

Mostafa Emad

Data Engineer

✉ mostafaelnagar.ds@gmail.com ☎ +201117084837 📍 Cairo, Egypt [in LinkedIn](#) [GitHub](#)

PROFESSIONAL SUMMARY

I am a results-driven Data Engineer with a solid foundation in mathematics and computer science, enriched by real-world experience in data systems. I excel at designing scalable ETL pipelines and creating interactive dashboards that drive strategic decisions. My expertise includes Spark, Airflow, Docker, and Power BI, resulting in data operations improvements of up to 50%. I communicate technical insights effectively to diverse stakeholders and I am dedicated to continuous learning through practical projects and formal training

PROFESSIONAL EXPERIENCE

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

- Currently undergoing intensive training in data engineering and business intelligence projects focused on real-world scenarios.
- Working with a modern tech stack including Spark, SQL, Docker, Airflow, Power BI, and Python to build scalable data pipelines and analytics dashboards.
- Collaborating with cross-functional teams to deliver project-based solutions involving data modeling, ETL workflows, and reporting automation.

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

- Managed the visitor registration system in a military rehabilitation center, handling sensitive personal data and optimizing daily operations.
- Generated monthly performance and visit summary reports to support leadership decision-making.
- Designed and implemented an ID card data extractor tool using scanning and OCR techniques, reducing registration time by 50% and improving overall workflow efficiency.

EDUCATION

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

- Graduated with Honors (Ranked 1st in the class)
- Coursework: Database Systems, Operating Systems, Artificial Intelligence, Data Structures & Algorithms, Image Processing, Operations Research, Linear Algebra, Calculus, Statistics, Probability, Numerical Analysis.

PROJECTS

FordGoBike Data Pipeline [GitHub](#)

- Designed an ETL pipeline with Airflow, PostgreSQL, and Docker to process 3M+ monthly trip records.
- Identified weekday rush hour peaks (+30% usage) and seasonal trends, improving resource planning.
- Reduced manual reporting time by 80% through interactive filtering and automated visual updates.

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.
- Found strong linear correlation between distance and duration for rides under 130 mins ($R \approx 0.8$), leading to segmentation insights by user type.
- 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

ONLINE COURSES

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