

SABA SUHAIL

(650)397-0079 | sabasuhail1112@gmail.com | [Github](#) | [LinkedIn](#) | [Portfolio](#) | San Francisco, CA

I'm a data scientist with an engineering background. I believe in drawing insights from data and using statistical principles and sound research design methodology to make critical business decisions. With over seven years' stint in public sector research organizations, I have extensive project management and strategy experience and can communicate well to technical and non-technical audiences alike.

SKILLS

Statistical Methods & Tools: Data Wrangling, Predictive Modeling, Causal Modeling, Stata, MATLAB, Hypothesis Testing

Machine Learning: Linear Regression, Classification(Logistic Regression, KNN, Random Forests, Decision Trees), Ensemble Methods, NLP, Neural Nets, Clustering

Programming Languages & Tools: Python (Numpy, Pandas, Sklearn, Scipy, Keras), Git, SQL, AWS

Data Visualization: Matplotlib, Tableau, Seaborn, Plotly, Excel, Google Sheets

RELEVANT EXPERIENCE

Data Science Immersive Fellow

GENERAL ASSEMBLY | REMOTE | APR. 2022 – JUL.2022

- Completed over 20 labs and 5 projects in a 12 week immersive data science course that included Python programming, exploratory data analysis, statistical modeling, and supervised and unsupervised Machine learning
- Used Natural Language Processing techniques, various classification algorithms and Ensemble Methods to predict the popularity of reddit posts scraped using Reddit API
- Collaborated with a team of fellow data scientists to use machine learning principles to make predictions about the climate awareness of a typical American citizen
- Used Natural Language Processing techniques, classical Machine Learning and deep learning methods to detect whether some text is sarcastic or not.
- **Skills:** Python, SQL, Data Wrangling, Logistic Regression, Random Forest, Neural Networks

Research Assistant

INDIAN INSTITUTE OF MANAGEMENT BANGALORE | AUG.2020- OCT.2021

- Responsible for data analysis, technical reading and presentation for multiple research topics.
- Collected sustainability data for Fortune 500 companies from 1955 to 2019 and GDP of various countries from 1960-2019 ,analyzed this data for adding in MOOC
- Analyzed census data of India for 1909 for female literacy data and microfinance institutions performance data for year 2019 in various districts in India to find causal relationship in various Indian districts for year 2019
- **Skills:** Data Cleaning, Causal Modeling, Visualization, Operation Research, Strategy Management, Econometrics, Innovation Management, Research Design, Stata, Python

Scientist (Data Analysis), Defense Research & Development Organization

DELHI, INDIA | AUG. 2014 – JUL. 2020

- Deployed and tested cryptographic systems, extensive data analysis of bulk data
- Built a new practical hybrid random generator using a low frequency oscillator and ISAAC pseudo RNG, checked the statistical properties of the output in MATLAB
- Developed an inhouse secure and reliable storage system to replace commercial storage devices
- **Skills:** Data Analysis, Statistical Analysis, Data Modeling, Report/Dashboards, SQL, Data Wrangling, Matlab

Research Intern

UNIVERSITY OF WISCONSIN MADISON | MAY. 2013 – JUL. 2013

- Developed an accessible tool for power electronics research by linking together a set of simple electrical components which makes power electronics research safer and more accessible

EDUCATION & CERTIFICATIONS

CERTIFICATE, DATA SCIENCE

GENERAL ASSEMBLY | REMOTE | APR. 2022 – JUL. 2022

PhD, STRATEGIC MANAGEMENT

INDIAN INSTITUTE OF MANAGEMENT BANGALORE | 2020-2021

BACHELOR OF SCIENCE, ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY JODHPUR, INDIA | 2010 - 2014

SIX SIGMA GREEN BELT

INDIAN STATISTICAL INSTITUTE | HYDERABAD, INDIA | 2014

SYSTEMS ENGINEERING PRINCIPLES

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY | 2014