Types of Data Science:

A Practical Map for Learners, Analysts, and Data Product Managers

- **Training goal:** one shared map of data science types
- **Framing:** descriptive, diagnostic, predictive, prescriptive, generative/agentic
- **Plus** specialization tracks and the infrastructure foundation
- **Why classify:** choose methods, tools, and expectations before work starts
- **How to use this deck:** compare question, typical data, deliverables, pitfalls, maturity

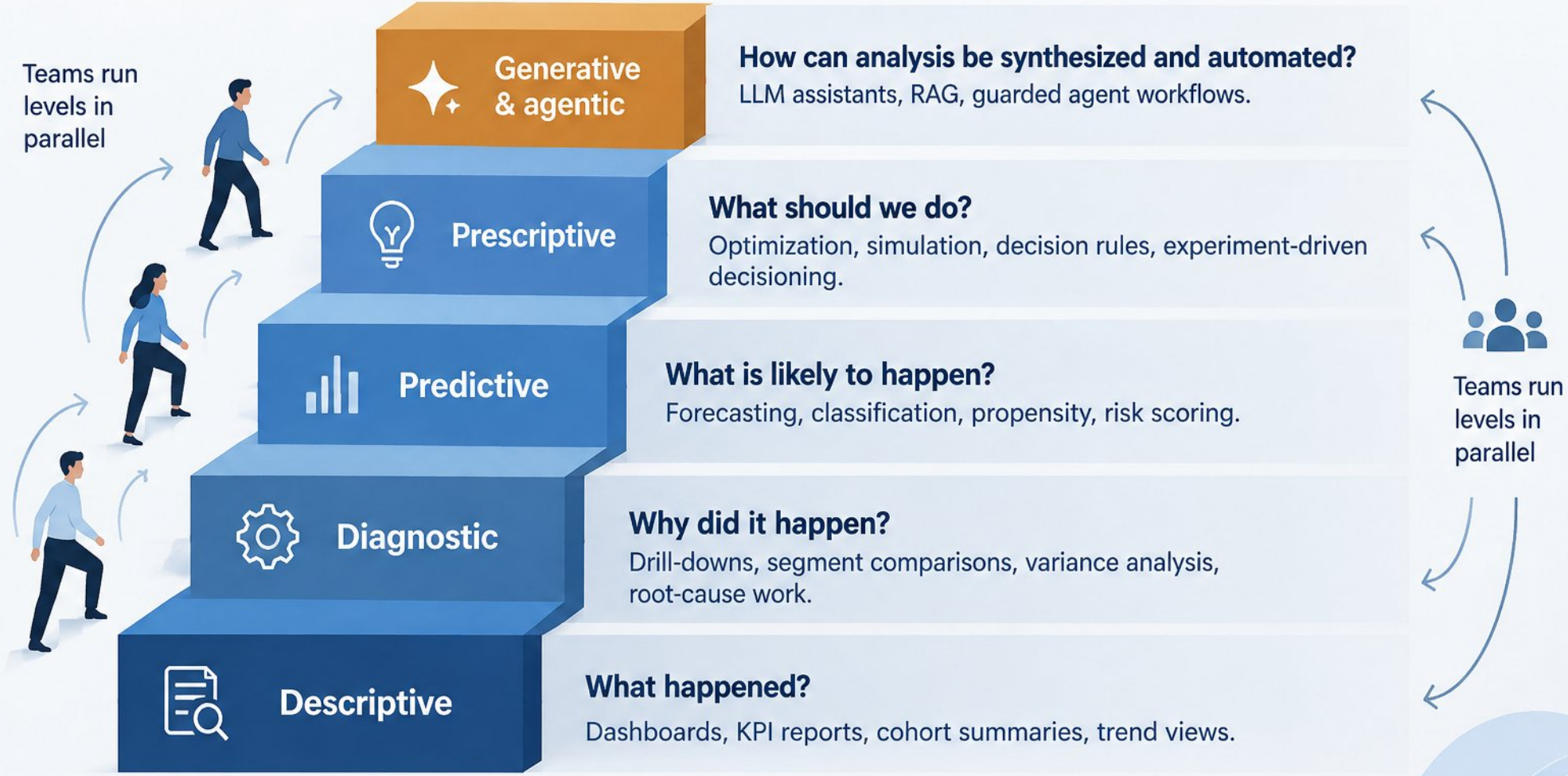


Why 'Data Science' Is an Umbrella, Not a Single Job

- Data science spans collecting, preparing, managing, analyzing, explaining, and disseminating data
- Data analytics is one component of data science, not a synonym
- Data science is not a subfield of machine learning or AI
- Industry usage conflates data science with BI, ML engineering, and data engineering
- Mis-scoped titles cause mis-hiring and mis-scoped projects
- Start with the application and the decision, not the tool



The Analytics Maturity Ladder: Five Levels of Questions



Maturity is not strictly sequential; teams run levels in parallel, and advanced work fails on weak foundations.

Shared Vocabulary: Data Types, Method Families, and Deliverables

Data Types

- structured
- unstructured
- streaming
- synthetic

Method Families

- statistics
- ML
- deep learning
- optimization
- causal
- NLP
- vision

Deliverables

- report
- dashboard
- model
- data product
- pipeline
- GenAI assistant

Data types favor different methods: structured, unstructured, streaming, synthetic

Method families: statistics, ML, deep learning, optimization, causal, NLP, vision

Deliverables: report, dashboard, model, data product, pipeline, GenAI assistant

Evaluate on accuracy vs. interpretability, latency, cost, governance, business value

Match the narrowest capable model to your data and task

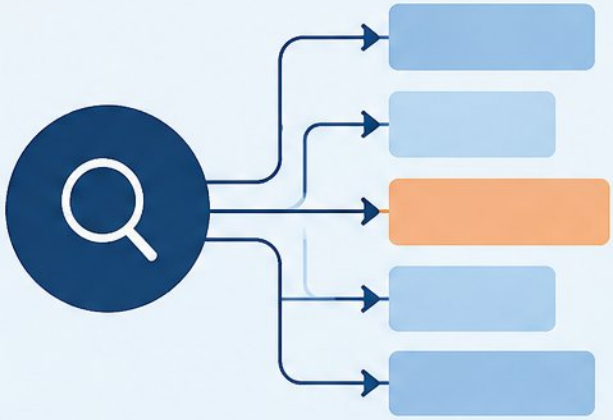
Descriptive and Diagnostic Analytics in Practice

Descriptive
What happened?



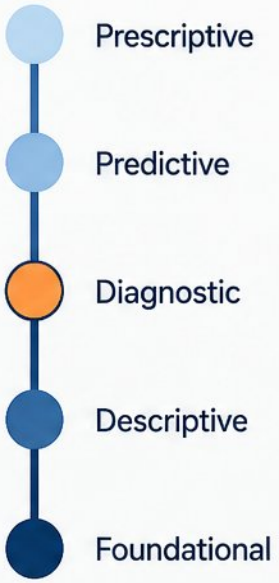
Descriptive: what happened?
Revenue, funnels, segmentation, dashboards, retention.

Diagnostic
Why did it happen?



Diagnostic: why did it happen?
Churn root cause, anomaly, cohort, A/B readout.

Analytics Maturity Ladder



Tools
SQL, BI platforms, statistical summaries, visualization, data quality checks.

Pitfalls
misleading aggregates, missing denominators, Simpson's paradox, untracked data issues.

Deliverables
self-service dashboards, executive briefings, investigation reports with a driver and a fix.

Churn diagnosis earns its keep as a prioritized driver list with severity and one action each.

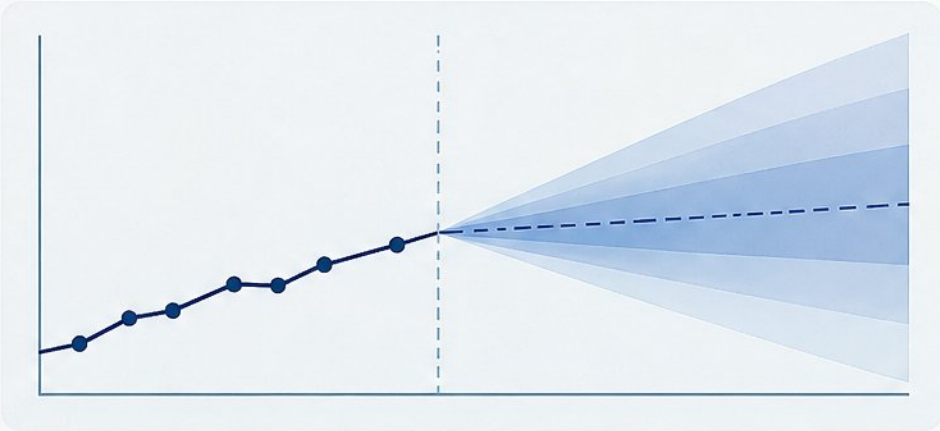
From Churn Dashboard to Churn Diagnosis: A Worked Example

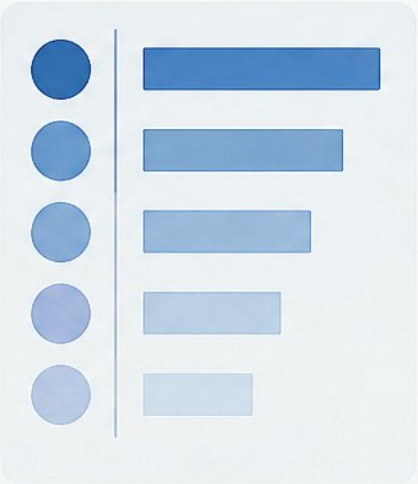


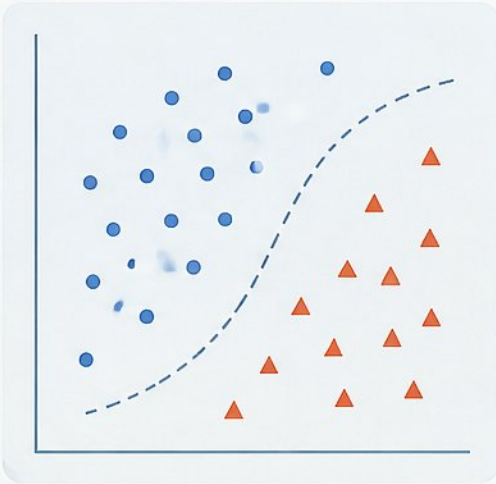
Descriptive	Diagnostic	Compare	Trap	Output
Descriptive: churn rose 15% last month; the dashboard shows trend, not cause	Diagnostic: which segments, cohorts, or behaviors changed before cancellation?	Compare churned versus retained accounts; classify drivers with two independent sources	Trap: exit surveys say "too expensive," but activation failure often drives loss	Output: prioritized driver list with severity and one concrete action each

Predictive Analytics: Forecasting, Scoring, and Classification

- Regression, classification, forecasting, ranking, propensity, anomaly detection
- Needs stable relationships and enough history; degrades quietly on change
- Outputs are probabilities and forecasts with confidence intervals, not certainties
- Lifecycle: framing, features, validation, deployment, monitoring, retraining
- Failure mode: predictions never reach the decision-maker in time to act







Prescriptive Analytics and Causal Reasoning: From Prediction to Action

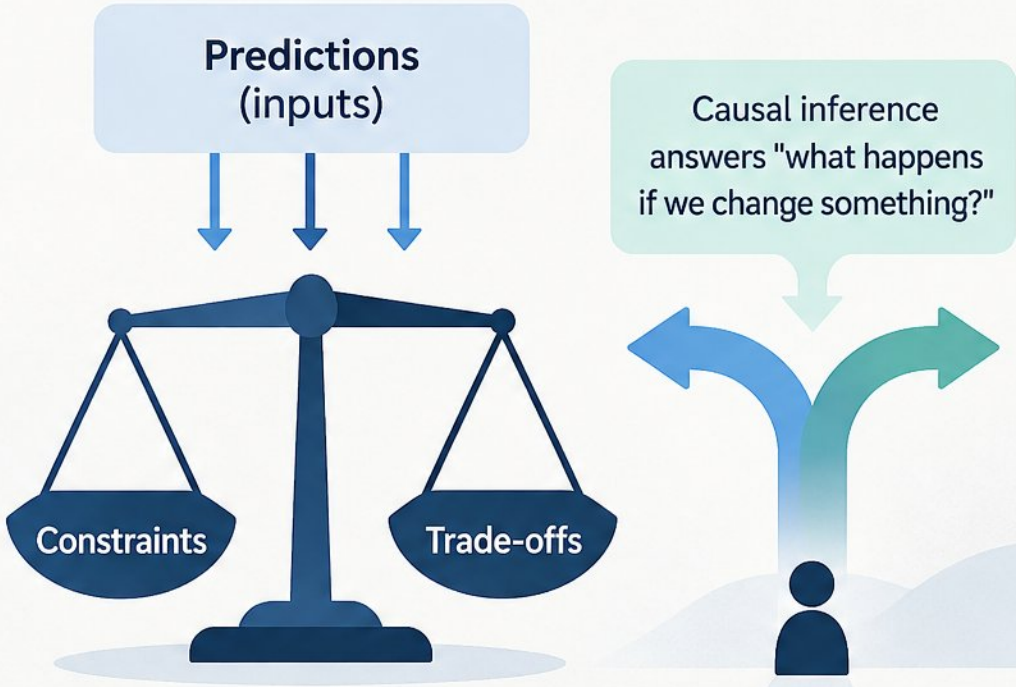
Prescriptive types: linear and integer optimization, simulation, reinforcement learning, experiment-driven decisioning

Predictions are inputs, not decisions; optimization makes trade-offs across many decisions

Use it when constraints, interacting decisions, and explicit trade-offs matter

Causal inference answers "what happens if we change something?" via experiments and quasi-experiments

Trade-offs: interpretability, fairness, latency, cost, and organizational trust.



Specialized Tracks: NLP, Computer Vision, and Generative AI

NLP



Consumes: text
Key risk: hallucination, bias

Computer vision



Comumes: images
Key risk: bias, domain fit

Generative AI

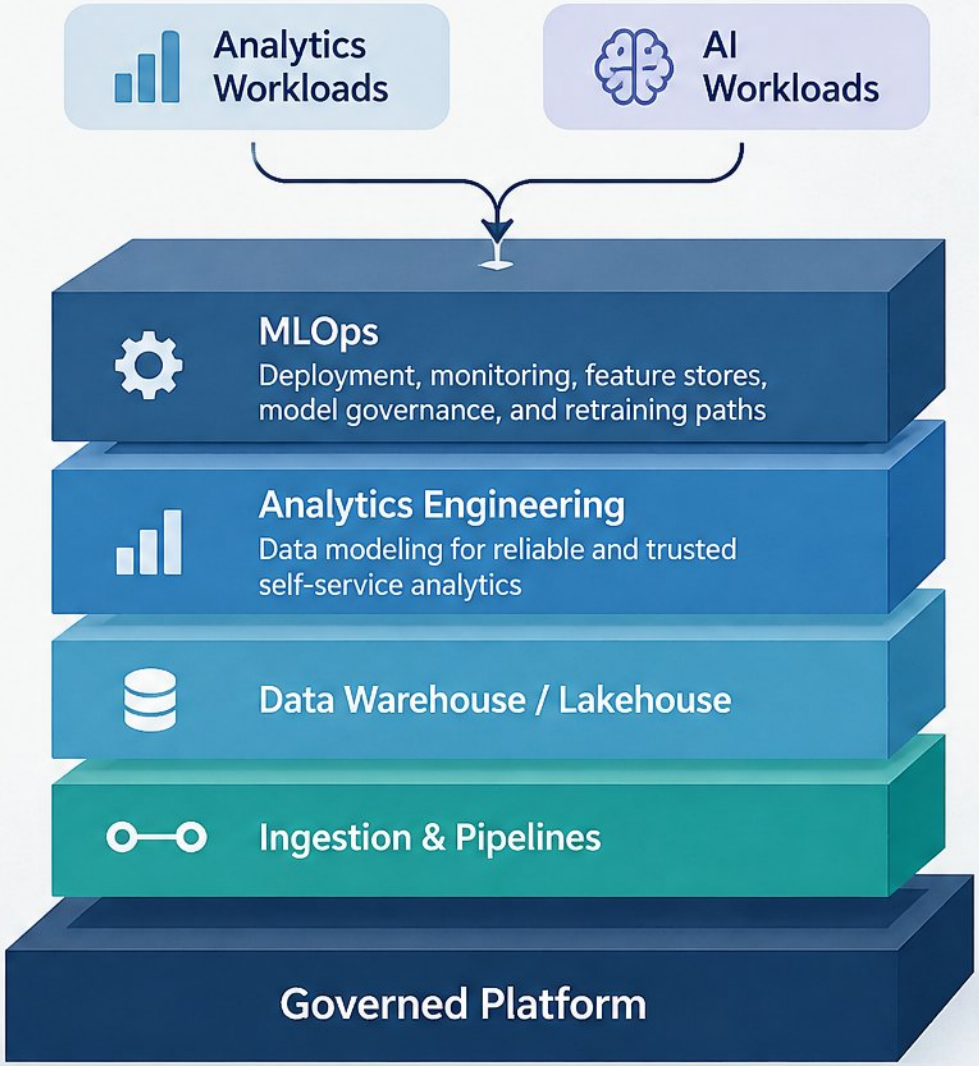


Consumes: text, code
Key risk: hallucination, prompt sensitivity, cost, latency

- **NLP:** text classification, entity extraction, sentiment, search, summarization
- **Computer vision:** image classification, object detection, OCR, document intelligence
- **Generative AI:** LLMs for content, code, RAG, and agent-assisted analysis
- **Match model to data:** LLMs for text, classical ML for tabular, vision for images
- **Evaluate** for hallucination, bias, prompt sensitivity, cost, latency, and domain fit
- **Graph and geospatial analytics:** niche, high-value specializations worth recognizing

The Infrastructure Track:
Data Engineering, Analytics Engineering, and MLOps

- Data engineering is the foundation: ingestion, pipelines, warehouses, lakehouses, orchestration, data quality
- Analytics engineering and data modeling make self-service analytics reliable and trusted
- MLOps covers deployment, monitoring, feature stores, model governance, and retraining paths
- Modern stacks converge analytics and AI workloads on governed platforms, not separate silos
- The type of data science a team can do is limited by its infrastructure



Choosing the Right Type: A Decision Framework

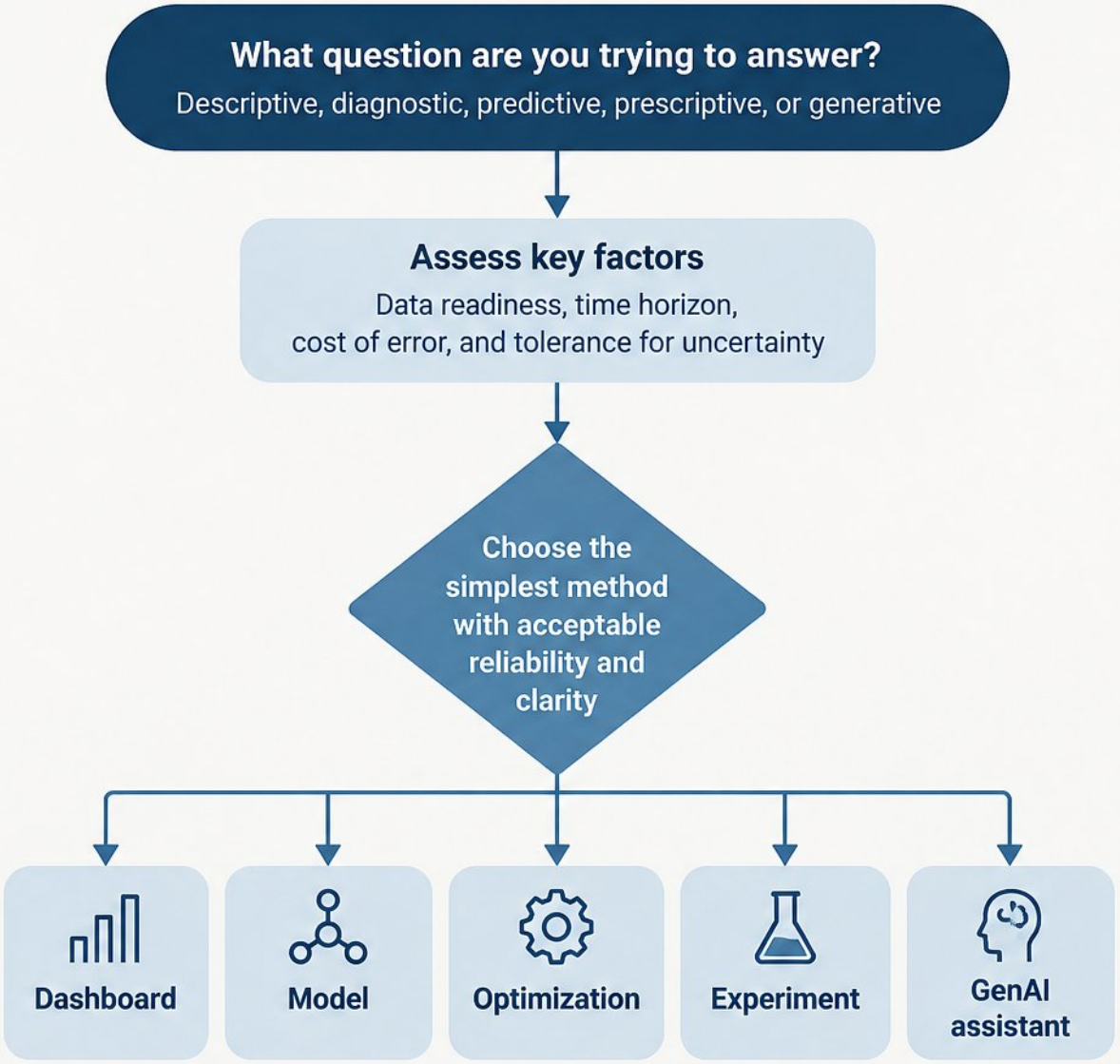
Name the question first: descriptive, diagnostic, predictive, prescriptive, or generative

Check data readiness, time horizon, cost of error, and tolerance for uncertainty

Match the method to the deliverable: dashboard, model, optimization, experiment, or GenAI assistant

Prefer the simplest method with acceptable reliability and clarity

Five-question executive screen: cost of being wrong, decision frequency, enumerable action space, known constraints, measurable feedback



Roles and Skill Expectations Across the Types

- **Analyst:** descriptive and diagnostic work; SQL, BI, and dashboards
- **Data scientist:** statistics, experiments, and predictive models
- **ML engineer:** productionizes and monitors models; software depth required
- **Data and analytics engineers:** pipelines, modeling, and trusted self-service data
- **Data product manager:** frames problems and outcomes, not implementation
- **2026 signals:** SQL, BI, and pipelines in at least 82% of postings
- **GenAI and LLMs:** about 37% of scientist postings, 7% of analyst postings
- **Career paths:** portfolio evidence matters more than titles



Applying the Map: Misapplied Methods and Their Consequences

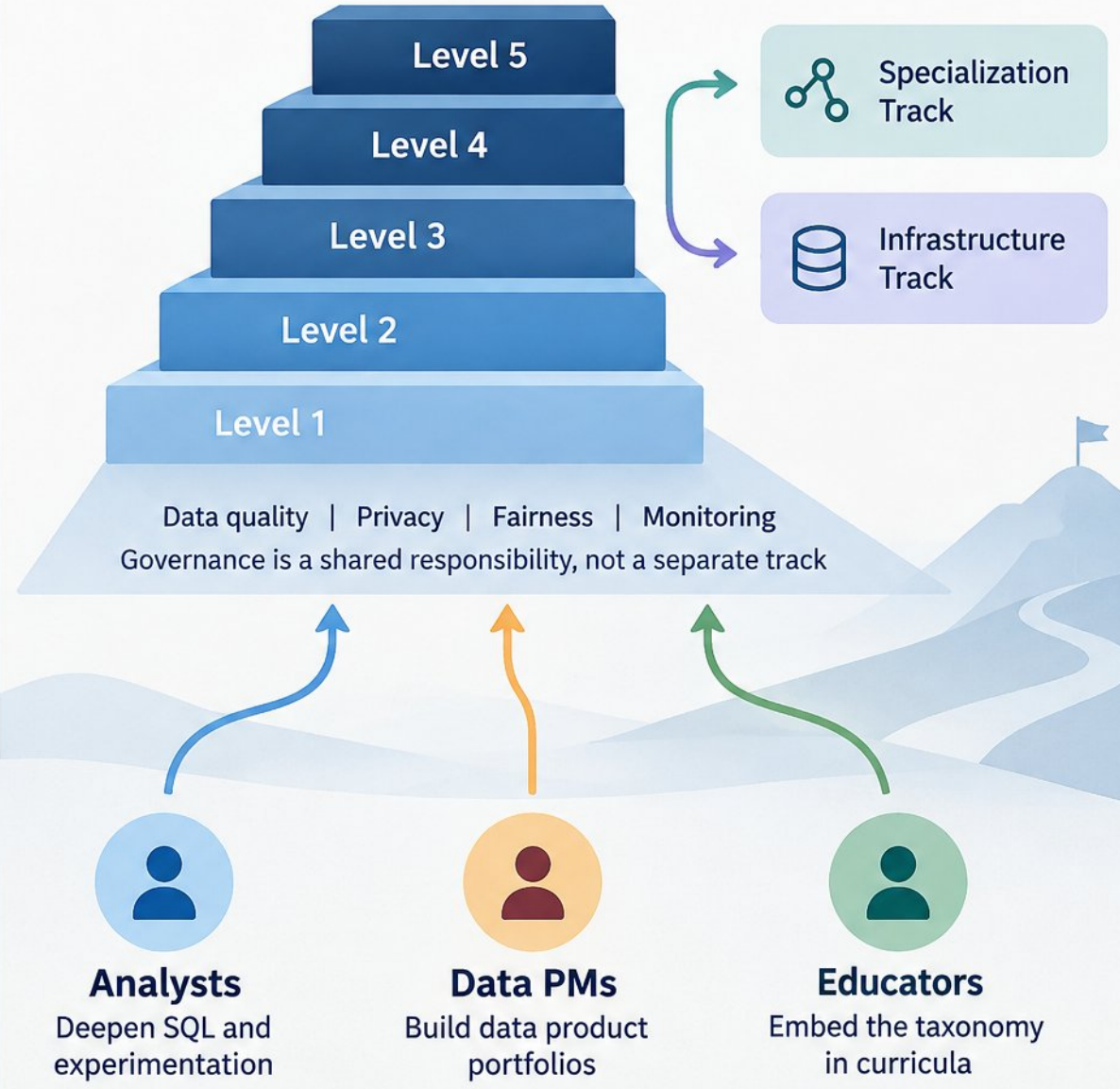
- Classic mistake: reaching for prediction when a clear description answers the question
- Equally costly: expecting a dashboard to prescribe an action
- Descriptive fails on data quality; predictive on drift and weak features
- Prescriptive fails on bad causal assumptions and weak constraints modeling
- Data product managers: map stakeholders to types and agree deliverables first
- Starter action: classify projects, audit skill gaps, define success metrics up front





Key Takeaways and Next Steps for Each Audience

- The map: five analytics levels plus specialization and infrastructure tracks
- Use the taxonomy diagnostically in meetings and project intake
- All types need data quality, privacy, fairness, and monitoring
- Governance is a shared responsibility, not a separate track
- Follow-up paths: analysts deepen SQL and experimentation, data PMs build data product portfolios, educators embed the taxonomy in curricula
- Reflection: which level does your project truly need?



PersonWise.

PersonWise • AI digital-human course platform

Scan the code or click anywhere to watch this course live, with voice Q&A

personwise.ai/t/5f7c8a20/public