

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 modeling pipelines, helping teammates do rigorous, thorough data analysis, and explaining data scientific methods to other members of the team to democratize data science work at CollegeVine
- Transitioned our chancing algorithm (used by hundreds of thousands of students to get their chances at over 1,500 colleges and universities) away from a rules- and heuristics-based model to an ML-forward approach using feature engineering that incorporates expert knowledge of college admissions and rigorous testing and validation of the model
- Implemented a dashboard with sequential testing tools to help CollegeVine product managers run tests more quickly without sacrificing statistical rigor
- Serving as a subject matter expert for everyday statistics- and data science-related questions

Open-Source R Developer

slackr & fitbitr

Oct 2020 – Present

- Current author and maintainer of *slackr*, an R package for connecting R to Slack with 210k+ downloads
- Author and creator of *fitbitr*, an R package that streamlines pulling Fitbit user data via the Fitbit API
- Responsible for all aspects of package development and maintenance, including general maintenance tasks, implementing new methods, improving error handling, writing unit tests and setting up checks in CI, writing descriptive documentation, and more
- 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