Brandon **Dey**

Senior Data Scientist

7428 N. Charleston Ave APT 402, Portland, OR 97203

(503) 866 - 8354 | brandonwdey@gmail.com | brandondey.github.io | brandondey | brandondey

Summary

I'm a creative **full stack data scientist** focusing on applications of state of the art Natural Language Processing. I thrive on complexity and know how to communicate to people at all levels of seniority and technical acumen. With expertise in end-to-end data analytics processes, from ETL to executive level presentations, my professional background has sharpened my **chops in R, Python, SQL, AWS, Git, and Hadoop**, which I've leveraged to **design and evaluate experiments**, **test client hypotheses, mine actionable, *timely* customer insights, and build data-driven, production-grade** systems that monetize data.

I compute my math and science on System76's **Thelio Major** for scientific computing. **RAM:** 64GB; **Hard drive:** 1TB; **Processor:** 24 threads @ 4.3 GHz; **GPU:** Radeon RX 590 with 2304 stream processors.

Education

University of Wisconsin - Milwaukee *Milwaukee, WI*

B.A. in Economics, GPA 3.73 *2010 - 2015*

THESIS: "How Has the Black-White Health Gap Changed Over the Last 40 Years?"

Continuing Education

Data Camp *2018*

- Statistical Modeling I, II
- ggplot2 I, II, III
- Machine Learning in R
- Reporting with R Markdown
- Correlation and Regression
- + a half-dozen more

Amazon Web Services *Forthcoming*

- Big Data Certificate
- Cloud Practitioner Certificate

Programming Skills

Languages

- SQL (Highly adept, 6 years)
- R (Highly adept, 6 years) {tidyverse, stats, MASS, custom functions}
- SPSS (Proficient, 2 years)
- LaTeX (Proficient, 3 years)
- STATA (Proficient, 1 year)
- Python (Proficient, 3 year) {Beautifulsoup4}

Technologies

- Microsoft SQL Server (Highly adept, 6 years) {SSMS, SSRS, T-SQL}
- MySQL (Highly adept, 5 years) {Workbench}
- Microsoft Visio & LucidCharts (Proficient 5 + years)
- R Markdown (Experienced, 4+ years)
- Git (Lab & Hub) (Adept, 3+ year)
- OpenStreetMap (Proficient, 1+ year)
- Hadoop (Proficient, 2+ years) {Hive, Impala}
- Tableau (Working knowledge, < 1 year)
- R Shiny (Working knowledge, < 1 year)
- AWS (Proficient) {Redshift}

Work Experience

The Kroger Company (84.51)

Portland, OR

Senior Data Scientist, Operations Research

March 2019 - PRESENT

- Team lead for measuring the effect of our proprietary stacked ensemble that predicts how likely a store clerk will churn in the short term. I architected and engineered an automated evaluation tool that operates within the quasi experimental context of our interventions. Under the hood, it's a network of Bayesian structural time series models, not unlike Google's CausalImpact package, but with greater transparency and improved flexibility for more precise selection of the synthetic control group. In production.
- The aforementioned HR model includes legally protected classes, so statistical equity is paramount. I built an automata that detects and addresses any bias in our machine by evaluating true/false positive rates across class groups and applying a novel post hoc weighting scheme when needed. In production.
- Responsible for engineering 15 features in our HR model with 100+ others, for three of which I used anomaly detection (anomalize) to root out more signal. 4-5 of my features land in the top 10 most predictive features every month. In production.
- Designed and engineered a suite of unit tests for a critical ETL pipeline that robustly checks various consistencies between databases and emails the team an HTML summary. In production.
- Lead team in software engineering best practices by functioning as gatekeeper who reviews all pull requests before committing to master/production. Spearhead meetings to keep team accountable to strong software practices.
- Improved forecasts of Pickup demand by adding engineered features for local weather conditions, resulting in more accurate estimates of how much labor a store ought to schedule, which enabled our optimization engine to create even better labor schedules for Pickup.

Fisher Investments

Portland, OR

Team Lead Data Science, Global Marketing

Oct. 2018 - March 2019

- Increased key sales quality metric by 9 percentage points while maintaining sales volume with two random forest models that discovered new predictive features from online retirement calculator which legacy solution missed. Solution gave marketing management control over cost and quality of sales leads for first time.
- Increased accuracy of propensity mail model by one percentage point for "free" by using active learning when downsampling non-responders who outnumbered responders 20 to 1.
- Optimized paramaters of legacy elastic net lead scoring model for remarketing mail campaigns, resulting in more precise targeting.

Direct Supply, Inc.

Milwaukee, WI

Data Scientist, Data Science

April 2018 - Sept 2018

- Served as technical contributor and editor of an internal data science journal focused on (i) discovering actionable customer insights, (ii) recommending what to do about them, (iii) unpacking how to monetize the data generally and (iv) demonstrating the power of machine learning, written for an executive audience.
- Identified an opportunity to create a new line of business after using a Random Forest decision tree to predict if a customer turnover would generate a net profit or loss.
- Designed, built, and evaluated a system that generates recurring monthly call and email campaigns by targeting customers most likely to buy in timeframe using logistic regression. Achieved 13 percentage point boost in clicks and 5 percentage point increase in purchase rate over a 20% holdout group in first 12 months in production (R::stats::glm).
- Built a Latent Dirichlet Allocation Topic Model that predicted what free-form text-based work orders were about, allowing engineering to trigger real time advertising on topical key words entered by mobile app users (R::topicmodels).
- Estimated causal effect of a new product line on existing margin using bayesian structural timeseries models (in absence of randomized experiment), enabling product managers to proceed to rollout more product (R::CausalImpact).
- Predicted date range a skilled nursing facility was to be inspected with 85% accuracy using Stratified Cox survival model to monetize publicly available data from the Centers for Medicare & Medicaid Services (CMS) (R::survival).
- Automated the analysis of a biannual Net Promoter Score survey in R (was living in SPSS), reducing 280 hours of work from 1.5 FTE's to 14 hours of work by 1 FTE using custom-developed library of R functions.
- Helped eCommerce executive leadership meet their KPI of retaining, growing, and finding customers by visualizing a time series of customer lifecycle (new, full, lapsed)(R::TraMineR).
- Estimated opportunity cost of discontinuing support of Internet Explorer 8 to help IT assess impact.

Direct Supply, Inc.

Milwaukee, WI

Analytics Advisor, Data Science

March 2017 - April 2018

- Partnered with key internal stakeholders for the development of scalable machine learning and predictive modeling solutions. Provided analytical consultation to partners across all departments, ranging from executive level to team members, using R and SQL.
- Created dashboards, reports, and analyses that summarized complex findings to help internal partners make smarter decisions.
- Developed high level data driven insights that confirm or refute partner's hypothesis. Responded to complex inquiries with actionable insights and advice.
- Defined analytical and reporting requirements with partners.

Direct Supply, Inc.

Milwaukee, WI

Business Intelligence Analyst, Data Science

Dec. 2015 - March 2017

- Authored an Abandonded Shopping Cart Report that informed sellers when customers in territory placed product in their shopping cart but hadn't purchased yet, which led to a 4 percentage point lift in sales.
- Authored complex SQL queries/views/functions and reports to meet key KPIs, which varied across teams, departments, and business units.
- Acted as an advisor to stakeholders by having knowledge of data structures and system capabilities.

Direct Supply, Inc.

Milwaukee, WI

Reporting Analyst Intern, Data Science

Oct. 2014 - Dec. 2015

- Fulfilled reporting requests by building and updating reports and dashboards using SQL, SSRS, and Microsoft Visual Studio.

University of Wisconsin - Milwaukee *Milwaukee, WI*

Economics Research Assistant, Undergraduate Research Fellowship *Sept. 2014 - Dec. 2014*

- Performed tasks to advance the research project of addressing whether a 5 cent plastic bag tax in Washington D.C. led to a decrease in food consumption, employing a difference-in-difference model, under the guidance of Dr. Laura Grant.

Literacy Services of Wisconsin *Milwaukee, WI*

Math Basics Coordinator, Intern *May 2014 - Aug. 2014*

- Developed a new curriculum of basic math courses for adult students and taught all math lessons 1:1.

Side Projects

Natural Language Processing

Dicto – The smart dictionary *Portland, OR*

Forthcoming

- Imagine a logophilic bot that fetches the precise word for the precise context. Search for words based on intuitive parameters like "words that sort of mean [adjective] but are less like [synonym]", word rareness, modernness, domain relevance (legalese, medical terminology, sports vernacular, male(female)ness, etc.), the word(s)' "[adjective]-iness", tensile strength, etc. It also displays each word's association to contentious topics like race, slavery, gender, and other extremist topics so the user is informed of – and can sidestep – the word's charge in the American zeitgeist.
- It's **DataMuse** meets **David Foster Wallace** meets **Aesop Rock**.
- Powered by: **Latent Semantic Analysis**, semantic similarity measures computed by a superior corpus & knowledge-based method (vs. vectorial models common in information retrieval), and **Convolution Neural Networks.**
- Fed data from: **Google nGram**, **Merriam-Webster's Collegiate® Thesaurus**, and **DataMuse.**

Text Mining 3 Years of My Family's Text Thread *Milwaukee, WI*

March 2015

- Created personalized visualizations for 12 family members who exchanged texts with one another over a three year period.
- Insights included who was most likely to initiate conversation, who was most exclamatory(!!!), most inquisitive(???), most emoji-tional, night owls, early birds, topical trends over time per person, simple word clouds, and predicted response time given who the initiating interlocuter.

Writing

Professional & Published

Machine Learning, Applied Statistics, Data Science *Cambridge, MA*

AI & Data Science Journalist, Open Data Science Conference (ODSC) *MAY 2018 - PRESENT*

- 'Active Learning: Your Model's New Personal Trainer'
- 'ODSC West 2018 Review'
- **'An Introduction to Sentence-Level Sentiment Analysis with sentimentr'**
- 'Where's Humanity's Algorithm to Solve Poverty?'
- 'A Beer Lover's BFF? An Introduction to Geospatial Interpolation via Inverse Distance Weighting'
- 'Comparing Five Different Smooths – Which One Rules Them All?'
- 'A Short Summary of Smoothing Algorithms'

Professional & Unpublished

'Consider The Fly'

Honors & Awards

2012 **Recipient,** Scholarly Award, Delta Sigma Pi. Highest score on pledge exam. *Milwaukee, WI*

2010 **Recipient,** Most Likely to Succeed Class of 2010 Coleman High School *Coleman, WI*

2010 **Recipient,** English Award Coleman High School *Coleman, WI*