

# Johannes Harkins

206.326.0961

Beacon, NY

[johannes.harkins@gmail.com](mailto:johannes.harkins@gmail.com)

[jmharkins.github.io](https://github.com/jmharkins)

[My Github Profile](#)

## Technical Skills

- I'm a Data Scientist primarily using Python daily for data analysis. I am proficient with many popular data science packages such as Pandas, Numpy, Scikit-Learn and PyMC. I've also performed many of the same data analysis concepts in R.
- Experience managing Data Science codebase in Git, specifically through GitHub and the Atlassian suite of tools.
- Familiarity interacting with data in a number of data formats including: XML, JSON, CSV, MongoDB, SQL-like databases, the Hadoop DFS, web scraping, various forms of APIs
- I have hands-on experience through various projects with Node, JavaScript (in particular d3.js), Flask, Airflow, Luigi, TensorFlow, Keras, Apache Spark, Tableau, and Scala

## Experience

- **Plated** New York, NY  
*Data Scientist* *March 2019 - February 2020*
  - Worked in cross-functional teams organized around business goals to provide model-building and data analysis expertise.
  - My work on these teams included modeling customer lifetime value (LTV) at time of sign-up, building and managing customer propensity models, modeling similarity between pieces of content using natural language processing techniques, and working on a recommendation system for customers.
  - Aided in set-up and analysis of A/B tests
  - Participated and presented in the knowledge-sharing meeting series run by the data guild
  - **Primary Tools:** PyData Stack (same as below), Amazon Redshift, Google Analytics, Amplitude
- **Opta Sports (Perform Group)** New York, NY  
*Data Scientist* *May 2015 - March 2019*
  - Responsible for developing innovative metrics from the stages of theory and data exploration to model building to production. I've participated in or owned models that are used to assess player and team performance, aid transfer market strategy, and predict match outcomes, among other applications
  - Improved the structure of the Data Science team's code according to coding best practices by introducing version control and standardizing workflow
  - Judged submissions for the 2016, 2017, and 2018 editions of the OptaPro Forum, and delivered a talk at the first North American regional OptaPro Forum in January 2018.
  - Guest lectured for 2016 and 2017 offerings of the *Analytics of Global Sport* course at Columbia University's Sports Management program and judged the end-of-year case competition by participating students
  - **Primary Tools:** PyData Stack (pandas, NumPy, IPython, Scikit-Learn, PyMC3), MySQL, OracleSQL, R
- **First Insight Corporation** Hillsboro, OR  
*Jr. Software Engineer - Data Analytics* *June 2014 - May 2015*
  - Responsible for the development of analytics reporting for a software for Optometry and Ophthalmology practices. Developed a dashboarding UI using AngularJS and D3 for at-a-glance management statistics for client practices
  - **Primary Tools:** FileMaker Pro, SQL, AngularJS, D3.js
- **Google** Remote  
*Project Analysis Contributor* *June 2014 (Temporary)*
  - Contributed to analysis and predictions for Google's 2014 World Cup prediction models summarized in [Google I/O presentation](#) and [related blog posts](#)
  - **Primary Tools:** R, BigQuery

## Education

- **Reed College** Portland, OR  
*Bachelor of Arts, Mathematics & Economics* *May 2014*
  - Thesis: *Testing the Efficiency of Sports Betting Markets*