

Sukaina D. Alkhaldy

US (United States) citizen +1(517) 706-1081 | alkhal13@msu.edu | www.linkedin.com/in/sukainaalkhaldy

EDUCATION East Lansing MI, Year Senior

Cumulative GPA: GPA 3.557 /4.00, Dean's List Honors (2 consecutive semesters)

Michigan State University | East Lansing, MI | Graduation Day May 2026

Bachelor of Science Double Majoring in Data Science and Statistics. East Lansing High School: Diploma

SKILLS

Programming Languages: Python, R Language, and C++.

Python Libraries: NumPy, Pandas, Scikit-Learn, Seaborn, Matplotlib

Big Data and Other Tools: Jupyter Notebook, GitHub, RStudio, Visual Studio Code, and Pycharm.

Other Skills: Machine Learning, Debugging, Statistical Analysis, Time Management, Teamwork, Data Organization, SQL, Tidy Data, Stacks, Queues, and Deques.

TECHNICAL EXPERIENCE

Forest Fire Analysis in Portugal

January,2024 – April, 2024

Analyzed wildfire spread using data preprocessing and visualization (Seaborn, Matplotlib). Applied Linear Regression and SVM models to predict burned area size. Found temperature and humidity as key drivers, with dataset imbalance affecting accuracy.

Pandemic Impact Analysis

October ,2023- December

2023 Analyzed U.S. employment and GDP trends pre- and post-lockdown and identified significant declines in GDP growth and employment rates. Learned importance of aligning datasets after challenges with mismatched timeframes.

Comparative Analysis of penguin populations

July ,2024 - August ,2024

- Analyzed penguin body mass across three islands using ANOVA and t-tests. And Identified Gentoo penguins on Biscoe Island as a key factor in weight differences. Found male penguins weigh ~683g more, aiding survival in cold climates. Concluded populations were balanced, but Torgersen Island required further research.

Neural Decoding of Identity and Subject Information in Macaque AM Face Patch

Spring ,2025

- Analyzed neural spike recordings from macaque AM face patch (Freiwald & Tsao dataset with 25 identities, 8 orientations, 193 files, 2 subjects). Engineered features from 400 ms post-stimulus activity and applied supervised models including logistic regression, SVC, random forest, and KNN. Achieved about 70% accuracy with random forest for face identity classification and 68% accuracy for subject classification, supported with cross-validation and feature importance visualizations.

NYC Yellow Taxi Fare Analysis

Spring ,2025

- Analyzed 2023 NYC Yellow Taxi dataset 6.9M cleaned records, 19 features to investigate factors influencing fare pricing. Applied multiple linear regression, ANOVA, Tukey HSD, and t-tests to assess effects of trip distance, duration, day of week, and time of day. Found trip distance as the strongest predictor of fare, with weekend night trips unexpectedly showing the lowest average fares.

SpartaHack X – Hypertension Risk Calculator

February,2025

- Developed a logistic regression model in Python to predict hypertension risk based on age, cholesterol, and BMI. Built an interactive GUI for user-friendly risk assessment.

Porto Seguro Safe Driver Prediction

Fall ,2025

- Built a machine learning model to predict insurance claim likelihood using 595k records with severe class imbalance. Applied data cleaning, one-hot encoding, scaling, and evaluated multiple models Logistic Regression, Balanced Random Forest, EasyEnsemble. Found EasyEnsemble performed best (AUC 0.636) and identified key vehicle and demographic features driving claim risk.

CERTIFICATIONS: Advanced Python: Working with Databases| LinkedIn|

Languages: English and Arabic