



Vaughn Johnson

Data Scientist

Seattle, WA
v@ughn.io
vaughnjohnson.me
github.com/vaughn-johnson

Skills

Data Science

- **Database** Amazon Redshift, MySQL, Postgres
- **Python** Pandas, scikit-learn, Jupyter, Anaconda, Plotly
- **R** Tidyverse, ggplot
- **Statistics Fundamentals**
Statistical machine learning, Parametric & Nonparametric modeling, Experimental design & methods
- Some Tableau & Spark

Full Stack Development

- **Backend** Ruby on Rails, RSpec, Node.js, Express
- **Frontend** React, React Native, jest, detox, redux
- **Google Cloud** Compute Engine, Cloud Storage
- **AWS** EC2, S3, Lambda, Cloudwatch & RDS
- Experience with **Docker**

Eager to deepen knowledge about

- MCMC methods & Probabilistic Programming
- Time Series data & Stochastic systems
- Transformer Architectures

Languages

- Native **English** Speaker
- **日本語**を三年間勉強しました

About Me

I am a **curious** person. I value **legibility**, **evidence**, and **getting things done**. My passion is making sense from data.

EXPERIENCE

2021 - Now **Data Scientist** | **GoDaddy**

- Communicating daily with business partners across the entire firm on statistics and experimental best practice
- Implemented False Discovery Rate adjustment, reducing the false positive rate of every experiment conducted on our platform
- Created a new unit testing framework in multiple repositories to provide coverage to critical and previously untested legacy code paths
- Set technical direction and research culture of the statistics team by introducing plaintext Jupyter notebooks used with version control and virtual environments.
- Direct contributions to our entire ETL, user interface, and analysis pipeline

2019 - 2020 **Full Stack Software Engineer** | **Tavour (An awesome craft beer retailer)**

- Close collaboration with product manager and finance team, including interpreting A/B test results, providing automated reports, and piping data into a data warehouse
- Developed a delivery window estimator. Enabled Tavour to communicate the variability of shipping durations across dozens of delivery services with one interface
- Refactored hundreds of RSpec unit tests in our legacy codebase and set testing precedents used by the rest of the team

2017 & 2018 **Data Warehousing and Analytics Summer Internships** | **Tavour**

- Developed the Redshift data warehouse and ETL pipeline the company currently uses to perform business intelligence
- Enabled Tavour to make major capital investments by developing metrics to assess the revenue impact of specific breweries
- Made our complex schema accessible to the entire organization by creating and testing a new simplified relational schema

EDUCATION

2015 - 2019 **Statistics & Data Science** | **University of Washington**

- Graduated in 2019 with a B.A. in **Statistics** and **Data Science** with minors in **Linguistics** and **Physics**
- Won "Best in Show" at the 2019 UW DataFest
- Presented at the 2019 UW Undergraduate Research Symposium on my research on police violence and crowd sourced data

PROJECTS

2020 **Talkspace Data Analysis** | **talkspace-notebooks**

- I use the online therapy service Talkspace and I wanted insights into my own data. I wrote an npm package (**talkspace-scraper**) so anyone can download his or her entire message history. I extracted features from my own data and made it available as a public API on Google Cloud. I also did an analysis and documented the process both in R and Python Jupyter notebooks and in a React client on **my website**, also hosted on Google Cloud