JOHN LA

535 Bowdoin Street, San Francisco, CA, 94134 • (415) 837-8095 • jl6105@columbia.edu •
johnla228.github.io • www.linkedin.com/in/johnla228/

EDUCATION

Columbia University in the City of New York   *New York, NY*
Master of Science in Applied Analytics, **GPA**: 3.83/4.00   *May 2023*
University of Southern California, Viterbi School of Engineering   *Los Angeles, CA*
Bachelor of Science in Computer Science/Business Administration, **GPA**: 3.55/4.00, Dean's List, Cum Laude   *May 2021*
- Relevant Coursework: Data Storytelling, Probability and Statistics, Research Design, Supervised Learning, Unsupervised Learning, SQL and Relational Databases, Python Programming, Managing Data, Software Development, Algorithms and Computing Theory, Data Structures and Object-Oriented Design, Software Engineering, Strategic and Operations Management, Applied Business Statistics, Machine Learning, Design and Construction of Large Software Systems

PROJECTS

Spotify Clustering Analysis & Recommender Systems   *May 2022*
- Extracted Spotify API data containing popular songs and their audio features and performed a K-means clustering model to understand the different audio attributes, genres, and variation among the clusters that characterize a popular song
- Applied Principal Component Analysis to reduce the number of variables in the data into factors and utilized a non-personalized recommendation system and item-based collaborative filtering to analyze the most popular songs and genres

Apple iPhone BNPL Sales   *December 2021*
- Conducted a simple random sampling with a sample size of 10,000 divided into four testing groups of 2,500 (no BNPL, 12 months, 18 months, 24 months) to determine the optimal Buy Now Pay Later scheme to aid Apple in maintaining its long-term dominance in the market amid increasing competition by looking at the impact on customers' purchase decisions
- Simulated the experiment data in R, devised research questions using null and alternative hypotheses, performed two-sample t-Tests, and analyzed the mean effect and p-value for each test to see whether there were statistical differences

NYC Airbnb Rental Prices Prediction   *December 2021*
- Created models, using linear and logistic regression, lasso, feature selection, decision trees, random forests, gradient boost machine, to predict the price of a set of 35,000 Airbnb rentals in New York City containing 90 variables
- Placed 1st in a class of 60 students based on accuracy of price predictions evaluated using root mean squared error

PROFESSIONAL EXPERIENCE

Genentech   *San Francisco, CA*
Incoming Data Analytics Intern, Strategy Analytics & Intelligence   *September 2022 – December 2022*
Health Care Service Corporation/Blue Cross Blue Shield of Illinois   *Chicago, IL*
Data Analytics Intern   *June 2022 – August 2022*
- Built machine learning models on provider and claims data, containing 35,400+ rows and 37 columns, to determine what features/predictors affect COVID-19 claim disputes getting rejected and dispute counter offers getting accepted
- Constructed views and extracted data using SQL on employee tables, containing 30 million+ rows and 28 columns, to analyze work quality and productivity through collaboration and meeting hours in similar and different office locations
- Conducted A/B Testing and designed two partition models based on tenure and performance ratings to assess if there is a statistical significance in the rate of turnover and engagement scores for employees who utilize tuition reimbursement

Columbia University in the City of New York   *New York, NY*
Data Analyst Intern   *January 2022 – May 2022*
- Extracted email performance campaign data from Slate to create weekly productivity reports on Excel that include 5 KPIs, such as opens, clicks, bounces, and opt-outs, to improve university marketing strategies and decision-making
- Designed data dashboards, using Google Data Studio, for email marketing campaigns of 30+ university programs
- Collaborated cross-functionally with school programs and department staff to build email templates, using HTML, CSS, and Adobe Dreamweaver to produce over 700+ emails catered to current and prospective university students

University of Southern California   *Los Angeles, CA*
CSCI 310 – Software Engineering Course Producer   *January 2021 – May 2021*
- Assisted a class of 225 students by holding office hours, staff collaborating, and grading assignments and exams
- Guided students in learning agile processes and tools, such as Git, GitHub, Maven, JUnit, Cucumber, and Travis CI
- Acted as a stakeholder/client for students' agile-based projects to simulate sprint review process of an application

SKILLS & INTERESTS

Technical: C++, Java, Python (NumPy, Pandas, Flask), JavaScript, HTML, CSS, Bootstrap, Tableau, SQL (Postgres), NoSQL (MongoDB), R (ggplot2, dplyr, stringr, lubridate), Git, supervised learning, unsupervised learning, Elasticsearch, Neo4j, A/B testing
Languages: Chaozhou (fluent), Mandarin Chinese (conversational), Japanese (conversational), Spanish (conversational)
Activities: Applied Analytics Club (VP of Alumni Relations), International Student Assembly (Chair of Professional Affairs)
Interests: East Asian culture, traveling, fitness, golf, basketball, EDM and pop music, gastronomy, stock trading, real estate