Gavin Wentzel

(607)592-6171 | gwentzel@binghamton.edu | https://github.com/gwentzel26

EDUCATION

Binghamton University, State University of New York   **Binghamton, NY**
Master of Science in Data Analytics   *June 2025*
GPA: 3.92/4.00

University of Denver   **Denver, CO**
Bachelor of Science in Business Administration   *June 2019*
Cumulative GPA: 3.43/4.00

University of Denver   **Denver, CO**
Certificate in Full Stack Software Development   *January - May 2022*
Course Grade*:* A

TECHNICAL SKILLS

Analytics Languages: Python, R, SQL, SAS, Alteryx
Visualization Tools: Tableau, Power BI, ggplot, Adobe Photoshop
Databases: MongoDB, AWS, DynamoDB, Redshift
Development Languages: Javascript, React, HTML, CSS, Tailwind, Node
Graduate Coursework: Machine Learning, Linear Predictive Modeling, Data Visualization, Large Databases, Optimization
Undergraduate Coursework: Sports Analytics, Digital & SEO Marketing, Sports & Entertainment Marketing
Certifications: DataCamp Tableau and PowerBI, Full Stack Web Development, Google Analytics, Microsoft Excel, Search Engine Optimization, Adobe Photoshop

PROFESSIONAL EXPERIENCE

Broome County Transit   **Binghamton, NY**
Data Engineer   *January - May 2025*
- Developed both SQL and R code to clean and preprocess hundreds of excel files, automate the removal of unnecessary columns, standardize date/time, and assign each file a corresponding bus number, eventually transforming 9 million rows of 15-second interval gps tracking data into a polished 2.3 million row dataset (Full presentation deliverable on Github)
- Utilized ArcGIS to create a color coordinated map of the transit route system to dynamically change each route path thickness by the number of transactions in a visually appealing manner in order to clearly show the most popular bus routes
- Utilized Tableau to create a Ridership Dashboard (Available on the website's Ridership Dashboard tab) to interactively explore ridership trends by route, month/year, day of week, time of day, and stop in order to give Broome County Transit a tool they can use to better understand the demand of their busses among these time segments
- Attended weekly meetings to brainstorm ideas of how to properly collaborate with team members and to determine our next steps

PROJECT EXPERIENCE - Portfolio on Github

SAS Case Report   **Binghamton, NY**
Solo Study   *May 2025*
- Compared performance of decision trees, random forests, neural networks, and logistic regression using SAS Visual Analytics on a marketing dataset to predict customer purchase behavior
- Generated and interpreted lift charts for each model to assess predictive power, helping determine which models best identified high-propensity customers
- Analyzed validation metrics including misclassification rate and ROC curves to recommend the most effective model for future targeting strategies

MLB WAR Predictor   **Binghamton, NY**
Solo Sports Project
December 2024
- Developed a predictive linear regression model in R to calculate WAR (Wins Above Replacement), integrating player performance metrics from the previous 3 seasons.
- The model had an R^2 value of .94 meaning 94% of the variability in WAR could be explained by the model's predictors
- Applied Box-Cox transformation to the predictive linear model in R to satisfy normality and homoscedasticity assumptions, ensuring the model adhered to linear regression requirements.
- Visualized model results and key insights using ggplot2, enabling clear communication of player value to MLB teams for the 2025 season.

Disney Stock Predictor   **Binghamton, NY**
Team Project   *November 2024*
- Preprocessed and analyzed historical stock market data to train models, achieving high accuracy in predicting stock price trends for the next 60 days
- Delivered an in-depth lecture on Recurrent Neural Networks (RNNs), including the architecture and functionality of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models.
- Demonstrated real-world applications of neural networks in financial forecasting, showcasing the advantages and disadvantages of machine learning for time-series data.