

Kartik Ullal

+1 (857) 334-3460 | kartikullal99@gmail.com | [linkedin.com/in/kartikullal](https://www.linkedin.com/in/kartikullal) | github.com/Kartikullal

Experience

Data Science Engineer

Feb 2024 – Present

KeelWorks Foundation

Boston, MA

- Cleaned and preprocessed contact information from 4 sources using Python and SQL, and stored it in AWS RDS, facilitating the sales team to send targeted marketing and fundraising emails, resulting in a 30% higher email deliverability
- Conducted A/B tests on email campaigns for donor acquisition, optimizing subject lines, content, and send times, resulting in optimized emails achieving 18% higher open rates and a 22% increase in donations compared to non-optimized campaigns

Data Engineer Co-op

July 2022 – Dec 2022

Wayfair LLC

Boston, MA

- Boosted service level objective by 6% and streamlined bug fixes by developing a procedure using BigQuery to de-duplicate tables during the ETL runs, and send alerts using a Python script
- Implemented caching and aggregate tables in LookML to improve the latency of our Looker dashboards across the team, achieving a 10% increase in performance
- Developed a forecasting model to detect anomalies in HomeServices request rates using Facebook Prophet, resulting in a 50% decrease in resolution time
- Designed and maintained ETL pipelines for Wayfair credit card data using Python, Airflow, BigQuery, and dbt, enabling stakeholders to make informed decisions on credit card acquisition and spending patterns

Data Science Intern

May 2020 – Dec 2020

Prayaas Corps

Jaipur, India

- Engineered a scheduled pipeline to automate extraction and preprocessing of student data using Excel, MS SQL server and SSIS, enabling analysis of student performance over time
- Created a dashboard to analyze and find at-risk students using PowerBI (DAX, M Query), identifying trends between student grades and engagement
- Developed an XGBoost model with an RMSE of 4.5 to predict student grades using Python, allowing teachers to provide timely support to at-risk students, improving academic outcomes

Data Science Intern

March 2019 – Dec 2019

Navlakhi Education

Mumbai, India

- Designed a database by integrating questions and student answers to streamline the JEE testing software using PostgreSQL, providing instant feedback to students and improving their performance
- Assisted with the implementation of an ETL pipeline of Mumbai Fashion Academy's student and course data for analysis, leading to data-driven decisions to optimize course offerings and enhance the overall student experience
- Deployed a course recommendation system for Mumbai Fashion Academy's course catalog using collaborative filtering, increasing course enrollment by 20%

Education

Northeastern University

Sept 2021 – Dec 2023

Master's of Science, Data Science

Boston, MA

University of Mumbai

Aug 2017 – May 2021

Bachelor's of Science, Computer Science

Mumbai, India

Projects

NYC TaxiFlow Analytics

- Designed an automated pipeline using Python to extract data from the NYC Gov website, loading it into Google Cloud Storage and BigQuery. Scheduled and monitored pipeline execution using Prefect
- Transformed raw data into a Star Schema model using DBT and BigQuery, ensuring efficient analysis. Implemented data validation and documentation for accuracy and maintainability
- Crafted an engaging Looker Studio Dashboard, translating raw New York taxi data into actionable insights for stakeholders

Adventure Works Analytics Pipeline

- Engineered an end-to-end ETL process, extracting data from Adventure Works to SQL Server using Python and Airflow, then transferring the data to Azure Data Lake Gen2 via Azure Data Factory
- Employed Databricks and PySpark for robust data cleaning and transformation, and implemented a Star Schema model using T-SQL in Azure Synapse for optimized analysis
- Designed a comprehensive Power BI dashboard featuring KPIs, revenue trends, product details, and customer details, empowering stakeholders with actionable insights

Real-time Data Streaming Pipeline

- Ingested real-time user data from randomuser.me API, utilizing Apache Kafka and Zookeeper for streaming into Kafka queue, managing data flow via Control Center and Schema Registry
- Orchestrated data processing pipeline with Apache Airflow, processing data from Kafka queue using Apache Spark, storing preprocessed data in Cassandra, containerizing entire pipeline using Docker

Technical Skills

Proficient: Python, SQL(MySQL, PostgreSQL), BigQuery, Data Analysis (PowerBI, Looker), DBT, Machine Learning, GCP

Familiar: Excel, Scheduling tools(Airflow, Prefect), AWS, PySpark, Docker, R programming, Deep Learning, Snowflake

Basic: Azure Cloud, Kubernetes