

How GSK Tech is Democratizing Data Science

Pragathi Kotha Venkata
Senior Manager, Data Science

Our Journey so far



Overview

1. Provision of modern self-service Data science stack - an easily accessible way for requesting data science tools in a self-service fashion
2. Training and development - leveraging platforms such as Pluralsight, Learnathons & Hackathons
3. Inspiring code reuse and collaboration - by promoting usage of platforms such as GitHub and Guilds on workplace



Guiding principle



Ask- Analyze - Act

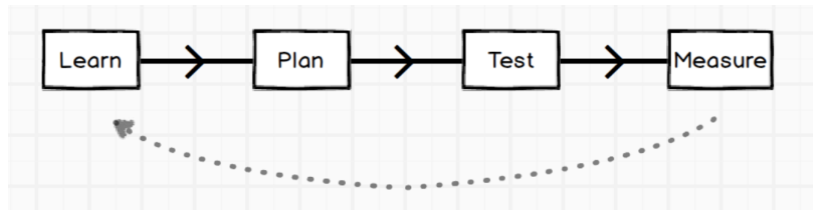
Define Initial Personas



Data Analyst



Data Scientist



Data Science User Research Survey: Core Technology



Page 1 of 1

Data Analysts, Scientist & Enthusiasts; the Data Product and Engineering team would love to hear from you to learn more about the tools you use at GSK for Data Science and Analysis. This short survey will take less than 2 minutes, and will be used to ensure we make user centric decisions around Data Science tools and capabilities.

1. Overall, how happy are you with current tools and technologies available to you for carrying out data and analytics tasks?

- ☐ Very satisfied
☐ Satisfied
☐ Neither satisfied or dissatisfied
☐ Dissatisfied
☐ Very Dissatisfied

2. What is the primary reason for your reply above?

3. On an average, what percentage of your work involves working with data - analyzing data or cleaning data or maintaining data and databases, etc?

- ☐ 25% or less
☐ More than 25% but less than 50%
☐ Over 50%
☐ My day job is working with data
☐ I don't work with data

4. What is your level of expertise with the following tools/languages?

	Never used it	Beginner	Intermediate	Advanced
SAS	☐	☐	☐	☐
SQL	☐	☐	☐	☐
Python	☐	☐	☐	☐
R	☐	☐	☐	☐
Spotfire	☐	☐	☐	☐
PowerBI	☐	☐	☐	☐
Qlik	☐	☐	☐	☐
Tableau	☐	☐	☐	☐

5. Please list any other tools you use that are not listed above.

6. If you use excel or other similar tools for your data and analytics jobs, would you like to learn other tools/technologies which can provide you with additional capabilities and skills?

- ☐ Yes
☐ No
☐ Not applicable

7. If you answered yes to the question above, what are other tools and technologies that you are interested in exploring and learning more?

- ☐ Advanced data modeling tools
☐ Machine learning tools
☐ Advanced Visualization tools
☐ Interested but I'm not sure
☐ Other, please specify

8. Please provide any other suggestions or comments regarding tools, technologies or platforms that will help improve your ability to work with data at GSK.

9. Which Organization do you belong to?

- ☐ Pharma R&D
☐ Pharmaceuticals
☐ GMS
☐ Vaccines
☐ Consumer Health
☐ Vliv
☐ Global Tech Organization
☐ Other, please specify

10. How do you identify yourself?

- ☐ Data Analyst
☐ Business Analyst
☐ Data Management Professional
☐ Data Engineer
☐ Data Scientist
☐ Data Science Enthusiast
☐ Other, please specify

11. What is your experience with data and analytics in years?

- ☐ I don't work with data or analytics
☐ Less than 1 year
☐ 1 to 3 years
☐ 3 to 10 years
☐ More than 10 years

Done

We learned



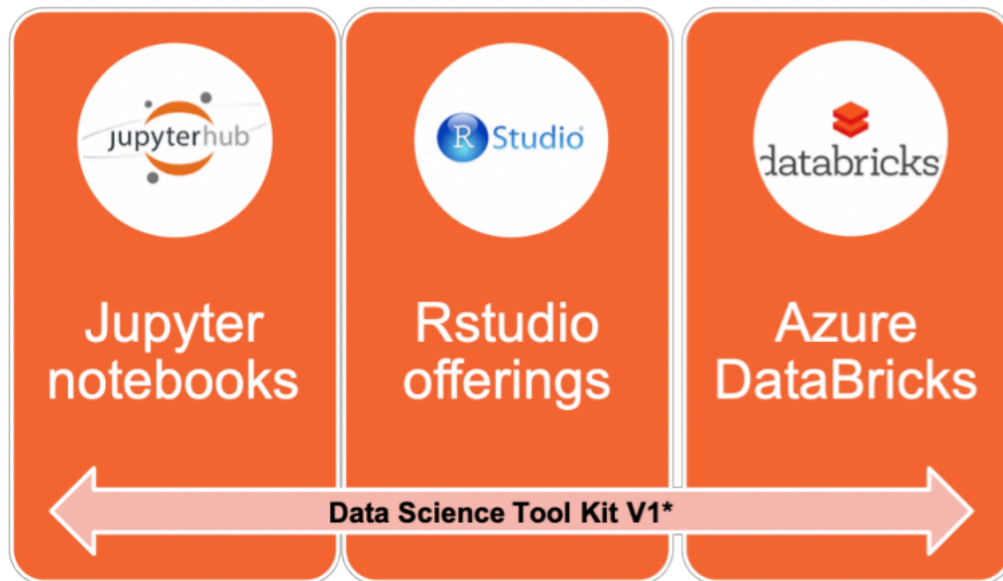
Current existing pain points

- Lack of access to tools across groups (Some teams have better access, others not much)
- Need for training and awareness around capabilities

1. Provisioning of modern Data science stack



MVP Tool set



Self-service portal and Enterprise Licensing



← → ↻ one.gsk.com/data/databricks 🔍 ☆ 🌐 📱 🌐 🌐 🌐 🌐 🌐

gsk CLOUD SERVICES APIs UI FRAMEWORKS DATA STORIES CONNECT **DEV PORTAL**

Databricks

Databricks Unified Analytics Platform, from the original creators of Apache Spark™, unifies data science and engineering across the Machine Learning lifecycle from data preparation, to experimentation and deployment of ML applications.

REQUEST ACCESS

Overview

Getting Access

FEEDBACK

v2.0.0

Request to try Databricks in a sandbox environment

Complete this form and we'll be in touch with the credentials you need to get started.

NAME	EMAIL
Please enter your name*	Please enter your email address*
MUD ID	BUSINESS UNIT
Please enter your Mud ID*	Select

Do you have a dataset you'd like to use?

☒ No ☐ Yes

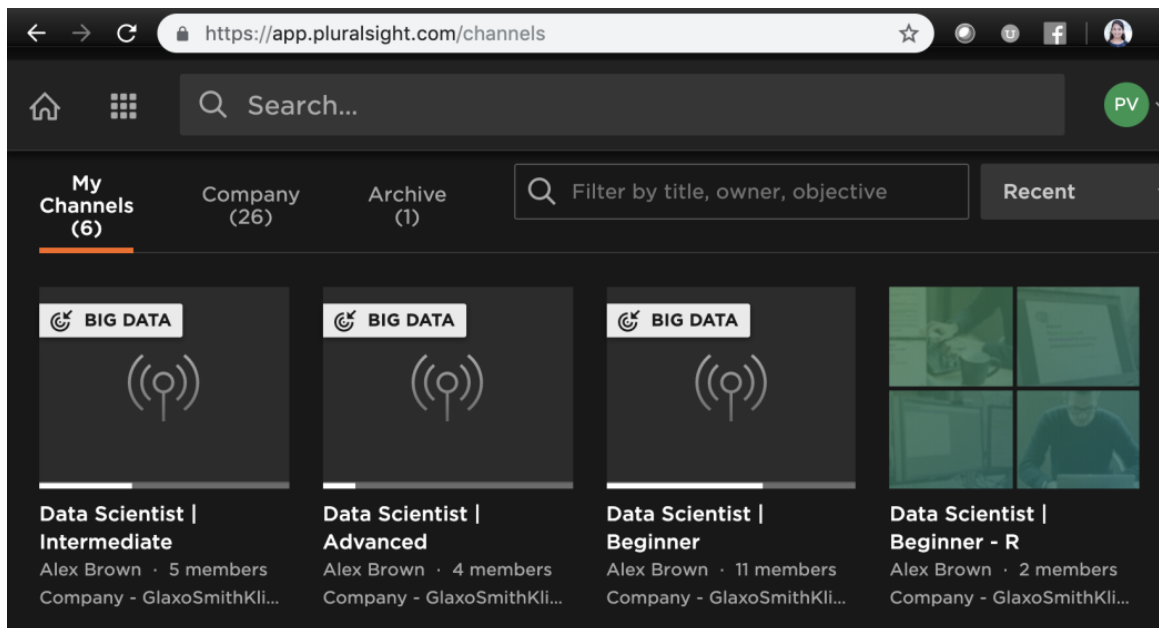
☐ By submitting the request, you agree to the **ONE Platform Data, Product, and Engineering Shared Responsibilities Agreement**.

SUBMIT **CANCEL**

2. Learning and Development

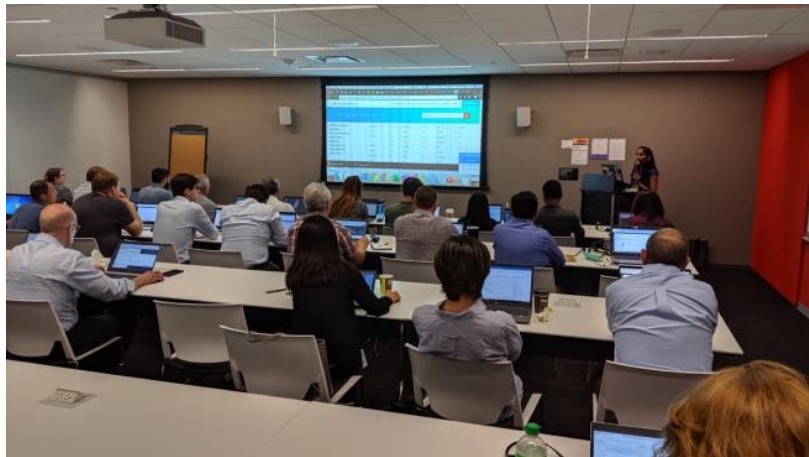


Self-paced learning through Pluralsight



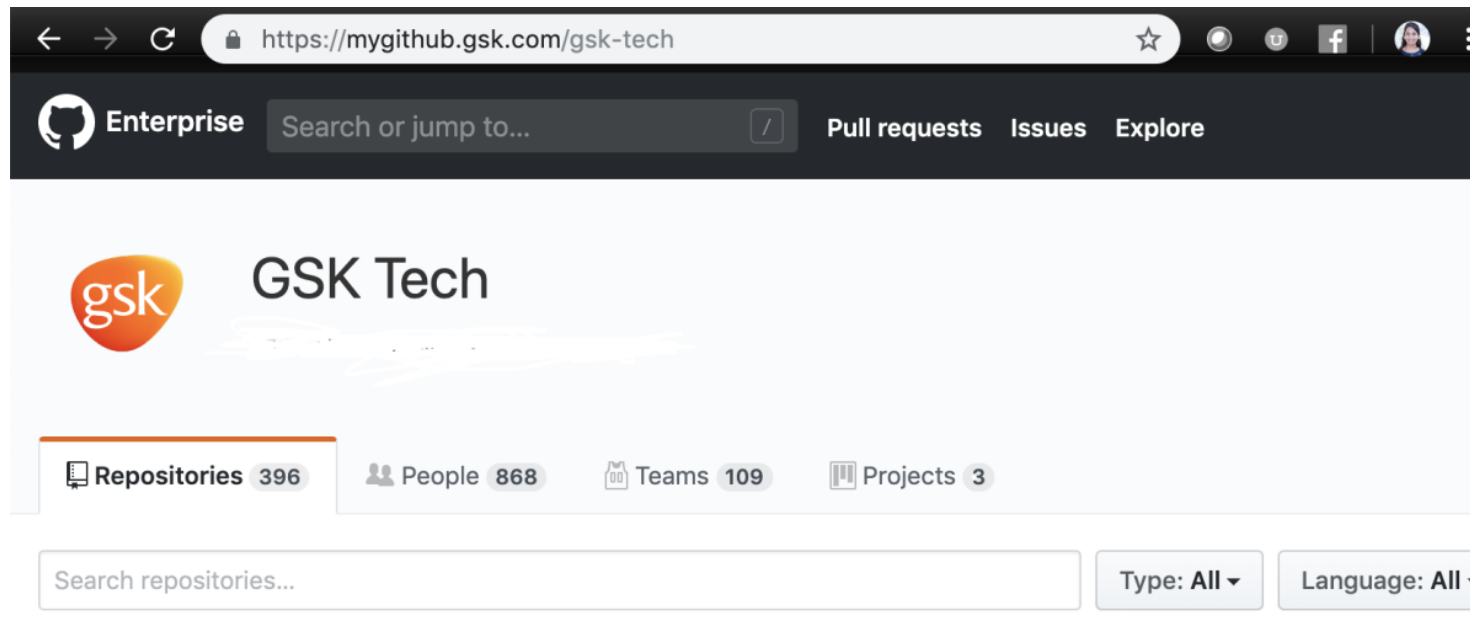
2. Learning and Development

Hands on Learnathons & Outreach programs



WOMEN IN DATA SCIENCE
DURHAM @GSK

3. Code reuse through GitHub and collaboration



Conclusion



Though we are in our initial stages of Digital Transformation journey, we are proud of the progress made so far. We are leveraging agile methodologies by collecting feedback, measuring progress and constantly refining our future strategy to increase speed and efficiency in democratizing Data science.

Questions?