



DataScienceSF

Data science for service change

Presented by Tania Jogesh
Data Scientist
City and County of San Francisco
[@datasf](#) | [datasf.org](#)



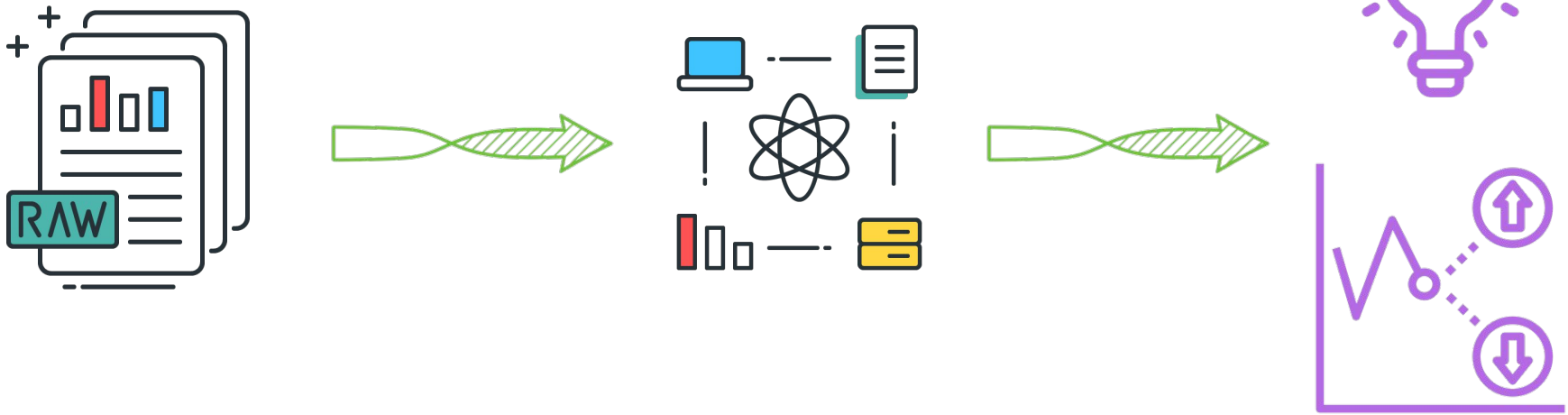
Data**SF**

Agenda

- Welcome from the CDO
- What is data science?
- What is DataScienceSF?
- Quiz!
- What can data science do for you?
- Activity: Start your application!
- Next steps

What is Data Science?

Using models built on existing data to generate insights and predictions

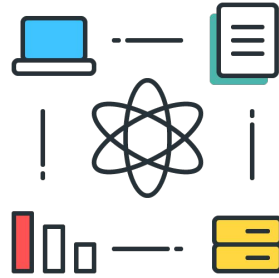


Data science is a component of analytics



Dashboards

vs



Models

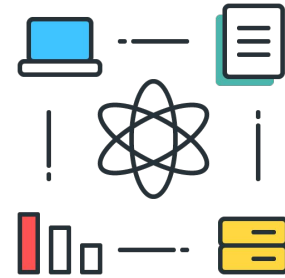
- A live dashboard built on clean automated datasets can get you a long ways!
- But in some cases, it makes sense to use a more complicated model to understand a problem or to make a prediction

Dashboarding is distinct from Data science



Dashboards

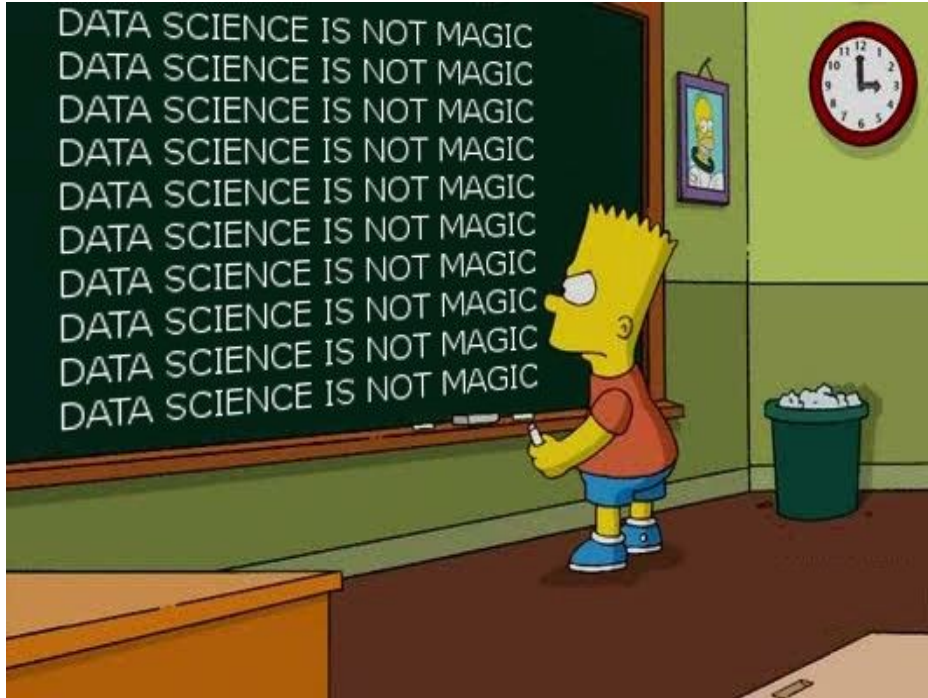
- ✔ Communicate insights
- ✔ Show past trends
- ✔ Based on “simple” analysis



Models

- ✔ Generate insights
- ✔ Continuous learning
- ✔ Predict & prescribe
- ✔ Complex analysis

Data science models are everywhere!



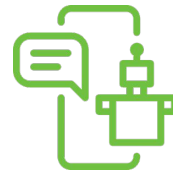
**We engage with data science models
on a regular basis**



Amazon recommendations



Uber/ Lyft rides



Chatbox/ Virtual Assistant

...and are expanding to public services

Targeting public benefits

Los Angeles Times

SUBSCRIBE

A computer model predicts who will become homeless in L.A. Then these workers step in



[Source](#)

Navigating public services

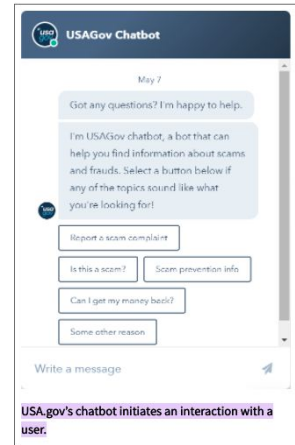
Federal Student Aid
An OFFICE of the U.S. DEPARTMENT of EDUCATION

UNDERSTAND AID

Meet Aidan®

Introducing Aidan, a virtual assistant that can answer questions using advanced technology—artificial intelligence and natural language processing. Aidan can help you find information about grants, or get help contacting your loan servicer.

[Source](#)



**What is
DataScienceSF?**

What is DataScienceSF?



Data Science

Applying advanced statistical tools to existing data to generate new insights



Service Change

Converting new data insights into (often small) changes to business processes



Smarter Work

More efficient and effective use of staff and resources & better services for residents

What complements data science?

(and is really good stuff to do)

Approach	Process	Outcome	Examples
Performance Management	Define, visualize, often using dashboards, and manage to KPIs	Meet goals and KPI targets	SF Scorecard, PublicWorks Stat & Stat starter kit
Evaluation	Assess a project, program or policy design or results	Better investment of resources; Better policy decisions	Evaluation of transitional-kindergarten in SF
Policy Analysis	Define and assess alternatives using a broad range of tools	Report or memo with policy or program recommendations	Shape Up SF Policy Analysis
Open Data	Publish civic data for use by the City and the public	Easier data sharing and reporting, new tools or services built on data	SFPUC Adopt a Drain
DataScienceSF	Identify insights using advanced statistics tied to a service change	Smarter work “on the ground” in real time	See rest of deck!

What complements data science?

(and is really good stuff to do)

Approach

Performance
Management

Evaluation

Policy Analysis

Open Data

DataScienceSF

All of these approaches are important
and usually interdependent!

What's in the DataScienceSF Toolkit?

Statistical Methods



Tools

User Experience Research

Time series analysis

Data mining

Multilevel modeling

Missing data imputations

Classification and clustering

Survival analysis

Pattern recognition

Principal component and factor analysis

AB testing

Machine learning

Forecasting

Sentiment analysis

Propensity score matching

Logistic, multinomial and multiple linear regression techniques

Network analysis

What's in the DataScienceSF Toolkit?

Statistical Methods

Tools

User Experience Research



Languages

Python

R

SQL

Javascript

NodeJS

Libraries

SciPy

Pandas

Scikit-learn

GPText

OpenNLP

Mahout

+many others

Data Engineering

Profiling

ETL

Job notices

APIs

Optimized data
pipelines

Optimized data
storage/access

Visualization

D3.js

Gephi

R

Leaflet

PowerBI

ggplot2

shiny

What's in the DataScienceSF Toolkit?

Statistical Methods

Tools

User Experience Research

**Iterative
Prototyping**

Photo journaling
and documenting



Service
blueprinting

Journey mapping

Ride-alongs

Process mapping

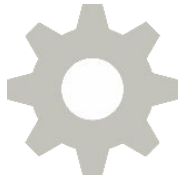
Ethnographic field
research and user
observation

Usability testing

What is **NOT** data science?



This



Service change



Small changes

Created by Danil Polshin
from Noun Project



Use existing data

Created by Arthur Shlain
from Noun Project

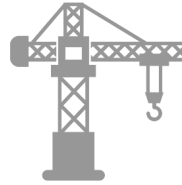


Not that



Academic research

Created by Rockicon
from Noun Project



Major overhauls /
service disruptions

Created by Hopkins
from Noun Project

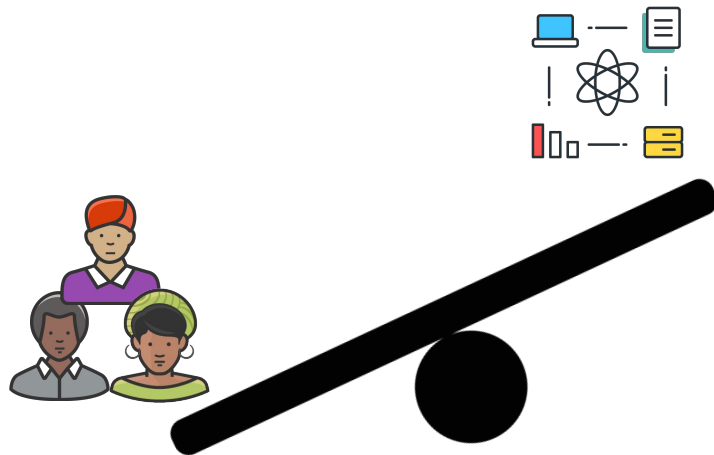


Collecting new data
(mostly ;)

Created by Chameleon Design
from Noun Project

In addition to the service change, we emphasize equity

- many sophisticated models can have inequitable outcomes
- we prioritize projects that have a positive equity impact



**What can data
science do for you?**

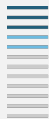
Data science project types

1



Find the needle in the haystack

2



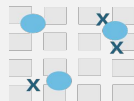
Prioritize your backlog

3



Flag “stuff” early

4



Optimize your resources

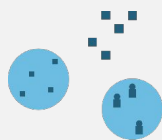
5



A/B test something

Data science project types

1



Find the needle in the haystack

2



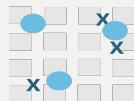
Prioritize your backlog

3



Flag "stuff" early

4



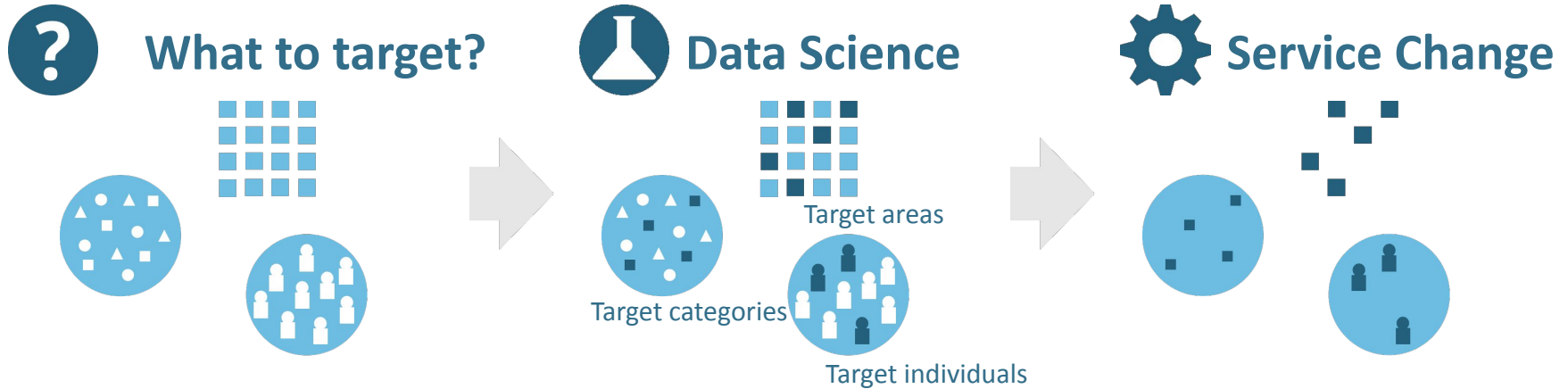
Optimize your resources

5



A/B test something

Project Type: Find the needle in the haystack



Service Issue:
Difficult to identify
targets in a population

Data Science Process:
Use existing data and
predictive modeling to
identify targets

Service Change:
Engage with target
subset of population

Increasing opt-ins for PUC's CleanPowerSF Supergreen program



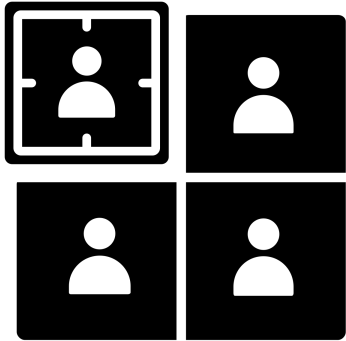
SuperGreen

100% RENEWABLE



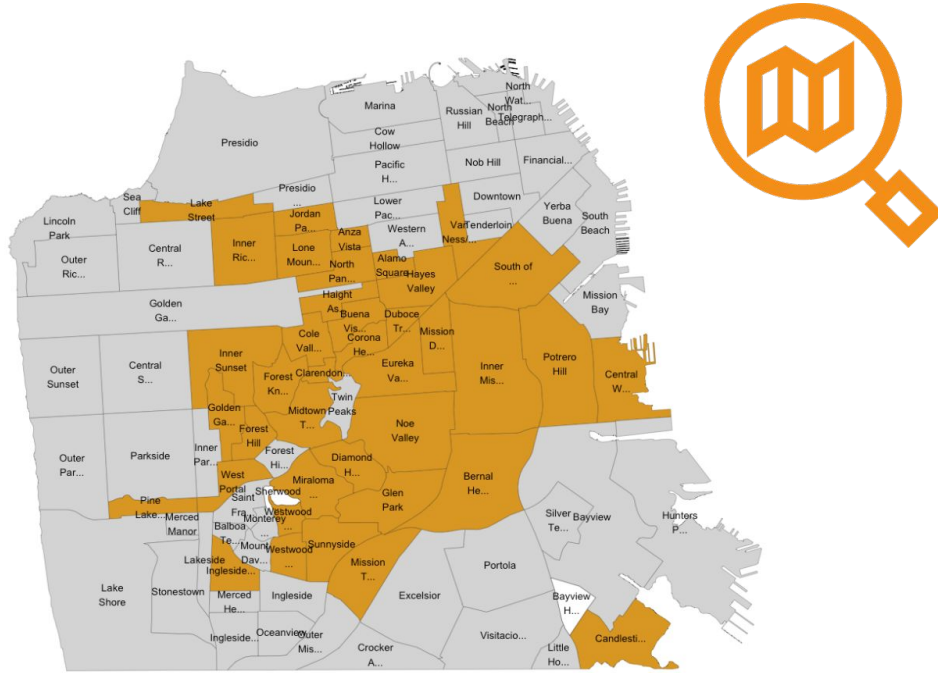
UPGRADE TO SUPERGREEN

Who should they target to increase supergreen sign-ups?



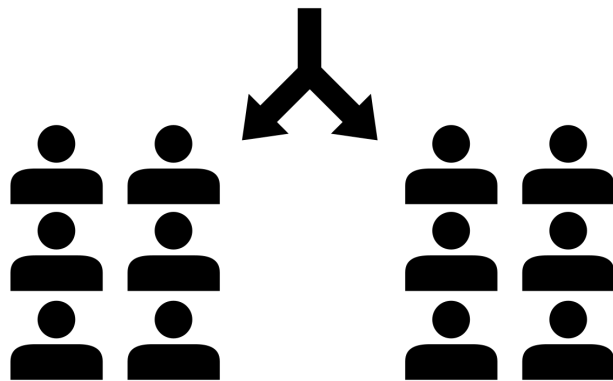
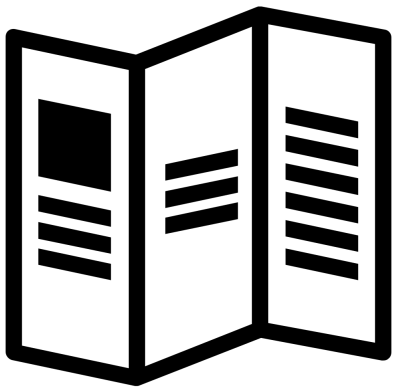
UPGRADE TO SUPERGREEN

Used existing data to segment customers and identify groups most likely to opt-in



- Customer locations, energy usage data
- Most likely to opt in were residents in the highlighted neighbourhoods and those who had opted for a PG&E “Time of Use” rate

Insights will help them design materials and A/B test their communications



Data science project types

1



Find the needle in the haystack

2



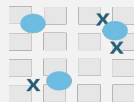
Prioritize your backlog

3



Flag "stuff" early

4



Optimize your resources

5



A/B test something

Project Type: Prioritize your backlog



Service Issue:

Backlog is tackled via first in, first out (FIFO)

Data Science Process:

Create a model to categorize and group past and current cases

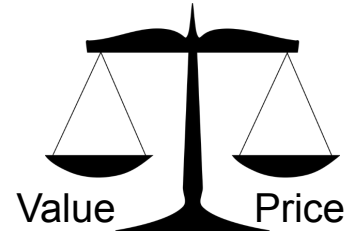
Service Change:

Prioritize cases based on categories in order of risk, need or opportunity

Streamlining property tax appraisals

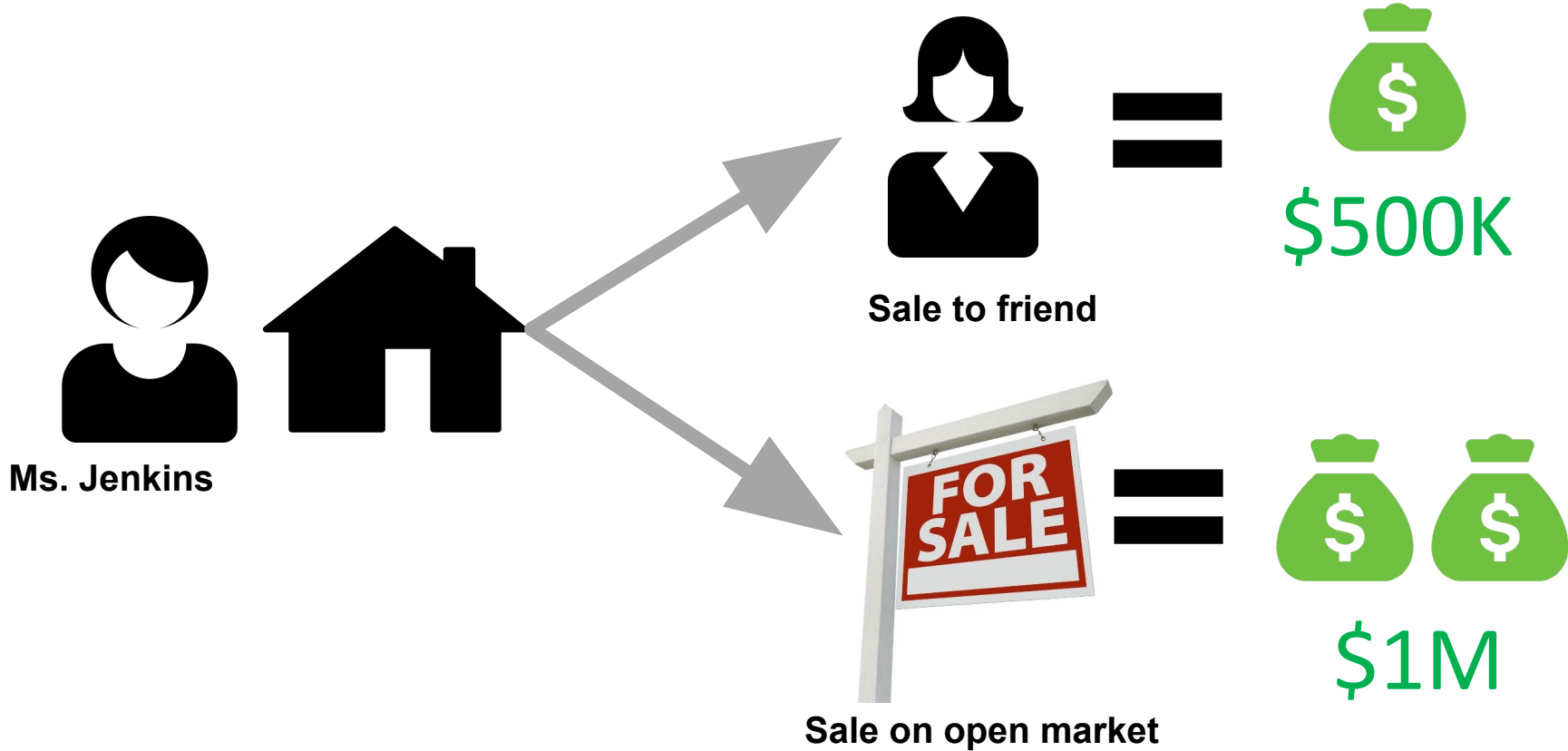


**Sale
price**



**Fair market
value**

Sale price may not always reflect fair market value...

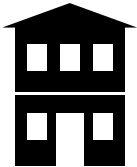


Doing a full appraisal for each sale is time consuming and ASR has heavy caseloads

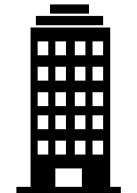


We developed models that generate a predicted fair market value

Property Type



Condo



Multi-family



Single family home



Property Characteristics



Date sold



Square feet



Location, Location, Location



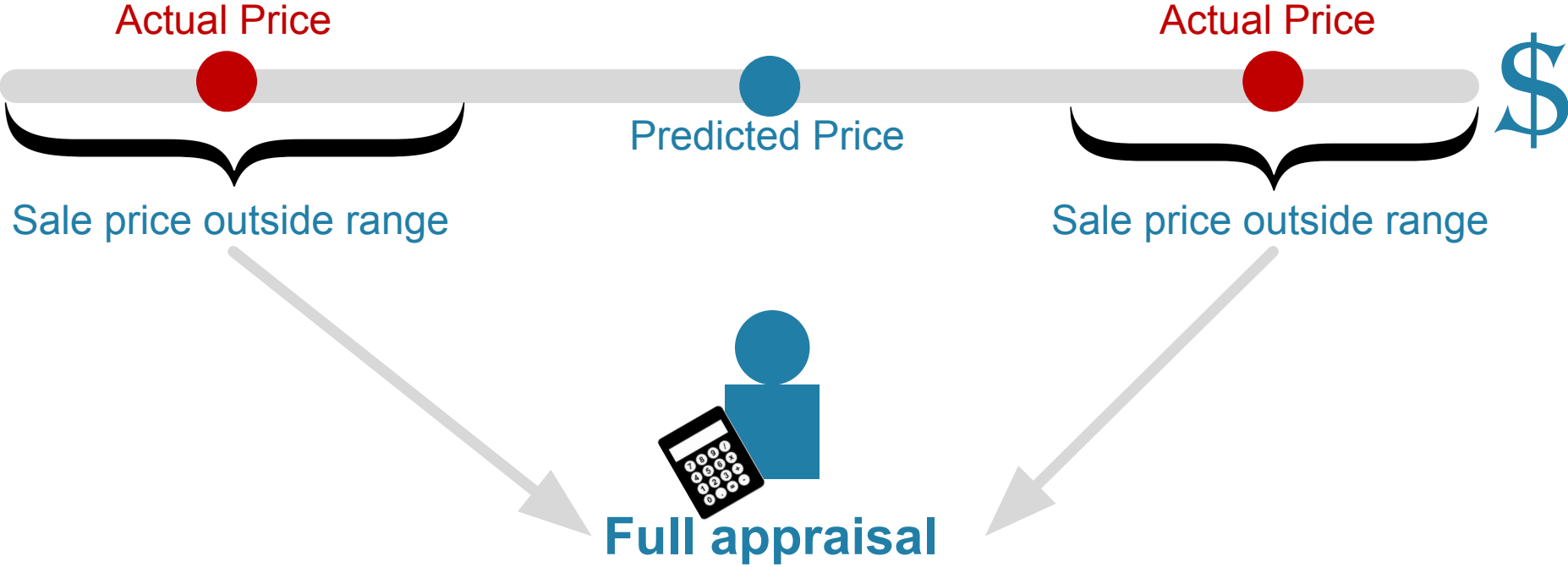
Models



Predicted Price



If the sale price is far away from the predicted price, ASR conducts a full appraisal



Result = Increased revenue!

⇒ The first run of the models reduced the workload by 166 properties or 10% giving the city immediate access to \$239 million in roll value leading to ~\$2.8M in revenue

+\$2.8 M

in tax

revenue

Data science project types

1



Find the needle in the haystack

2



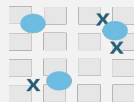
Prioritize your backlog

3



Flag “stuff” early

4



Optimize your resources

5



A/B test something

Project Type: Flag “stuff” early



Service Issue:

Hard to predict future condition which leads to reactive services

Data Science Process:

Use historical and current data to create estimate ranges for potential outcomes

Service Change:

Use estimates to change and tailor intervention points

Improving outcomes for residents in the Workforce program



SAN FRANCISCO
HUMAN SERVICES AGENCY

SERVICES ▾

ABOUT ▾

CONTACT ▾

PARTNERS ▾

LANGUAGES ▾

Q I am looking for...

GO

[Home](#) > [Our Services](#) > [Jobs + Money](#) > [JobsNOW!](#)

JobsNOW!

JobsNOW! provides employment and training services to income-qualified San Franciscans and offers wage reimbursement to employers when they hire program participants.

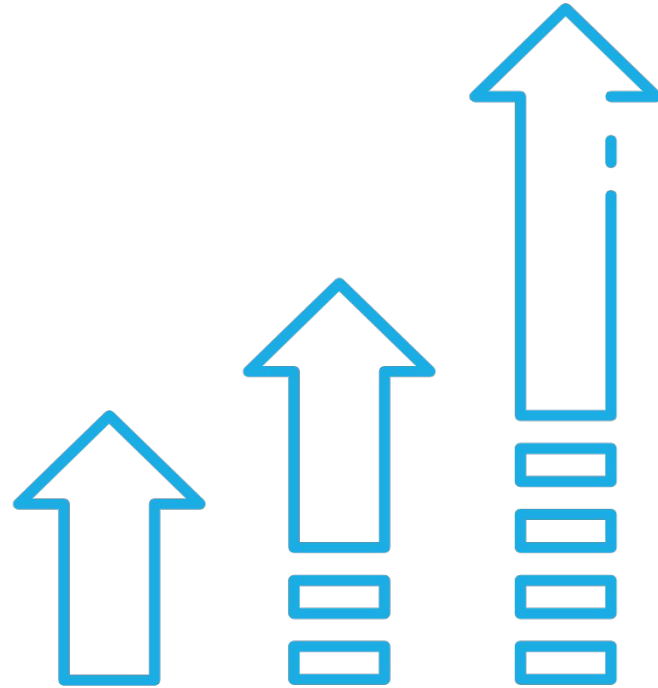


HSA Goal

Help HSA's Jobs Now program target its efforts and resources toward opportunities that are most likely to lead clients toward career advancement

Data science can help

to understand which industry, employer, occupations resulted in better client outcomes



Incorporate labor market lens into outreach & strategy efforts

Earnings increased with:

Employer size

Industry rep in SF

Healthcare jobs

Admin roles

- HSA should incorporate an industry strategy for outreach efforts for future employers



These insights helped HSA redesign their jobsNow program

Untargeted
partnerships with
employers



Targeted partnerships
that improve potential
earnings for clients

Data science project types

1



Find the needle in the haystack

2



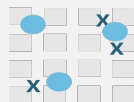
Prioritize your backlog

3



Flag “stuff” early

4



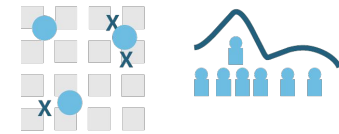
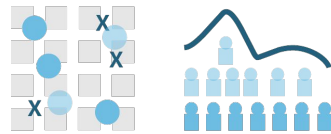
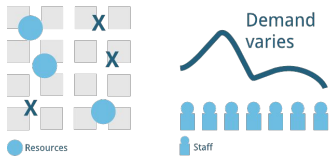
Optimize your resources

5



A/B test something

Project Type: Optimize your resources



Service Issue:

Difficult to identify where to place or distribute resources to be most effective

Data Science Process:

Use geospatial and/or other data to identify optimal distribution of resources

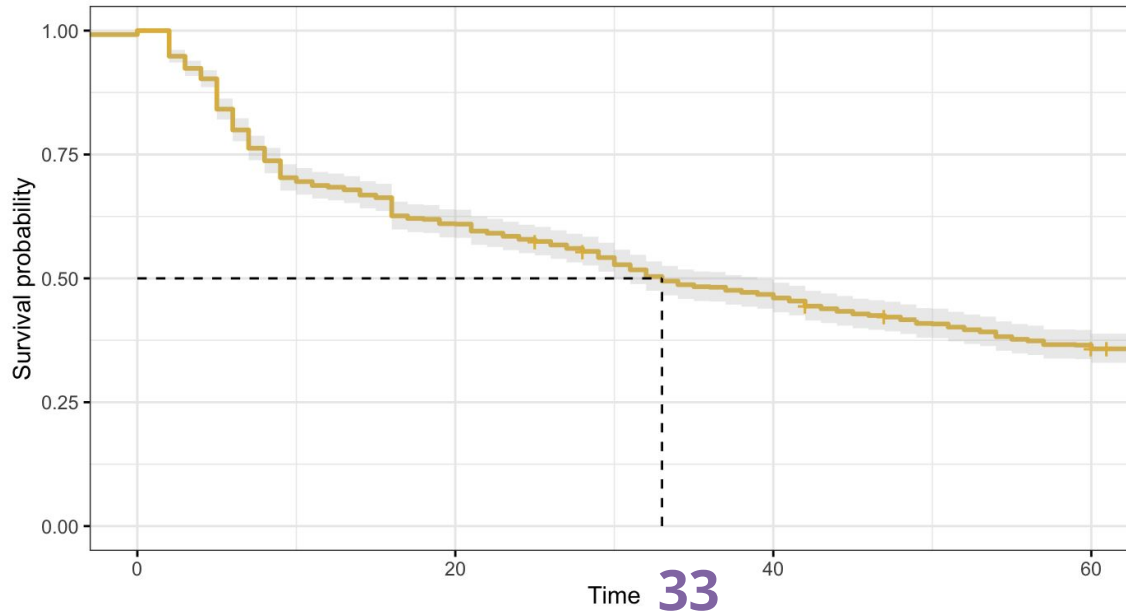
Service Change:

Re-allocates resources to optimal distribution

Where are PW sensors likely to break?

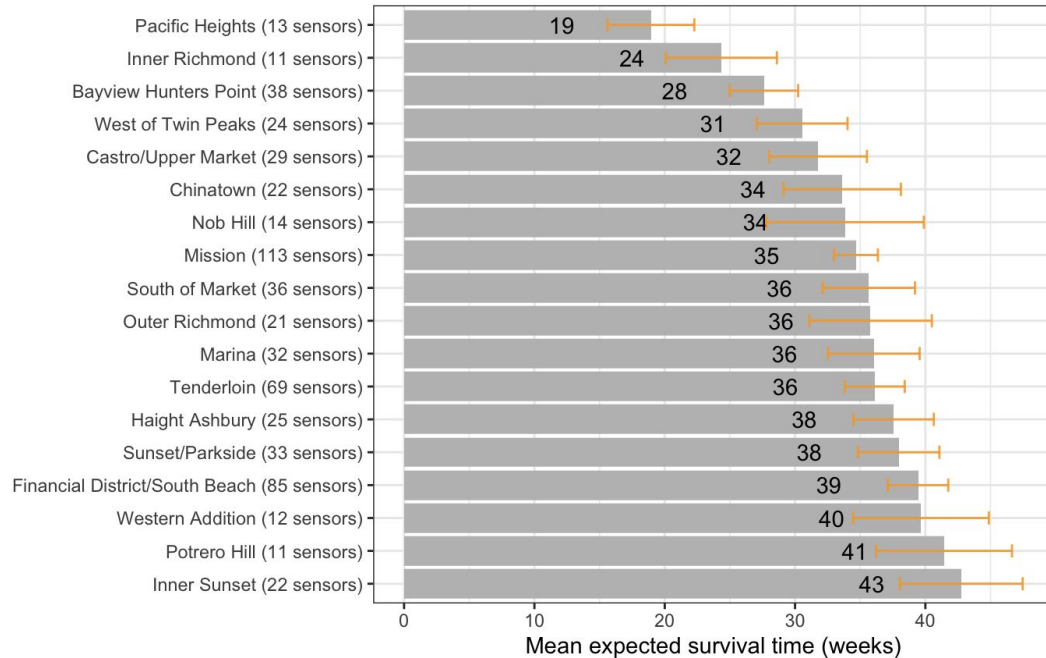


Estimated that 50% of sensors will break after 33 weeks overall



- Steady decrease in working sensors
- Are there specific places that sensors are breaking?

Sensor survival rate varies by neighborhood



- In Pacific Heights, Inner Richmond, and Bayview sensors survive the shortest amount of time
- Cellular connectivity and vandalism could be reasons why

Insights provided guidance for procuring new sensors



- Current sensors were breaking frequently and having cellular connectivity issues
- New contracts should include failure reason (vandalism, missing, etc) to track issues for easy reporting

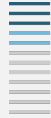
Data science project types

1



Find the needle in the haystack

2



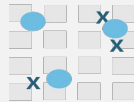
Prioritize your backlog

3



Flag “stuff” early

4



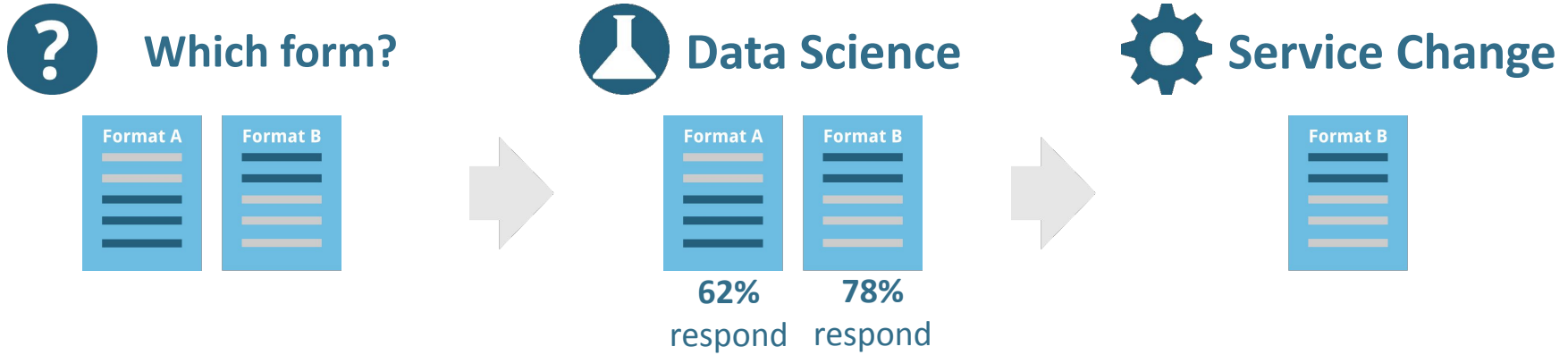
Optimize your resources

5



A/B test something

Project Type: A/B test something



Service Issue:

Costly outreach methods are not tested before implementation

Data Science Process:

Statistical testing on outreach methods to identify which, when, and to whom to send

Service Change:

Use statistically validated outreach method

TTX: Increase response to tax letter

? Service Issue

TTX wanted to use behavioral economics and A/B test to increase effectiveness of collection letter for unsecured personal property (a traditionally difficult type to collect on).

🧪 Data Science

DataSF helped organize a Behavioral Insights Training (BIT) workshop and provided guidance on A/B test

⚙️ Service Change

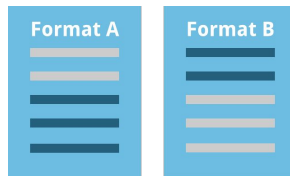
Use whichever letter gets the best response

✓ Result

Improved response rate by 17%. TTX continuing to apply BIT principles to other taxpayer communications



Treasurer & Tax Collector



A/B test something

Activity

- Take 5 minutes in your breakout room
 - Introduce yourself
 - Discuss: What are some potential projects you think data science could help with?

Next steps for you

Overview of Project Phases

Cohort 5

Aug 2022 - Sept 2023

2022

Aug-Oct

Oct 7

Oct

Nov 7

Nov-Dec

2023

Jan-Jul*

Sept

Application period

Application due!

Project selection

Applicants notified!

Cohort kickoff

Project refinement, scheduling, & cohort kickoff

Project delivery

Charter sign off, analysis, modelling, & service change

Present findings & lessons learned

**We work with you to agree project timing*

Phase: Solicitation

Opportunities to learn more:

- Brown bags (8/12 and 9/23)
- Office hours
- Ad hoc presentations

Dates at datasf.org/science

Aug-Oct

Oct 7

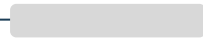
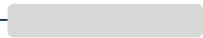
Oct

Nov 7

Nov-Dec

Jan-Jul

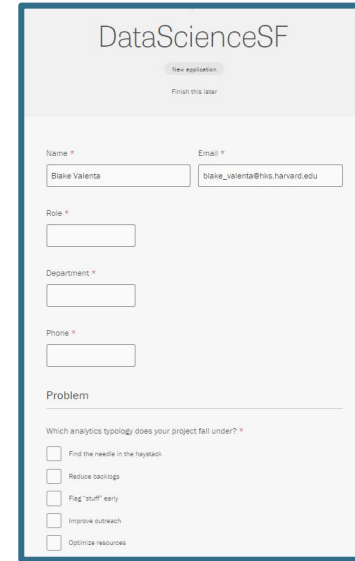
Sept



Phase: Application

Brief online form

- Problem statement (200 words)
- Impact statement (100 words)
- Service change statement
- Data overview
- **Due Oct 7**



The screenshot shows a web form titled "DataScienceSF". At the top, there are two buttons: "New application" and "Finish this user". Below these are input fields for "Name *" (containing "Blake Valenta") and "Email *" (containing "blake_valenta@hks.harvard.edu"). There are also input fields for "Role *", "Department *", and "Phone *". A section titled "Problem" contains a question: "Which analytics typology does your project fall under? *". Below this question are five radio button options: "Find the needle in the haystack", "Reduce backlogs", "Flag 'stuff' early", "Improve subwatch", and "Optimize resources".

Available at datasf.org/science

Aug-Oct

Oct 7

Oct

Nov 7

Nov-Dec

Jan-Jul

Sept

Successful applications are made during office hours!

- Brainstorm projects
- Get lingering questions answered
- High correlation between strong applications and prior attendance of office hours



Phase: Selection

Process

- Initial review
 - Criteria assessment
 - Application scoring
- Department follow-ups, as needed
 - Be available for questions



Blog Post

Part 1: How to solicit and select data science projects

This is the 1st of a 4 part series on managing data science projects in government.

1. [Part 1: How to solicit and select data science projects](#)
2. [Part 2: How to scope data science projects](#)
3. [Part 3: How to deliver a data science project](#)
4. [Part 4: How to tell your data science story](#)

If you are starting a data science service in your jurisdiction, your first task will be to develop a backlog of projects. This article walks through how we solicit and select data science projects.

How to solicit data science projects

[Read about our selection process](#)
datasf.org/blog/part-1-how-to-solicit-and-select-data-science-projects/

Aug-Oct

Oct 7

Oct

Nov 7

Nov-Dec

Jan-Jul*

Sept



Phase: Winners Announced

And gentle off-ramps for the rest...

Some projects may not be appropriate for data science or for our timeline. We will help identify other opportunities that may be a better fit:

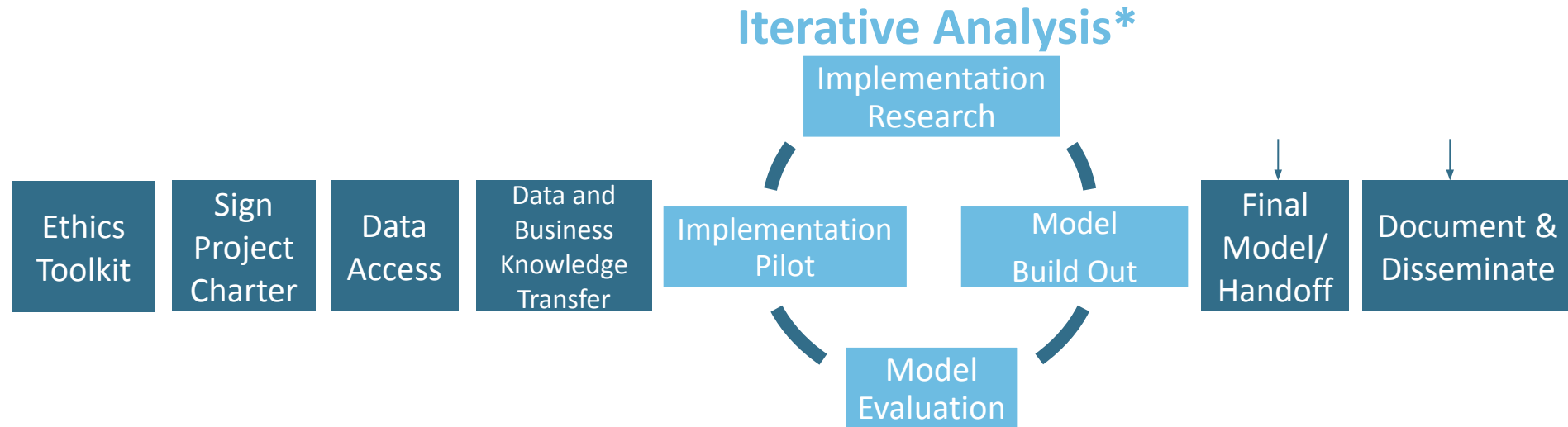
- Civic Bridge
- STIR
- DataSF Dashboarding Services
- Controller's Performance Unit
- Data Academy classes (Fall 2022!)
- External Data Science groups or volunteers
- Other technical assistance



Selected? Key things to know

- Service is free!
- Dept provides *project champion*
 - Prepare for, on average, **25% time spent on DataScienceSF**
 - Some weeks less, some weeks more
 - Act as our main point of contact through the DataScienceSF process
- **Project logistics**
 - Projects tackled sequentially, not concurrently
 - 1-2 months on average per project
 - **During kick off meeting we will set order (you have a say!)**

Phase: Analysis and service change



**Things can change, project team adapts to meet goals and improve service delivery*

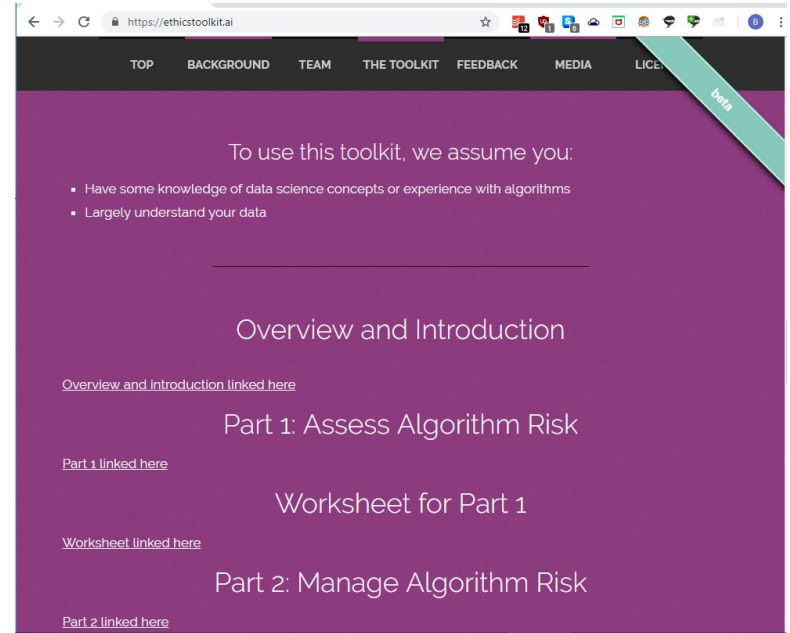
Ethics & Algorithm Toolkit

Why is this important?

- Outcomes can have important consequences for people in the city
- Understand biases in the data so we can use data responsibly
- Understand biases in the models so we can use algorithms responsibly

How?

Done jointly with DataSF & Department at start of project



Developed in partnership with:

- DataSF
- Harvard Ash Center
- John Hopkins Center for Government Excellence
- Data Community DC

Phase: Analysis and service change



Final Product:

Algorithm + Tool

scripted and automated algorithm

tied to some service change tool (e.g. list, service, alert)

implemented together

Aug-Oct

Oct 7

Oct

Nov 7

Nov-Dec

Jan-Jul

Sept



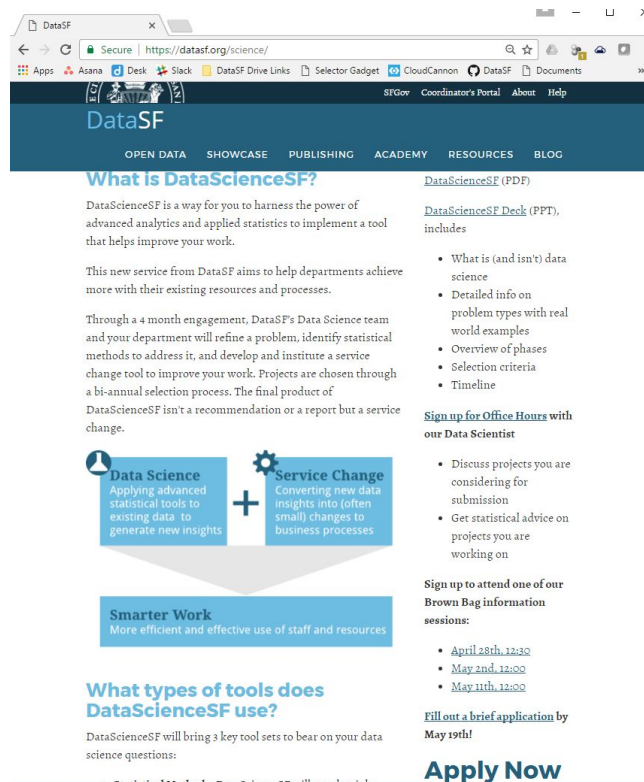
Your next steps:

Visit datasf.org/science:

- Sign up for office hours
- Apply by **Oct 7!**

 Questions?

email/ Teams: tania.jogesh@sfgov.org



The screenshot shows a web browser displaying the DataScienceSF website. The page has a dark blue header with the DataSF logo and navigation links: OPEN DATA, SHOWCASE, PUBLISHING, ACADEMY, RESOURCES, and BLOG. The main content area is white with a blue sidebar on the right. The main text explains the purpose of DataScienceSF, which is to help departments achieve more with their existing resources and processes. It details a 4-month engagement process where the Data Science team refines a problem, identifies statistical methods, and develops a service change tool. A diagram shows 'Data Science' (Applying advanced statistical tools to existing data to generate new insights) plus 'Service Change' (Converting new data insights into (often small) changes to business processes) leading to 'Smarter Work' (More efficient and effective use of staff and resources). The sidebar contains links to 'DataScienceSF (PDF)', 'DataScienceSF Deck (PPT)', and 'Sign up for Office Hours with our Data Scientist'. Below the sidebar, there are 'Sign up to attend one of our Brown Bag information sessions' for April 28th, May 2nd, and May 11th, and a 'Fill out a brief application by May 19th!' link. A large blue button labeled 'Apply Now' is at the bottom right.

Meet the team & acknowledgements



Michelle
Chief Data Officer
...and data mastermind



Tania
Data Scientist
...and bug doctor



Cody
Analytics Strategist
...and jack of all trades



Helen
Analytics Engineer
...and tooling master

Many thanks to New Orleans' NOLAlytics team, New York City's MODA, and Harvard's DataSmart for their resources, tools and templates!



Data, for the love of the City

Thank you! Questions?

@datasf | datasf.org | datasf.org/blog