BURTON DEWILDE

burtondewilde@gmail.com • linkedin.com/in/burtondewilde • 646.241.2381 • Chicago, IL

End-to-end data scientist with 7+ years' experience in web analytics, data product development, media research, and nonprofit consulting. Proven ability to scope and complete projects; solve meaningful problems with data; write good, clean code; and work productively while remote.

SKILLS

- **Data Engineering:** general programming; cloud computing; streaming data pipelines; database design and management; Python, git, Amazon Web Services, Apache Kafka, PostgreSQL, MongoDB, Elasticsearch
- **Data Analysis:** machine learning; deep learning; natural language processing; time series analysis; data visualization; `pandas`, `scikit-learn`, `pytorch`, `statsmodels`, `spacy`, `matplotlib`
- **Product Development:** project scoping and management; scrum and kanban; REST API design and implementation; technical writing; Jira, Balsamiq Wireframes, `flask` and `fastapi`, markdown and LaTeX

PROFESSIONAL EXPERIENCE

Senior Data Scientist   **Nov 2017 – Present**

Data Scientist   **Sep 2015 – Nov 2017**

Chartbeat; New York, NY

Developed data products and performed data analyses to provide critical insights used by online publishers around the world. Designed, implemented, and maintained multiple data systems operating at web scale to power key product features; applied novel statistical analysis and machine learning to web content, engagement, and audience datasets to answer new questions and solve new problems for journalists and editors; developed internal tools used widely by both the data science and backend engineering teams in day-to-day work; scoped and led consulting-style data projects to empower internal stakeholders.

DataCorps Volunteer, Data Ambassador, and Data Creative   **July 2014 – Jan 2020**

DataKind; New York, NY

Scoped, implemented, and managed a variety of pro bono data science projects for mission-driven client organizations. Built out data acquisition and transformation pipelines; trained and validated machine learning models for both offline and production use; designed and implemented an extensive, public-facing web app to facilitate researchers' systematic reviews; produced a 50-page technical report on recidivism in the criminal justice system; led volunteer teams on both short- and long-term engagements.

Data Scientist   **July 2012 – Aug 2015**

Harmony Institute; New York, NY

Performed mixed-methods research and developed data products examining the social impact of narrative media. Built data pipelines powering an interactive web app; led adoption of new technology and quantitative methods in response to research questions; assisted with user research, conceptual frameworks, and UX design on multiple projects.

PROJECTS

textacy: NLP, before and after spaCy

Open-source Python package for natural language processing, built on and around spaCy. An early and unusually broad contribution to spaCy's ecosystem, with functionality for language identification, text preprocessing, information extraction, topic modeling, and more.

dragnet: web page content extraction

Open-source Python package for extracting an article's main body text from HTML documents using machine learning. Notable for its excellent run-time performance and, for a time, its classification accuracy. Still widely used to create text corpora from web-scraped datasets.

EDUCATION

Ph.D. in Physics, Specialization in Elementary Particle Physics   **2007 – 2012**

Stony Brook University, Stony Brook, NY

B.A. in Physics and Spanish, Minor in Applied Math   **2003 – 2007**

Kalamazoo College; Kalamazoo, MI

Graduated *summa cum laude*, with honors in Spanish.