

MATTHEW WATSON

Chicago, IL | (331) 777-0957 | watsonmatt2003@gmail.com | [LinkedIn](#) | [GitHub](#)

EDUCATION

Illinois Institute of Technology

M.A.S. Data Science; B.S. Computer Science; Minor in Information Technology & Management

Chicago, IL

Expected May 2027

- **GPA:** 3.5
- Heald Scholarship Recipient
- **Relevant Courses:** Intro to Machine Learning, Probability & Statistics, Data Structures & Algorithms, Database Organization, Computational Mathematics, Discrete Mathematics, Software Engineering, Regression, Online Social Network Analysis, Coding for Data, Data Mining, Big Data Technologies, Statistical Learning

Mahidol University International College

Study Abroad — Data Mining

Bangkok, TH

Fall 2026

PROFESSIONAL EXPERIENCE

RSM US LLP

Advanced Analytics & Artificial Intelligence Consulting Intern

Chicago, IL

June 2026 – August 2026

- Delivered a runtime speedup of more than 10x by building a Python FP-Growth engine to mine association rules from 22,000+ work orders and 129,000+ service-line records across 1,630 unique parts, paired with LLM classification into 14 fault categories.
- Replaced a crash-prone Excel process with a governed, self-service Power BI reporting layer by authoring a BRD and 30+ KPI catalog and building SQL and Azure Data Factory medallion pipelines.
- Built a 60-action Power Automate pipeline using 5 connectors to validate row counts across 178 tables in a 4,393-entity scope; the firm productized it as a reusable client data-quality solution.

CNO Financial Group

Information Technology Intern (Software Engineering Specialization)

Chicago, IL

May 2025 – August 2025

- Implemented an NLP system to flag insurance risk in real time across 100+ data streams from 16 sources, combining keyword matching, sentiment analysis, and weighted scoring to generate confidence scores.
- Built a Monte Carlo simulation engine using binomial, Poisson, and gamma models to calculate 95% Value at Risk (VaR) and simulate loss frequency-severity distributions for actuarial decision support.
- Built and deployed a full-stack React/FastAPI dashboard for 5-minute real-time data visualization, loss histograms, and scenario management, integrated with secure pipelines and Agile CI/CD workflows.

The Global Career Accelerator

Data Science Extern

Chicago, IL

August 2025 – November 2025

- Built SQL workflows using CRUD operations, CTEs, and multi-table joins to analyze 2M+ records, including datasets of 22,000+ GameJet Users, 23,000+ NBA games, and 4.8M+ TfL journeys.
- Performed OLAP-style aggregations to calculate KPIs, including a 6.76% first-day purchase rate, and analyzed more than 1,900 user sessions.
- Used Python notebooks to clean data, engineer features, and conduct exploratory analysis on College Football, Capital Bikeshare, and Olympic performance datasets, visualizing trends and validating SQL insights.

PROJECTS

Statistical Process Control of Chicago Daily High Temperatures

CHE 560 Semester Project

Chicago, IL

Spring 2026

- Analyzed 75 weekday Chicago/O'Hare high-temperature observations across 15 five-day subgroups using X-bar/R and EWMA control charts to assess process stability.
- Calculated a 46.40°F X-bar centerline with 32.13°F-60.67°F control limits and a 52.27°F R-chart UCL, identifying 6 weekly mean violations while all weekly ranges remained in control.
- Configured EWMA with $\lambda=0.20$ and $L=3$, detecting 19 control-limit violations that quantified sustained seasonal shifts from late-winter cold to April warming.

Credit Card Fraud Detection Using Big Data Machine Learning

Big Data Technologies, CSP554

Chicago, IL

Spring 2026

- Engineered a distributed Spark MLlib fraud-detection pipeline on AWS EMR/S3 for 1,296,675 credit-card transactions with 0.58% fraud, applying class weighting, preprocessing, and feature engineering for rare-event classification.
- Benchmarked 8 models across Spark MLlib, H2O, and TensorFlow/Keras; Spark Random Forest achieved 0.9947 AUC-ROC, 0.8511 AUC-PR, and 95.22% fraud recall.
- Engineered 5+ temporal/geospatial features including age, time variables, cyclical encodings, and Haversine distance; the strongest H2O alternative reached 0.8452 AUC-PR and 97.20% recall.

Crime Analytics Across Chicago

Online Social Network Analysis, CSP579

Chicago, IL

Fall 2025

- Integrated 2021 ACS and City of Chicago crime data in R across 2,311 block groups and all 77 community areas using sf, tidy census, dplyr, and spatial joins.
- Built a queen-contiguity network with 2,311 nodes and 5,862 edges in igraph, measuring an average degree of 5.07 and producing geospatial maps plus community-level metrics.
- Quantified socioeconomic relationships through correlation analysis, identifying crime-rate associations with unemployment ($r=0.75$), poverty ($r=0.72$), renter share ($r=0.46$), and median income ($r=-0.53$).

Linthurst Biomass Regression & Variable Selection

Chicago, IL

Statistical Learning, MATH564

Fall 2025

- Modeled 45 Linthurst observations across 14 soil predictors with OLS and multicollinearity diagnostics, identifying VIFs up to 551 despite an R^2 of 0.807.
- Applied PCA and principal component regression using the first 6 components, reducing multicollinearity while retaining an R^2 of 0.68 and stabilizing coefficient estimates.
- Compared stepwise selection, all 2-predictor subsets, and ridge regression; selected a pH + Na model with $R^2=0.658$, adjusted $R^2=0.642$, VIF ≈ 1.0014 , and maximum condition index 1.0385.

Housing Valuation Machine Learning Project

Chicago, IL

Intro to Machine Learning, CS484

Spring 2025

- Benchmarked Linear Regression, XGBoost, and LightGBM for Boston and Chicago housing valuation; XGBoost reached $R^2=0.8380$ on Boston data and optimized LightGBM reached $R^2=0.7485$ on Chicago data.
- Cleaned Chicago sales data with $1.5\times$ IQR outlier filtering, removing 9,785 observations, and engineered 3 structural features spanning garage size, residence type, and building-to-land area.
- Ran 1,000 Hyperopt Bayesian trials across 9 LightGBM hyperparameters, reducing Chicago RMSE from \$95,052 to \$85,463 ($\sim 10.1\%$) with mean cross-validation $R^2=0.7390$.

Player Force & Velocity Estimation (AIMHI)

Chicago, IL

Interprofessional Project, IPRO497

Fall 2024

- Built a computer-vision pipeline with YOLOv5x, PyTorch, OpenCV Farneback optical flow, Norfair tracking, and per-player Kalman filters to detect athletes and estimate motion from video.
- Converted pixel motion to m/s using 0.006 m/pixel calibration, estimated player mass across a 60-120 kg range, and calculated impact force with a mass-adjusted 1,500 N danger threshold.
- Generated a 20-FPS annotated video plus a 4-column player-level CSV (ID, frame, velocity, force) and per-player time-series plots for downstream impact analysis.

Ticket Management System

Chicago, IL

ITMD 411 Course Project

Spring 2024

- Developed a Java Swing/JDBC ticket-management application backed by MySQL, implementing 2 role-based interfaces and 2 relational tables for user credentials and ticket lifecycle data.
- Built 5 admin ticket workflows - create, view, update, delete, and close - with SQL CRUD operations, date tracking, and role-based login validation.
- Packaged the system as a runnable JAR with automatic schema initialization and 2 default role accounts for repeatable startup and testing.

MLS 2024 Database

Chicago, IL

Database Organization, CS425

Spring 2024

- Designed a MySQL MLS database with 7 relational tables and 136 seeded records spanning players, clubs, managers, matches, referees, stadiums, and player-match statistics.
- Built a Python/mysql.connector CLI with a 7-option interactive menu supporting CRUD operations, table views, data validation, and player/club analytics.
- Implemented OLAP and window-function queries using RANK, CUME_DIST, NTILE(6), GROUP BY ROLLUP, and 2 database views to analyze standings, stadium distributions, and scoring/foul trends.

ACTIVITIES & LEADERSHIP

Illinois Institute of Technology Men's Soccer Team

Chicago, IL

Team Captain; Member of Team Leadership Group

May 2025 – May 2026

Illinois Institute of Technology Student-Athlete Advisory Committee (SAAC)

Chicago, IL

Men's Soccer Team Representative (1/2)

March 2025 – May 2026

SKILLS & INTERESTS

Skills: Programming: Java, Python, SQL, C, R, JavaScript, CSS; **Web:** React/Vite; **Data & Analytics:** MySQL, SQLite, Jupyter Notebook, Pandas, Matplotlib/Seaborn, Power BI, ETL/ELT, DAX, Power Query, Apache Spark, A/B testing, experimental design, hypothesis testing; **Machine Learning:** PyTorch, OpenCV, Sklearn, YOLO, MLflow, model monitoring, cross-validation, TensorFlow, Azure Machine Learning, MLOps, model explainability, feature store; **Cloud & Platforms:** AWS, Microsoft Fabric, Databverse, Dynamics 365 F&O, Power Automate, Azure Data Factory, Docker; **Tools:** GitHub, Microsoft Suite, Visual Studio, Eclipse IDE, Android Studio.

Interests: AI/ML, Data Analytics, Data Visualization, Technology Consulting, Financial Technology, Sports Analytics.