

Logan Cooper

Portfolio: logan-cooper.com

Github: github.com/ldtcooper

LinkedIn: linkedin.com/in/ldtcooper/

Email: LDTCoop@gmail.com

Phone: (323) 698-6487

SKILLS

- **Languages:** Python, SQL, JavaScript
- **Frameworks:** Pandas, scikit-learn, numpy, scipy, PyTorch, spacy, Flask, BeautifulSoup, D3.js, Node, Puppeteer.js
- **Tools:** Git, PostgreSQL, MongoDB, Amazon Web Services (AWS), Google Cloud Platform (GCP), Excel, Tableau

EDUCATION

- **Duke University** Durham, NC
 - *M.S. Economics and Computer Science* *Aug 2021 - May 2023*
 - Selected Coursework: Machine Learning, Natural Language Processing, Databases, Topics in Data Science, Mathematical Statistics
- **U.C. Santa Cruz** Santa Cruz, CA
 - *B.A. Global Economics (cum laude)* *Sep 2012 - Dec 2016*

EXPERIENCE

- **Duke University** Durham, NC
 - *Graduate Research Assistant* *Feb 2022 - May 2022, Sep 2022 - Present*
 - Created a novel research dataset by computationally scraping and cleaning 80GB of news broadcast transcripts into a PostgreSQL database.
 - Identified subsets of data relevant to specific research questions with BERT-based keyword extraction and topic modelling algorithms.
- **Pearson** Durham, NC
 - *Data Scientist Intern* *May 2022 - Aug 2022*
 - Developed Python data pipelines to download, clean, and transcribe 11.5 million text, audio, and video documents from internal APIs and store them in Google BigQuery.
 - Prototyped a BM25-based search and recommendation model to match student queries with resources from over 1,200 e-textbooks.
- **Federal Reserve Bank of San Francisco** San Francisco, CA
 - *Software Developer* *Jul 2020 - Jul 2021*
 - Reduced access time for one particular dataset from hours to minutes by building a custom Java ETL tool to transform CSV data into a format our data warehouse could ingest.
- **Forio** San Francisco, CA
 - *Software Developer* *Jan 2018 - Jun 2020*
 - Built React/Redux frontends for 9 educational simulations collectively worth over \$990,000 in revenue, including data visualizations and grade dashboards.
 - Saved 8-16 hours of developer time per project while improving adherence to internal coding standards by creating a Node CLI tool to auto-generate project and component boilerplate.

PROJECTS

- **Debunked! – Modelling the LOCO Conspiracy Corpus:** Group project. Uses the LOCO Conspiracy Corpus to identify topics and keywords in conspiracy theory articles and how closely those terms are linked to each other. Tech: Python, pandas, spaCy, BERT
- **PollyDarton – NLP-driven Data Curation for Polymer Nanocomposite Data Resources:** Group project. NLP-based classification of sentences in published materials science papers. Improved F1 score over current models by 0.05 points. Tech: Python, PyTorch, scikit-learn, NLTK, Gensim, RoBERTa
- **Obesity and Corruption in the Former USSR:** Working research paper with Prof. Charlie Becker which corrects for omitted variables in a 2021 paper which uses machine learning estimates of politicians' BMIs as a proxy variable for corruption. Finds that upon correcting for age and sex, the corruption-BMI relationship is weaker but still significant between the 5% and 10% levels. Tech: Python, pandas, Statsmodels
- **The Vocabulary of Buck 65, Quantified:** Blog post with analysis of rapper Buck 65's vocabulary inspired by Matt Daniels' Raper Vocab Chart. Finds that Buck 65 should be the third highest vocabulary in hip-hop. Tech: Python, Pandas.
- **Linear Regression API:** API which can perform and visualize simple OLS linear regressions. Tech: Python, Flask, JavaScript, D3.js, AWS Lambda