Jenny Rhee

(337) 794-4923 | jennyirhee@gmail.com | jennyrhee.github.io | jennyrhee | jenny-rhee-84ab5576

Projects

Severe Storm Events in Louisiana

- *(In progress)* Using two NOAA data sets (severe storm events and meteorological) for EDA and modeling storm events in Louisiana
- Technologies used: Python (pandas, NumPy, pandarallel, seaborn, matplotlib, BeautifulSoup, requests), SQLite

Skills

Languages Python (NumPy, pandas, matplotlib, seaborn), SQL, Java, MATLAB

Tools Git, Microsoft SQL Server, VS Code

Technical Skills Machine learning (scikit-learn), experimental design, statistics, NLP (NLTK), time series (statsmodels)

Experience

Research Associate — Baton Rouge, LA

Louisiana State University — *June 2019 to present*

- Compiling data from various sources to build an econometric model to analyze the effects of driving forces on a variety of environmental impacts in the U.S. and Germany over the past two centuries
- Developing a simulation model in Python to illustrate the relationship between embodied energy and fitness of a K-selected species
- Technologies used: Python, Excel

Data Analyst Intern — Lafayette, LA

Acadian Ambulance — *Aug 2018 to May 2019*

- Led the first data science projects to be conducted at the company
- Exploratory analysis (clustering using scikit-learn and topic modeling using Gensim and MALLET) of rejected medical records to categorize and understand the common reasons for rejection
- Analyzed the "virality" of medics with low medical documentation accuracy on their partners with the goal of changing training procedures
- Time series analysis (statsmodels) to forecast future daily numbers of billable calls for more efficient medic scheduling recommendations to operations managers
- Technologies used: Python, T-SQL, Microsoft SQL Server

Great Lakes Summer Fellow — Ann Arbor, MI

University of Michigan — *May 2018 to Aug 2018*

- Developed a data processing script in Python to normalize historical time series data from 15 stations and buoys in the Great Lakes (2015-2017; 196 million observations)
- Proposed a data management solution for 15 stations and buoys in the Great Lakes. Data were in inconsistent formats and units over time and between stations, ranged over a decade, and stored as flat text files on a server.
- Designed and implemented a time-series database (TimescaleDB) prototype to manage historical and real-time streaming data from the Great Lakes
- Technologies used: Python, TimescaleDB

NSF REU Fellow — Dauphin Island, AL

Dauphin Island Sea Lab — *May 2017 to Aug 2017*

- Won 1st place in the REU poster symposium, awarding full funding to present research at Ocean Sciences Meeting in Portland, OR
- Collaborated with a physical oceanographer and biological oceanographer to design an experiment using existing data that had yet to be analyzed
- Processed, sanitized, and compiled several years (2009-2012) of CTD data from 15 stations
- Calculated Model-I linear regressions and statistics to make novel conclusions about the Mobile Bay to shelf transect
- Technologies used: MATLAB, SeaBird SBE Data Processing, Excel

Education

Data Science Career Track, Certification — Online

Springboard — *July 2019 to present*

- 6 month intensive course in data science, data visualization, machine learning, hypothesis testing, Python, SQL, and Spark
- Estimated completion: November 2019

Bachelor of Science in Biology — Lafayette, LA

University of Louisiana at Lafayette — *May 2018*

- Relevant coursework: Advanced Data Structures and Software Engineering, Linear Algebra, Calculus I and II