Manye Dong

md4270@cumc.columbia.edu | (+1) 510-365-1468 | New York, NY | linkedin.com/in/manye-dong215/

EDUCATION

Columbia University, New York, NY *Expected Graduation: May 2025*
Master of Science in Biostatistics, Data Science Track
Relevant Coursework: Data Science (R), SQL Programming, SAS, Probability, Statistics

University of California, Berkeley, Berkeley, CA *Aug 2019 - May 2023*
Bachelor of Arts in Data Science (Dean's List), double major in Comparative Literature *GPA: 3.8 / 4.0*

SKILLS SUMMARY

- *Programming Language*: SQL (MySQL, MS SQL Server), Python (Keras, Matplotlib, scikit-learn), R (ggplot2, dplyr)
- *Software*: Tableau, Power BI, Microsoft Office Excel (vlookup, Pivot Tables, VBA, macros), GitHub, Jupyter Notebook
- *Methodology*: A/B testing, Predictive modeling, Exploratory data analysis, Data storytelling, Database management

PROFESSIONAL EXPERIENCE

Joblogic-X Corporation | Service Consulting Firm Plano, TX
Data Analyst Intern *Mar 2023 - Sept 2023*
- Minimized the risk of electricity theft and prevented an estimated revenue loss by 5% by implementing a real-time data system with **SQL** to query over 4M+ rows of electricity usage data in **SAP HANA** to detect stealing-suspect accounts
- Improved 20% data **ETL** and analysis efficiency by designing data models such as stored procedures and events in **MySQL workbench** to automate real-time transaction of electricity user behavior data for 50M+ residential houses
- Reported suspicious customers to the client's teams to collect surcharge or terminate service to prevent future fraud and monitored electricity usage trends and outlier identification in visualization dashboards with **Tableau**
- Performed **ad-hoc analysis** and market research, presenting 30+ **Tableau** reports with actionable insights to cross-functional stakeholders (business operations, leadership) to optimize strategies for business process improvements

WonderCV | YC Accepted Consumer Tech Firm with 5M+ Users Beijing, China
Data Science Analyst Intern *Jun 2023 - Jul 2023*
- Optimized marketing strategies, saved $2K+ in the budget by forecasting user purchases with **logistic regression** developed in **Python** to identify high-intent user groups and communicate insights to marketing team
- Ensured data quality with completeness and consistency by developing reusable **Python** programs to standardize data schema, integrating 150K+ user-uploaded resumes from **MySQL** with corresponding customer data from **Excel**
- Streamlined membership service reporting by crafting and validating 20+ real-time visualization dashboards in **Power BI**, tracking KPIs and reducing daily manual labor by 70% for the business team

FunPins | Social Media Platform Startup Alhambra, CA
Product Operation and Analyst Intern *Jun 2021 - Oct 2021*
- Increased daily click-through rate (CTR) by 40% through co-designing and implementing **A/B testing** using **Python** to test user conversion effects of different banner slogans and providing design team with suggestions for UI changes
- Sustained an average monthly retention rate of 22% for 4 consecutive months by analyzing daily active users (DAU) and engagement times for 1K+ users with **SQL** and providing critical insights to CRM team for customer strategies
- Provided analytical guidance on the mobile app's feature improvement by collecting and transforming data from 300+ user surveys into **Excel** in close collaboration with UX team to gather business requirements

RESEARCH & PROJECT

Cardiovascular Mortality and Air Pollution Causality Modeling Berkeley, CA
Voluntary Data Science Researcher *Jan 2023 - May 2023*
- Investigated the absence of causality between PM2.5 and ozone levels and the crude mortality rate of cardiovascular disease, applying causal inference on three CDC public datasets, each with 100M+ observations
- Predicted the crude mortality rate of cardiovascular disease from PM2.5 and ozone concentration levels by developing various **GLMs** and the **K-NN** method (RMSE between 35 and 40) with **R**