

Swapnil Deore

315-883-7097 | sdeorework@gmail.com | [linkedin.com/in/swapnildeore](https://www.linkedin.com/in/swapnildeore) | github.com/Swapppyy

SUMMARY

Data Science professional with 3 years of experience in Data Engineering and Analytics, proficient in designing scalable data pipelines, dashboards and ML models. Proficient in Python, SQL, Databricks, PySpark, Tableau, AWS, Snowflake, Power BI and Dbt.

EXPERIENCE

Data Analyst, Syracuse University

September 2024 - Present

- Implementing Databricks PySpark ETL job on Azure Data Lake to process 10 million alumni records into Snowflake, driving 60% faster data processing for dashboard, enhancing networking opportunities between students and alumni.
- Developing Tableau dashboard integrated with Snowflake to evaluate sales trends and restocking cycles for different quarters, facilitating data-driven inventory strategies and contributing to over \$25K in annual cost savings by minimizing excess stock.
- Collaborating with cross-functional teams utilizing Jira and Confluence to gather requirements and create Business and Functional Requirement Documents (BRDs/FRDs), driving delivery of data-driven solutions in agile sprint sprints.
- Leveraging Gitlab for version control and CI/CD pipeline integration for deploying PySpark scripts and Dbt models.

Data Science Intern, Prompt Marketing Agency

June 2024 - August 2024

- Led end-to-end analysis of follower interactions on the Lidl CEO's LinkedIn using SQL and ML models; uncovered peak engagement times and top-performing content themes, leading to 80,000+ increase in profile engagement.
- Performed advanced data modeling in BigQuery by applying CTEs, window functions, and complex data transformations to convert high-volume marketing data into analytics-ready views, powering scalable and efficient social media reporting.
- Built campaign KPI dashboards in Power BI utilizing DAX formulas and calculated columns, streamlining marketing performance reporting and directing 12% improvement in ROI by enabling quicker insights and data-driven decision-making.
- Analyzed over 70K marketing interaction records using Python libraries (NumPy, Pandas, Matplotlib), performing granular EDA surfaced high-performing segments and enabled 15% increase in ad engagement rates.

System Engineer, Tata Consultancy Services

September 2020 - July 2022

- Led User Acceptance Testing (UAT) for FIX API integrations with clearing corporations and various broker firms, configuring message sessions and troubleshooting issues related to latency, network configurations at National Stock Exchange.
- Managed 100 million trades via FIX protocol, ensuring secure transactions and end-to-end data integrity across systems.
- Delivered 30+ ad hoc trading insight reports leveraging Excel (Pivot Tables, VBA, LOOKUPS) and Tableau, accelerating decision-making and strengthening compliance oversight by significantly reducing report turnaround time and improving transparency.
- Devised robust checksum-based data validation mechanism for broker and clearing firm trades, ensuring end-to-end data integrity and 100% accuracy, leading to reduced manual reconciliation efforts by 70%.

EDUCATION

Master of Science in Information Systems

CGPA: 3.8/4.00

May 2024

Syracuse University, Syracuse, NY, USA

- Relevant Coursework: Database Systems, Data Governance, Business Intelligence, Data Visualization, Business Analytics, Data Analysis & Decision Making, Introduction to Data Science, Data Engineering, Data Warehousing, Machine Learning.

Bachelor of Engineering in Computer Science

CGPA: 3.6/4.00

June 2020

University of Mumbai, Mumbai, Maharashtra, India

- Relevant coursework: Data Structures, Analysis of Algorithms, Big Data, Data Warehousing, Computer Networks.
- Publication: Published paper on data mining algorithms, evaluating the Top-K algorithm's efficiency in supermarket sales analysis and demonstrated its superior 2x scalability/resource optimization over the UP-Growth algorithm. [View Paper](#).

PROJECTS

Sentiment Analysis using Databricks | Python, Spark, Databricks, SQL, MLflow

March 2024 - April 2024

- Transformed 200K tweets with PySpark and Delta Lake architecture for real-time analytics.
- Deployed pre-trained HuggingFace LLM model via MLflow for scalable sentiment classification.

Analyzing Subreddits | Airflow, AWS S3, Redshift, Postgres, Git, Tableau

June 2023 - July 2023

- Built real-time data pipeline to process 10K+ Reddit posts, providing structured data to AWS Redshift via AWS S3.
- Uncovered key insights using AWS tools (Athena, Redshift) identifying 42% growth in AI/ML related topics.

SKILLS

- Programming and Analytics: Python, SQL, R, C++, HTML/CSS, NumPy, Pandas, Matplotlib, Seaborn, PySpark, NLP, PyTorch, TensorFlow, Jupyter, PyCharm, Visual Studio, Data Wrangling, Data Mining, Visualization, Statistical Modeling, EDA, A/B Testing
- Data Engineering/Big Data: Airflow, Kafka, Hadoop, Spark, Informatica, Dbt, Kubernetes, EMR, AWS Glue, Azure Data Factory
- Business Intelligence Tools: Tableau, Power BI, Looker, Qlik, Alteryx, GIS, Gitlab, Docker, SDLC, Agile/Scrum, Jenkins, Metabase
- Database and Cloud Platform: Databricks, Snowflake, AWS Redshift, PostgreSQL, MySQL, Google BigQuery, CRM, SAP, AWS (S3, IAM, Glue), GCP, Azure, ERP systems, Salesforce CRM, MongoDB, SAS, SSRS, SSIS, Fivetran, Matillion
- Compliance: GDPR, CCPA, Data Encryption, Data Masking, FINRAA