

# Yedige Ashmet

+7 (747) 244 9725 | yedigeaa@gmail.com | github.com/gradedSystem | linkedin.com/in/yedige2000/ | k kaggle.com/yedigeashmet

Data Engineer with over 2 years of experience in data processing, pipeline automation, and analytics, skilled in building scalable workflows, integrating datasets, and optimizing scripts for efficient data management.

## EDUCATION

**Data Science Tech Institute**, *Applied MSc in Data Engineering for Artificial Intelligence* |

Sophia-Antipolis, France GPA: 3.4 / 4.0

March 2024

**Astana IT University**, *Bachelors in Big Data* | Astana, Kazakhstan GPA: 3.28 / 4

May 2022

Tech Finalists: Tech-Garden IT-kvartal Finalist creating QR-code Queue system for Halyk Bank

Courses: Cloud Platforms | Semantic Web Programming | Data Pipelines | ML | Deep Learning | Cybersecurity | Artificial Intelligence | Data Governance | AWS

## EXPERIENCE

**Datopian**, *Data Engineer* | Remote (Redwood city, USA)

Aug 2023 - Present

- Website Improvements: Spearheaded improvements to Datopian's website using TypeScript and modern front-end development practices, enhancing functionality, boosting performance, and improving overall user experience, which led to a 25% faster load time and a smoother user journey.
- FedEx Project: Collaborated on delivering a structured, cleaned dataset of postal codes to FedEx, ensuring data accuracy and seamless integration.
- What-to-Fix Project: Designed and implemented a scalable data architecture for the What-to-Fix platform, building automated data pipelines that ingested, processed, and analyzed data from multiple sources. This automation increased processing speed by 40% and supported continuous data updates, providing the client with real-time actionable insights.
- Workflow Automation: Led the automation of daily data processing workflows, leveraging Google BigQuery, GitHub Actions, and Cloudflare R2, resulting in a 30% reduction in manual tasks and improved system reliability by automating critical pipeline tasks. This ensured data availability and accuracy for key stakeholders.
- Open-Source Dataset Contribution: Contributed significantly to the open-source Awesome Dataset project, building efficient workflows to automate the integration and processing of datasets. Replaced legacy scripts with a fully automated pipeline, improving both performance and maintainability. These efforts enabled the project to handle larger data volumes while reducing operational overhead.

**Halyk Bank**, *Data analytics intern* | Almaty

May 2022 - Aug 2022

- Implemented daily SQL scripts to improve departmental satisfaction and operational efficiency.
- Assisted in the development and deployment of Grafana dashboards to monitor daily performance metrics, revenue, and statistical data for the bank.

**Kazakh Temir Zholy**, *Data Engineer intern* | Astana, Kazakhstan

Jan 2022 - March 2022

- Led a team of three members to conduct analysis and identify potential transaction anomalies.
- Delivered Power BI dashboards for anomaly analysis, facilitating client insights and the identification of transactional irregularities pertaining to specific individuals.

## SKILLS

Languages Python, SQL, R, Git, Bash, LaTeX, Javascript

Software BigQuery, Neo4j, Docker, Kafka, Confluent, AWS, Cloudflare, Terraform

Certifications Neo4j Certified Professional – (2022) | DSTI University – (2022)

## PROJECTS

**IT-Kvartal challenge queue system** – *Queue system* (2nd place)

January 2019

- Developed a real-time queue system enabling customers to track their order progress efficiently, freeing up their time for other activities while eliminating wait times at the bank.

**Real-time-Data-Analytics-Pipeline** - *Continuous Analytics*

Aug 2023 - September 2023

- Developed a real-time analytics solution utilizing Apache Kafka to seamlessly update an S3 bucket, ensuring continuous data processing and delivering actionable insights in real-time.

**Goodreads dataset** – *Kaggle challenge*

Jan 2023 - April 2023

- Analyzed the Goodreads dataset containing information about books, authors, and user reviews.
- Conducted exploratory data analysis to understand the distribution of book ratings, review counts, and other variables.
- Cleaned the dataset by handling missing values and removing duplicates.
- Developed and trained a predictive model to forecast book ratings based on various features.