

# Saba Amin

onetreesaba@gmail.com | (732) 857-9113 | Holmdel, NJ | linkedin.com/in/sabaamin-atta | GitHub: github.com/Wamuza1  
Portfolio: [https://wamuza1.github.io/Saba\\_Portfolio/](https://wamuza1.github.io/Saba_Portfolio/) | Website: <https://unraveldatascience.com/>

---

## SUMMARY

Analytical Data Scientist leveraging extensive academic and bootcamp experience to produce quantitative insights for real-world business problems. Ambitious learner, constantly enrolling in challenging programs to analyze, interpret and present crucial findings from structured, unstructured, and semi-structured data sets of all sizes. Excels in areas such as statistical and predictive analysis using Machine Learning.

---

## TECHNICAL SKILLS

**Languages:** Python, R, JavaScript, HTML, CSS, SQL, NoSQL  
**Applications:** GitHub, MongoDB, PostgreSQL, AWS, Flask, Tableau  
**Tools:** Excel, SQL Alchemy, Databasing, Pandas, Regression, Classification

---

## ACADEMIC PROJECTS

**Rutgers Data Science Certificate** NJ  
*Credit Risk Analysis* | [https://github.com/Wamuza1/Credit\\_Risk\\_Analysis](https://github.com/Wamuza1/Credit_Risk_Analysis) Sept. 2022

Tools: Python, Pandas, NumPy, Supervised Machine Learning

- Evaluated machine learning models such as random forest, cluster centroid, oversampling and undersampling techniques to determine which model was better at predicting credit risk for a class imbalance with bias.
- Selected the easy assemble model to meet business requirements and avoid bias with a 91% recall rate of credit risk.

*Program Evaluation Analysis* | [https://github.com/Wamuza1/Program\\_Evaluation](https://github.com/Wamuza1/Program_Evaluation) Dec. 2022

Tools: Pandas, Tableau, Scikit-learn, Jupyter Notebook

- To determine which program has greater user engagement, a descriptive data analysis and visualizations were performed.
- Performed multiple classification models and suggested stakeholders to conduct A/B testing in order to assess the effectiveness of the various models.

*Big Data* | [https://github.com/Wamuza1/Amazon\\_Vine\\_Analysis](https://github.com/Wamuza1/Amazon_Vine_Analysis) Aug. 2022

Tools: PySpark, GoogleColab, PgAdmin, AWS RDS, S3

- Utilized PySpark, Extract, Transform, and Load (ETL) on S3 to analyze Amazon reviews written for wine products by paid vs. non-paid members to determine if variables influenced rate of positive reviews.
- Stored data with SQL and Pgadmin and loaded it into Google Cloud. Compared reviews and determined the difference in 5-star rating percentage for paid vs. non-paid members (40.61% and 45.34%) was not large enough to correlate to bias.

*Data Visualization* | <https://github.com/Wamuza1/bikeSharing> July 2022

Tools: Tableau, Python, Jupyter Notebook

- Analyzed CitiBike data in Tableau to transplant the idea of a bike-sharing system to a new city. Visualized insights for an investor presentation to demonstrate target customers and peak times for ridership of the bike-sharing program.

*Movie ETL* | [https://github.com/Wamuza1/Movies\\_ETL](https://github.com/Wamuza1/Movies_ETL) June. 2022

Tools: Python, Jupyter Notebook, Pgadmin

- Extracted, loaded the Wikipedia JSON, Kaggle Metadata, and Ratings CSV files into a Pandas Data Frame.
- Transformed, cleaned, and organized the data in structured format before loaded into database.

---

## EDUCATION

**Bachelor of Business Administration, Data Analytics** May 2024  
Rutgers University Camden, NJ

Honors: Dean's List 2x, GPA 3.45

**Rutgers Data Science Certificate** Oct. 2022  
Rutgers University New Brunswick, NJ

24-week intensive program focused on technical programming skills in Excel, Python, R, JavaScript, SQL/ NoSQL Databases, Tableau, Big Data, Neural Networks, and Machine Learning.

**Data Science, Machine Learning and AI Program** April. 2023

30-week rigorous program focused on Data Science and Machine Learning Algorithms, Regression, Classification, Anomaly Detection, Time Series, NLP, and Deep Learning.