

# Ying Liu

- Tel: (341)-345-7248
- Email: [ying.liu0331@gmail.com](mailto:ying.liu0331@gmail.com)
- Address: [40 Upper Rock Cir, Rockville, MD, 20850](#)

- Github: <https://github.com/yingliu1206>
- Portfolio: <https://yingliu1206.github.io/portfolio/index.html>

## EDUCATION

<b>Georgetown University</b> <i>Master of Science in Data Science and Analytics, GPA: 4.0/4.0</i>	<b>08/2020 - 05/2022</b> <i>Washington DC, US</i>
<b>Shanghai International Studies University</b> <i>Bachelor of Arts in Business English, GPA: 3.82/4.0</i>	<b>09/2016 - 06/2020</b> <i>Shanghai, China</i>
• <b>Awards:</b> Province-level Merit Graduate (Top 1%), Leadership Excellence Award (Top 5%), Merit Scholarship (Top 5%)	

## EXPERIENCE

<b>Kaiser Permanente</b> <i>Data Analyst, Financial and Operational Department</i> Working Tools: <b>Python, Oracle, SQL Server, SAS, Tableau, Excel</b>	<b>06/2022 - present</b> <i>Rockville, MD</i>
<ul style="list-style-type: none"><li>• Created effective Python programs, SAS programs and macros to meet requests from regional specialty departments and streamline data analysis and reporting processes for internal clients</li><li>• Optimized database triggers and stored procedures for the budgeting tool and collaborated with the software development team to enhance the front-end functionality, improving the efficiency and accuracy of roster and panel size management</li><li>• Conducted data cleaning and exploratory data analysis on a dataset over 1 million records, and designed interactive Tableau dashboards to support regional executive decision-making on service distribution, capital planning, and investment priorities</li></ul>	
<b>Georgetown University: Massive Data Institute</b> <i>Research Assistant, Web-Crawling &amp; Digital Data Collection Team</i> Working Tools: <b>Python, AWS, Git, Docker, MongoDB</b>	<b>09/2021 - 05/2022</b> <i>Washington, DC</i>
<ul style="list-style-type: none"><li>• Designed and implemented three end-to-end pipelines to automate the storage of images, files, and raw text scraped from websites into a MongoDB database on Docker</li><li>• Collaborated with a cross-functional team to design and initiate a large-scale data collection campaign to gather information from 100,000 major U.S. public schools</li><li>• Released an alpha version of Crawl4All, a web interface and application for automated web scraping, and documented the entire development process</li></ul>	
<b>Didi Chuxing Technology Company</b> <i>Data Analyst Intern, International Business Group</i> Working Tools: <b>Python, SQL, Excel, Tableau</b>	<b>04/2021-07/2021</b> <i>Beijing, China</i>
<ul style="list-style-type: none"><li>• Collected and analyzed 30 key performance indicators from 60 million overseas user trip records, identifying abnormal fluctuations and developing strategies to improve user experience</li><li>• Automated manual reporting processes by creating interactive Tableau reports using Python and SQL, reducing labor costs by 75% and improving efficiency</li><li>• Proactively coordinated with product managers across the organization to streamline data-related processes in product experiments, including troubleshooting A/B testing, correcting data inconsistencies, and developing new indicators</li></ul>	

## PROJECTS

<b>Analyzing Factors for Advertisement Click Rate</b> Utilized machine learning and cloud computing to analyze factors that impact the click rate of advertisements	<b>04/2021-05/2021</b>
<ul style="list-style-type: none"><li>• <b>Cloud Service Startup:</b> developed a cloud-based environment for analysis on <b>AWS EMR</b> and <b>PySpark</b></li><li>• <b>Data Preprocessing:</b> processed 70 million data points stored in a S3 bucket, cleaned and merged the data tables using <b>spark.sql</b></li><li>• <b>Feature Engineering:</b> applied Spark's <b>StringIndexer</b> and <b>OneHotEncoder</b> functions to encode the features including title, geo-targeting information, transformed the data by <b>VectorAssembler</b>, and established a pipeline to streamline this process</li><li>• <b>Machine Learning:</b> trained and fine-tuned <b>XGBoost</b> and Random Forest models, achieving an AUC score of 82%</li></ul>	
<b>Obesity Classification and Data Analysis via Machine Learning</b> Conducted data analysis with ensemble methods to identify important factors contributing to obesity	<b>08/2020-12/2020</b>
<ul style="list-style-type: none"><li>• <b>Data Preprocessing:</b> preprocessed data using Pandas and NumPy, encoded categorical features, and normalized the dataset</li><li>• <b>Exploratory Data Analysis:</b> explored target variable distribution and relationship with factors such as age, height, weight, and calorie consumption through data visualizations using <b>Seaborn</b> and <b>Matplotlib</b></li><li>• <b>Machine Learning:</b> trained and fine-tuned a <b>Voting Classifier</b> ensemble of high-performing models including <b>Random Forest</b>, <b>Gradient Boosting</b>, and <b>XGBoost</b>, achieving an accuracy of 95% (41% increase)</li></ul>	

## SKILLS

- Programming Skills: **Python, SQL**
- Big Data Skills: **AWS, Azure, Hadoop, Spark, MongoDB**
- Others: **Excel, Git, Tableau, Docker, Bloomberg**