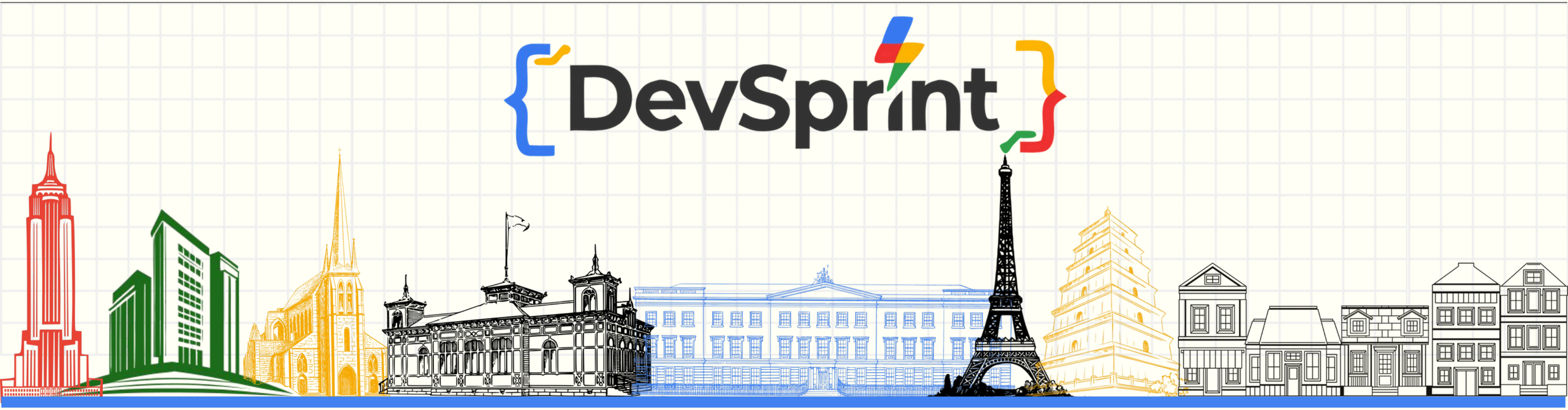
Google Developer Groups
On Campus · MITS-DU Gwalior



Team name: Mind Mesh

Team leader name: Pulatsya Bhagwat

Problem Statement: Open Innovation “Intelligent Data Science Copilot”

Google Developer Groups
On Campus · MITS-DU Gwalior

DevSprint

Brief about solution and problem statement

Problem Statement

Data science workflows are complex and manual, requiring multiple tools, high expertise, and significant time.
This makes it difficult for beginners to start and forces professionals to repeat routine tasks

Solution – Data Science Agent

Data Science Agent is an autonomous AI system that automates the entire data science pipeline using natural language.
Users upload a dataset, describe the goal, and the agent handles data preparation, modeling, and visualization—making data science faster, simpler, and accessible.

Opportunities

a. How different is it from any of the other existing ideas?

Existing Tools	Data Science Agent
Chat-based suggestions only	Autonomous execution
Manual ML pipelines	Fully automated workflows
No memory across steps	Session memory & context awareness
Fixed AutoML pipelines	Dynamic reasoning + tool chaining
Limited explainability	Dedicated reasoning & Business summaries

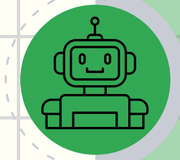
Google Developer Groups
On Campus · MITS-DU Gwalior

DevSprint

b. How will it be able to solve the problem?

Natural Language Input
Users describe goals in plain English

01

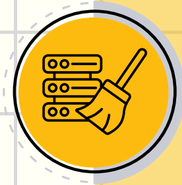


End-to-End Automation
Automates the complete data science workflow

02

Context Awareness
Maintains session memory across steps

03



Smart Data Processing
Handles data cleaning and feature engineering

04

Intelligent Tool Orchestration
Automatically selects and executes the right ML tools

05

Self-Correcting Execution
Detects errors and retries with fixes

06

Google Developer Groups
On Campus · MITS-DU Gwalior

DevSprint

List of features offered by the solution

Intelligent Agent Capabilities

- Dual LLM support (Gemini + Groq)
- Natural language task understanding
- Autonomous tool execution
- Session memory across conversations
- Automatic error recovery
- Built-in code interpreter

Performance & Scalability

- Polars & DuckDB for fast processing
- Token-optimized LLM communication
- SQLite caching for efficiency
- BigQuery integration for large datasets

Multiple Interfaces

- Web UI (React.js)
- Command Line Interface (CLI)
- REST API (Cloud Run ready)
- Python SDK

Complete Data Science Pipeline

- Dataset profiling & quality checks
- Intelligent data cleaning
- Advanced feature engineering
- Model training & ensemble learning
- Hyperparameter tuning (Optuna)
- Explainable AI (SHAP, feature importance)
- Interactive & static visualizations
- Automated EDA reports

Google Developer Groups
On Campus · MITS-DU Gwalior

DevSprint

Google Technologies Used

Google Cloud Run

Deploys the agent as a scalable REST API

Google AI Studio

Used to integrate and experiment with Gemini models for AI reasoning.

Google BigQuery

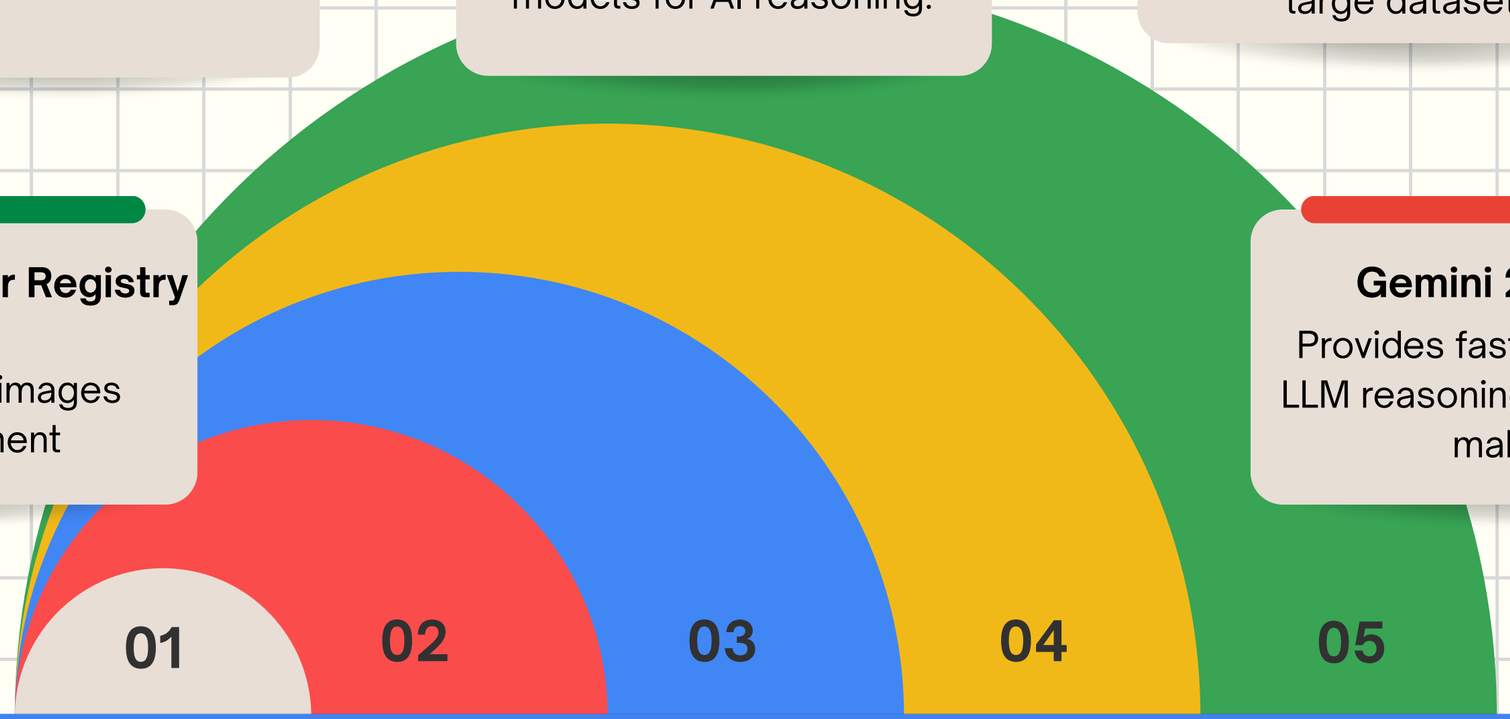
Enables efficient querying, profiling, and analysis of large datasets

Google Container Registry (GCR)

Stores Docker images for deployment

Gemini 2.0 Flash

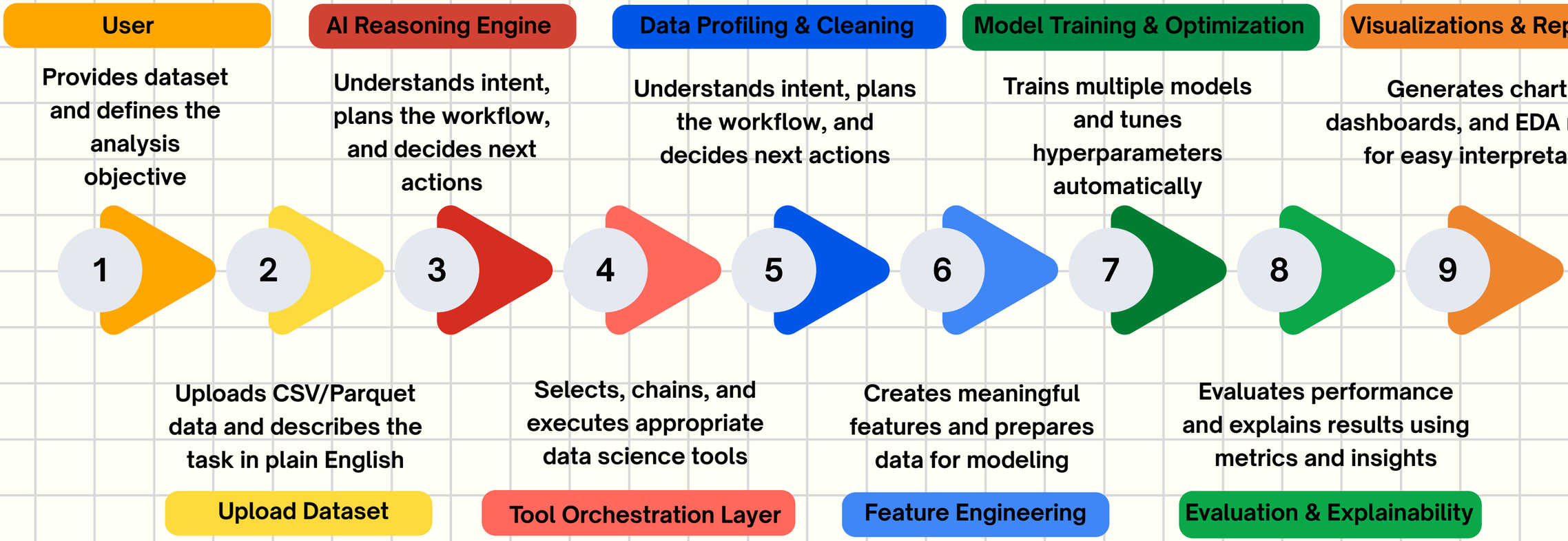
Provides fast, low-latency LLM reasoning for decision-making.



Google Developer Groups
On Campus · MITS-DU Gwalior

DevSprint

Process Flow Diagram

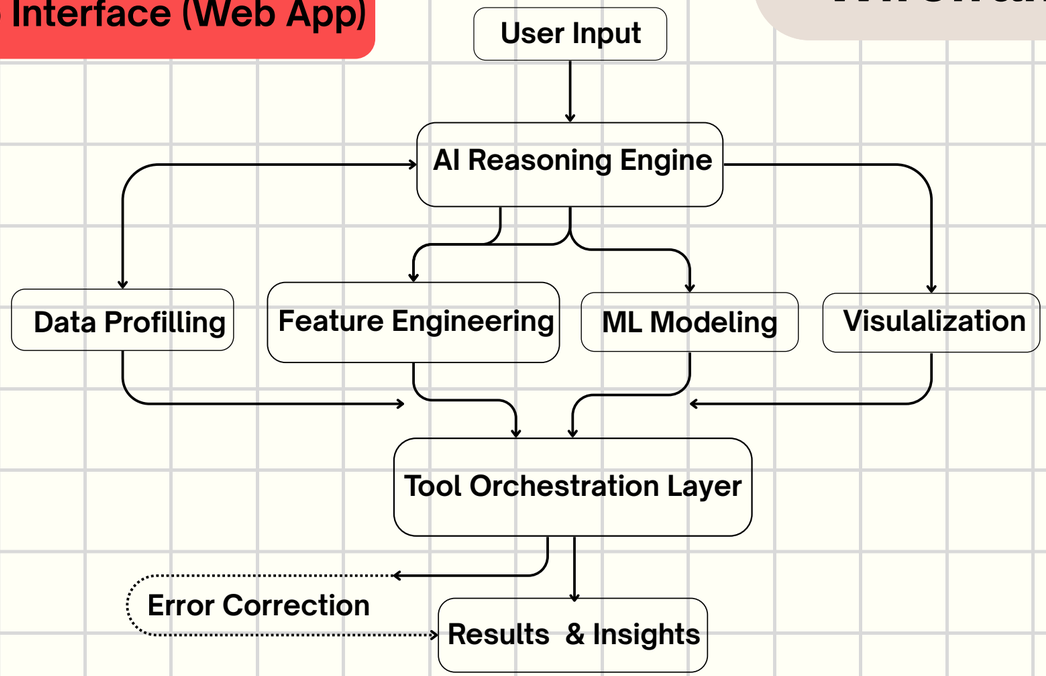


Google Developer Groups
On Campus · MITS-DU Gwalior

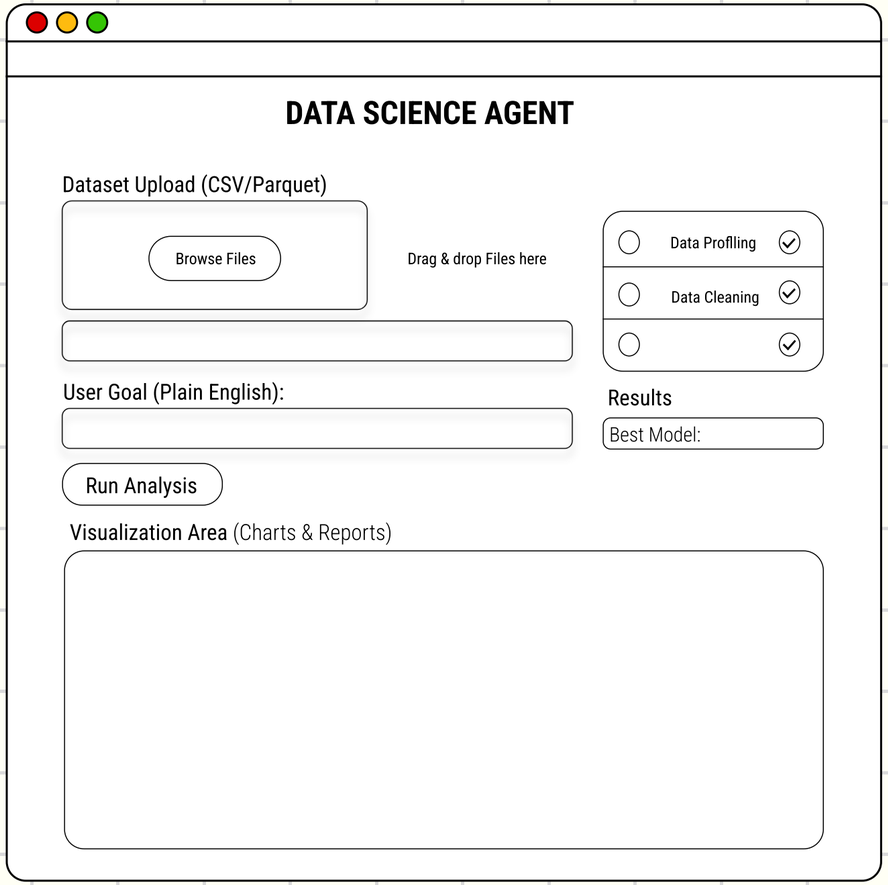
DevSprint

Wireframes / Mock Diagram

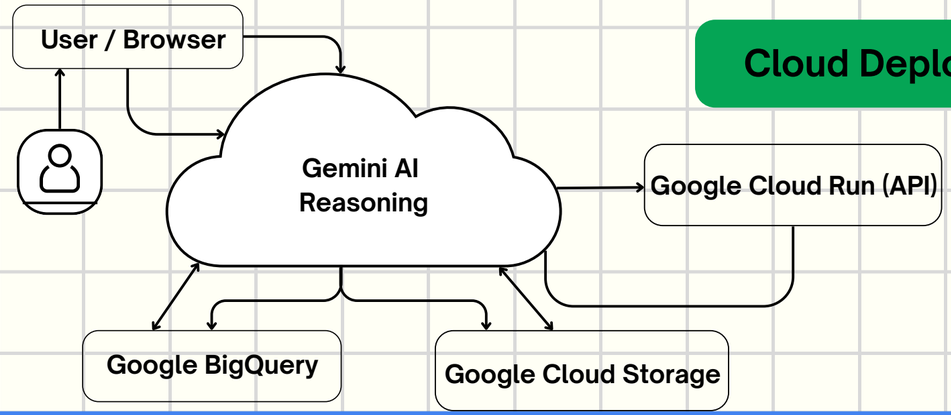
Main Web Interface (Web App)



Backend AI Processing Flow (System View)



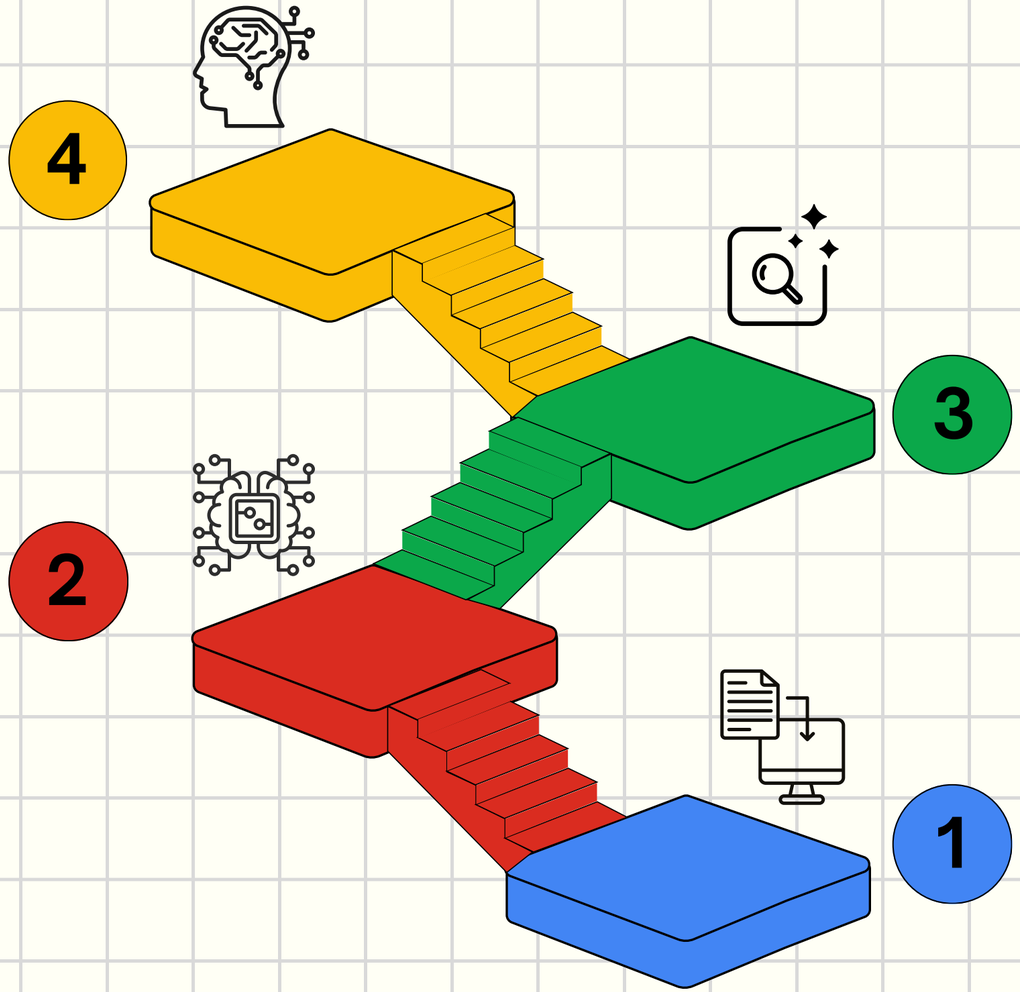
Cloud Deployment Diagram



Google Developer Groups
On Campus · MITS-DU Gwalior

DevSprint

Architecture diagram of the proposed solution



4

The final stage where AI generates and presents meaningful insights to the user.

3

The phase where AI triggers and managed various tools to perform specific tasks.

2

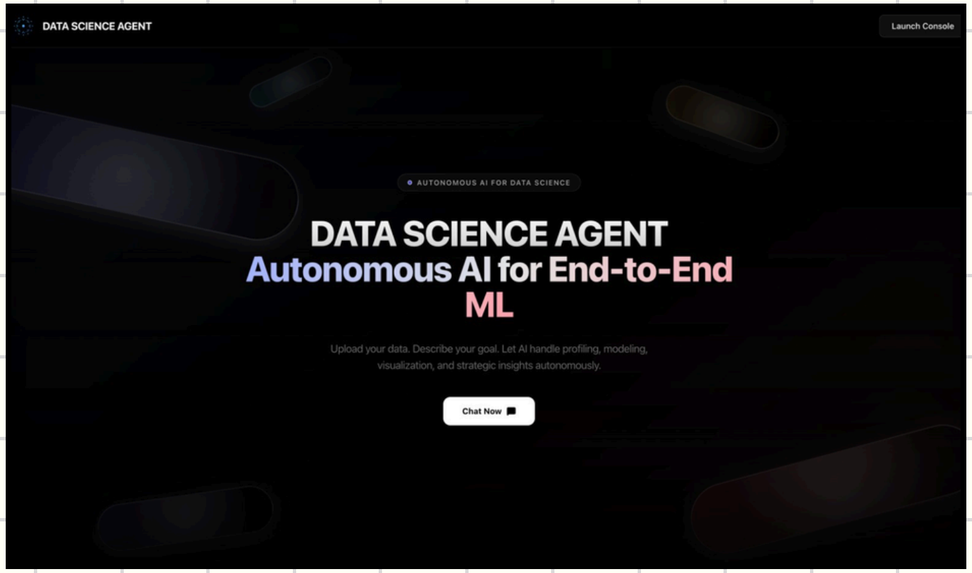
The core of the pipeline where AI processes and analyzes the input data.

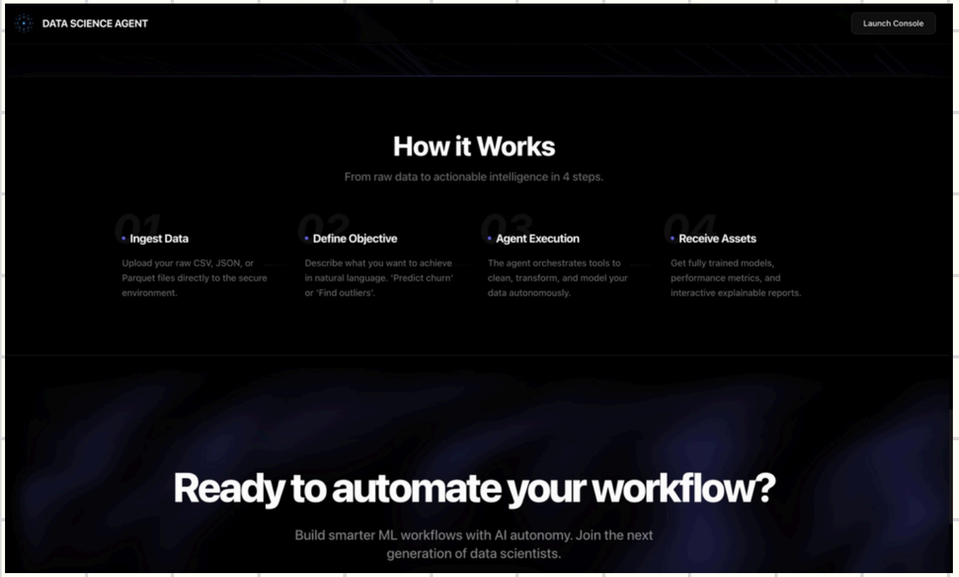
1

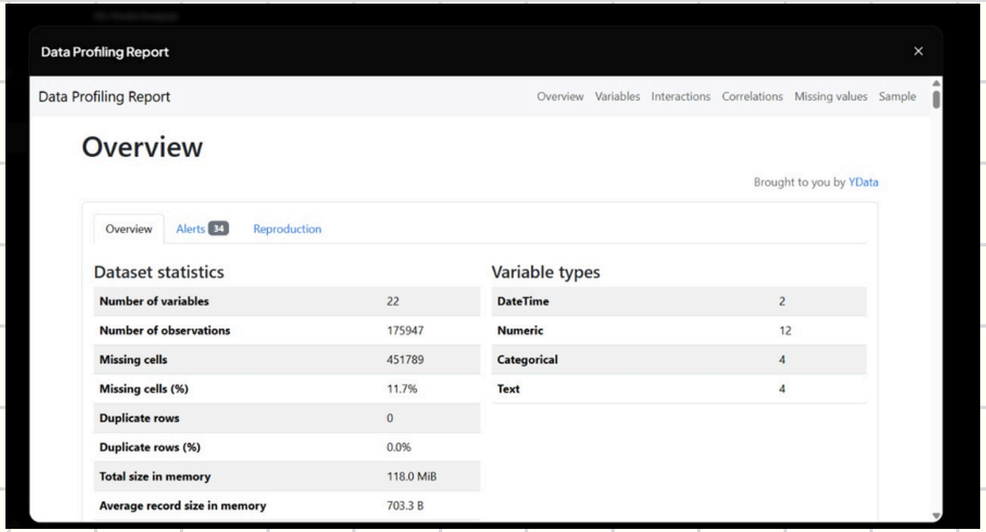
The intial stage where users interact with the system through a web interface.

Google Developer Groups
On Campus · MITS-DU Gwalior

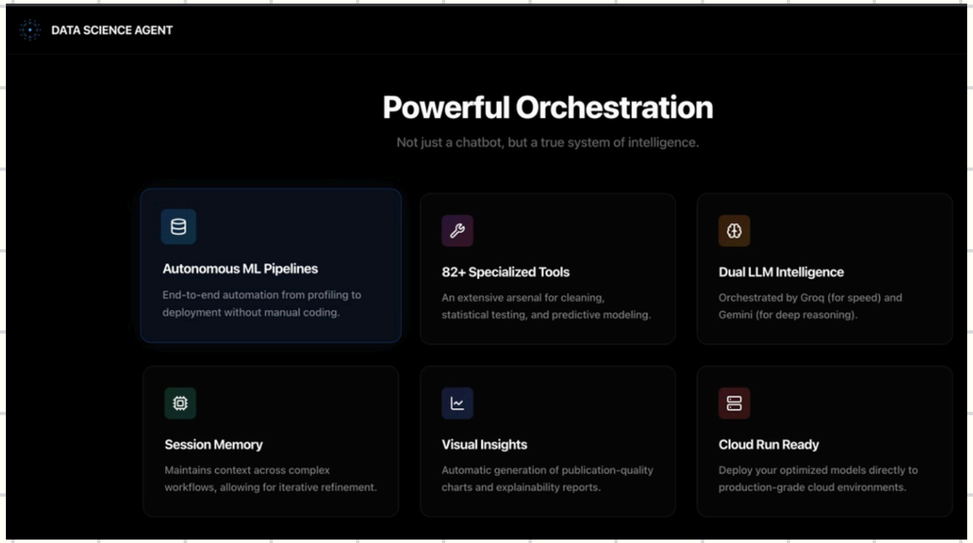
DevSprint







Snapshots of the MVP



Google Developer Groups
On Campus · MITS-DU Gwalior

DevSprint

Future Development

Kaggle Integration

Real-time performance, data drift, and model health tracking

Model Registry

Versioned model storage with performance tracking

Vertex AI Integration

Scalable training, experiment tracking, and MLOps pipelines

Team Collaboration

Shared workspaces for teams to build and review models together

Automated Retraining

Periodic model retraining using new data and drift detection

Multi-User Authentication

Support secure login, user roles, and project isolation

Google Developer Groups
On Campus · MITS-DU Gwalior

DevSprint

Links of Project:

GitHub Public Repository : https://github.com/Pulastya-B/DevSprint-Data-Science-Agent

Demo Video Link (3 Minutes):

https://drive.google.com/drive/u/0/folders/18bfEhsLfah9mZaSST6XVbvH0Yl_t434Q

MVP Link: https://data-science-agent-cezn.onrender.com

Google Developer Groups
On Campus · MITS-DU Gwalior

{DevSprint}

Thank you!