MATT KAYE

mrkaye97.github.io | mrkaye97 | mrkaye97@gmail.com | (646) 853-5997 | kayem20

EXPERIENCE

Data Scientist

CollegeVine

Sept 2020 – Present | Cambridge, MA

- First member of the CollegeVine data science team. Responsible for establishing data science workflow best practices, building out a modeling pipeline, and providing easy-to-use data science tools to other members of the team
- Transitioned our chancing algorithm (used to tell hundreds of thousands of students their chances of admission at their favorite schools) away from a rules- and heuristics-based model to an ML-forward approach, using tree boosting, rigorous validation, and clever feature engineering
- Helped hundreds of thousands of students find new schools they were likely to be interested in by creating a recommender system to suggest schools based on our knowledge of them and their interests
- Advised on survey work at CollegeVine and analyzed results for use in published reports on college admissions
- Serving as a subject matter expert for everyday statistics- and data science-related questions

Open-Source R Developer

slackr

Oct 2020 – Present

- Current author and maintainer of *slackr*, an R package for connecting R to Slack with 200k downloads
- Working on package improvements with much help from the R community, including improving documentation to help with setup and debugging, improving error messaging, setting up pre-packaged Slack apps to make setup easier, improving the way the package interfaces with the Slack API, and setting up CI (GH Actions) and unit testing to test the package when changes are made
- Helping users work through bugs and questions submitted in issues, reviewing PRs, and doing other everyday package maintenance tasks

Baseball Operations Fellow

Baltimore Orioles

Mar 2020 – Sept 2020 | Baltimore, MD

- 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
- Implemented a robust-to-multimodality version of the Metropolis-Hastings algorithm to determine the optimal way to position our defense against a specific hitter
- Worked on a variety of day-to-day data science tasks related to game strategy and player evaluation

SKILLS

Languages:

Haskell | Java | Python | R | SQL

Frameworks, Software, and Tools:

AWS | brms | Docker | Excel | Git | Heroku | Numpy | Pandas | R Shiny | Scikit-Learn | Tidyverse + Tidymodels

Data Science:

Bayesian Modeling and Inference | Data Wrangling | Data Visualization | Machine Learning | Statistical Inference | Time Series Modeling

LANGUAGES

English

Spanish

Swedish

EDUCATION

B.A. Economics and Mathematics

Carleton College

Sept 2016 – Nov 2019

Economics Thesis: *The Effect of Transit on Life Satisfaction: A Multilevel Modeling Exploration of Urban Happiness*

Mathematics Capstone: *Dynamic Linear Models and the Kalman Filter*

Choate Rosemary Hall

Sept 2013 – June 2016

INTERESTS

Baseball | Browsing r/AskReddit | Cooking | Running | Minimalism | Skiing | Data Visualization | Fantasy Novels | Solo Travel | Nature & Architecture Photography