



# David Jair Landeo Vargas

Data Engineer | Data Analyst

## Profile

**Data Engineer** with **3 years of experience** designing and optimizing scalable **data pipelines** and analytics solutions. Strong background in **Big Data**, building **ETL** processes in **Python, PySpark**, and **Databricks** across Bronze and Silver layers. Experienced in developing **Power BI dashboards** that enable data-driven decision-making, and implementing data quality checks on production pipelines. Proven ability to **automate workflows, improve performance**, and deliver **high-quality data products** in enterprise environments.

## Work Experience

### Data Engineer | Procter & Gamble

Sep 2023 – present

- Built and maintained ETL pipelines in Databricks (PySpark) to load data across Bronze and Silver layers.
- Refactored a legacy PySpark codebase, reducing pipeline runtime from several hours to ~1 hour.
- Designed and developed 2 Power BI dashboards to track key business metrics, used by business and engineering stakeholders.
- Adapted CI/CD bundles from existing projects to ensure consistent deployments across data pipelines.
- Implemented data quality checks and alerting on production pipelines.
- Collaborated with the team using Git and Databricks Repos, including branching workflows and code reviews.
- Managed and ensured the accuracy and functionality of Master Data.
- Designed and implemented automated workflows in Power Automate, reducing workload and streamlining processes.

### AI & Big Data Analyst | Instituto Escalae

Sep 2023 – February 2025

- Implemented AI-driven code for translating files into Spanish, English, Portuguese, Finnish, and Swedish.
- Created a web API on Azure that transcribes videos into text via web requests.
- Automated and simplified the process of sending certificates to thousands using Open Badge Factory.

### Data Engineer Intern | Rentals United

May 2023 – July 2023

- Created web scraping scripts for acquiring contact details of prospective clients.
- Implemented automation scripts to optimize processes and reduce working time.
- Automated pipelines connected with Google Sheets, BigQuery and APIs for data extraction.

## Languages

Spanish - Native

English - Advanced

## Additional Information

- [Water polo player](#) for the national team of Peru for a period of 3 years
- Recognized as a qualified athlete in Peru
- [Student Representative](#) before the University Assembly
- Swimmer on the university swim team

✉ davidwp37@gmail.com

☎ +34 695 12 87 62

📍 Madrid, Madrid

🌐 [linkedin.com/in/david-landeo/](https://www.linkedin.com/in/david-landeo/)

🐙 [github.com/ockdavid](https://github.com/ockdavid)

## Education

- Barcelona Technology School | Universitat de Barcelona  
[Master of Big Data and Artificial Intelligence Solutions](#)  
October 2022 - July 2023
- Pontificia Universidad Católica del Perú  
[Electronic Engineering](#)  
March 2013 - June 2019
- Universidad de Vigo  
[Student Exchange Program](#)  
January 2019 - June 2019

## Expertise

- Python, Pyspark, SQL, Git, Github, Linux, Ubuntu
- Databricks, Airflow, Docker, Flask
- Scikit-learn, PyTorch, Selenium, Streamlit
- PowerBI, Excel, Tableau, Power Automate, VBA

## Projects

- [Developed](#) an [AI-powered](#) app
- [Developed](#) a word cloud generator [app](#) for WhatsApp chats
- [Built](#) an [NLP notebook](#) for sentiment analysis of whatsapp chat
- [Created](#) a machine learning [model](#) to predict used car prices
- [Wrote](#) a web scraper [script](#) to analyze shared bedroom prices