Matt Shadish

https://github.com/mshadish
mshadish@sbcglobal.net | (925) 487-1921

https://www.linkedin.com/in/matt-shadish/

EXPERIENCE

ENGAGE3 | Davis, CA

Data Science & Analytics Lead | Aug 2015 - Current
Analytics Consultant | Jun 2014 - Aug 2015
Data Scientist | Dec 2013 - Jun 2014

- Data Science & Analytics Lead on team of 10-15 full-time employees where we turned company around from 2-week runway (2015) to closing Series B funding round (2017)
- Designed, developed, and implemented pipeline of in-store data collection, management, analysis, and on-demand reporting; achieved via SQL databases, Python processes and endpoints, and bash shell scripts; system supported 50%+ of total 2017 revenue
- Led team of 2-4 contractors and employees to scale-up data processing pipeline and reporting
- Directly drove 3% of 2017 revenue through analytics offering of reverse-engineering competitor pricing strategies using 10M+ data points; methods include unsupervised clustering and classification algorithms using Scala and Spark on Amazon EMR
- Interfaced directly with clients regularly to outline client needs and requirements; created and presented Tableau and Powerpoint analyses directly to clients
- Designed, developed, and deployed realtime processing of 1M+ daily crawl records in Amazon Redshift and Amazon RDS Postgres
- Deployed and actively maintained 5+ Python Flask and CherryPy endpoints for external and internal needs on Amazon Elastic Beanstalk and company hardware
- Created fully-automated QA system to process millions of price records hourly using Oracle PL/SQL

GENERAL ELECTRIC | San Ramon, CA

Data Analytics Intern | Nov 2014 - Mar 2015

- Created easy-to-use search flow in parts database using Elasticsearch, Python Flask, and basic JavaScript

ATAC | Sunnyvale, CA

Aviation Data Analyst | Jun 2012 - Jun 2013

- Analyzed hundreds of thousands of flight data points to identify opportunities for noise and emissions reduction in standard departure/arrival routes using Microsoft Excel extensively

EDUCATION

UNIVERSITY OF SAN FRANCISCO | MS in Analytics

2015
Cum. GPA: 3.88

UNIVERSITY OF CALIFORNIA, LOS ANGELES | BS in Mechanical Engineering

2013
Cum. GPA: 3.75 • Phi Beta Kappa • Tau Beta Pi

SKILLS

TOOLS & TECHNOLOGIES
AWS
EMR • Elastic Beanstalk • Redshift • S3 • EC2
Tools
Spark • Tableau • Apache Solr • Elasticsearch • Redis
Project Management
Git • Apache Subversion • Assembla • Basecamp

LANGUAGES
10,000 lines
Python • SQL (Oracle, PostgreSQL, Redshift)
1,000 lines
Bash • Scala • LaTeX
100 lines
JavaScript (D3.js) • R

PROJECTS

LENDINGCLUB PORTFOLIO PERFORMANCE ANALYSIS
`https://github.com/mshadish/lendingclub/`
- Hosted at `http://mshadish.github.io/lendingclub/`
- Wrote methods to automatically collect LendingClub personal portfolio data nightly and store time-series loan data in a SQLite database using Scala and LendingClub's REST API
- Currently analyzing trends in borrower solvency using visualizations in D3.js

TREASURY YIELD CURVE VISUALIZATION
`https://github.com/mshadish/yield-curve/`
- Hosted at `https://mshadish.github.io/yield-curve/`
- Download and transform historical XML treasury yield data using Python to feed into a D3.js visualization to show yield curve changes over time
- Automatically pushing new data using cron and bash

AUTOMATED FUND RECOMMENDER
`https://github.com/mshadish/morningstar-scrape/`
- Scraped Morningstar daily for a select set of mutual funds and stored time series data locally with Python and cron
- Identified buy and sell opportunities using fundamental valuation metrics and delivered insights via SMS using Python

KAGGLE: DRIVER TELEMATICS ANALYSIS
`https://github.com/mshadish/kaggle_drivers/`
- Extracted and created features from driver positional data using R to establish driver fingerprints
- Detected anomalies in driver data using a combination of a bagged logistic regression estimator and a random forest in Python with scikit-learn
- Scored in the top 30% of contestants out of 1500+ with an AUC of 0.85