Mostafa Emad

Data Engineer

✉ mostafaelnagar.ds@gmail.com 📞 +201117084837 📍 Cairo, Egypt 💼 [LinkedIn](#) 🐙 [Github](#)

PROFESSIONAL SUMMARY

Data Engineer with a background in mathematics and computer science. Experienced in building ETL pipelines and dashboards using tools like Spark, Airflow, Docker, and Power BI. Proven ability to improve data workflows and communicate findings clearly. Focused on hands-on work and continuous learning through real-world projects.

PROFESSIONAL EXPERIENCE

Data Engineer Trainee – Information Technology Institute (ITI)   June-2025 – Present
Port Said, Egypt

- Currently undergoing an intensive 4-month training in data engineering and business intelligence.

Data Specialist – Egyptian Army (Rehabilitation Center)   May-2024 – June-2025
Cairo, Egypt

- Managed a high-volume visitor registration system at a military rehabilitation center, overseeing the secure processing of sensitive data for 300+ daily visits.
- Generated monthly performance and visit reports for senior leadership, aiding in operational planning and decision-making.
- Built an ID card data extractor using OCR and scanning tools, reducing post-registration data verification workload by 2+ hours daily.
- Redesigned the data entry workflow for the registration team, accelerating the daily registration process by 3 hours and boosting staff efficiency.
- Achieved a 50% reduction in overall registration time by integrating automated ID extraction and streamlining the entire workflow.

EDUCATION

Bachelor of Science, Pure Mathematics and Computer Science   2019 – 2023
Al-Azhar University, Egypt

- Graduated with **Excellent with Honors**, ranked #1.
- Relevant coursework:
  - **Computer Science:** Data Structures & Algorithms, Operating Systems, Database Systems, System Design, Image Processing, Artificial Intelligence
  - **Mathematics:** Linear Algebra, Calculus, Statistics, Probability, Numerical Analysis, Operations Research, Discret Mathematics, Functional Analysis

PROJECTS

FordGoBike Data Pipeline   [GitHub](#)

- Built a data pipeline with Airflow, PostgreSQL, and Docker to process over 3 million trip records monthly.
- Found that weekday peak hours had 30% more rides, helping improve bike placement and staffing.
- Cut manual reporting by 80% using automated visuals and filters.

Bike Sharing Data Exploration – San Francisco Bay Area   [GitHub](#)

- Performed exploratory data analysis on 183,412 bike rides using Python and Pandas to uncover usage trends by time, gender, age, and user type.
- Discovered that rides under 30 minutes are most common during work commutes (7–9 AM and 4–6 PM), indicating usage for short-distance daily transport.
- Revealed that customers ride longer (avg duration 22 mins) than subscribers (avg 12 mins), and male customers take longer rides than other groups.

Customer Churn Analysis Dashboard [GitHub](#)
- Built an end-to-end Power BI dashboard to analyze customer churn across 7 years (2018–2024), enabling granular filtering by year and gender.
- Monitored KPIs across 7043 customers, revealing a churn rate of 26.54%, retention rate of 73.46%, and total revenue of $5.47M.
- Quantified $1.67M in revenue loss and tracked LTV before and after churn ($2.28K vs $2.55K), highlighting upsell opportunities and long-term value patterns.
- Uncovered that customers with fewer services (1–2) had the highest churn rate (¿40%) using a combo chart of churn vs. service count.

LinkedIn Data Visualization Dashboard [GitHub](#)
- Designed and deployed a Plotly dashboard analyzing 18 months of LinkedIn connections (Jan 2022 – Jul 2023) from exported CSV data.
- Identified that 98% of connections were in data-related fields, with a majority affiliated with ITI as trainees or employees.
- Visualized top 10 affiliated companies via bar charts; ITI ranked highest, indicating strong industry alignment.

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL, Bash, DAX, Scala, Java
- **Data Engineering:** Apache Airflow, Spark, Power Query
- **Databases:** PostgreSQL, SQL Server, MongoDB, MySQL, NoSQL
- **Business Intelligence:** Power BI, Excel, Tableau, Jupyter Notebooks
- **DevOps & Cloud:** Docker, AWS, Git, GitHub Actions, CI/CD

ADDITIONAL TRAINING

Udacity – Data Analyst Nanodegree
- Intro to Data Analysis, Advanced Data Wrangling, and Data Visualization using Pandas, NumPy, Matplotlib, and Seaborn

Coursera – Introduction to Data Science in Python
- Covered lambdas, CSV parsing, DataFrames, statistical tests, and data exploration techniques

MIT OpenCourseWare – Computer Science for Data Science
- Topics: Optimization, Graph Models, Monte Carlo Simulation, Confidence Intervals, Classification, Clustering, and Statistical Pitfalls

LANGUAGES

- **English:** Fluent
- **Arabic:** Native