



# MOUSAMI SONI



Daimlerstrasse 169 Berlin, Germany | +4915158852196 | [soni.mousami7@gmail.com](mailto:soni.mousami7@gmail.com) |  
[https://mousami7.github.io/MousamiSoni\\_Portfolio\\_Airflow/](https://mousami7.github.io/MousamiSoni_Portfolio_Airflow/)  
<https://www.linkedin.com/in/mousami7/>  
<https://github.com/Mousami7>

## PROFESSIONAL SUMMARY

Certified **Apache Airflow Professional** and passionate Airflow Developer with 5+ years of experience in Python, SQL, and data workflows, including 2.5+ years of hands-on Apache Airflow. Skilled in **Big Data integration, ETL pipeline design, and workflow orchestration**. Adept at working with **single-node Airflow deployments** for lightweight, production-ready pipelines and integrating cloud services such as AWS S3, DynamoDB, and PostgreSQL. Strong background in **Data Science and ML modules**, supporting model training and automation alongside data engineering tasks. Experienced trainer who has taught Python, SQL, and Airflow fundamentals, and delivered projects ranging from **event data automation (Ceamless)** to **audio data pipelines with ML workflows (Xounds)**.

## SKILLS

**Programming Languages:** Python, SQL.

**Workflow Orchestration:** Apache Airflow (DAG Authoring, scheduling. Operators, sensors, XComs dbt (basic concepts).

**Databases:** PostgreSQL, DynamoDB, MySQL, SQLite, Snowflake (basic concepts).

**Machine Learning & AI Libraries:** Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, SpeechBrain, Whisper, TensorFlow, PyTorch.

**Backend Frameworks:** FastAPI, Flask, Streamlit.

**Cloud & DevOps:** Amazon Web Services (AWS – Lambda, S3, EC2), Git, Docker, RESTful APIs, CI/CD.

**Specialized Expertise:** Audio AI (Speaker Diarization, Abuse Detection), Natural Language Processing (NLP), Predictive Analytics, ETL Pipelines.

## WORK EXPERIENCE

Xounds Berlin, Germany

**Airflow & ML AI Developer**

March 2025 – Present

Dec 2024 – Feb 2025

- **Developed** Apache Airflow workflows on a single-node setup to schedule and manage small-scale audio processing pipelines.
- **Designed and scheduled** workflows to handle and process incoming audio data efficiently.
- **Implemented** data pipelines to collect files from storage, transform/clean them, and prepare them for downstream use.
- **Built** Python-based tasks for organizing and managing incoming data streams.
- **Implemented** basic metadata storage and tracking for audio files in AWS DynamoDB using Python scripts wrapped as Airflow tasks.
- **Developed** DAGs to automatically pick up audio files from Amazon S3, clean/transform them, and store the results back to S3.
- **Configured** monitoring and retries in Airflow to ensure pipeline resilience and detect failures early.
- **Contributed** to backend service development and cloud integrations, ensuring scalability and reliability.
- **Built** scalable backend APIs with FastAPI, AWS Lambda, S3, and EC2, and **documented APIs using Swagger (OpenAPI)** for clear communication with frontend and ML teams.
- **Led** ML feature development for an AI-powered mobile audio app, including abuse detection, speaker diarization, and speaker verification pipelines.
- **Developed** ML data pipelines supporting training and deployment of core audio models.



**Ceamless Events, Berlin, Germany**  
**Airflow/Data Developer**

June 2024 – November 2024

- **Set up** a lightweight Apache Airflow instance (single node, LocalExecutor) to automate event data workflows.
- **Developed** DAGs to ingest attendee data (CSV/Excel files) from Amazon S3 and load it into PostgreSQL.
- **Implemented** FileSensor and S3KeySensor to automatically trigger pipelines on file arrival.
- **Built** Python-based tasks for data cleaning, including duplicate removal, phone number formatting, and name normalization.
- **Scheduled** event-driven jobs to generate summary reports (attendee statistics, ticket sales) and distribute them via email.
- **Collaborated** with cross-functional team members to identify and address key business data needs.
- **Enhanced** teamwork and collaboration skills.

**Yess Infotech (Training Institute) – Pune, India**  
**Data Science & Python Trainer**

Aug 2022 – Aug 2023

- Performed in-house training sessions on **Python, SQL, Data Workflow fundamentals and Data Science** topics.
- Designed and delivered introductory Airflow Sessions, covering
  - DAG creation.
  - Using PythonOperator and BashOperator for small tasks.
  - Introduction to Sensors (file arrival triggers) and XComs.
- Trained learners on best practices for **performance optimization** in data tasks.
- Introduced students to **Apache Airflow** with hands-on examples.
- Guided participants through **POC projects in Airflow** to demonstrate automation.
- Designed project frameworks for ETL, reporting, and basic data workflows.
- Supported learners in building end-to-end mini-projects.
- Created training material, exercises, and evaluations for classroom delivery.

**Vocational Education Project – MP Govt. – India**  
**IT Trainer**

Sep 2016 – Sep 2019

- Conducted training on **Python, SQL, database management, and IT skills** for 200+ students.
- Enhanced analytical and technical skills of learners.
- Supported the **District Magistrate office's IT department** in **election database management**.
- Assisted in maintaining departmental datasets and ensuring data accuracy.

## KEY PROJECTS

- **User Data ETL Pipeline with API & PostgreSQL (POC Project, 2025)**
  - ❖ Designed a **production-style Airflow DAG** using the **TaskFlow API**.
  - ❖ Extracted data from an **external API**, transformed it into a standard format, and loaded it into **PostgreSQL**.
  - ❖ Added **data validation checks**, retry logic, and monitoring to ensure reliability.
  - ❖ Containerized the solution with **Docker & Docker Compose** for easy deployment.
  - ❖ Demonstrated **best practices**: error handling, logging, cleanup, and security (parameterized queries).
- **Audio Abuse Detection & Speaker Diarization (Xounds) –**
  - ❖ Developed workflows in **Apache Airflow** to collect audio files from AWS S3 and process them for further use.
  - ❖ Used **PythonOperator** for audio processing tasks and **custom operators** for integration with **DynamoDB** metadata storage.



- ❖ Deployed everything on a **single-node Airflow instance**, for small workloads.
- ❖ Organized and tracked basic metadata using AWS DynamoDB.
- ❖ Set up simple monitoring with retries and alerts to ensure reliability.
- ❖ Built ML pipelines for real-time audio intelligence; deployed with AWS + FastAPI.
- **AI Event Management System (Ceamless Events)** –
  - ❖ Built Airflow pipelines to process attendee files (CSV/Excel) and load cleaned data into PostgreSQL.
  - ❖ Deployed workflows on a **lightweight single-node Airflow setup**.
  - ❖ Developed Python tasks for data cleaning (removing duplicates, formatting, normalization).
  - ❖ Ran pipelines on a single node Airflow Setup (LocalExecutor) to keep infrastructure lightweight.
- **Introductory Airflow Projects (Yess Infotech)** –
  - ❖ Loading and cleaning CSV data into SQL databases.
  - ❖ Guided learners through **POC projects** in Airflow to demonstrate workflow automation.
  - ❖ Created sample Airflow DAGs as part of Python/SQL training programs.
  - ❖ Taught students the basics of DAGs, Operators, Sensors, XComs, and Scheduling, providing early exposure to workflow orchestration platform.
  - ❖ Classification models to predict student conversions; Flask/Streamlit deployment.

---

## EDUCATION

- 
- **Masters in Computer Science**  
*International University of Applied Sciences Germany* **2022 – 2024**  
*with distinction (75%)*
  - **Master of Technology in Digital communication**  
*Sagar Institute of Research & Technology Indore* **2015-2017**  
*with distinction (76%)*
  - **Bachelor of Engineering in Electronics and Communication**  
*Jawaharlal Institute of Technology Borawan* **2010-2013**  
*with distinction (75%)*

---

## INTERNSHIPS & CERTIFICATIONS

- 
- Astronomer Certification for Apache Airflow 2 Fundamentals **Sept 2025**
  - Data Analytics and Machine Learning | Puneri Pattern Pvt. Ltd. Pune India **Jan 2022 - June 2022**
  - Data Science with Python | SimpliLearn | **Apr 2021 - Sep 2021**

---

## LANGUAGES

- 
- English (proficient) IELTS C1 certified.
  - German (basic conversational) A2.

---

**References can be provided upon request.**

---