MATT KAYE

mrkaye97.github.io mrkaye97 mrkaye97@gmail.com (646) 853-5997

EXPERIENCE

Data Scientist

CollegeVine

Sept 2020 – Present

- First member of the CollegeVine data team. Responsible for infrastructure, processes, and execution of technical work including maintaining the code base, introducing new tooling to help the team ship quickly and painlessly, mentoring junior teammates, consulting on a wide variety of projects, and getting data products over the finish line and into production.
- Devised and implemented an acceptance probability model used by hundreds of thousands of students and families every year to understand their chances at their favorite schools. Chancing is a primary driver of traffic, new accounts, and retention on CollegeVine and powers other products like CollegeVine Applications, which automatically matches students with schools where they're a good fit and submits applications for them.
- Created a recommender system to suggest new colleges to students based on their schools list and preferences. School recommendations also powers Collegevine Applications by generating automated matches between thousands of students and colleges they would be interested in.
- Built a sequential testing toolkit to enable our team to call A/B tests more quickly without sacrificing statistical rigor. Wrote extensive documentation and tutorials to help teammates onboard to the new tooling, and gave talks introducing and explaining complex A/B testing topics in an approachable way.
- Built out a data warehouse for analytics purposes, enabling our team to iterate more quickly, repeat less SQL, be less prone to bugs and inconsistencies in metric definitions, and easily perform BI tasks that were previously infeasible.
- Introduced new tooling (dbt, Airflow, MLFlow, AWS stack) and owned its adoption, maintenance (when applicable), and integration with the code base.
- Owned the deployment and monitoring of machine learning models in production, generally as REST APIs packaged up as Dockerized microservices and deployed on Heroku.

Baseball Operations Fellow

Baltimore Orioles

Mar 2020 – Sept 2020

- Created a fully Bayesian, simulation-based projection system for MLB player performance over a six year time horizon
- Modeled free agent salaries with a gamma hurdle regression framework
- Devised a Markov Chain Monte Carlo approach to determining optimal shifts against opposing hitters
- Worked on a variety of day-to-day data science tasks related to game strategy and player evaluation

Open-Source Contributor

mlflow, slackr, lightMLFlow, fitbitr

Oct 2020 – Present

- Contributor to MLFlow, an open-source platform for managing the machine learning lifecycle.
- Current author and maintainer of $slackr$, an R package for connecting R to Slack with 250k+ downloads.
- Author and creator of $lightMLFlow$, a lightweight, user-friendly R wrapper for the MLFlow REST API.

EDUCATION

Bachelor of Arts
Economics, Mathematics

Carleton College

Sept 2016 – Nov 2019

Choate Rosemary Hall

Sept 2013 – June 2016

SKILLS

Programming:

Bash Python R SQL

Frameworks, Software, and Tools:

AWS Tooling (Redshift, S3, Batch, etc.) Airflow CI/CD (Circle, GHA) Docker Git Heroku MLFlow {{plumber}} {{shiny}}

HOBBIES

Distance running
Learning Haskell
Low and slow cooking
Nature & architecture photography
Reading fantasy novels
Skiing chop and powder
Solo traveling + hostel hopping