

# BENJAMIN OFURHIE

Lisbon, Portugal | Ofurhieb@yahoo.com

[LinkedIn](#) | [Portfolio](#)

## TECHNICAL SKILLS

- **Languages:** Python (Pandas), SQL (BigQuery, MySQL, Postgres), PySpark.
- **Databases:** PostgreSQL, MySQL, BigQuery, MongoDB (basic familiarity)
- **ETL & Data Processing:** Data cleaning, transformation pipelines, workflow automation
- **Cloud & Tools:** AWS (S3, Redshift, Glue, Lambda), GCP, Databricks, Docker
- **Environment:** Git, Docker,
- **Visualization:** Power BI, Looker Studio, Tableau, Excel, Google Sheets.

## Education & Certification

- **Bachelor of Medicine and Surgery** | VN Karazin National University.
- **IBM Data Engineering** Professional Certificate.
- **Google Data Analytics** Professional Certificate.

## EXPERIENCE

### Data Operations Specialist | Dealroom.co

Lisbon, Portugal | 2024

- Designed and maintained Python-based ETL processes to extract, clean, and transform API data into structured datasets for dashboard and KPI reporting.
- Wrote and optimized complex SQL queries in BigQuery to support reporting, business analysis, and metric definition.
- Built automated data validation checks and monitoring routines to ensure data integrity and consistency across reporting layers.
- Performed advanced data cleaning, transformation, and exploratory analysis using SQL and BigQuery to generate actionable insights for leadership.
- Orchestrated and monitored data workflows using Apache Airflow to ensure reliable and timely data delivery.

### Data Analyst | Meri Skill

Remote | 2022 – 2023

- Designed and automated reporting workflows using Power Query, Excel, and SQL to streamline recurring business reports.
- Built and maintained Python (Pandas) data pipelines to process and structure large datasets for operational and performance analysis.
- Analyzed sales and revenue data to define KPIs and support cross-functional decision-making across marketing and operations teams.
- Developed interactive dashboards and KPI visualizations in Power BI and Excel to communicate business performance to non-technical stakeholders.
- Identified and resolved data inconsistencies through structured analysis, improving reporting accuracy and reliability

## **DATA ENGINEERING PROJECTS- Project Repo**

### **Uber ETL Pipeline (Airflow & GCP) – Github Repo**

- Designed a scalable ETL pipeline to extract, transform, and load Uber ride data into Google Cloud Storage.
- Automated daily data workflows using Apache Airflow with retry logic and task orchestration to ensure reliable data availability.
- Built Python (Pandas) data cleaning and validation routines to standardize fields and handle missing values before analysis.
- Structured transformed data into analytics-ready tables to support reporting and performance analysis.
- Deployed the pipeline using Docker to ensure consistent and reproducible execution across environments.

### **Data Warehouse & Modeling Project – Github Repo**

- Built end-to-end ETL pipelines to ingest and process CRM and ERP data into structured analytics-ready datasets.
- Designed star-schema fact and dimension tables to optimize SQL querying and support business reporting.
- Developed complex SQL transformations using joins, window functions, and conditional logic to prepare BI-ready data.
- Cleaned and standardized datasets through deduplication and field validation to improve data quality and consistency.
- Implemented a scalable data warehouse structure to support dashboards, KPI tracking, and automated reporting.

### **Real-Time Flight Data ETL Pipeline – GitHub Repo**

- Built an end-to-end ETL pipeline to ingest and process real-time aviation data from the OpenSky API into structured, analytics-ready datasets.
- Built API connections to collect raw data and convert it into clean tables for analysis and reporting.
- Automated workflow scheduling and monitoring using Apache Airflow to ensure consistent and reliable data updates.
- Implemented data enrichment and aggregation logic to generate country-level flight metrics, including total flights, average velocity, and ground status.
- Applied error handling, logging, and retry mechanisms to improve pipeline reliability and data consistency.
- Deployed the solution using Docker to enable reproducible execution across environments.