

David C. Ramsey - Philadelphia, PA

Dave@DataJazzDave.com | (215) 459-9287 | www.DataJazzDave.com

JAZZ TO DATA SCIENCE: I'm a jazz trumpet player who began Python during the Covid shutdown and was hooked after FastAI neural network training. I find striking similarities between data science projects and jazz composition, fueled by a passion for collaboration.

EXPERIENCE

Seek: New York Recommendations (app)

Remote

User Data Science Consultant

August - December 2022

- Trained Random Forest Regression (66% accurate) model predicting user longevity (5+ sessions) from most important behaviors (features)
- Conducted hypothesis test ($p < 0.001$) at a 95% confidence level, showing that "following a curator" in a user's first session reduced the likelihood of becoming a "super-user" (5+ sessions)
- Identified error in ETL pipeline where user data was unknowingly split into two different unique ID's skewing all ML inference

Decent Espresso Machines

Remote

[Youtube Data Consultant Project](#)

June - August 2022

- Found 20 "most important" non-coffee related channels with Geo Locations using Random Forest Regression model
- Pulled Subscriber data using Youtube API for 600+ Subscribers
- Found 20 most important coffee related channel subscriptions Random Forest model

StrataScratch

Remote

[Youtube NLP Consultant](#)

February - July 2022

- Discovered the key predictive indicators/features for a user's likelihood to subscribe to StrataScratch YouTube channel
- Discovered niche client competitors using subscription data pulled using YouTube API
- Used NLP topic modeling using Top2Vec to analyze which topics and words were related to popular and unpopular videos

Bluebonnet Data

Remote

Fellow & Mentor

January - August 2022

- Scraped and wangled valuable election data for candidate in a newly drawn state senate district from local gov. websites
- Created signature verification search tool (5x faster) and population density analysis for targeted efficient door knocking

La Colombe Coffee Roasters

Philadelphia, PA

Data Analysis Internship

January - March 2021

- Presented optimal number of retail coffee options for 30 cafes in 5 states using AWS Redshift and SQL with Tableau

Philadelphia Medical Reserve Corps (MRC)

Philadelphia, PA

Volunteer

March - June 2020

- Supported COVID-19 specimen collection at Philadelphia Federal Testing site

PROJECTS

Projects noted below can be found at www.DataJazzDave.com or www.github.com/daveramseymusic

VA Election Prediction Using Tweets: Validating Twitter Opinion Trends against Real Clear Politics Poll Aggregates in McAuliffe vs. Youngkin 2021 Virginia Governor Election

- Built an NLP neural net model: fine-tuning a WikiText-103 language model, using ULMFiT FAST.AI approach, transfer learning from my other previously trained Twitter language models, following the methods of the Bovet, Morone, and Makse 2016 Clinton v Trump Election published [nature.com/scientificreports](https://www.nature.com/scientificreports) on the 2016 Clinton vs. Trump election. Achieved a 95% accuracy on the validation set and 93.6% accuracy on hand-labeled test set of 500 items from scraped tweets.

Other Projects:

- Using [GPT-3](#) and Hugging Face Large Language Models to structure LinkedIn resume data with automatic subject tags
- Personal AWS RDS SQL cloud database to hold data pulled from multiple sources for multiple ML model training projects
- [Movie Recommender](#) using MovieLens Dataset, Stochastic Gradient Descent, and Collaborative Filtering with Latent Factors

TECHNICAL SKILLS

Programming Languages: Python, PostgreSQL, SQL, Regex; **Machine Learning:** FastAI, PyTorch, TensorFlow, Scikit-Learn, Top2Vec, HuggingFace, NumPy; **Visualization:** Matplotlib, Tableau, R Studio, Plotly; **APIs:** OpenAI, BioBERT/BERN, Youtube, Square, Tweepy, ReadIt, ; **Other:** Pandas, Scrapy, BeautifulSoup, Selenium, Google Ads, Azure

EDUCATION

Temple University

Bachelor of Arts in Jazz Trumpet Performance

Philadelphia, PA