Hsiang-Lun (David) Kao

646-750-2493 | davidkao030994@gmail.com | github.com/hfepbook1 | linkedin.com/in/hsianglunkao | New York, NY

Work Experience

IQVIA *Parsippany, NJ*

Data Scientist *Mar. 2020 – Jan. 2025*

- **Executive-Facing Analytics & Dashboarding:** Engineered interactive, executive-facing Power BI dashboards (Python, SQL, R) to consolidate multi-source behavioral and engagement data for our team's flagship product. Delivered real-time insights to senior leadership that increased client engagement by 30 %, shaped feature prioritization, and accelerated data-driven decisions.
- **Predictive & Prescriptive Insights:** Led technical initiatives with product and client services leaders. Translated high-level ideas into detailed requirements and uncovered 5 new, high-impact use cases. Built propensity, cost prediction, and fraud detection models on Snowflake with Python, converting analytics into clear presentations and communicating insights to cross-functional stakeholders.
- **Customer Journey Mapping & Segmentation:** Designed an end-to-end journey analytics pipeline that unified clickstream, CRM, and support-ticket data to analyze every customer touchpoint. Applied K-Means clustering and survival analysis to segment power, core, and at-risk cohorts. Armed Client Services with predictive insights that cut churn by 15% and increased upsell conversion rates by 10%.
- **Advanced Analytics Tools & Market Insights:** Designed and developed 10+ sophisticated analytics tools quantifying competitive mentions, sentiment, and mind-share for senior leaders at global pharma clients, such as Sanofi and Pfizer, to deliver evidence-generation and KOL-influence metrics that sharpened campaign strategy and lifted performance by 30 %.
- **A/B Testing & Feature Optimization:** Led cross-functional teams to design and implement A/B tests on the company's core product, collaborating with Product and Marketing to define hypotheses and roll out data-driven changes, resulting in a 10% increase in saved searches creation and a more optimized user experience.
- **End-to-End Machine Learning Pipeline:** Developed a comprehensive machine learning pipeline for claims data using FastAPI, Scikit-learn (XGBoost), and Streamlit. This internal-facing solution flags potentially fraudulent claims and cuts manual review time by up to 70%. The project encompassed feature engineering, model training, API deployment on Render, and a user-friendly frontend, all managed via GitHub and deployed on Streamlit Cloud.
- **Real-Time Sentiment & Trend Analysis:** Developed and deployed a full-stack Python ETL pipeline on Dataflow that streams EHR data into BigQuery, powering a real-time Looker dashboard for sentiment and trend analysis. The new workflow cut reporting turnaround by 40% and delivered Vertex AI–driven NLP insights to more than 15 clinical and product stakeholders, speeding evidence-based decisions on patient outcomes and emerging trends.
- **ETL Optimization & Workflow Automation:** Developed a self-service Python web app with REST API for live data ingestion, auto-generate Boolean queries and other repetitive tasks. Centralized scattered Jira ticket data into a unified data warehouse, creating a single source of truth that slashed retrieval time and boosted cross-team efficiency.

DIA Associates *New York, NY*

Analyst *Aug. 2019 - Jan. 2020*

- Drove critical decision making by utilizing **Hive SQL**, **Python**, **R** and **Excel** through modeling, profiling and delivering results in a client-ready format in a big data environment.
- Designed and created an enhanced monthly scorecard and measurement frameworks for Content Team with streamlined KPIs for 5 different director teams to articulate the business value and the performance that the team drives and further increased their reporting efficiency by 20%.
- Translated 10 complex sets of customer and marketing data, each contains over billions of data, into meaningful insights; shared learnings to stakeholders at all levels across the organization in an effective way.

NYC Department of Correction *New York, NY*

Data Science Summer Intern *Jun. 2018 - Aug. 2018*

- Developed a multifunctional, interactive **R Shiny** app for **exploratory data analysis (EDA)** and visualization, enhancing data accessibility for non-technical DOC employees and displaying inmate population data.
- Leveraged **machine learning** techniques in **R** to predict acts of violence and assess gang activity across ten DOC facilities, averting an estimated 5-10 potential violent incidents monthly.

Education

Columbia University in the City of New York *New York, NY*

M.S. in Biostatistics *Aug. 2017 - May. 2019*

Taipei Medical University *Taipei, Taiwan*

B.S. in Pharmacy *Sep. 2012 - Jun. 2016*

Technical Skills

Languages & Scripting	Python, SQL (MySQL, MS SQL Server), R, SAS, PowerShell
Cloud & Big Data	Hive, Google BigQuery, AWS, Redshift, Spark, Azure, GCP, Snowflake, Databricks, dbt
Library & Packages	PySpark, TensorFlow, PyTorch, Scipy, Scikit-Learn, Keras, Pandas, Numpy, Seaborn
BI & Visualization Tools	Power BI, Tableau, R Shiny, Looker, Qlik, Business Intelligence
Data Analysis & Modeling	A/B Testing, Prescriptive Analytics, Predictive Modeling, Statistical Modeling, Sentiment Analysis, Causal Inference, LLM