DataLegos
CONNECTING DATA, CRAFTING INNOVATION.

Project Streams Overview

Career Catalyst @ DataLegos offers three dynamic streams for project-based experiential learning. Each project is carefully designed around **UI**, **API**, and **Database** layers, anchored by a clear functional problem. Students explore real-world scenarios while building deployable solutions in one of the following three categories:

🔷 1. Data Engineering Stream

The Data Engineering track equips learners with core programming skills, data transformation techniques, and modern data architecture practices. Projects in this stream are rooted in solving data-centric business problems using Python, ETL pipelines, cloud platforms, and distributed processing frameworks. Students gain hands-on experience across ingestion, modeling, and automation workflows.

🔷 2. AI Automation (No-Code / Low-Code) Stream

This stream is tailored for those eager to implement AI-driven automations with minimal coding. Learners explore workflow design, API integration, and AI-enabled tasks using tools like n8n, Zapier, and LangFlow. By the end of the stream, students can independently build and deploy robust, scalable, and intelligent automations for real-world operations.

🔷 3. AI, LLM, RAG & Knowledge Graph Stream

Projects in this stream dive deep into the evolving world of Generative AI. Students learn to work with large language models (LLMs), Retrieval-Augmented Generation (RAG) pipelines, and knowledge graphs. From prompt engineering and vector search to agent orchestration and Neo4j integration, learners build next-generation AI systems with real-world applicability and multilingual intelligence.

⚠️ Note on Technology Stack

Not all technologies and tools listed are implemented in every project.
The **technical stack is carefully chosen** based on the **nature and requirements of the problem being solved**.
Students are exposed to tools and practices **relevant to their assigned project** while gaining foundational knowledge across all domains.

DataLegos
CONNECTING DATA, CRAFTING INNOVATION.

Details of the topics:

DATA ENGINEERING STREAM:

- **Design Thinking**
  Introduction to problem identification, empathy mapping, and iterative solution design.
- **Python Programming Foundation**
  Core programming concepts, syntax, data types, functions, and modules.
- **Data Structures & Algorithms**
  Logic building, time complexity, arrays, lists, stacks, and recursion basics.
- **Problem Solving with Python**
  Hands-on exercises focused on real-world data and automation challenges.
- **Data Engineering:**
  - **Data Engineering Overview** – Understand the role, tools, and industry demand for Data Engineers.
  - **Problem Solving with Data** – Learn to translate business challenges into data engineering solutions.
  - **Data Fundamentals** – Grasp structured, semi-structured, and unstructured data types and workflows.
  - **Python for Data Engineering** – Master Python libraries and best practices for data pipeline development.
  - **Data Extraction and Ingestion** – Ingest data from APIs, files, and databases using ETL/ELT concepts.
  - **Data Cleaning and Transformation** – Clean, reshape, and standardize raw data for analysis or storage.
  - **Storage & Databases** – Understand relational, columnar, and lake-based storage systems.
  - **Data Modeling** – Learn to build star, snowflake, and dimensional models for analytical querying.
  - **Big Data Processing with PySpark** – Process large-scale data using distributed computing with PySpark.
- **Cloud Fundamentals**
  Basics of AWS and Azure; understanding compute, storage, and deployment models.
- **AI/ML Foundations**
  Introduction to supervised, unsupervised learning, and ML model lifecycle.
- **End-to-End ML Project Deployment**
  Building and deploying a machine learning model on the cloud.
- **Edge AI and IoT Basics**
  Understanding sensor data, hardware integration, and deploying AI to the edge.
- **API Design & Integration**
  Working with REST APIs, Webhooks, authentication, and third-party services.
- **UI/UX Basics**
  Building user interfaces, low-fidelity wireframes, and front-end logic.
- **Three-Tier & N-Tier Architectures**
  Concepts of separation of concerns in scalable software systems.

- **Git & DevOps Fundamentals**
  Version control, GitHub workflows, and CI/CD basics.
- **Low-Level System Design**
  Component-level decomposition using mock APIs and flowcharts.
- **Agile and Scrum Practices**
  Daily standups, sprint reviews, retrospectives, and backlog grooming.
- **Data Processing & Automation**
  ETL pipelines using Python (Pandas, NumPy) and tools like n8n.
- **Test-Driven Development (TDD)**
  Writing unit tests and integration tests from the start of development.
- **Continuous Integration & QA**
  Manual and automated test suite creation, debugging, and validation.
- **Documentation & Technical Writing**
  Writing project reports, architecture notes, and user manuals.
- **Presentation and Tech Storytelling**
  Building a portfolio-ready presentation and publishing on LinkedIn/YouTube.
- **Mock Interviews & Career Readiness**
  Resume review, peer interviews, and aligning with real-world job roles.

No-Code / Low-Code:

🤖 AI Automation (No-Code / Low-Code) – Core Topics

- **Introduction to No-Code/Low-Code Platforms**
  Overview of automation platforms, use-cases, and ecosystem.
- **Workflow Design & Automation Thinking**
  Fundamentals of flow-based design and event-driven automation logic.
- **APIs & Webhooks**
  Understanding REST APIs, authentication, and triggering flows with webhooks.
- **Data Handling in Workflows**
  Data formats (JSON, CSV), variable handling, transformation, and parsing.
- **Popular Tools Mastery**
  Hands-on with tools like **n8n, Make (Integromat), Zapier, Power Automate**.
- **Trigger-Action Architecture**
  Designing reactive workflows using triggers, conditions, and branching.
- **AI Integration in Workflows**
  Integrating OpenAI, Hugging Face, Google Vertex AI for intelligent responses.
- **Prompt Engineering Basics**
  Writing structured prompts, chaining calls, and handling responses.
- **Agentic Workflows**
  Designing multi-step workflows with memory and reasoning (LangChain, CrewAI).
- **Form Builders & Input Interfaces**
  Integrating with Typeform, Google Forms, or custom UIs for user input.
- **Document Automation**
  Auto-generating PDFs, spreadsheets, reports using tools like Google Docs/Sheets APIs.

- **Email, Slack, WhatsApp Integration**
  Sending automated messages and alerts across channels.
- **Data Persistence & Storage**
  Connecting with Notion, Google Sheets, Firebase, or SQL for data storage.
- **Custom Code Blocks**
  Adding Python/JavaScript code where no-code logic is insufficient.
- **Error Handling & Logging**
  Setting up try/catch, fallback paths, retry policies, and alerting.
- **Scheduling & Cron Jobs**
  Setting up recurring tasks and time-based triggers.
- **Use of AI Agents & RAG Pipelines**
  Building Retrieval-Augmented Generation workflows with LangGraph or LangFlow.
- **Automation Deployment & Monitoring**
  Deploying automation on n8n cloud/self-hosted, Make or Zapier; monitoring executions.
- **Security, Authentication & Token Management**
  OAuth2, API key handling, data privacy practices.
- **Portfolio Projects & Career Readiness**
  Building end-to-end automations (e.g., newsletter generator, lead tracker) and showcasing them.

AI, LLM, RAG & Knowledge Graph

- **Introduction to LLMs and NLP** – Understand how large language models process and generate human-like text.
- **Prompt Engineering** – Learn to design effective prompts for zero-shot, few-shot, and structured outputs.
- **RAG Fundamentals** – Combine retrieval systems with LLMs to generate accurate, context-rich answers.
- **Embeddings and Vector Stores** – Generate and search embeddings using FAISS, Pinecone, or Weaviate.
- **Document Loaders & Chunking** – Ingest PDFs, websites, and other sources into RAG pipelines.
- **Knowledge Graph Concepts** – Learn the basics of entities, relationships, and semantic modeling.
- **Text-to-Graph Extraction** – Use LLMs and NER to build knowledge graphs from raw text.
- **Graph Database Basics (Neo4j)** – Design and query property graphs using Cypher.
- **Graph + LLM Integration** – Use graphs to enhance LLM responses through GraphRAG.
- **LangChain Tool & Agent Framework** – Orchestrate multi-step workflows using LangChain tools and agents.
- **RAG vs Fine-tuning** – Explore when to use retrieval versus model training for specific applications.
- **Evaluation of LLM Outputs** – Apply human and automated metrics to assess hallucination and accuracy.

- **Security & Prompt Injection Protection** – Ensure safe LLM interactions with input validation and redaction.
- **Deploying LLM Apps** – Use Streamlit, FastAPI, or LangServe to launch AI-driven applications.
- **Multilingual & Multimodal AI** – Build systems that handle Indian languages and multiple content types.
- **No-code LLM Orchestration** – Use tools like LangFlow or CrewAI for visual, low-code automation.
- **Enterprise AI Architectures** – Design knowledge-aware agents for domain-specific use-cases.
- **Storage & Caching in RAG** – Persist data and responses for faster, cheaper future queries.
- **End-to-End Capstone Projects** – Build and deploy real-world apps like AI chatbots or semantic search tools.
- **Career Prep & Tech Branding** – Showcase projects, write posts, and prepare for LLM-centric job roles.