

Data science without the data

Rhian Davies*

Résumé (max 300 mots)

As data scientists, we sometimes find ourselves faced with the daunting task of writing code without actually seeing the data we are working with. Whether it's due to data privacy concerns, limited access, or simply data that has not yet been collected, we often have to rely on incomplete or synthetic data to develop and test our code.

In a recent project, we worked on patient-level data. As such, the controls around the data and analysis (were rightfully) tightly controlled. We'll share how we used dummy data and mock-ups to inform code development, maintaining flexibility and adaptability in the face of changing data requirements. We'll also discuss the importance of and collaboration between developers and subject experts to ensure that code is developed with a deep understanding of the data domain

By understanding these challenges and developing effective strategies for overcoming them, we can ensure that our code is robust, reliable, and effective, even in the absence of direct data access.

Développement

Références

*Jumping Rivers, rhian@jumpingrivers.com