

Adam T Houlette

athoul01@gmail.com | ahoulette.com | Mastodon

I'm a data scientist with 8 years of experience solving problems and answering questions with messy data. Whether I'm getting my hands dirty cleaning datasets, building predictive models and simulations, or deploying APIs and building clear, intuitive data visualizations, I love extracting meaning and value from the chaos and then communicating the results to do my small part to elucidate some of the world's many mysteries.

Projects

- [ahoulette.com](#): 8 years of data analysis and communication including:
- **Netflix Viewing Analysis**: Developed novel cumulative viewing visualization to examine Netflix viewing distribution, finding just 100 shows account for 25% of all viewing. Analysis of period-over-period viewing changes including new season lift, viewer decay, and new release-over-new release growth metrics with bootstrapped confidence intervals.
- **Seattle Public Library Checkouts**: Wrangled 50M+ rows of library checkout data using DuckDB to facilitate rapid exploration including analysis of checkouts by medium, subject, and author. Mined the physical to digital shift for insights, discovering the wrench said shift throws into standardization of book subject classification across mediums. Developed end-to-end modeling pipeline to predict monthly checkouts of titles using XGBoost and tidymodels and deployed model to API using vetiver.
- **Investment Income Simulation**: Developed Monte Carlo simulation pipeline to investigate the role of starting capital in final investment income. Implemented modular simulation logic to generate over 2.2 million simulated investors under a variety of starting conditions and assumptions. Built 'racing' plot visualizations to highlight the 'winners' and interrogated the results, finding that the impact of starting capital is of greater importance than skill.
- **Kaggle Kernels Award**: Won Kaggle Kernel award for processing, visualizing and modeling interview attendance using random forests.

Technical Skills

- **Machine Learning Methods**: Developed classification, regression and recommendation system models to predict monthly book checkouts, classify crime types, predict Yelp restaurant ratings, and recommend favorable beers using a variety of methods including XGBoost, linear methods and recurrent neural networks(RNN). Generated monthly thefts forecast using ARIMA and exponential smoothing methodology.
- **Statistics/Analytics**: Developed Monte Carlo simulations to investigate investment income. Rigorously applied hypothesis testing, t-test/Poisson tests, bootstrapping, Gini coefficients/Lorenz Curves, and other statistical methodology to determine the significance of increases in Louisville crime and develop confidence intervals around period-over-period Netflix show growth. Quantified changes in presidential news coverage and Yelp reviews data using NLP and text mining analytics including TF-IDF and sentiment analysis.
- **Data Visualization**: Created novel, publication ready visualizations detailing the cumulative share of views or [checkouts](#), highlighting [anomalous events](#), and [communicating topline results](#) as well as quick, iterative visualizations to rapidly explore new datasets and generate insight from raw data. Well-versed in the full suite of R visualization tools with a keen eye for building detailed, well-formatted, intuitive visualizations and tables.
- **Data Wrangling/Munging/Mining**: Experience manipulating and exploring 100M+ row datasets using DuckDB and pins. Fluency gathering, verifying and processing messy data whether that's geocoding millions of rows of Louisville crime data to form clean, tabular data or developing efficient, readable scripts to process and feature engineer raw Seattle library or Netflix viewing data for ingestion into the modeling pipeline.
- **Software and Programming Languages**: *R(primary language)*(tidyverse, tidymodels, ggplot2, vetiver, DuckDB); *Python*(pandas, numpy, scikit-learn); *SQL*; *Git*; *Excel*

Education

Boston University School of Law - *J.D. Program*

University of Louisville - *B.S. Economics, 3.98 GPA, Summa Cum Laude*

Vanderbilt University - *Blair School of Music, Vocal Performance*

Certifications:

- **Machine Learning Specialization**: 4-course specialization by the University of Washington on Coursera

Experience

Independent Data Scientist & Technical Blogger: *ahoulette.com, 2017-Present*

- Self-directed analytics research spanning the full data landscape, including end-to-end predictive modeling pipelines with XGBoost and vetiver, Monte Carlo simulations of millions of investors, and exploratory data analysis on 100M+ row datasets using DuckDB.

Gig Work

- Various gig work including freelance data work, delivery, and logistics

Substitute Teacher: *Jefferson County Public Schools*

- Assisted with high school lesson plans ranging across a variety of topics, but usually pertaining to math or science.