

## Baptiste Gorteau

4710 rue Fulton  
Montréal, QC, Canada  
[LinkedIn](#), [GitHub](#)

baptiste.gorteau@gmail.com  
[bgorteau.github.io](https://bgorteau.github.io)  
+33 7 83 47 87 18

---

### FORMATION

**Université Rennes 2**, Rennes, France

*Master*, Mathématiques Appliquées et Statistiques (MAS), septembre 2023

*Enseignements principaux* : Python, R, Modèles Linéaires, Séries Temporelles, Deep Learning

**Université Rennes 2**, Rennes, France

*Licence*, Mathématiques et Informatique Appliquées aux Sciences Humaines et Sociales (MIASHS), juin 2021

*Enseignements principaux* : Python, Gestion de Bases de Données, Algèbre, Data Mining, Économie, Sociologie

### EXPÉRIENCE

**Consultant Data**

avril 2024 – avril 2025

Équipe du Rapport GEM de l'UNESCO

Paris, France (À distance)

- Développement d'un pipeline de données en utilisant R et SQL pour créer une base de données d'aide à l'éducation à partir des données de l'International Aid Transparency Initiative (IATI).
- Conception de rapports Excel automatisés à l'aide de R et Python pour l'extraction et l'analyse de données.
- Collaboration avec l'UNICEF pour l'élaboration d'un rapport sur l'usage des données IATI dans l'aide à l'éducation.

**Développeur – Data Scientist**

mai 2022 – septembre 2023

Capgemini T.S.

Rennes, France

- Implémentation de règles de gestion avec PySpark (Python) et validation via des tests unitaires automatisés (Pytest).
- Mise en œuvre d'un pipeline ETL Big Data avec PySpark, Nifi, HDFS et environnements Cloud.
- Développement d'un système de détection d'anomalies dans des fichiers de log à l'aide de méthodes de machine learning, en utilisant Scikit-Learn et TensorFlow.
- Débogage de code Python, R et SQL.

**Data Scientist – Stage**

avril 2021 – juillet 2021

Laboratoire Mouvement, Sport et Santé

Bruz, France

- Analyse et visualisation de données de tracking avec Python dans le cadre d'un projet de recherche sur la performance en football et la réalité virtuelle.
- Analyse de séries temporelles issues de données de tracking pour comparer les trajectoires des joueurs et développer des modèles prédictifs de mouvements.
- Création de rapports et de présentations à l'aide de LaTeX et PowerPoint.

## PROJETS

**Création d'une base de données des performances de fond et de demi-fond au niveau international** – Projet Personnel (avril 2025)

Extraction automatisée de données à partir de fichiers PDF avec Python pour constituer une base de temps intermédiaires (100 m) sur des courses de fond et demi-fond en compétitions internationales (2019–aujourd'hui).

**Finaliste de la Big Data Cup 2025 organisée par Stathletes** – Compétition Data (février 2025)

Développement de métriques à partir de données de tracking pour évaluer la disponibilité des joueurs de hockey sur la glace.

**COMPÉTENCES** **Langues:** Anglais (C1), Espagnol (B1)

**Logiciels et langages de programmation:** Python, R(Studio), SQL, HTML, Microsoft Office — (Word, Excel, PowerPoint, Outlook),  $\text{\LaTeX}$ .

**Packages Python:** Pandas, Numpy, BeautifulSoup, Selenium, Scikit Learn, TensorFlow, Matplotlib