

# Justin Chen

[chen.justi@northeastern.edu](mailto:chen.justi@northeastern.edu) • 650-388-0985 • <https://jyhchen.github.io/>

Looking for full-time positions starting July 2023

## EDUCATION

### Northeastern University, Khoury College of Computer Sciences

*Candidate for Bachelor of Science in Data Science and Business Administration*

**Awards and Honors:** Dean's Scholarship, Dean's List, University Honors Program

September 2019 - Present

Expected: June 2023

GPA: 3.51/4.0

**Coursework In:** Statistics; Object Oriented Design; Database Design and Information Storage; Machine Learning and Data Mining; Data Visualization; Sports Media, Marketing, and Communications; Financial Data Analytics

**Activities:** Baseball Student Manager, WRBB Sports (student radio), Intramural Sports

## SKILLS AND INTERESTS

**Computer Languages:** Java, Python, Racket, SQL, HTML, CSS, Javascript, R

**Frameworks and Packages:** Django, React.js, D3.js, SQL Server, AWS

**Software:** Git, IntelliJ, Jupyter, Snowflake, Adobe Analytics, Visual Studio Code, Tableau

**Interests:** Movie Soundtracks, Documentaries, Hats, Public Transportation, Biking

**Other:** Social Media, Sports Writing and Broadcasting, Chinese (native speaker)

## WORK EXPERIENCE

### Student Manager and Analyst

*Northeastern University Baseball*

Boston, MA

January 2020- Present

- Modeled expected-run values with Rapsodo and Statcast metrics for pitch sequencing and arsenal design
- Charted pitch-by-pitch game data with Excel and visualized trends and statistics with Python and Powerpoint
- Created scouting reports with Synergy and assisted with deciding game strategy
- Analyzed player video for pitch tipping and mechanical inconsistencies

### Web Analytics and A/B Testing Co-op

*Harvard Business Review*

Brighton, MA

July 2022- December 2022

- Created interactive reports with Adobe Analytics to summarize A/B test findings and analyze user engagement
- Designed experiments to analyze article recommendations, newsletter engagement, and SEO article performance
- Developed model to predict website visitors' conversion to paying subscriber with 80% accuracy
- Automated dashboard backup and creation with R and Python, utilizing Adobe APIs

### Baseball Systems Development Co-op

*Baltimore Orioles*

Baltimore, MD

July 2021- January 2022

- Rebuilt data pipelines for player stat views on the team's internal site with Django
- Developed amateur player evaluation forms and tables for scouts with React
- Made quality checks and internal ID mapping for imported third-party data and created Slack notification bots
- Automated web scrapers with AWS lambda functions

## PROJECTS

### NCAA Basketball Prospect Grader (Python/Jupyter Notebook)

- Built models to rank NBA Draft prospects and project their NBA win shares based on their collegiate careers
- Collected dataset of 15,000 NCAA players with over 45 features and their NBA contributions with data scraping
- Used one-hot encoding and recursive feature elimination find the optimal feature subset
- Trained regression models with cross-validation, including Random Forest, AdaBoost, and Multilayer Perceptron
- Ranked players for the 2022 Draft, with results similar to ESPN's projections

### MLB Repository (MySQL/Python)

- Created a relational database to store biographies and statistics for players and teams
- Implemented functions and triggers to update tables with additions or deletions of records
- Used Python to create a command-line application to access player and team information and statistics

### Shot Grader: an Alternative to xG (WHKYHAC 2022) (Python/Jupyter Notebook)

- Graded PHL and Olympic shots based on shot destination results, different than the two-true outcomes of expected goals
- Explored importance of skater traffic and shot type when analyzing the quality of a shot
- Discussed data constraints in the realm of women's hockey