

SUMMARY

Data Scientist with a background in Biology and Pharmacology and over four years of experience as a researcher, interpreting and analyzing data in order to drive departmental and business decisions. Brings a strong quantitative skill set, passion for creating lasting impact, and the ability to build exceptional interpersonal relationships and cross-functional collaborations

CONTACT

✉ sgdas32@gmail.com
🌐 sodas32.github.io/
☎ 901.335.4785
📍 San Francisco, CA 94109
in sonali-dasgupta
🌐 sodas32

EDUCATION

Stanford School of Engineering
Foundations for Data Science Certificate 2020

University of Tennessee Health Science Center
Master of Science Pharmacology 2016

Northwestern University
Bachelor of Arts Biological Sciences (Neurobiology) 2014
Bachelor of Arts Psychology (minor) 2014

SKILLS

LANGUAGES & LIBRARIES: Python, SQL, NumPy, Pandas, Scikit-Learn, Flask, BeautifulSoup, Keras

MACHINE LEARNING: Regression, Classification, Clustering, NLP, Neural Networks, Recommendation Systems

DATA MANAGEMENT: PostgreSQL, MongoDB, AWS, Git

VISUALIZATION: Matplotlib, Seaborn, Tableau

EXPERIENCE

Metis

Data Scientist

San Francisco, CA
Mar. 2020 to Current

- Completed an immersive 12-week accredited data science bootcamp with a focus on developing skills in Python, machine learning, statistical modeling, and data communication
- Developed hands-on, practical experience by completing five full projects from idea inception, data acquisition, and exploratory analysis to implementing machine learning algorithms, visualizing, and presenting results
- See project section for project highlights

Unity Biotechnology

Senior Research Associate-In vivo Pharmacology

South San Francisco, CA
Jan. 2020 to Mar. 2020

- Generated key reports for business development team's partnership efforts and pre-IND filings

Research Associate II-In vivo Pharmacology

June 2018 to Jan. 2020

- Project leader for glaucoma indication and representative in departmental core team
- Analyzed RNA sequencing experiments using R; created visualizations with Genialis to distinguish important biomarkers and biological trends
- Developed internal imaging capabilities with Micron IV system; did image quantification with Fiji & ImageJ
- Trained and mentored four research associates

Research Associate I-In vivo Pharmacology

Sept. 2016 to June 2018

- Conducted statistical analysis and visualization of data using excel and Prism; communicated conclusions to core team
- Managed collaborations with international academic labs and contract research organizations
- Supported three ocular indications by independently designing and executing experiments; composed standard operating procedures for projects
- Developed strong leadership, and communication skills through cross-functional relationships across multiple departments

University of Tennessee Health Science Center

Intern-Department of Preventative Medicine

Memphis, TN
June 2016 to Aug. 2016

- Conducted health outcomes research for diabetic patients using Medical Expenditure Panel Survey
- Used STATA to sort and analyze data assessing the role of provider gender in patient symptom management

Feinberg School of Medicine

Research Assistant-Department of Ophthalmology

Chicago, IL
June 2012 to June 2014

- Analyzed retinal layer data in disease model from stained tissue using ImageJ software and Excel
- Participated and presented in lab meetings and with collaborators to discuss experimental findings

Methodist University Hospital

Scribe-Emergency Department

Aug. 2014 to Aug. 2016

- Responsible for writing physician notes for EMR

PROJECTS

Pneumonia Detection with Deep Learning

Used X-ray image dataset to build a convolutional neural network using Keras library to classify whether image is from a patient with normal lungs or pneumonia.

EdinBuzzed: A Bar Guide

From a large Yelp dataset with over 5 million reviews, took a subset of data representing bars in Edinburgh, UK to analyze text using natural language processing (NLP) libraries and topic modeling. Used flask to create a recommender system that responds to user input and recommends top 5 bars with similar qualities to what user is looking for.

Diagnosing Diabetic Retinopathy

Used a dataset from UC Irvine Machine Learning repository to create a classification model to predict whether a patient has diabetic retinopathy or not. Based of AUC/ROC values, features that had most predictive ability were exudate presence and micro-aneurysms.

Predicting Domestic Gross Revenue for Films

Scraped data for 1000 movies from Box Office Mojo using BeautifulSoup to create a linear regression model. Budget, opening weekend gross, and international gross were the selected features and the target variable, gross domestic revenue (amount of money made after opening weekend) was feature engineered.

Mapping Trends in New York Subway Ridership

Using data from New York City's MTA dataset, conducted exploratory data analyses to create a list of top subway stations to place street team for a tech event.