DATA ENGINEER

Bet Bardají

Data Engineer at Schneider Electric with ~3 years of experience building scalable data pipelines in Databricks (PySpark) for finance reporting. Previously collaborated with research groups in Argentina and Malaysia, gaining expertise in statistical analysis and medical data. Holder of two Master's degrees in Data Science. I aspire to combine my expertise in data engineering with my background in biology to drive advances in biomedical research and contribute to finding data-driven solutions for neurological diseases.

PERSONAL DETAILS

☎ 0034 652919807

✉ bardajibet@gmail.com

🕊 1ST OF FEBRUARY 2001

🌍 Location: Barcelona, Spain (open to hybrid/remote)

🔗 LinkedIn: https://www.linkedin.com/in/bet-b-320431251/

SKILLS

Programming & Big Data
Python (Advanced), PySpark, SQL, R, Spark SQL, YAML, Pandas, NumPy, Scikit-learn

Data Engineering & Cloud
Databricks (Workflows, DLT, SQL Notebooks), Delta Lake (time travel, optimization, Azure Data Lake (ADLS), Azure Data Factory (ADF), AWS (Redshift), ETL/ELT pipelines, CI/CD with Git

Data Governance & Security
Row-Level Security (RLS), GDPR/PII compliance, Data Quality validation frameworks, Approval workflows, Metadata management, Data Lineage tracking

Data Warehousing
Synapse Analytics, Amazon Redshift

Visualization & BI
Tableau, Power BI, Alteryx, Excel (pivot tables, advanced formulas, macros)

Other Tools & Methods
Web Scraping, API Integration (REST/JSON), Machine Learning (clustering, prediction models, CNN basics), Automation with Power Automate & Power Apps, Docker (basic), Agile/Scrum collaboration

LANGUAGES

- Catalan — Native
- Spanish — Native
- English — C2 (Full Professional Proficiency)
- German — A2 (Basic)
- French — A1 (Beginner)

EXPERIENCE

DATA ENGINEER

Schneider Electric

March 2024 - Present

- Built finance data pipelines in Databricks (PySpark, SQL, Delta Lake), ensuring reliable reporting for global finance teams.
- Automated recurring processes, reducing manual workload by ~40%.
- Developed Tableau dashboards to support financial analysis and decision-making.
- Optimized pipeline performance, improving processing speed and reliability.

DATA SCIENTIST INTERN

Schneider Electric

2023 -2024

- Designed algorithms to clean duplicates across finance databases, improving data accuracy.
- Developed Tableau dashboards and visualizations for finance KPIs.
- Conducted clustering, web scraping, and ML experiments (DeepL, NLP).
- Built KPI tracking with SemRush to support marketing analytics.

ASSOCIATE RESEARCHER

Universidad Abierta Interamericana

2021 - 2022

- Conducted research on the Renin-Angiotensin system to explore new strategies for Parkinson's treatment.
- Built and validated prediction models (CNNs, statistical models) using a patents database with thousands of records.
- Applied text mining to extract insights from scientific literature and patent data.
- Designed and maintained data pipelines to preprocess biomedical datasets for modeling.
- Collaborated with an international research team, presenting weekly updates and ensuring reproducibility of analyses.
- Contributed to research publications by preparing figures, data visualizations, and summaries of results.

DATA ANALYST INTERSHIP

Monash University,Malaysia

2020 - 2021

- Assisted in research on epilepsy development in zebrafish models.

- Designed experiments and implemented data collection and filtering protocols.

- Processed large volumes of biological data, applying statistical analysis and ML models to predict patient outcomes.

- Built and tested ML models (classification/regression) to identify developmental patterns.

- Used Python, R, and visualization tools to explore datasets and communicate findings to the research team.

- Supported publications and academic reports with analysis, ensuring data integrity and accuracy.

EDUCATION

BACHELOR'S IN BIOLOGY

Universitat de Barcelona

2018 - 2022

Specialization in Neurobiology and Genetics.

MASTER IN BIOMEDICAL DATA SCIENCE

Universitat Rovira Virgili

2022 -2024

Focus: Statistical modeling, Python, R, data analysis in healthcare.

MASTER IN DATA SCIENCE

Nuclio Digital school

2022-2023

Focus: Machine Learning, Big Data, Advanced Analytics.

ADDITIONAL EDUCATION

Johns Hopkins University (2014–2015, 2016–2017):
Online certifications in Algebra, Mathematical Problem Solving, Mathematical Logic, Awarded High Honours.

Coursera (2021–2022):
Master in Clinical Neuropsychology,

Universitat de Barcelona(2022):
Neuro-linguistic Programming (NLP), Emotional Wellbeing, Rare Diseases.

AWARDS

- Award of High Honours, by the Johns Hopkins University, 2016.
- Award for the Best Research Project at High School Badalonés (Barcelona) for the ¨Epigenetics on monozygotic twins¨ research

VOLUNTEERING

- Association of Single Parent Families (2015–2020) — support tasks and childcare during holidays.
- CorsAvi (2020–2021) — volunteer writing letters to seniors in nursing homes.