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
- Google sheet/Excel
- PowerBi /Tableau
- Microsoft Office Suite
- Apache Airflow
- 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 Intern- 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 Repository
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.

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

MACHINE LEARNING PROJECTS ---Github Repositiory

Customer Churn Prediction | Python, scikit-learn, Pandas, Matplotlib

- Objective: Predicted customer churn for a telecom company to help reduce attrition.
- Approach: Cleaned and engineered features from usage data, trained a Random Forest model.
- Impact: Improved churn identification logic and supported targeted interventions to reduce attrition.

Link: GitHub

Used Car Price Prediction | Python, scikit-learn, Pandas, NumPy, Matplotlib, Jupyter Notebook

- Objective: Predicted used car prices based on features like brand, mileage, year, and engine size to support pricing decisions.
- Approach: Performed data cleaning, feature engineering, and trained regression models including Linear Regression/Random Forest.
- Impact: Achieved an R^2 score of 0.92 with MSE value of 0.057 on the test set, enabling more accurate pricing insights and reducing average prediction error.

Link: GitHub