

Datathon Manual

→ Introduction

Congratulations on being selected to participate in Citadel's Data Open, in partnership with Correlation One! Below, we will familiarize you with this "Datathon" event and provide some pointers on how to maximize your performance.

First, what is a Datathon? A Datathon is a new type of live competition for data scientists, analogous to "Hackathons" for software engineers. But rather than using your coding skills to build a product, you will use a real-world dataset that we provide in tandem with your data analysis skills to solve a socially impactful problem. At the end of the competition, you will then submit a report detailing your solution.

→ Preparation

You should bring your laptop and charger. Make sure to have AT LEAST 1GB of free RAM on your computer for the event. Otherwise, it will be difficult for you to work with the provided dataset.

→ Dataset Download

We will provide your team with a USB drive that contains the dataset, from which you can copy the files onto your computer. If you do not have a USB portal, you may download the sets from the folder in your team's Dropbox account (more below) titled "Datathon Materials". In that folder, simply click on the "..." at the right of any given file row, and click "Download" from the drop-down menu.

House Rules

1/

Registration will run from 8:00 - 8:15AM. Once you arrive, please check in with a Datathon team member.

2/

After registering, please look for your team number in the room. If you do not remember your team number, please see a Datathon representative.

3/

You are allowed to start downloading the dataset using your team's USB drive immediately after you register. However, please do not begin working with the dataset until after the hacking session begins.

→ Submission

Submissions will take place via Dropbox. Your team will be given a Dropbox account, along with the associated username and password, at registration.

To submit, simply log in to your account, then upload your submission content to the folder titled “Team [x]”, where [x] is your team number. You can do this via the “Upload Files” button on the right of the “Files” page. However, please note that you **MUST** add your submission files to the “Team-[x]” folder already created for you. Otherwise, the judges **WILL NOT** be able to view and evaluate your submission.

→ Evaluation

Your submission should contain all the key components of a thorough data analysis investigation, which include:

- A clearly defined topic question, including an explanation of why it is important
- A summary of your key findings, including an explanation of why they are important
- An overview of your data manipulation methods – in particular, what you did to wrangle the data into a form amenable to your analysis, and why
- An exposition of your exploratory data analysis process, including why you made the investigative choices you did and what you were able to conclude from them
- A walkthrough of your modeling approach and rigorous substantiation of any results
- Any summary statistics and visualizations that are helpful in communicating your key points

House Rules (Cont'd)

4 /

You are allowed to use your laptops and any static reference materials (e.g. books, papers, notes, etc.) during the hacking session. However, penalties will occur if teams are caught communicating with sources outside the competition. This includes, but is not limited to, online real-time Q&A, people outside of your team, and voice-activated bots.

5 /

You may continue working on the problem statement after the competition for personal purposes, provided that you do NOT distribute your solution, the statement, or the dataset to people outside of your team.

→ FAQ

Q. Can I use the Internet during the hacking session?

A. Yes, you may use the Internet to access any static reference materials and help manuals. However, you may not use dynamic consultation resources like real-time Q&A.

Q. Can I access my cell phone/email/Facebook during the hacking session?

A. No, you are NOT allowed to access your phone, email, Facebook, or other electronic communication channels. If you need to transfer files amongst your team, use the Dropbox account that we have provided you.

Q. Can I leave the competition vicinity to go to the bathroom?

A. Yes, though we ask that you please use the one closest to the competition premises.

Q. Will there be food?

A. Yes, we will provide breakfast and lunch for all Datathon participants.

Q. What can I expect from the dataset?

A. The dataset will be reasonably pre-cleaned, and sized such that it is compatible with a standard, healthy laptop that has at least 1GB of free RAM. We will provide access to the dataset on the day of the competition.

Q. Do I need to stay for the entire event?

A. Yes, all participants will need to stay for the duration of the event.

Tips + Recommendations

1/

Spend some time with your teammates prior to the event dividing up responsibilities and coming up with a game plan. Three to four of you will be working together to solve a problem within a tight time limit, and it makes sense to have each person work on aspects of the problem that cater to their strengths.

2/

Do NOT stay up late trying to do last-minute preparation. The hacking session will be a sprint and it is best to be well-rested for it.

3/

Do NOT try to learn new techniques in the middle of the competition. Stick to what you know and how you can best make use of your knowledge.

4/

Make sure to start your final report AT LEAST four hours before the deadline. In the past, many teams have spent a lot of time conducting great analyses, only to realize that they left almost no time for actually writing up and presenting their results.

5/

Make sure your report is well-presented AND well-substantiated. This cannot be stressed enough – merely having a good model will NOT win a top prize. You must defend the model's rigor and clearly explain why its results are significant.