

Siddharth Pandit

Portfolio: siddharthpandit.cv

Github: github.com/siddharthp1997

Email: siddharthpandit@zohomail.in

Mobile: +91-852-6606-666

EDUCATION

- George Mason University** Fairfax, USA
Master's of Science - Data Analytics Engineering; GPA: 3.87 2022 - 2024
Courses: Big Data to Information, Decision Analysis, Visualization for Analytics, Data Mining, Big Data Essentials, Fraud Examination
- RV College Of Engineering** Bangalore, India
Bachelor of Technology - Electronics and Communication; GPA: 6.63 2015 - 2021
Courses: Object Oriented Programming in C++, Graph Theory, Database Management Systems, Cryptography and Network Security

SKILLS SUMMARY

- Programming & Databases:** Python, SQL (MySQL, PostgreSQL), Pandas, NumPy, PySpark
- Data Engineering:** Airflow, dbt, Databricks, Snowflake, AWS Glue, ETL Pipelines, Kafka
- Cloud Platforms:** AWS (S3, Lambda, EC2, DynamoDB), Azure, GCP, Docker, CI/CD, GitHub Actions
- Data Visualization & BI:** Power BI, Tableau, Looker Studio, Streamlit, Plotly
- Machine Learning & AI:** scikit-learn, TensorFlow, PyTorch, SHAP, NLP, Model Deployment
- Tools & Systems:** Linux, Git, RESTful APIs, Bash, JSON, YAML, Jupyter
- Soft Skills:** Leadership, Communication, Collaboration, Problem Solving, Time Management

EXPERIENCE

- Inficloud** Virginia, USA
Backend Developer Intern (Full-time) Jan 2024 - Jan 2025
 - Contributed to AWS-based ETL pipelines processing 500GB+ of multi-source data monthly: Assisted with data preparation and testing to improve pipeline reliability.
 - Supported the development of Tableau and Power BI dashboards: Ensured accurate reporting and faster trend identification for marketing analysts.
 - Collaborated with engineers on RESTful API integration for live data streams: Documented endpoints and validated outputs for business use.
 - Automated repetitive validation tasks in Python: Reduced manual data errors and saved significant team hours.
 - Wrote production-ready Python scripts for automation: Saved 200+ engineer hours annually.
- Aryagami Cloud Services** Hyderabad, India
Programmer (Full-time) July 2021 - July 2022
 - Assisted in designing and operating backend infrastructure for SaaS analytics platform: Contributed to efficiency improvements across the platform.
 - Helped optimize MySQL databases: Enhanced application response times and reliability.
 - Learned and applied Docker containerization: Simplified CI/CD processes under senior engineer guidance.
 - Supported improvements that reduced API response times by 25% (400ms → 300ms): Enabled better scalability and faster performance.
 - Automated basic logging and monitoring tasks: Contributed to faster bug diagnosis and reduced downtime.
- Aryagami Cloud Services** Hyderabad, India
Backend Developer Intern (Full-time) June 2020 - July 2021
 - Supported the development of RESTful APIs: Contributed to a 30% improvement in data retrieval speed for analytics dashboards.
 - Assisted in managing and indexing MySQL relational databases: Ensured smooth integration with front-end systems.
 - Helped automate ETL pipelines for large-scale data transformations: Reduced data preparation time by 20%.
 - Gained experience deploying backend services on AWS and Docker: Improved scalability and uptime of services.
 - Contributed to pre-deployment CI/CD checks: Reduced rollbacks by 25% and improved deployment reliability.
 - Learned SQL performance tuning by refactoring queries and re-indexing large tables: Reduced CPU load by 20%.

PROJECTS

- **Credit Card Fraud Detection (Machine Learning, Data Engineering, Explainable AI):** Developed an end-to-end fraud detection system using PySpark and AWS Glue to process 10M+ transactions. Achieved 92% recall and 90% F1-score with Gradient Boosted Trees (GBTClassifier). Integrated SHAP explainability to interpret model decisions and automated ETL pipelines using Airflow, reducing manual processing by 60%. Designed Power BI dashboards for real-time fraud KPIs. Tech: PySpark, Databricks, Airflow, AWS Glue, SHAP, Power BI, Python. (May '23)
- **Collision Analysis in Manhattan Borough (Big Data Analytics, Visualization, Azure Cloud):** Analyzed NYC collision data (20M+ records) using Azure Databricks, Docker, PostgreSQL, and Apache Spark. Discovered 73% of accidents occurred on Thursdays and Fridays, with 84% between 14:00–16:00. Built Tableau dashboards visualizing temporal and spatial trends, and applied predictive analytics to identify high-risk zones. Tech: Azure Databricks, Apache Spark, Tableau, SQL, Docker, Power BI, Python. (May '23)
- **Personalized Content Recommendation System (Data Pipeline, AWS, Machine Learning):** Spearheaded a personalized movie recommendation engine leveraging AWS Lambda, DynamoDB, and Amazon SageMaker. Implemented collaborative filtering and content-based algorithms in Python with 10% accuracy improvement. Automated ingestion with dbt + Snowflake, and visualized engagement metrics in Looker Studio showing 20% higher session time. Tech: AWS Lambda, SageMaker, DynamoDB, Snowflake, dbt, Looker Studio, Python. (Feb '24)
- **Air Quality Dashboard (Data Pipeline Automation, Streamlit, CI/CD):** Automated real-time air quality tracking using AirVisual API, AWS Lambda, and Python scripts, reducing manual retrieval by 95%. Built an interactive Streamlit dashboard with Plotly visualizations, improving user engagement by 75%. Integrated CI/CD via GitHub Actions and automated alert notifications through AWS SNS. Tech: Streamlit, AWS Lambda, Plotly, GeoPandas, GitHub Actions, Python. (Jun '24)
- **Fentanyl Precursor Chemical Analysis (Data Mining, NLP, Network Analysis):** Led data-driven research to identify synthetic opioid precursor manufacturers and complicit state officials. Extracted and analyzed 10K+ trade records using spaCy, BeautifulSoup, and Neo4j, mapping corporate-government networks. Findings were used in a U.S. House Committee investigation on U.S.-China strategic competition. Tech: Python, NLP (spaCy), Neo4j, Power BI, NetworkX, Web Scraping. (Dec '23)

CERTIFICATIONS

- Tableau Desktop Specialist, Nov 2024
- Post Graduate Certificate in Software Engineering for Data Science, 2022