Alan Flint

San Francisco, CA | 425-273-4884 | alan@flintagan.org
alanflint.com | linkedin.com/in/alan-flint/ | github.com/alan-flint

Education

University of San Francisco San Francisco, CA
Master of Science in Data Science *June 2020*
- Deep Learning Certificate
- Relevant Coursework: Python, Databases (SQL), Machine Learning, A/B Testing, Deep Learning, Distributed Computing (Spark), Data Structures & Algorithms, Time Series Analysis, Data Visualization

University of Maryland College Park, MD
Bachelor of Science in Mathematics and Economics *May 2019*
- Relevant Coursework: Statistics and Probability, Linear Algebra, Econometrics, R
- Studied abroad at the University of Sydney, Australia for one year

Work Experience

Trulia San Francisco, CA
Data Science Intern *November 2019 - Present*
- Discovered the most influential website features for converting visitors to home buyers using an interpretable machine learning model and SHAP feature importance
- Built an internal power analysis tool in Python and Streamlit to calculate how long to run an A/B test for various in-house KPI's, providing key stakeholders a resource to efficiently and accurately run experiments
- Deployed an ETL on AWS to automatically query for key business metrics and validate data sources using Presto and MySQL, saving hours of manual querying time

Projects

- Context Recommender System - Code Notebook
  - Predicted users' ratings on local businesses with a matrix factorization model coded from scratch in PyTorch; achieved 0.65 MSE on the Google Local reviews dataset
  - Improved upon the baseline model by incorporating written reviews into the model framework with a GRU language model and pre-trained GloVe embeddings
- Indigo Music Studio - Indigo Music Studio
  - Developed a web application for users to generate new music with machine learning using Google's Magenta platform; collaborated with six other students
  - Implemented Google Analytics tracking to monitor user behavior, wrote PyTests to ensure code functionality and accuracy, and established proper code documentation
- Classified Damage to Buildings Hit by the Nepal Earthquake - Code Repository
  - Modeled damage level to buildings with Random Forest by engineering features from geographic and building structure variables
  - Currently placed top 4% in the drivendata.org competition among 2300 participants with a 75% F1-score

Skills and Technologies

- Python (scikit-learn, NumPy, pandas, PyTorch, matplotlib, Plotly, Flask, spaCy), SQL, Spark (PySpark, SparkSQL, SparkML), AWS (EC2, EMR, EBS), git, R (tidyverse, ggplot2)