Swanand Barsawade
Data Scientist

Jersey City, New Jersey • 619-552-5943 • swanandsb@gmail.com

SUMMARY

With 5+ years of experience in data science, machine learning, and data mining, complemented by a strong command of Excel for managing structured and unstructured datasets, I bring expertise to any project. Proficient in Agile and Waterfall methodologies, I seamlessly adapt to varying project requirements, ensuring successful outcomes. My skill set includes advanced knowledge in hypothesis testing, regression analysis, clustering, classification techniques, decision trees, SVM, and time series analysis, empowering precise predictive modeling. Moreover, I excel in crafting impactful visualizations through Tableau and Power BI, enhancing data interpretation and presentation. Additionally, I excel in managing and optimizing MySQL and MS SQL Server databases, streamlining data retrieval for efficiency.

SKILLS

Methodologies: SDLC, Agile, Waterfall
Programming Language: Python, R, SQL
Natural Language Processing: NLTK, SpaCy
Visualization Tools: Tableau, Power BI, Microsoft Excel
Data Science and Machine Learning: Hypothesis Testing, Classification Techniques, A/B testing, Regression, Decision Trees, Support Vector Machines (SVM), ETL (Extract, Transform, Load), Neural Networks, Time-series analysis, Unsupervised Clustering, Exploratory Data Analysis (EDA)

Database Management: MySQL, SQL Server
Cloud Platforms: AWS, Azure, Google Cloud Platform, QuickSight
Version Control: Git, GitHub
Collaborative Tools: Jupyter Notebooks, Trello
Communication Skills: Data Storytelling, Visualization, Technical Documentation
Packages: NumPy, Pandas, Matplotlib, SciPy, ggplot2, Seaborn, Plotly, Scikit-learn, TensorFlow, PyTorch, XGBoost

WORK EXPERIENCE

Process Intelligence, Consultant | Ashling Partners, Chicago, IL   *Sep 2024 - Present*
- Led successful process mining deployments, developing, documenting, validating, and streamlining **91TB** of data models and analyses.
- Managed end-to-end project lifecycles, from kickoff through training, ensuring timely delivery and alignment with customer requirements
- Optimized **$60M in customer savings** by applying **Celonis OCPM** to uncover insights and support strategic decision-making

Analyst | Allure Marketing, NYC, NY   *Jun 2024 - Sep 2024*
- Integrated **Agile practices** within the Waterfall model to enable iteration, cut product cycles by **27.3%**, and accelerate project deliveries.
- Fortified a data strategy on **AWS**, deploying cost-efficient cloud solutions to lower infrastructure costs
- Orchestrated ETL processes by overseeing data extraction and transformation, reducing processing time by **19.4%**

Data Scientist | Evolution International, NYC, NY   *Jul 2023 - Jun 2024*
- Achieved a **31.4%** increase in consumer revenue by designing and implementing a data pipeline, leveraging **BigQuery** as the database and **Celonis** as the analytical tool to facilitate value oriented financial analysis for the organization
- Spearheaded implementation to automate the process by **Jenkins and Airflow DAGs**, processing **5TB** data daily for small enterprise
- Slashed **3 hours** of developer's time by creating a Python and R solution for Excel data pre-processing

Data Science Intern | Evolution International, NYC, NY   *Jun 2022 - Aug 2022*
- Engineered an **ETL pipeline** to streamline data ingestion and transformation processes, enhancing efficiency by **29.6%** through the development of **47 data visualizations** in Tableau.
- Collaborated with **cross-functional teams** to ensure **data integrity** and smooth workflow integration while keeping **stakeholders** informed about the status of the product
- Built an **OLAP predictive model** for forecasting recommendations for **a/b tests** for campaign improvements

Software Engineer - Data | San Diego State University, San Diego, CA   *Aug 2021 - May 2023*
- Led a project for quantifying blood flow velocity by analyzing videos of blood flow in veins using **Image processing** and recognition
- Secured **80.7%** accuracy in proving that blood velocity is higher in the center of blood vessels, leveraging OpenCV's optical flow algorithm.
- Performed **statistical analysis** using linear and logistic regression, implemented data preprocessing and pipelining with **AWS** and **Databricks** for data transformation, utilized Synapse for analytics and Blob storage for data management.

PROJECTS

Advertisement Recommendation System   *Oct 2022 - Dec 2022*
- Developed a centralized website leveraging NLP to consolidate information on advertisement platforms
- Established a centralized data lake on AWS S3, optimizing structured and semi-structured data retrieval time by **25%**
- Accomplished a **53.7%** improvement in efficiency and accuracy of data storage and extraction by gaining insights from users
- Designed and executed a QuickSight dashboard for detailed analysis of video categories and trending metrics
- **Data/Tech Stack:** Python, ETL, AWS s3, Athena, Lambda, QuickSight

Yelp Data Analysis   *Apr 2022 - May 2022*
- Evaluated the Yelp dataset by application of k-means algorithms, VADER lexicon, classification, logistics regression, clustering, and neural networks for identifying metrics and parameters to ensure restaurant success
- Revamped an interactive Looker Studio dashboard to enable in-depth analysis across various parameters
- Implemented use cases utilizing Google Maps API to find ideal locations.
- **Data/Tech Stack:** Python, ETL, mage-ai, Google Cloud Platform, BigQuery, Looker Studio

EDUCATION

Master of Science in Computer Science, San Diego State University, San Diego, CA (GPA: 3.8/4.0)   *Aug 2021 - May 2023*
Bachelor of Engineering in Computer Engineering, University of Pune, Pune, India (GPA: 3.9/4.0)   *Aug 2016 - May 2020*