BENJAMIN OFURHIE

Lisbon, Portugal
[LinkedIn](#) | [Portfolio](#) | Ofurhieb@yahoo.com

EDUCATION

VN Karazin National University   July 2023
Bachelor of Medicine and Surgery

CERTIFICATIONS
Google Data Analytics
IBM Data Science

TECHNICAL SKILLS

- Python
- Scikit-Learn/Linear Reg
- Google Cloud platform
- Looker Studio
- Amazon Webservice
- Google sheet/Excel
- Power BI /Tableau
- Microsoft Office Suite
- Apache Airflow
- Azure Data Studio/Postgres
- BigQuery/ SQL
- Tableau
- GitHub / Version control
- Docker

INDUSTRY EXPERIENCE

Data Moderator – Dealroom.co   **Lisbon, Jan 2024**

- Partnered with data engineers, analysts, and product teams to define data requirements and optimize end-to-end data pipelines, ensuring smooth and reliable data flow.
- Developed scalable Python ETL pipelines and web scraping solutions (BeautifulSoup, Pandas) to extract, process, and integrate high-volume datasets from APIs and web sources, improving processing efficiency by 30%.
- Designed and implemented automated data quality checks, validation routines, and alert systems, ensuring accuracy and integrity of datasets for reporting and analytics.
- Leveraged SQL and BigQuery on Google Cloud Platform for data cleaning, transformation, aggregation, and exploratory data analysis to support strategic decision-making.
- Built AI-powered orchestration services using existing APIs (e.g., Gemini) to automate content generation and analytical workflows, enhancing productivity and reducing manual intervention.
- Collaborated cross-functionally with data curation, engineering, and product teams to deliver actionable insights and promote a data-driven culture.
- Managed pipeline scheduling, orchestration, and monitoring to guarantee timely, accurate, and scalable data delivery across platforms.

Data Analyst - Meri Skill.   **Jul 2022 - Nov 2023**

- Designed and built automated reporting workflows using Power Query, Excel, and SQL, reducing manual effort and boosting reporting efficiency.
- Conducted ad hoc and structured data analysis on sales and revenue metrics to support cross-functional decision-making.
- Developed dynamic dashboards and KPI visualizations to communicate business trends to non-technical stakeholders.
- Built scalable data pipelines using Python, Pandas, and MySQL to process and manage structured datasets.
- Investigated data inconsistencies and trends through independent analysis, enabling process improvements and strategic insights.
- Collaborated with product and operations teams to align data-driven insights with evolving business goals.

DATA ENGINEERING ETL PROJECTS

Uber ETL Pipeline with Airflow & GCS – GitHub
End-to-end ETL project for Uber ride data leveraging Airflow, Python, and GCP

- Designed and implemented a scalable ETL pipeline to extract, transform, and load Uber ride data into Google Cloud Storage (GCS) for analytics and visualization.
- Automated daily data workflows using Apache Airflow, including retry logic and task orchestration, ensuring reliable and consistent data delivery.
- Developed data cleaning and transformation routines with Python and Pandas, standardizing columns, handling missing values, and validating datasets for quality assurance.
- Leveraged XCom in Airflow to pass metadata between tasks, improving workflow efficiency and maintainability.
- Built Looker Dashboard connected to cleaned datasets enabling analytics such as trip trends, revenue reports, geospatial analysis, and driver performance metrics.
- Implemented containerized deployment options with Docker for reproducible and portable ETL execution.

Data Warehouse ETL - GitHub
Tools: SQL, MySQL, ETL, Star Schema, Data Modeling, Data Cleansing, Data Pipeline

- Built end-to-end ETL pipelines (Bronze → Silver → Gold) for CRM and ERP data ingestion.
- Cleansed, deduplicated, and standardized datasets; validated dates, genders, and product categories.
- Designed star-schema dimension (dim_customers, dim_products) and fact (fact_sales) tables.
- Developed SQL transformations with window functions, joins, conditional logic, and aggregations.
- Created scalable, maintainable data warehouse architecture for analytics, BI, and reporting.

Key Skills: ETL Pipeline Development, Data Engineering, Workflow Automation, Apache Airflow, Python, Pandas, Google Cloud Storage, Data Quality, Looker Studio, Docker

Data Analytics SQL - GitHub
Tools: Azure Data Studio, SQL

- Built customer segmentation models (VIP, Regular, New) using behavioral and demographic data.
- Analyzed product performance with time-series metrics, rolling aggregates, and moving averages.
- Calculated key KPIs including recency, average order value, monthly spend, and customer lifespan.
- Developed advanced SQL scripts using joins, CTEs, window functions, and conditional logic.
- Created recruiter-ready datasets for BI reporting and dashboards.

Key Skills: SQL, Joins, Subqueries, CTEs, Window Functions, Aggregation, Data Analysis, KPI Calculation, Data Visualization, Schema Design, Git/GitHub

MACHINE LEARNING PROJECTS

Telco Customer Churn Prediction - GitHub
Tools: Python, scikit-learn, Pandas, Matplotlib

- Built a machine learning model to predict customer churn using service usage, contract type, and demographics.
- Performed data preprocessing, exploratory data analysis, and feature encoding with DictVectorizer.
- Identified key churn drivers: month-to-month contracts, short tenure, high monthly charges, and service features.
- Trained and evaluated a Logistic Regression model achieving **81% test accuracy**.
- Produced actionable insights for customer retention strategies and predictive analytics dashboards.

Key Skills: Python (Pandas, NumPy), Scikit-learn, Logistic Regression, EDA (Matplotlib, Seaborn), Feature Engineering, Accuracy Evaluation, Predictive Modeling