Max Ma

mm2559@cornell.edu • Max-Ma- • MaxM2559 • Personal Website

EDUCATION

Cornell University, College of Arts and Sciences, Ithaca, NY
Bachelor of Arts in Economics and Information Science *2021-2025*

Relevant coursework: Pract Tools for Operations Research • Object Oriented Programming • Data Structures & Functional Programming • Probability Models • Applied Econometrics • Language & Information • Data-Driven Web Applications • Data Mining & Machine Learning • Spreadsheet Modeling • Interactive Info Visualization • Intro to Data Science • Networks

SKILLS

Programming Languages: Python (NumPy, Pandas, Scikit-Learn, NLTK, OpenPyXl, Matplotlib, pdfplumber, Statsmodels), **R** (ggplot2, Tidyverse), **Javascript** (D3.js, Leaflet), Stata, Java, OCaml, HTML-CSS, Git

Data Analysis/Visualization: SQL, Excel, Stata, Power BI, Tableau, JupyterLab, Google Colab, LaTeX

PROFESSIONAL EXPERIENCE

NYC Office of Budget and Management - *Data Analyst* *June '25 - current*

- **Created a Python package** to scrape bank earnings reports data, clean said data, and run custom summary statistics. Used libraries like **pdfplumber** and **OpenPyXl** to scrape data from **Excel** and PDF files. Used **Pandas** to help clean/process the data and run data analysis for econometrics analysis. Utilized **Fuzzy Search** for data redundancy.
- Creating a **MS Power App** to display visualizations from daily collections data. Visualizations created and imported from **Power BI** and using **DirectQuery** to pull data directly from SQL servers daily. Includes Time Series graphs, Year over Year (YoY) charts, forecasting, and more.

Nexus Scholars - *Economics Researcher* *May '24 - Aug '24*

- Worked with Cornell University professors on creating standardized metrics for measuring 1000+ college student's comprehension of macroeconomics. Analysis included **Data Collection**, **Data Cleaning**, **Data Discrimination**, and **Linear/Multivariable Regressions** conducted through **STATA** and **R**. Analysis and reports were written with LaTeX
- Collaborated with members of the St. Louis Federal Reserve to develop the first ever standardized college-level Intermediate Macroeconomics curriculum, aligning academic content with the Federal Reserve's current operations
- Gave a presentation on our results to the Nexus Program Scholar donors, which included visualizations made with **Matplotlib, Excel,** and STATA. My team has been asked to present at **CTREE** (Economics Education Conference)

Edugion - *Data Science Intern* *May '22 - Aug '22*

- Developed an interactive web map with 10,000+ **geospatial data** points using **JavaScript** and the **Leaflet API**. Utilized a custom **auto-clustering algorithm** to drastically reduce latency and improve user data exploration. Included a global address search bar, map layer control, and data toggling features. Data gathering/cleaning done in **Python** and **Excel**
- Created a Python script to convert raw sources of 1000+ international industry codes into a unified list enabling users to efficiently translate between dozens of international industry codes. Required preliminary research on a novel topic, Web scraping with **Tidyverse** (in R), and **Documentation** on novel Python functions that talked with **Excel** files (**OpenPyXl**)

Learn2Smart - *Data Science Intern* *Nov '20 - Mar '21*

- Built a Python program that generates personalized SAT study guides and practice exams based on users' previous performance, helping approximately 250+ users improve their overall SAT scores within 4 months of taking assessment
- Collected and cleaned 1,000+ SAT reading passages with **Pandas** and **NLTK**, and effectively categorized each reading passage using tools from **Scikit-Learn** and NLTK such as CountVectorizer and TfidfVectorizer.

PROJECT EXPERIENCE

TED Finds: Online Search Engine for TED Talks *Spring '24*

- Created a Custom Search Engine that recommends TED Talks based on user's queries by analyzing the scripts and comments of thousands of existing TED Talks.

Analysis of Energy Usage in Buildings on Cornell's Campus *Fall '24*

- Collected and analyzed data on energy usage across Cornell's campus buildings with Python, identifying patterns in energy consumption by energy type and building category to uncover key insights into energy-intensive areas of campus

Easy Maps: Easily Create Interactive Cluster Map with Automatic Categories *Summer '23*

- Developed an interactive mapping tool to visualize thousands of geospatial data points with an automatic clustering algorithm, ensuring efficient and user-friendly navigation. Data categories are automatically created and visualized