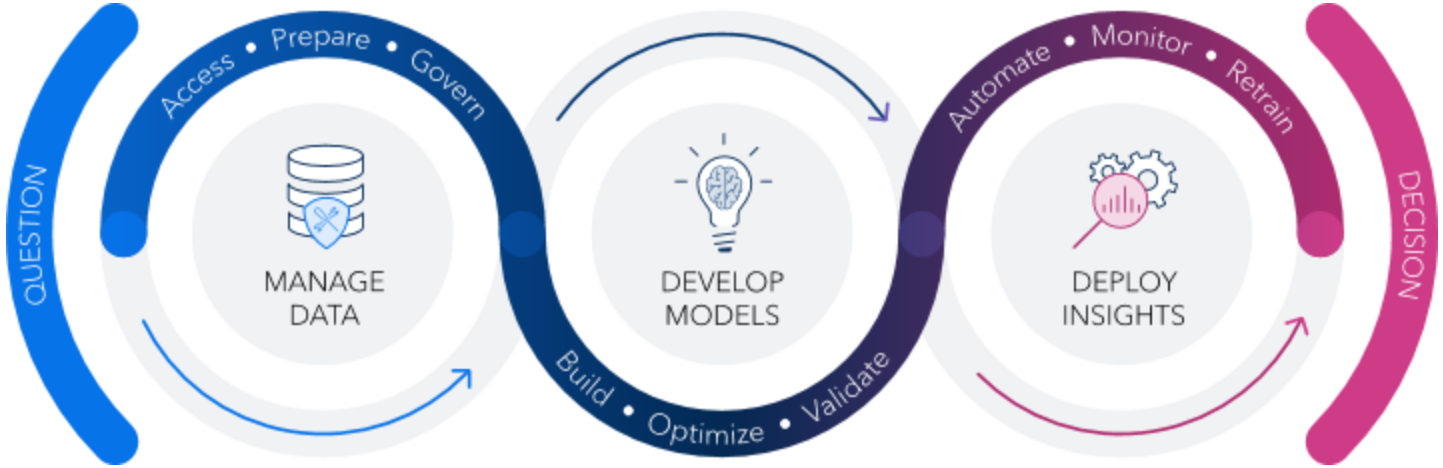
Harnessing Generative AI to Transform Clinical Trials

Soundarya Palanisamy, Sr Industry Advisor, Global Health and Life Sciences Practice, SAS Institute

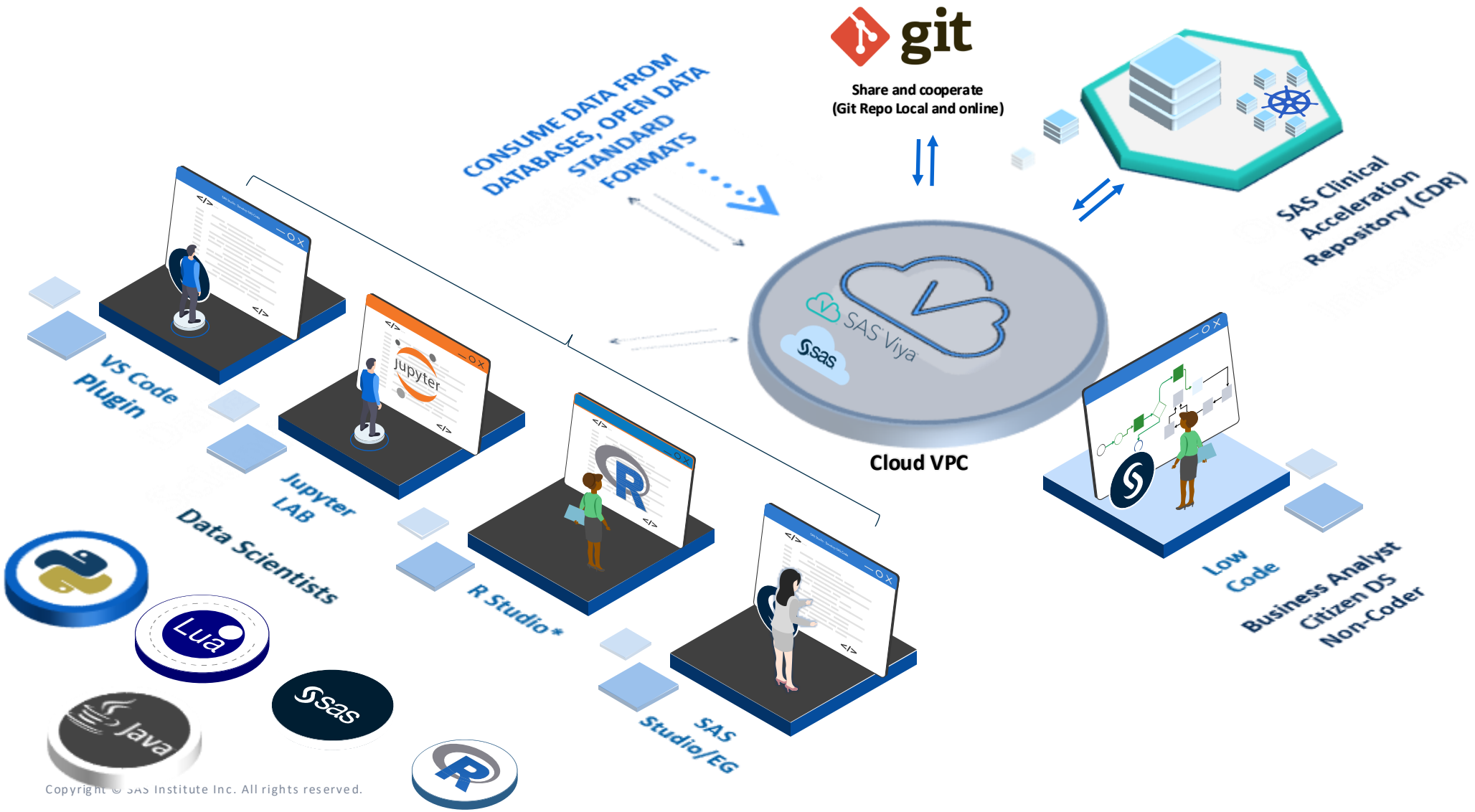
Viya is the Latest SAS Platform That Connects All Aspects of the Data and AI Lifecycle



DATA FOR ANALYTICS AND DECISIONING	ANALYTICAL VISUALIZATION	ADVANCING WITH AI	AUTOMATE MODELING	OPERATIONALIZING ANALYTICS
Accelerate data storytelling and increase speed to value.		*Thread automation and intelligence throughout the data & AI lifecycle.*	*Deliver fast, self-service power to every user, and deliver embedded smart insights.*	

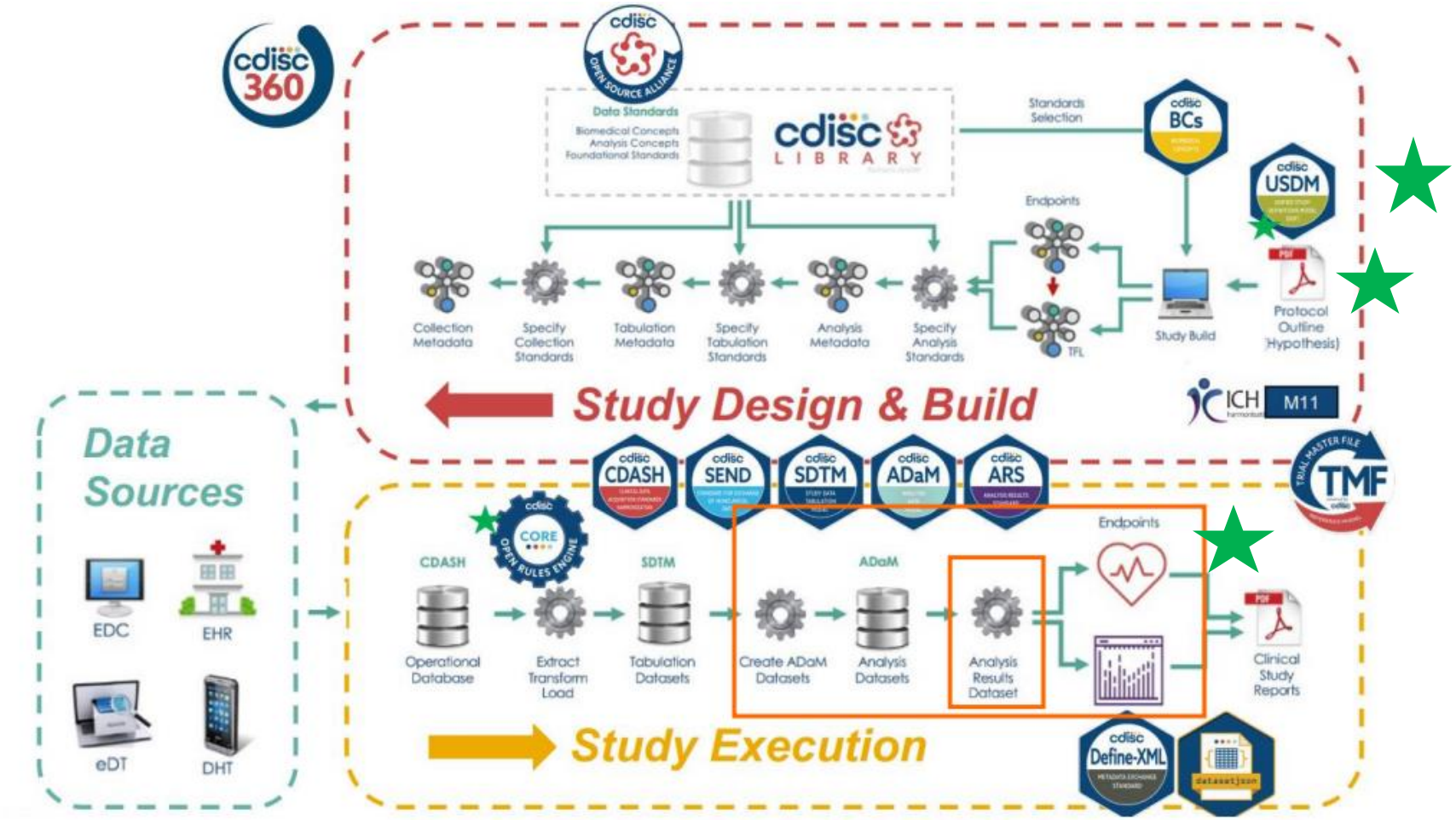
SAS Viya

SAS Viya is the latest release from SAS. It is your one-stop-shop for all your analytic needs. Think of it as suites of product that does things like data preparation, data visualization, model building, Programming (SAS\open-source) etc., all running on a very powerful in-memory & SAS 9 compute engine.

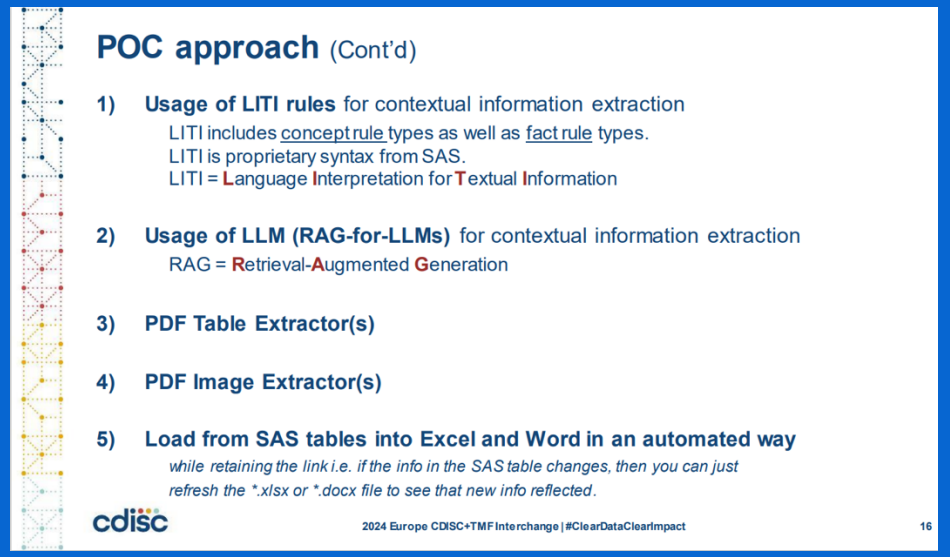


Case Study

CDISC 360 Vision



Case Study: Development of USDM through translation of human-readable protocols



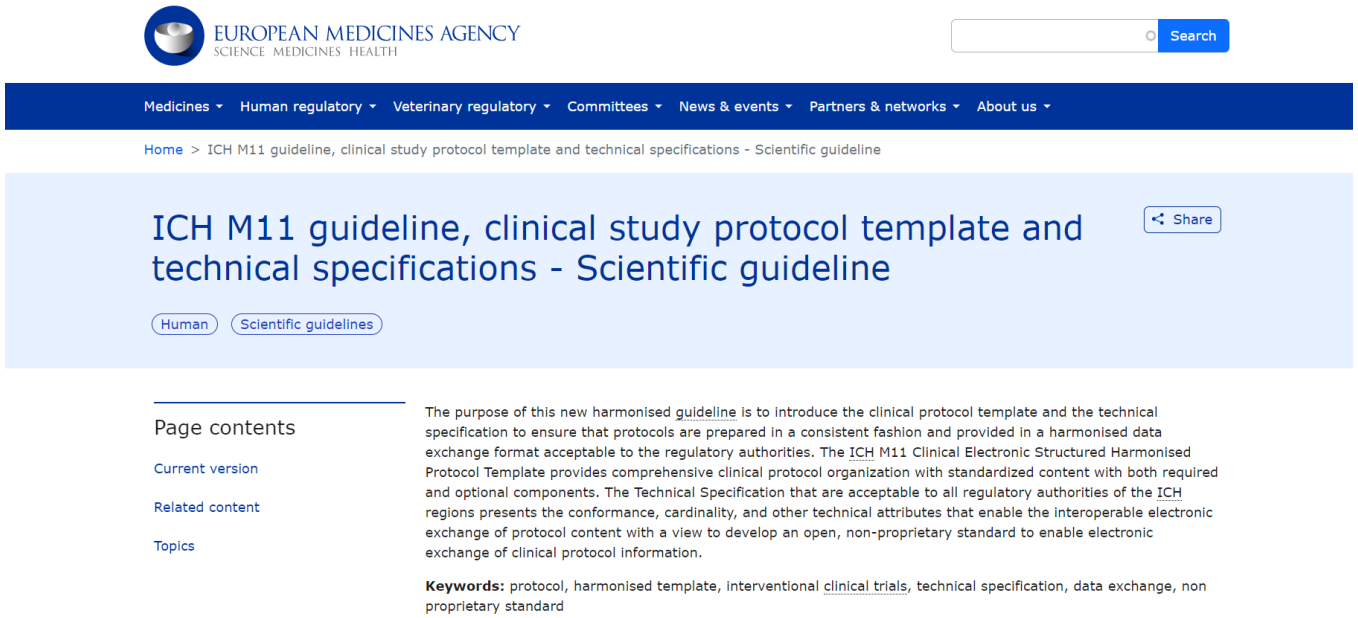
Motivation and background

Need to standardize protocols

[ICH M11 guideline, clinical study protocol template and technical specifications - Scientific guideline | European Medicines Agency (europa.eu)]

Explore new technology

ICH M11 guideline



AIM – Perform **contextual extraction** from clinical trial protocols and **store relevant information** into the USDM workbook in the **right place** (in the right field)

studyDesignOE Sheet

objectiveName	objectiveDescription	objectiveLabel	objectiveText	objectiveLevel	objectiveDictionary	endpointName	endpointDescription	endpointLabel	endpointText	endpointPurpose	endpointLevel	endpointDictionary
OBJ1	Primary		The primary efficacy objective for this study is to evaluate the efficacy of TCZ compared with placebo in combination with SOC for the treatment of severe COVID-19 pneumonia	Study Primary Objective		END1	Day 28, 7 category scale		Clinical status assessed using a 7-category ordinal scale at Day 28		Primary Endpoint	
OBJ2	Secondary		The secondary efficacy objective for this study is to evaluate the efficacy of TCZ compared with placebo in combination with SOC for the treatment of severe COVID-19 pneumonia over the age of [min_age]	Study Secondary Objective	OE_Dict	END2	TTCI		Time to clinical improvement (TTCI) defined as a National Early Warning Score 2 (NEWS2) of <=2 maintained for 24 hours		Secondary Endpoint	
						END3	Time to Imporvement		Time to improvement of at least 2 categories relative to baseline on a 7-category ordinal scale of clinical statu		Secondary Endpoint	

USDM Excel Sheet Formats & Links Infographic
17th January 2024, USDM Package v0.43

Details the excel workbook format as used by the **USDM** python package.

Details of the package can be found at https://github.com/data4knowledge/usdm. Details for using the package and the sheet formats are detailed within the readme file within the repository.

studyDesign Sheet

studyDesignName	Study Design 1				
studyDesignDescription	The main design for the study				
therapeuticAreas	SPONSOR:T2_DIABETES=Type 2 diabetes, SNOMED: 73211009=Diabetes mellitus (disorder)				
studyDesignRationale	Basic study				
studyDesignBlindingScheme	OPEN LABEL				
trialIntentTypes	BASIC SCIENCE, DEVICE FEASIBILITY				
trialTypes	Efficacy Study				
interventionModel	C82639				
mainTimeline	mainTimeline				
otherTimelines					
Epoch/Arms	Screening	Baseline	Treatment	Follow-Up	
Active	EL1	EL2	EL3, EL5	EL4	
Placebo	EL1	EL2	EL3, EL5	EL4	

studyDesignElements Sheet

name	studyElementDescription	Label	transitionStartRule	transitionEndRule
EL1	Screening Element	Screening	Study Start	Screened
EL2	Baseline Element	Baseline	Screened	Radomized
EL3	Treatment Element 1	Treatment 1	Radomized	Completed treatment 1
EL4	Follow Up Element	Follow Up	Treated	Leave Study
EL5	Treatment Element 2	Treatment 2	Radomized	Completed treatment 2

studyDesignArms Sheet

name	description	label	type	dataOriginDescription	dataOriginType
Active	Active Substance	Active Substance	Active Comparator Arm	Data collected from subjects	Data Generated Within Study
Placebo	Placebo	Placebo	Placebo Comparator Arm	Data collected from subjects	Data Generated Within Study

<timeline name> Sheet

Name	Main Timeline	name	SCREEN	PRE DOSE	DOSE	D14	PROG	D28	FU
Description	This is the main timeline for the study design.	description	Screening	Pre Dose	Dosing	Day 14	Check Opt In	Day 28 Optional	Follow Up
Condition	Potential subject identified	label	Screen	Baseline	Treatment	Day 14	Check opt In by subject	Day 28	Follow Up
		type	Activity	Activity	Activity	Activity	Decision	Activity	Activity
		default	PRE DOSE	DOSE	D14	PROG	D28	FU	(EXIT)
		condition					FU: if opted out		
		epoch	Screening	Baseline	Treatment	Treatment		Treatment	Follow-Up
		encounter	E1	E2	E3	E4		E5	E6
Parent Activity	Child Activity	BC/Procedure/Timeline							
-	Demographics	BC:Age, BC:Sex, BC:Race, BC:Body Weight	X						
-	Procedures	PR: PR1, PR: PR2	X	X	X	X			X
-	Optional Weight	BC:Weight	X			X			
	Optional	BC:SYSBP, BC:DIABP						X	

Accuracy Enhancements

Result

Example << primary / secondary endpoint >>

Endpoints

6.2.1. Primary Endpoint

- Percentage reduction in total IgG levels, compared to baseline, at day 29 (week 4), 7 days after the fourth IV or SC administration of efgartigimod

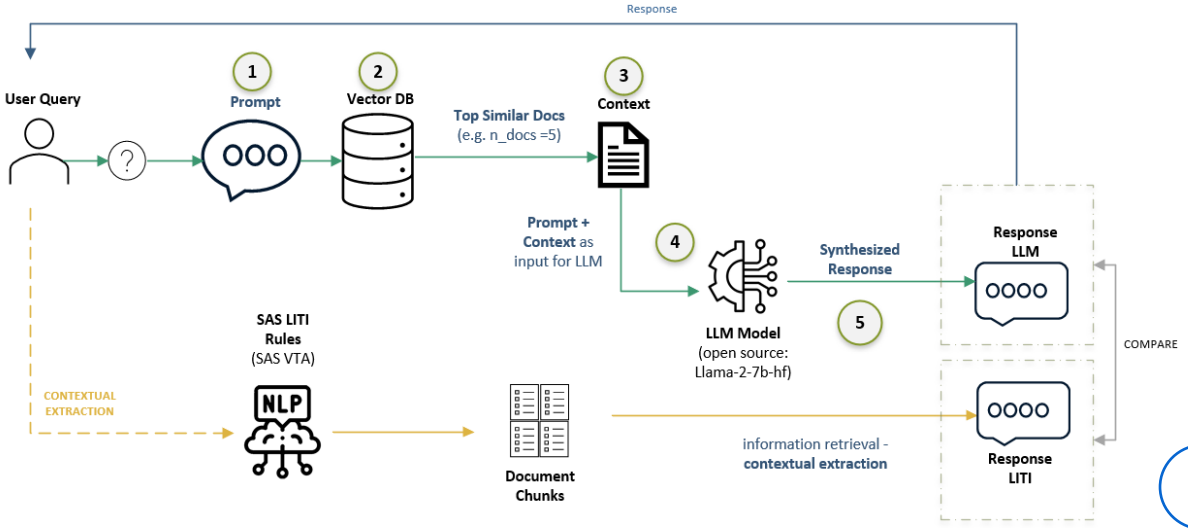
6.2.2. Secondary Endpoints

- Percentage reduction in total IgG levels at all other assessment timepoints as of week 4
- Percentage reduction in levels of IgG subtypes (IgG1, IgG2, IgG3, and IgG4) at all assessment timepoints
- Absolute values and changes from baseline in total IgG levels and levels of IgG subtypes (IgG1, IgG2, IgG3, and IgG4) at all assessment timepoints
- AUEC for percentage reduction in total IgG levels and for each subtype per weekly interval after each dose (week 1, week 2, week 3, and week 4), over the interval week 1 to week 4, and over the entire study period (week 1 to week 11)
- Serum levels of efgartigimod and derived PK parameters
- Clinical laboratory evaluations, vital sign measurements, ECG recordings, and incidence and characterization of TEAEs

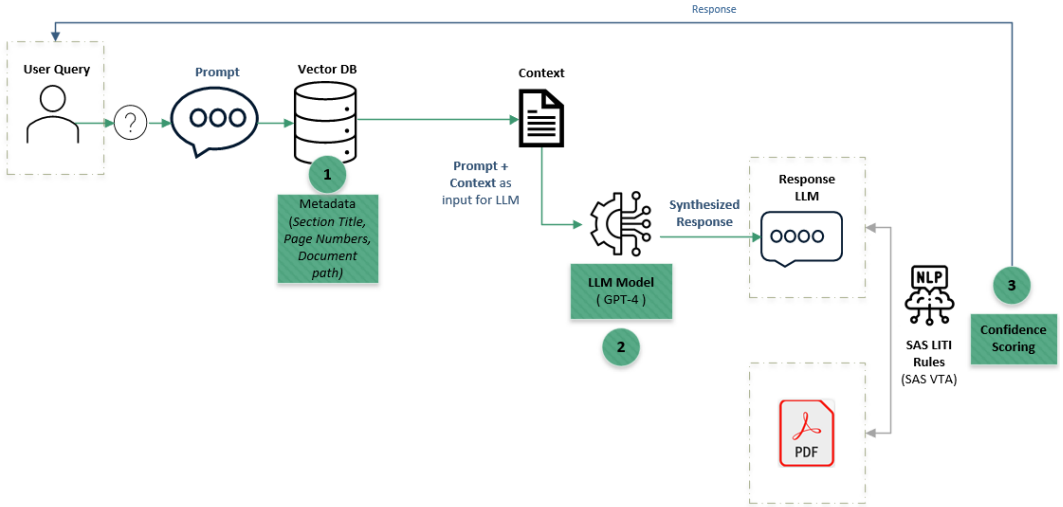
```
query_text = "List all Primary and Secondary endpoints of the study"

context = retriever.get_relevant_documents(query_text)
prompt = prompt_template.format(context=context, question=query_text)

with get_openai_callback() as cb:
    response = llm.invoke(prompt)
    print(f"Total Cost (USD): ${format(cb.total_cost, '.6f')}")
    print(f"Model: {response.response_metadata['model_name']}, Response Metadata: {response.response_metadata['token_usage']}")
    print(response.content)
#     print(response)
```

```
Total Cost (USD): $1.602420
Model: gpt-4, Response Metadata: {'completion_tokens': 256, 'prompt_tokens': 52902, 'total_tokens': 53158}
Primary Endpoint:
- Percentage reduction in total IgG levels, compared to baseline, at day 29 (week 4), 7 days after the fourth IV or SC administration of ef
gartigimod.

Secondary Endpoints:
- Percentage reduction in total IgG levels at all other assessment timepoints as of week 4.
- Percentage reduction in levels of IgG subtypes (IgG1, IgG2, IgG3, and IgG4) at all assessment timepoints.
- Absolute values and changes from baseline in total IgG levels and levels of IgG subtypes (IgG1, IgG2, IgG3, and IgG4) at all assessment t
imepoints.
- Area under the effect curve (AUEC) for percentage reduction in total IgG levels and for each subtype per weekly interval after each dose
(week 1, week 2, week 3, and week 4), over the interval week 1 to week 4, and over the entire study period (week 1 to week 11).
- Serum levels of efgartigimod and derived PK parameters.
- Clinical laboratory evaluations, vital sign measurements, ECG recordings, and incidence and characterization of TEAEs.
```

1 SAS LITI (NLP) + LLM for contextual extraction



2 LLM (Open AI) + SAS LITI (NLP + Confidence Scoring)



3 Evaluating Gen AI and Text Analytics for creating a USDM

Some initial results after pre processing, highlights the extractions of **inclusion** and **exclusion criteria** by LITI rules and LLMs.

Legend :

- perfect
- > 75 % hit rate
- between 60 % and 75 % hit rate
- < 60% hit rate

Select any image to see a larger version.

Mobile users: To view the images, select the "Full" version at the bottom of the page.

		INCLUSION CRITERIA			EXCLUSION CRITERIA		
Documents (clinical trial protocols)	document partitioning (training / validation / test)	Number	Number Captured With LITI	Number Captured with LLM	Number	Number Captured With LITI	Number Captured with LLM
Protocol - 13 Feb 2020.pdf	TRAINING	15	15	12*	22	22	19
Protocol_22Jul2020.pdf	TEST	7	7	6	20	20	34*
EliLilly_NCT03421379_Diabetes.pdf		10	10	10	26	26	26
Roche_NCT04320615_COVID.pdf		7	7	7	12	12	12
CDISC_Pilot_Study.pdf		8	8	8	23	23	23

There are several challenges to overcome, such as overfitting in LITI rules and the need for extensive pre-processing.

Comparison of 2 ways for information retrieval / contextual extraction

LITI – rules (SAS VTA)

- It's "Regular Expressions on steroids".
- Knowledge (and effort) required to write linguistic rules with correct syntax.
- Some patterns are hard to come by (because they are complicated, non-standard, with a lot of variety in the language …).
- You need some preliminary knowledge on the topic at hand and on what you want to catch!
- Rules are easily overfit (too specific) if n° of training docs is low.
- Library of LITI-rules can / should grow very big to guarantee a high hit rate.
- Results are "proven", non-debatable (there's a clear match in the text).
- Relatively light in terms of resource usage (and cost).
- page breaks and headers and footers are no gift.
- special and non-printable characters pose no problem.

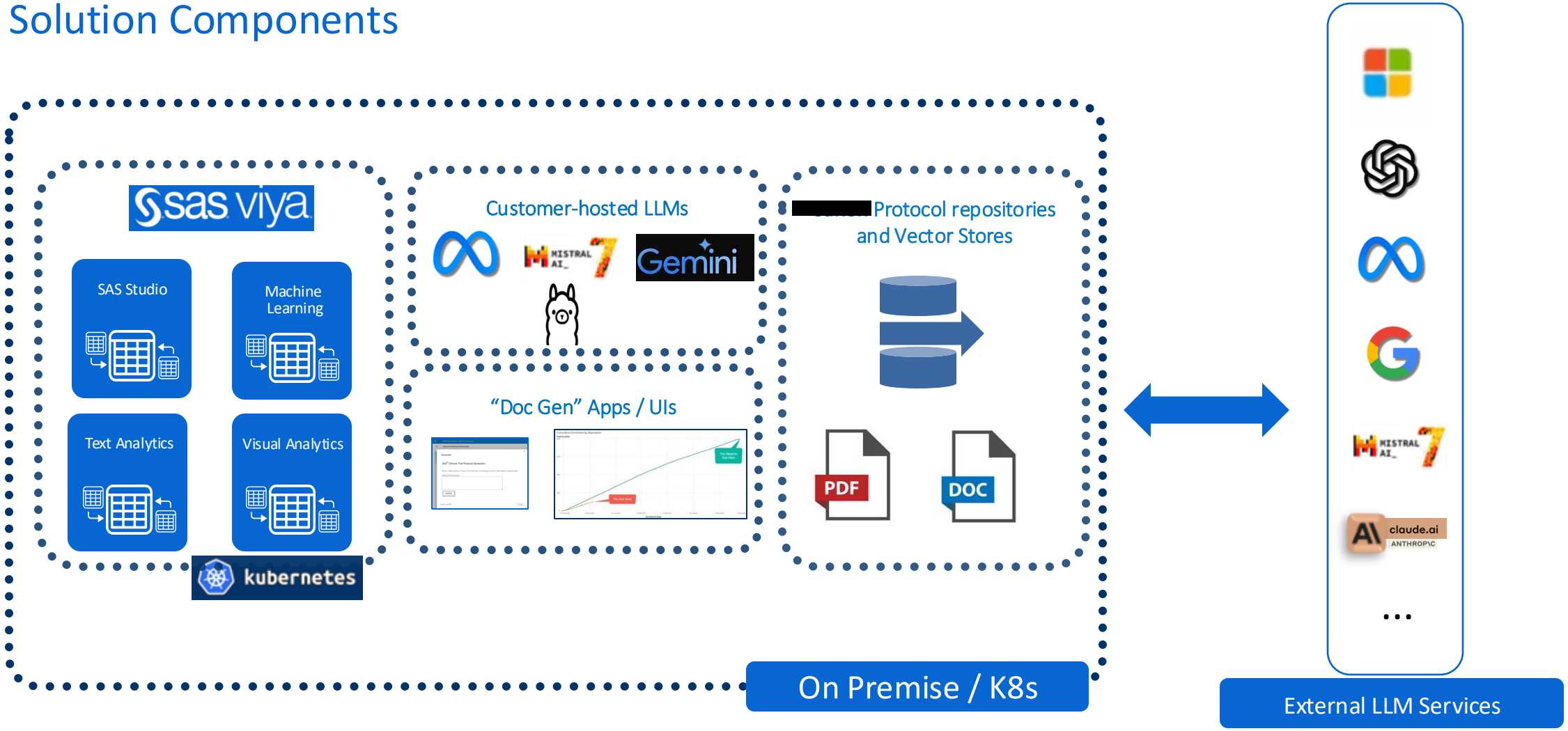
RAG – LLM (Open Source / Commercial Models)

- It's generative AI.
- Knowledge required to do proper prompt engineering (designing the user query). But much less effort required. It's just an API-call to a pre-trained model.
- If the context around a particular topic can vary widely from document to document, there is a clear advantage.
- It's generative AI, so some of the info returned can be made up (hallucination).
- Implementing offline models (internalization) and maintaining them is labor-intensive.
- Heavy and intensive in terms of resource usage (even for one or a few documents). Often, several GPUs are needed, and run-time is considerable.
- Scaling (expand the scope) is easier here.
- page breaks and headers and footers are no gift.
- special and non-printable characters pose no problem.

Case Study: AI-Powered Clinical Trial Protocol Benchmarking, Simulation, Development

This is the house that we built.

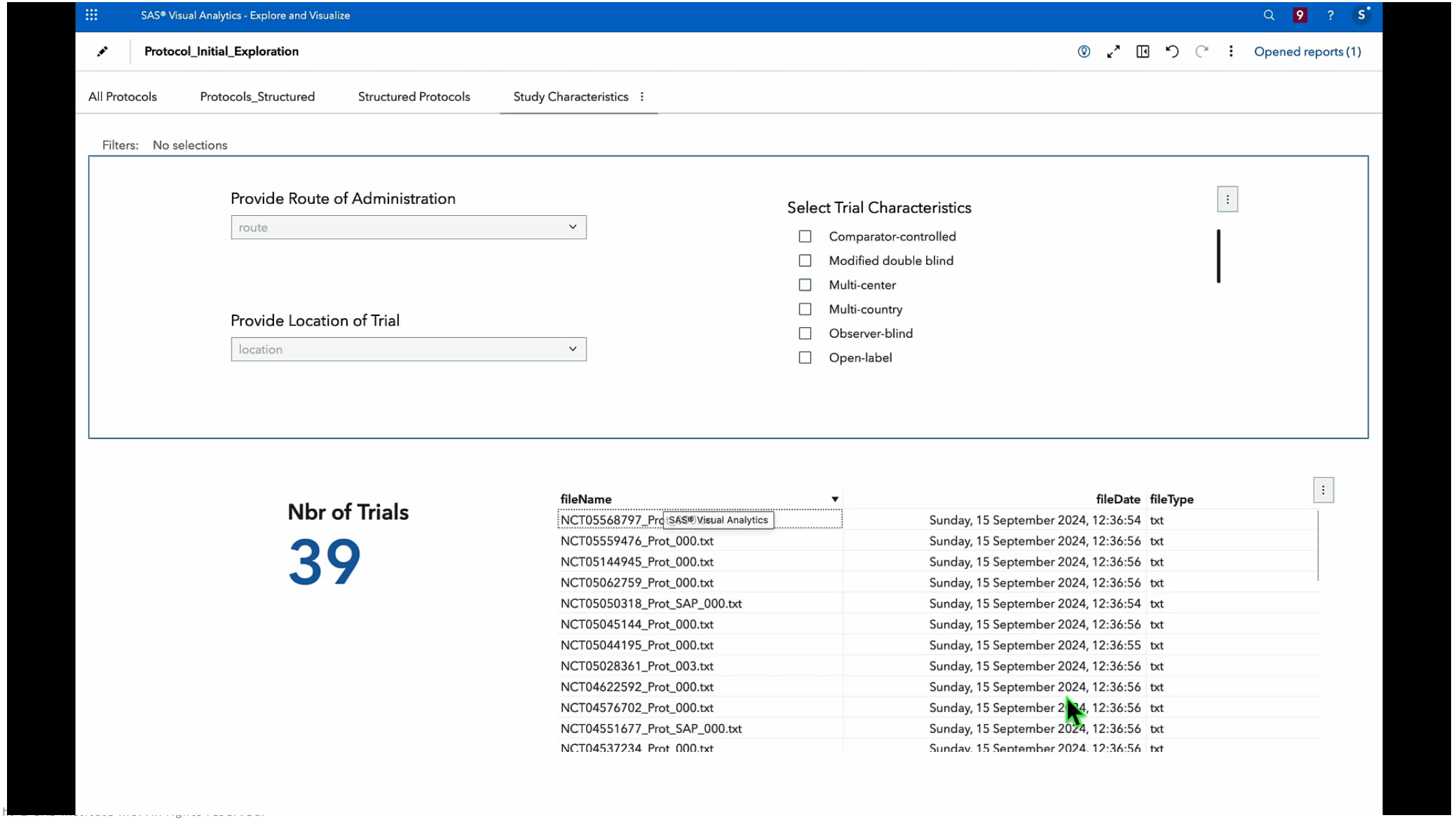
Solution Components



Note: Icons representing LLMs are purely indicative and not recommendations

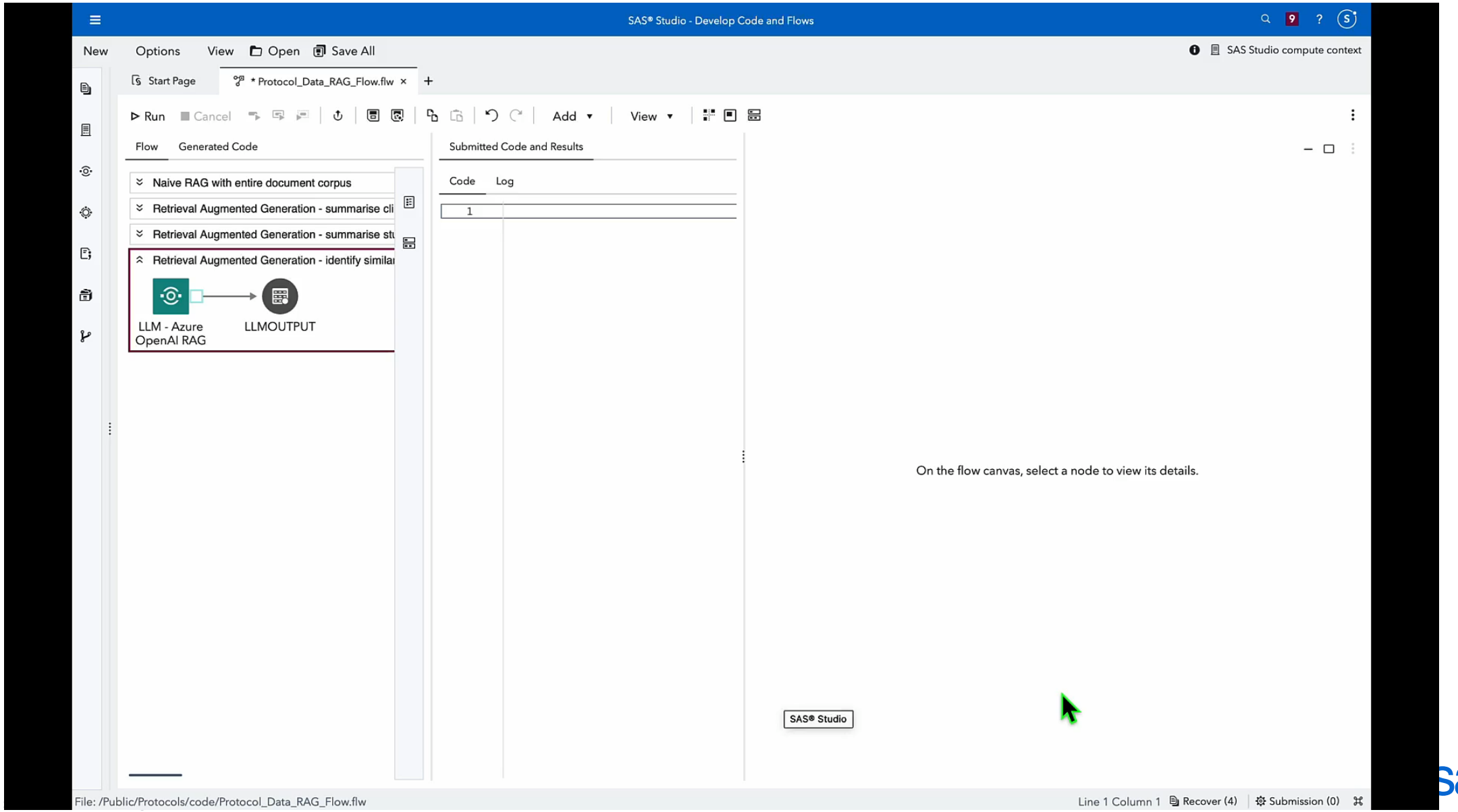
Enter initial study characteristics (incl. objectives and endpoints)

Enter Initial Study Characteristics



Identify similar studies (internal and external)

Identify Similar Studies



Compare study characteristics of similar past and ongoing studies to generate insights

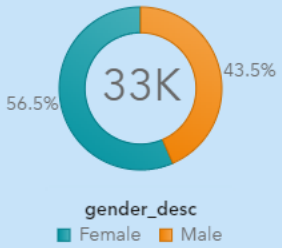
Compare Study Characteristics

Baseline Characteristics (Stringent Vs. Relaxed Vs. ML/SME) | Geographical Distribution (Stringent Vs. Relaxed Vs. ML/SME) | Automated Explanation | Logistic Regression | Decision Tree | Gradient Boosting | Neural Network | Model Comparison

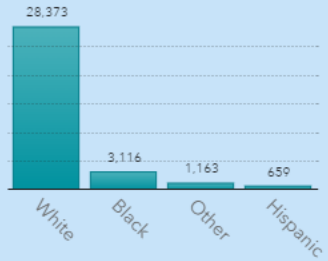
Filters: No selections

STRINGENT ELIGIBILITY CRITERIA (per Study Protocol IE criteria)

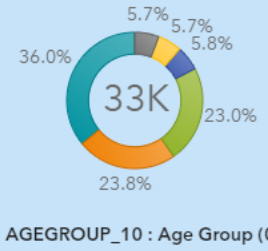
Gender - S



Race - S

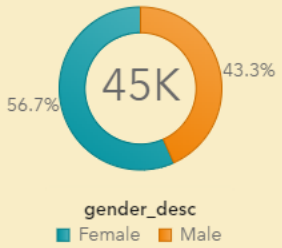


Age Groups - S

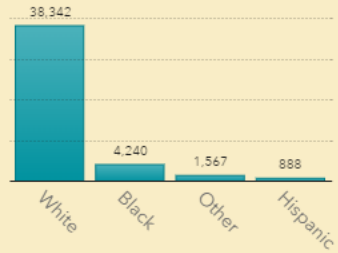


RELAXED ELIGIBILITY CRITERIA (per SME - dropped CHEST PAIN excl. criteria)

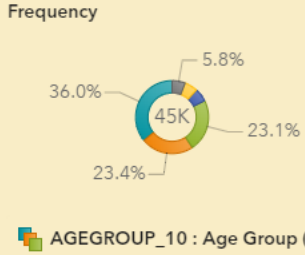
Gender - R



Race - R

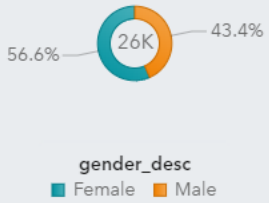


Age Groups - R

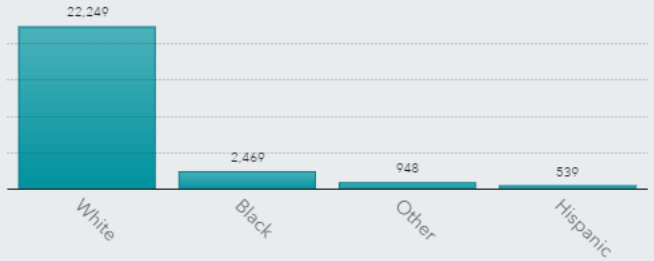


ML & SME driven ELIGIBILITY CRITERIA

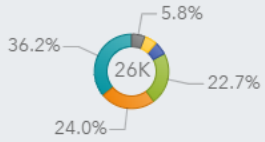
Gender - ML & SME



Race - ML & SME



Age Groups - ML & SME



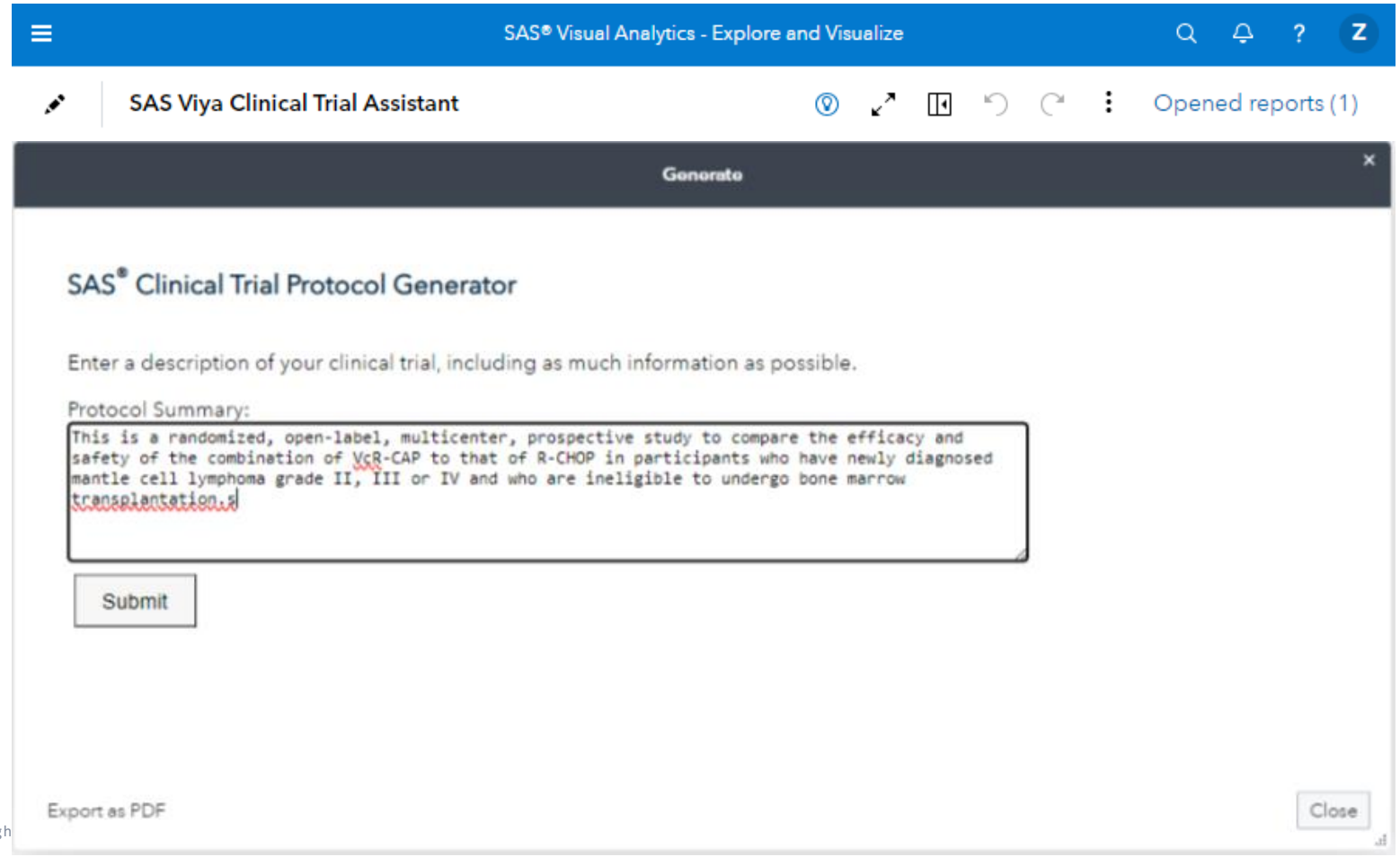
AGEGROUP_10 : Age Group (0-9, 10-19, 20-29, 30-39, 40-49, ...)
70-79 | 60-69 | 80-89 | 50-59 | 90+ | Other

Generate Protocol document

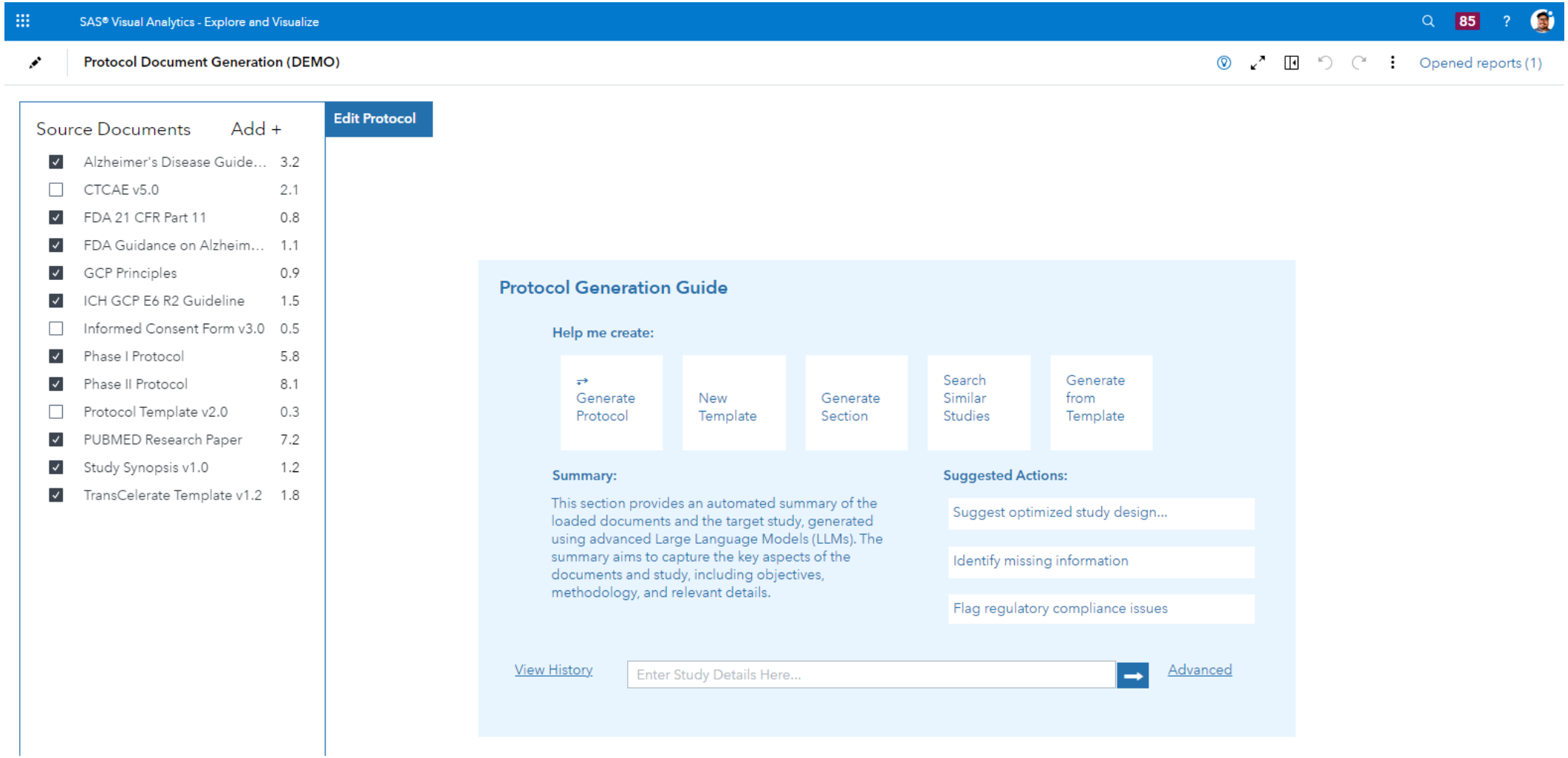
Protocol Draft Generation

SAS seamlessly integrates with LLMs to generate draft protocols

Our demo uses a custom Open AI model that is trained to write protocols based on the **ICH M11 Protocol Template**.



Generate Full Protocol Package & Other Documents (Proposed)



Sample Protocol Generation Results

A Phase III Randomized, Double-blind, Placebo-controlled Multicenter Study in Adults to Determine the Safety, Efficacy, and Immunogenicity

[redacted]

1.3 Relevant Clinical Data:

Phase I and II st[redacted]trated [redacted] with [redacted] convalescent serum. These studies informed the dose selection and regimen for this Phase III trial.

References:

Grade	Description
1	Mild; asymptomatic or mild symptoms; clinical or diagnostic observations only; intervention not indicated.
2	Moderate; minimal, local or non-invasive intervention indicated; limiting age-appropriate instrumental activities of daily living.
3	Severe; inability to perform activities of daily living; hospitalization or prolongation of hospitalization indicated.
4	Life-threatening; urgent intervention indicated.

Relationship to Study Drug:

- The following table will be used to assess the relationship between AEs and study drug:

Relationship	Description
Unrelated	No reasonable possibility that the AE is related to study drug.
Possible	Cannot rule out a reasonable possibility that the AE is related to study drug.
Probable	A reasonable possibility that the AE is related to study drug.
Definite	A clear temporal relationship and/or known mechanism of action suggests a high probability that the AE is related to study drug.

11.2 Serious Adverse Events (SAEs)

Definition: An SAE is any AE that:

- Results in death
- Is life-threatening
- Requires hospitalization or prolongation of hospitalization
- Results in persistent or significant disability

8.3 Formulation of Test and Control Products

- AZD1222: non-replicating ChAdOx1 vector vaccine, 5 × 10^10 viral particles (VP)/dose
- Placebo: sterile saline solution
- M[redacted]
- Formulation:

Product	Active Ingredient (mg/mL)	Other Ingredients (mg/mL)	pH
AZD1222	0.5 mg/mL	Histidine, sucrose, polysorbate 80	7.4
Placebo	-	Sodium chloride, water for injection	7.0

- Packaging and labeling: identical for AZD1222 and placebo, labeled by [redacted]

Label Mock-up: [Insert label mock-up]

8.4 Supply of Study Drug at Site

- Initial supply: shipped to site upon IRB approval
- Replacement kits: provided through IWRS upon notification of subject withdrawal
- Dosage/Dosage Regimen:
- Dose: 2 doses of 5 × 10^10 VP/dose, administered IM on Day 1 and Day 29
- Route: intramuscular injection
- Timing: optimal timing between doses 28 days

8.5 Dispensing

- Authority: designated site pharmacist or investigator
- Requirements: verify subject eligibility, record dispensing in study records

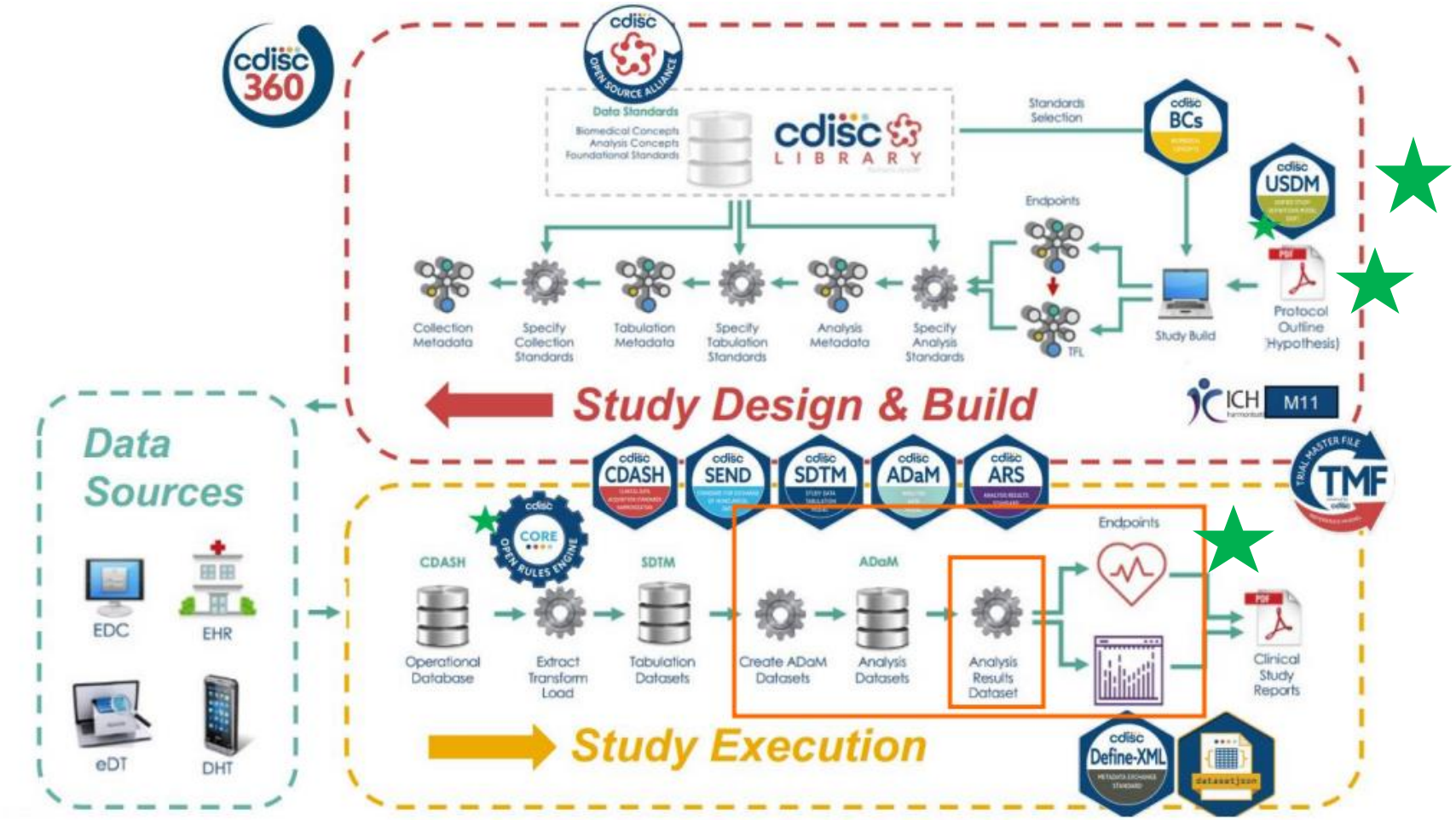
8.6 Administration Instructions

- Step-by-step instructions:

1. Verify subject identity and eligibility
2. Prepare study drug according to packaging instructions
3. Administer study drug via IM injection
4. Record administration in study records

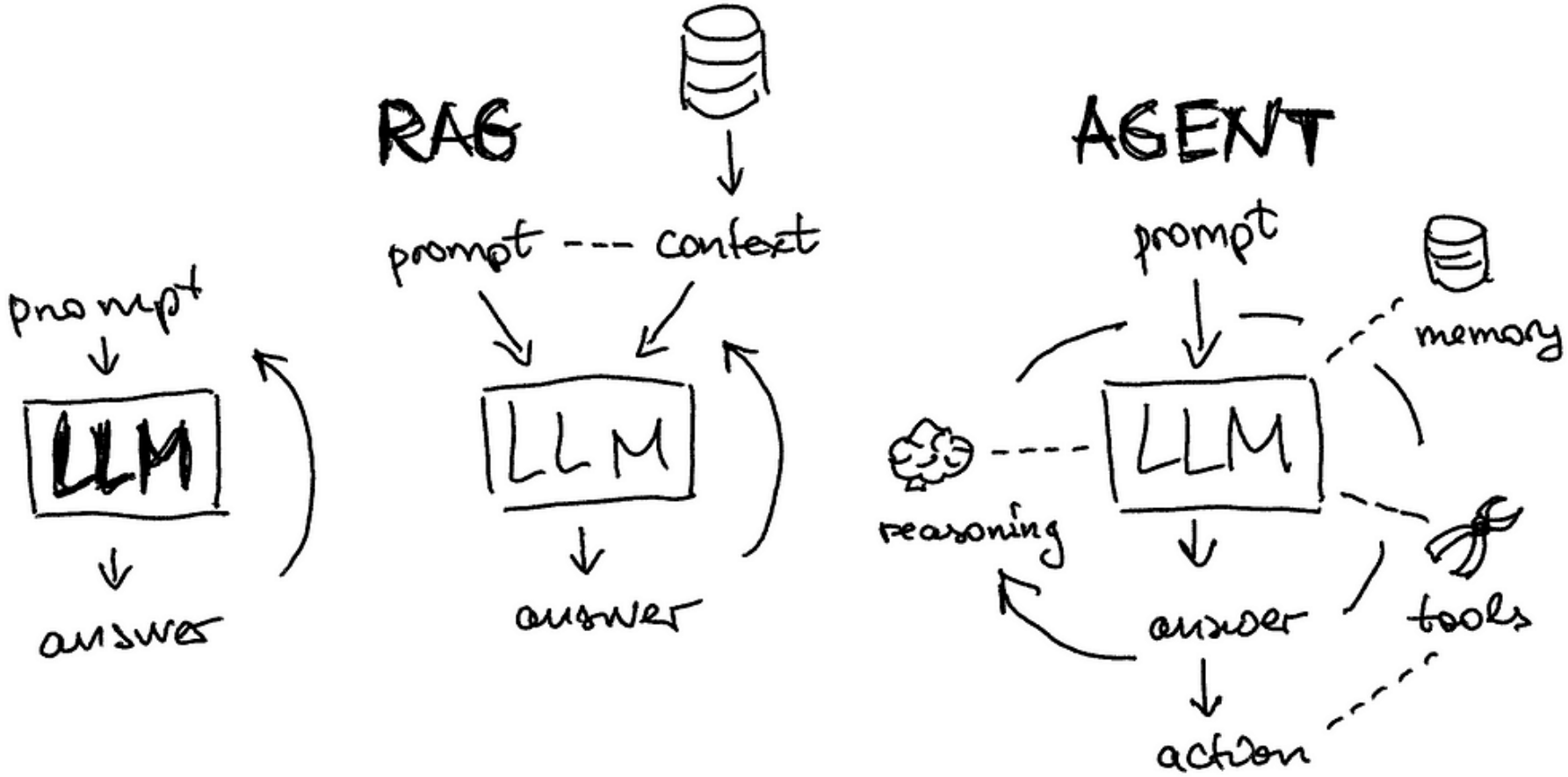
Case Study: SAS agents for clinical trial submission

CDISC 360 Vision



The Art of Building Capable AI Assistants

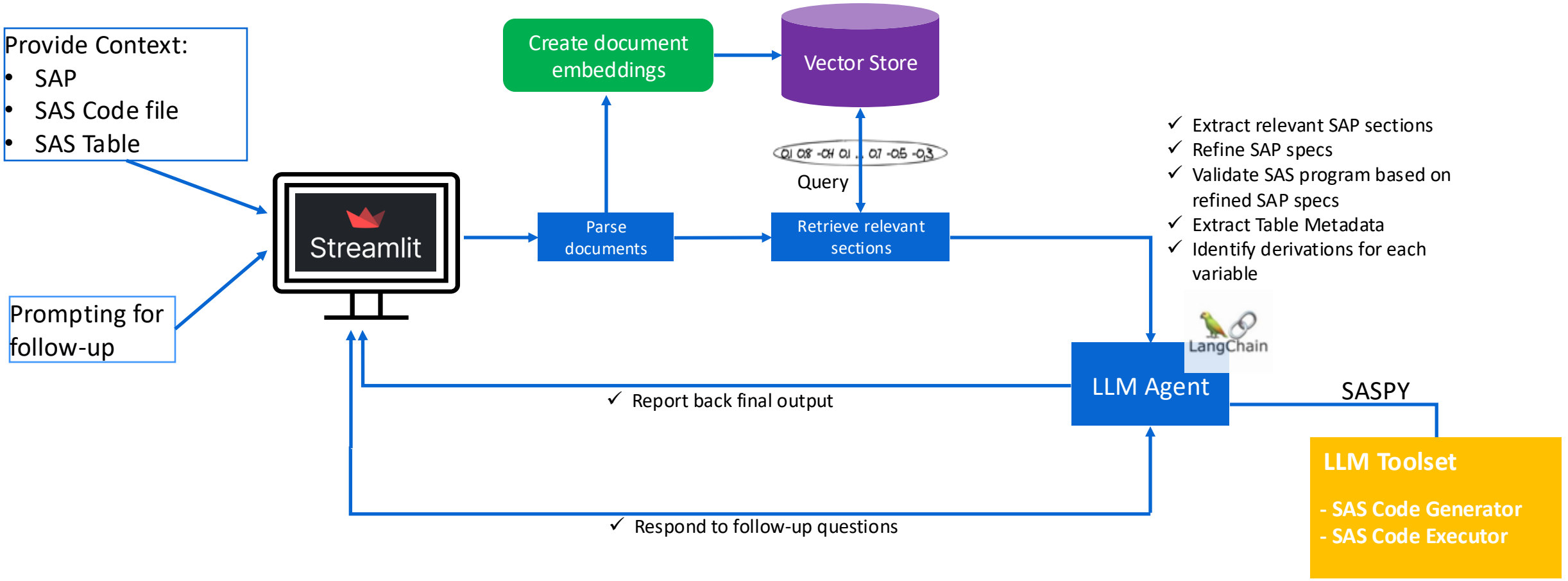
From Foundational LLM to “LLM app”



From:

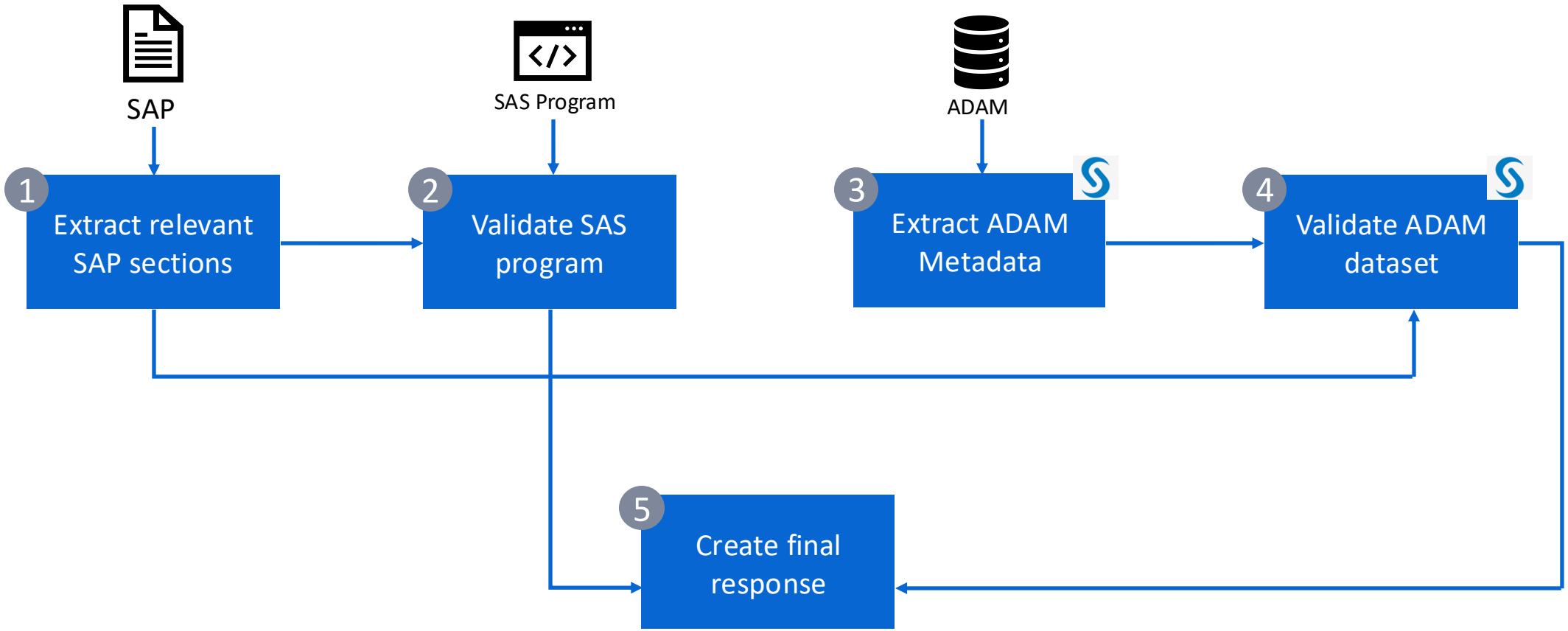
Intro to LLM Agents with Langchain: When RAG is Not Enough | by Alex Honchar | Mar, 2024 | Towards Data Science

High Level Design



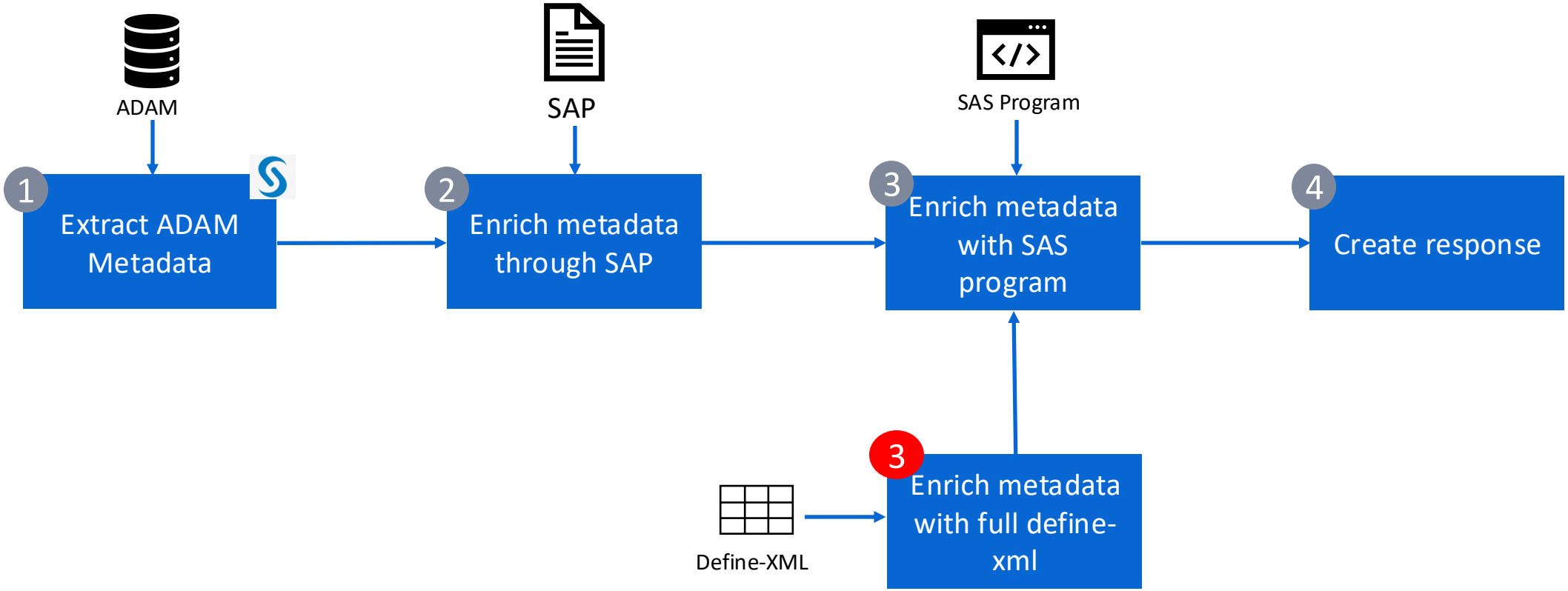
SAS Code Validation

Process flow



Define-XML

Process flow



Connect to SAS Viya

Username

sbxwew

Password

Press Enter to submit form

Login

Deploy

SAS Agents for Clinical Trials

- Label: Criterion 2 Evaluation Result Flag

Validation Findings for ADaM Dataset "EUETS.ADVS"

1. Frequency Distributions for Categorical Variables:

- All categorical variables were checked for frequency distributions. No unexpected categories or missing values were found in key variables like `STUDYID`, `USUBJID`, `SUBJGRI`, `FASFL`, `TNTRIP`, `VSTESTCD`, `PARAMCD`, `AVISIT`, `ATPT`, `APHASE`, `DTYPE`, `ATOXDSCL`, `ATOXGRL`, `ATOXDSCH`, `ATOXGRH`, `ABLFL`, `AML01FL`, `CRIT1FL`, and `CRIT2FL`.

2. Summary Statistics for Continuous Variables:

- Continuous variables were summarized, showing expected ranges and distributions. Key variables like `VSSTRESN`, `AVAL`, `BASE`, and `CHG` were within expected ranges, except for some threshold violations noted below.

3. Outlier Detection in Continuous Variables:

- The `PROC UNIVARIATE` procedure identified extreme values in `VSSTRESN`, `AVAL`, and `CHG`. The minimum value for `VSSTRESN` and `AVAL` was 5, which is below the expected range for some parameters.

4. Validation of Derivations for Change from Baseline:

- The derivation of `CHG` (Change from Baseline) was validated. No mismatches were found between `CHG` and the calculated difference `AVAL - BASE`.

5. Threshold Violations for Abnormalities:

- Numerous threshold violations were detected for parameters like `DIABP` and `RESP`, indicating values outside the expected ranges specified in the SAP. For example, `DIABP` had values below 91 and above 100, and `RESP` had values below 17 and above 25.

Disconnect from SAS Viya

Navigation

SAS Code Validation

Define-XML

7. Handling of Missing Values:

- Missing values were checked for key variables. `BASE` and `CHG` had missing values, which aligns with the SAP's specification of no imputation for missing values.

Summary:

- **Mismatches and Issues**:
  - **Threshold Violations**: Significant threshold violations were found for `DIABP` and `RESP`, which need to be addressed to align with SAP specifications.
  - **Outliers**: Extreme values in `VSSTRESN` and `AVAL` need further investigation to ensure data integrity.
- **Suggestions for Corrections**:
  - Review and correct the data for `DIABP` and `RESP` to ensure they fall within the specified thresholds.
  - Investigate the source of extreme values in `VSSTRESN` and `AVAL` to determine if they are valid or require correction.

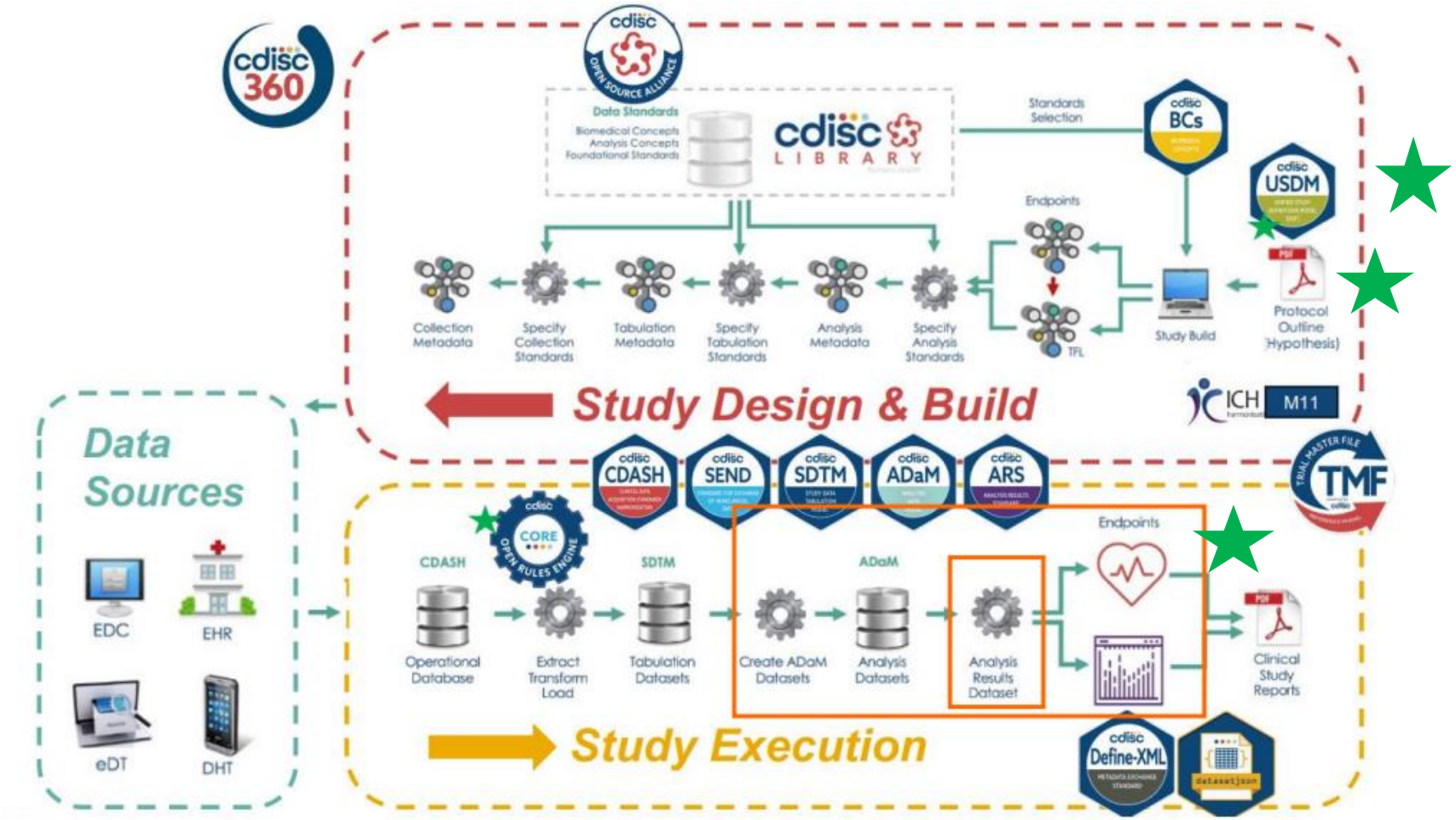
Overall, the dataset requires adjustments to meet the SAP specifications, particularly concerning threshold violations and outlier management.

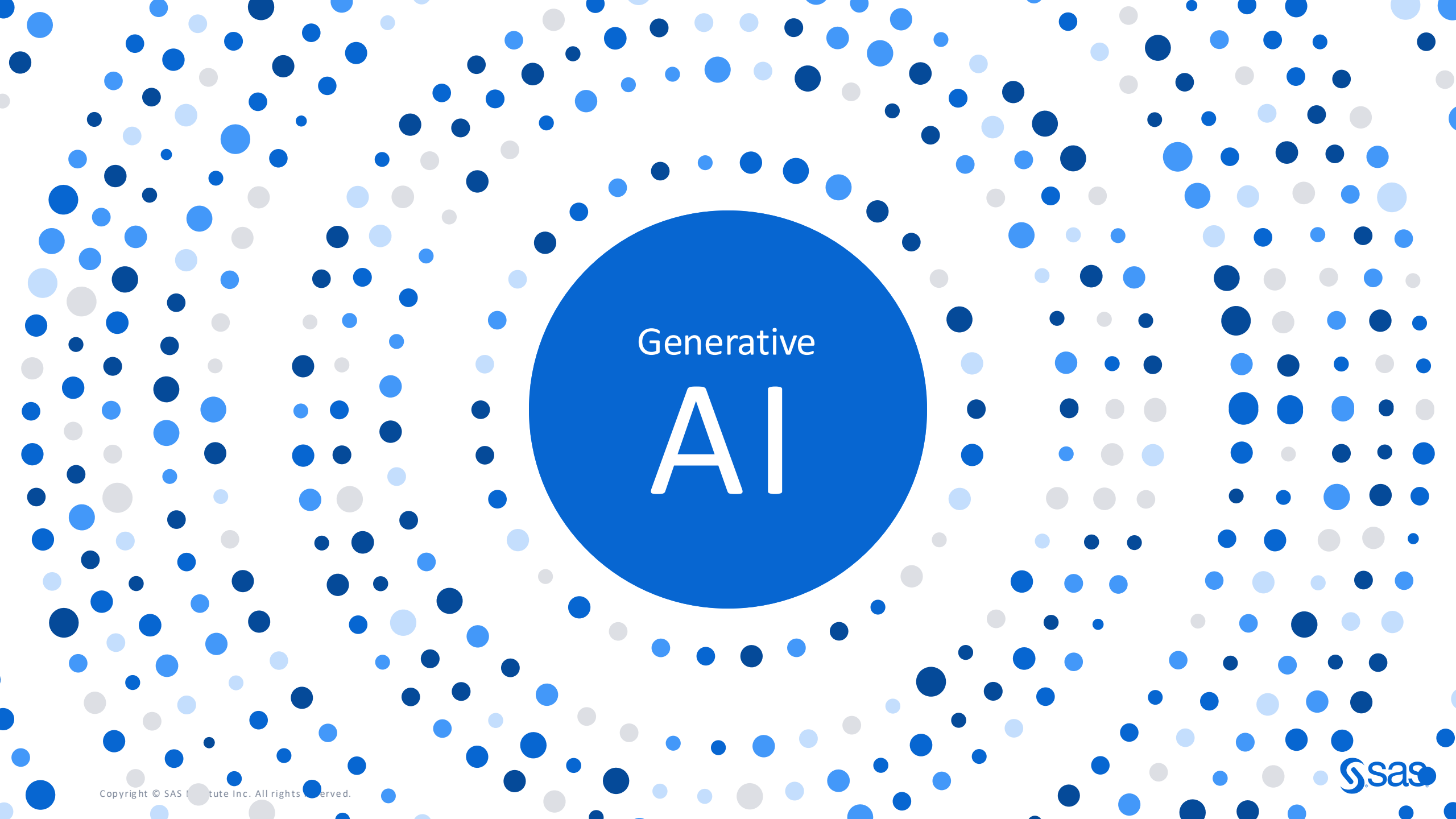
Combined Findings from Dataset and Code Validations

Areas of Compliance:

1. **Tachycardia and Bradycardia**:
   - Both the dataset and the SAS program correctly implement the grading for tachycardia and bradycardia as per the SAP specifications.

CDISC 360 Vision





Generative AI in action with SAS

SAS Viya

Explain, Govern,
Orchestrate LLMs

Viya Copilots

Accelerate AI
Lifecycle / Industry/
Business Tasks

Synthetic Data Generation

Mitigate Data
Quality and Scarcity

Generative AI in Action with SAS

SAS Viya

Explain, Govern,
Orchestrate LLMs

Viya Copilots

Accelerate AI
Lifecycle / Industry/
Business Tasks

Synthetic Data Generation

Mitigate Data
Quality and Scarcity

SAS® Viya® Copilot to help you use SAS Viya

Viya Copilot to further democratize analytics as a suite of GenAI productivity tools to help you conversationally query the end-to-end SAS Viya AI & Analytics Life Cycle platform

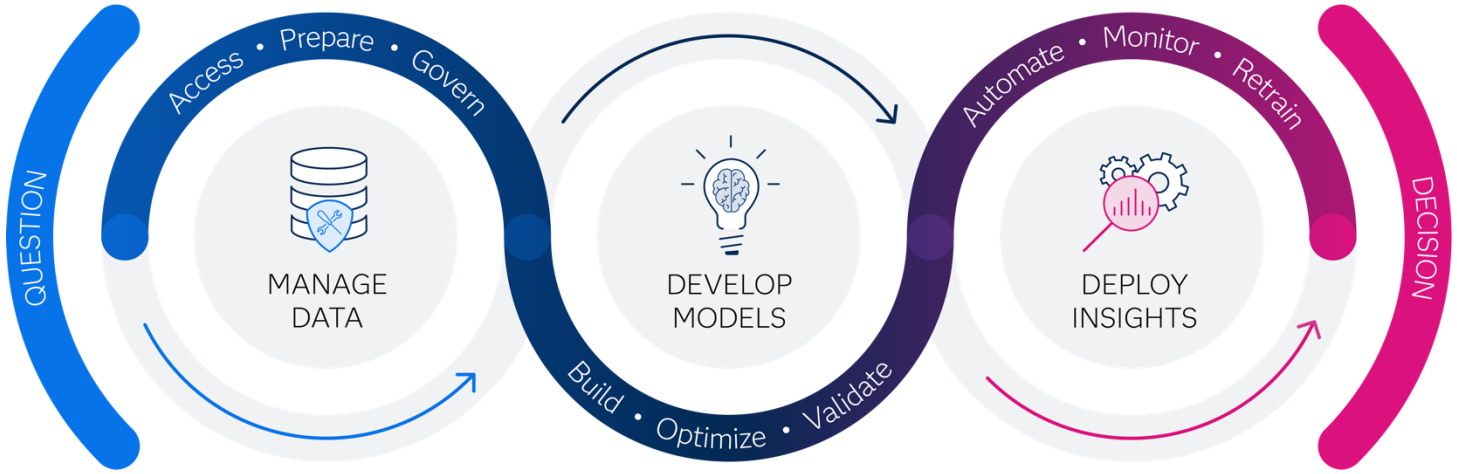
Check Data Quality

Generate ETL

Generate Dashboard

Explain & Predict

Test & Validate Prompts



Generate SAS Code

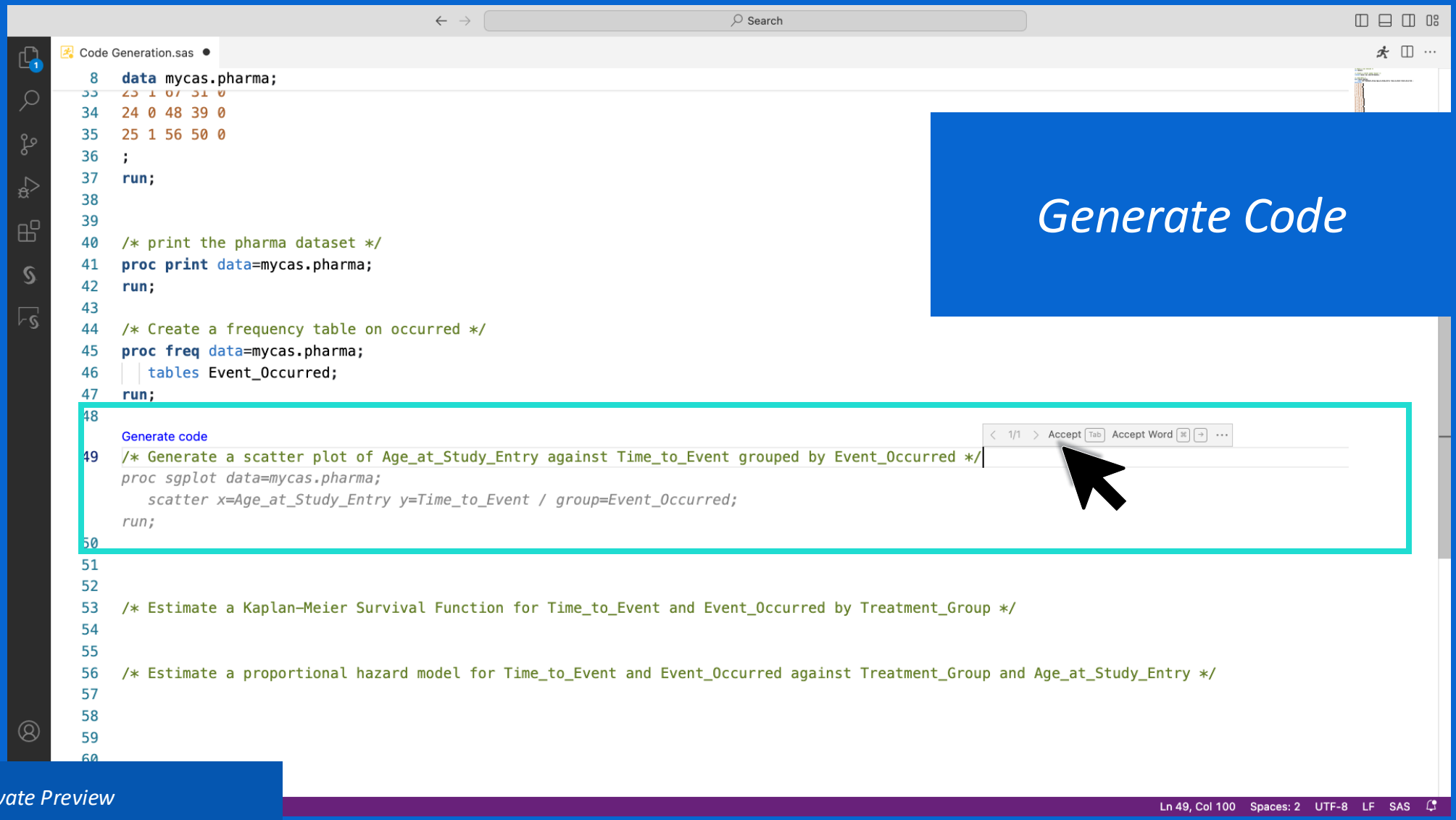
Model Reporting & Assessment

Tune Rules

Generate Compliance Report

Proof of concept
Subject to change

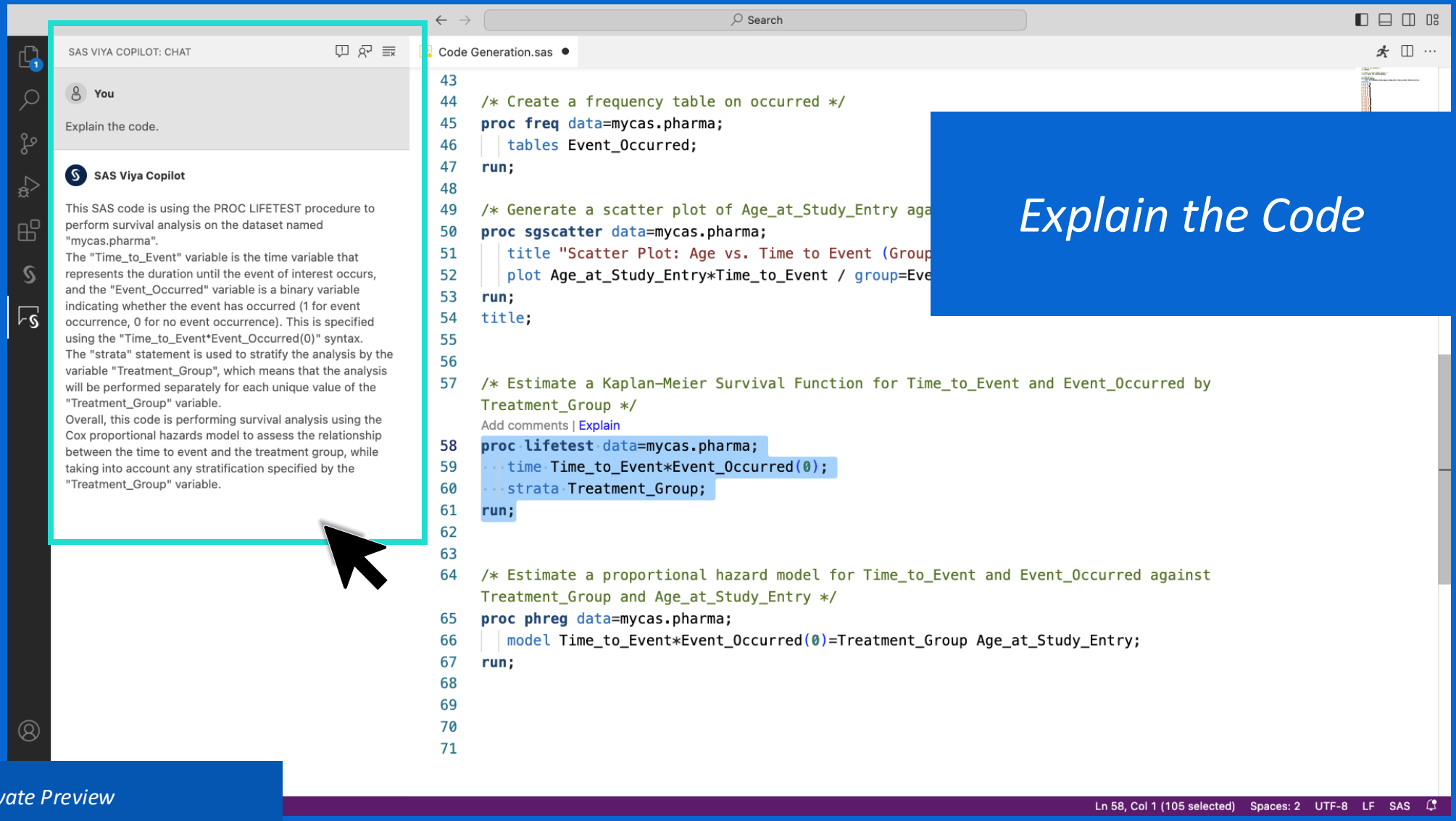
SAS Code Generation



Generate Code

Private Preview

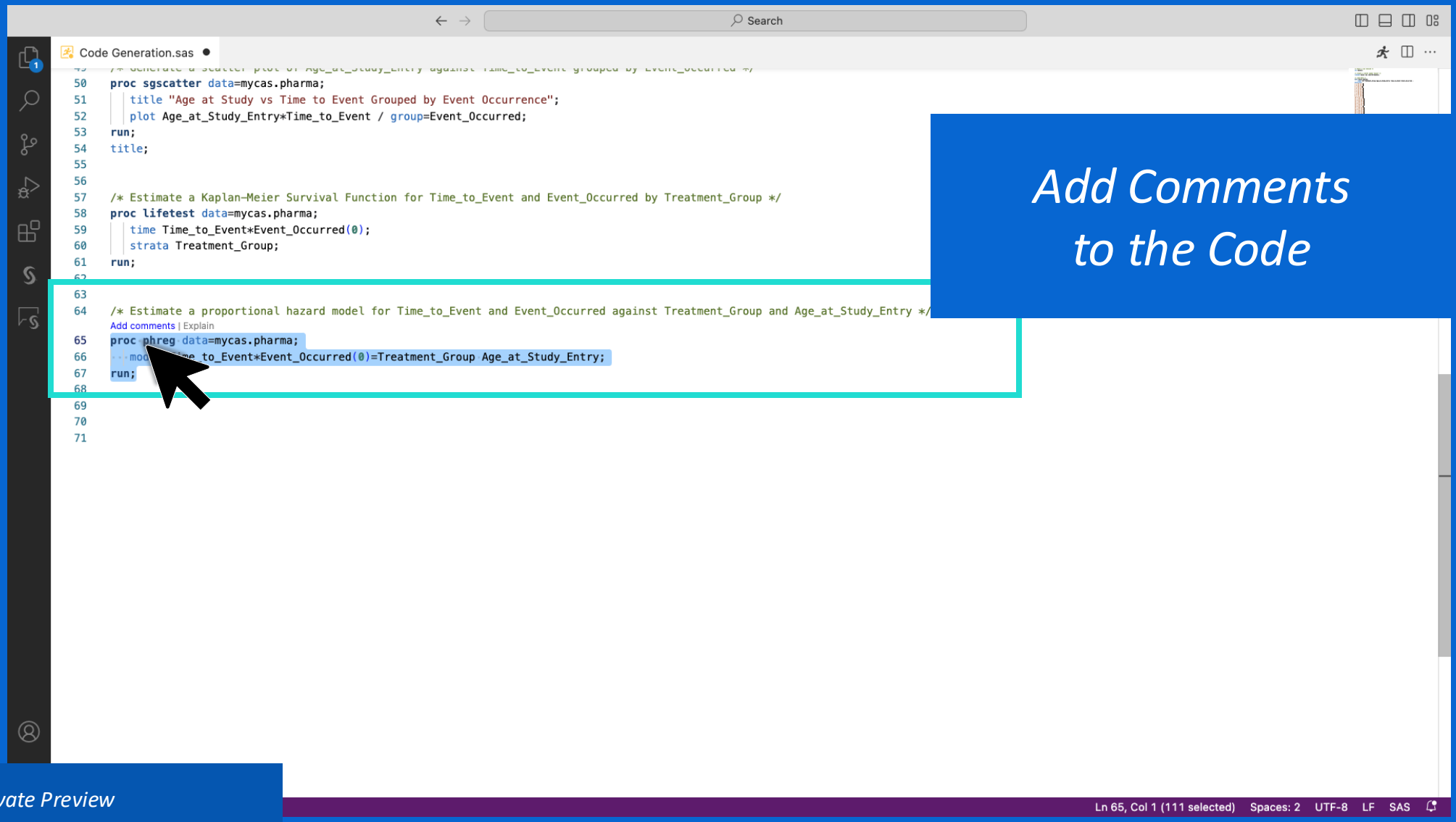
SAS Code Generation



Explain the Code

Private Preview

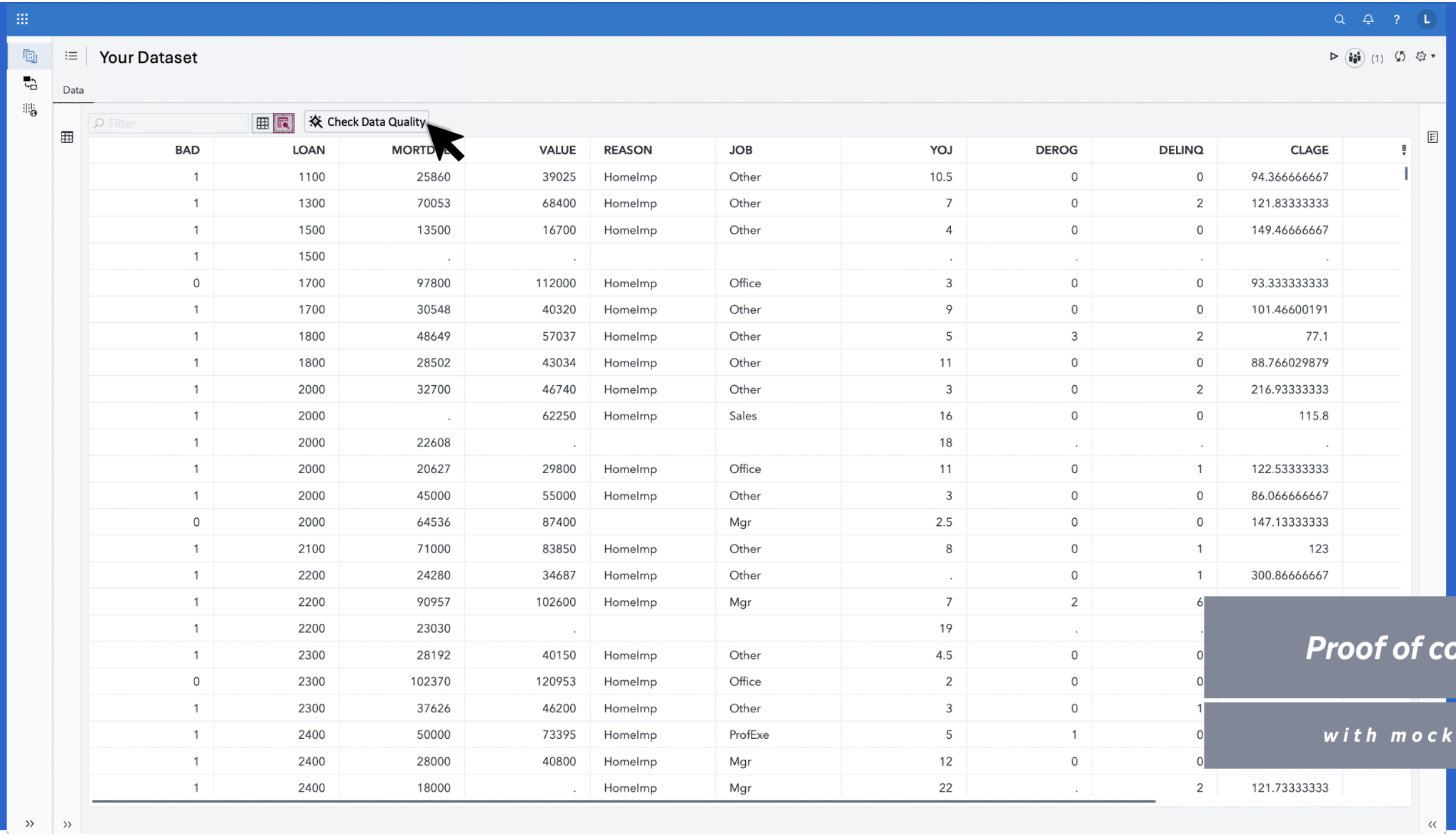
SAS Code Generation



Add Comments to the Code

Private Preview

Prep & Analyze Data Feature



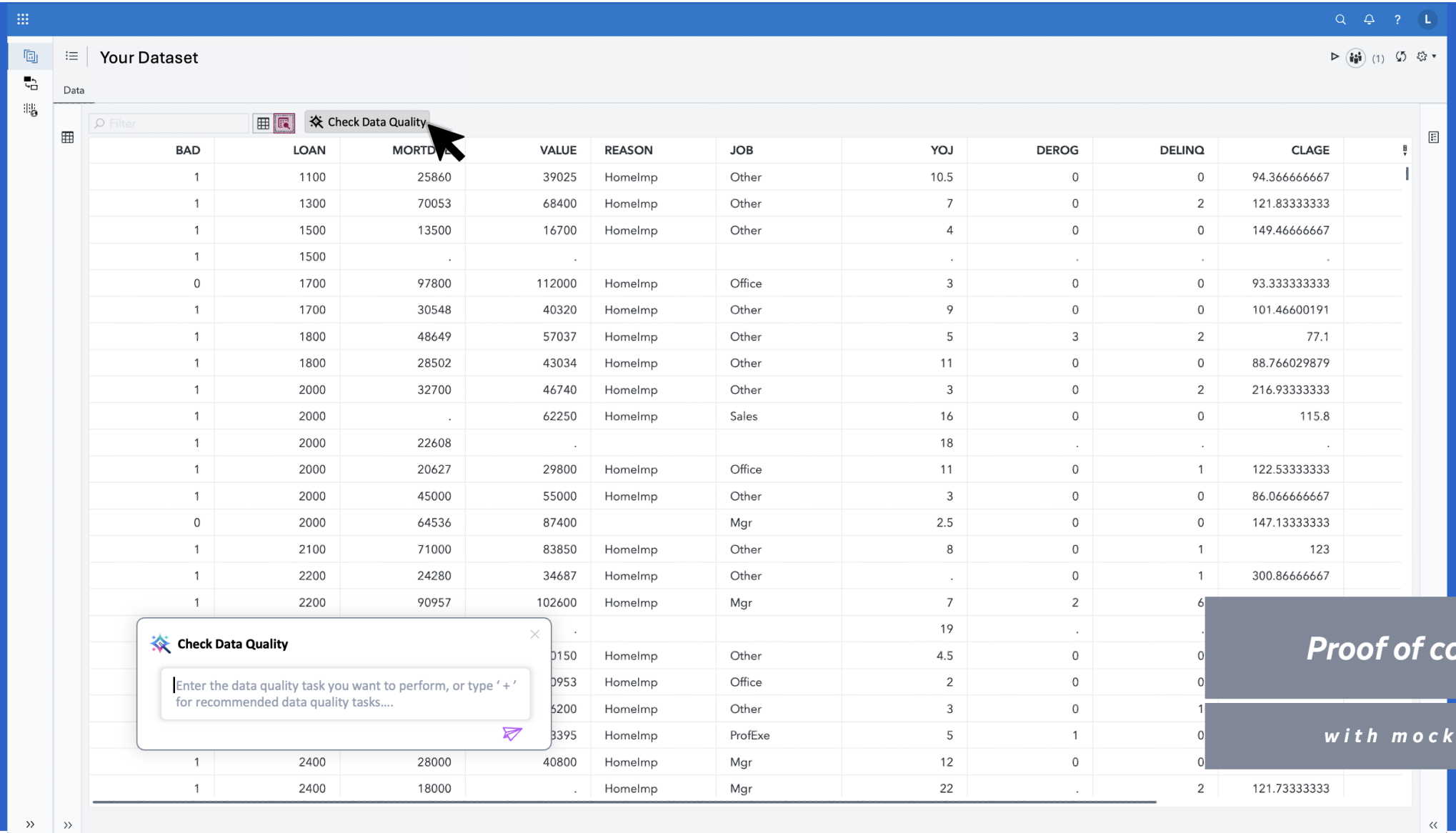
BAD	LOAN	MORTDUE	VALUE	REASON	JOB	YOJ	DEROG	DELINQ	CLAGE
1	1100	25860	39025	HomeImp	Other	10.5	0	0	94.366666667
1	1300	70053	68400	HomeImp	Other	7	0	2	121.83333333
1	1500	13500	16700	HomeImp	Other	4	0	0	149.46666667
1	1500	.	.			.	.	.	.
0	1700	97800	112000	HomeImp	Office	3	0	0	93.333333333
1	1700	30548	40320	HomeImp	Other	9	0	0	101.46600191
1	1800	48649	57037	HomeImp	Other	5	3	2	77.1
1	1800	28502	43034	HomeImp	Other	11	0	0	88.766029879
1	2000	32700	46740	HomeImp	Other	3	0	2	216.93333333
1	2000	.	62250	HomeImp	Sales	16	0	0	115.8
1	2000	22608	.			18	.	.	.
1	2000	20627	29800	HomeImp	Office	11	0	1	122.53333333
1	2000	45000	55000	HomeImp	Other	3	0	0	86.066666667
0	2000	64536	87400		Mgr	2.5	0	0	147.13333333
1	2100	71000	83850	HomeImp	Other	8	0	1	123
1	2200	24280	34687	HomeImp	Other	.	0	1	300.86666667
1	2200	90957	102600	HomeImp	Mgr	7	2	6	
1	2200	23030	.			19	.	.	
1	2300	28192	40150	HomeImp	Other	4.5	0	0	
0	2300	102370	120953	HomeImp	Office	2	0	0	
1	2300	37626	46200	HomeImp	Other	3	0	1	
1	2400	50000	73395	HomeImp	ProfExe	5	1	0	
1	2400	28000	40800	HomeImp	Mgr	12	0	0	
1	2400	18000	.	HomeImp	Mgr	22	.	2	121.73333333

Proof of concept

with mock data

Prep & Analyze Data Feature

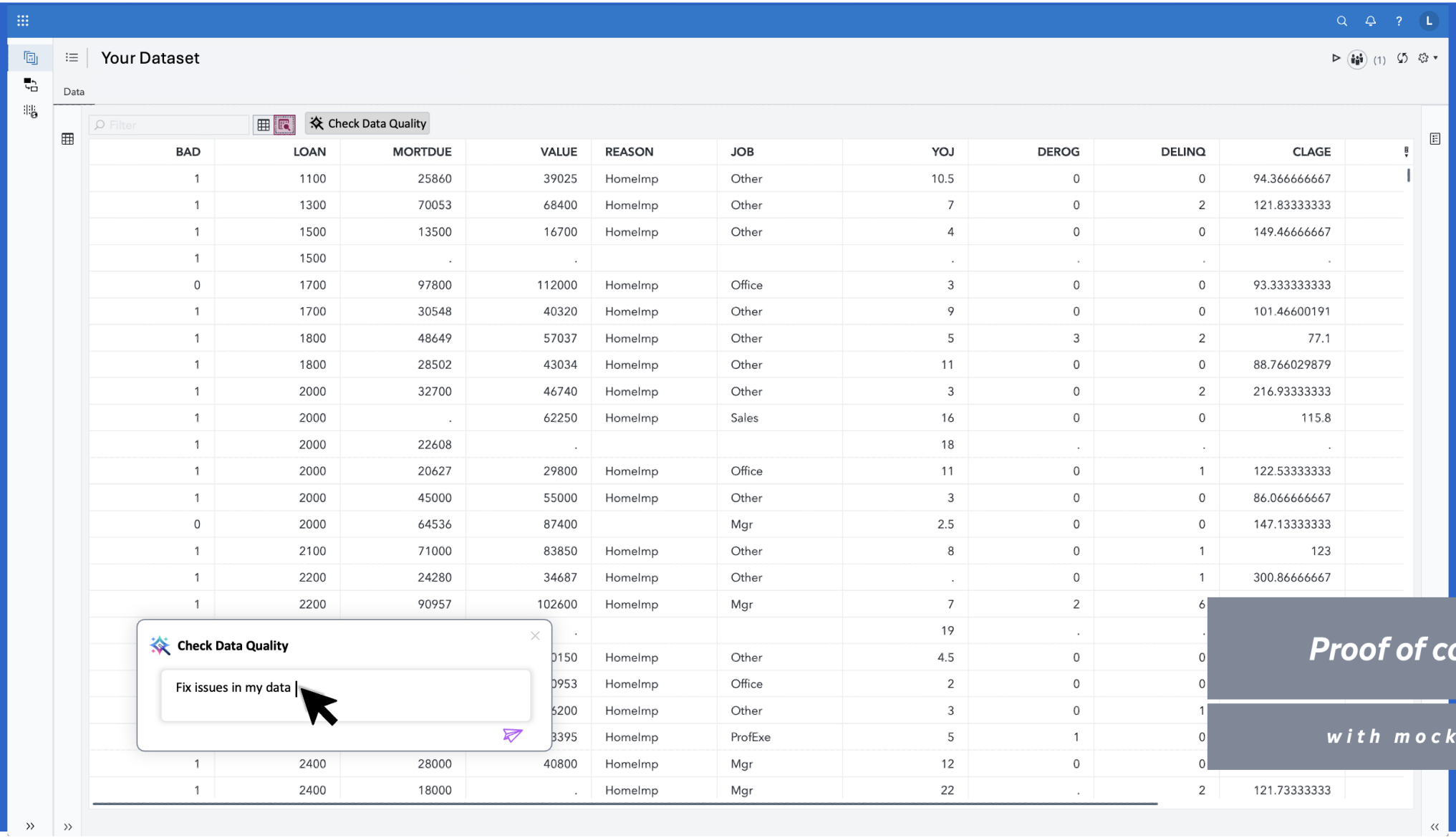
SAS® Viya® Copilot



Proof of concept

with mock data

Prep & Analyze Data Feature

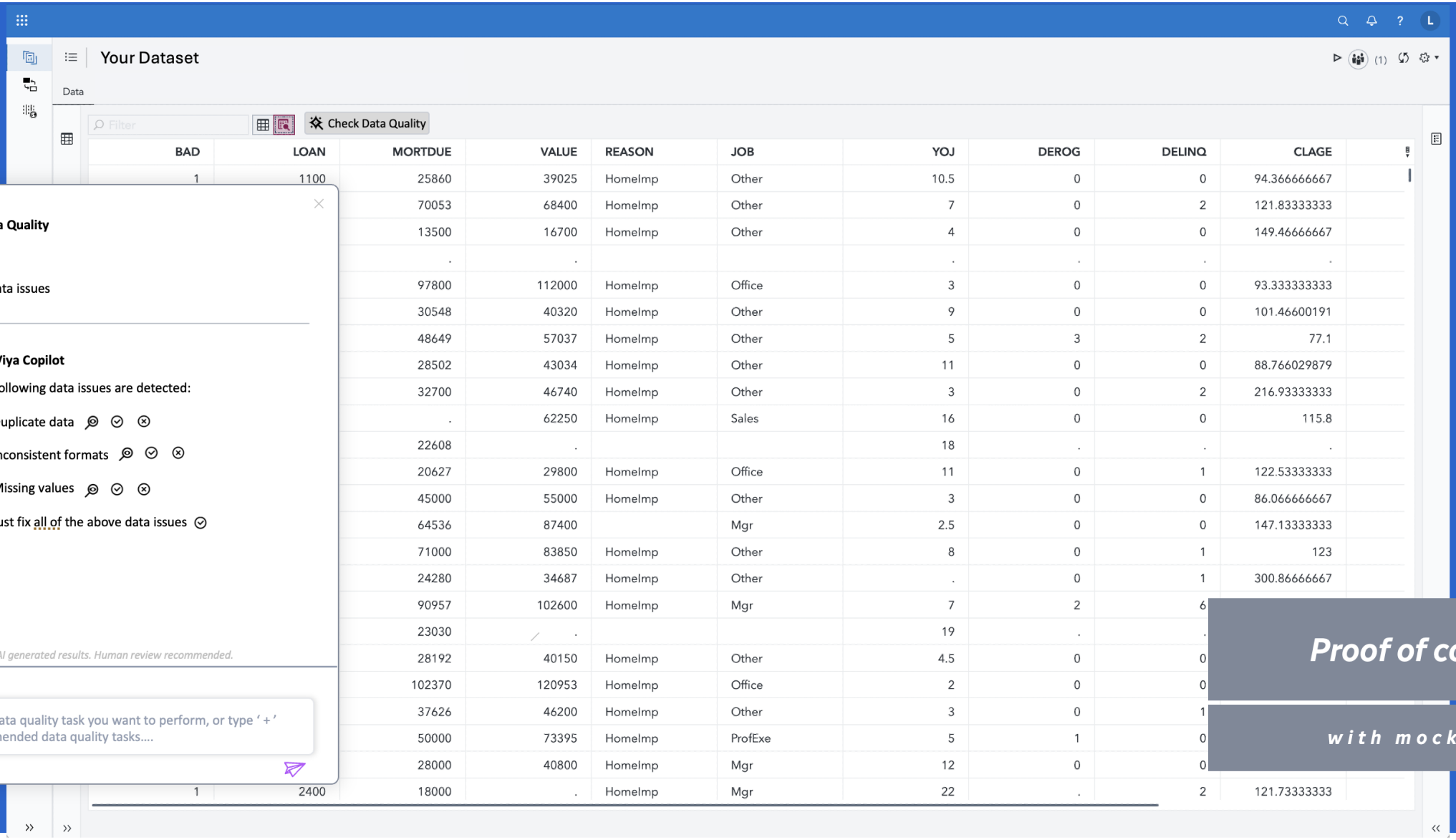


Proof of concept

with mock data

Prep & Analyze Data Feature

SAS® Viya® Copilot



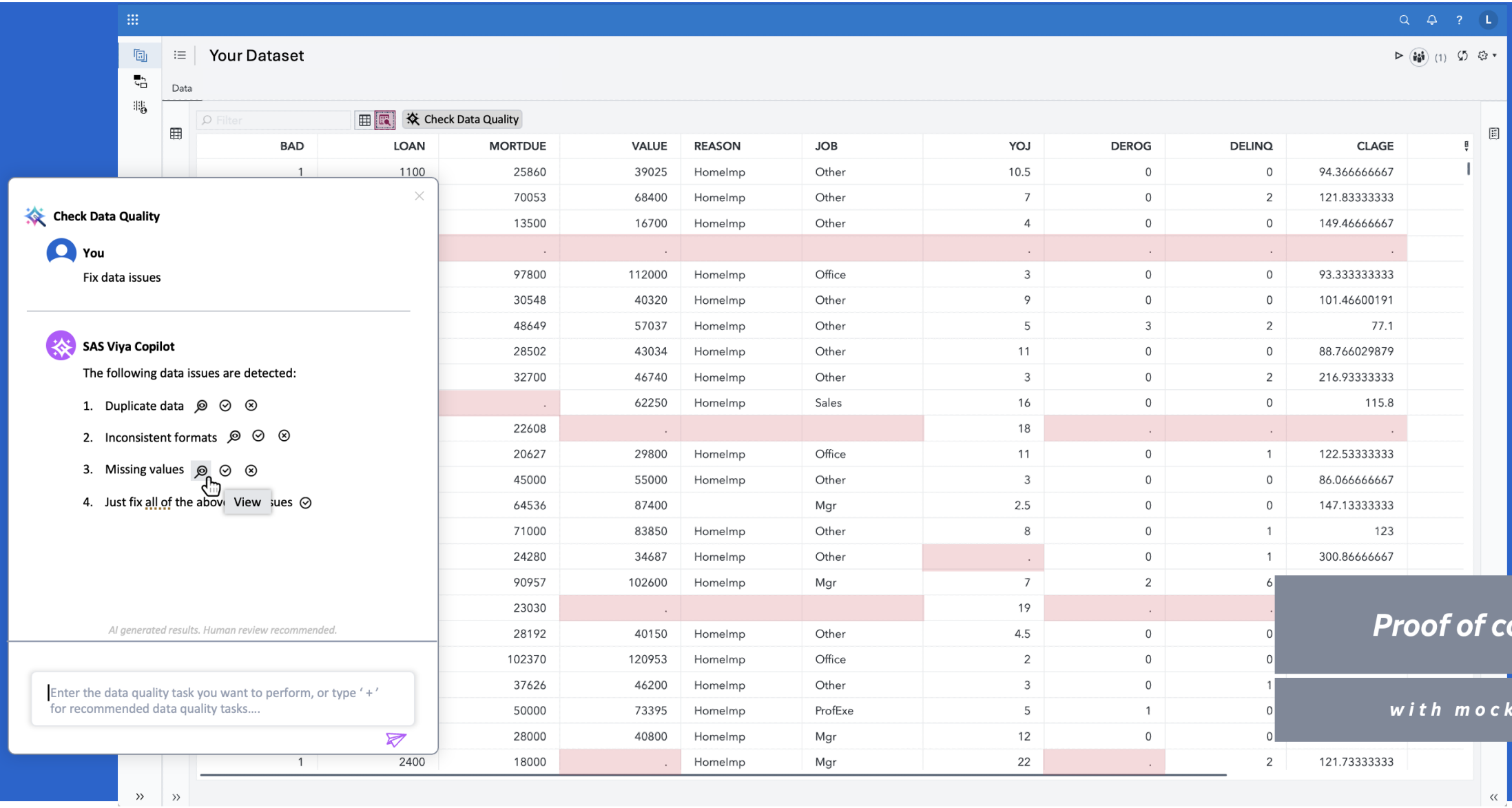
BAD	LOAN	MORTDUE	VALUE	REASON	JOB	YOJ	DEROG	DELINQ	CLAGE
1	1100	25860	39025	HomeImp	Other	10.5	0	0	94.366666667
		70053	68400	HomeImp	Other	7	0	2	121.83333333
		13500	16700	HomeImp	Other	4	0	0	149.46666667
		.	.			.	.	.	.
		97800	112000	HomeImp	Office	3	0	0	93.333333333
		30548	40320	HomeImp	Other	9	0	0	101.46600191
		48649	57037	HomeImp	Other	5	3	2	77.1
		28502	43034	HomeImp	Other	11	0	0	88.766029879
		32700	46740	HomeImp	Other	3	0	2	216.93333333
		.	62250	HomeImp	Sales	16	0	0	115.8
		22608	.			18	.	.	.
		20627	29800	HomeImp	Office	11	0	1	122.53333333
		45000	55000	HomeImp	Other	3	0	0	86.066666667
		64536	87400		Mgr	2.5	0	0	147.13333333
		71000	83850	HomeImp	Other	8	0	1	123
		24280	34687	HomeImp	Other	.	0	1	300.86666667
		90957	102600	HomeImp	Mgr	7	2	6	
		23030	.			19	.	.	
		28192	40150	HomeImp	Other	4.5	0	0	
		102370	120953	HomeImp	Office	2	0	0	
		37626	46200	HomeImp	Other	3	0	1	
		50000	73395	HomeImp	ProfExe	5	1	0	
		28000	40800	HomeImp	Mgr	12	0	0	
1	2400	18000	.	HomeImp	Mgr	22	.	2	121.73333333

Proof of concept

with mock data

Prep & Analyze Data Feature

SAS® Viya® Copilot

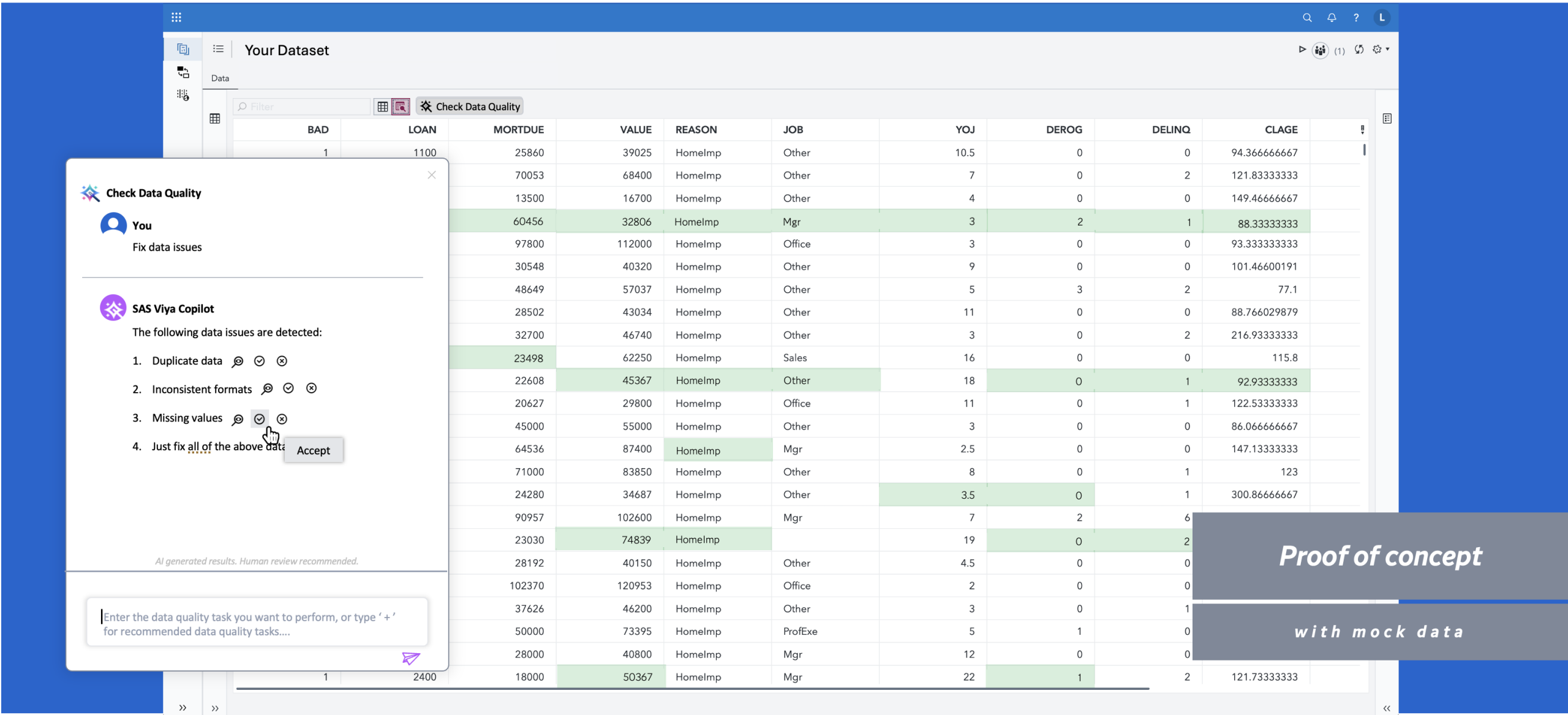


BAD	LOAN	MORTDUE	VALUE	REASON	JOB	YOJ	DEROG	DELINQ	CLAGE
1	1100	25860	39025	HomeImp	Other	10.5	0	0	94.366666667
		70053	68400	HomeImp	Other	7	0	2	121.83333333
		13500	16700	HomeImp	Other	4	0	0	149.46666667
		.	.			.	.	.	.
		97800	112000	HomeImp	Office	3	0	0	93.333333333
		30548	40320	HomeImp	Other	9	0	0	101.46600191
		48649	57037	HomeImp	Other	5	3	2	77.1
		28502	43034	HomeImp	Other	11	0	0	88.766029879
		32700	46740	HomeImp	Other	3	0	2	216.93333333
		.	62250	HomeImp	Sales	16	0	0	115.8
		22608	.			18	.	.	.
		20627	29800	HomeImp	Office	11	0	1	122.53333333
		45000	55000	HomeImp	Other	3	0	0	86.066666667
		64536	87400		Mgr	2.5	0	0	147.13333333
		71000	83850	HomeImp	Other	8	0	1	123
		24280	34687	HomeImp	Other	.	0	1	300.86666667
		90957	102600	HomeImp	Mgr	7	2	6	
		23030	.			19	.	.	
		28192	40150	HomeImp	Other	4.5	0	0	
		102370	120953	HomeImp	Office	2	0	0	
		37626	46200	HomeImp	Other	3	0	1	
		50000	73395	HomeImp	ProfExe	5	1	0	
		28000	40800	HomeImp	Mgr	12	0	0	
1	2400	18000	.	HomeImp	Mgr	22	.	2	121.73333333

Proof of concept

with mock data

Prep & Analyze Data Feature



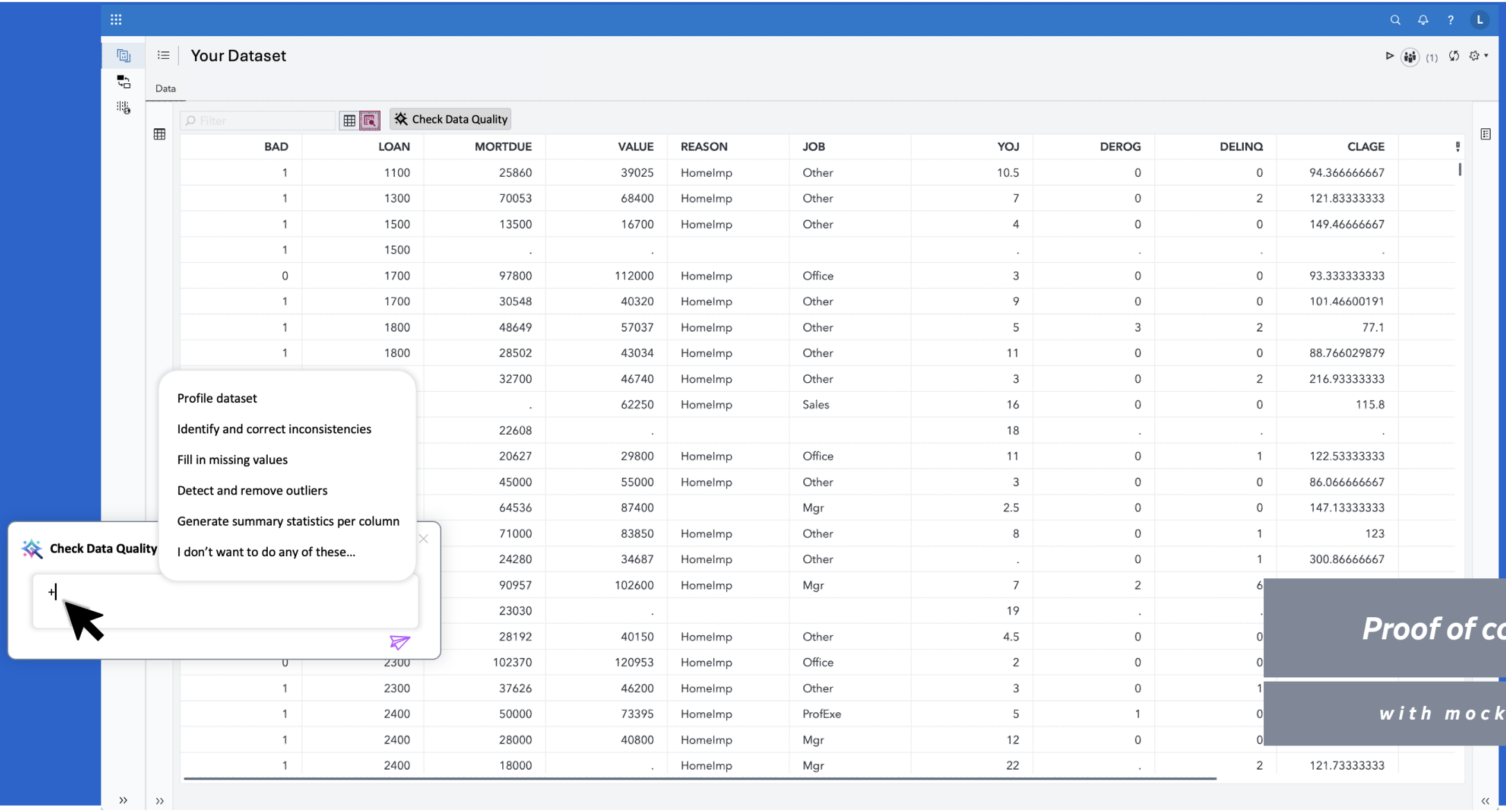
BAD	LOAN	MORTDUE	VALUE	REASON	JOB	YOJ	DEROG	DELINQ	CLAGE
1	1100	25860	39025	HomeImp	Other	10.5	0	0	94.366666667
		70053	68400	HomeImp	Other	7	0	2	121.83333333
		13500	16700	HomeImp	Other	4	0	0	149.46666667
		60456	32806	HomeImp	Mgr	3	2	1	88.33333333
		97800	112000	HomeImp	Office	3	0	0	93.333333333
		30548	40320	HomeImp	Other	9	0	0	101.46600191
		48649	57037	HomeImp	Other	5	3	2	77.1
		28502	43034	HomeImp	Other	11	0	0	88.766029879
		32700	46740	HomeImp	Other	3	0	2	216.93333333
		23498	62250	HomeImp	Sales	16	0	0	115.8
		22608	45367	HomeImp	Other	18	0	1	92.93333333
		20627	29800	HomeImp	Office	11	0	1	122.53333333
		45000	55000	HomeImp	Other	3	0	0	86.066666667
		64536	87400	HomeImp	Mgr	2.5	0	0	147.13333333
		71000	83850	HomeImp	Other	8	0	1	123
		24280	34687	HomeImp	Other	3.5	0	1	300.86666667
		90957	102600	HomeImp	Mgr	7	2	6	
		23030	74839	HomeImp		19	0	2	
		28192	40150	HomeImp	Other	4.5	0	0	
		102370	120953	HomeImp	Office	2	0	0	
		37626	46200	HomeImp	Other	3	0	1	
		50000	73395	HomeImp	ProfExe	5	1	0	
		28000	40800	HomeImp	Mgr	12	0	0	
1	2400	18000	50367	HomeImp	Mgr	22	1	2	121.73333333

Proof of concept

with mock data

Prep & Analyze Data Feature

SAS® Viya® Copilot



BAD	LOAN	MORTDUE	VALUE	REASON	JOB	YOJ	DEROG	DELINQ	CLAGE
1	1100	25860	39025	HomeImp	Other	10.5	0	0	94.366666667
1	1300	70053	68400	HomeImp	Other	7	0	2	121.83333333
1	1500	13500	16700	HomeImp	Other	4	0	0	149.46666667
1	1500	.	.			.	.	.	.
0	1700	97800	112000	HomeImp	Office	3	0	0	93.333333333
1	1700	30548	40320	HomeImp	Other	9	0	0	101.46600191
1	1800	48649	57037	HomeImp	Other	5	3	2	77.1
1	1800	28502	43034	HomeImp	Other	11	0	0	88.766029879
		32700	46740	HomeImp	Other	3	0	2	216.93333333
		.	62250	HomeImp	Sales	16	0	0	115.8
		22608	.			18	.	.	.
		20627	29800	HomeImp	Office	11	0	1	122.53333333
		45000	55000	HomeImp	Other	3	0	0	86.066666667
		64536	87400		Mgr	2.5	0	0	147.13333333
		71000	83850	HomeImp	Other	8	0	1	123
		24280	34687	HomeImp	Other	.	0	1	300.86666667
		90957	102600	HomeImp	Mgr	7	2	6	
		23030	.			19	.	.	
		28192	40150	HomeImp	Other	4.5	0	0	
0	2300	102370	120953	HomeImp	Office	2	0	0	
1	2300	37626	46200	HomeImp	Other	3	0	1	
1	2400	50000	73395	HomeImp	ProfExe	5	1	0	
1	2400	28000	40800	HomeImp	Mgr	12	0	0	
1	2400	18000	.	HomeImp	Mgr	22	.	2	121.73333333

Proof of concept

with mock data

Generative AI in Action with SAS

SAS Viya

Explain, Govern, Orchestrate LLMs

Viya Copilots

Accelerate AI Lifecycle / Industry / Business Tasks

Synthetic Data Generation

Mitigate Data Quality and Scarcity

hazy

sas Software Learn Support Partners About Us

SAS acquires Hazy synthetic data software to boost generative AI portfolio

Strategic move reinforces SAS leadership in data and AI, enhancing customer innovation and data security

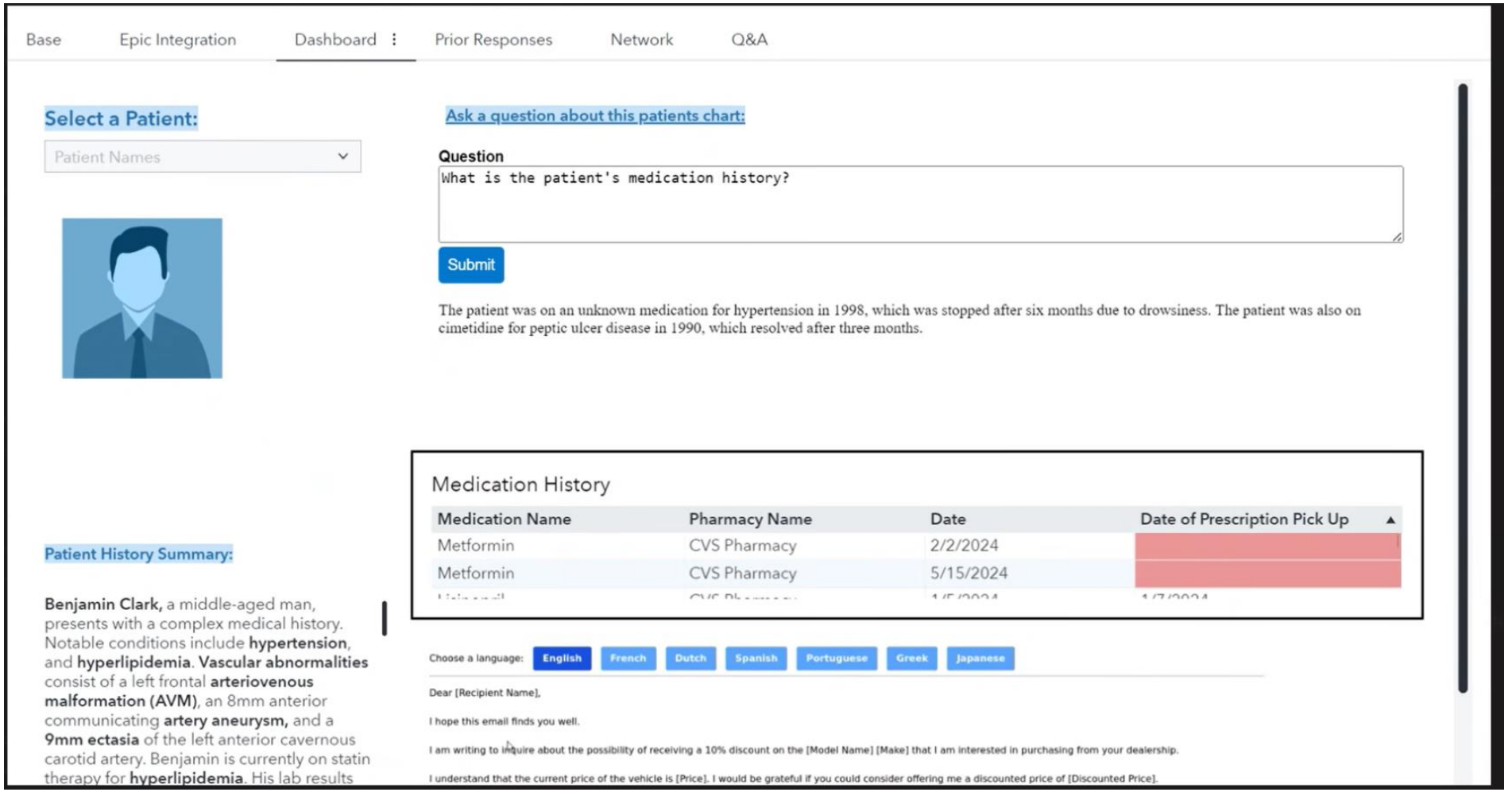
Copilot Use Cases – 2025 Roadmap

HLS

#	Use Case	User Story
1	**Patient cohort creation**	**As a**: Clinical Researcher **I want**: to use natural language processing (NLP) to generate SQL queries for creating patient cohorts **So that**: I can efficiently and accurately create patient cohorts without needing extensive SQL knowledge
2	**Clinical record summary**	**As a**: Healthcare Provider **I want**: to use natural language processing (NLP) to generate concise summaries of clinical records from structured clinical notes data **So that**: I can quickly understand patient history and current condition without reading through extensive clinical notes
3	**SAS Code Creation in SCE**	**As a**: Biostatistician **I want**: to use a SAS Viya Copilot to generate SAS code for biostatistical analysis in clinical trials **So that**: I can efficiently perform complex statistical analyses without manually writing extensive SAS code
4	**Synthetic SDTM data creation**	**As a**: Data Manager **I want**: to use Data Maker to generate synthetic Study Data Tabulation Model (SDTM) data **So that**: I can create realistic datasets for testing and validation without compromising patient privacy

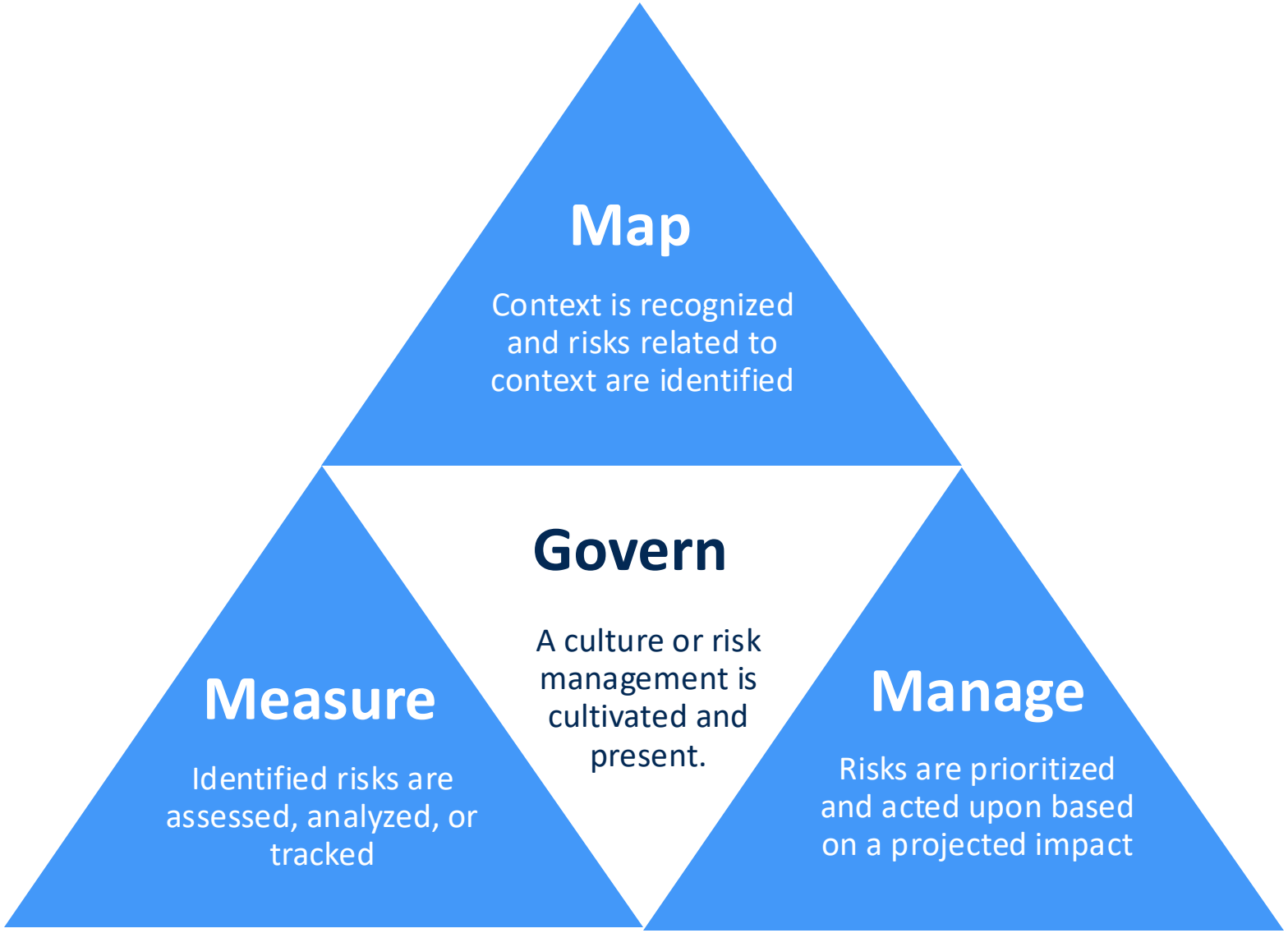
Real-world use cases for Copilot

Unlocking new possibilities with the latest SAS Health features

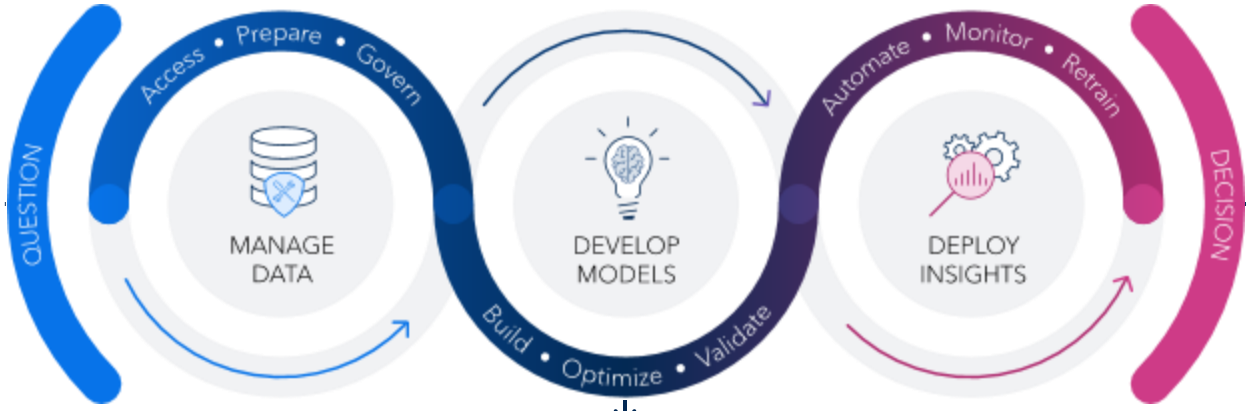


What is Trustworthy AI?

AI Risk Management Framework



SAS Viya Trustworthy AI capabilities

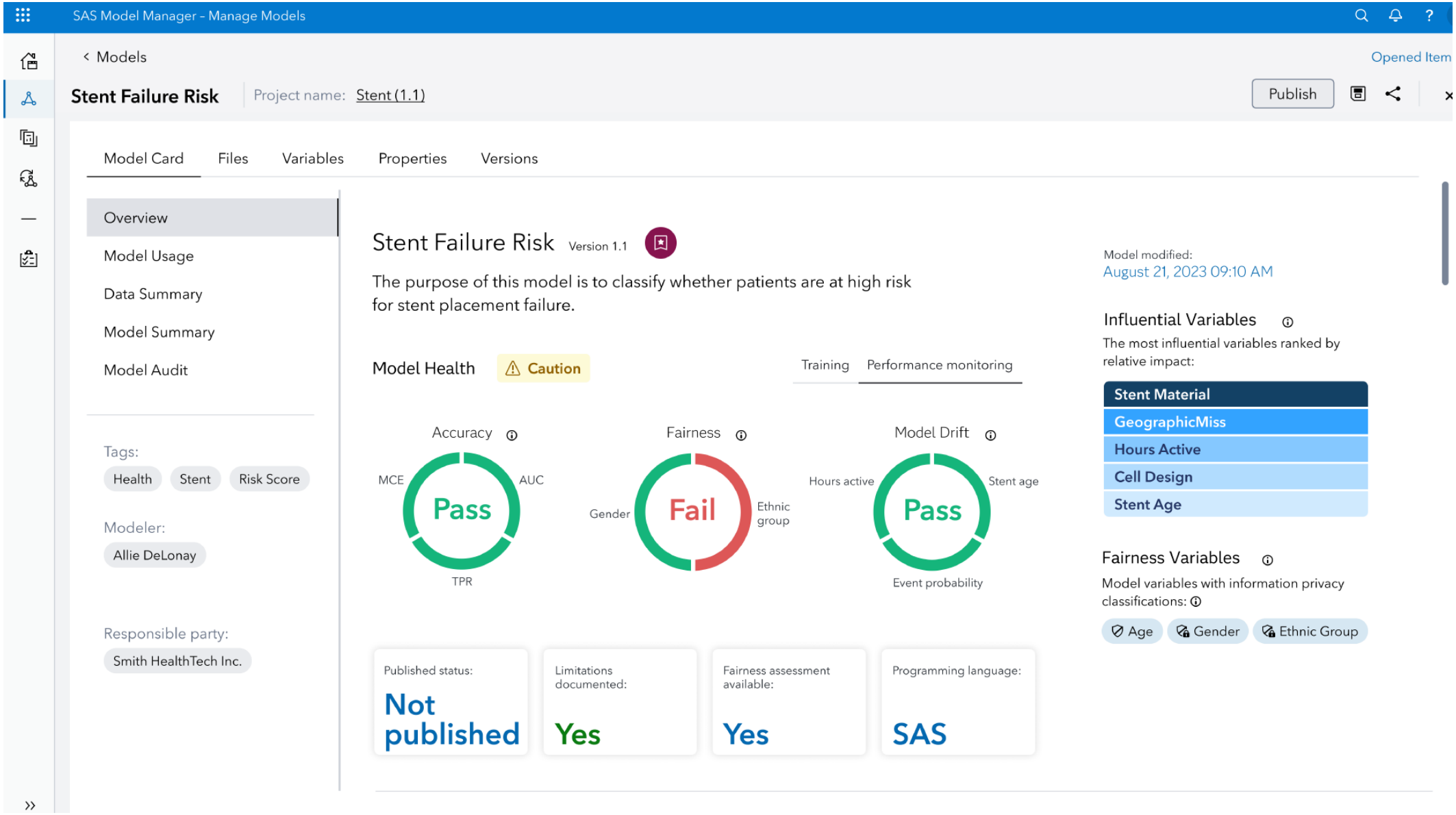


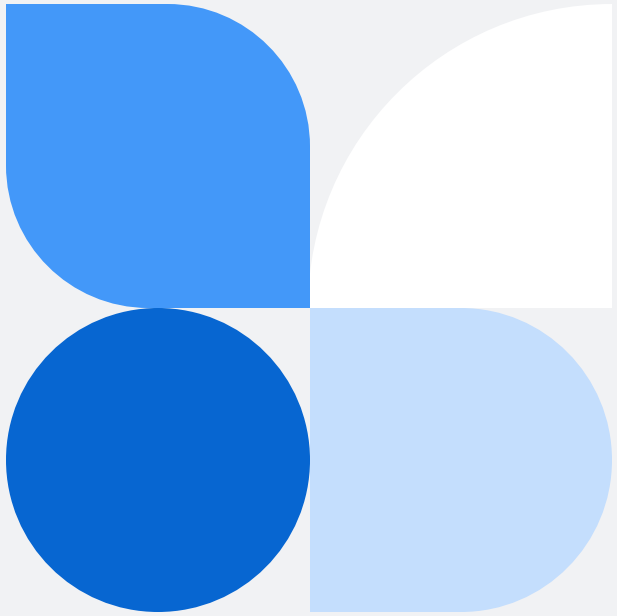
- Data Quality
- Data Exploration
- Information Privacy
- Data Masking
- Data Suppression
- Data Lineage
- Synthetic Data Generation
- Semantic Type Remediation

- Natural Language Insights
- Model Interpretability
- Fairness Assessment
- Bias Mitigation

- Model Governance
- Model Monitoring
- Decision Accountability
- Model Cards*

Trustworthy AI Workflow





Benefits for Generative AI

Accelerated Innovation

Quickly transform LLMs into actionable insights by seamlessly integrating GenAI models into decisioning workflows, AI/ML applications and existing business processes

Data Protection

Uphold user privacy and security with robust data protection measures, including data minimization, anonymization and encryption, ensuring that sensitive information remains safeguarded

Trustworthy Results

Apply SAS natural processing techniques to preprocess data, ensuring only high-quality data are fed to LLMs, minimizing computational waste, reducing costs and producing reliable outcomes

Enhanced Governance

Use built-in workflows that validate the entire life cycle of LLMs, from regulatory compliance to model risk management. Additionally, SAS offers responsible ideation, experimentation and operationalization support

Precise Decisioning

Quantitative decisioning capabilities, critical for successful GenAI reasoning, are built into the Viya platform.

Thank you