

Ferdie Taruc

U.C. Berkeley Graduate, Data Science and Economics

@ ferdie@berkeley.edu

📍 California, United States

🌐 <https://ferdie.org>

in www.linkedin.com/in/ferdiataruc/

🔗 www.github.com/ftaruc

Experience

Presto Data Science Intern

Treasure Data

📅 Jun 2019 - Aug 2019

📍 Tokyo, Japan

- Researched as proof of concept of query workload optimization through the usage prediction of CPU, memory, and query duration through past Presto queries – with the inclusion of new engineered features
- Compared performance and efficiency of different, cross-validated tree ensemble models: xgboost, lightgbm, gbr, random forest, and stackedCV – where stackedCV outperforms all models for $RMSE$ and R^2 in both validation and test sets, but still overfits the most
- Researched clustering with new features, workload metrics and metadata stored in S3, among query signature to improve the performance of their naive query simulation framework

Undergraduate Lab Assistant

CS88, Department of Computer Science, UC Berkeley

📅 Jan 2019 - May 2019

📍 Berkeley, CA

- Tutored and monitored sixty students toward the weekly completion of computer-oriented lab assignments
- Helped create and implement a semester's worth of curriculum to teach computational structures for Data Science using Python and Git

Business Development & Finance Supervisor

Associated Students of the University of California Student Union

📅 Jan 2018 - May 2019

📍 Berkeley, CA

- Regulated financial transactions and reimbursements with over 1,000 student organizations using Excel and Abila MIP Fund accounting system
- Reviewed all invoices for appropriate documentation and approval prior to payment of all reconciliations into MIP and the ASUC's official accounting system
- Acted as liaison between students, faculty, and employees in the school to resolve issues regarding their invoices

Involvement

Director of Finance

Partnership for Pre-Professional Pilipinx [P4]

📅 Aug 2017 - Aug 2018

📍 Berkeley, CA

- Allocated and managed a tentative \$5,000 budget aligned with ASUC's Spring Budget process, contingency funding requests, and grants for the use of future events
- Fundraised an additional \$2,000 for promoting the growth of minorities (mainly Pilipinx identifying individuals) in professional spaces through networking opportunities, professional development, and other resources.

Volunteer Income Tax Assistant

Cooperate Center Federal Credit Union

📅 Oct 2018 - June 2019

📍 Berkeley, CA

- Passed an IRS-sponsored basic VITA Certification Test to legally prepare tax returns
- Provided a free tax filing service to 50+ clients that are either low-income, with disabilities, or with limited English

Skills

Programming Languages

- SQL, Java, R, Python

Programs

- MongoDB, SQLite, Jupyter Notebooks, Carto, RStudio, Git

Education

Bachelor of Arts in Data Science

University of California, Berkeley

📅 Sep 2016 - Dec 2020

📍 Berkeley, California

Bachelor of Arts in Economics

University of California, Berkeley

📅 Sep 2016 - May 2020

📍 Berkeley, California

Technical Coursework:

- | | | |
|---------------------------------------------|-------------------------------------------|--------------------------------------------|
| • Calculus | • Principles & Techniques of Data Science | • Psychology |
| • Linear Algebra and Differential Equations | • Data Mining | • Introduction to Artificial Intelligence |
| • Econometrics | • Urban Data Analytics | • Computational Structures in Data Science |
| • Introduction to Time Series | • Data Science in Linguistics | • Interference and Decisions |
| • Probability in Data Science | • Data Science in | |

Projects

Spotify Playlist Recommendation System

- Used Spotify's MDP dataset and their web API to build a playlist recommendation system using K-means clustering

Analysis of Evictions and Development in San Francisco

- Wrangled and cleaned administrative data filed by San Francisco Rent Board to analyze trends in evictions, buyout declarations, and agreements, from 1997 to 2020
- Used Carto and Seaborn for displaying interactive visualizations to explain how unfair evictions and buyouts are reasons for the growing displacement and gentrification in the area

Analysis of Proposition 13 in Alameda County

- Wrangled and cleaned data from Alameda County Assessor's Office that describes every property's assessed value and property taxes they paid in the county in 2013
- Scraped FHFA House Price indices as property growth rates to calculate total lost assessed value and total tax value loss if Proposition 13 was never legislated in California
- Calculated gini coefficients (that models property tax per square foot value) for each census block group in the county to explore discriminatory correlations of Proposition 13 on different demographics of the county's block groups

Sports Analytics: LeBron James' "Attack"

- Compared and cross-validated four different time-series (two parametric signal and two seasonal differencing) that models LeBron James' "willingness to attack the paint"
- Forecasted the next ten "attack" points following his 3/2/2019 career game, using the best performing model: a cubic, sinusoidal model that uses SMA[1][32] to capture remaining noise

Sentiment Analysis: Lyrics versus Instrumentation

- Scripted a pipeline that generates a .csv file that summarizes any artist's discography with audio features, song metadata, and lyrics provided by both Spotify and Genius's API
- Compared Spotify's engineered valence and the lyrical valence using VADER, an open source lexicon rule-based sentiment analysis, across Frank Ocean's discography