ALEXANDRA (JIAHAO) HUO

jiahao.h@columbia.edu | (424) 302-8221 | Portfolio Website | GitHub | LinkedIn

Data-driven and analytical individual with experience of leveraging **Statistical Modeling**, **Machine Learning methods**, **A/B Testing**, and **Data Visualization tools** to identify actionable insights and suggest business recommendations. Well-versed in communicating key findings to cross-functional groups. Proficient knowledge in **Analytics**, **Statistics**, **Mathematics**, and **Programming**.

EDUCATION

Columbia University: Columbia Engineering & Columbia Business School　　　Aug 2018 – Dec 2019
M.S. Business Analytics, GPA 3.84/4.00　　　New York, NY
- **Teaching Assistant:** Analytics in Python (two semesters)

University of California, Los Angeles (UCLA)　　　Sep 2014 - Mar 2018
B.S. Statistics & B.S. Financial Actuarial Mathematics, GPA 3.91/4.00 SUMMA CUM LAUDE　　　Los Angeles, CA

WORKING & RESEARCH EXPERIENCE

Activision Blizzard　　　May 2019 – Aug 2019
Data Scientist Intern (Player Science A/B Testing, Global Analytics)　　　Los Angeles, CA
- Developed Call of Duty (CoD) engagement metrics to track impact of in-game item releases using SQL and D3 visualization
- Implemented dashboards of weapon performance metrics for game designers to monitor in-game ecosystem with Spark
- Customized statistical models to predict players' survival prospects in game, used by producers to tune game features in CoD
- Performed retrospective cohort analysis on in-game avatar usage in CoD using statistical testing with Spark SQL and Python, and communicated key business insights to project managers and marketing teams in future titles' marketing strategies

OCP Group　　　Nov 2018 – Present
Research Data Scientist (Text Mining and Nature Language Processing)　　　New York, NY
- Utilized information extraction systems to generate relational tuples from fertilizer reports using Python on Google Cloud
- Constructed network graphs to visualize relationships between entities to facilitate text summarization of the reports

WiZR (Video Analytics Platform Startup)　　　Feb 2018 – June 2018
Data Scientist　　　Los Angeles, CA
- Introduced new data preprocessing techniques for Convolutional Neutral Network which boosted detection accuracy by 7.6%
- Built a pipeline to transform raw data into scheduled performance reports on video analytics metrics using SQL and Python

Vale S.A.　　　June 2017 – Aug 2017
Assistant Data Analyst　　　Shanghai, CN
- Assessed impact of demographic trends of China on economy and steel consumption using Python and SQL

UCLA Department of Statistics: NBA Players Analytics　　　Mar 2017 – June 2017
Research Assistant (Sports Analytics)　　　Los Angeles, CA
- Devised a ranking algorithm with self-defined KPIs and distance metrics to evaluate players' performance using R
- Visualized sweet spots and unsweet spots for the NBA Finalists with Spatial Heat Maps using Tableau

PROJECT & LEADERSHIP EXPERIENCE

CreditRiskMonitor Capstone: Identify CAR (Customer-At-Risk)　　　Sep 2019 - Present
Data Scientist　　　New York, NY
- Created a customer churn model using Random Forest, and consulted with account managers on customer retention strategies

The Data Open Datathon: Renewable Energy & Respiratory Disease　　　Feb 2019
Project Manager　　　Princeton, NJ
- Applied statistical tests to examine casual relation between non-renewable energy consumption and asthma prevalence

Game Analytics: Catch Cheaters in game PlayerUnknown's Battlegrounds (PUBG)　　　Dec 2018
Team Lead　　　New York, NY
- Implemented unsupervised anomaly detection algorithms (One-Class SVM and Isolation Forest) to detect cheaters with Python
- Defined metrics to evaluate performance of the unsupervised models, and calculated proportion of cheating population in game

DataFest 2018: Introduce Revenue Maximization Strategy for *Indeed.com*　　　Apr 2018
Best Insight Award | Team Lead　　　Los Angeles, CA
- Delivered a XGBoost regression model to compute business values of job postings to optimize job posting rank and revenue

Kaggle Competition: Predict Average Response Time of the Los Angeles Fire Department　　　Mar 2017 – June 2017
Ranked 3rd/92 teams | Team Lead　　　Los Angeles, CA
- Implemented XGBoost in R, and tuned hyperparameters using random search in parallel with parallelMap

SKILLS & CERTIFICATIONS

- **Technical Skills**: Python, SQL, Spark, Hadoop, Scala, R, Tableau, Git, Google Cloud, Qubole
- **Packages:** Pandas, NumPy, SciPy, StatsModels, scikit-learn, XGBoost, Spark MLlib, Spark SQL, Spark GraphX, Plotly
- **Society of Actuarial Exams**: P: Probability Theory; FM: Financial Mathematics; MFE: Models for Financial Economics