Jason Chen

718-866-7738 | jachen7@bu.edu |Boston, MA| https://jachen2956.github.io/Portfolio/

Education:

Boston University Boston, MA
Bachelor of Science in Data Science Graduation Date: May 2026

- Minor: Business Administration & Management
- Honors: Dean's List (Fall 2022 – Present)

Skills:

Programming Languages: Python, SQL, R, Rust, HTML/CSS/JavaScript
Frameworks & Tools: Git, ArcGIS, Power BI, Tableau, Microsoft Azure, MySQL, Google BigQuery, Docker

Professional Experience:

Boston University, Kolachalama Lab Boston, MA
Machine Learning Research Assistant (UROP) September 2024 – Present

- Engineered a **transformer-based AI model** to analyze real-world EHR data, optimizing masking strategies to address **40% missing data inconsistencies**, achieving **92% accuracy** in early dementia detection.
- Pioneered a dynamic masking framework, testing **30+ feature permutations** and **9 masking strategies**, identifying an optimal masking threshold that boosted model robustness by **15% AUC**.
- Synthesized insights from **10+ machine learning research papers** to design a novel permutation-masking hybrid technique, reducing false positives by **22%** and accelerating inference speed by **1.7x**.

University of Massachusetts Amherst Amherst, MA
Data Science Research Assistant (REU CEET) June 2024 – August 2024

- Developed **scalable scripts** to merge datasets, extract daily files, and impute missing hourly features, accelerating forecast generation by **40%** and enabling real-time analysis for **500+ daily data points**.
- Designed and automated Python pipelines to retrieve and process missing EU regional weather data, achieving data coverage for **40+ US/EU regions** and reducing manual processing time by **35%**.
- Introduced novel evaluation metrics to diagnose model bias, reducing over-prediction in non-renewable forecasts by **25%** and aligning outputs with grid demand patterns.

Authentic Caribbean Foundation Inc Boston, MA
Data Analyst Intern November 2023 – February 2024

- Analyzed **10K+** monthly restaurant transactions using Python and Excel, pinpointing underperforming menu items and driving a **15% sales increase** in Q3 2023 through targeted promotions.
- Mapped U.S. Census income and population data to restaurant locations, identifying 3 high-potential suburban markets that generated **$1.2M in first-year revenue** post-expansion.

Project Experience:

Second Nature (Python, ArcGIS)

- Uncovered 2008 as the most frequent baseline year for carbon reduction goals by analyzing **1,500+ institutional targets**, linking its adoption to post-Kyoto Protocol sustainability frameworks and SIMAP tool standardization.
- Streamlined emissions reporting for 72 institutions by restructuring offset datasets, resolving **100%** of **naming inconsistencies** and filling **85%** of **null "Carbon Neutral"** flags to clarify progress.
- Visualized regional emission hotspots via **ArcGIS spatial analysis**, revealing a **40%** faster progress lag in Southeastern states post-COVID and prompting state-level policy advocacy for renewable energy incentives.

DC City Government (Python, Tableau, Google BigQuery, SQL)

- Deployed **4 interactive Tableau dashboards** featuring 7+ dynamic visualizations and 4 filters, enabling city planners to reduce backlog by **45%** via targeted crew deployments.
- Implemented a shapefile ingestion system to process geospatial data for Gallery Place, reducing missing data errors by **40%** and enabling localized interventions that improved streetlight repair speeds by **22%**
- Automated data extraction by leveraging Python, SQL, and Google BigQuery to process over **500k records** monthly that resulted in **25% boost** in operational efficiency for service request analysis.