Amol Khade

khadeal@mail.uc.edu ❖ San Jose, California ❖ (513) 652-2242 ❖ linkedin.com/in/khadeamol ❖ github.com/khadeamol

PROFILE SUMMARY

Data professional with 4+ years of experience and a hybrid set of skills spread across software engineering, data engineering and big data analytics. Possess experience in collaborating across business functions. Excellent communication skills and presentation skills. Proficient in modern tech stack for backend and data engineering. **Open to remote work, as well as relocation for on-site.**

WORK EXPERIENCE

xFusion Technologies
February 2022 – March 2023

Data Engineer *Sacramento, CA*

- Product: **Low-code Data Management and ETL automation Platform**
- Collaborated with sales and product teams to develop data ingestion, ETL design and development features for the platform
- Developed Python APIs to process data in PostgreSQL, S3 and Snowflake
- Achieved 80% faster processing on Data APIs by integrating Vaex Python library
- Developed custom Airflow operators to support ETL Automation

Conceptualized and developed user-facing data management features:

- **AWS S3 Bucket Explorer:** Python API based service within the platform to manage user content stored in Amazon S3
- **Data Quality Engine:** Displays data quality metrics for customer data
- **SQL Query Execution Engine**: DuckDB powered feature, allows users to run SQL queries on S3 data from within the platform and display results in the platform UI

Avant Financial
October 2019 – June 2021

Data Management Analyst - Data Engineering *Chicago, IL*

- Led the migration of Credit Card Application data from legacy systems through redesigned data models
- Led the collaboration with stakeholders to implement business use-cases and eliminate bottlenecks by developing and implementing a new Business Intelligence and ETL processes
- Saved thousands of $s in computation costs by achieving improvement in ETL run-time by optimizing data assets and use cases
- Provided support and bug resolution for ETL processing, demonstrating comprehensive knowledge in programming and data modelling
- Created and owned LookML definitions for Looker dashboards to streamline reporting
- Worked across business teams to provide periodic reporting using SQL, also automated repetitive requests

Gloo, LLC
June 2018 - July 2019

Data Analyst - Data Science and Strategic Analytics *Boulder, CO*

- Enhanced coverage for Insights platform by 3x by integrating new data sources
- Performed exploration of novel data source using R scripts (dplyr, tidyr, ggplot2)
- 3x-ed longitudinal tracking for attributes through automated R auditing scripts
- Developed **interactive Tableau dashboards to** understand sentiment analysis

EDUCATION

- **MS in Information Systems** - University of Cincinnati
- **BE Computer Science** – University of Pune

TECHNICAL SKILLS

- **Python** for Software Development, Data Engineering, Object Oriented Programming, ETL/ELT development (Boto3, Vaex, PyTorch, TensorFlow, PySpark, Pandas, NumPy, FastAPI)
- **Workflow Orchestration and deployment:** Docker, Kubernetes, Jenkins, Airflow
- **SQL (PostgreSQL, NoSQL, SQL Server, MySQL), R** (Tidyverse – tidyr, dplyr, ggplot2) for data analysis
- **Big data tools: Spark, Hadoop, Presto, Hive**
- **Machine Learning:** Scikit-learn, Keras, MLFlow, CNNs
- **BI and Data Visualization:** Dremio, Alation, Looker, Tableau, Matplotlib, Seaborn
- **Experienced with AWS and GCP products**

PROJECTS

Formula 1 Racing dashboard – Google Cloud, Streamlit, Python, Matplotlib, Seaborn (Ongoing Project)

- Built entirely on Google Cloud Instances and deployed using Streamlit. Available for public consumption at f1analysisdash.streamlit.app
- Created in Python using FastAPI, Flask and Pandas, displays telemetry and timing data for F1 Races
- Allows user to understand impact of weather, tire life/compound, strategy, DRS on lap times, and enable them to make better decisions to improve performance
- Features head-to-head driver comparison for timing, driving styles, understand scope for improvement, pitfalls, time and distance delta down to microsecond level precision
- Constantly adding new features for deeper analysis

Car Efficiency Prediction: End-to-end Machine Learning Pipeline

- Engineered an efficient MPG prediction model for cars using scikit-learn, optimizing Linear Regression, Decision Trees, and Random Forests models
  - Implemented hyperparameter tuning and K-fold Cross Validation, achieving superior model performance, evaluated using Root Mean Squared Error
  - Implemented scikit-learn for Stratified Shuffle split, Train Test split, One Hot encoding, Simple Imputer, Standard Scaler, Column Transformer