

# Welcome to Factored 2023 Datathon

July 24th - August 04th 2023



Factored

Welcome aboard **01**

Keynote **02**  
Kevin Pope, VP of Engineering

Quick Overview **03**

# Agenda

**04** Details

**05** Prizes




01

# Welcome Aboard

---



# Timeline

- 
- 24/07/2023, 5:30 PM EST - **Datathon Starts**
  - 28/07/2023 - **Data streaming begins**
  - 05/08/2023, 23:30 EST - **Last Submission deadline**
  - Technical - Business Review phase
  - 11/08/2023 - Closing Event - **Winners announcement**

# We're here to support you!



Andrés Rojas  
Senior Data Analyst



Jose Baene  
Senior Geospatial  
Analyst

# 02

# Keynote

---

Kevin Pope, Vice President of Engineering



03

# Quick Overview

---



# 493

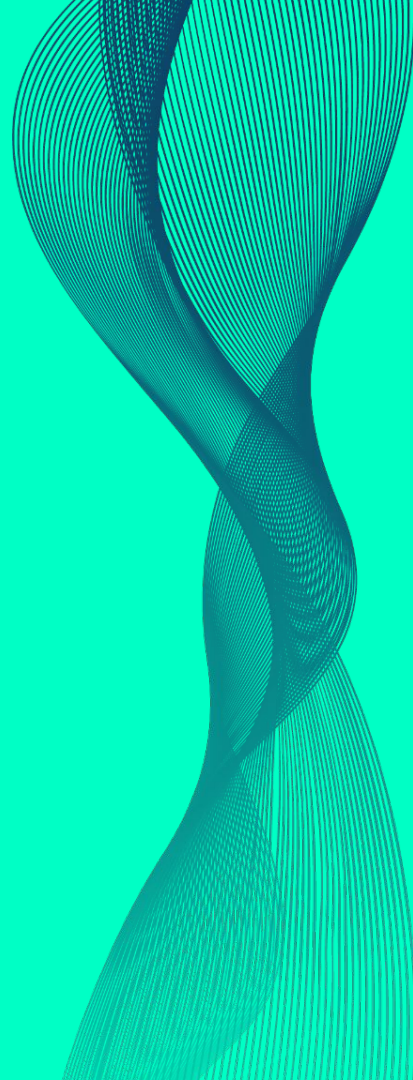
---

Participants



111

Teams



# The Challenge

Create an innovative data solution (web apps, chatbots, dashboards, model interfaces...) to empower businesses with insights from product reviews.

With an Amazon products review dataset <sup>1</sup> you will embody a product company's data team and develop an end-to-end data solution, to generate actionable insights to inform stakeholders' decisions.

- What has the most significant impact on customer satisfaction?
- Can we infer customer segment based on their review patterns?
- It's possible to build a recommendation system based on recommendations?

**<sup>1</sup> Justifying recommendations using distantly-labeled reviews and fined-grained aspects**

Jianmo Ni, Jiacheng Li, Julian McAuley  
Empirical Methods in Natural Language Processing  
(EMNLP), 2019

# The Challenge

Overall this challenge is divided into two parts

## Technical

- Overall quality of your solution. Does your Data Pipeline works correctly, regarding ingestion, data quality, delivery?
- How reliable are your model results?
- Does your solution works as expected?



## Business

- Does your solution answers a business question?
- Are the insights from the workflow be used to take actions?
- Think that you're part of a company data team or consultants required to provide value to a client!

**<sup>1</sup> Justifying recommendations using distantly-labeled reviews and fined-grained aspects**

Jianmo Ni, Jiacheng Li, Julian McAuley  
Empirical Methods in Natural Language Processing (EMNLP), 2019

# The Challenge

Details on the Data Delivery

Dataset is splitted in two parts:

- 80% will be handled in batch format
- 20% will be handled through a streaming starting on **Thursday July 27th**

**We will share with you a document detailing more information about the dataset**

**<sup>1</sup> Justifying recommendations using distantly-labeled reviews and fined-grained aspects**

Jianmo Ni, Jiacheng Li, Julian McAuley  
Empirical Methods in Natural Language Processing (EMNLP), 2019

# 04

## Details

---



# Event Logistics

## Communication!

Remember to join the Datathon community in **Slack**:

<https://factored-datathon-23.slack.com>

## Recommendations:

- Keep in touch always with your team!!
- We have some preset Channels, join and explore them!
- We might create additional ones to share specific topics that may help you going through with the challenge
- We've set up specific channels to get you in touch with mentors!
  - #de-mentorship : Everything related to Data Engineering
  - #ml-mentorship : Everything related to Machine Learning
  - #da-mentorship : Everything related to Data Analytics



If you don't have a team we will take care of setting up teams with people without them

# Event Logistics

## Document your code

All code should be stored in a Github Repository specific to your project

1. Set up the repo with the following formatting  
`factored-datathon-2023-[your team's name]`
2. Add **anferojasr** and **jnbaene** as collaborators on your repository
3. We will share a Forms to collect each team link to the repo, only the team leader should answer this form
4. Further instructions will be provided during the datathon

**Remember we have commit windows that are every day between 19:30 EST and 21:30 EST. (Not mandatory but you will get extra points)**



# Event Logistics

**You're free to use any language and/or set of tools you deem necessary!**

Some (maybe) valuable resources:

[Microsoft Azure](#)

[Databricks](#)

[Snowflake](#)

[AWS](#)





# Submission Details

For your submission to be considered successful you should:

1. Share the link to the GitHub Repository
2. Share a link to where your tool is deployed
3. Share a 4 - 6 Slides presentations with details on your tool

Submit your tool no matter what!!!

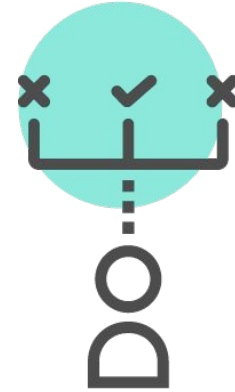
Submit all of this to [datathon.admin@factored.ai](mailto:datathon.admin@factored.ai)



# Evaluation Criteria

First and foremost our solution should work

1. Overall project rationale and documentation
2. Data Engineering: How are you dealing with the Extraction and Transformation of the data
3. Machine Learning: Model selection, optimization, implementation and tracking
4. Data Analytics: Data Quality and Providing relevant insights from the solution
5. Software Engineering: Backend, Frontend and Deployment.



# Code of Conduct

We are an inclusive community.

Violence and discrimination will not be tolerated.

Remember we do have a [code of conduct](#).

# 05

## Prizes

---



# Prizes

**US\$ 2,500**

2nd

**US\$ 5,000**

1st

**US\$ 1,500**

3rd

**Be INTERVIEWED by our engineering & talent team at FACTORED!**



**Thank you!**  
**Happy Coding!**

