

✉ gabrielequitz@gmail.com 🌐 equitzcode.wordpress.com 📞 415-592-4819 📍 San Francisco, CA, 94116
in <https://www.linkedin.com/in/gabriel-equitz-b39b48135/> 🔄 gequitz

SKILLS

COMPUTER LANGUAGES: Python, SQL, HTML, CSS

LIBRARIES: Pandas, Numpy, Scikit-Learn, NLTK

DATA VISUALIZATION: Matplotlib, Seaborn, Tableau, Plotly, Dash

MACHINE LEARNING ALGORITHMS: Logistic Regression, XGBoost, Random Forest, KNN, Naive Bayes, Ensembling

CLOUD AND DATA STORAGE:

WEB APPS: Streamlit, Heroku, Flask, Dash

TOPIC MODELING: LDA, NMF

MACHINE LEARNING SKILLS: Regression, Classification, Natural Language Processing, Neural Networks, Deep Learning, Convolutional Neural Networks

CLOUD STORAGE AND DATA MANAGEMENT: Pickle/Serialization, Colab, Spark, Git, Postgres

EXPERIENCE

Metis

Data Scientist · Jan. 2021 to Mar. 2021 · San Francisco, CA (remote work)

- Metis is an immersive 12 week immersive data science bootcamp focused on project oriented learning
- The core curriculum is centered around Python, statistics, supervised and unsupervised machine learning, exploratory data analysis, databases, and visualization techniques
- Completed five self-designed data science projects from conception to presentation; including data collection, data management, exploratory data analysis, modeling, and visualizations
- Project highlights include:

Sovereign Risk Model (Supervised Machine Learning)

- Created a Classification Model that predicts probability whether a country will default on its debts or not in the following years
- Analyzed how features such as inflation, exports, and Government Expenditures % of GDP affect likelihood of Sovereign Default and Restructuring
- Developed a Web App allowing users to explore model, adjust features. Intended to help investors avoid buying bonds in countries likely to default, and understand what factors lead to Sovereign Default
- Data was economic data from the IMF since 1980 and Historical Sovereign Default and Restructuring dataset since 1970
- Tools: Binary Classification, Supervised Learning, Web App. Matplotlib, Numpy, Pandas, Scikit-learn, Seaborn. Tableau for visualizations. XGBoost, Random Forest, KNN, Naive Bayes, Logistic Regression. Streamlit for Web App. Hosted on Heroku
- Best results were achieved with Logistic Regression. Debt/GDP seems to be most correlated with a country defaulting on its bonds

State of The Union Analysis NLP (Unsupervised Machine Learning)

- Analyzed how topics in the State of the Union addresses have changed over more than 200 years of American history and compared which speeches were most similar topically using Natural Language Processing (1790-2018)
- Used data pulled from Kaggle on the State of the Union corpus
- Tools: Python Pickle, NLTK, Pandas, Scikit-learn, Matplotlib. Count Vectorizer, TF-IDF, NMF
Tableau used for plots in presentation slides

Issuing Credit Cards (Classification with Supervised Learning)

- Created a model for a credit card company to avoid issuing credit cards to users who default
- Pulled data from a Kaggle data set consisting of credit card client data
- Tools: Matplotlib, Numpy, Pandas, Scikit-learn, Seaborn, XGBoost, Random Forest, KNN, Naive Bayes, Logistic Regression
- Higher Education, Marriage, and Women were correlated with lower risk of credit card default
- **Completed** a set of 4 challenges on SQL. This contributed to a greater knowledge and confidence on Structured Query Language

BoxOfficeMojo Box Office Analysis (Web Scraping & Regression)

- Utilized regression to find which features are correlated with the highest Movie Box Office returns
- Data was pulled from Box Office stats on Top Lifetime Grossing movies and was analyzed after web scraping from BoxOfficeMojo.com
- Created features from scraped HTML including: Distributor (studio), Budget (USD), Foreign Market Count, Genres (10), and Rating (MPAA)
- Tools: Pandas, Numpy, Scikit-learn, Seaborn, Matplotlib, BeautifulSoup, lxml
- Model showed Higher budgets correlated with higher box office gross

MTA Exploratory Data Analysis

- Provided recommendations on optimizing deployment of a tech client's team to collect email addresses of potential attendees for their upcoming gala event
- Turnstile data from New York City subway stations (MTA) in combination with NYC demographic data from the census was used for the project
- Tools: Pandas, Numpy, Scikit-learn, Matplotlib, and Seaborn
- Identified five MTA stations with the most traffic over our 4 week period
- The data was aggregated to find the average weekly activity based on time and weekday and demographics were analyzed for the top five stations

San Francisco Zoo

Summer Counselor · June 2012 to Aug. 2013, June 2012 to Aug. 2013 · San Francisco, CA

- Summer camp counselor for children ages four to ten
- Supervised camp activities to ensure children's safety
- Communicated any updates or issues with supervisors on staff
- Collaborated with other counselors to maintain a fun and safe environment for the children

EDUCATION

San Francisco State University · Aug. 2014 to Dec. 2019

Bachelors of Science Computer Science 2019