SWANAND BARSAWADE

619-552-5943 • swanandsb@gmail.com • GitHub • LinkedIn • Portfolio

Skills

- **Programming Languages**: Python, SQL, R, C, C++, Java
- **Machine Learning**: Scikit-learn, NumPy, Sklearn, Pandas, Linear Regression, Logistic Regression, XGBoost, Anomaly Detection, Database Management, Unstructured Data
- **Tools**: Android Studio, VSCode, Jupyter, Git, Google Collab
- **Cloud**: Google Cloud Platform, AWS, QuickSight, Docker
- **Data Visualization**: PySpark, geoPy, Swifter, Matpoltlib, Seaborn, Streamlit, Plotly, Altair Charts, Folium, React

Work Experience

Data Scientist, **Evolution International** – New York, NY 06/2023 to Present

- Designed & implemented a data pipeline to facilitate financial analysis of the organization, leveraging **BigQuery** as the database and **Celonis** as the analysis tool resulting in 31% increase in revenue
- Created automation for the process using **Jenkins** and **Airflow DAGs**, processing a massive 5TB of data daily
- Engineered a solution in Python and R to pre-process data from **Excel** which saved approximately 3 hours of developer's time

Data Science Intern, **Evolution International** – New York, NY 06/2022 to 08/2022

- Utilized **Tableau** software to develop over 47 data visualizations, strategically enhancing data comprehension by 31%.
- Applied supervised and unsupervised learning techniques for actionable insights, boosting revenue by 17% via data-driven strategies.
- Efficiently processed and cleaned 7+ terabytes of data on **AWS** using scripting, resulting in a 41% reduction in data processing time, effectively addressing a critical performance problem while keeping stakeholders informed of the status of the product.

Teaching Assistant, **SDSU** – San Diego 08/2021 to 05/2023

- Researched on Python project to calculate velocity of blood by analyzing blood flow videos in veins by **Image processing** techniques
- Proved with accuracy of 81% that blood flows faster from middle of blood vessel using optical flow algorithm of the **OpenCV library**
- Conducted **ad-hoc statistical analysis** using Python libraries to evaluate the efficacy of current medications in treating pediatric patients with bipolar disorder by analyzing correlations between subcortical brain regions

Data Analyst, **Capgemini Technologies** – Pune, India 07/2020 to 07/2021

- Utilized technical skills to enhance applications, transforming, validating, and loading big data into storage systems using the **ETL** tool **Fivetran**. This written solution optimized data processes.
- Demonstrated leadership by collaborating with the testing team, overseeing end-to-end testing, and implementing process improvements in applications. Standardized coding practices and reduced running time by 21.4% through **quantitative analysis**.
- Applied advanced analytics and statistics to create 13 data cubes and **OLAP** models, advancing data analysis by 9%. This self-driven innovation contributed to the development of predictive models for future growth.

Projects

Covid-19 Predictive Data Analytics | *Tableau, Python, R, AWS, Streamlit*

- Analyzed the stagewise pandemic data for the USA in Python/R using linear regression and spark on parameters like age, gender, ethnicity, etc. and visualized the weekly new cases and deaths using Tableau, Streamlit while processing was done on AWS.

Finding Ideal Location for a New Restaurant | *Jupyter Notebook, Machine Learning Models, Google Maps API*

- Scrutinized the Yelp dataset by applying k-means algorithm, VADER lexicon, logistic regression to find parameters that ensure the restaurant will be successful and used Google Maps API to find ideal locations.

Advertisement Recommendation System | *HTML, CSS, JavaScript, PHP, MySQL, SQL*

- Created a website to centralize information about advertisement platforms with their charges sourced from the users resulting in 53% increase in efficiency and accuracy of data storage and extraction as well as designed a SQL database for system.

Education

Master of Science: Computer Science, San Diego State University, San Diego, CA (GPA: 3.8) 08/2021 to 06/2023

Coursework: Data Science, Cloud Computing, Algorithm Analysis, Data Management, Computer Vision, Artificial Intelligence

Bachelor of Engineering: Computer Science, University of Pune, India (GPA: 3.9) 08/2016 to 06/2020

Certifications

- AWS (Amazon), Machine Learning (IBM), Core Java (Zensar Technologies), Data Science (Udemy), C/C++ , PHP, SQL (IIT Bombay)