

# EDITH JOHNSTON

## » CONTACT

✉ ediejn@gmail.com

☎ 408-750-7071

📍 San Jose, CA 95125

in linkedin/in/edith-johnston/

🌐 edithalice

## » EXPERIENCE

### Metis Data Science Bootcamp

Online · 07/2020 - Present

Highly intensive project based accredited data science training program

### Starbucks

*Barista*

San Jose, CA · 06/2018 - 09/2018

*Shift Supervisor*

San Jose, CA · 09/2018 - 03/2020

Coordinated and deployed teams of 2-6 people in high pressure environments  
Ordered, tracked and managed store inventory

### Reputation.com

*Data Acquisition and Entry Temp*

San Mateo, CA · Summer/2017

Researched and input data in order to onboard new criteria of customer base

### Various

*Tutor*

2009 - 2019

Taught subjects including mathematics, statistics, chemistry, physics, biology, psychology, writing, history and more, at levels ranging from elementary to collegiate

## » SKILLS

**LANGUAGES:** Python, SQL, JAVA, MATLAB

**LIBRARIES:** Sci-kit learn, Numpy, Pandas, Dask, Xarray, Tensorflow, Keras, NLTK, SpaCy, Gensim, Flask, BeautifulSoup, Selenium

**MACHINE LEARNING TECHNIQUES/ALGORITHMS:** Linear Regression, Logistic Regression, Decision Trees/Random Forest, XGBoost, HistGradientBoost, LSA/NMF, LDA, Word2Vec, FastText

**DATA VISUALIZATION:** Matplotlib, Seaborn, Tableau, Cartopy

**CLOUD/DATA STORAGE:** Google Colab, Google Cloud Platform

**MATHEMATICAL BACKGROUND:** Linear Algebra, Ordinary and Partial Differential Equations, Numerical Analysis, Probability, Vector Calculus, Complex Analysis, Population Modeling

## » VOLUNTEERING

### Women in Tech International

*Annual Summit Assistant*

San Jose, CA · 06/2015 - 06/2018

Helped plan and run the coaching segment of the annual summit  
Handled cancellations and unexpected changes to the schedule, website, and app  
Directed volunteers and changes during the summit

## » EDUCATION

### University of New Mexico

B.S. Applied Mathematics 2019

Minors in Computer Science, Chemistry, and Evolutionary Anthropology

## » PROJECTS

### Modeling Global Carbon Dioxide Levels with Vegetation Data

Objective: Regression Model

- The goal of this project was to build a global scale model to quantify the impact of vegetation levels on local carbon dioxide atmospheric concentrations.

Data Sources

- For this project, I acquired two datasets from NASA's Earthdata portal, the Terra MODIS VCF yearly product, was annual vegetation data with a resolution scale of 250m, which meant that the full global span of the data contained almost 1.5 billion data points for each year. The second of these was the CO2 data from TES Level 2 data product.

Main Techniques

- Regression modeling, big data computation, time series autoregression

Main Tools and Packages Used

- Python, Jupyter, Xarray, Dask, NetCDF4, HDF-EOS, Numpy, Pandas, PyProj, Sci-kit learn, Statsmodels, Cartopy, Matplotlib

### Stellar Spectral Type Classification

Objective: Classification Model

- The goal of this project was to use stellar spectroscopic data to determine Morgan-Keenan spectral type, in an effort to automate the process of classifying stars. (MK Spectral Types)

Data Sources

- For this project, I acquired spectroscopic data acquired from the Sloan Digital Sky Survey's SQL portal (CasJobs), containing both spectroscopic line index measurements, and stellar type labels.

Main Techniques

- Classification modeling, class balancing, model selection

Main Tools and Packages Used

- Python, Jupyter, SQL, CasJobs, SciDrive, SciServer, Sci-kit learn, Imblearn, XGBoost, Matplotlib, Seaborn, Pandas, Numpy

### Jane Austen Style Text Generation

Objective: Text Generation

- The goal of this project was to custom train word embeddings and use LSTMs in a neural network model to generate text in the semantic style of Jane Austen. In addition to this, I built a Fill in the Blank style web app. (Web app demo!)

Data Wrangling and Modeling Process

- I acquired the text of Jane Austen's six novels from Project Gutenberg.

Main Techniques

- NLP, unsupervised/deep learning, regular expressions

Main Tools and Packages Used

- Python, Jupyter, TensorFlow, FastText, Word2Vec, Gensim, NLTK, Google Colab, Flask, Bootleg

### Modeling NYC Public School Success

Objective: Regression Model

- The purpose of this project was to build a model predicting the academic success of New York City public schools, based on available demographic information, as well as funding and spending reports. I also wanted to attempt to determine the most influential factors in a school's success (or lack thereof), for the eventual hypothetical purpose of mitigating or improving such factors.

Data Sources

- To quantify school success, I used data from the New York City Department of Education's annual Quality Review. I was also able to acquire demographic data from this source. In order to acquire funding data, which was publicly posted, but not easily downloadable in a readable format, I wrote a web scraping script to acquire the relevant data from New York State's Education Department's data site.

Main Techniques

- Web scraping, regression modeling, feature selection

Main Tools and Packages Used

- Python, Jupyter, Pandas, BeautifulSoup, Selenium, Sci-kit learn, Statsmodels, Matplotlib, Seaborn, Yellowbrick

