Swapnil Deore

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

SUMMARY

Data Engineering professional with 3 years of experience in ETL/ELT pipelines, data modeling and business intelligence solutions. Proficient in Python, SQL, PySpark, DBT for efficient data processing with additional expertise in Tableau and Power BI.

SKILLS

Programming Languages: Python, SQL, NoSQL, NumPy, Pandas, Matplotlib, Seaborn, PySpark, Plotly
Data Engineering: Spark, Dbt, Snowflake, Fivetran, Hive, Kafka, Airflow, HDFS, Hadoop, Airflow
Cloud Platforms: Azure (Blob Storage, ADLS, Data Factory, Databricks, Synapse), AWS (S3, Redshift, Glue, Athena, EMR)
Databases: PostgreSQL, MySQL, CRM, SAP, SAS, ERP, MongoDB, Microsoft SQL server, Oracle, SSRS
BI Tools: Tableau, Power BI, Looker, Qlik, Alteryx, GIS, Excel (Power Query, VBA, Pivot Tables, LOOKUPs)
Other: Github, Docker, VS Code, Jupyter Notebooks, PyCharm, MS 365, Visio, Gitlab, Docker, SDLC, Agile/Scrum, Waterfall, Jira

EXPERIENCE

Data Engineer, Syracuse University   September 2024 – Present
- Built Power BI dashboards integrated with Azure Databricks to process TBs of data to analyze supply chain demand trends, driving data-driven inventory strategies resulting in $200K+ annual profit by restocking trending items.
- Configured and maintained Unity Catalog in Databricks to centralize metadata management, enabling fine grained access control, data lineage tracking, and improved data governance across cross-functional analytics teams.
- Created DBT data models in Snowflake to transform raw alumni data into star schema, making it easier to build materialized views / tables for Tableau dashboard supporting alumni engagement and networking initiatives for university students.
- Optimized existing data warehouse structures and implemented data validation tests, reducing data inconsistencies by 25% and improving overall data quality and reliability for business decision-making across various business departments.

Data Science Intern, Prompt Marketing Agency   May 2024 - August 2024
- Conducted conversational analysis of follower interactions on Lidl CEO's LinkedIn using MySQL, Talkwalker and OpenAI Langchain models uncovering top-performing content themes and trending topics, leading to 80,000+ increase in engagement.
- Managed MySQL ETL processes integrating Talkwalker APIs and campaign performance data to monitor competitor activity and influencer trends, enabling real-time dashboards for brand risk alerts and strategic insights increasing share of voice.
- Performed exploratory data analysis (EDA) on 70K+ marketing records utilizing Python (NumPy, Pandas, Matplotlib) identifying top-performing channels leading to 15% increase in campaign engagement and informed reallocation of ad spend.
- Conducted A/B testing on landing pages, advertisements and marketing emails for nationwide retail client, analyzing open and click-through rates to identify optimal messaging strategies, increasing campaign engagement by 45%.
- Evaluated product features across 50+ locations using statistical analysis in Python, resulting in improved user engagement metrics and data-driven product strategy recommendations for one of my client.

Data Engineer, Tata Consultancy Services   June 2020 - July 2022
- Developed ETL pipelines with Azure Data Factory and PySpark for transforming and loading structured and unstructured financial data into the data lake, processing over 10 TB of raw data daily, improving data processing efficiency.
- Implemented data quality frameworks with Python and SQL, ensuring 95% data accuracy and consistent compliance with GDPR and CCPA data protection standards, maintaining high data integrity and data validity across all pipelines.
- Optimized Financial Information Exchange (FIX) API ETL pipelines for India's National Stock Exchange (NSE), reducing latency from 300ms to 25ms via UAT testing and session tuning, enabling real-time execution of millions of daily trade orders.
- Improved query performance by refactoring complex SQL queries, shuffle operations / joins, indexing critical columns and partitioning large datasets, enhancing real-time analytics efficiency for TBs of financial trading data.
- Automated ETL processes via Apache Airflow, orchestrating data collection, cleaning and integration across 12 business units.

EDUCATION

Master of Science in Information Systems   CGPA: 3.8/4.0   August 2022 - May 2024
Syracuse University

Bachelor of Engineering in Computer Science   CGPA: 3.6/4.0   June 2016 - June 2020
University of Mumbai

PROJECTS

Analyzing F1 Driver Dominance with Databricks and Power BI (Link)   April 2024
- Built an end-to-end data pipeline on Databricks leveraging PySpark, Delta Lake, and Power BI to ingest data from Ergast API, transform and model Formula 1 race data, enabling high-performance data processing.
- Developed Power BI dashboards to visualize driver performance trends, showcasing Lewis Hamilton's dominance since 2012 to 2020, including metrics on yearly wins, total races, and points earned, providing insightful analysis of F1 racing results.

Data-Driven Sales Strategy Dashboard (Link)   January 2024
- Designed interactive Power BI dashboard leveraging manufacturing company's data, supporting executives to monitor regional sales, product performance, and customer profitability, leading to data-driven decisions improving sales strategy.
- Implemented snowflake schema data model (Fact tables, dimension tables) and crafted over 20 DAX measures to deliver advanced insights into revenue trends, high-value customers, and product-level contributions across time and geography.