

Adept and resourceful statistician experienced in biological research, sports analysis, and digital marketing. Committed to identifying and communicating clear solutions to complex problems.

PROGRAMMING LANGUAGES

C++ LaTeX
SAS SQL
HTML CSS
R (astsa, qpcR, rpart, glmnet)
Python (Pandas, NumPy, Seaborn)

TECHNOLOGIES

Git Ver. Control/Collaboration
Google Analytics
MS Office
Jupyter Notebook
UNIX
Databricks

TECHNICAL KNOWLEDGE

Regression/Classification

- Multivariate Linear
- Decision Trees

Generalized Linear Models
Stochastic Processes
Time Series

- ACF & PACF Interpretation
- Spectral Analysis

Machine Learning

- Support Vector Machines
- K-Nearest Neighbors

Experimental Design
Data Retrieval & Preprocessing
Web Scraping/Crawling
Feature Engineering

INTERESTS

I enjoy camping and hiking; my favorite trail is the Narrows at Zion National Park.

EDUCATION

University of California, Santa Barbara 2020
| B.S. Statistics & Data Science

- Dean's Honors List - Fall 2016, Winter 2017

WORK EXPERIENCE

Piedmont Racing Ltd. Dec 2019 – Present
| Analytics Platform Design & Development Intern

- Developed psychological index -- an indication of a rider's mental state based on individual recent history and standing/status through time.
- Collaborated with team members and domain experts remotely using Slack and Databricks to develop and discuss numerical indexes and simulations.
- Cleaned and gathered data on individual riders and over three years of races for index development use.

Channel Islands YMCA Nov 2019 – Mar 2020
| Data & Analytics Marketing Intern

- Utilized data gathered from Google Analytics, Facebook Insights, and market research to verify and expand on previous demographics reports and projections.
- Created visualizations and generated insights on customer characteristics from gathered data using R (grDevices, ggplot2) to help make data-driven decisions.
- Assisted marketing manager in competitive analysis and SEO by identifying key competitors across six branch locations, determining competitor analytic platforms, and comparing rival Facebook content engagement with other metrics.

PROJECTS

Web Scrape & Visualization of Used Car Prices [apang782.github.io/vroom1](https://github.com/apang782/vroom1)
| Python (BeautifulSoup, Pandas, NumPy, Selenium, Seaborn, Matplotlib)

- Designed automated scraper using Selenium and BeautifulSoup to pull data from over a thousand dynamic Javascript used car pages with randomized URLs.

Regression Modeling & Prediction of Used Car Prices [apang782.github.io/vroom2](https://github.com/apang782/vroom2)
| R (car, MASS, glmnet, rpart, randomForest, gbm)

- Devised multiple regression analyses on over 20 predictor variables using ridge/LASSO regression, random forests, and other methods. Best tuned model able to reliably predict used car prices with a \$2400 margin of error.

Time Series Analysis of US Monthly Candy Production [apang782.github.io/CTSA](https://github.com/apang782/CTSA)
| R (MASS, forecast, astsa, qpcR, tseries, TSA, GeneCycle)

- Generated SARIMA model using AICc, ACF and PACF interpretation and diagnostics. Model is proven to accurately forecast a full year of future production values.