

BENJAMIN OFURHIE

Lisbon, Portugal | Ofurhieb@yahoo.com

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

TECHNICAL SKILLS

- **Languages:** Python, SQL (BigQuery, MySQL, Postgres), PySpark.
- **Data Engineering:** ETL/ELT Pipeline Development, Data Modeling (Star Schema), Data Quality Assurance.
- **Cloud & Tools:** Google Cloud Platform (GCP), AWS (S3, Redshift), Databricks, Azure Data Studio, Data Build Tool(dbt), Apache Airflow, Docker, GitHub/Version Control.
- **Visualization:** Looker Studio, Power BI, Tableau, Excel.

EXPERIENCE

Data Moderator | Dealroom.co

Lisbon, Portugal | 2024

- Developed scalable Python ETL pipelines and web scraping solutions using BeautifulSoup and Pandas to integrate high-volume datasets from APIs, improving efficiency by 30%.
- Partnered with data engineers and product teams to define data requirements and optimize end-to-end pipelines, ensuring smooth data flow.
- Designed and implemented automated data quality checks, validation routines, and alert systems to ensure accuracy and integrity for analytics.
- Leveraged SQL and BigQuery on Google Cloud Platform for complex data cleaning, transformation, and exploratory analysis to support strategic decision-making.
- Managed pipeline scheduling, orchestration, and monitoring using Apache Airflow to guarantee timely and scalable data delivery across platforms.

Data Analyst | Meri Skill

Remote | 2022 – 2023

- Designed and built automated reporting workflows using Power Query, Excel, and SQL, reducing manual effort and boosting reporting efficiency.
- Built scalable data pipelines using Python, Pandas, and MySQL to process and manage large, structured datasets for business operations.
- Conducted structured data analysis on sales and revenue metrics to support cross-functional decision-making and business strategy.
- Developed dynamic dashboards and KPI visualizations to communicate complex business trends to non-technical stakeholders effectively.
- Investigated data inconsistencies through independent analysis, enabling critical process improvements and strategic organizational insights.

DATA ENGINEERING PROJECTS- Project Repo

Uber ETL Pipeline (Airflow & GCP) – Github Repo

- **Implementation:** Designed a scalable ETL pipeline to extract, transform, and load Uber ride data into Google Cloud Storage (GCS).
- **Automation:** Automated daily workflows using Apache Airflow, incorporating retry logic and task orchestration for consistent data delivery.
- **Transformation:** Developed data cleaning and validation routines with Python and Pandas to standardize columns and handle missing values.
- **Workflow:** Utilized XCom in Airflow to pass metadata between tasks, improving overall workflow efficiency and maintainability.
- **Deployment:** Implemented containerized deployment options with Docker for reproducible and portable ETL execution.

Data Warehouse & Modeling Project – Github Repo

- **Architecture:** Built end-to-end ETL pipelines following a Bronze → Silver → Gold architecture for CRM and ERP data ingestion.
- **Modeling:** Designed star-schema dimension tables (dim_customers, dim_products) and fact tables (fact_sales) for optimized querying.
- **Transformation:** Developed complex SQL transformations using window functions, joins, and conditional logic to prepare data for BI reporting.
- **Quality:** Performed deduplication and standardization of datasets, validating fields such as dates, genders, and product categories.
- **Infrastructure:** Created a scalable data warehouse architecture designed to support analytics, BI, and automated reporting.

Real-Time Flight Data ETL Pipeline – GitHub Repo

- **Architecture:** Engineered a production-grade ETL pipeline using a Bronze → Silver → Gold medallion architecture to process real-time aviation data from the OpenSky Network API.
- **Ingestion:** Built robust REST API integrations to extract live flight state data, transforming nested JSON payloads into analytics-ready tabular formats.
- **Orchestration:** Automated daily workflows with Apache Airflow, leveraging XCom for inter-task data exchange between pipeline layers.
- **Transformation:** Developed data enrichment and aggregation logic to compute country-level flight metrics such as total flights, average velocity, and ground status.
- **Reliability:** Implemented production-ready retry policies, failure handling, and logging to ensure consistent pipeline execution and high data reliability.
- **Deployment:** Containerized the pipeline with Docker to enable reproducible execution across local and cloud environments.

Education & Certification

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