Andrew Cai

M.S. Data Science Candidate at Northeastern University

Foster City, CA 94404 | cai.and@northeastern.edu | (585) 754-7891 | LinkedIn: andrew-cai1225 | GitHub: acai747

Skills & Abilities

Programming & Databases: Python, SQL, R, Snowflake, MySQL, SQL Server, PostgreSQL, Oracle SQL Developer
Statistics & Data Visualization: SAS, Tableau, Power BI, Excel
Development & Collaboration: Visual Studio Code, Google Colab, Git/GitHub, BitBucket
Core Skills: EDA and ETL, Statistical Modeling, A/B Testing, Experimental Design, Supervised/Unsupervised ML, NLP, Project Management, Adaptability, Effective Communication, Stakeholder Management

Career Experience

Data Science Intern | San Francisco International Airport | San Francisco, CA | July 2025 – Present
- Building a forecasting model using XGBoost and pandas that feeds into a Power BI dashboard from passenger and flight data to predict the number of people in the airport, gates, and terminals throughout the day.
- Creating a Tableau expiring contracts dashboard to alert stakeholders on upcoming deadlines and bottlenecks.
- Migrating Power BI and Tableau dashboard data sources from static data to SQL databases for dynamic updates.

Data Science Intern | Northrop Grumman Corporation | Redondo Beach, CA | May 2025 – August 2025
- Developed a clustering NLP algorithm and tool that grouped quality data that streamlined corrective actions.
- Created an automatic ETL script that pulled from multiple sources including SQL Server that fed into a Tableau Dashboard to track effectively project milestones saving 2 hours of daily manual processing.
- Designed a script and executable file for the manufacturing team that compiled inspection data that were in non-standard formats into a structured table and allowed for quick analyses saving weeks of labor.
- Created an automatic script that scanned and cross-checked our planning tool data tables for any potential discrepancies which allowed the database engineers to proactively resolve urgent issues.

Data Analyst Intern | AT&T Corporation | Bothell, WA | June 2024 – August 2024
- Reduced data processing time by 300% and saved over $10K monthly costs by optimizing and automating AT&T project initiative financial spending reports with key table joins.
- Designed several interactive Power BI dashboards of internal financial data from QuickBase, Excel, and Snowflake that provided rapid insights for budget committee meetings.

Program Management Analyst | Intel Corporation | Santa Clara, CA | March 2023 – March 2024
- Achieved 99% reduction in data processing time by optimizing SQL queries in product monitoring dashboards.
- Partnered with data engineers to fix pipelines into dashboards to fix several errors such as character mismatches.

Microelectronic Packaging R&D Engineer | Intel Corporation | Hillsboro, OR | May 2022 – March 2024
- Automated equipment data processing using Python, completing a critical project 1 month ahead of schedule.
- Enabled 50% more proactive issue resolution through SQL-based Power BI monitoring dashboards.

Projects

Data Engineering:
- ***Data Warehouse Construction*** (R, MySQL, XML) – Designed and built a pharmaceutical sales data warehouse, and wrote SQL queries for business intelligence.

Machine Learning:
- ***Lego Set Rating Prediction*** (Python, Scikit-Learn) – Predicted audience reception towards Lego sets based on features such as number of pieces, theme, and pricing.

Education

Northeastern University | M.S. Data Science | GPA: 4.00	December 2025
Case Western Reserve University | M.S. Materials Science and Engineering | GPA: 4.00	August 2020