

**Participant Survival Guide**

*Visit this guide regularly for the latest updates. The live version of the document can be accessed at:* [*http://tiny.cc/DUCguide*](http://tiny.cc/DUCguide)

# 2020 DUC Datathon

The SPE Calgary Data Science Engineering Analytics (DSEA) SIG has teamed up with Untapped Energy to organize a special data event: **the 2020 DUC Datathon**! Aimed at data practitioners and data enthusiasts from all levels of backgrounds and experiences, participants will:

* Learn and refine data science skills and techniques
* Meet new people and be exposed to new ways of collaborating and working together
* Tackle a real industry challenge using data science

The event consists of a number of features:

* An exclusive high-quality dataset
* A series of data science bootcamps
* An entry to into the data competition and the chance to win great prizes
* Access to the entire DataCamp learning platform for the duration of the Datathon
* Use of world class tools for the Datathon: TIBCO Spotfire, Tableau, Cartofact

**What are DUC’s?**

The topic of the Datathon is ‘Drilled but Uncompleted Wells’, or DUCs. DUCs are new wells after the end of the drilling process, but before its first completion process has been conducted. They offer flexibility to producers, but face potential issues related to case integrity, reservoir compromise and stranded capital. This flexibility gives DUCs the potential to greatly impact the global supply and demand of oil & gas.

To learn more about DUCs, view a replay of the Datathon Kick-off: <https://youtu.be/M3wQmcpir9k> (starting at 11:22)

# The Datathon

A Datathon brings data enthusiasts from all backgrounds and experiences to use data and analytics techniques to solve problems. For this Datathon, participants will solve problems related to DUC (drilled uncompleted well) using a provided dataset. Competitors will be evaluated on quantitative (feature engineering) and qualitative (insights and application) criteria. The competition will occur between August 2 and August 23, and competitors will provide their submissions online as a team. Outcomes may include visualizations, insights into specific variables, predictions or analyses involving external datasets.

To better prepare the participants to compete in the Datathon, a series of bootcamps have been organized. These bootcamps are not mandatory to attend, though each bootcamp will be using a subset of the actual dataset. Topics covered by the bootcamps include: Python, R, Cartofact, Spotfire and Tableau.

A number of tools are used for the Datathon, and all participants are provided will proper licensing to these tools for the duration of the Datathon. Additionally, all participants receive access to the entire DataCamp online learning platform for the duration of the Datathon. Participants are encouraged to leverage this feature to hone their data skills.

# Open Source Guiding Principle

The Datathon is built on open source guiding principles – the full definition can be found here: <https://opensource.org/osd>. By participating in the event, all submissions fall under the Apaches License 2.0 (<https://opensource.org/licenses/Apache-2.0>). The dataset used for this event is proprietary, but has been made available for this event by JWN Energy/CanOils. Participants in the event must complete the data licensing agreement in order to use the dataset (more details below in The Dataset section of this guide).

# Bootcamps

All 2020 DUC Datathon bootcamps can be found here: <http://tiny.cc/DUCbootcamps> – all bootcamps can be added to individual calendars and the Zoom link is embedded.

1. The bootcamp instructors may be using Slack to enable communications before, during and after the bootcamps. Please join this Slack workspace using this link: <http://tiny.cc/DUCjoinslack> You will be automatically added to the DUC Datathon channels were you will find message threads related to the respective bootcamps, as well as the rest of the DUC Datathon, posted.
2. The bootcamp instructors may be using Github to host relevant data/code files; these repositories are sorted by bootcamp instructor. All Github repositories for the bootcamps are found here: <https://github.com/untappedenergy?tab=repositories>

Details about the bootcamp will also be posted in Slack. All bootcamps are recorded and the YouTube recording will be posted to Slack afterwards.

**DataCamp – Learning Platform and Learning Competition!**

All registered participants receive access to the entire DataCamp platform for the duration of the Datathon. Additionally, DataCamp is sponsoring a Learning Competition: the participant who completes the greatest number of courses throughout the Datathon will receive a one-year premium subscription to DataCamp. The winner must achieve an XP score above 60%.

 [www.datacamp.com](http://www.datacamp.com)

# Teams

To be eligible to win the Datathon competition, all entries must be submitted as a Team. A Team must consist of a minimum of 2 people, but there are no limits for how large a team can be. Submissions must be received by **Monday August 17, 2020**. Submissions are done online at: <http://tiny.cc/DUCsubmission>

# Competitions, Judging and Prizes

There are numerous ways to be recognized during the 2020 DUC Datathon:

The Datathon Competition

This competition is broken into two main sections, designed to test your ability to model and gain insights from real world oil and gas data:

1. Quantitative (20%)
2. Qualitative (80%)

You will be challenged to explain the implications of your findings, and discuss how they can be leveraged towards improving industry’s ability to address the issue of uncompleted wells building up in Western Canada.

A screenshot of a computer

Description automatically generated

### **Quantitative Criteria: Feature Engineering**

### In this section of the competition you are tasked with building regression machine learning (or other) model that will be able to predict TVD (True Vertical Depth). A few rules:

### Teams can only use the data provided by the organizers of the competition.

### Top 3 performing teams (based on private leaderboard) will be asked to open-source their code and model to ensure that other participants and broader Calgary Data Science Community can learn from their solutions.

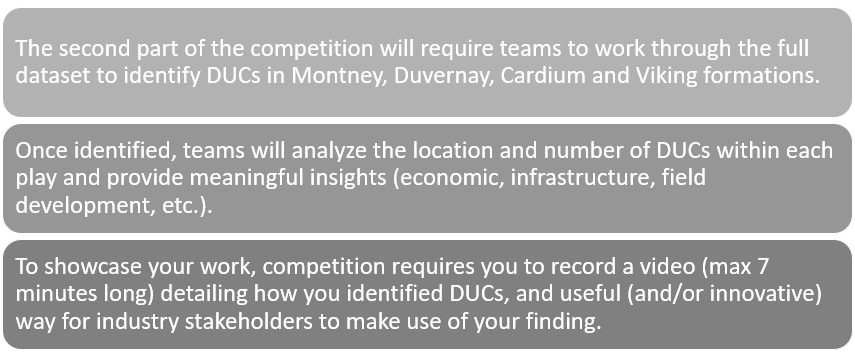
### No AutoML services.

A close up of a logo

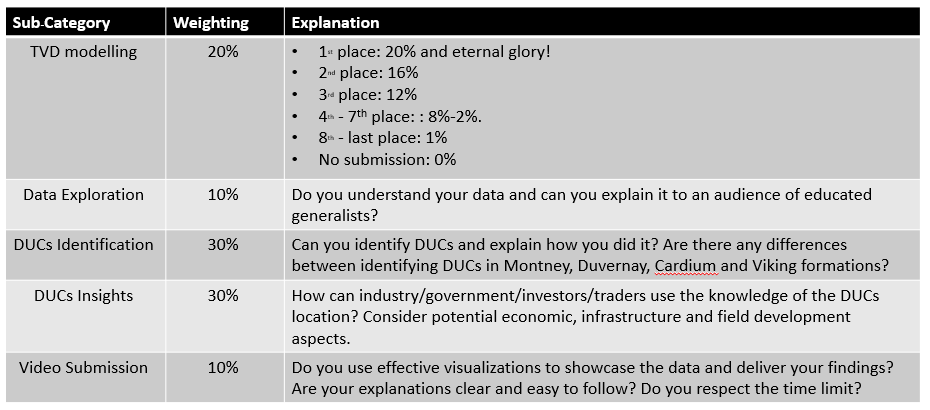
Description automatically generated

### **Qualitative Criteria: Insights and Application**

The objective here is to come up with new insights using the dataset provided. Such insight might include an analysis of a specific variable, a new way to visualize the data, a geographical analysis, time series, prediction, segmentation, etc. Participants submit a slide deck or video, visually explaining their ideas.



Competition Scoring Detail:



## Submissions and Judging

A panel of esteemed judges will apply an evaluation criterion to determine the winning submission. The winning team will be announced at the competition Closing Ceremony on Sunday August 23, 2020. Submissions must be received by **Monday, August 17, 2020** in order to qualify for judging. Submissions, which must include the Team’s name and member’s names, are done online at: <http://tiny.cc/DUCsubmission>

## Prizes

Prizes for this Datathon are made possible through the generosity of our valued partners.

## IT-IQ Grand Prize

The winning team for the Datathon Competition, as selected by a panel of esteemed judges, will receive $1,000 and an exclusive meeting with the senior leader of an organizing hiring data talent. This prize is generously provided by IT-IQ.

The winning team will meet with Evan Ferguson, founder and CEO of Quantifi and its proprietary machine learning base credit adjudication platform across Western Canada. Quantifi has 130+ active platform partners and deployed $25M in consumer loans.

 [www.it-iq.com](http://www.it-iq.com)

Cartofact People’s Choice Award

This prize will be awarded to the team as decided by voting by those attending the Closing Ceremony on Sunday August 23, 2020, where the finalist team submissions will be showcased. The team determined to the People’s Choice will win $500, generously provided by Cartofact.

 [www.cartofact.com](http://www.cartofact.com)

Visualization Competition

Individual participants are also able to compete in 2 data visualization competitions:

* Best Visualization using TIBCO Spotfire – the individual submissions using TIBCO Spotfire for data visualization will be judged by a TIBCO data scientist. The individual winner will win a set of Bose noise-cancelling headphones (valued at $500), generously provided by TIBCO. All submissions must be received by **Wednesday August 19** at: <http://tiny.cc/DUCspotfiresubmission>

 [www.TIBCO.com](http://www.tibco.com)

* Best Visualization using Tableau – the individual submissions using Tableau for data visualization will be judged by a Tableau specialist. The individual winner will win a pair of Bose Noise Cancelling Bluetooth headphones (valued at $500), generously provided by Tableau. All submissions must be received by **Wednesday August 19** at: <http://tiny.cc/DUCtableausubmission>

 [www.tableau.com](http://www.tableau.com)

# The Dataset

The dataset for the competition has been generously provided by JWN Energy/CanOils. In order to utilize the data, each participant is required to complete the data licensing agreement before the data can be used. The participant agrees to not post the data, or use the data other than for the Datathon and must delete the data once the Datathon is over. The data licensing agreement is found here (<http://tiny.cc/DUCdataagreement>) and the completed agreement is to be sent to: [ktrempner@spe.org](mailto:ktrempner@spe.org) or [info@untappedenergy.ca](mailto:info@untappedenergy.ca)

Prior to the release of the competition dataset on August 2, 2020, a subset of the data is made available for use in the bootcamps by the instructors and to allow participants to become familiar with the competition data. This can be found at: <http://tiny.cc/ducdata1>. Included in this location is a video describing what DUC’s are and the problem to be solved during the Datathon.

The complete dataset that is used for the Data Competition will be released on **August 2, 2020**. All registered participants will receive an email with the location of the complete dataset. Participants using the dataset must have completed a data licensing agreement (<http://tiny.cc/DUCdataagreement>) ) beforehand (as part of the registration process). Access and use of the dataset is governed by the data licensing agreement. If you need support accessing the data, please post on Slack or email: [info@untappedenergy.ca](mailto:info@untappedenergy.ca)

[www.canoils.com](http://www.canoils.com)

# Tools for Participants

## Slack

The Untapped Energy (datascienceoilgas.slack.com) Slack channel will be used for this Datathon. To join, use: <http://tiny.cc/DUCjoinslack>.To join, each participant must have a Slack account (it is free to join Slack).

All of the channels related to the 2020 DUC Datathon begin with “#2020ducdatathon...”. The primary channels are:

* #2020ducdatathon\_data - <https://app.slack.com/client/TA6PWQ4LT/C0171532N73/details/top>
* #2020ducdatathon\_general - <https://app.slack.com/client/TA6PWQ4LT/C016NG451V0>

## Github

All Github repositories for the bootcamps are found here: <https://github.com/untappedenergy?tab=repositories>

## Networking

Hangouts will be organized and occur until August 14, 2020. It will be a combination of short large group discussion and breakout sessions (randomly assigned) and is a great opportunity to hear more about the Datathon, to discuss the problem and the data with others and to meet other participants. Zoom details are emailed to all registered participants.

## TIBCO Spotfire

All participants will receive the free use of TIBCO Spotfire for the duration of the Datathon! Additionally, students will receive 1-year free access. TIBCO is also providing access to much training content and will host a bootcamp on July 28 bootcamp instructed by one of their own data scientists – Neil Kunungo – Dr. Spotfire!

To register for your complimentary free trial of TIBCO Spotfire:

* Sign-up here for a 30 day [**Spotfire free trial**](https://slack-redir.net/link?url=https%3A%2F%2Faccount.cloud.tibco.com%2Fsignup%2Fspotfire)
* Using the same email as the trial sign up above, complete this [**registration form**](https://slack-redir.net/link?url=https%3A%2F%2Fdocs.google.com%2Fforms%2Fd%2Fe%2F1FAIpQLSfEb4yF9nTWUd6ek2OAjng7-eucwfi5mrhnJom4MCvDLaB8UQ%2Fviewform) by July 31, 2020, to receive a bonus 30-day trail extension for Spotfire.
* If you sign up after July 31, 2020 you will only have access to Spotfire for 30 days.

if you are a student, educator, or non-profit, sign up at [TIBCO Spotfire single-license request](https://slack-redir.net/link?url=https%3A%2F%2Fwww.tibco.com%2Facademic%2Fspotfire) for 1 year of free access.

## Tableau

Participants have access to Tableau Desktop and Tableau Prep for the Datathon! Access details have been posted in Slack or via this link: [Tableau Licensing Program](https://www.tableau.com/tpep/suncor)

## Cartofact

Participants have access to Cartofact for the Datathon! Participants will access this tool through a web address and access code, creating an email-based account for access. Access details have been posted in Slack or: https://map.cartofact.com (Activation Code: SPEDUC2020). For best results, use Firefox or Chrome and a fast internet connection. Consider using a computer with at least 16 GB RAM.

**Timeline**

|  |  |  |
| --- | --- | --- |
| Event (all events are virtual) | Date/Time (all times are MST) | Details |
| **Kick-Off Event** | Tuesday, June 30 2020 – 5:30 p.m. | https://youtu.be/M3wQmcpir9k |
| **Bootcamps** | Throughout the month of July 2020 | http://tiny.cc/DUCbootcamps |
| **Competition Opening Ceremony** | Sunday, August 2 2020 – 3:00 p.m. | <https://www.meetup.com/untappedenergy/events/272004058/> |
| **Competition period** | August 2 2020 to August 23 2020 |  |
| **Submission deadline** | Teams Submission: Monday, August 17, 2020  Individual Visualization Submission: Wednesday August 19, 2020 | <http://tiny.cc/DUCsubmission>  <http://tiny.cc/DUCspotfiresubmission>  <http://tiny.cc/DUCtableausubmission> |
| **Competition Closing Ceremony** | Sunday, August 23 2020 – 3:00 p.m. | (Meetup page) |

# Valued Partners

Thanks to our valuable partners for helping to make the 2020 DUC Datathon a reality!

 [www.datacamp.com](http://www.datacamp.com)

 [www.it-iq.com](http://www.it-iq.com)

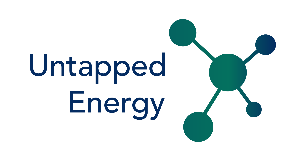
 [www.cartofact.com](http://www.cartofact.com)

 [www.TIBCO.com](http://www.tibco.com)

 [www.tableau.com](http://www.tableau.com)

[www.canoils.com](http://www.canoils.com)

 [www.connect.spe.org/calgary/home](https://connect.spe.org/calgary/home)

 [www.untappedenergy.ca](http://www.untappedenergy.ca)