

AWANTIKA SRIVASTAVA

Data Scientist | Statistical Modeling, Machine Learning

+91-8920482037 | sawantika81@gmail.com | [LINKEDIN](#) | [Github](#)

PROFILE SUMMARY

Associate Data Analyst with **2+ years of hands-on experience** in **data preparation, validation, automation, and reporting** using **Python and SQL**. Strong expertise in **pandas-based data manipulation**, building **reusable automation scripts**, and performing **data quality checks** on large datasets. Experienced in working with cross-functional teams to deliver **clean, reliable, and business-ready data** that supports operational and analytical decision-making. Detail-oriented, process-driven, and comfortable working in fast-paced environments.

CORE TECHNICAL SKILLS

- **Programming & Data Analysis:** Python (Advanced), Numpy, Pandas, C++, SQL, Data Manipulation, Data Transformation, Jupyter Notebook.
- **Statistics & Mathematics:** Statistical Modeling, Hypothesis Testing, Confidence Intervals, A/B Testing, Scenario Analysis.
- **Machine Learning:** Supervised & Unsupervised Learning, Regression, Classification, Clustering, Random Forest, Decision Trees, SVM, KNN, K-Means, XGBoost, Model Evaluation Metrics.
- **Deep Learning :** Neural Networks, CNN, RNN, LSTM, Transformers (BERT), Transfer Learning, Model Fine-tuning, Computer Vision, TensorFlow, PyTorch.
- **Data Preparation & Quality:** Data Cleaning, Data Validation, Quality Checks, Data Readiness, Structured Datasets.
- **Automation & Scripting:** Python Automation Scripts, Reusable Workflows, Scheduled Jobs, Process Optimization.
- **Reporting & Analysis:** Exploratory Data Analysis (EDA), Business Reporting, Insight Generation, Documentation.
- **Tools & Collaboration:** Git, GitHub, Version Control, Cross-functional Collaboration.
- **Professional Skills:** Attention to Detail, Problem Solving, Clear Communication, Requirement Understanding.

EXPERIENCE

Data Scientist / Machine Learning Engineer | PPS International Pvt. Ltd.

January 2024-Present

- Cleaned, validated, and structured **large datasets using Python (pandas)** to meet internal data quality and readiness standards.
- Built and maintained **reusable Python scripts** to automate recurring tasks such as data refreshes, file updates, and preprocessing workflows.
- Managed end-to-end **data transformation pipelines**, ensuring seamless integration with downstream systems.
- Performed routine and ad-hoc **data quality checks**, identifying inconsistencies, missing values, and anomalies.
- Generated **regular business reports** and extracted actionable insights to support operational and analytical teams.
- Created clear **documentation for data processes, scripts, and reports**, improving knowledge sharing and maintainability.
- Collaborated closely with **Data Analysts and Data Scientists** to understand data requirements and support analytics use cases.

PROJECTS

Railway Driver Assistance System (RDAS) | Enterprise ML Project

- **Prepared, cleaned, and validated large volumes of sensor and image-based data to ensure high data quality and readiness.**
- Built Python-based data preprocessing pipelines for **data cleaning, structuring, and transformation.**
- Automated recurring data preparation and validation tasks using reusable Python scripts, reducing manual effort.
- Performed exploratory data analysis (EDA) to identify data inconsistencies, missing values, and anomalies.
- Generated structured datasets and basic reports to support downstream analysis and operational decision-making.

Chatbot Using LLM & RAG | Applied ML Project

- Designed data ingestion workflows to collect and preprocess unstructured documents for NLP-based applications.
- Cleaned, validated, and transformed text data using Python to ensure consistency and reliability.
- Automated document parsing, embedding preparation, and data refresh workflows using reusable scripts.
- Performed exploratory analysis on ingested data to evaluate data coverage and quality.
- Documented data processing steps and maintained clear records to support collaboration and maintenance.

YouTube Comments Sentiment Analyzer | link - <https://youtube-ai-analyzer-ndzqo6r2mepjtsdjmwxl.streamlit.app/>

- Collected large-scale user comment data and performed **data cleaning and validation** using Python and pandas.
- Structured raw text data into analysis-ready formats through preprocessing and transformation workflows.
- Conducted exploratory data analysis to identify sentiment patterns and trends across datasets.
- Automated text preprocessing and dataset preparation using reusable Python scripts.
- Generated structured reports and documented findings to clearly communicate insights to stakeholders.

CERTIFICATION

- IBM Data Science & AI Certification
- AWS Generative AI with Large Language Models
- OpenCV Computer Vision Certification

EDUCATION

IMS Engineering College, Ghaziabad

September - 2020

Bachelor of Technology (Electrical and electronics engineering)