

Ashwini Khedkar

Boston, MA(Open to Relocate)| (857)-544-9921 | khedkar.as@northeastern.edu | [Linkedin](#) | [Github](#)

EDUCATION

Northeastern University

Master of Science in Information Systems

Boston, MA

May 2024

TECHNICAL SKILLS

Programming Languages: Python, C/C++, SQL, C#, Java , JavaScript, R, NoSQL, CI/CD, MongoDB, Cassandra

Frameworks: Flask, Django, Apache Nifi, Spark, Streamlit, Dash, Kubernetes, Docker, DBT, Kafka

Other Tech: Github/Gitlab, GCP, AWS, Azure, Jupyter, Excel, RStudio, Airflow, Databricks, Tableau, Snowflake, PowerBI, Alteryx, Talend, ER Studio, SAS, Azure Data Factory, SSIS, Bigquery

Libraries: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Tensorflow, Plotly

PROFESSIONAL EXPERIENCE

Data Engineer

Aug. 2019 – Jul. 2022

Virtusa Corporation

Bangalore, India

- Streamlined ETL processes with Nifi for data ingestion by implementing Root Cause Analysis and **reduce data processing time by 10%** through custom python transformation scripts
- Developed **70+ Python Flask and Jinja-based** programs to improve MongoDB data mining and mapping, improving risk analysis engine's accuracy and got awarded with Excellence Work Award
- Performed **profiling, EDA, data profiling and statistical modeling & data analysis** of raw P&C internal and external data sets for the enhancement of client's analytic engine in MongoDB and Python
- Collaborated with internal and external stakeholders to gather, validate, and implement data requirements, ensuring **data quality and accuracy** through the development and monitoring of data validation tools

Software Engineering Intern (Data Science - Autonomy)

May 2023 – Aug. 2023

Rivian Automotive

Palo Alto, CA

- Engineered ETL pipelines to profile PCAP vehicle sensors(fleet) data with **Airflow, Databricks(Spark) and AWS S3**, resulting in a **30% reduction** in data processing duration and improved data consistency
- Built a Streamlit dashboard with Kubernetes deployment for autonomous vehicles KPIs, **surpassing Tableau by 11%**, emphasizing data visualization principles, & enabling data-driven adhoc analysis for cross-functional teams
- Collaborated with Data Engineering teams to define use cases and identified key performance indicators (KPIs) that optimized autonomous vehicle testing strategies and enhanced operational efficiency

Graduate Research Assistant (Machine Learning & Analytics)

Sep. 2022 – Present

Northeastern University

Boston, MA

- Develop a data pipeline using python and unity scripts to collect and synchronize eye-gaze, joystick keypress and screen recording data to analyze player's performance using **advanced predictive modelling and analytics**
- Integrated **custom YOLO model with CVAT using Docker containers** to facilitate partial annotations, significantly improving the efficiency of image data labeling by reducing manual effort Published a paper at **IEEE-COG 2023 - Measuring gaze behavior to characterize spatial awareness skill in Rocket League**

ACADEMIC PROJECTS

IMDb Data Analytics and Integration | SQL, Talend, Alteryx, PowerBI, Tableau

Sep. 2023 – Dec. 2023

- Led a collaborative team to integrate multiple data sources(CSV,XML) for a data-driven IMDb warehouse (SQL, Azure SQL) by profiling the data by Alteryx & utilizing PowerBI and Tableau for data visualization
- Executed efficient & scalable ETL workflows with Talend to **optimize data flow by 35%** by emphasizing data architecture(Kimball) & modeling, data quality,integration principles,dimensional modeling and BI Tools

Instacart Product Recommendation System | AzureCosmosDB, Synapse, PowerBI

Jan. 2023 – Apr. 2023

- Utilized a document data-model in Azure CosmosDB to perform product market analytics using **Azure Devops**
- Employed a **document-graph hybrid cloud model** to visualize customer-product intuition links, a graph database to show client, product, and availability connections, and most-ordered product details and likelihood

Consumer-Business Marketing Analytics | Snowflake, Tableau, Machine Learning

Sept. 2023 – Dec. 2023

- Developed interactive Streamlit app for forecasting & anomaly detection using Snowflake's Forecast ML models
- Integrated **Customer Lifetime Value , Customer Spend and ROI prediction** into a multipage Streamlit app, leveraging Snowflake, Snowpark, and ML models like XGBoost and RandomForest with UDF functions