|  |  |
| --- | --- |
| DAVID **CASTILLO** | Brooklyn, NY  718.415.6767  dcast822@outlook.com |

|  |  |  |
| --- | --- | --- |
| **LinkedIn:** [linkedin.com/in/davidcastillo822](https://www.linkedin.com/in/davidcastillo822/) | **GitHub:** [github.com/dcast822](https://github.com/dcast822) | **Portfolio:** [datadavidc.com](https://www.datadavidc.com/) |

**>LANGUAGES AND TECHNOLOGIES**

* **Proficient:** Python, Pandas, NumPy, SciPy, Seaborn, Matplotlib, Machine Learning, Scikit-Learn, NLP, NLTK, Statsmodels, Deep Learning, Keras, Tensorflow, SQL, PostgreSQL, SQLite, MYSQL
* **Exposure:** Spark, Streamlit, Flask, Heroku, AWS, GCP, Tableau

**>PROFESSIONAL EXPERIENCE**

**T-Mobile** **| Manager | New York, NY** July 2013 - October 2021

* Conducted workshops on best practices for sales, using Dale Carnegie and Kaizen principles, leading to ~15% less escalations from the prior year and a decrease in ~$1.5K in credits.
* Established metrics for the sales team, using Power BI and Excel, resulting in 5 goals met at ~105% of target.
* Ensured compliance with the Operations Department leading to the resolution of discrepancies that shifted losses from my company to the logistic company in the average amount of ~$15K a year.
* Conducted training, using Dale Carnegie principles, and Kaizen, leading to 40+ personnel having promotions or improving ability to meet their targets.

**>PROJECT WORK**

**Housing Sales Price Prediction with Machine Learning | Data Scientist |** [**Code URL**](https://github.com/dcast822/Housing_salesprice_prediction_with_Machine_Learning)2022

*Predict housing prices based on ~30 different attributes using Linear Regression.*

* Applied imputing techniques to address null values, using Python, Scikit-learn, Simple & Iterative Imputer, Pandas, and NumPy, achieving an R2 score of 90% while including an additional ~200 rows.
* Used regression analysis to gain inference on model, using Python, Scikit-Learn, Pandas, and NumPy, gaining insights such as; with all else held constant, a house in the Greenhills neighborhood increases value by $102K.

**NLP Analysis with Machine Learning for Coffee & Tea Subreddits | Data Scientist |** [**Code URL**](https://github.com/dcast822/Coffee-Tea_subreddit-NLP-analysis)2022

*Predict origin of posts between two distinct subreddits using logistic regression.*

* Utilized web-scraping and NLP to process 20K posts from two subreddits, using Python, BeautifulSoup, Scikit-Learn, NLTK, and Logistic Regression, achieving a specificity score of 97% and recall score of 96%.
* Applied NLP techniques on two datasets, using Python, Pandas, NumPy, NLTK, word\_tokenize, stopwords, Scikit-Learn, Count Vectorizer, allowing me to identify the top 10 bigrams in each subreddit.

**Generate Music with Deep Learning | Data Scientist |** [**Code URL**](https://github.com/dcast822/Generate_music_with_deep_learning)2022

*Feed integers converted from MIDI file notes to Neural Networks and learn those notes and generate new music.*

* Utilized web-scraping to collect MIDI Files and convert notes into integers, using Python, BeautifulSoup, Music21, and NumPy, yielding ~1.2K unique notes from 1K songs produced by Neural Network architectures.
* Tuned hyper-parameters for neural networks, adjusting for drop out rate, learning rate, layer shape, dense layer, and dense shape, using Python, NumPy, Keras, TensorFlow, resulting in a value loss of 3.5.

**>EDUCATION**

**Data Science Bootcamp,** *General Assembly* Feb 2022

**MBA (Accounting),** *University of Arizona Global Campus* March 2021

**Bachelors (Political Science),** *CUNY Lehman*  May 2010