

Anwar Mousa Anwar

📍 Cairo, Egypt ✉ anwarmousa100@gmail.com ☎ +201144162459

in LinkedIn 🐙 GitHub

Summary

As a Data Enthusiast with a passion for data engineering and analytics, I have strong skills in data modeling, ETL processes, and cloud technologies. With hands-on experience using AWS services, I am eager to apply my technical skills to data engineering roles. My goal is to optimize data pipelines, improve data management processes, and deliver actionable insights that drive business success.

Education

Bachelor of Computer Science and Informatics *Aug 2022 – Jun 2026*
Delta University for Science and Technology

Data Analytics Scholarship - Digital Egypt Pioneers Initiative (DEPI) *Oct 2024 – May 2025*
Ministry of Communications and Information Technology (MCIT), Egypt

Experience

Data Analyst Intern *Sep 2024 – Oct 2024*
NeuronetiX, Hybrid

- Performed statistical analysis on large datasets using Python and SQL to identify trends and patterns for strategic decision-making.
- Conducted Exploratory Data Analysis (EDA) using Pandas, NumPy, and Matplotlib to extract meaningful insights from raw data.
- Created visualizations and reports using Power BI and Tableau to present data analysis results.

Business Intelligence Intern *Jun 2024 – Aug 2024*
PwC, Remote

- Built interactive dashboards using Power BI to improve data visualization for better decision-making
- Analyzed data to identify business trends and patterns, delivering actionable insights for strategic decisions
- Collaborated with cross-functional teams to align data strategies with business objectives, ensuring seamless project execution.

Projects

Real-time Forex Data Pipeline & Analytics Dashboard (Github Link [🔗](#))

- Developed a comprehensive real-time forex data pipeline using Python, **Snowflake**, and **Streamlit** that processes and visualizes currency exchange rates, with a focus on USD/EGP analysis.
- Implemented automated workflows with **Apache Airflow** for real-time data ingestion, transformation, and storage in Snowflake.
- Created an **interactive dashboard** with real-time candlestick charts, technical indicators (RSI, MACD, EMA), volatility and trend analysis, auto-refresh visuals, and custom export options.
- Engineered a robust data architecture with **multi-layer modeling (Bronze, Silver, Gold)**, automated data quality checks, error handling, logging, and secure credential management.

NASA Log Analytics Pipeline (Github Link [🔗](#))

- Developed an end-to-end data processing pipeline for NASA HTTP server logs using Apache Airflow, Spark, and **PostgreSQL**.
- **Processed over 3.4 million log entries (61+ GB)** from NASA's web server logs (Jul-Sept 1995)
- Implemented automated workflow orchestration with Airflow DAGs for data ingestion, transformation, and loading.
- Built distributed processing capabilities with **Apache Spark** for efficient large-scale log analysis
- Created comprehensive analytics features including traffic pattern analysis, error rate monitoring, and resource usage trends.
- Designed interactive dashboards in **Grafana** to visualize key metrics and insights.

Car Price Prediction ETL Pipeline (Github Link [↗](#))

- Engineered a prediction model achieving **98% accuracy** using **Scikit-learn's** gradient boosting algorithms and advanced feature engineering techniques
- Implemented robust **ETL processes** with Pandas for data manipulation and **PostgreSQL** for efficient database storage and retrieval
- Created comprehensive performance monitoring system with Matplotlib visualizations for model evaluation and error analysis
- Developed specialized depreciation metrics and non-linear transformations in Python to enhance prediction capabilities.

Real-Time API Data Pipeline (Github Link [↗](#))

- Built an end-to-end real-time data pipeline using **Python, PostgreSQL, and Apache Airflow** to ingest and process live API data.
- Automated data cleaning and transformation using **Pandas** and **Airflow DAGs** to ensure data consistency.
- Containerized the pipeline using **Docker**, enabling seamless deployment and scalability across environments.
- Implemented logging and monitoring mechanisms to track data freshness and pipeline performance proactively.

E-commerce Sales Data Analysis (Github Link [↗](#))

- Performed **Exploratory Data Analysis (EDA)** on e-commerce sales data using **Python** to uncover trends, patterns, and performance insights.
- Used multiple Python libraries, including **Pandas, NumPy, Matplotlib, Seaborn, and Plotly** for data cleaning, analysis, and visualization.
- Ensured high data quality through robust **data validation** and **automated cleaning pipelines**, enhancing reporting accuracy.

Technical Skills

Programming Languages:	Python (Pandas, NumPy, scikit-learn), SQL, Scala
Database Systems:	PostgreSQL, Microsoft SQL Server, MySQL
Data Processing:	ETL Pipeline Design, Apache Airflow, SSIS
Cloud & Big Data:	AWS (EC2, S3, RDS), Hadoop, Spark, Kafka
Data Architecture:	Data Warehouse, Dimensional Modeling, Data Lake
Visualization:	Power BI, Tableau
Development Tools:	Docker, Git, GitHub, Linux, Bash

Certifications

- **AWS Certified Cloud Practitioner** [↗](#)
- **Mathematics for Machine Learning and Data Science Specialization** [↗](#)
- **IBM Data Analyst Specialization** [↗](#)
- **Data Analyst Professional** [↗](#)
- **Google Data Analytics Certificate** [↗](#)