SAS User Group – Quebec City

Trustworthy AI Using SAS®: A Technical Discussion

Vrushali Sawant, Data Scientist, Data Ethics Practice

Vrushali Sawant

Data Scientist, Data Ethics Practice

Vrushali Sawant is a data scientist with the SAS Data Ethics Practice (DEP), steering the practical implementation of fairness and trustworthy principles into the SAS platform. She leads the DEP's marketing initiatives from a responsible innovation perspective. With a background in technology consulting, data management and data visualization, she has been helping customers make data-driven decisions for a decade.

Trustworthy AI: A Technical Discussion

Vrushali Sawant
Data Scientist, Data Ethics Practice

Agenda

- Introduction to Trustworthy AI
- Principles based approach to Trustworthy AI
- Trustworthy AI for fair business decisions
- From Principles to Practice
- Trustworthy AI across AI & Analytics Lifecycle
- Q & A’s

Introduction to Trustworthy AI

Definition of Trustworthy AI

What is Trustworthy AI?

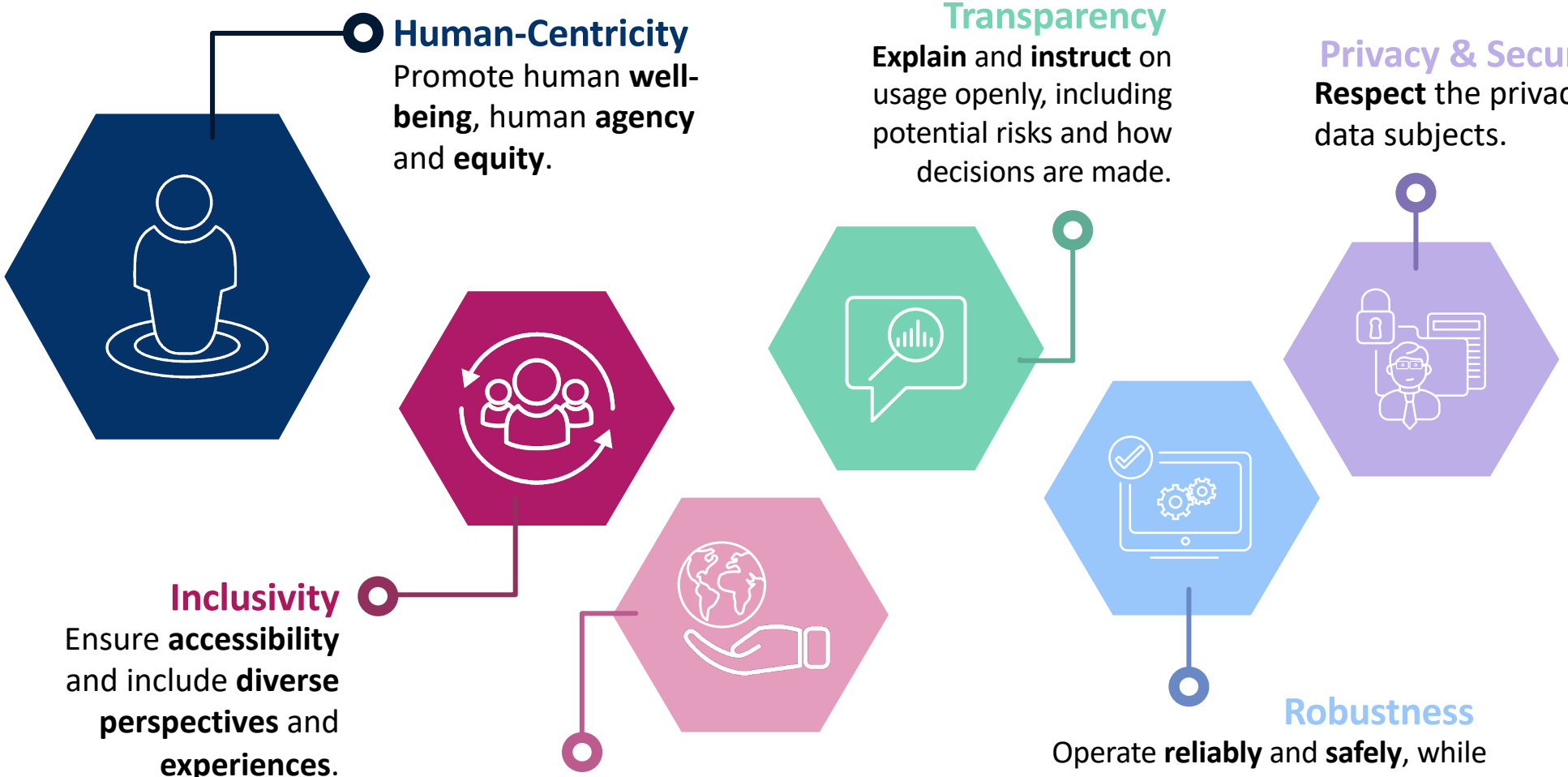
- Asking not just, “Could we?” but also, “**Should we**?”
- Ensuring AI **does not harm** people
- Building AI that **reflects our values** as a society

Trustworthy AI requires a socio-technical approach.

Principles based approach to Trustworthy AI

SAS Principles

Principles based approach for Trustworthy AI



Human-Centricity
Promote human **well-being**, human **agency** and **equity**.

Transparency
Explain and **instruct** on usage openly, including potential risks and how decisions are made.

Privacy & Security
Respect the privacy of data subjects.

Inclusivity
Ensure **accessibility** and include **diverse perspectives** and **experiences**.

Accountability
Proactively identify and mitigate adverse impacts.

Robustness
Operate **reliably** and **safely**, while enabling mechanisms that assess and manage potential risks throughout a system's lifecycle.

Trustworthy AI for Fair decisions

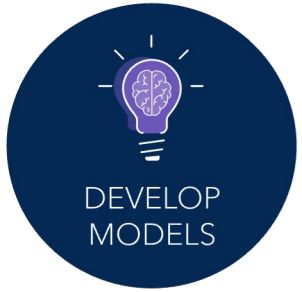
Examples from Healthcare, Banking and Government

Trustworthy AI approach for Healthcare

Triaging Patients

Gather Health Records Data → Build Triage Prediction Models → Publish and Deploy Triage Prediction Models

Availability Bias



Proxy Bias
Measurement Bias

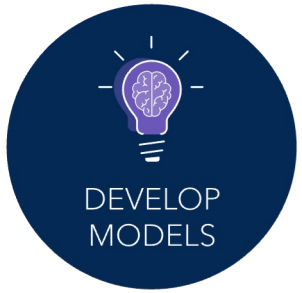
Drift Bias
Reporting Bias

Trustworthy AI approach for Government

Tax and Finance

Taxpayer data collected → Tax models developed → Decisions deployed

Historical Bias
Pre-processing Bias



Confirmation Bias
Cause-Effect Bias
Selection Bias

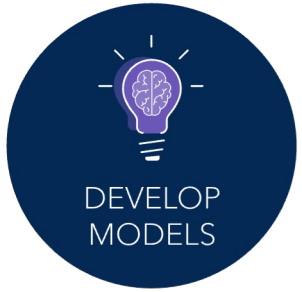
Deployment Bias
Reporting Bias

Trustworthy AI approach for Banking

Credit Lending

Gather Loan Application Data → Assess Risk for Loan → Loan Approved/Rejected

Availability Bias



Historical Bias
Prediction Bias
Exclusion Bias

Automation Bias
Drift Bias

From Principles to Practice

A Comprehensive Approach

Trustworthy AI : A Comprehensive Approach

Activation

Oversight
AI Governance, Strategy and Enforcement

Compliance
Performance and Risk Management

Operations
Operating Procedures with Infrastructure

Culture
Systems, Norms and Practices

Technology

Data Management
- Data Quality
- Variable Metadata
- Data Preparation
- Data Asset Catalog

Explanation
- Natural Language Explanation
- Explainable ML
- Counterfactual Explanation
- Surrogate Model Interpretation
- Causal Inference

Detection
- Bias Detection
- Fairness Assessment

Privacy & Security
- Privacy Preservation
- Model Security
- Autonomy Preservation
- Consent & Control

Mitigation
- Bias Mitigation
- Bias Prevention
- Synthetic Data Generation

ModelOps
- Model Cards
- Decisioning
- Lifecycle Management
- Metric Monitoring
- Model Robustness

Solutions (industry & domain)

Accessible and Intuitive UI

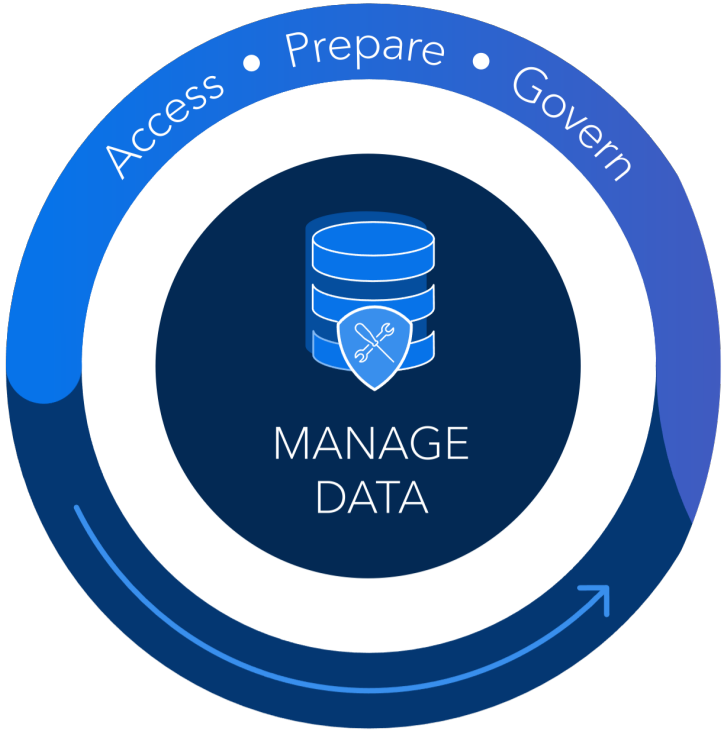
Logging & Auditability

Interoperability

Trustworthy AI Across AI & Analytics Lifecycle

SAS Viya end-to-end Trustworthy AI Capabilities

Robust Data Governance

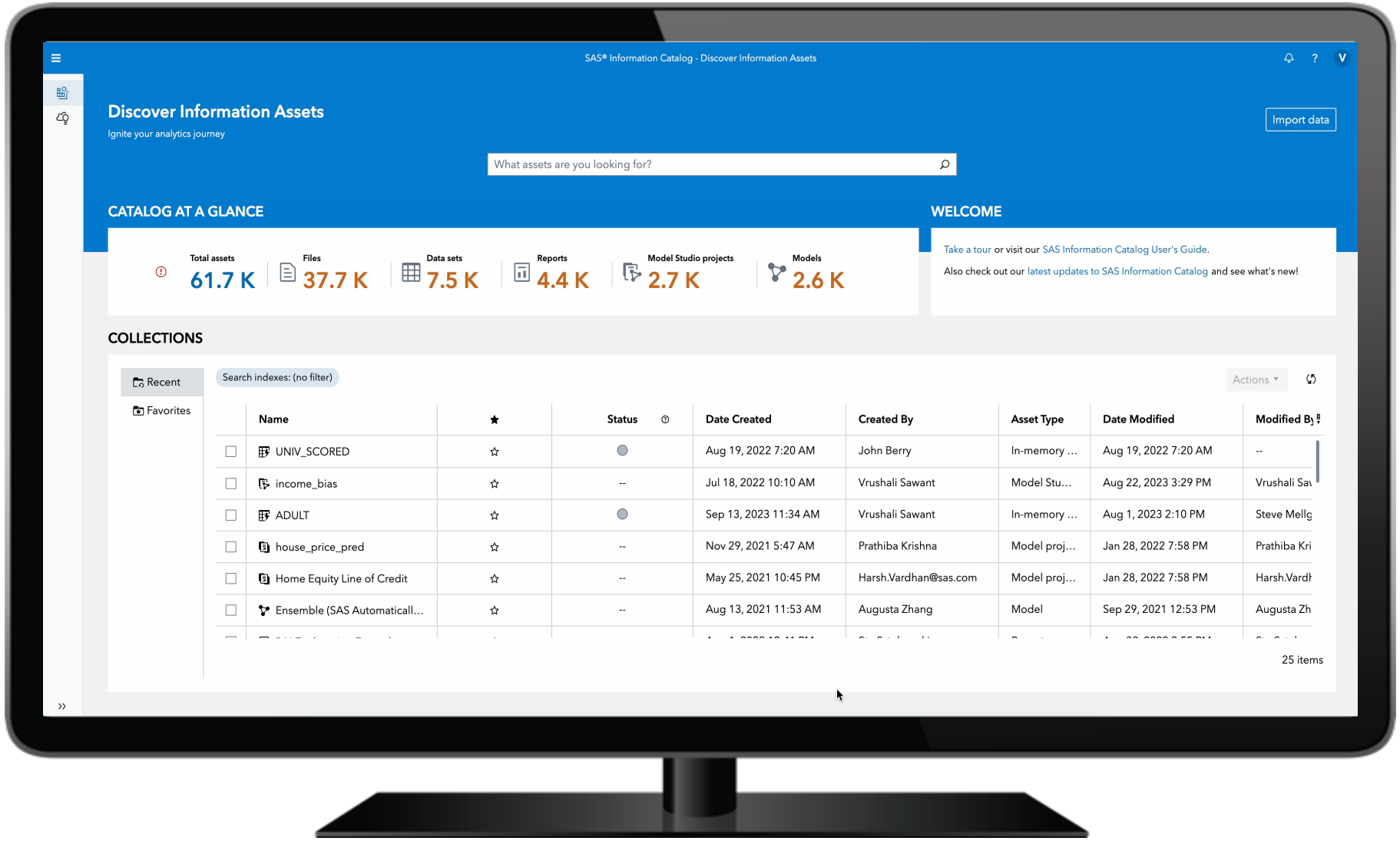


Trustworthy decisions start with good data.

- Data Quality
- Automated Data Exploration
- Information Privacy
- Data Masking
- Data Suppression
- Data Lineage
- Synthetic Data Generation

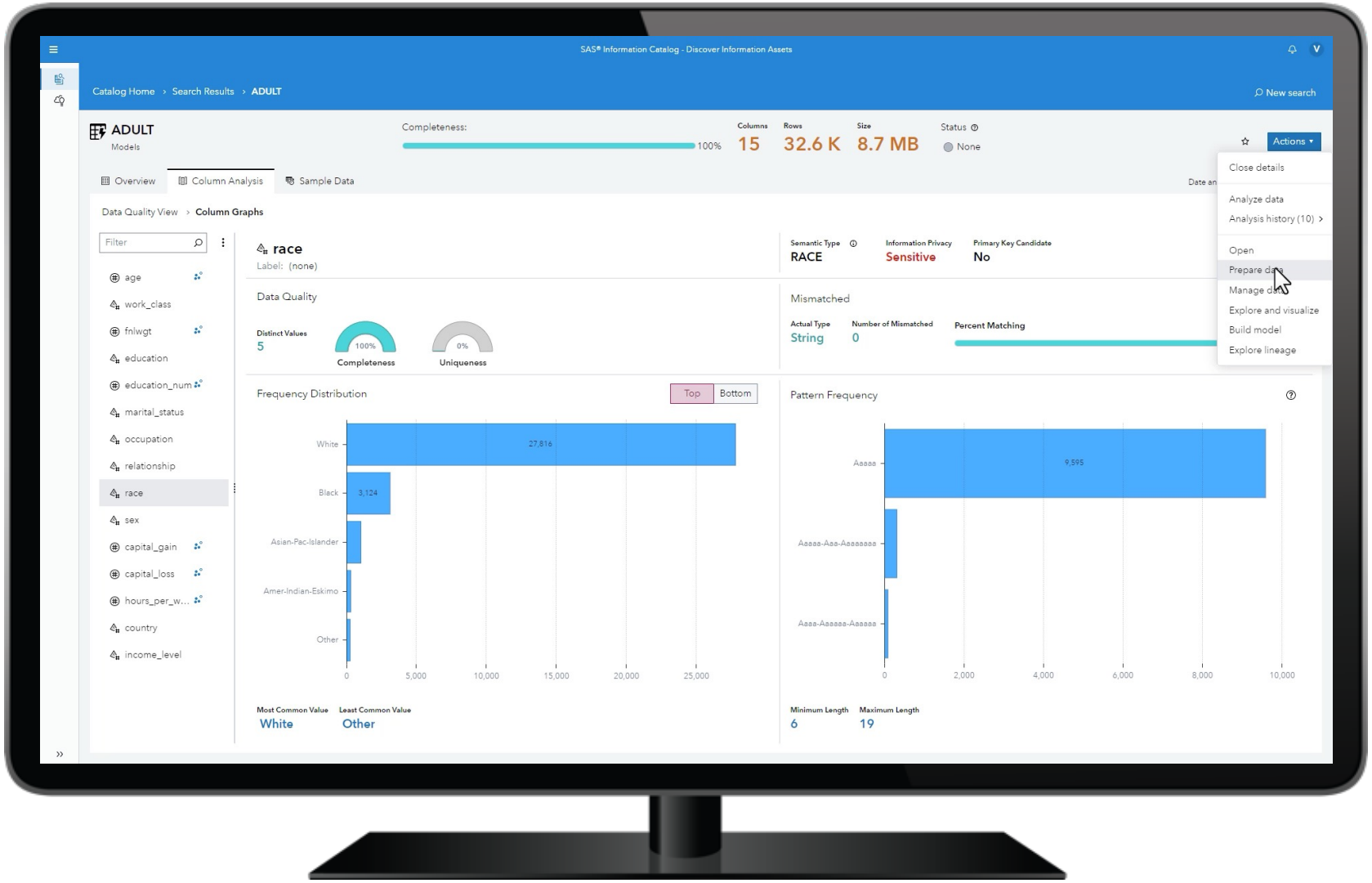
Robust Data Governance

Information Privacy & Data Quality



Robust Data Governance

Data Masking



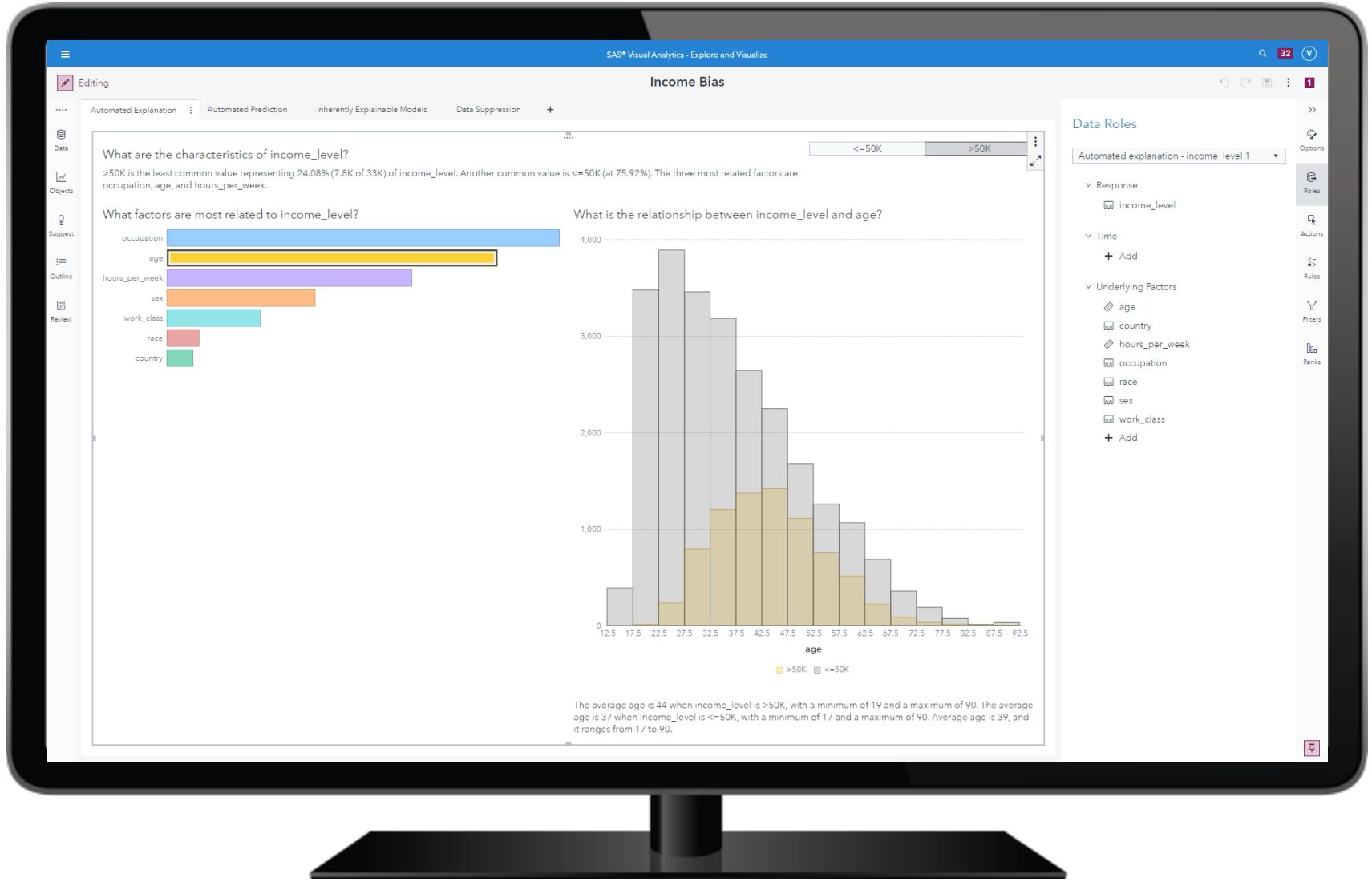
Transparent AI

AI comes with both risk and reward.

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

Transparent AI

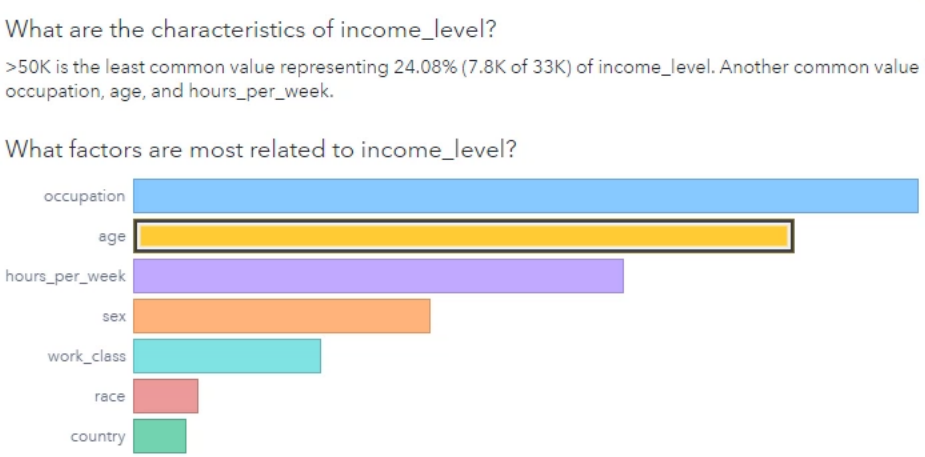
Natural Language Insights



Automated Explanation | Automated Prediction | Inherently Explainable Models | Data Suppression

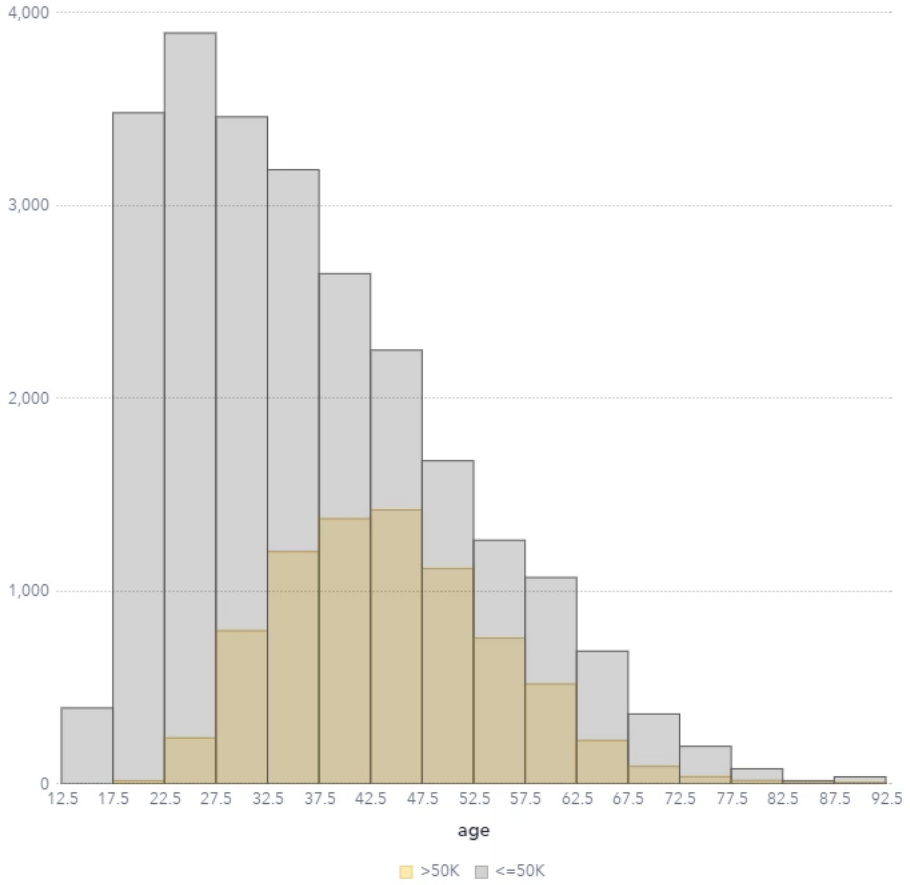
What are the characteristics of income_level?

>50K is the least common value representing 24.08% (7.8K of 33K) of income_level. Another common value is <=50K (at 75.92%). The three most related factors are occupation, age, and hours_per_week.



What factors are most related to income_level?

What is the relationship between income_level and age?



The average age is 44 when income_level is >50K, with a minimum of 19 and a maximum of 90. The average age is 37 when income_level is <=50K, with a minimum of 17 and a maximum of 90. Average age is 39, and it ranges from 17 to 90.

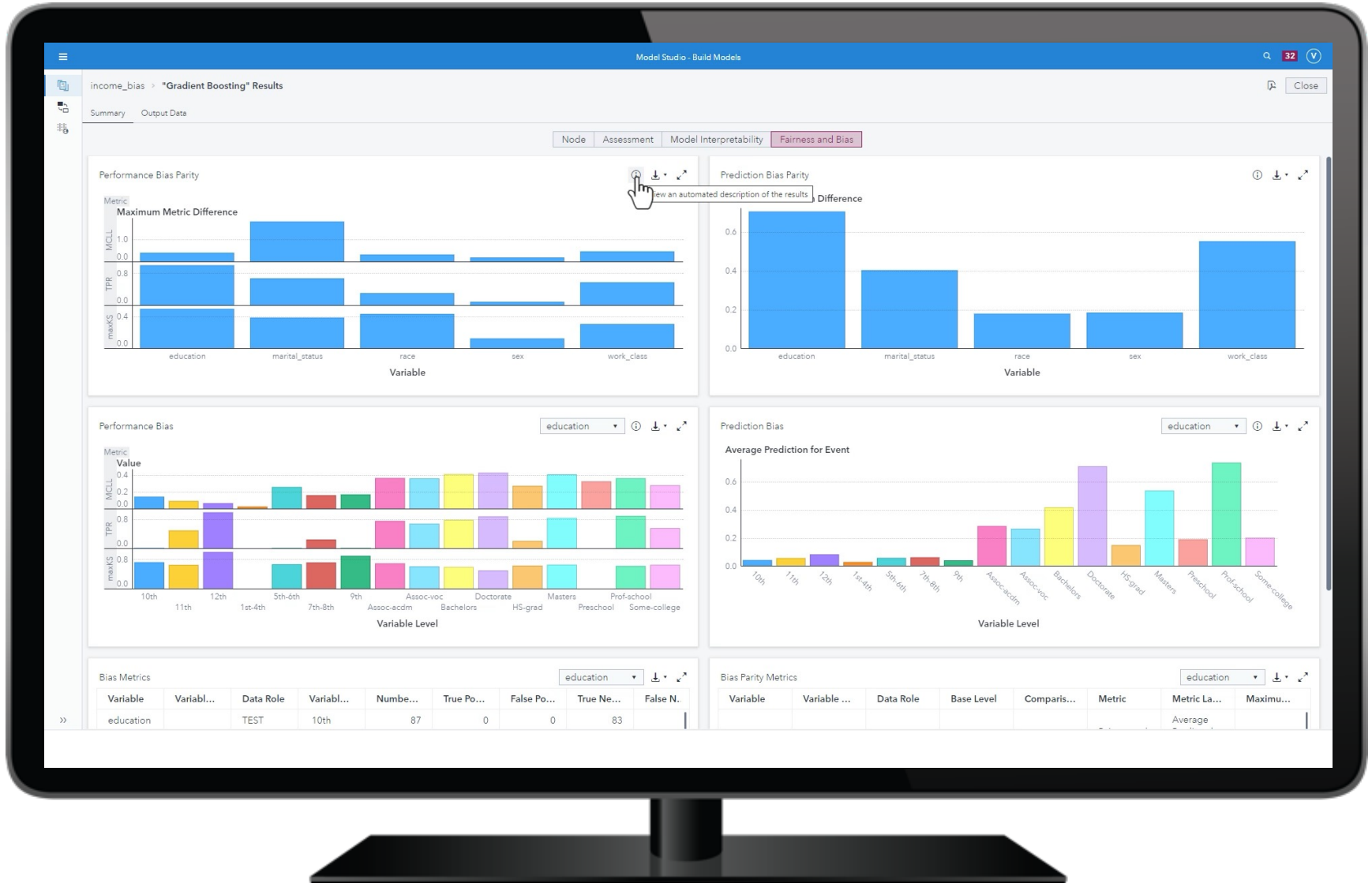
Data Roles

Automated explanation - income_level 1

- Response
  - income_level
- Time
  - Add
- Underlying Factors
  - age
  - country
  - hours_per_week
  - occupation
  - race
  - sex
  - work_class
  - Add

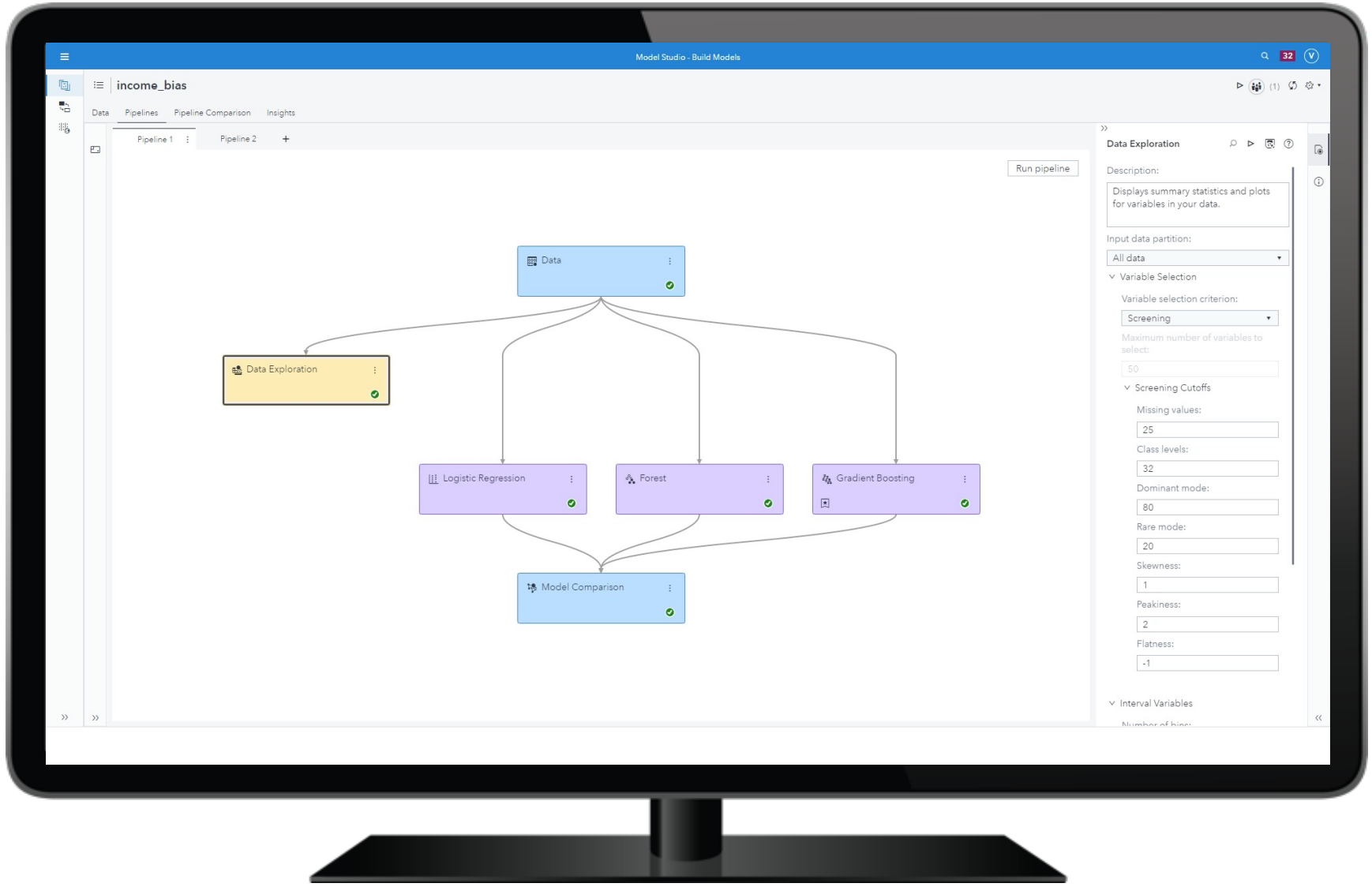
Transparent AI

Fairness Assessment and Bias detection

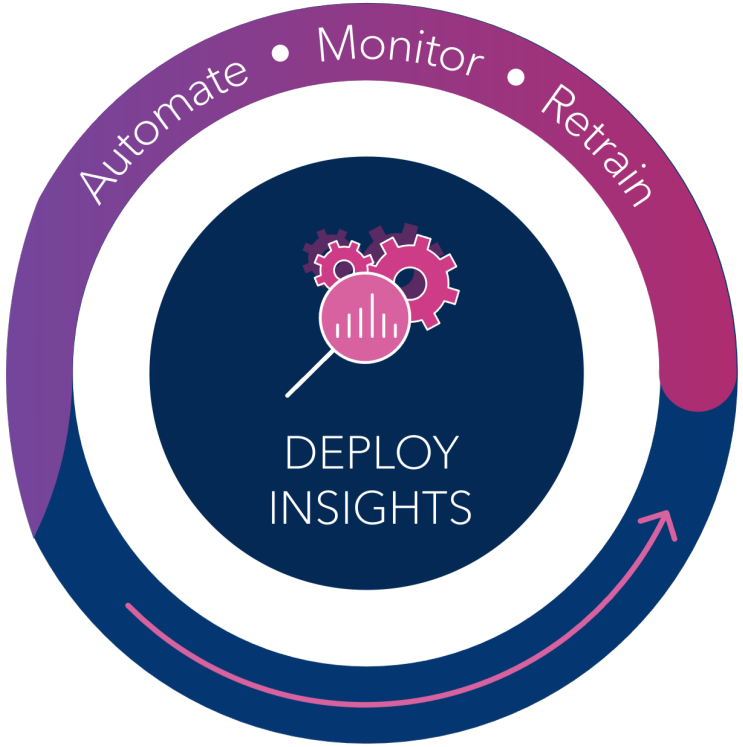


Transparent AI

Model Interpretability



Secure ModelOps Processes

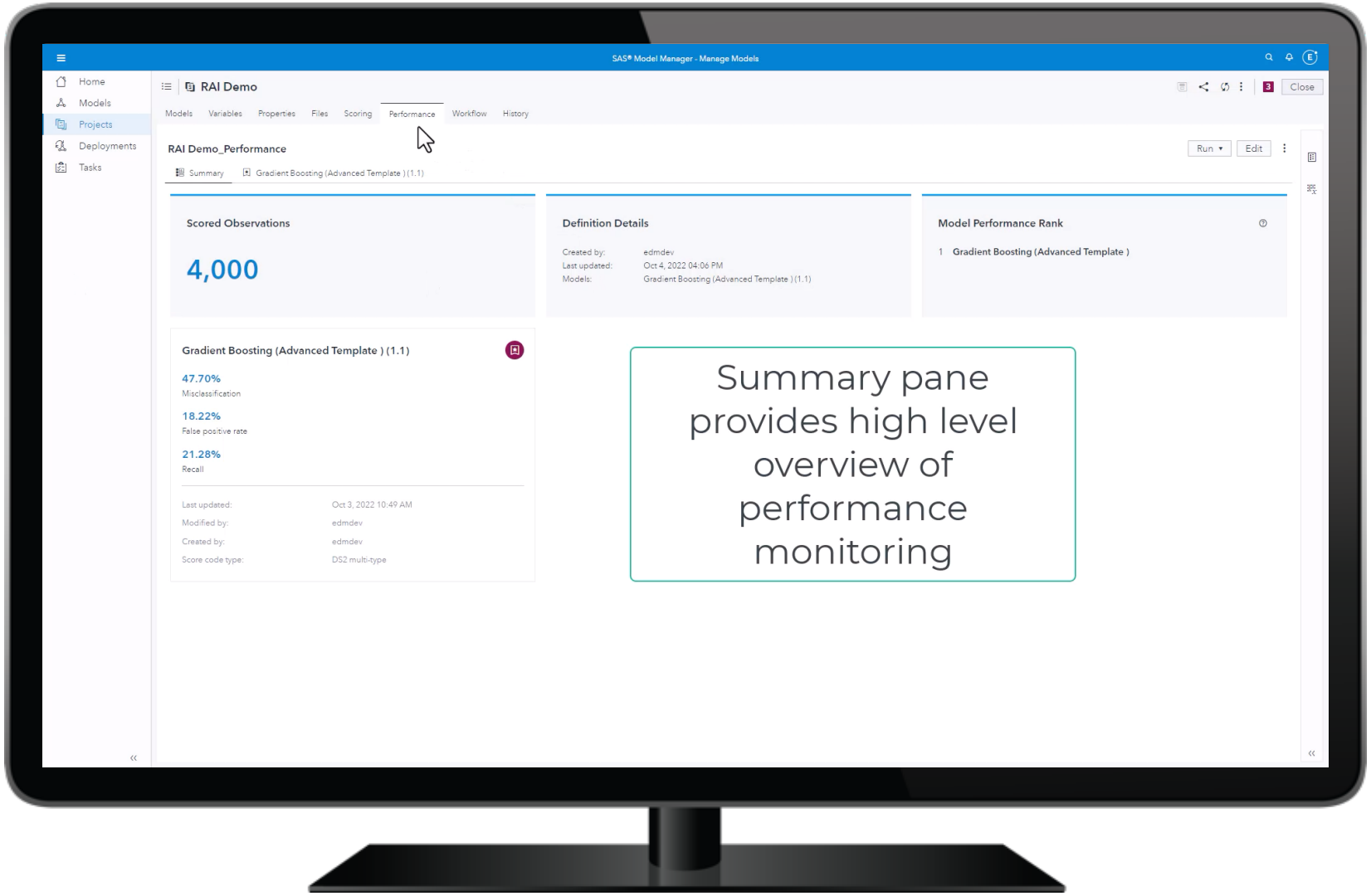


Deploy models into production with confidence.

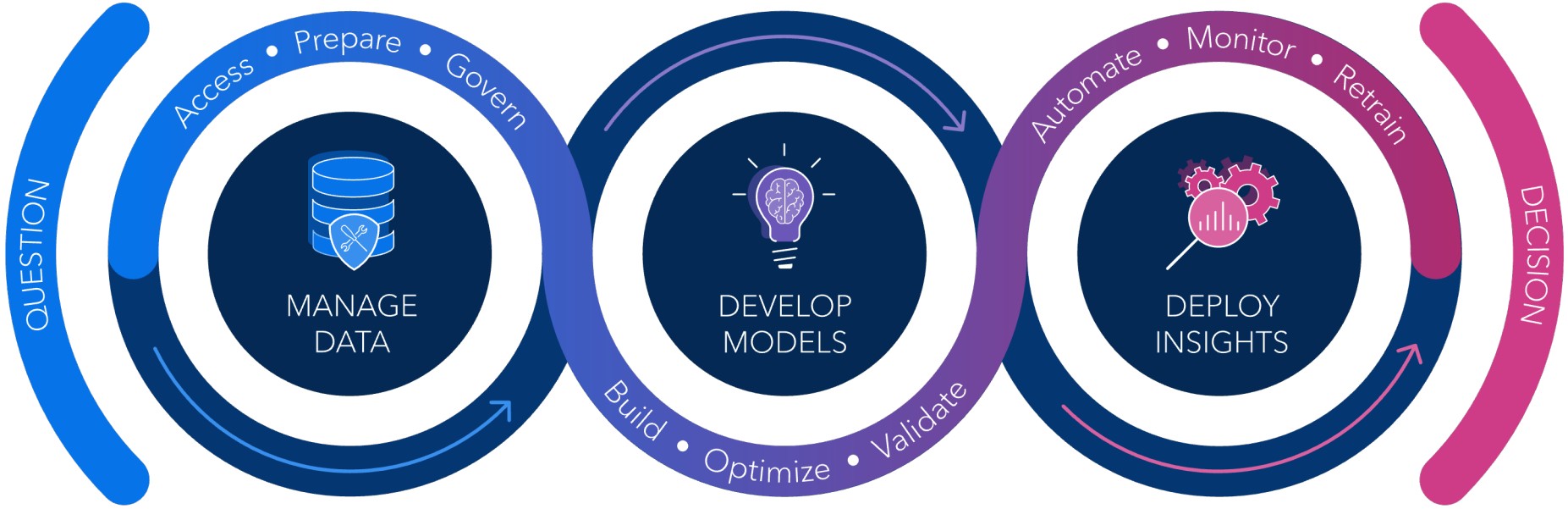
- Model Governance
- Model Monitoring
- Decision Accountability

Secure ModelOps Processes

Model Monitoring



SAS Viya Trustworthy AI end-to-end capabilities



- Data Quality
- Data Exploration
- Information Privacy
- Data Masking (1) (2)
- Data Suppression
- Data Lineage
- Synthetic Data Generation

- Natural Language Insights (1) (2)
- Model Interpretability
- Fairness Assessment & Bias Mitigation

- Model Governance
- Model Monitoring
- Decision Accountability

Resources

Watch

Trustworthy AI features demos on SAS Viya

Read

Data Ethics stories on SAS Blogs

Try

SAS Viya Free Trial

Engage

Ask the Expert SAS Community

Questions?