generated from the base and extension project work.

SIDEWALKS - STANTEC

Client: Ramandeep Josen, Senior Asset Management Specialist, Stantec

Project Description: The goal of this project is to revolutionize how the City of Boston understands and assesses their ==sidewalk== network by developing an ==accessibility score.== This score will be created for each sidewalk block based on existing trip hazards, sidewalk obstructions, the existence and condition of pedestrian ramps, etc.

Data & Sources: Boston shapefiles (ramps, roadway centerline, sidewalk hazards), Public Works Active Work Zones Data, Climate Ready Social Vulnerability Data, 311 Requests, and Census Demographic Data

Some questions to explore:
- How can we leverage existing data to create an accessibility score for each sidewalk block in the City? Can we use this score to identify which sidewalks are the most inaccessible in the City?
- Can we create routes from residential parcels to the nearest important pedestrian destinations to see how many residents can get to these destinations without coming across any inaccessible features? How do these routes differ for citizens that live in different parts of the City?
- Based on all the findings above, how should the City re-strategize its sidewalk repair strategy to be most impactful?

Expected Final Deliverable:
- Development of an accessibility score which helps identify sidewalks that are currently inaccessible ==based on existing data for trip hazards and pedestrian ramp existence/conditions, crosswalks.==
- Develop a way to route from residential parcels to important destinations and quantify how inaccessible those routes are based on accessibility score.
- Create metrics/KPIs based on accessibility score and routing to important destinations for neighborhoods/districts for City officials to understand areas of City which are more inaccessible than others

CRADLE TO PRISON PIPELINE - NORTHEASTERN UNIVERSITY

<u>Client</u>: Erin Stewart, Skadden Fellow, Citizens for Juvenile Justice (CJJ)

Northeastern University - [Cradle-to-Prison Pipeline Project](#)

<u>Project Description</u>:

The Cradle-to-Prison Pipeline Research Team at Northeastern University conducted an expansive survey of incarcerated people in Massachusetts. The survey was designed to test the hypothesis that children of color are disproportionately impacted by an array of systems—child welfare, public education, mental health, school discipline, and juvenile justice—which operate to increase the likelihood of adult incarceration. The collection of data through this survey is intended to support advocacy efforts and inform policy decisions to help dismantle the cradle-to-prison pipeline and address mass incarceration in meaningful ways.

The goal of this project is to further analyze the extracted data and to answer the client's key questions, and to help clean and compile additional data on school-based arrests.
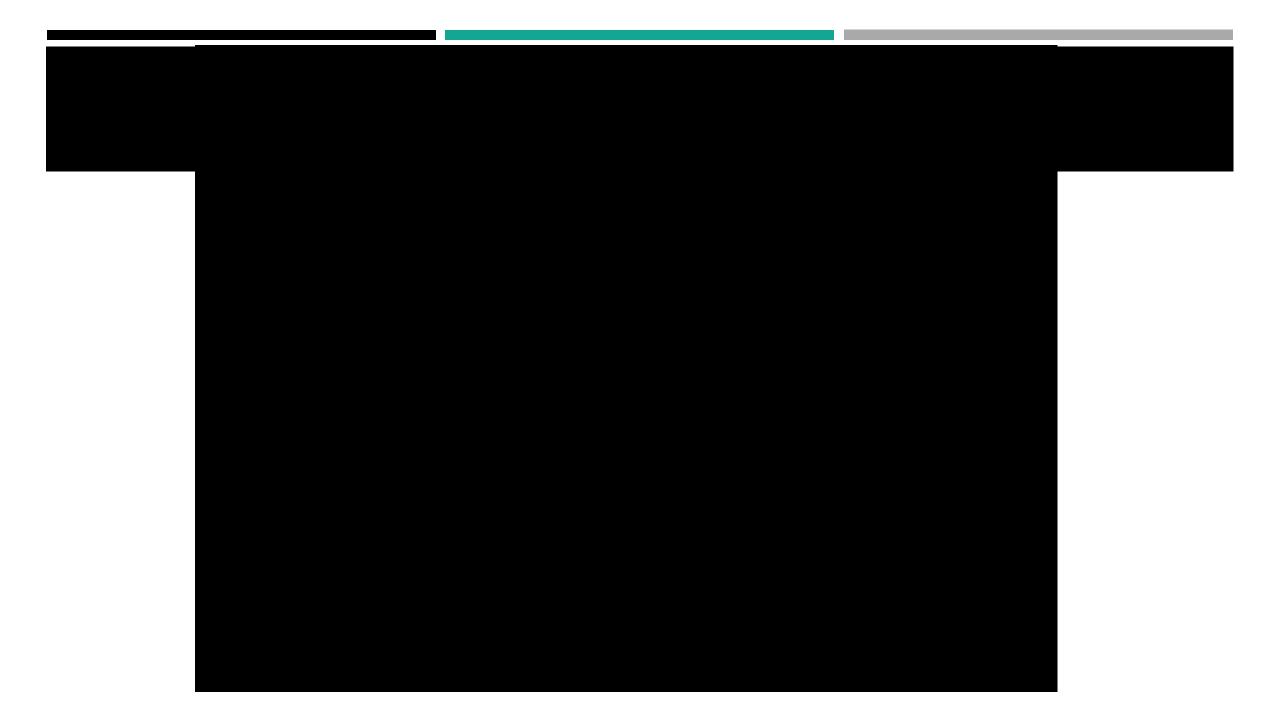
<u>Data & Sources</u>: School-based arrest data and project repos in Github created by Spark! students in previous semesters

<u>Some Questions to Explore</u>:
- How disproportionate are the instances of arrest based on ethnicity, school, etc?
- What percentage of survey respondents answered yes to "arrested before age 18" separated by race, gender, and sexual orientation?
- What percentage of survey respondents experienced school discipline (suspension/expulsion) while in school separated by race, gender, and sexual orientation?

<u>Expected Final Deliverable</u>:
Create a report with key visualizations and answers to your client's questions.

Content Warning:

The following project is centered around issues of gun violence

GUN VIOLENCE - BOSTON CITY COUNCILOR: BRIAN WORRELL

Client: Councilor Worrell, Boston City Councilman for District 4, which includes Mattapan, Dorchester, and parts of Jamaica Plain and Roslindale.

Project Description: Councilor Worrell seeks to understand violence including gun violence in the community first focusing on District 4 and then the city as a whole and its impact. There has been an uptick in gun possession amongst the youth which is concerning. This project seeks to understand what variables may be associated with increases or decreases in (police presence, population movements, greenspace, school performance, etc). Ultimately the project goal is to determine what drivers of violence looks like in District 4 as compared to other areas, but as a foundation, we need to understand the comparative volume and geographic distribution of gun violence in the district and beyond.

> Extension: This is an opportunity to get creative with additional data that you think would tell an interesting story, one your team finds compelling. You will create a mid-semester proposal for this extension project that the client will approve.

Data & Sources: Shapefile, Boston, Massachusetts Police Districts | GIS Map Data OR Districts — bpdnews.com, Police Records of Violence/Firearm Activity, Police reports, Police Presence

Some questions to explore:
- What is the current state of violence including gun violence in District 4? How does this compare to the rest of the city?
- Are there patterns of violence (ex. Locations, types of violence) in District 4? How does this compare to the rest of the city?
- What are potential drivers of violence in District 4? How does this compare to the rest of the city?

Expected Final Deliverable:
The results of this analysis will hopefully inform policies that can be beneficial for improving the district.

BAD LANDLORDS- BOSTON CITY COUNCILOR: LIZ BREADON

Client: Councilor Breadon, Boston City Councilor for District 9

Project Description: One of Councilor Breadon's missions is to increase the accessibility and transparency of the Boston planning, zoning, and development process so her constituents and residents across Boston can better understand the evolving landscape of their city. The office is interested in building a trackable system for property violations and a matrix to determine if a landlord is a bad landlord. The goal of this project is to figure out what ordinance is needed to define what threshold for bad landlord list e.g. X amount of violations in the calendar year or severity of violations associated with a given landlord.

      Extension: This is an opportunity to get creative with additional data that you think would tell an interesting story, one your team finds compelling. You will create a mid-semester proposal for this extension project that the client will approve.

Data & Sources: Building and Property Violations Dataset, Worst Landlord Formula, Property assessment fy22, Shapefile by district, Income-Restricted Housing Inventory - Datasets - Analyze Boston

Some questions to explore:
- What is the spectrum of violations and severity in regards to worst landlords classifications?
- What landlords are non-compliant? Overall volume, severe violations
- Which landlords are impacting vulnerable communities the most (based on social vulnerability index/ climate resilience and college students)

Expected Final Deliverable:
Create a report with key visualizations and answers to your client's questions. The vision is to create the Boston version of this site 2022 Worst Landlord Watchlist

UNIVERSAL BASIC INCOME - BOSTON CITY COUNCILOR: KENDRA LARA

Client: Kendra Lara, Boston City Councilor, District 6

Project Description: This project seeks to understand what a <mark>universal basic income</mark> program should look like for Boston by looking at the number of people under the poverty line, what the different levels of poverty are, which public assistance programs are being utilized and the amount of assistance, and how much money per household could have the greatest positive impact. Boston City Councilor Kendra Lara has the opportunity to provide 1000 of these people with assistance so the analysis from this project can be used to help inform her of a potential selection process for.

     Extension: This is an opportunity to get creative with additional data that you think would tell an interesting story, one your team finds compelling. You will create a mid-semester proposal for this extension project that the client will approve.

Data & Sources: Public assistance programs: DTA Data and Research | Mass.gov, Shapefile, WIC data, SNAP policy database (food assistance), Section 8 Housing Assistance

Some questions to explore:
- What is the income distribution for Boston residents living poverty line
- How many people participate in assistance programs - focus on WIC, Snap, and Housing assistance (section 8)?
- How much assistance is provided to residents at different income levels (under poverty level)
- What amount would pull the largest number of households out of the poverty level without sacrificing existing assistance benefits

Expected Final Deliverable:
Create a report with key visualizations and recommendations/ answers to your client's questions.

ANALYZING 311 DATA - BOSTON CITY COUNCILOR: JULIA MEJIA

Client: Councilor Mejia, City Councilor, At-Large, is currently the Chair of the Committee on Education, the Chair of the Committee on Government Accountability, Transparency, and Accessibility as well as the Chair of the Committee on Labor, Workforce, and Economic Development.

Project Description: At-Large City Councilor Julia Mejia seeks to understand the relationship between empowered citizens and city services by examining the city's responses to 311 service requests, and if they are resolving them in an equitable manner. You will use the voter file and 311 data to understand which communities feel empowered in Boston to demand services and how the city responds to empowered residents across the city.
> Extension: This is an opportunity to get creative with additional data that you think would tell an interesting story, one your team finds compelling. You will create a mid-semester proposal for this extension project that the client will approve.

Data & Sources: Social Vulnerability Index, Dataverse 311 Data, 311 data for Boston from 2011-2022

Some questions to explore:
- Which communities are most empowered based on 311 service requests and voting participation?
  - Which geographies submit the most request? What is the profile of these communities, i.e. voting history and demographics?
- How does the city respond to its residents based on the ratio of resolved to unresolved 311 requests?
  - Are there differences in the rate of closure for different types of residents (e.g. high voting. high requesting, and demographics)
  - Are there differences in the types of service requests the same different communities?

Expected Final Deliverable:
Create a report with key visualizations and answers to your client's questions.

ATHLETICS PERFORMANCE ANALYSIS - BU ATHLETICS

BU Athletics is dedicated to excellence in athletics, academics and community. Over the long history of the department our student-athletes have represented BU formidably, achieving and maintaining a high level of excellence both on and off the field while at the same time displaying commitment, integrity and sportsmanship.

Client: Kristie Bowers, BU Athletics, Senior Associate Director

Project Description:
The BU Athletics department has conducted their own "in-house" analysis for years around student athletes academic performance. The goal of this project is to analyze the correlation between admissions data (i.e. test scores and high school GPA) and the performance of student-athletes throughout their careers at BU. By understanding potential correlations, the athletics department can learn more insights about the students they serve and support.

Data & Sources: BU Athletics dataset containing information about GPA information, demographics, major, etc

Some Questions to Explore:
- What is the range of accepted SAT/ACT scores and highschool GPA for student athletes?
- What is the range of BU GPA for student athletes?
- How do these students with high SAT/ACT scores or high GPA perform academically at BU compared to student athletes with low SAT/ACT scores or low GPA?
- Are there any significant differences in the academic performance of student-athletes based on their sport?

Expected Final Deliverable:
Create a report that answers the client's questions. Determine which visuals (i.e. linear progression, etc) the client would like to see in their report - be creative. The end goal for the client is to have information to share with the rest of the BU Athletics department and Admissions.

SW BID DISTRICT PROFILE - WASHINGTON DC CITY COUNCILOR CHARLES ALLEN

Council member Charles Allen represents DC's Ward 6 and works closely with the Southwest BID and the SWBID Partnership to solve problems for neighbors and build a thriving and equitable Southwest. Clients: Jeanne Mattison, DC Ward 6 Councilmember & Laura Marks, Southwest BID's Partnership
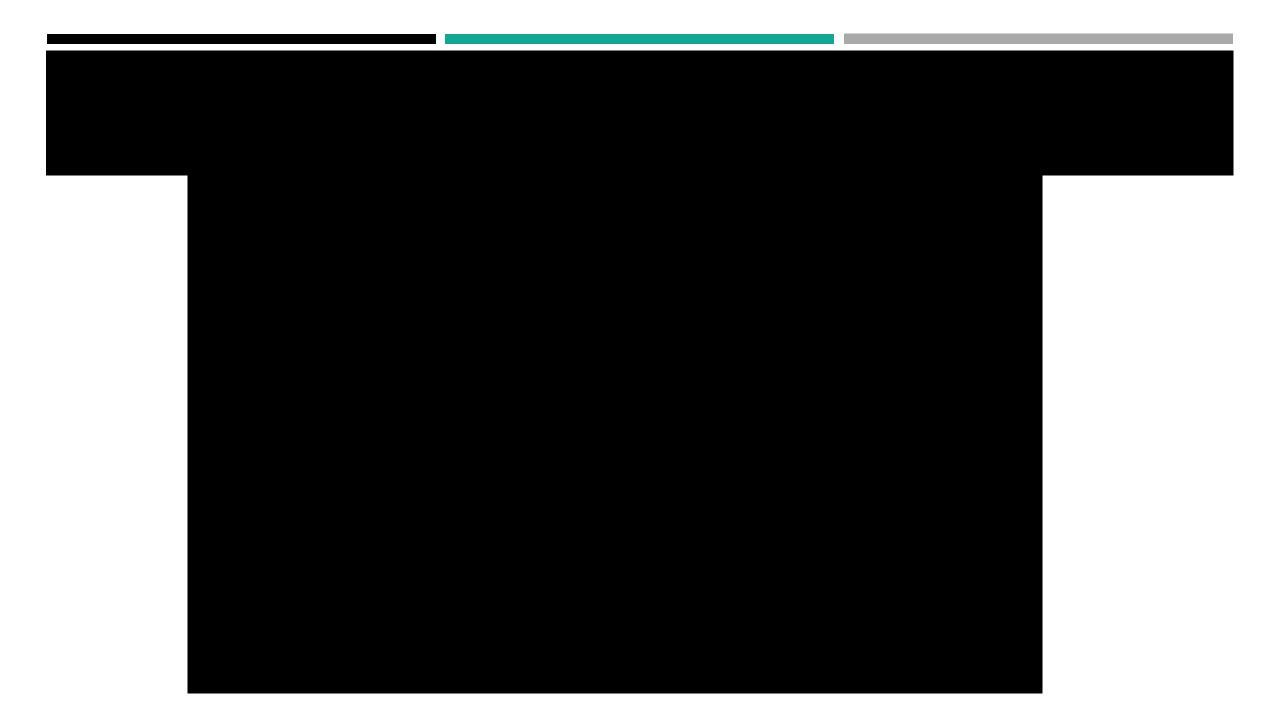
Project Description: Southwest DC is a neighborhood that has seen tremendous change over the last 60 years. Long home to a thriving African American community, Southwest was the first site of so-called "urban renewal," resulting in the displacement of that community and the destruction of 99% of all buildings in the quadrant during the 1950s-60s. Southwest is home to three large public housing communities which comprise the largest public housing community in DC. This area faces many of the common problems associated with concentrated generational poverty – high rates of unemployment, community violence, substance use, chronic disease, teen pregnancy, and food insecurity. These low income communities exist alongside an increasingly super wealthy areas which is confounding the demographic profile of the neighborhood overall. The goal of this project is to provide a more granular understanding of the demographics of the southwest district. [more information here]

Some Data & Sources: US Census Data, CDC Data, DC Racial Wealth Gap report, DC Open Data Catalog, and others

Some questions to explore:
- What is the demographic and quality of life profile of sub-communities within SW district, looking at the data at the census block group level? (e.g. race/ethnicity, age, income/ poverty, health equity, access to green space, violence, food insecurity, etc.)
- How have these indicators changed over the last 5-10 years to identify the rate of gentrification
- Which variables are most highly correlated with each other

Expected Final Deliverable: A comprehensive Data Dashboard with a racial equity and economic mobility lens would help confirm this understanding of the neighborhood, provide a clear image of the neighborhood's challenges, resources, and disparities, and help define a path forward toward a more equitable, inclusive, and thriving community.

POLICE BUDGET - ALCU

The ACLU of Massachusetts is a private, nonpartisan organization with more than 82,000 supporters across the Commonwealth and over 100,000 online activists—is a state affiliate of the national ACLU. We defend the principles enshrined in the Massachusetts Declaration of Rights, as well as the U.S. Constitution and Bill of Rights. [See video]

  Clients: Natasha Dhamankar, ALCU Tech Fellow & Kade Crockford, ALCU Director Technology for Liberty Program

Project Description: Every year a budget is approved for the city including the Boston police department (BPD). It is important to understand this data and to figure out where and how BPD is allocating these funds. You will first focus on BPD budget analysis and create a report based on the requested information below. This analysis will be used by City Council members to advocate for future budget allocations during the Spring budget negotiations with the Mayor and the Boston Police.

Data & Sources: (FY23) Operating Budget - Datasets - Analyze Boston, Employee Earnings Report - Datasets - Analyze Boston, Roster (list of active officers)

Some Questions to Explore:
- Who are the top highest earning BPD employees? Are they the same as last year?
- How does BPD employee compensation compare to other departments? Overall and overtime?

Expected Final Deliverable:
The final deliverable should be a report that includes the following information:
- Full analysis of the budget
- Visualizations and dashboards of key findings per MIRO and ACLU brand guide
- Templates for future budget analyses (a repo with instructions on installation/running, code written that can be easily run with future budgets)

HOARDING DEFINITION, DETECTION & "TREATMENT" IN BOSTON - CISS

The Center for Innovation in Social Science (CISS) at BU is dedicated to forging collaborative research and training future generations of students in the social sciences and related fields by using state-of-the-art methods and approaches and demonstrating the promise of the social sciences to improve human lives. Client: Meghann Lucy, CISS Graduate Affiliate, PHD Candidate, Sociology [see video]

Project Description: Hoarding can be defined as a condition characterized by the acquisition of and inability to discard possessions, regardless of their value that results in the extensive buildup of clutter, is a complex psychological and social problem. "Community-referred clients" are those whose hoarding behaviors are detected by the community, typically by neighbors or family reporting conditions to city agencies, city or landlord inspections, wellbeing checks, or handy workers. Community-referred clients tend to be of lower socioeconomic status. By asking questions about sociodemographic differences in hoarding experiences, the goal is to learn if the definition, detection, and "treatment" or sanctioning of hoarding functions exacerbate existing racial, class, and gender inequalities.

Data & Sources: Mass Courts Housing Data, years of 311 reports, inspectional services citation data

Questions to explore:
- How does a person who hoard's neighborhood, social class, race, and gender influence whether their hoarding is more likely to be detected?
- Do hoarding cases have different outcomes (e.g., evictions, appointment of a guardian ad litem, receiver, mediation, etc.) based on the social class, race, gender, and neighborhood/zipcode/census tract of the individual who hoards?
- Are evictions as an outcome of hoarding influenced by neighborhood, race, class, and gender?

Expected Final Deliverable:
Final report with answers to the aforementioned questions and any key visualizations, alongside a geographic map of Boston with different overlays.

Content Warning:

The following project is centered around issues of mental health and suicide

MENTAL HEALTH - CONNECT ED

Connect Ed's mission is to educate, encourage, and connect individuals, teams, organizations, and communities who want to play a productive role in their network. We believe in a shared narrative in which every one of us is a character with an important role to play. All of us are storytellers. When we learn how to share our stories in ways that contribute to belonging, our teams, organizations, and communities are healthier, more equitable, and more able to grow. Each of us has a powerful role to play. [see VIDEO next slide]

Client: Curtis Miller, founder of Connect Ed.

Project Description:
This project seeks to understand the mental, physical, and emotional health and feelings on suicide attempts. The client is particularly interested if there is a correlation with experiences of belonging or lack thereof and risk of emotional mental behavioral illness. By collecting data from a student population of 1000, the goal will be to provide analysis of the data and look for specific trends and conclusions that can be made.

Data & Sources:

Some questions to explore:
- Are there key words in the answer to the survey question "how would you describe belonging?" that can inform us of the student perspective of belonging?
- How does the sense of community of both staff members and students collected on the google form survey correlate with the number and strength of connections made and/or received?
- Is there a correlation between connectedness (outgoing and incoming) and club, activity, and/or sports environment?

Expected Final Deliverable: Create a report with key visualizations and answers to your client's questions.

PREFERENCE FORM & NEXT STEPS

Complete this form to indicate your project preferences for this course. Students who do not complete the form, or who submit it late will be placed on a team at Spark!'s discretion.

Please note that while we will make every effort to match your preferences, we can not guarantee that you will be matched to the project of your choice.

The Spark! team will take your preferences into consideration as they form teams to work on each project.

PREFERENCE FORM: https://forms.gle/ERSRxR3jiDSEWaVg8
DUE: Wednesday, Feb. 1 by 8 a.m.

QUESTIONS?