LEON YIN

hello.leonyin@gmail.com
(781) 439-2875
yinleon.github.io

RECENT WORK EXPERIENCE

THE ORCHARD
January 2016 to Present

Data Science Support Analyst
I make data of all formats, sizes and origins machine-readable using ETL's that employ distributed systems, cloud computing and workflow managers. I take part in exploratory analysis, data modeling and dashboard creation. I have productized the following:

- A Looker microsite to quickly navigate to relevant reports.
- Dashboards that join user input (from Google Sheets) to SQL queries and allow hot-linking values to external workflows.
- PySpark job that transforms raw data via user-defined functions, joins metadata from Redshift, and writes to s3.
- Market share dashboard extracted from 6 stores.
- ETLs from the Youtube and Spotify API.
- Random Forest for binary classification of deliverable content.

The stack: ***Jupyter, Pandas AWS (EC2, S3, Redshift), Looker, SQL, Spark, Scikit-Learn and Airflow.***

NASA GISS
July 2015 to August 2015

Climate Science and Oceanography Intern
Worked on a Fortran numerical model and performed linear regression, jackknife analysis and 2D-interpolation on global seawater measurements using a modern data science stack: ***Jupyter, Pandas, NumPy, SciPy, Matplotlib and Basemap Toolkit***.

October 2015 to August 2016

Web Development Consultant (Part-time)
Built QC pipeline for data and an interactive ***d3.js*** map[1]. Research published at the American Geophysical Union[2] and map live on NASA (soon).

HOBBY PROJECTS

FEDERAL FUND TRACKER 2016

NSF research lacks transparency— records are not stored in a format conducive to analysis at scale. I parsed XML-formatted NSF awards into ***SQLite*** databases to trace the funding history of institutions, research topics, and PIs with interactive visualizations using ***Jupyter, Pandas, Seaborn and Plot.ly***[3]. Used ***Scikit-Learn*** Bag of words to look at topic trends. Created network graph of investigators based on award co-authorship.

PALMER LTER INORGANIC ANALYSIS 2016

Performed spatial-temporal analysis on 15-years of physical and biochemical measurements taken along the Antarctic's Western Peninsula using ***Jupyter, Pandas, NumPy, SciPy, Seaborn, Matplotlib and Basemap Toolkit*** to understand how the marine carbon cycle— which generates 50% of atmospheric O_2, will respond to climate change[4]

FUNDED PROJECTS

JAPANESE BREAKFAST WEBSITE 2016

This project addresses the financial and technical barriers for indie bands to have a professional website. Tour and merchandise tables generated live via ***Google Sheets Javascript API*** on ***HTML5*** templates.

DEAN'S UNDERGRADUATE RESEACH FUND 2015

Designed an experiment on the physiological effects of Ocean acidification on oysters. Learn more about Ocean acidification from NOAA[5]

NYU GREEN GRANTS 2015

Wrote and produced "One Man's Trash," a short film inspired by Portlandia to spread awareness of composting[6]

EDUCATION

NEW YORK UNIVERSITY

BS in Chemistry, 2011 - 2015
Minor in Computer Science and Math

COMPUTER SKILLS with years of experience >1

Languages	Python[4], MySQL[3], Bash, HTML, Fortran
Programs	Matlab[2], AWS, Spark, Airflow, Jupyter[2]
OS	Mac OS X, Linux/Unix

LINKS

[1]	bit.ly/d18o_d3js	Bl.ocks page
[2]	bit.ly/agu_poster	AGU poster
[3]	bit.ly/nsf_notebook	Jupyter Notebook
[4]	bit.ly/palmer_notebook	Jupyter Notebook
[5]	https://www.youtube.com/watch?v=MgdlAt4CR-4	Youtube Video
[6]	bit.ly/1_mans_trash	Vimeo Video