SHANTANU DAHIPHALE
Data Analyst

Dayton, OH | https://www.linkedin.com/in/shantanu-dahiphale-749190201/ | cdahiphale@gmail.com | +1 (937) 516-7375

SUMMARY

Data Analyst with **~2 years** of combined experience in **SQL**, **Python**, and BI tools (**Power BI**, **Tableau**), delivering measurable business improvements including reducing reporting time **40%**, improving resource allocation **25%**, and cutting classification errors **15%**. Skilled in building automated pipelines, **KPI** dashboards, and applying statistical Modeling to large-scale datasets (800K+ records). Strong foundation in **ETL** workflows, **data governance**, and predictive analytics with proven results in healthcare and business domains. Capstone project: built ML pipeline on SEER cancer data achieving AUC **0.89** for survival prediction.

SKILLS:

- **Programming Language:** Python, R, SQL (MYSQL, PostgreSQL, Redshift, Oracle)
- **BI & Reporting:** Tableau, Power BI, Microsoft Excel (PivotTables, Power Query, VLOOKUP, VBA)
- **Data Analysis & Modeling:** Data Cleaning, Transformation, Statistical Modeling, Forecasting, Ad Hoc Analysis, A/B testing
- **ETL and Warehousing:** ETL Practices (SSIS, Alteryx), Dimensional Modeling, Star/Snowflake Schema
- **Cloud & Tools:** AWS (EC2, S3, RDS), Microsoft Office Suite (Excel, Word, PowerPoint) | Familiar: Azure, GCP, Git, Jira, Confluence
- **Data Governance & Quality:** Data Validation, Data Lineage, Data Quality Checks, Data Stewardship
- **Soft Skills:** Data Storytelling, Cross-Functional Collaboration, Problem-Solving, Stakeholder Communication

EXPERIENCE:

Roesch Library, University of Dayton | *Data Analyst*　　　　　　　　　　　　　　**May 2024 – Jun 2025**

- Managed and maintained 100K+ bibliographic and circulation records in Sierra ILS (**Oracle/SQL**), applying metadata standards, authority control, and compliance audits to ensure catalog integrity and accurate downstream reporting.
- Automated **SQL** pipelines with joins, triggers, and scheduled queries, reducing manual effort and cutting reporting time by **40%**, while increasing reliability and consistency of leadership reports used for operational planning.
- Designed and deployed **Power BI** dashboards (DAX, role-based security, data models) from Sierra circulation logs, providing real-time visibility and driving a **25%** improvement in allocation decisions.
- Designed ETL workflows (**SQL + Power Query + Sierra APIs**) to cleanse, transform, and standardize catalog metadata, reducing classification errors by 15% and ensuring consistent downstream analytics.

Secvera Technologies, Pune, India | *Data Science Intern*　　　　　　　　　　　　**Jan 2023 – Jun 2023**

- Built and tuned **Python ML** models (scikit-learn, Pandas) for churn and sales prediction, improving baseline accuracy by **~15%** and enabling more reliable forecasting.
- Automated **PostgreSQL** queries and developed **Tableau** dashboards to track KPIs such as retention and product usage, reducing recurring manual reporting and supporting weekly stakeholder reviews.
- Contributed to **Spark-based** batch pipelines and applied **regression analysis** / **A-B testing** on marketing data, generating actionable insights that influenced campaign strategy.

MRND LAB PVT LTD, Pune, India | *Java Developer Intern*　　　　　　　　　　　　**Feb 2022 - Apr 2022**

- Optimized **MySQL/PostgreSQL** databases, reducing query latency by **20%** and improving stored data accuracy.
- Built Java (**AWT/Swing**) product showcase with **SQL** backend, cutting product search time by **15%**, improving UI experience.
- Designed **Tableau** dashboards on **AWS** (**EC2/RDS**), improving leadership decision-making efficiency by **30%**.
- Streamlined backend data flows with SQL optimizations, improving query performance by 25% and enhancing system scalability to support **10K+ daily transactions** with greater accuracy and reliability.

EDUCATION:

University of Dayton, OH　　　　　　　　　　　　　　　　　　　　　　　　　　**Sep 2023 - Jun 2025**
Master of Business Analytics　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　GPA - 3.60
SPPU University, India　　　　　　　　　　　　　　　　　　　　　　　　　　　　**Aug 2017 - Jun 2021**
Bachelor of Engineering in Computer Engineering　　　　　　　　　　　　　　　　　　　　　GPA - 3.70

ACADEMIC PROJECTS:

Cancer Survival Prediction Using SEER Data | Capstone Project – University of Dayton　　**Mar 2025 - Jun 2025**
Built ML pipeline (Logistic Regression, Random Forest, LightGBM, CatBoost) with preprocessing on SEER data; achieved AUC 0.89 for survival prediction. Enhanced model robustness with feature engineering and NLP applied to physician notes.

Depression Detection Using AI | Capstone Project – Savitribai Phule University　　　　　**Jan 2023 - Apr 2023**
Designed Built multimodal ML pipeline (facial recognition, NLP sentiment); trained **SVM** and **Random Forest** models with **85–90%** accuracy (precision 0.83, recall 0.81). Validated results with cross-validation to ensure reliability of early detection indicators.