

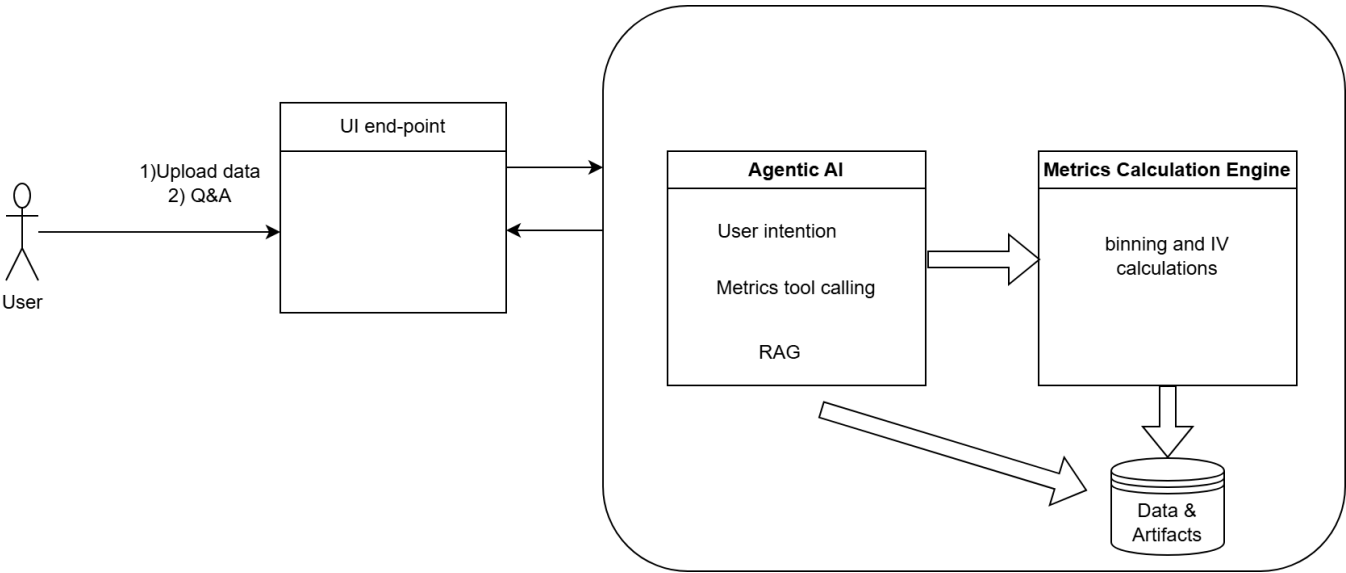
POC Plan and Solution Design (Scenario Information Value)

POC Plan:

Input dataset : training_data_20251007_v4.parquet

Item	Components	Goal	Provided Calculation logic and scorecard model	Success Criteria
1	Foundational Metric Engine (IV)	Prove that IV and related metrics can be automatically calculated from a dataset using the provided metric calculation logic.	iv_20251208 model_performance_metrics_code mtb_scorecard_3.pkl	Given a test dataset (CSV or parquet), the engine correctly produces the below objects containing the IV for each feature, which can be verified manually. <ul style="list-style-type: none">• outputs/MTB_features_IV.csv• outputs/YNB_features_IV.csv
2	Report (IV) relevant Artifacts generation	Prove that IV for each segment (MTB / YNTB) for both training & combined datasets can be automatically calculated and formatted into pic with text description using the provided metric calculation logic.		Given a test dataset (CSV or parquet), the below report artifacts are successfully generated into Pic format which can be verified manually. <ul style="list-style-type: none">• Table 3.3.1.1 Information Value (IV) for MTB Segment Features• Table 3.3.1.2 Information Value (IV) for YNTB Segment Features• <u>Feature Stability</u>: Appendix 2 & 3
3	Interactive AI Agent & Q&A	Enable a user to control the process and ask questions about the documents and results using natural language.	NA	The user can trigger metric calculation (IV) and report artifacts generation (IV) via a chat command (automated model performance runs) The user can ask Professional model Q&A and receive an accurate, context-aware answer.

Workflow



Implementation steps:

Step Index	Purpose	Task Breakdown	Dependency	Estimated efforts	Output
1	auto-calculation of IV (triggered by user upload training file)	<ul style="list-style-type: none"> feature functions for binning method algo functions for IV calculation artefact functions for tabular data and feature stability generation aggregation process based on use cases 	<ul style="list-style-type: none"> binning method logic (how to define user split? any other source apart from scorecard) clarity on use cases and aggregation method from Client 	09 Dec 2025 to 16 Dec 2025	
2	collect and land necessary artefacts	<ul style="list-style-type: none"> define necessary artefacts based on use cases proper landing process based on artefacts types resource configuration (DB / Data lake) 	<ul style="list-style-type: none"> step 1 to enable IV calculation 	16 Dec 2025 to 22 Dec 2025	
3	generate and store index for RAG	<ul style="list-style-type: none"> landed artefacts and IV metrics as input multimodal data processing (convert multimodal info to text info) build chunking / vectorised process under index framework (LlamaIndex) resource configuration (VectorDB) 	<ul style="list-style-type: none"> step 2 to generate artefacts No dependency on index related step 	11 Dec 2025 to 20 Dec 2025	
4	build chatbot backend and UI	<ul style="list-style-type: none"> build backend API build frontend UI integrate backend API with frontend UI Function call – Metric Engine (IV) 	None	11 Dec 2025 to 23 Dec 2025	
5	Test report generation of IV	<ul style="list-style-type: none"> build test dataset build evaluation metrics system prompt design 	<ul style="list-style-type: none"> step 1 to 4 	24 Dec 2025 to 06 Jan 2026	
6	Test Q&A with IV related topic	<ul style="list-style-type: none"> build test dataset build evaluation metrics system prompt design 	<ul style="list-style-type: none"> step 1 to 4 	24 Dec 2025 to 06 Jan 2026	