Women's Datathon 2025

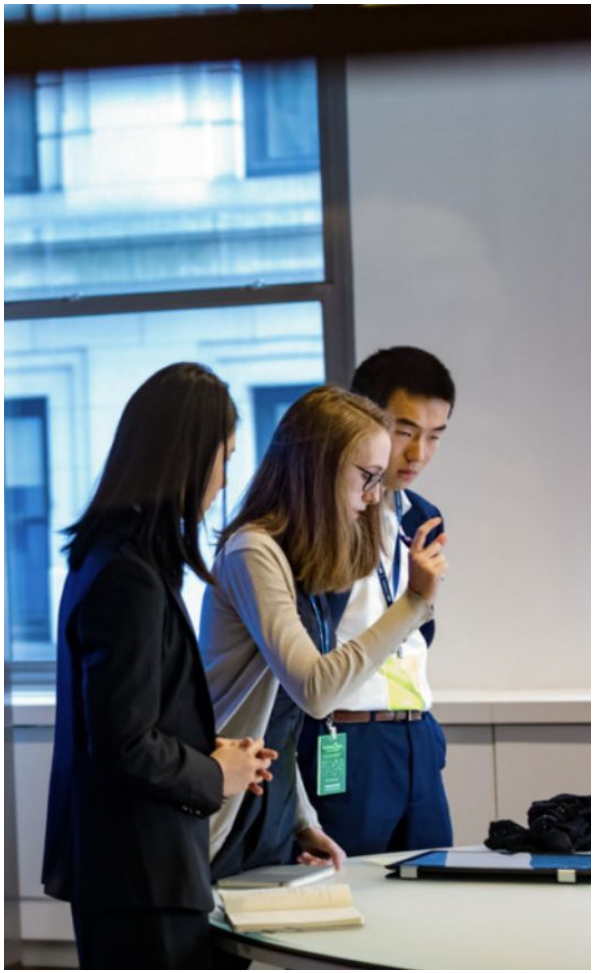
Competition Schedule and Guidelines

CITADEL | CITADEL | Securities | correlation one

Today's Agenda

- Brief Introductions
- Datathon Overview
- Problem Statement
- Datathon Evaluation
- Competition Schedule and Tips for Success
- Q&A

What is a Datathon?



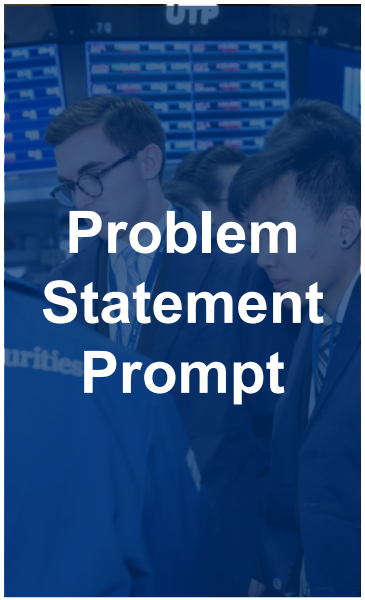
- **Sponsored by Citadel** and Citadel Securities in partnership with Correlation One, Datathons are globally recognized as premier university-level data science competitions, attracting top talent worldwide.
- **Since its launch in 2017**, the series has challenged participants to explore complex datasets and deliver innovative, topic-specific insights.
- Participants analyze the provided problem statement and datasets, submitting their findings to a panel of judges who evaluate entries based on originality, research depth, clarity, technical rigor, and accuracy.

Problem Statement

- The problem statement, provided as a PDF, outlines the Datathon's topic, your task, the datasets you'll use, and key details like evaluation criteria, submission format, and success tips.
- Materials will be distributed tomorrow morning. If you haven't already, we encourage you to introduce yourself in the Slack channel ahead of the event.
- **The Datathon team is here to support you!** During the competition week, C1 team members will be available via Slack.

For technical questions, our data team will be available on Slack throughout the week. For additional support, feel free to post in the relevant channel or contact Mert Paker at mert@correlation-one.com.

Problem Statement Intro



When Edwin Drake drilled the first well in Titusville over 160 years ago, there's no way he could have anticipated the weight and influence it would hold in modern society!

Gasoline remains one of the most heavily consumed fuels in the United States and is the primary output of domestic oil refineries. According to the 2020 U.S. Census, approximately 92% of American households own at least one vehicle, underscoring the country's deeply ingrained car culture. Even with the surge in electric-vehicle adoption, fully electric cars made up less than 10% of new-vehicle sales in Q1 2025, meaning more than 9 out of 10 vehicles still rely on gasoline for at least part of their power.

Given that car travel will continue to dominate U.S. transportation for the foreseeable future, understanding the behavior of gasoline prices remains economically and politically significant. ***With the wealth of available data, can we better anticipate these shifts, and more importantly, understand their ripple effects across the economy?***

You will be provided several pre-cleaned datasets, including:

- Weekly Gasoline Prices
- Weekly Supply Estimates
- Monthly Gasoline Makeup Percentages
- Energy Sector Stocks and Commodities (HLOC)

You are also welcome to utilize relevant external datasets (< 2GB) with proper citation.

Report Submittal and Evaluation Process

On Sunday, your submission will be sent to a group of judges. Judges will be evaluating your report without your team there to explain it; therefore, your submission must "speak for itself." Please ensure that your main findings are clear and that any visualizations are functionally labeled. Below are the key criteria for evaluation:

Non-Technical Summary
(What question are you answering? What are your key findings and their significance?)

- Select a question that aligns with the problem statement, providing clear, compelling reasons for your choice. Highlight how your question addresses a meaningful problem or adds value.
- Present precise and nuanced conclusions, avoiding blanket generalizations. Ensure that your insights are actionable and rooted in the analysis conducted.

Technical Exposition
(What was your methodology/approach used to answer your key questions?)

- **Wrangling & Cleaning Process:** Handle missing values, outliers, errors, and non-normalized fields effectively to ensure data quality. Make sure to apply thoughtful transformations to data fields when appropriate, clearly justifying any unconventional techniques used to better integrate or utilize the data.
- **Investigative Depth:** Conduct a multi-step Exploratory Data Analysis (EDA) process with proper visualizations at each stage. Explain why specific visualizations were chosen and how the insights informed subsequent steps in your analysis.
- **Analytical & Modeling Rigor:** Employ sound quantitative methodologies and incorporate qualitative considerations such as outlier analysis, residual analysis, and the examination of mediator or instrumental variables where applicable

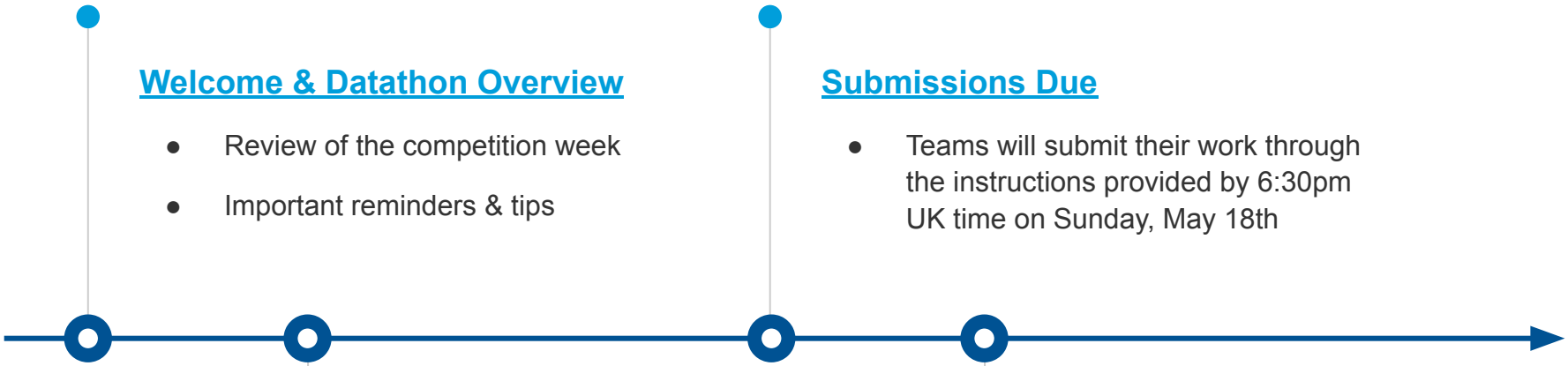
Important: You will be asked to include all relevant code that was used to generate your results.

Example Project Elements

A range of strategies exist for crafting a winning datathon submission. While many projects are commendable, certain key elements consistently distinguish the winning entries.

Project Element	Key to Differentiation
Novelty of Research Question	• Poses a clear, innovative question grounded in a thorough understanding of the underlying problem
Real-World Relevance & Effective Storytelling	• Provides actionable insights for businesses or policymakers • Presents findings in a clear, compelling narrative that is easy to follow
Technical Rigor & Complexity	• Demonstrates advanced modeling or hypothesis-testing methods and/or creative feature engineering • Clearly states, justifies, and explains all statistical assumptions and validation approaches

Competition Schedule



Welcome & Datathon Overview

- Review of the competition week
- Important reminders & tips

Coding Begins

- All the competition materials will be sent on Wednesday morning

Submissions Due

- Teams will submit their work through the instructions provided by 6:30pm UK time on Sunday, May 18th

Final Day Webinar

- The winning teams will receive $15,000 in cash prizes

Competition Guidelines

TIPS AND SUGGESTIONS FOR PARTICIPANTS

Here are the guidelines to ensure a smooth completion and delivery of your submission:

1

Schedule your work time and plan ahead

Dedicate the initial few days to exploring the data, understanding the topic, and testing hypotheses. Proper preparation can set the foundation for your analysis.

2

Commit to an approach

After completing the exploratory stage, narrow your focus to a specific topic that differentiates your submission. A well-defined scope will set your submission apart.

Be cautious about drawing conclusions. Even if your model passes cross-validation checks, this does not necessarily imply causality. Highlight limitations and justify any causal claims carefully.

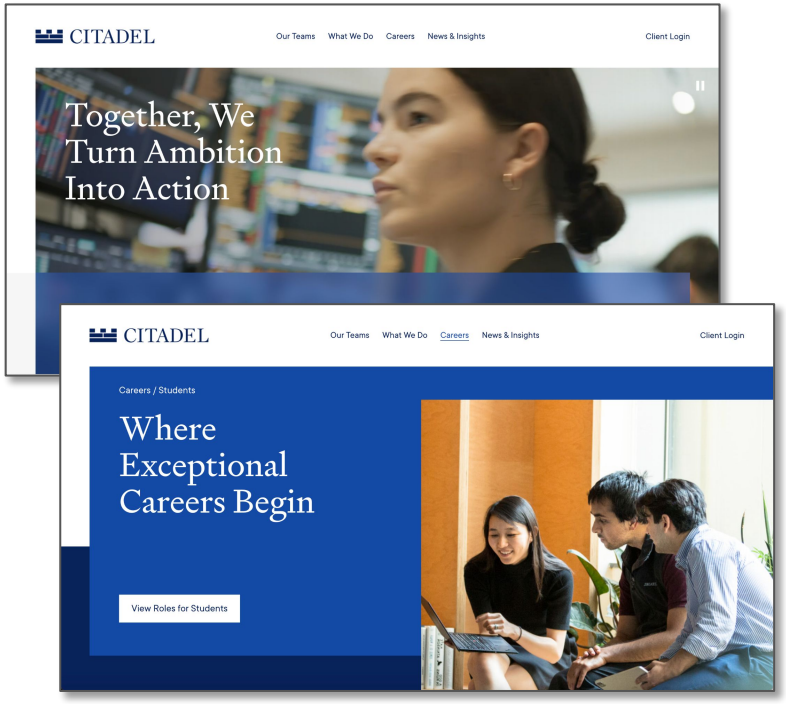
3

Start working on your report early

Judges will evaluate your technical report without your presence to provide context. Ensure your findings are clear, visualizations are appropriately labeled, and the submission effectively communicates your insights.

Event Follow Up

- **Interested in learning more about Citadel and Citadel Securities?** Join us for an exclusive online information session next week, where you'll gain firsthand insights into teams, roles, and career paths.
- **During this session, business representatives** will share details about their roles, discuss what it's like to work at Citadel and Citadel Securities, and answer your questions in a live Q&A.
- **Whether you're exploring potential career paths or looking for guidance on the application process**, this is your chance to connect directly with the Citadel team.
- **Stay tuned for more information!**



Good Luck!

Q&A

CITADEL | CITADEL | Securities | correlation one