Eric Son

www.ymericson.com| ym.eric.son@gmail.com | 703-501-8614

EXPERIENCE

Data Engineer | QuestBridge June 2021 – Present
- Migrated data processing and delivery pipeline to GCP using Compute Engine, Cloud Storage, and Cloud SDK
- Evaluated current architecture and developed improved processing through unit testing and multiprocessing
- Developed data pipeline architecture to collect and clean student application data and deliver to clients

Data Engineer Fellow | National Institutes of Health May 2020 – Feb 2021
- Designed Django web APIs to display NIH's $70M healthcare budget and utility consumption data
- Developed data modeling and data warehouse architecture, wrote MySQL queries to migrate 12 years of data
- Automated R Shiny visualization and reporting tool that summarizes energy usage and cost avoidance data

Data Research Assistant | University of Chicago Nov 2019 – July 2020
- Built an automated report pipeline from education survey outcome for a local school with 2,000 students
- Analyzed property tax data using machine learning pipelines to generate 3,000 localized, automated reports
- Standardized reports with Seaborn and Altair visualizations on Cook County property assessments

Strategic Targeting Analyst | EAB Aug 2016 – July 2019
- Key achievement: automated data collection process for 250+ clients by 90% through Python and Selenium, leading to one billion marketing emails being sent earlier than in previous years
- Wrote SQL queries to identify opportunities for 80+ clients, developed ArcGIS and Tableau reports
- Advised feature engineering on 50 schools to develop statistical models and improve marketing performance

PROJECTS

Chicago Streets Stats (Scala, Hive, HBase, Kafka, AWS, Java, Node.js) – Link
Full-stack big data web API that displays Chicago street congestion, crash, and traffic violation data
- Implemented the Lambda Architecture to store historical data in HBase and ingest real-time data into Kafka
- Used S3 to host the static website, EC2 instance to run the application, and CodeDeploy to release

School Metrics Dashboard (D3, JavaScript, HTML/CSS) – Link
Dynamic visualization that shows Chicago Public School performance metrics and demographics
- Interactive dashboard using D3.js to display enrollment, graduation rates, and ethnicity data

City of Chicago Salaries (R, Shiny, JavaScript, Heroku) – Link
Web application that shows annual salaries for Chicago municipal employees
- Used JavaScript and D3 to make graphs, and R to make searchable table of each annual/overtime salaries

EDUCATION

University of Chicago | Chicago, IL 2021
Master of Science, Computational Analysis and Public Policy (CS Department & Harris School)
Coursework: Python Programming , Machine Learning, Data Visualization, Database Systems, Big Data Applications Architecture, Cloud Computing

College of William & Mary | Williamsburg, VA 2015
Bachelor of Arts, Economics and Film/Media Studies

SKILLS

Languages: Python, R, Java, HTML/CSS, JavaScript, Node.js
Frameworks and Databases: Django, Flask, MySQL, NoSQL, PostgreSQL, Slate
Tools: GCP, AWS, Hadoop, MapReduce, Hive, Spark, HBase, Kafka, Thrift
Visualization: D3, Shiny, Seaborn, Altair, Tableau, ArcGIS, Qlik, Power BI