

PROFILE

Data Engineer at P&G, building data pipelines for customer data (from 600+ retailers into a unified data lakehouse), enabling over \$800M in business value.

Proficient in Python (Polars, PySpark, Pandas), SQL (BigQuery, SparkSQL, DBT), Databricks (Certified Data Engineer Associate) and cloud technologies (Azure and GCP certified).

Proven ability to implement batch and streaming data architectures, automate processes, optimize cloud costs and enhance data quality and discoverability.

Experienced in AI and ML (Scikit-learn, TensorFlow, Keras, XGBoost) and author of a master's thesis on machine learning time-series forecasting models.

PROFESSIONAL EXPERIENCE


Data Engineer

Procter & Gamble 

Sep 2022 – present

- Developed and maintained scalable batch data pipelines using Pyspark on Databricks.
- Created real-time streaming data architecture for Amazon retail and search data using Apache Kafka and Spark Streaming, achieving sub-second latency.
- Configured CI/CD with Databricks asset bundles deployment (configuration as code), automated testing (Pytest), linting/formatting (Ruff) and Static code analysis (SonarQube and Snyk)
- Optimized data processing with Polars & DuckDB, achieving 40% cloud cost savings over PySpark pipelines
- Automated data ingestion from Carrefour, using BigQuery Analytics Hub, reducing backfill time from weeks to hours.
- Automated publication of metadata and data quality checks, promoting data discoverability.

Data Analyst

Samba TV 

Oct 2021 – May 2022

- Automated people analytics from 5+ sources with REST API data integration, AWS Lambda serverless computing, Pandas, secure secret storage in AWS Secret Manager, and scheduling with AWS CloudWatch.
- BI reporting using Sisense and Google Data Studio.
- Promoted from intern to associate within 4 months.

CERTIFICATES

- Databricks Data Engineer Associate
- Azure Data Fundamentals
- GCP Data to Insights



Leonardo Bocci

Data Engineer



 leonardobocci99@hotmail.com

 leobocci.pages.dev

 +48 696351025  Warsaw, PL

SKILLS

Data Engineering

Python, SQL, Databricks, ETL/ELT, Data Warehousing, Data Modeling, Data Quality, APIs, Data Pipelines, Cloud Architecture.

Data Science/ML Engineering

AI/ML, Regression, Classification, Clustering, Statistical Inference, AB Testing, Experiment tracking.

EDUCATION

Kozminski University

- MSc, Finance, Honours. Thesis on ML timeseries forecasting models.
- BSc, Finance, Honours. Thesis on value relevance of Formula 1 results.

Notable Projects

- Winner in RoboCup and RomeCup robotics competitions, national finalist in CFA Equity Research Challenge, finalist in the AI in ESG investing competition.
- Contributed to the Kozminski AI club by consulting cloud architecture for a local start up.

LANGUAGES

English — Fluent

Polish — Fluent

Italian — Native/Bilingual