Lawrence Liu

lwrncliu@gmail.com | linkedin.com/in/laurieliu | (000) 000-0000

EDUCATION

New York University, College of Arts & Science   **New York, NY**
Bachelor of Arts (BA) in Economics   *Graduated: May 2021*
- Coursework: Data Science, Machine Learning for NLP, Statistics, Econometrics, Web Design

TECHNICAL SKILLS

- Tools: AWS, GCP, SQL Server, PowerBI, Report Builder (SSRS), SSIS, Snowflake, dbt, Dash
- Languages: Python, SQL, JavaScript, HTML/CSS
- Certifications: GCP Data Engineer, AWS Machine Learning Specialty, Snowflake SnowPro, dbt Analytics Engineering

EXPERIENCE

phData – Data Engineering Services   **Remote**
Data Engineer | Python, SQL, AWS   *May 2022 – Present*
- Migrated and refactored machine learning pipelines from Lambda functions into production pipelines using Airflow, Glue, and Redshift, decreasing end-to-end pipeline inference runtime by 80%, and possible failure points by thousands
- Designed Redshift object schemas and implemented Flyway into GitHub actions to version control database migrations
- Continued upskilling on data products, obtaining certifications and writing blogs for company SEO outreach

Epic – Healthcare Software   **Madison, WI**
Analytics Engineer/BI Developer | SQL, Python, JavaScript, HTML/CSS, Report Builder, PowerBI   *June 2021 – May 2022*
- Developed an automated notification framework for the 3500+ Solutions Engineering division allowing for easy project tracking and compliance monitoring for managers
- Directly reported to the SVP, tasked with conducting deep dives into existing processes to deliver data solutions for areas needing improvement, such as automating the annual hiring target calculation for Staffing Leads
- Built out the ETL of the reporting SQL Server database using SSIS packages
- Created pixel-perfect reports and dashboards for customers and internal stakeholders using Report Builder and PowerBI

ERm Research – Market Research   **New York, NY**
Market Research Intern   *Aug 2019 – Mar 2020*
- Conducted the quantitative data analysis of current fan base perceptions including youth programs, associated with the Knicks and Rangers franchises by demographics for the Madison Square Garden Company
- Supported a project with Samsung to measure digital billboard attractiveness in Times Square by collecting field data, hard-coding participants gaze on quarter-second intervals, and creating a heat map of billboard allure
- Coded open-ended qualitative responses from theater screening surveys into discrete categories for analysis
- Proofed for accuracy and coherency of pre-release film performance reports for clients including NEON and A24

PROJECTS

3D Basketball Shot Charts | *Adding another dimension to the traditional 2D shot charts*
- Processed 21 years of NBA data using PySpark, designed table schemas on BigQuery and built interactive data visualizations of field goal attempts using Dash, a python data visualization web framework written on top of Flask, Plotly.js and React
- Re-learned quadratic functions, originally taught in 8th grade, in order to calculate ball flight path coordinates in 3D space

Data Augmentation in Document Understanding Tasks | *Python, HuggingFace*
- Implemented data augmentation onto the optical character recognition (OCR) output of restaurant receipts to create a larger sample size when fine-tuning the LayoutLM machine learning model to build model resiliency and improve label prediction
- Found that data augmentation of layout-rich documents improved a model's F1 score for small training sizes
- Challenges include error handling of faulty ground-truth labels provided with the receipt dataset and integrating the pre-trained model into the data augmentation pipeline

PERSONAL INTERESTS

- I enjoy playing tennis and running moderate distances