Swapnil Deore

315-883-7097 | sdeorework@gmail.com | linkedin.com/in/swapnildeore | github.com/Swapppyy

---

SUMMARY

Data Engineering and Analytics professional with 3+ years of experience in building scalable data pipelines, ETL processes, data modeling, tracking KPIs and designing data-driven solutions. Proficient in Python, SQL, Azure Databricks, PySpark, Tableau, AWS, Snowflake, Power BI and dbt with a strong focus on optimizing data workflows and enabling business insights.

EXPERIENCE

Data Analyst, Syracuse University   **September 2024 - Present**

- Engineered ETL pipelines in Azure Databricks using PySpark to process 10M+ alumni records from Azure Data Lake into Azure Synapse Analytics, reducing data latency by 60% and enabling real-time insights for enhanced alumni networking opportunities.
- Monitored and optimized data pipeline performance by reducing Spark job run times by 40% using advanced techniques such as data partitioning, in-memory caching, and SQL query tuning to improve overall processing efficiency.
- Developed Tableau dashboard integrated with Azure SQL Data Warehouse to analyze sales trends and restocking cycles, driving data-driven inventory strategies that resulted in $50K+ annual savings by minimizing excess stock.
- Implemented orchestration workflows using Azure Data Factory to automate multi-source data pipelines, ensuring timely, reliable delivery to downstream analytics systems with 99.9% SLA compliance and improved operational efficiency.

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 Snowflake 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 Tableau utilizing parameters and calculated fields, 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.
- Utilized Dbt Cloud for version control, testing, and deployment of SQL models in Snowflake, improving the efficiency and accuracy of data transformation pipelines across development, staging, and production environments for marketing analytics.

System Engineer, Tata Consultancy Services   **September 2020 - July 2022**

- Led User Acceptance Testing (UAT) for Financial Information Exchange (FIX) API data pipelines, configuring message sessions, troubleshooting latency and network issues for clearing corporations and broker firms at the National Stock Exchange.
- Managed 100 million daily trading transactions, ensuring data integrity and security across ETL systems.
- Created and maintained comprehensive data lineage and metadata documentation to support regulatory audits, ensuring full transparency in data transformations, and meeting compliance requirements for accurate financial reporting and audits.
- Devised robust checksum-based data validation mechanism in ETL pipelines for broker and clearing firm trades, ensuring end-to-end data integrity and 100% accuracy, leading to reduced manual reconciliation efforts by 70%.
- Optimized SQL queries and database performance for large financial datasets in PostgreSQL and Snowflake, reducing query runtime by 30% and improving report generation efficiency for investment managers.

EDUCATION

Master of Science in Information Systems   **CGPA: 3.8/4.00**   **May 2024**
Syracuse University, Syracuse, NY, USA

Bachelor of Engineering in Computer Science   **CGPA: 3.6/4.00**   **June 2020**
University of Mumbai, Mumbai, Maharashtra, India

PROJECTS

Sentiment Analysis using Databricks | Python, Spark, Databricks, SQL   **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, Waterfall, MongoDB, SAS, SSRS, SSIS, Fivetran, Matillion