Scott Cole

San Francisco, CA   803-370-9183   scott.cole0@gmail.com

https://srcole.github.io   www.github.com/srcole   www.linkedin.com/in/scott-cole/

Skills

Languages: Python - pandas, sklearn, seaborn, matplotlib, pyspark, dash, scipy; SQL; Git
Techniques: data mining, data visualization, data wrangling, machine learning, statistics, signal processing

---

Experience

Data Scientist   Square   Jan 2020 - present
- Fraud detection on the Risk team for Cash App. More details TBD.

Data Scientist   Samba TV   Apr 2019 - Jan 2020
- Designed and implemented a framework to benchmark the accuracy of core data processing pipelines
- Conducted large amounts of exploratory data analysis to identify systematic errors by our algorithms
- Collaborated with engineers and researchers to implement improvements and assure quality of changes

Data Science Fellow   Insight Data Science   Jan 2019 - Mar 2019
- Created a gradient boosting model to authenticate messages based on users' unique keystroke patterns
- Deployed an interactive dashboard to visualize results of the authentication model to aid security decisions

Data Science Intern   Crime Lab New York   Sep 2017 - Dec 2017
- Furthered a machine learning pipeline to predict police misconduct from officer reports and activity data
- Mined data to uncover systematic differences in police behavior to identify problematic districts and officers

Graduate Student Researcher   University of California, San Diego   Sep 2014 - Dec 2018
- Identified limitations in standard signal processing techniques. Developed new analysis methods to overcome gain new information from noisy brain signals. Published python code to reproduce publications.
- Led development of 3 collaborative open-source python packages on GitHub (1000+ downloads)
- Mentored 7 students by directing projects and teaching Python, machine learning, and statistics

---

Personal Data Science Projects (See more at https://srcole.github.io/dataprojects/)

10-dimensional burrito review system
- Developed a system to systematically review over 350 burritos at over 75 taco shops across San Diego
- Analyzed the importance of and relationships between burrito features using linear regression and PCA

Measurement of open science
- Mined the text of over 100,000 academic papers to quantify the prevalence of code sharing
- Coordinated with a team of 8 at a hackathon to extend the project to explore data sharing and 10x papers

---

Education

University of California, San Diego	**Ph.D. Neuroscience**	La Jolla, CA	Dec 2018
Clemson University	**B.S. Bioengineering**	Clemson, SC	May 2014