

# One-Click ETL Generator - One Pager

## Problem Statement

Building ETL pipelines often requires deep knowledge of SQL, PySpark, and the underlying datasets. Most data practitioners struggle with the steep learning curve, lack of documentation, and time-consuming trial and error. There is a need for an intuitive, self-serve solution that allows users to describe data transformations in plain English and generate valid, optimized code automatically.

## Preliminary Solution

We built a web-based GenAI application where users select datasets from a catalog, describe transformation logic in natural language, and get back:

- Plain-text transformation steps (via LLM reasoning)
- Auto-generated SQL
- Executable PySpark code

It is designed for simplicity, clarity, and speed, enabling even non-technical users to build pipelines.

## GitHub Copilot Usage Scenario

While the LLM-powered backend generates transformation logic dynamically, Copilot assisted during development by:

- Auto-suggesting function structures and Python/Streamlit patterns
- Optimizing regular expression-based validation
- Helping structure reusable generation logic (SQL and PySpark templates)

## Team Roles and Responsibilities

Prajwal Goswami - Team Lead, GenAI Integration Expert, Backend, Logic & Prompting (Major contribution)

Swarali - Team coordination, Meeting facilitation, Testing support

Saloni - UI Development - Streamlit page styling, navigation

Nirupama - UI Development - Dataset catalog UI and interactivity

Divya - Research & Validation of transformation patterns, GenAI prompting logic