Yuna Liu

Data Engineer/Data Scientist

Yuna Liu is an experienced data engineer/data scientist with a PhD in Economics. She has strong skills in cloud platforms like Azure, Snowflake, and Databricks, ETL tools such as Matillion, and data visualization with Power BI. Yuna focuses on building scalable, efficient data pipelines and applying advanced machine learning with Python, PySpark, and Databricks ML to solve complex business problems. At Volvo Cars, she mainly works as a data engineer and GenAI developer, supporting Sales Dashboard projects, developing Streamlit chatbots integrated with Snowflake, and creating insightful Power BI visuals. Previously, she built automated Azure data pipelines for battery cell data from scratch and led the development of ML models for credit default prediction. She completed a two-year AI and ML developer program and is structured, solution-oriented, enthusiastic, and self-driven.

Project experience

09.2023 - 09.2025

VOLVO CARS CORPORATION (VCC)

About the customer: Volvo Cars is a Swedish multinational manufacturer of vehicles.

Project: Work in Sales Dashboard project, to empower users from all over the organization with stable and accurate analytical data surrounding the ordering process, financial and physical flows.

Roles and responsibilities: Data Engineer

As a Data Engineer, I manage backend processes for the Sales Dashboard team, working with diverse datasets such as sales, stock, order-to-delivery, occurrences, options, and finance data. Develop and maintain data engineering pipelines using Azure, Snowflake, and Matillion to automate data processing and ensure high-quality, accurate reporting. Provide user support in Power BI reports and Azure Analysis Services (Cube), performing maintenance and debugging using SQL in the Snowflake database.

Roles and responsibilities: GenAI Developer

As a GenAI Developer, built pilot versions of Snowflake Cortex Analyst and Cortex Search—AI-powered self-service analytics tools enabling natural language interaction with structured and unstructured data. Cortex Analyst generates accurate answers, SQL code, and visualizations (line/bar charts), while storing queries, responses, and feedbacks for continuous improvement.

Method and technology: Data Engineering, Data Visualization, Business Intelligence, Snowflake, Snowpipe, Azure Analysis Service, Azure Data warehouse, Matillion, Services, Microsoft Power Automate, Microsoft

Power BI, SQL, Python

01.2023 - 05.2023

Volvo Cars, Gothenburg

About the customer: Cell durability team, at Battery cell and module department under Traction Battery R&D at Volvo Cars in Torslanda. The team works on optimizing individual Li-Ion Battery systems.

Project: This project explores how to use data engineering techniques and cloud tools to design and implement data pipelines for cell durability data, and in the next step, automation of a dashboard in Power BI for the Battery team at Volvo Cars. The business value is to use data engineering techniques to improve the efficiency and effectiveness of data-driven decision making.

Roles and responsibilities: Data Engineer

I was responsible for designing and implementing data pipelines that automate the integration of data from local disk to Azure Blob Storage, and later to Azure SQL Database, and transform it into a usable format. This pipeline required me to utilize my knowledge of cloud-based data warehousing, ETL/ELT development and automation tool such as Microsoft Power Automate.

- Worked on cell test and simulation data at cell durability team;

- Built data pipelines using Azure to automate deployments of cell data in dashboard using Python, SQL, Azure technologies, Power BI and automation tools such as Microsoft Power automate;

- Developed and maintained ETL processes, and data pipelines to ensure data accuracy and completeness, which improve performance and reduce processing time by more than 70%;

Roles and responsibilities: Data Analyst

I was responsible for creating dashboard which connected to a regularly automatically refreshed dataset, in conveying insights to stakeholders. I created compelling visualizations using tools like Tableau and Power BI, which allowed me to communicate complex data insights in a clear and concise manner to cross-functional teams. Once the dashboard was created, I deployed it as an app so that end-users could easily access it without the need for either a Power BI account or Azure SQL Database credentials.

- Provided regular reports to stakeholders in Power BI and Tableau;

- Collaborated with cross-functional teams to gather requirements, prioritize tasks, and deliver high-quality solutions;

- Conducted workshops with senior cross-functional team and stakeholders to present findings and gather feedback for continuous improvement

Method and technology: Microsoft Azure SQL Database, Microsoft Azure SQL Server, Microsoft Azure Blob Storage, Microsoft Azure Data Factory, Microsoft Power Automate, Microsoft Power BI, Tableau, Presentation Skills, Communication & interpersonal skills

09.2022 - 11.2022

Collector Bank, Gothenburg

About the customer: Collector Bank AB specializes in financing solutions for corporate and private individuals, with particular focus on small and

medium-sized companies. The main business is conducted through four segments namely Corporate, Real estate, Consumer and Payments.

Project: This project is first to clean, transform, integrate and visualize credit information data which stored at Azure Data Lakehouse; and later on to predict default probability using old credit information bought from UC credit, to be more precisely, I used Pyspark and Python programming in Databricks machine learning algorithms to estimate the default rate of a customer who was rejected previously and then make a next application.

The motivation is to save the cost on buying credit information from UC credit for customers applying for private loans repeatedly within short periods.

Roles and responsibilities: Data Scientist/Machine Learning Engineer

- Databricks Machine Learning with Pyspark, Spark, and Pythong programming.

- Focused on private loans' customer credit data which stored in Data Lakehouse.

- Predicted the default probability based on credit information.

- Conducted exploratory data analysis and statistical analysis;

- Developed machine learning models to solve default detection problem, using tools such as scikit-learn, TensorFlow, and pyTorch;

- Applyed Logistic Regression, Random Forest, XGBoost, MLP, SVM, K-means etc.

Roles and responsibilities: Data Analyst

- Data cleaning, data transformation and data integration of different datasets;

- Visualize raw data into insightful findings and visualizations using Python, Matplotlib, Seaborn, Plotly;

Method and technology: Microsoft Azure Databricks, PySpark, Python, SQL, Plotly, Matplotlib, Seaborn, Machine Learning (ML), XGBoost, Random forest, Regression Analysis, K-means, PyTorch, TensorFlow, Databricks

Education

2021 - 2023	**Developer in AI and Machine learning** IT University, Gothenburg, Sweden
2009 - 2016	**Ph.D. in Economics** Umeå University, Umeå
2005 - 2008	**Master in Economics and Management** Humboldt University of Berlin, Germany

Employments

09.2023 -	Capgemini
01.2023 - 05.2023	Volvo Cars, Gothenburg
09.2022 - 11.2022	Collector Bank, Gothenburg
01.2018 - 06.2021	University West, Trollhättan
09.2009 - 06.2016	Umeå University, Umeå

Certifications

SnowPro Advanced Data Engineer
Databricks Certified Machine Learning Professional
Databricks Certified Machine Learning Associate
Databricks Certified Generative AI Engineer Associate
Microsoft Certified: AI-900 Azure AI Fundamentals
Microsoft Certified: DP-900 Azure Data Fundamentals
Microsoft Certified: DP-203 Azure Data Engineer Associate
Microsoft Certified: AZ-900 Azure Fundamentals
Matillion ETL Foundations
SnowPro Core Certification

Courses

12.2021	Credit risk modeling in Python, Udemy

Honors and awards

12.2011	Stiftelsen JC Kempes Minnes Stipendiefond (SJCKMS), JC Kempes Minnes Stipendiefond (SJCKMS)
10.2011	NASDAQ OMX Nordic Foundation grant, NASDAQ OMX aktiebolag

Competencies

Microsoft Azure Databricks, PySpark, Python, SQL, XGBoost, Regression Analysis, Microsoft Azure SQL Server, Microsoft Azure Blob Storage, Microsoft Azure Data Factory, Microsoft Azure Key Vault, Microsoft Power BI, Tableau, Pandas, Microsoft Azure SQL Database, Microsoft Power Automate, Machine Learning (ML), PyTorch, Data Visualization, Business Intelligence, Snowflake, Snowpipe, Matillion, Databricks

Languages

Swedish: Fluent
English: Fluent
Chinese Mandarin: Mother tongue