Xueshan (Kevin) Peng | 彭雪山 | ホウセツザン

United States | (858) 531-6701 | kevinpeng2025@gmail.com | Website | LinkedIn | GitHub

Highly strategic **Data Scientist** with 2+ YOE, specialized in **machine learning** and **experimentation**.

EDUCATION

UCLA Anderson School of Management — Los Angeles, CA
M.S. in Business Analytics (STEM OPT) — Expected 12/2025

- VP of Student Onboarding | Stats (Hypothesis Test, Regression), Causal Inference (A/B Testing, Experiment Design), Traditional ML (Supervised, Unsupervised, Ensemble Methods), Deep Learning (CNN, RNN), LLMs (Prompt Engineering, Agentic Workflow), Operation Research, Data Visualization, Data Management (SQL), Customer Analytics (LTV, Churn & CVR)

University of California San Diego — San Diego, CA
B.S. in Economics, Mathematics, Minor in Music — 06/2024

- GPA: 3.9/4.0 | Provost Honors (2020-2024) | Pianist (95JC Jazz Ensemble)

PROFESSIONAL EXPERIENCE

John Burns Research and Consulting | *Data Scientist Intern* | *Irvine, CA* — 06/2025 – 08/2025

- Engineered advanced time-series clustering pipeline (STL, PCA, K-means, Hungarian algo) that enabled cluster-specific labor market insights, enabling clients to make informed product planning decisions amid shifting regional housing market dynamics.
- Built bootstrap-based panel time-series framework to quantify contemporaneous and lagged relationships between housing and labor market metrics, integrating findings into cross-metric forecast validation tool and reducing review time 35+ hours annually.
- Developed vector autoregressive models with impulse response analysis to simulate economic shocks to highly cyclical housing and macroeconomic time series, establishing foundational insights to support researchers in addressing complex client inquiries.
- Optimized XGBoost pipeline by implementing feature selection, hyperparameter tuning, and K-means clustering to segment metros with distinct economic dynamics, training tailored models for each cluster and improving forecast MAE by 15.4%.

UCLA Anderson Center for Impact | *Data Scientist Intern* | *Los Angeles, CA* — 12/2024 – Present

- Implementing ESG data analysis for S&P 500 firms, analyzing 50+ sustainability performance metrics using Python and creating dynamic data visualization tools with Tableau and Python-DASH to support research and sustainability strategy development.
- Deployed a custom LLM pipeline (OpenAI APIs, Python) to extract sustainability metrics, slashing manual validation effort by 80% (from 1,000 to 200 doc reviews) and eliminating the manual workflow post-tuning in favor of AI pipeline maintenance.

Center for Community Energy | *Data Scientist Intern* | *San Diego, CA* — 02/2023 – 06/2023

- Developed regression model to analyze relationship between vehicle usage frequency and various features, such as vehicle age and fuel types, validating employer's hypothesis and improving accuracy of key metric by 9.9%.
- Created Python template for parking analysis and metric optimization, reducing costs by 12% with only 4.8% profit impact.

Sage Automotive Interiors | *Business Analyst Intern* | *Greenville, SC* — 08/2022 – 09/2022

- Collected, cleaned, and analyzed data from 20+ websites, presenting market forecasts and automotive interior design trends through Excel data visualizations, including insights on product placement, customer segmentation, and market share.
- Pioneered electric vehicle market research report, adopted as pivotal tool for informing global business strategies and client presentations (Honda Japan, etc.), influencing global-level decision-making in support of $630M revenue generation in 2022.

DATA ANALYTICS PROJECTS & CHALLENGES[1]

AI-Powered Admissions Chatbot with Automated Knowledge Pipeline

- Built automated RAG pipeline scraping/vectorizing UCLA MSBA pages weekly into Supabase, integrated with Claude Sonnet 4 agent using multi-tool retrieval (vector store, web search) and PostgreSQL memory via n8n webhook.

Driver Incentive Program Evaluation - Causal Inference Analysis

- Assessed impact of ride-share incentive rollout on driver engagement using A/B testing, Difference-in-Differences (DiD), and Instrumental Variables (2SLS), addressing self-selection bias and staggered enrollment in complex experimentation setting.

TECHNICAL SKILLS

Languages: Python (Pandas, NumPy, Scikit-Learn, Gurobi, Matplotlib, Seaborn, PySpark, TensorFlow), SQL, R (ggplot2, dplyr)
Software: Bash/Git, Jupyter, Tableau, Power BI, Azure, AWS & AWS SageMaker, Docker, SAS, Advanced Excel

INTERESTS

- **Pianist**: Classically trained pianist for 15 years who loves jazz. Performed with UCSD's 95JC Jazz Ensemble.
- **Tennis Medalist**: Played tennis for 12 years. Won silver and bronze medals in USTA 16U tournaments.

[1] My full analytics and data science portfolio is available on my website.