



Columbia Data Science Society

Presents

The 6th annual Data Science Hackathon!

Sponsored By:

Two Sigma

NYC Data Science Academy

Who are we?



CDSS is an interdisciplinary society that promotes data science across Columbia University and the New York City community.

We are a team of undergraduate and graduate students from many majors and disciplines, and we work together to plan and host a variety of events including workshops, recruiting events, panels, discussions, as well as our annual hackathon!

Quick Plug...

We are now taking applications for new E-board members!



<https://forms.gle/qii15Z37hckFTNip9>

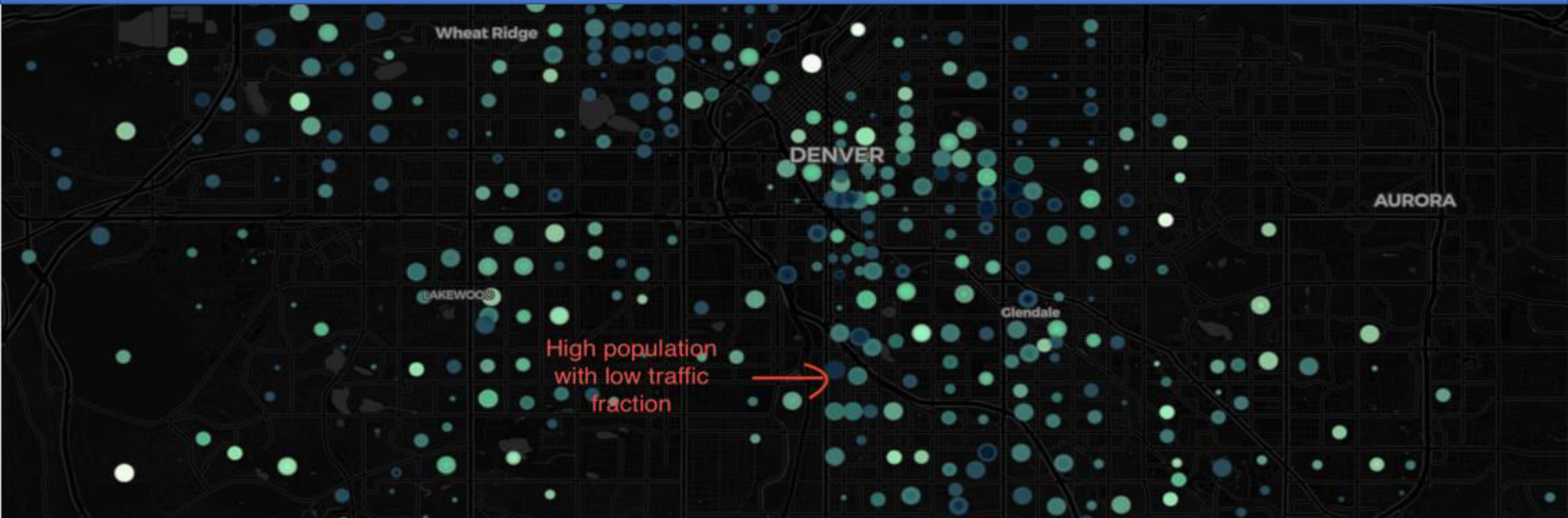
What is the Data Science Hackathon?



The Data Science Hackathon is CDSS's largest annual event: a friendly competition that encourages students of all skill levels to work together and engage in a meaningful data-problem, honing their skills along the way.

We hope that by the end, you'll not only have a new project under your belt, but you'll also feel more confident in your ability to collaborate on data and communicate your findings, and you'll

What is the Data Science Hackathon?



We select a dataset (or multiple datasets) that contains information that we could use to solve a problem. We want you and your team, through the computational methods at your disposal, to learn new insights from this dataset, and present that knowledge to us and a panel of judges.

Patterns in the data can take many forms, and often the most interesting patterns are relationships between variables. A presentation should evaluate these relationships, judge their strength, visualize them, and if possible, model them.

The Process

1. Frame the problem and look at the big picture
2. Get the data
3. Explore the data
4. Prepare the data
5. Short-list promising models
6. Fine-Tune the System
7. Summarize and Visualize your model
8. Present your solution

Recommendations

1. Divide and conquer the work
2. Don't underestimate visualization during EDA
 1. Scatter-matrices, residual plots, heatmaps, geo-plots, etc.
3. Setup a collaborative environment
 1. Google Collab, Github
4. Set a deadline for starting presentation
5. Ask questions!
 1. We can't answer questions that assist you in your project, but we can answer general questions and logistics questions
 2. Contact us during virtual office hours at cdss_executives@Columbia.edu

Schedule

6:00 PM	Logon to Zoom, introduction of Hackathon and sponsors
6:15 PM	Data overview, Demo to get data from Google Cloud, clarification of rules and schedule
6:45 PM	Q/A
7:00 PM	Hackathon Begins, Two Sigma Recruiting Talk
7:30 PM	Two Sigma Technical Mentor Office hours Begins
9:00 PM	CDSS Virtual Office Hours Begins
11:00 PM	Virtual Office Hours Ends

Sunday, September 20nd

6:00 AM	Virtual office hours begin
8:00 AM	Virtual office hours end
9:00 AM	Hacking ends, judging begins
11:00 AM	Judging ends, finalists announced
11:30 AM	Presentations begin
1:00 PM	Presentations end
1:30 PM	Winners announced
1:45 PM	Event ends

And this year's theme is...



How has COVID-19 affected mobility of people and goods across the globe?

Let's take a look at the data...



<https://console.cloud.google.com/marketplace/product/bigquery-public-datasets/covid19-public-data-program?project=cdss-hackathon-test&folder=&organizationId=>

Data Usage

Always free usage limits

As part of the [Google Cloud Free Tier](#), BigQuery offers some resources free of charge up to a specific limit. These free usage limits are available during and after the free trial period. If you go over these usage limits and are no longer in the free trial period, you will be charged according to the pricing on this page.

Resource	Monthly free usage limits	Details
Storage	The first 10 GB per month is free.	BigQuery ML models and training data stored in BigQuery are included in the BigQuery storage free tier.
Queries (analysis)	The first 1 TB of query data processed per month is free.	Queries that use BigQuery ML prediction, inspection, and evaluation functions are included in the BigQuery analysis free tier. BigQuery ML queries that contain CREATE MODEL statements are not. BigQuery flat-rate pricing is also available for high-volume customers that prefer a stable, monthly cost.
BigQuery ML CREATE MODEL queries	The first 10 GB of data processed by queries that contain CREATE MODEL statements per month is free.	BigQuery ML CREATE MODEL queries are independent of the BigQuery analysis free tier, and only apply to BigQuery ML built-in models (models that are trained within BigQuery).

How to Submit

- Submission must be a PowerPoint (preferably >2013)
 - Presentation should be no longer than 10 minutes
- Must submit before 9:00 a.m. EST Sunday, September 20th
- Email cdss_executives@columbia.edu
 - Include the names of all team members
 - Attach your powerpoint
 - Attach all relevant code
 - Confirm your availability to present if you are selected as a finalist
 - Include a team name (Optional)

And the winner is...

- Finalists will present 11:30 a.m. – 1:00 p.m. EST
- The top three teams will be eligible for cash prizes from our sponsor, Two Sigma
 - 1st Place: \$500
 - 2nd Place: \$350
 - 3rd Place: \$150
- Winners will be announced at 1:30 p.m. EST

Stick Around

Two Sigma Recruiting Talk: 7:00 pm – 7:30 EST (same zoom)

Two Sigma Technical Office Hours: 8:00 pm – 9:00 pm EST (same zoom)

CDSS Virtual Office Hours: 9:00 pm – 11:00 pm EST (email cdss_executives@columbia.edu)

CDSS Virtual Office Hours: 6:00 am – 8:00 am EST (email cdss_executives@columbia.edu)

Happy Hacking!



Columbia
Data Science
Society

Questions?