Andy Pang

Los Angeles, CA
apang782.github.io
☎: (626) 533-3157
✉: apang782@gmail.com

Adept statistician experienced in biological research, sports analysis, and digital marketing.

EDUCATION

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

PROGRAMMING LANGUAGES

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

TECHNOLOGIES

Git Ver. Control/Collaboration
Google Analytics
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

WORK EXPERIENCE

Piedmont Racing Ltd. Remote
Analytics Platform Design & Development Intern Dec 2019 – Sep 2020

- Developed psychological index: an indication of a rider's mental state based on individual recent history and standing/status through career.
- Utilized domain expert knowledge to perform statistical case studies on key riders.

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

- Expanded on past demographic reports and projections with 2019 Google Analytics and Facebook Insights data.
- Created visualizations on customer characteristics across 6 branch locations using R (grDevices, ggplot2).
- Conducted competitive analysis by investigating rival analytic platforms and Facebook content engagement.

PROJECTS

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

- Designed automated scraper using Selenium and BeautifulSoup to pull data from 1440 dynamic Javascript used car pages with randomized URLs.
- Investigated ranges, distributions, and outliers of 22 different car attributes through visual analysis and general statistics.
- Determined archetypal listed car to be a 2019 Infiniti SUV, with 28,000 miles, 1 previous owner, and a price tag of around $19,000 or $26,000.

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

- Found torque, horsepower, and city MPG as the most important predictors of list price during decision tree, linear modeling, and ensemble method development.
- Predicted 218 car prices with $2400 margin of error using random forest model.
- Decided against removing any observations after testing outlier and high leverage data removal effects.

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

- Achieved stationarity and normality as seen in model decomposition and the augmented Dickey–Fuller test by differencing and using the Box Cox transformation.
- Generated SARIMA model by using AICc, ACF and PACF interpretation.
- Predicted Sep 2016 to Aug 2017 production values fall within a 95% CI and are nearly identical to actual values.