Abraham Abate

Data Scientist

Contact

Phone: (712) 749-5301
Email: avram.abate@gmail.com
GitHub: /abraham-22
[Data Science Portfolio](#)

Education

Bellevue University:
Oct.2023 – May 2025, Master of Science Program in Data Science

----

Rutgers University:
Oct.2022 – March 2023, Advanced Certificate Program in Data Science/Analytics
[Digital Badge](#)

SUMMARY

Accomplished and detail-oriented Senior Data Analyst with experience across healthcare and life sciences safety sectors. Proven expertise in data analytics, statistical modeling, and process improvement. Adept at applying machine learning, data engineering, and visualization tools to generate insights that drive clinical and operational decision-making. Strong leadership background with a track record of delivering impactful data solutions, optimizing workflows, and ensuring regulatory compliance in healthcare environments.

RELEVANT EXPERIENCE

IE/ Continuous Improvement Lead, Tyson Foods Inc.
2022 – present

- o Analyze data to identify areas for process improvement and efficiency gains.
- o Use data-driven insights to develop and implement strategies for optimizing business processes.
- o Apply statistical methods to analyze data and identify trends, patterns, and areas for improvement.
- o Track and evaluate the effectiveness of improvement initiatives using data metrics.

TECHNICAL SKILLS

- o Languages: Python, R, SQL, HTML/CSS, JavaScript (Basic)

- o Tools & Platforms: Tableau, Power BI, GitHub, Jupyter, Google Colab, Visual Studio Code, CLI

- o Analytics & ML: A/B Testing, Customer Segmentation, Predictive Modeling, NLP, Regression, Clustering

- o Cloud & Deployment: AWS (EC2, S3), Google Cloud Platform, RESTful APIs

- o Data Engineering: SQL, NoSQL, MongoDB, Postgres, PySpark, ETL pipelines

PROJECT HIGHLIGHTS

COVID-19 Fatality Prediction

- o **Goal:** Predict the fatality rate of COVID-19 using patient-level clinical and demographic data.
- o **Tools**: Python, Pandas, Scikit-learn, TensorFlow, Matplotlib
- o **Approach**: Built and compared machine learning models (Logistic Regression, Random Forest, Neural Networks) to assess prediction accuracy.
- o **Outcome:** Achieved high classification accuracy; valuable insights for resource planning and risk assessment.

Heart Disease Risk Prediction Using R

- o **Goal**: Develop a predictive model to assess individual risk for heart disease based on medical records and lifestyle data.
- o **Tools:** R, ggplot2, caret, glmnet
- o **Approach:** Performed regression analysis and feature selection to identify key risk factors such as cholesterol, age, and blood pressure.
- o **Outcome:** Enabled early intervention strategies through data visualization and model interpretation.

Alzheimer's Disease Detection & Prediction

- o **Goal:** Identify early signs of Alzheimer's disease using cognitive assessment scores and health data.
- o **Tools:** Python, Scikit-learn, Pandas, Seaborn
- o **Approach:** Implemented data preprocessing, feature engineering, and classification models including SVM and Decision Trees.
- o **Outcome:** Improved model performance and interpretability, aiding clinical decision support systems.

Early Detection of High-Risk Diagnoses Leading to Inpatient Readmissions

- **Goal:** Predict inpatient readmission risk using clinical, diagnostic, and demographic data to support early intervention strategies.
- **Tools:** Python, Pandas, Scikit-learn, Matplotlib, Seaborn
- **Approach:** Preprocessed patient data, engineered features, and trained multiple classifiers (Logistic Regression, Random Forest, XGBoost) to identify high-risk cases.
- **Outcome:** Improved readmission prediction accuracy; identified key risk factors.

ADDITIONAL PROJECTS

- **GHG Emissions Clustering:** Conducted unsupervised clustering of vehicle emission profiles to support environmental impact analysis.
- **Insurance Claim Risk Modeling:** Explored healthcare cost anomalies using classification and anomaly detection techniques.
- **White House Visitors & TSA Complaints Analytics**: Visualization-driven insights using Python and public datasets.
- **Credit Risk & Sentiment Classification (Amazon Vine Reviews):** Processed large-scale review data using NLP techniques.
- **Customer Segmentation Using Clustering Algorithms**: Applied K-Means, DBSCAN, and Hierarchical Clustering on behavioral and demographic data to segment users.

LEADERSHIP & BUSINESS IMPACT

- **Strategic Planning & Execution** – Drove analytics initiatives aligned with business goals.
- **Team Collaboration** – Mentored peers and facilitated data-driven decision-making across departments.
- **Process Optimization** – Identified opportunities to improve workflows and customer touchpoints.

References

Available upon request