

Data-Driven Decisions

From Raw Data to Executive Plan

Laurent Perier



Copyright

Please note that this work is copyrighted, and I am the sole license holder.

Dissemination or sale of any part of this work (including on the World Wide Web) will destroy the integrity of the work and is not permitted.

Any content extracted from this document is to be approved by me and should include appropriate referencing.

Using any part of this work for the current course is allowed and encouraged.

Data-Driven Decisions with Excel

Content



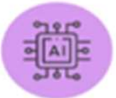
- Data
- Project
- Execution

Future of Jobs

Data: Key Enabler



1. IA & Big Data



2. Networks and Cybersecurity



3. Technological literacy



4. Creative thinking



5. Resilience and agility



6. Curiosity and lifelong learning



7. Leadership and social influence



8. Talent management



9. Analytical thinking



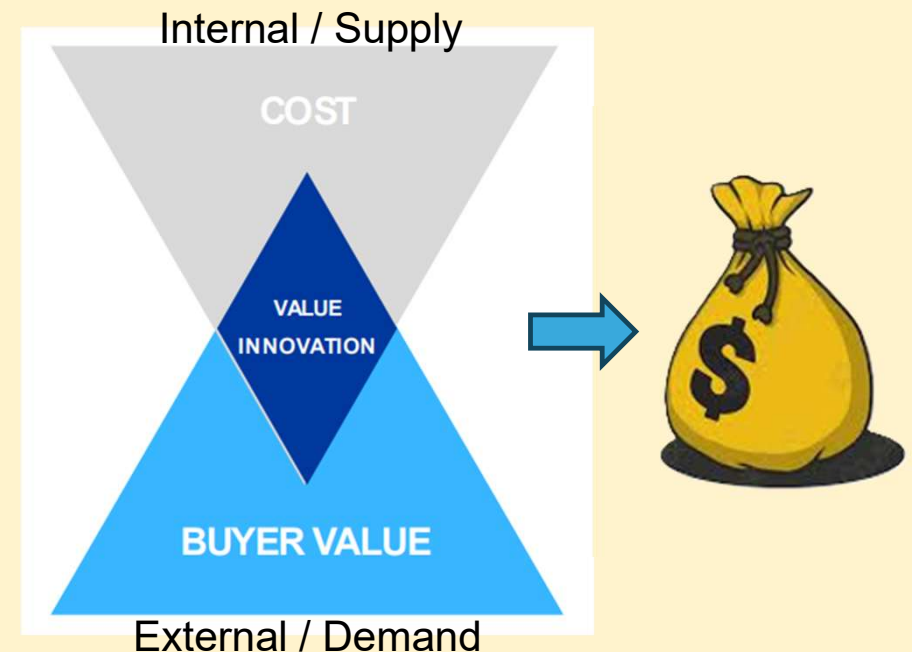
10. Systems thinking



*Source: *The Future of Jobs Report 2025, Forum Economique Mondial*

Business Profit

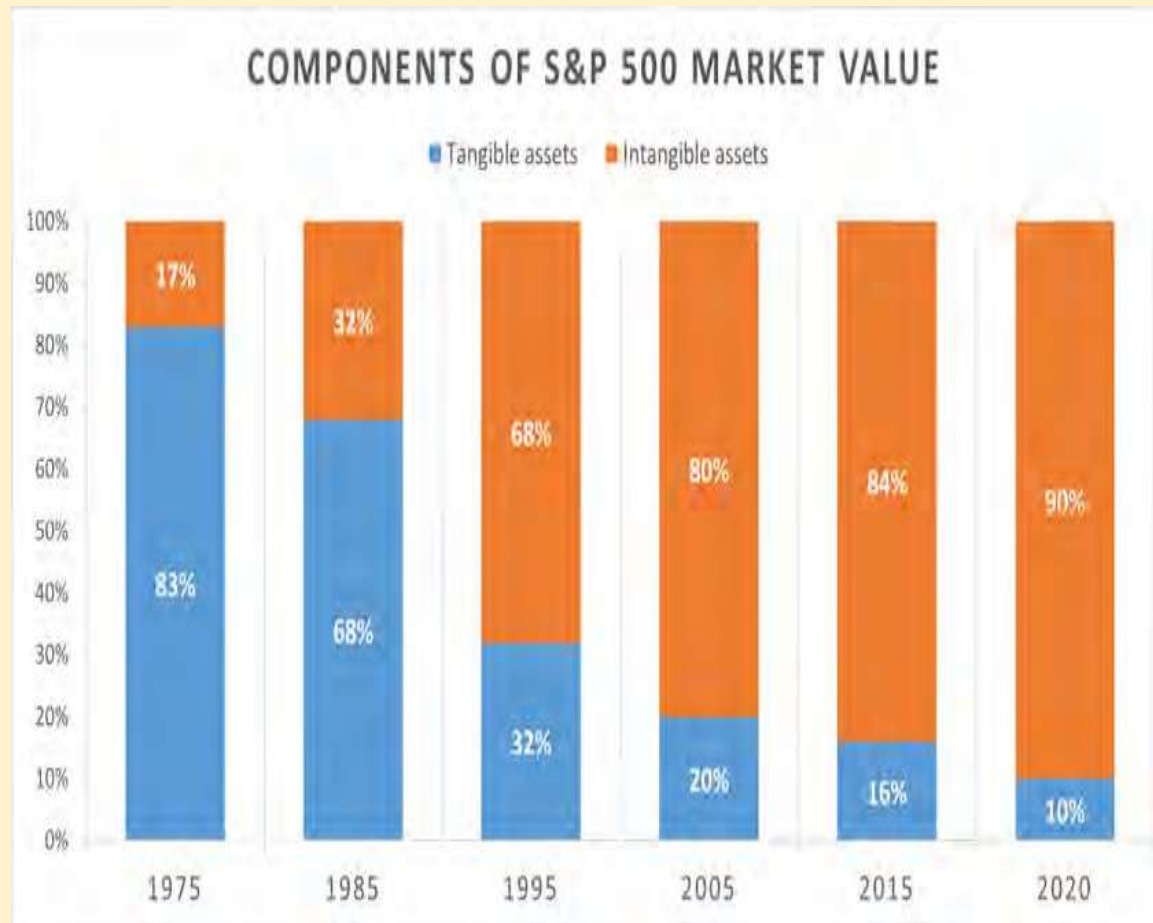
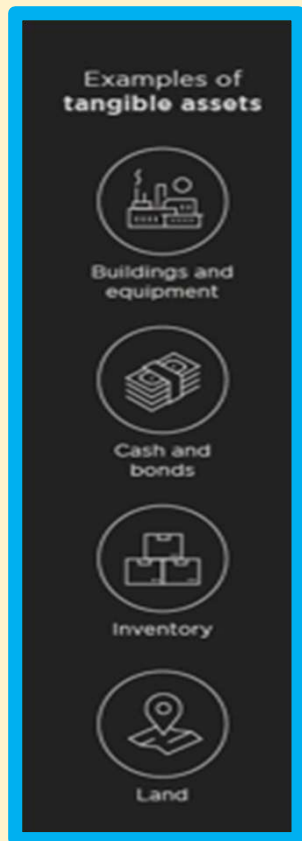
High-performing companies generate more **profit** through a **unique** positioning, made possible by **differentiated** processes and presenting the highest economic **value**.



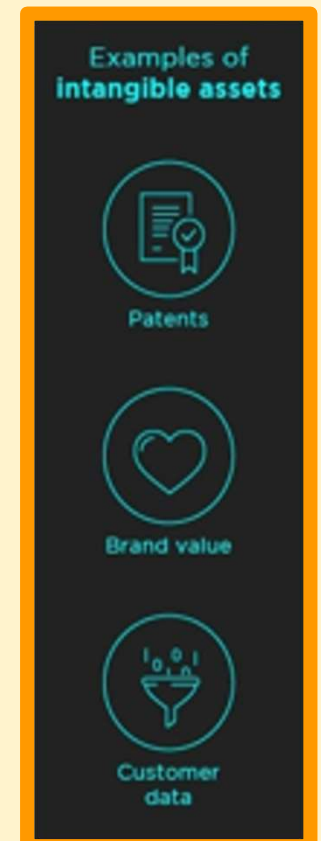
Enterprise Value

From Tangibles to Intangibles

Hard (Mortal)
Machine, Goods



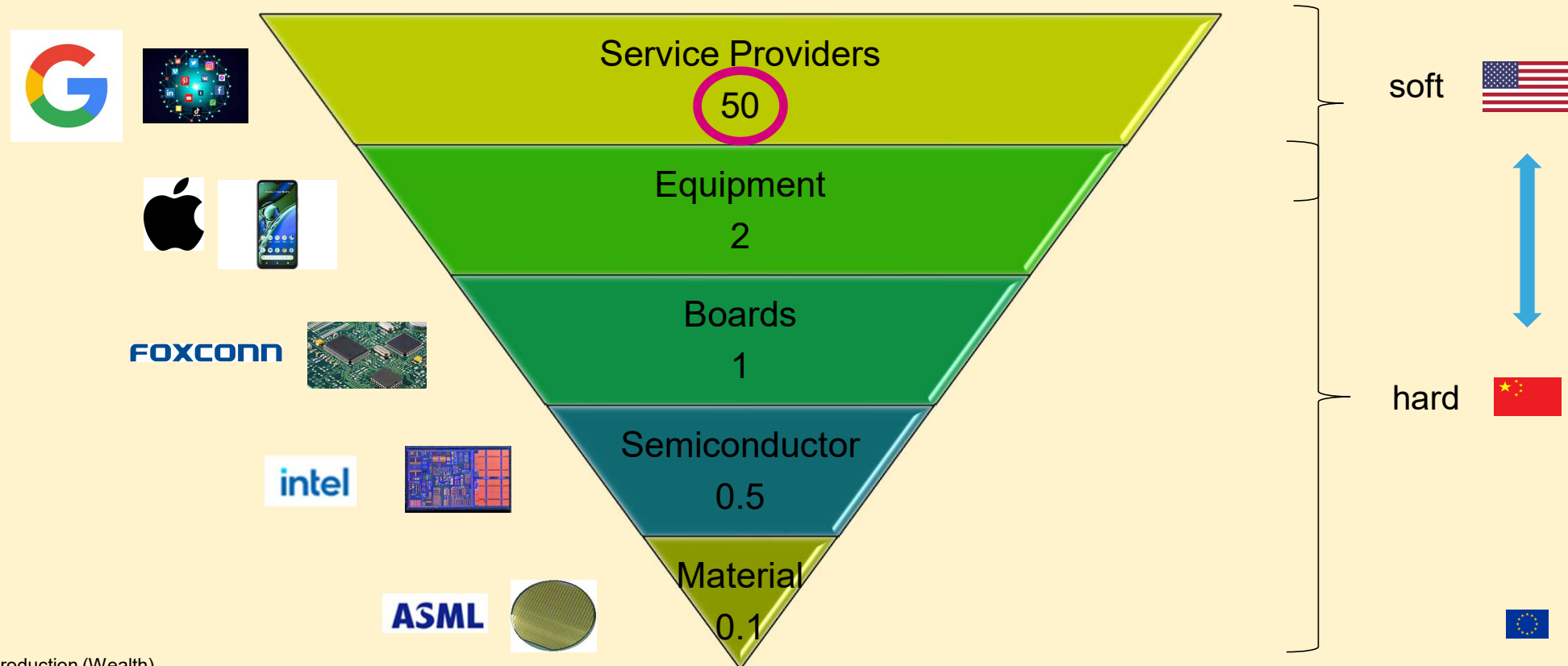
Soft (Click)
Services, Innovation



Digital Value Chain

+50% World Wealth

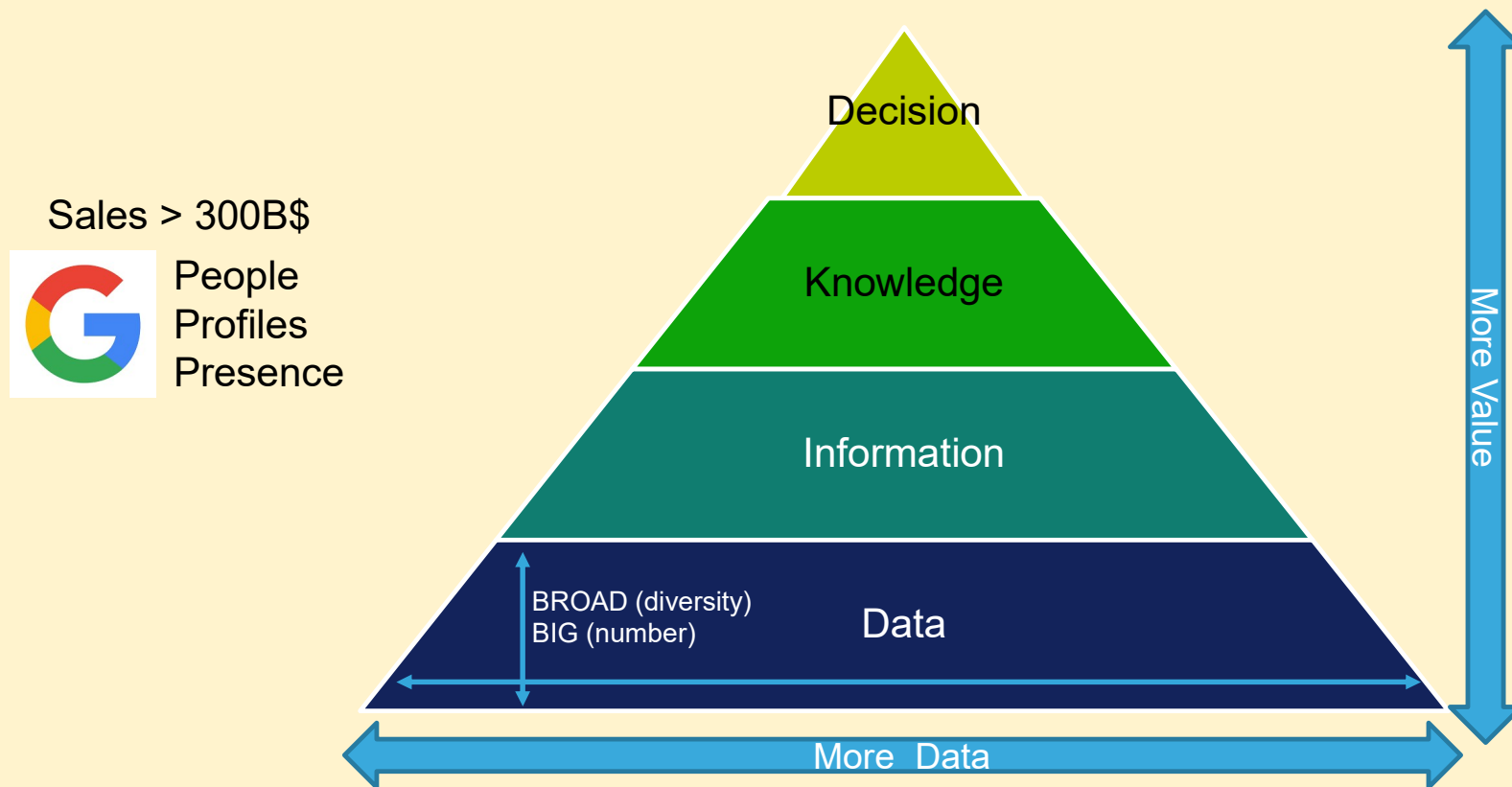
World GDP 2024: 100 T€ (1T=1000 B)



GDP: Gross Domestic Production (Wealth)

Add Value to Data

Scale: Big & Broad



From Data to Value

Example

Data

17

Information

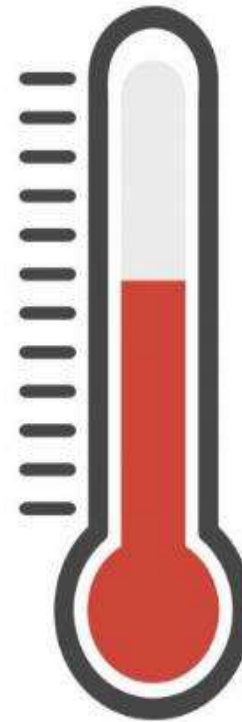
17 degrees
Celsius

Knowledge

Cold for
summertime

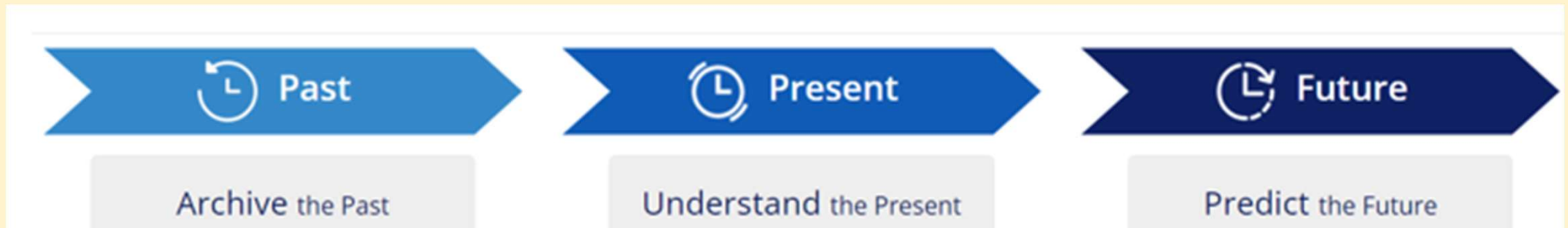
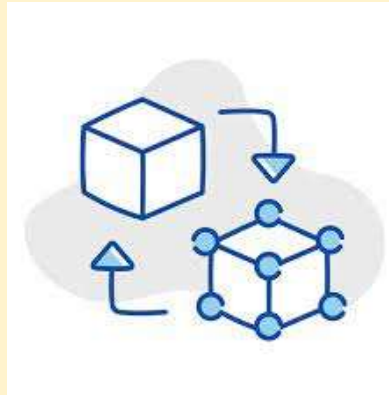
Decision

My coat



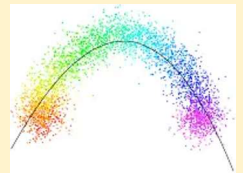
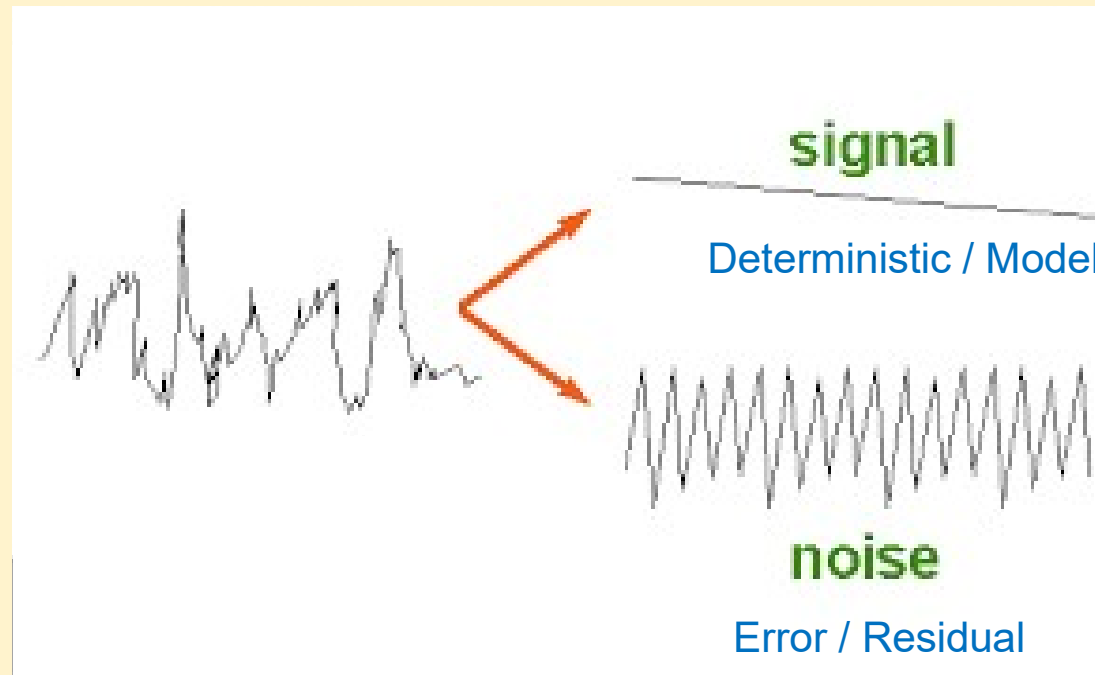
Add Value to Data

Modelization: Digital Twin



Data Analytics

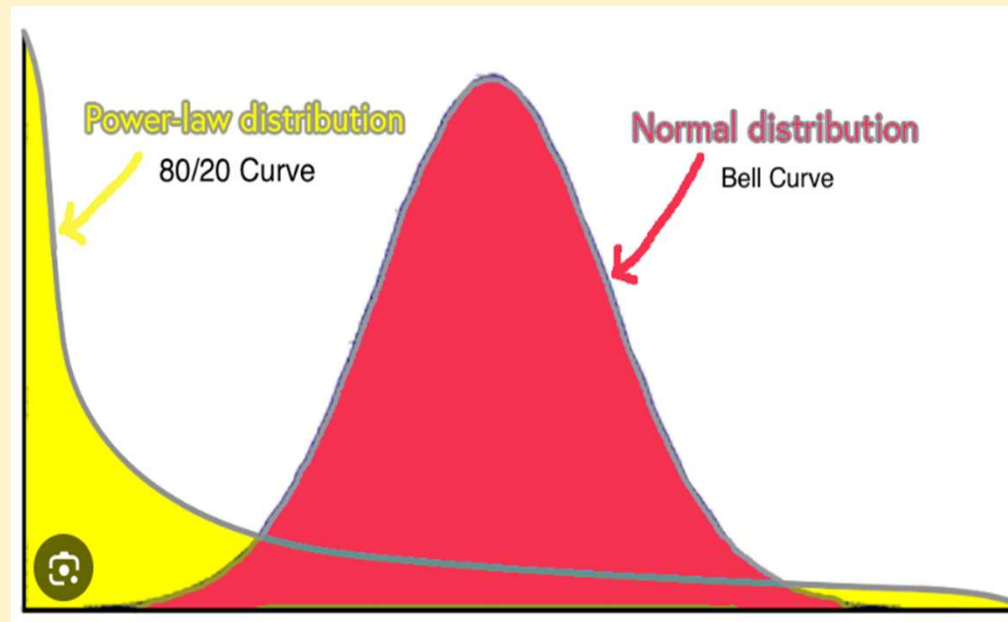
Detective: Split Signal from Noise



Data are Numbers

Predictive Probability Models

Power Law
80-20: Wealth

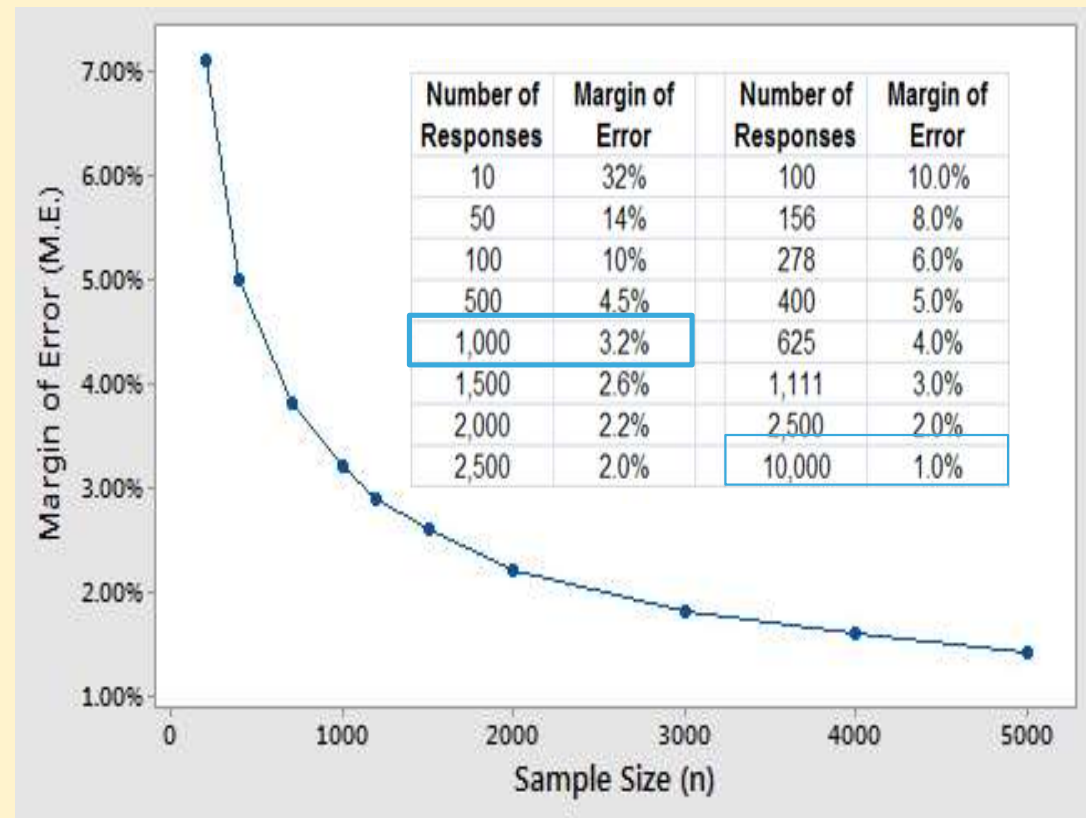


Normal Distribution
Bell Curve: Humans Height

Maths : XXI century science

Add Value to Data

More Data = More Knowledge



$$\text{Error} \propto 1/\sqrt{n}$$

Data Analytics Limits

The future cannot be predicted ; it must be prepared for.

Transforming uncertainty into competitive advantage requires a robust framework.

Mathematical rationality has its limits: it is imbued with values and remains imperfect.

Human cannot be mathematized, it creates novelty, identifies problems and chooses solutions.

Prediction tools rest on three pillars: data mining, technical expertise, computing power/energy.

The best use cases rely on reliable data, clear processes, measurable results, and controlled risk.

Data-Driven Decisions with Excel

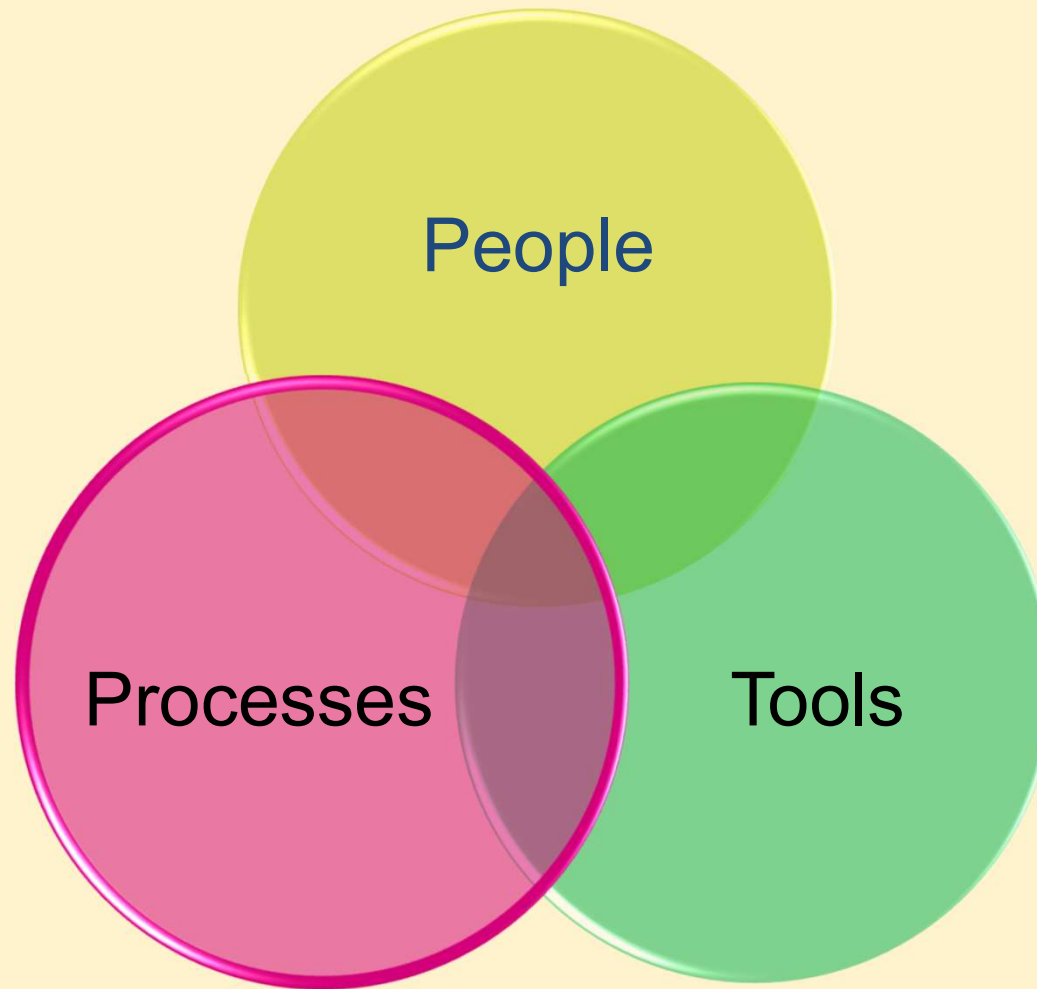
Content



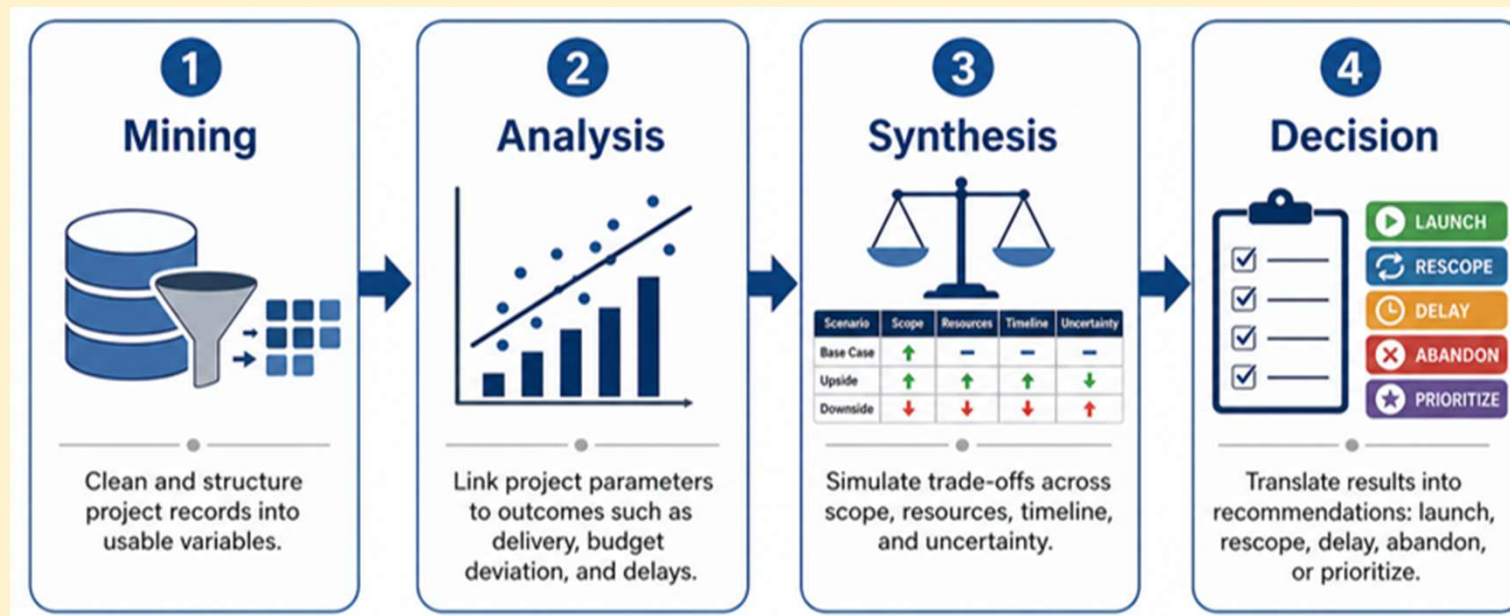
- Data
- Project
- Execution

Manager Role

Enabler



Predictive Analytics Workflow



Predictive Analytics Process

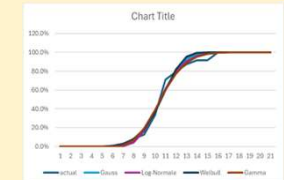
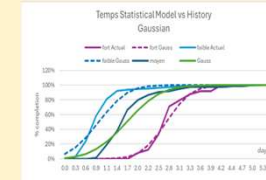
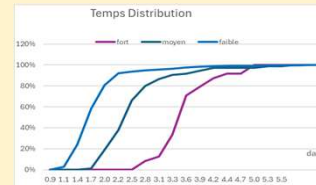
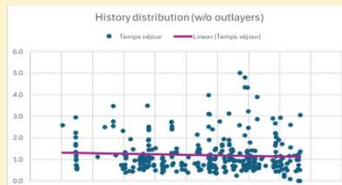
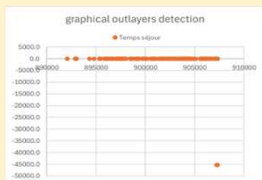
Database

Clean-Up

Distribution

Model

Optimization



Business Case

Your Project

- Team: Project Management
- Objective: Predictive Model
- Process: Apply the Framework
- Data Source: “Project_IN”
- Deliverable: “Project_OUT”
- Size: 4 people



Problematic

*You are a consultant for a business Development company.
Propose a solution to improve project portfolio decision-making using predictive tools.*

Objective:

Predict the projects Time To Market using the provided data and a predictive model to be constructed.

Business Case

Project Management

Problem: Avernixa Group faces recurring project delays, budget overruns, and missed strategic opportunities due to reliance on managerial judgment, static checklists, and fragmented files. Despite extensive historical data (e.g., delivery dates, budgets, complexity ratings), early project decisions lack quantitative rigor, leading to late discovery of risks and costly mitigations.

Solution: Reduce Time-to-Market (TTM) and improve project portfolio decision-making by leveraging historical project data to predict success probabilities, risks, and resource impacts.

Tool: PLUTO turns Avernixa's historical data into actionable insights, enabling evidence-based governance while preserving human accountability. The workflow is designed to be disciplined, explainable, and measurable, addressing the gap between data availability and decision-making rigor.

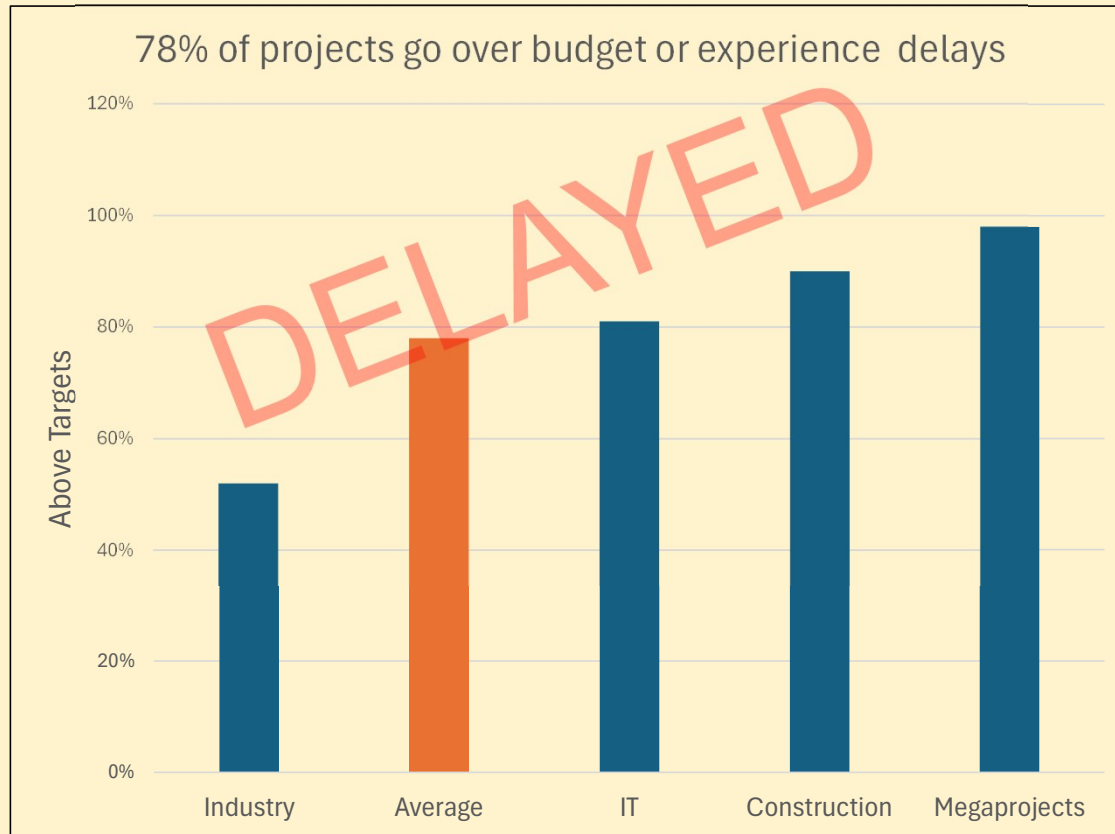
Learning Steps (Diagnostic Process): A 4-stage DA+AI workflow is implemented :

1. Mining: Observation → Clean history project data
2. Analysis: Symptoms → Predictive models + best-fit selection
3. Synthesis: Diagnostic → Evidence-based prioritization
4. Decision: Treatment → Portfolio optimization

Deliverable: A reusable Excel analytical plan documenting the full decision process, from data mining to predictive model outputs with a transparent, auditable workflow that improves debate quality, exposes assumptions, and supports better project choices without replacing managerial judgment.

Problem: Too Many Projects go Wrong

Late Projects, High Costs, Lost Sales



78%

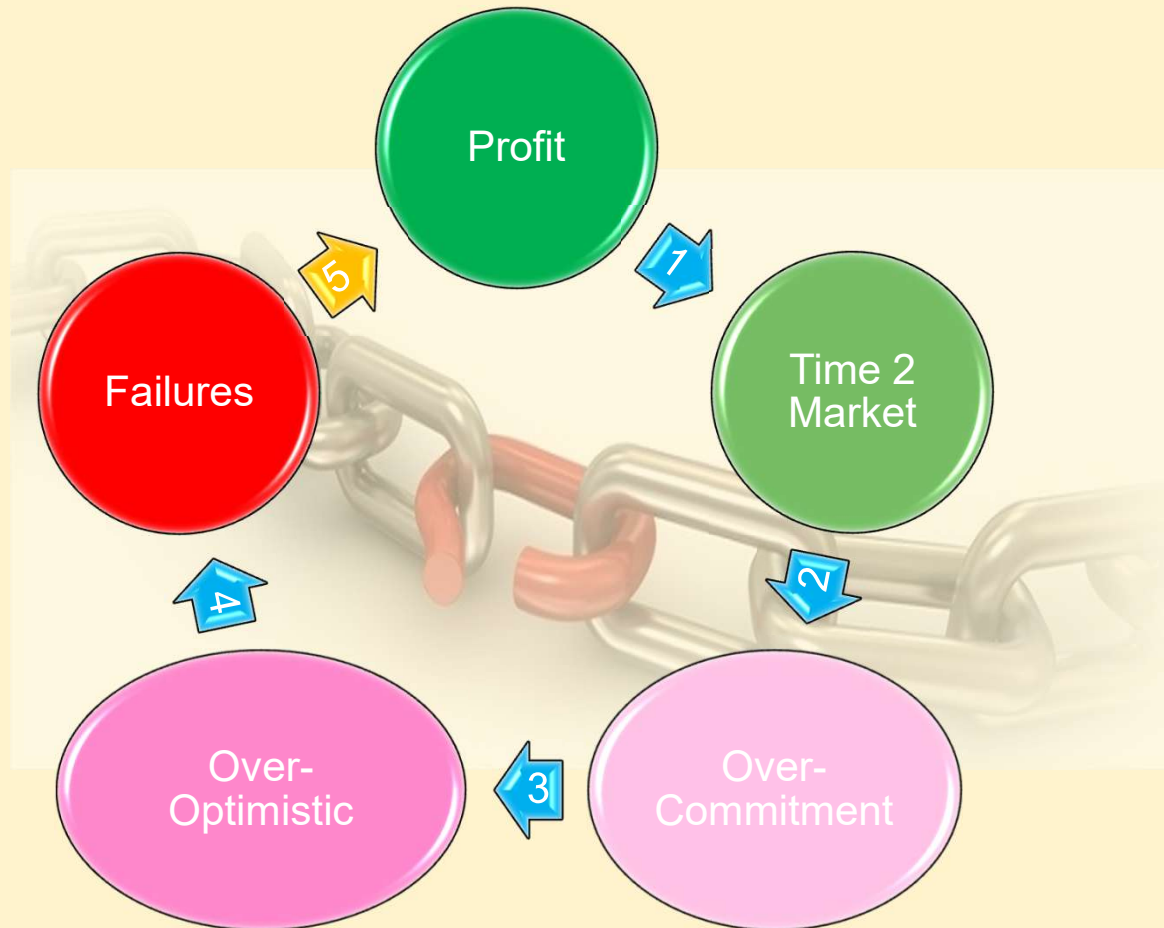
Not reducing

Project management only an operational task in past becomes critical for competitiveness

Source: Gartner

Failure Root-Cause

Over-Commitment

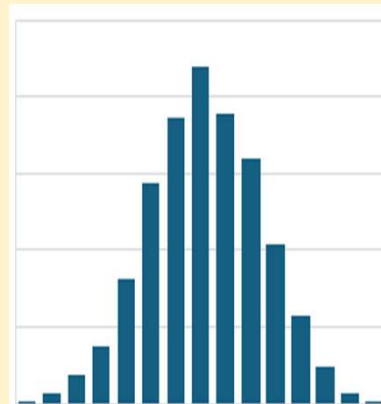


Answer: Plan Better

Data Statistics



Precision
Number
Centering
Average – Median
Distribution
Stdev
Process Capability
Cpk
Optimal Stat Model
Khi-Deux (χ^2)_Test

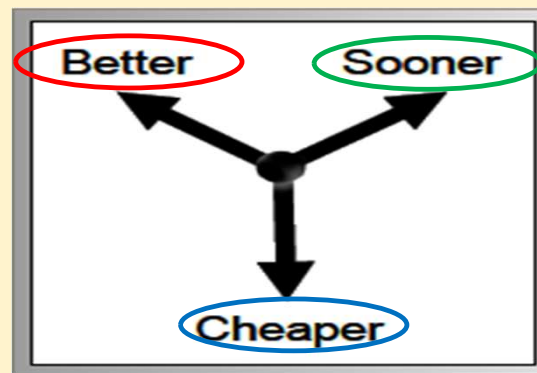


Innovation
Duration
Effort
Cost
Revisions

Innovation Complexity Assessment

1st in family
(new Architecture / Complex IPs)

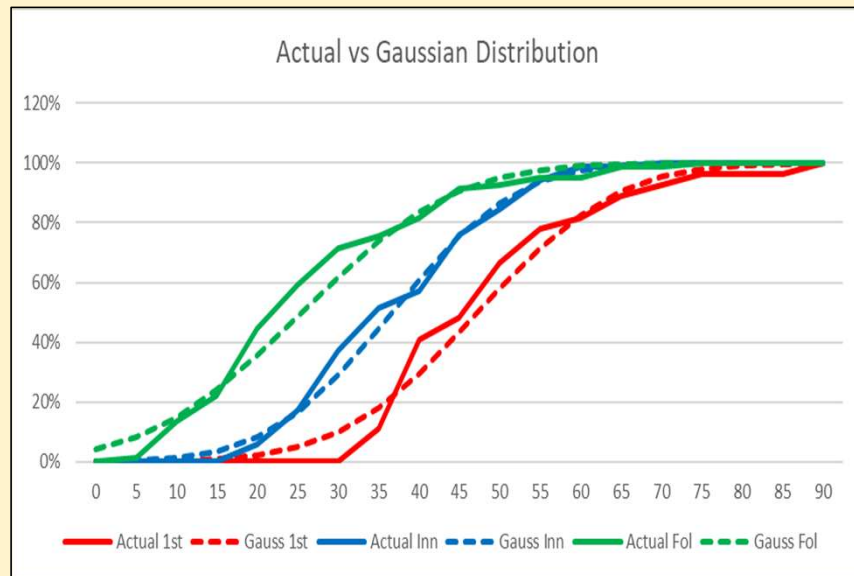
Innovative
(new IPs / Experiment)



Follower
(less IPs, Cost Reduction, Scale-Up)

Statistics Extraction

Process Quality



	1st in family	Innovative	Follower
number of projects (precision)	27	70	81
average	47	37	26
stdev	14	12	15
max	87	69	72
min	31	15	5
median	45	35	21
median/average	95%	94%	84%
Process Analysis (Audit)			
cp (min 1.33, goal 1.67)	0.69	0.74	0.76
cpk (min 1.33, goal 1.67)	0.41	0.60	0.46
Correlation Test (Model extraction)			
X2 Test proba (Gaussian Model)	100%	100%	100%
> 5%, hypothesis accepted: Normal Distribution applicable			

Benefit :Stakeholders Alignment

Less Uncertainty, Less Silos, More Trust



Data-Driven Decisions with Excel

Content

- Data
- Project
- Execution



Data Analytics Process

By structuring the analysis in successive steps, from data exploration to the overall evaluation of a project :

- History (database)
- Clean-up (outliers)
- Segmentation (criteria)
- Distribution (statistical)
- Modeling (first statistical model)
- Optimization (error minimization)
- Exploration (testing other models)
- Prediction (best model)

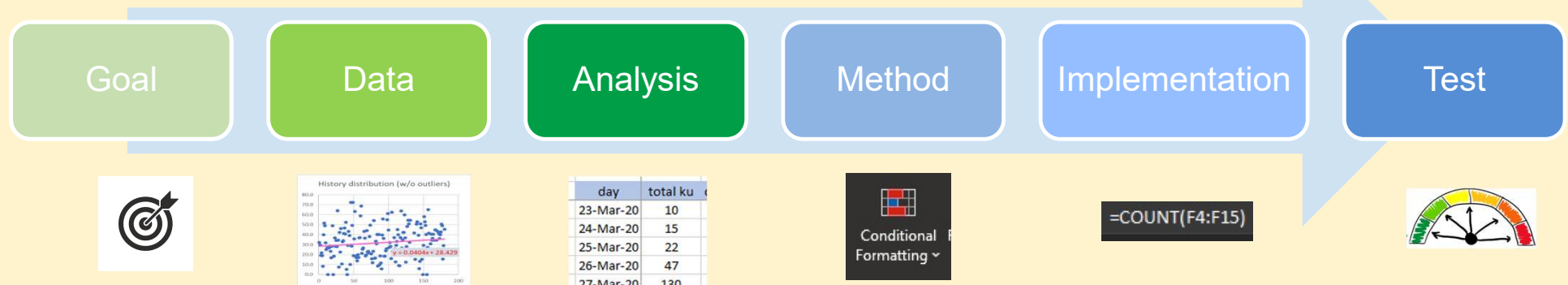
Hypothesis

- A good model is a trade-off between accuracy and simplicity
- Project complexity defines duration, effort and number of redesigns
- Process assuming all resources interchangeable (no split by job, skills)

Data Analytics Execution

Effectiveness / Right Thing
WHY

Efficiency / Thing Right
HOW



Going too fast induces Delay

Action 1 : Historical Database

Data Mining

B	C	D	E	F	G	H	I
PLUTO : Predictive Project Management (Deal with Facts, not Hopes)							
INPUT (history)			© Copyright 2026, ThinknSolv - Laurent Perier All Rights Reserved laurentperier@thinknsolv.com				
Project	complexity	duration months					
1	Innovative	32					
2	follower	30					
3	follower	19					
4	follower	22					
5	Innovative	19					
6	follower	8					
7	follower	40					
8	follower	52					
9	follower	49					
10	Innovative	32					
11	follower	25					
12	Innovative	27					
13	1st in family	87					
14	follower	27					
15	follower	27					
16	follower	29					
17	follower	23					
18	follower	17					
19	Innovative	28					
20	Innovative	49					
21	follower	1203					
22	follower	18					
23	follower	37					
24	Innovative	28					
25	Innovative	34					
26	Innovative	28					
27	follower	64					
28	Innovative	35					
29	follower	7					
30	Innovative	42					
31	Innovative	22					
EULA	Glossary	Duration_History_IN	Duration_Cleaning	Duration_Distribution	Duration_G		

Data only indicative

Action 2 : Database clean-Up

Define Outliers Limits



Fill the yellow cells

=COUNT(D44:D696)

=COUNTIFS(C44:C696,"<>"&"",D44:D696,"")

=MIN(D44:D696)

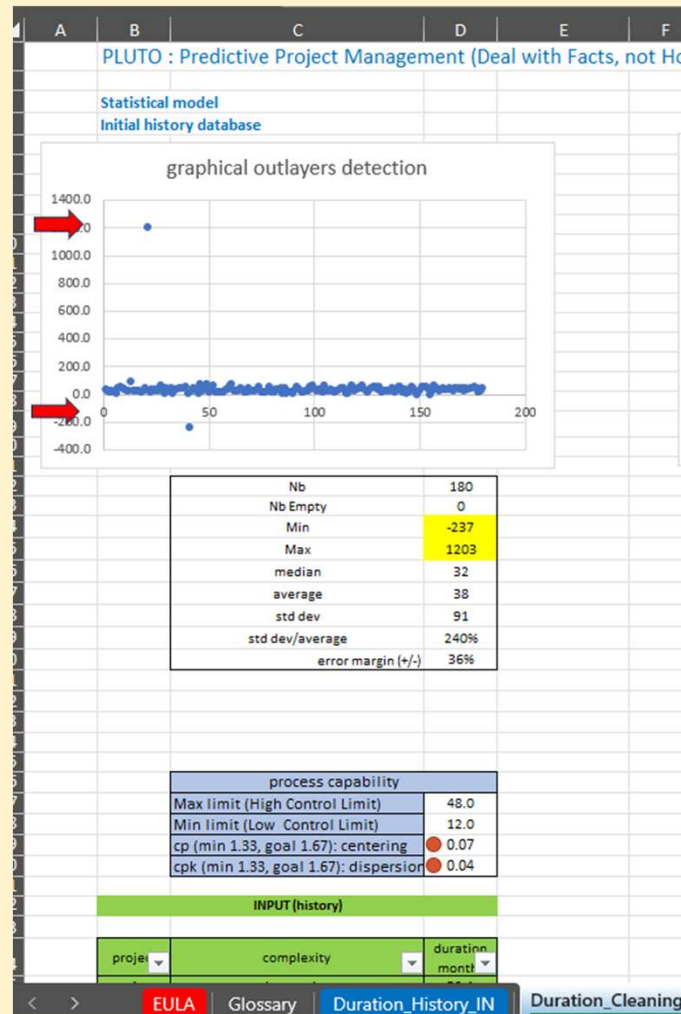
=MAX(D44:D696)

=MEDIAN(D44:D696)

=SUBTOTAL(1,D44:D696)

=SUBTOTAL(7,D44:D696)

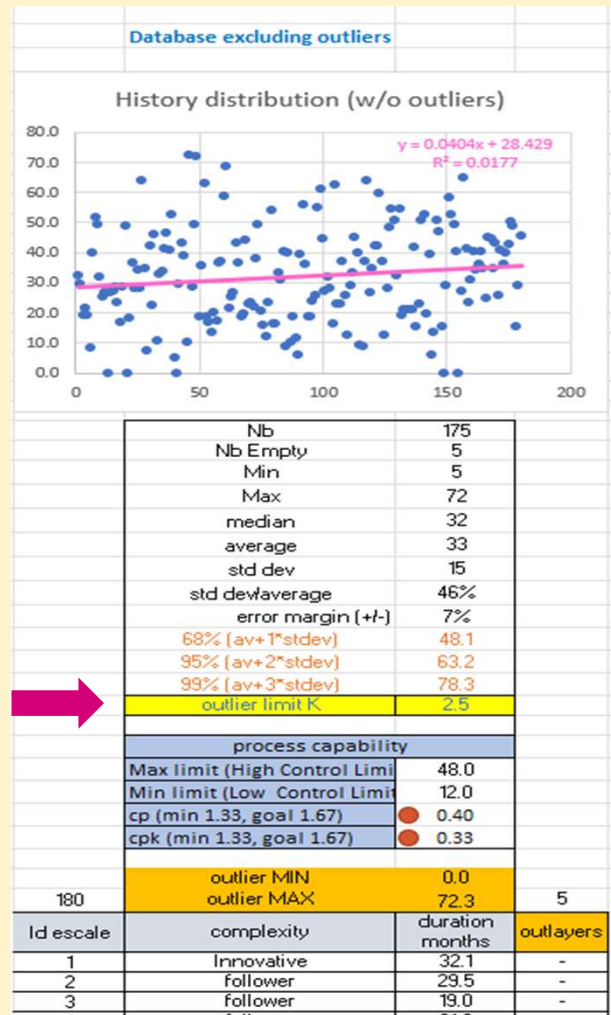
=1/SQRT(D21)



Data only indicative

Action 3 : Database Clean-Up

Removal Outliers

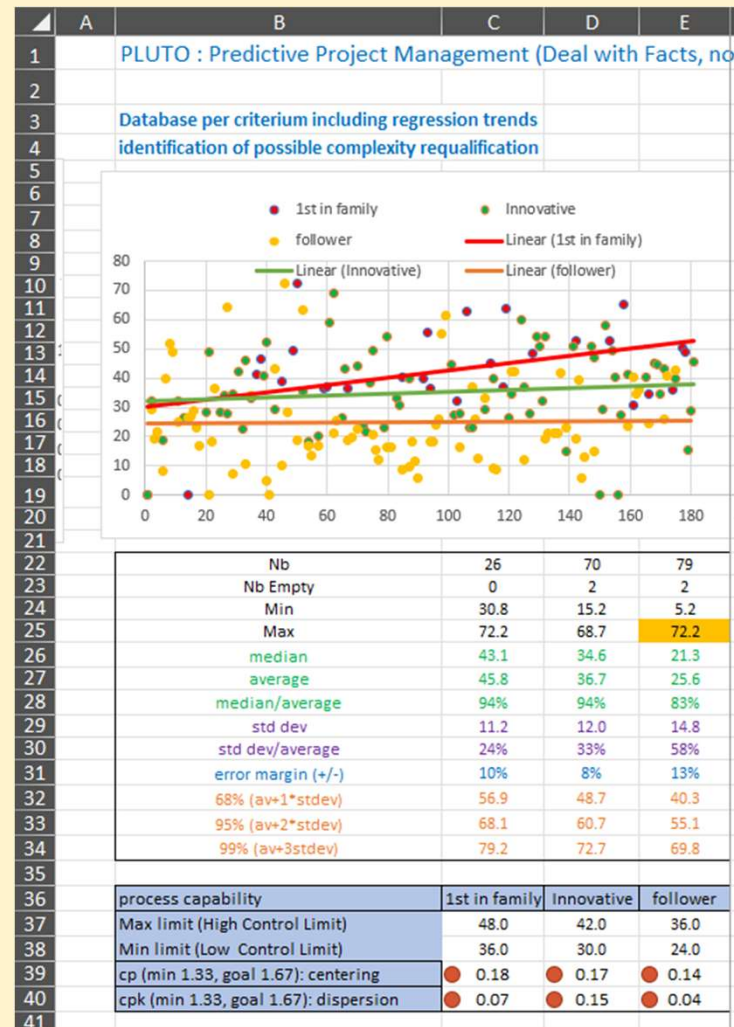


$$=(D37-D38)/(6*D28)$$

$$=MIN((D27-D38)/(3*D28), (D37-D27)/(3*D28))$$

Data only indicative

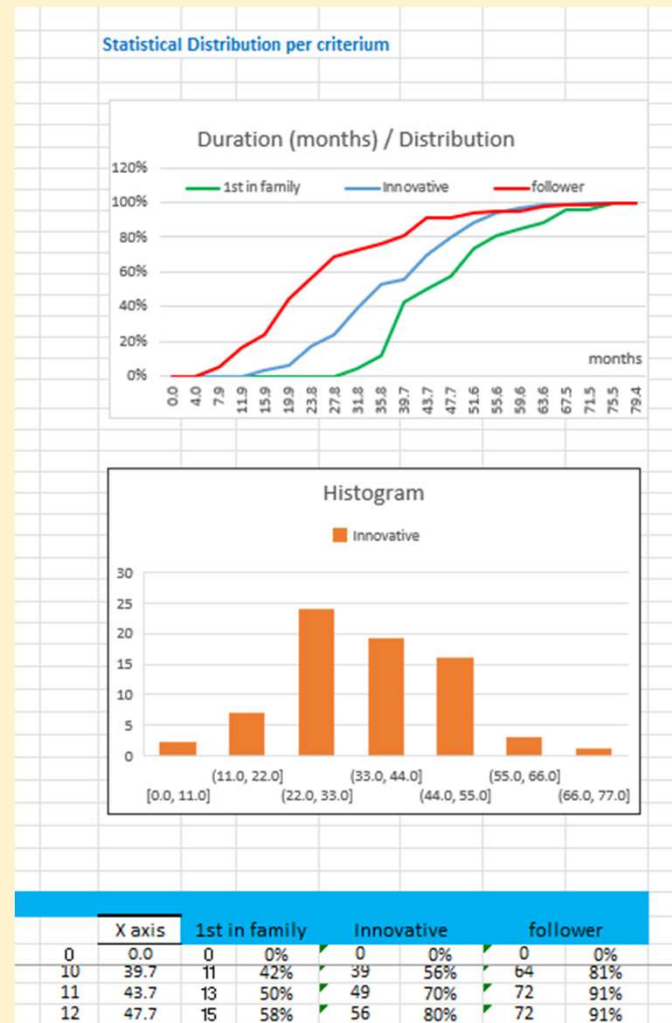
Action 4 : Analysis per KPI Split per criteria



=IF(\$C44=Q\$43,\$M44,NA())

Data only indicative

Action 5 : KPI Distribution Statistical Distribution



=COUNTIF(Q\$43:Q\$412, "<"&W44)

=X44/X\$64

Number of data samples

Few guidelines

The more diverse the process, the more data are required.

Rule 1: Total number of observations : $N \geq 20 \times k$

N = total number of observations;
k = number of relevant categories

Here: $20 \times k = 20 \times 3 = 60$ et $348 \geq 60$

OK

Rule 2: Critical or rare category : $N \times p_{\min} \geq 30$

The category of interest should contain at least **30 observations**.
Where $p_{\min}$ is the minimum frequency of the category under analysis.

Here: $348 \times 0,069 = 24 < 30$

NOOK

To obtain 30 cases in the “high” category while keeping the same proportion of 6.9%: $n \geq 30 / 0,069 \approx 435$
Therefore, approximately **87 additional calls** are required.

Rule 3: Stability of diversity : $H(X) = - \sum p_i * \log_2(p_i)$

Shannon Entropy $H(X)$ measures the average number of bits required to represent or transmit the values of a signal.
The distribution is considered stable if entropy varies by less than **0.05 bit** across three successive checkpoints.

Example 1: two equally frequent categories If there are two categories, each representing 50% of observations: $H = -[0.5 \log_2(0.5) + 0.5 \log_2(0.5)] = 1$

Example 2: two uneven categories If there are two categories, representing 90% and 10%: $H = -[0.9 \log_2(0.9) + 0.1 \log_2(0.1)] = 0.47$

Application: three categories

- $H_{\max} = \log_2(3) = 1,585$ bit
- $H_{348} = 1,08$ bit = $H_{\text{norm}} = 1,084 / 1,585 = 0,68$ (**moderate diversity**)
- $H_{200} = 1.18$, $H_{250} = 1.145$, $H_{300} = 1.12$
- The successive changes are below 0.05 bit: OK

Therefore, the diversity of the distribution can be considered **stable**.

Interpretation

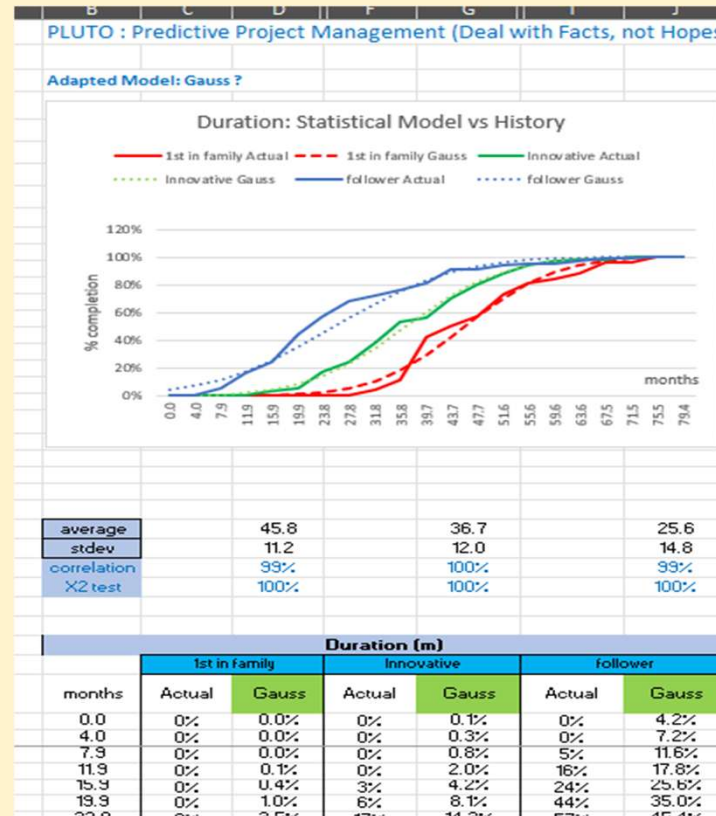
The overall sample size of 348 observations is sufficient for analysing the three predefined categories. The distribution of categories also appears stable.
However, the rare “high” category has only about 24 observations. It is therefore too small for a robust standalone analysis under the chosen 30-observation rule.
Reaching approximately **435 total observations** would provide the minimum of 30 “high” cases, assuming its observed share remains close to 6.9%.

Categories	Number of data	Part
Low	250	71,8%
Mid	74	21,3%
High	24	6,9%
Total	348	100%

Data only indicative

Action 6 : 1st Model Analysis

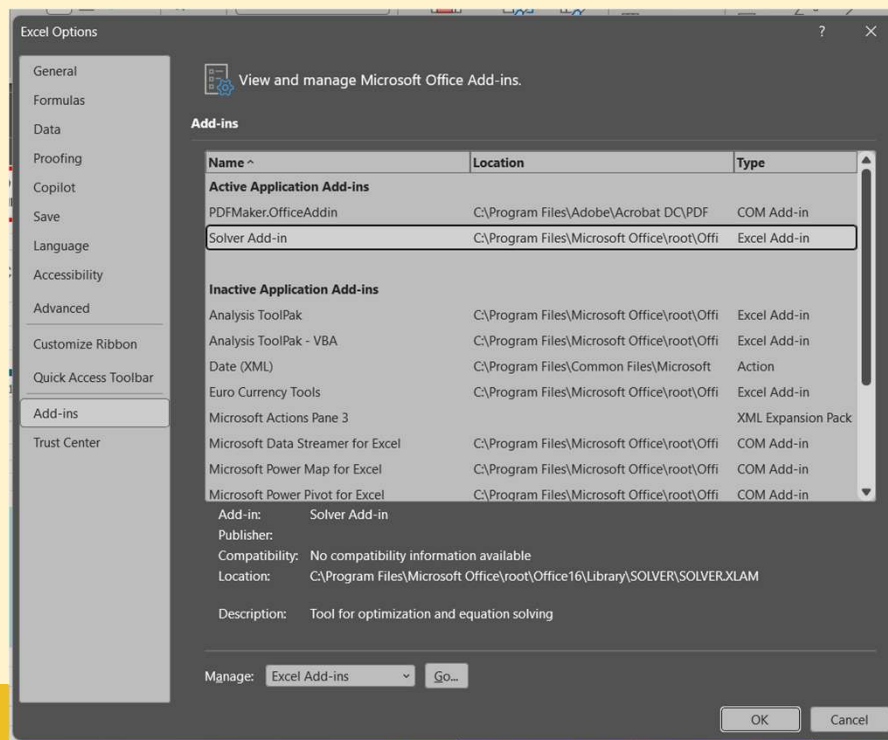
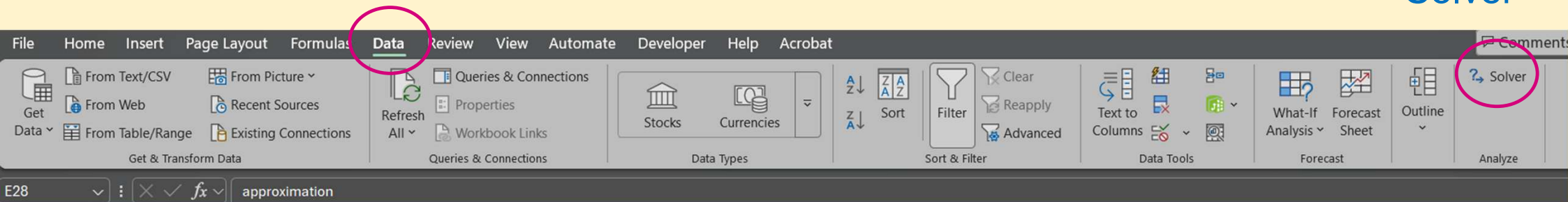
Parameters Selection (Gaussian)



=NORM.DIST(\$AE44,AG\$24,AG\$25,TRUE)

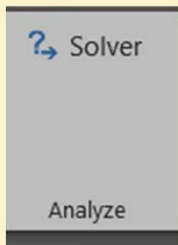
=IFERROR((AG44-AF44)^2,"")

Optimization Solver



Action 7 : 1st Model Optimization

Error Minimization (solver)



Solver Parameters

Set Objective: ↑

To: ☐ Max ☒ Min ☐ Value Of:

By Changing Variable Cells: ↑

Subject to the Constraints:

☒ Make Unconstrained Variables Non-Negative

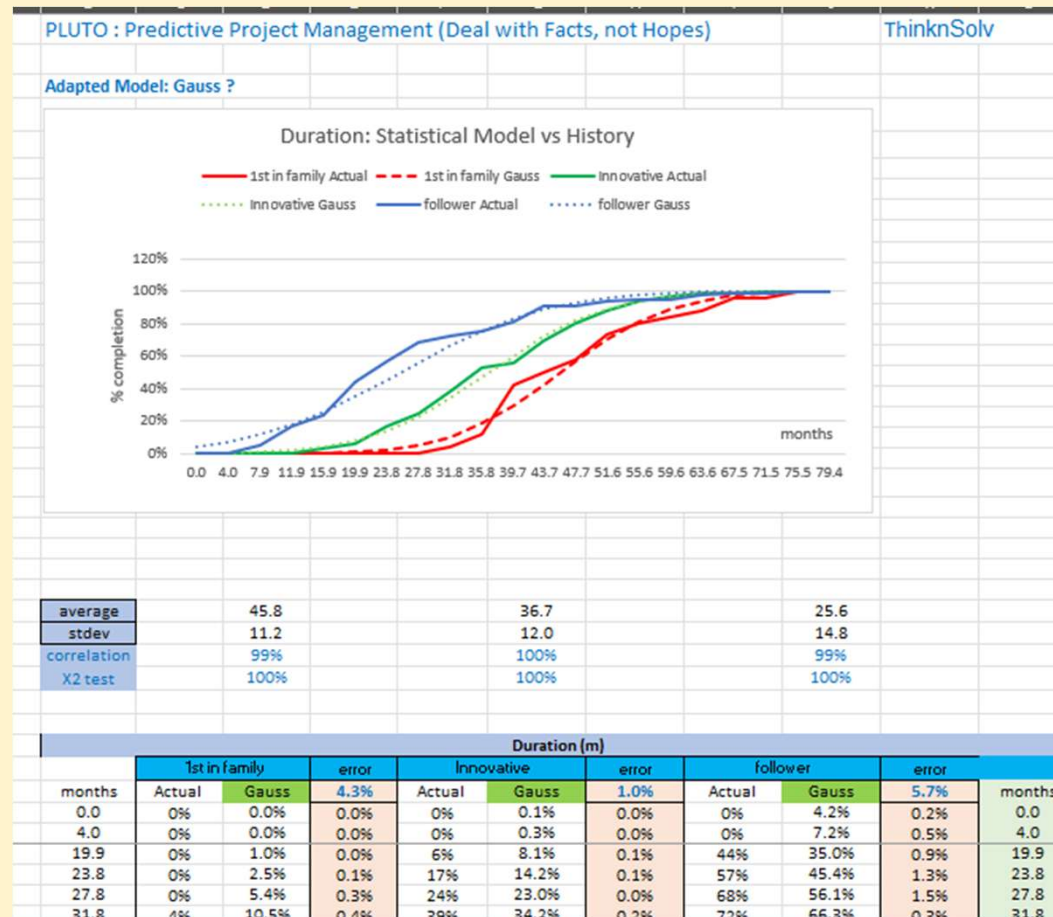
Select a Solving Method: Options

Duration (m)									
months	1st in family			Innovative			follower		
	Actual	Gauss	error	Actual	Gauss	error	Actual	Gauss	error
0.0	0%	0.0%	0.0%	0%	0.1%	0.0%	0%	4.2%	0.2%
4.0	0%	0.0%	0.0%	0%	0.3%	0.0%	0%	7.2%	0.5%
19.9	0%	1.0%	0.0%	6%	8.1%	0.1%	44%	35.0%	0.9%
23.8	0%	2.5%	0.1%	17%	14.2%	0.1%	57%	45.4%	1.3%
27.8	0%	5.4%	0.3%	24%	23.0%	0.0%	68%	56.1%	1.5%
31.8	1%	10.5%	0.1%	39%	34.2%	0.2%	72%	66.3%	0.3%

Data only indicative

Action 8 : Model Optimisation

Gap Minimisation Model vs History



Action 9 : 1st Predictive Model

Gaussian

% Estimate duration vs parameter		
Volume	duration (days)	
moyen	2.0	to fill
Probability	78%	
relative error margin (+/-)	12%	

=IF(\$AQ\$6=Q43,NORM.DIST(\$AR6,AG24,AG25,TRUE),IF(\$AQ\$6=R43,NORM.DIST(AR6,AJ24,AJ25,TRUE),IF(\$AQ\$6=S43,NORM.DIST(AR6,AM24,AM25,TRUE))))

=IF(\$AQ\$6=Q43,Q28,IF(\$AQ\$6=R43,R28,S28))

Statistical Models

Tools Box



Gauss



Gamma



Poisson



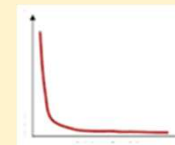
Log Normale



Exponential



Weibull



Power Law-Pareto

Statistical Models

Excel Formulas

```
=IF(EXPON.DIST(N36,R$27,TRUE)<>0,EXPON.DIST(N36,R$27,TRUE),"")
```

```
=IFERROR(IF(LOGNORM.DIST(N36,V$27,V$28,TRUE)<>0,LOGNORM.DIST(N36,V$27,V$28,TRUE),""),"")
```

```
=IFERROR(IF(1-(X$30/N36)^X$29<>0,1-(X$30/N36)^X$29,""),"")
```

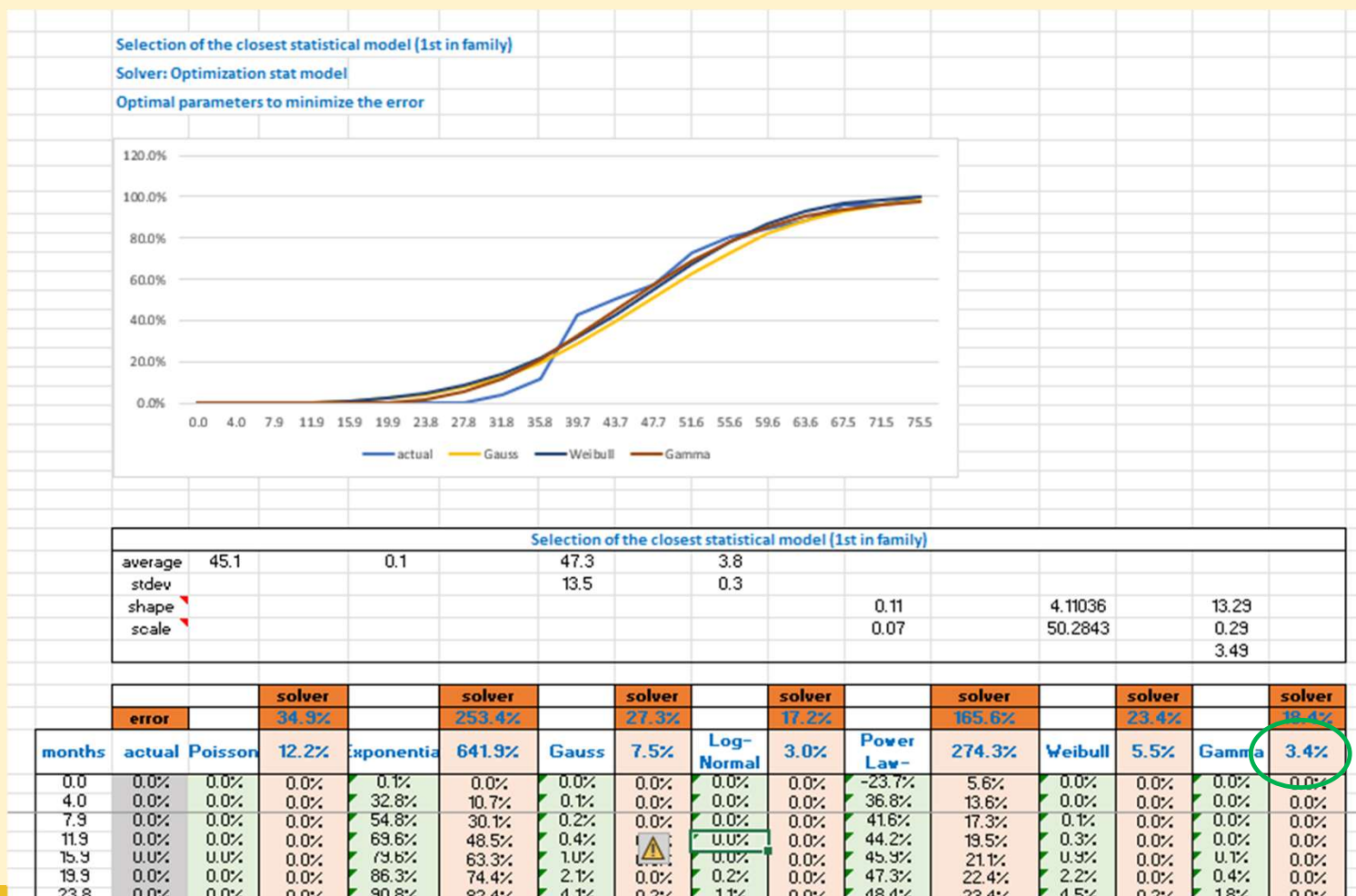
```
=IF(WEIBULL(N36,Z$29,Z$30,TRUE)<>0,WEIBULL(N36,Z$29,Z$30,TRUE),"")
```

```
=IF(GAMMA.DIST(N36,AB$29,AB$31,TRUE)<>0,GAMMA.DIST(N36,AB$29,AB$31,TRUE),"")
```

Data only indicative

Action 10 : Exploration of other Models

Parameters Optimization



Data only indicative

Optimal Model Selection

Khi-2 (χ^2) test



hypothesis: same data between dataset and model

Model	A	B	total
1	68	35	103
2	70	50	120
Total	138	85	223

dataset	A	B	total
1.0	68.0	35.0	103
2.0	70.0	50.0	120
total	138	85	223

Test Xi2

100%

Dependency hypothesis OK
(link between variables > 95%)

=CHISQ.TEST(U3:V5,U7:V9)

hypothesis: different data between dataset and model

model	A	B	total
1	68	35	103
2	70	50	120
total	138	85	223

dataset	A	B	total
1.0	25	70	95
2.0	113	7	120
total	138	77	223

Test Xi2

0%

Dependency hypothesis NOT OK
(link between variables < 95%)

adequation dataset vs reference model : χ^2 test # 100%

Data only indicative

Action 11 : Optimal Model Selection

Khi-2 (χ^2) test

days	actual	Poisson	76.4%	Exponential	49.6%	Gaussian	12.2%	Log-Normal	38.9%	Power Law	58.7%	Weibull	11.4%	Gamma	8.8%
0.0	0.0%	37.7%	14.2%			0.5%	0.0%								
0.2	1.2%	37.7%	13.3%	21.2%	4.0%	2.9%	0.0%	0.0%	0.0%	-16.9%	3.3%	1.8%	0.0%	0.5%	0.0%
0.4	6.6%	37.7%	9.6%	37.9%	9.8%	11.3%	0.2%	2.0%	0.2%	41.0%	11.8%	11.4%	0.2%	8.9%	0.1%
0.7	31.5%	37.7%	0.4%	51.1%	3.8%	29.8%	0.0%	14.7%	2.8%	60.4%	8.4%	31.1%	0.0%	31.4%	0.0%
0.9	58.9%	37.7%	4.5%	61.4%	0.1%	55.9%	0.1%	37.0%	4.8%	70.2%	1.3%	56.2%	0.1%	57.9%	0.0%
1.1	78.8%	74.4%	0.2%	69.6%	0.9%	79.6%	0.0%	58.9%	4.0%	76.1%	0.1%	78.4%	0.0%	78.1%	0.0%
1.3	90.5%	74.4%	2.6%	76.1%	2.1%	93.4%	0.1%	75.2%	2.3%	80.0%	1.1%	92.1%	0.0%	90.0%	0.0%
1.5	94.6%	74.4%	4.1%	81.1%	1.8%	98.6%	0.2%	85.6%	0.8%	82.8%	1.4%	98.0%	0.1%	95.9%	0.0%
1.8	94.6%	74.4%	4.1%	85.1%	0.9%	99.8%	0.3%	91.9%	0.1%	85.0%	0.9%	99.6%	0.3%	98.4%	0.1%
2.0	95.4%	74.4%	4.4%	88.3%	0.5%	100.0%	0.2%	95.5%	0.0%	86.6%	0.8%	100.0%	0.2%	99.4%	0.2%
2.2	95.4%	92.4%	0.1%	90.8%	0.2%	100.0%	0.2%	97.5%	0.0%	87.9%	0.6%	100.0%	0.2%	99.8%	0.2%
2.4	96.7%	92.4%	0.2%	92.7%	0.2%	100.0%	0.1%	98.6%	0.0%	89.0%	0.6%	100.0%	0.1%	99.9%	0.1%
2.6	97.9%	92.4%	0.3%	94.3%	0.1%	100.0%	0.0%	99.2%	0.0%	89.9%	0.6%	100.0%	0.0%	100.0%	0.0%
2.8	98.3%	92.4%	0.4%	95.5%	0.1%	100.0%	0.0%	99.5%	0.0%	90.7%	0.6%	100.0%	0.0%	100.0%	0.0%
3.1	99.2%	98.2%	0.0%	96.4%	0.1%	100.0%	0.0%	99.7%	0.0%	91.3%	0.6%	100.0%	0.0%	100.0%	0.0%
3.3	99.6%	98.2%	0.0%	97.2%	0.1%	100.0%	0.0%	99.8%	0.0%	91.9%	0.6%	100.0%	0.0%	100.0%	0.0%
3.5	99.6%	98.2%	0.0%	97.8%	0.0%	100.0%	0.0%	99.9%	0.0%	92.4%	0.5%	100.0%	0.0%	100.0%	0.0%
3.7	100.0%	98.2%	0.0%	98.3%	0.0%	100.0%	0.0%	99.9%	0.0%	92.8%	0.5%	100.0%	0.0%	100.0%	0.0%
3.9	100.0%	98.2%	0.0%	98.6%	0.0%	100.0%	0.0%	100.0%	0.0%	93.2%	0.5%	100.0%	0.0%	100.0%	0.0%
4.2	100.0%	99.7%	0.0%	98.9%	0.0%	100.0%	0.0%	100.0%	0.0%	93.6%	0.4%	100.0%	0.0%	100.0%	0.0%
test correlation		90%		97%		100%		98%		91%		100%		100%	
Test Xi2		100.00%		100.00%		100.00%		100.00%		100.00%		100.00%		100.00%	

=CORREL (\$AV44:\$AV63,AW44:AW63)

=CHISQ.TEST(\$AV44:\$AV63,AW44:AW63)

Process Capability Control

6 Sigma Method

A process needs to be both precise and centered to be truly capable.

◆ **Cp (Process Capability Index):** Measures potential capability. It assumes the process is centered between specification limits. It compares the spread of the process to the tolerance (USL - LSL).
Formula: $Cp = (USL - LSL) / (6 \times \text{Standard Deviation})$

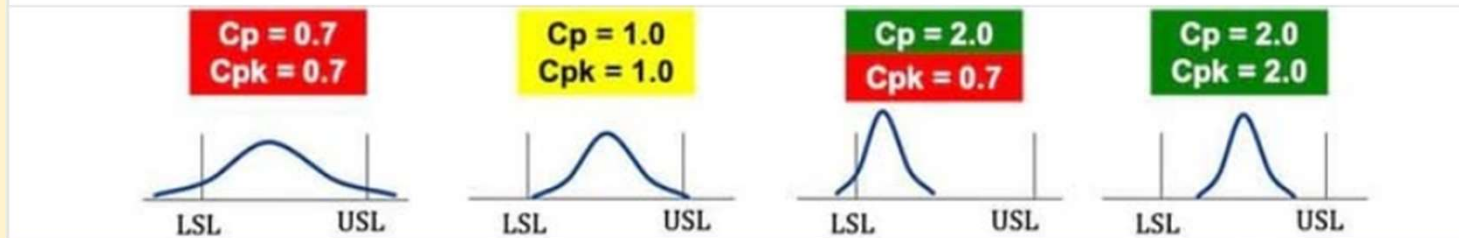
◆ **Cpk (Process Capability Index adjusted for centering):** Measures actual capability. It considers whether the process is centered. It takes into account how close the process is to the USL or LSL.

Formula: $Cpk = \min(Cpl, Cpu)$

where :

$Cpl = (\text{Process Mean} - LSL) / (3 \times \text{Standard Deviation})$

$Cpu = (USL - \text{Process Mean}) / (3 \times \text{Standard Deviation})$



Data only indicative

Action 11: Processus Capability Control

6 Sigma Method

process capability	1st in family	Innovative	follower
Max limit (High Control Limit)	48.0	42.0	36.0
Min limit (Low Control Limit)	36.0	30.0	24.0
cp (min 1.33, goal 1.67): centering	● 0.18	● 0.17	● 0.14
cpk (min 1.33, goal 1.67): dispersion	● 0.07	● 0.15	● 0.04

Monte-Carlo Simulation

A Monte Carlo simulation is a statistical technique used to model the probability of different outcomes in a process that is uncertain and complex. It involves running a large number of simulations (random experiments) to predict the range of possible results based on a set of input variables and their probability distributions. This method allows us to understand the impact of risk and uncertainty in models and is often applied in fields like finance, engineering, project management, and science.

How It Works

1. Define the Model: Set up a model that represents the system or process you want to analyze. This model should include key variables that affect the outcome.

2. Assign Probabilities: For each variable, assign a probability distribution (like normal, uniform, or exponential) that represents the uncertainty around that variable.

3. Simulate Outcomes: Using a random number generator, "sample" values for each variable according to their assigned distributions, and then calculate the result. This is repeated thousands or even millions of times.

4. Analyze Results: The simulations produce a range of possible outcomes, and you can calculate statistics, like the mean, variance, and confidence intervals, to understand the likelihood of various results.

Benefits: It captures a range of possible outcomes, is flexible, and allows for complex modeling.

Limitations: It can be computationally intensive, and accuracy is only as good as the model and input data.

Data only indicative

Action 12: Monte-Carlo Simulation

KPI Forecast Example

For project management KPIs, Monte Carlo is often used to forecast:

- Schedule completion probability.
- Budget overrun probability.
- Resource utilization risk.
- Delay risk for individual milestones.

A project manager can use the simulation to decide:

- Whether the current deadline is realistic.
- How much contingency budget to reserve.
- Whether to add resources or reduce scope.
- Which milestone is most likely to slip.

Monte-Carlo simulation uses the optimal model defined step 11 “comparison”. It is more precise than the history database due to the higher number of trials.

Prediction: Monte-Carlo Simulation (Duration)				
Dumation (months)	1st in family	Innovative	Follower	
average	45.8	36.7	25.6	
std dev	11.2	12.0	14.8	
nb	1000	1000	1000	
max	77.7	71.1	67.8	
confidence 80%	54.5	46.1	37.1	
confidence 50%	45.1	35.9	24.6	
confidence 20%	36.5	26.7	13.3	
min	10.3	-1.5	-21.4	
error margin (+/-)	2%	2%	4%	
std dev	10.9	11.7	14.4	
trial	random nb	1st in family	Innovative	Follower
1	65%	49.9	41.2	31.1
2	70%	51.5	42.9	33.2
3	19%	36.0	26.2	12.7
4	3%	25.0	14.3	-1.9

Data only indicative

Action 13 – Prediction

Selected Statistical Model



Target Project						
Project name	Complexity Innovation	duration (years)	Effort (mxy)	Revisions (nb)	Cost (k€)	Target Success %
Pluto	Innovative	2.5	6	3	2500	80%
Project success probability	29%	29%	42%	48%	51%	Not OK
relative error margin (+/-) on KP	12%					

Action 14 - IA Prompt

Explicit Workflow

the objective is to extract a predictive statistical model from the attached history database.

Process step by step

- 1) remove outliers (explain criteria),
- 2) extract key parameters (number, min, max, median, average, standard deviation, margin of error)
- 3) define limits min and max for the data to keep to generate the model,
- 4) split the table by complexity (1st in family, innovative, follower)
- 5) extract the same key parameters for each category
- 6) propose a gaussian/normal statistical model , indicate average and standard dev for each complexity minimizing the error vs the extracted history database (within min/max limits)
- 7) propose another statistical model which could better fit, objective minimum error between the history distribution and the proposed model, for each complexity.
- 8) suggested models: Poisson, Exponential, Gauss Normal, Log-Normal, Power Law-Pareto, Weibull, Gamma. Other models OK too.

Use Python for the calculations.

Report:

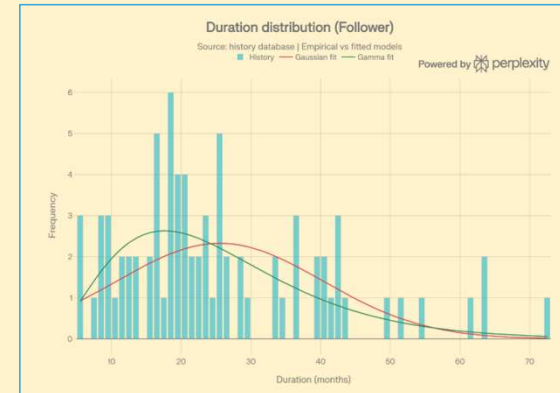
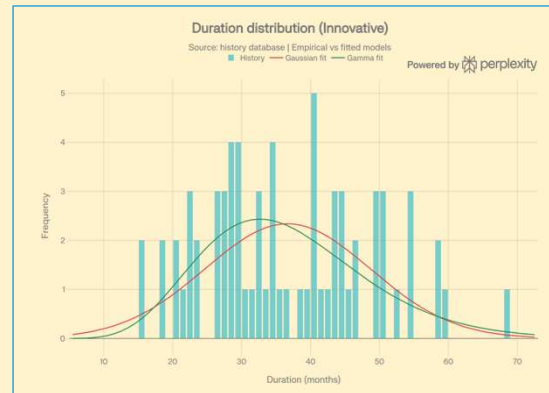
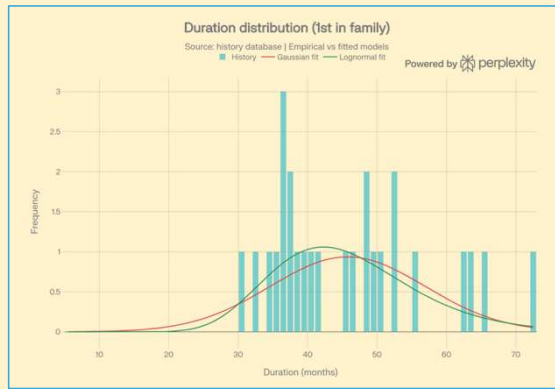
Follow the process. Provide a preliminary report at each step.

Generate a graph comparing the historical distribution for each class and the optimized statistical model.

is it clear ?

Data only indicative

AI report extract



Complexity	Distribution	Param 1 name	Param 1 value	Param 2 name	Param 2 value	Param 3 name	Param 3 value	Notes
1st in family	Gaussian	Mean (μ)	45.76301242	Std (σ)	11.15152924			Use NORM.DIST with mean, sd
Follower	Gaussian	Mean (μ)	25.55414017	Std (σ)	14.75273646			Use NORM.DIST with mean, sd
Innovative	Gaussian	Mean (μ)	36.67155426	Std (σ)	12.00361443			Use NORM.DIST with mean, sd
1st in family	Log-normal	Shape	0.22904017	Loc	0	Scale	44.54815055	lognorm(shape, loc, scale)
Follower	Gamma	Shape (a)	3.2189102	Loc	0	Scale	7.93875522	gamma(a, loc, scale)
Innovative	Gamma	Shape (a)	9.1505742	Loc	0	Scale	4.00756864	gamma(a, loc, scale)

Data only indicative

Action 14: Comparison AI vs DA

IA a bit more precise and much faster (if precise prompt)

AI model

Complexity	Distribution	Param 1 name	Param 1 value	Param 2 name	Param 2 value
1st in family	Gaussian	Mean (μ)	45.8	Std (σ)	11.2
Follower	Gaussian	Mean (μ)	25.6	Std (σ)	14.8
Innovative	Gaussian	Mean (μ)	36.7	Std (σ)	12.0

Excel Gaussian

average
stdev

1st in family

47.38

13.5

Innovative

36.3

13.0

follower

23.8

13.4

Data only indicative



Pluto research summary
Document - PDF



Open in Adobe Acrobat

Action 15: AI Report

Thinknsolv.net

PLUTO

Predictive Project Management — Full Methodology & Results Report

1. HISTORICAL DATASET

151 completed R&D projects — 3 complexity classes, duration in months. Two extreme outliers removed (1,203 mo and -237 mo) via IQR-k cleaning, leaving 149 clean records. The IQR coefficient k is user-selectable (default 2.5) in the HTML tool.

Complexity	n raw	n clean	Mean (mo)	Std Dev	CV	Error margin	Pearson r
1st in Family	21	-20	48.2	13.9	29%	±19%	0.9935
Innovative	57	-56	36.4	12.6	35%	±12%	0.9984
Follower	72	-70	25.5*	14.9	58%	±11%	0.9887

* After removal of outliers, DB reference values (n=27/70/79) from full Excel dataset.

2. DATA CLEANING — IQR * k METHOD

Outlier detection uses the IQR * k method (default k=2.5, user-selectable 1.0-5.0). Projects outside [Q1-k*IQR, Q3+k*IQR] are excluded, as are negative values and nulls. At k=2.5: Q1=19.8 mo, Q3=43.2 mo, IQR=23.4 mo, upper limit=101.8 mo. The standard k=1.5 (Tukey) is more aggressive; k=2.5 preserves more valid high-duration projects.

3. STATISTICAL MODEL COMPARISON — 6 MODELS

Six distributions are fitted per complexity class. Error metric: CDF mean squared error (lower = better fit). Parameters estimated via method of moments. Excel reference uses Solver optimisation (exact MLE/LSQ). Gaussian is retained as primary model for interpretability.

Model	1st in Family	Innovative	Follower	Best for
Gaussian	0.27325	0.09854	0.05738	—
Log-Normal	* 0.17245	0.12392	—	1st in Fam
Weibull	0.23405	* 0.09154	—	Innovative
Gamma	0.18380	0.10131	—	—
Poisson	0.34934	0.43918	—	—
Exponential	2.53354	1.99880	—	—

* = best fit per complexity (Excel Solver). Poisson and Exponential are consistently poor fits for R&D project durations.

Gaussian model validation statistics (from Excel):

Complexity	Pearson r	Chi-square d	Cp	Cpk	Interpretation
1st in family	0.9935	0.99999	0.179	0.067	High variability — normal for R&D.
Innovative	0.9984	1.00000	0.167	0.148	High variability — normal for R&D.
Follower	0.9887	0.99999	0.136	0.038	High variability — normal for R&D.

Cp/Cpk < 1.33 indicates high process variability — expected for R&D durations. Pearson > 0.99 confirms Gaussian is a valid working model despite not being the best fit.

4. KPI SUCCESS PROBABILITIES — PLUTO PROJECT (INNOVATIVE)

$P(\text{success}) = \Phi((\text{target} - \mu) / \sigma)$. Overall probability = min of all four KPIs — each is a binding constraint. Duration is the binding factor: 2.5 yr target is 18% below DB average 3.1 yr.

KPI	Pluto Target	DB Average	DB Std Dev	P(success)
Duration	2.5 yr (30 mo)	36.7 mo / 3.1 yr	12.0 mo	23%
Effort	6 mxy	7.2 mxy	6.3 mxy	42%
Revisions	3	3.1	1.1	49%
Cost	2,500 k€	2,469 k€	959 k€	61%
Overall (min)	—	—	—	23%

PLUTO — Predictive Project Management | ThinknSolv | Full Methodology Report

To reach 80% success on duration, target must be extended to ~3.8 years ($= \mu + \sigma \cdot \text{norminv}(0.80) = 36.7 + 12.0 \cdot 0.842$).

5. RESOURCE LOADING & HIRING PLAN — 110% HARD CAP

Six active projects (A, B, C ongoing + Pluto, D, E new). Utilisation capped at ≤110% per quarter. Values corrected from Excel (exclusive layers A+B+C / Pluto / D+E — previous version double-counted Pluto). Extra hires computed as $\lceil \text{load} - \text{HC} \cdot 1.10 \rceil / 1.10$.

Quarter	Base HC	A+B+C	Pluto	D+E	Total	Util%	Extra Hire	Status
3Q25	6	6.60	0.00	0.00	6.60	110%	0	tight
4Q25	7	4.97	1.74	0.00	6.71	96%	0	OK
1Q26	8	4.33	5.39	0.00	9.72	121%	+1	HIRE
2Q26	8	3.35	6.10	0.00	9.45	118%	+1	HIRE
3Q26	9	1.86	5.45	2.09	9.40	104%	0	tight
4Q26	9	0.90	2.83	5.80	9.53	106%	0	tight
1Q27	9	0.00	5.29	6.35	11.54	129%	+2	HIRE
2Q27	9	0.00	4.17	7.17	11.34	126%	+2	HIRE
3Q27	10	0.00	1.96	3.80	5.76	58%	0	OK
4Q27	11	0.00	4.14	6.37	12.51	114%	+1	HIRE
1Q28	12	0.00	2.80	9.41	12.21	102%	0	tight
2Q28	12	0.00	2.83	5.04	7.88	66%	0	OK

Mandatory hires: 1Q26: +1 → HC 9 | 2Q26: +1 → HC 9 | 1Q27: +2 → HC 11 | 2Q27: +2 → HC 11 | 4Q27: +1 → HC 12

Peak utilisation: 129% — 5 quarter(s) exceed 110% — total additional hires: 7 — recommended SoD: 1Q26.

6. START DATE RECOMMENDATION

Parameter	Target	Recommended	Status
Start of Design (SoD)	ASAP	1Q26	Feasible (with hiring)
Start of Production (SoP)	4Q28	1Q29	+1 quarter late
Duration	2.5 yr (DB avg)	3.1 yr (DB avg)	Extend target
Effort	6 mxy	7.2 mxy (DB avg)	Increase budget
Tech Platform	Test Vehicle+	Samples (stage 10)	Already met
Loading cap	≤110%	With hiring plan	8 quarters need hire
KPI success probability	80%	29% (current targets)	Revise KPIs

7. MONTE CARLO SIMULATION

1,000 trials per complexity via Box-Muller transform on $N(\mu, \sigma)$. P20/P50/P80 provide conservative, median and optimistic planning ranges.

Complexity	DB Mean	P20 — conservative	P50 — median	P80 — optimistic
1st in Family	48.2 mo	~36 mo	~48 mo	~60 mo
Innovative	36.7 mo	~26 mo	~37 mo	~47 mo
Follower	25.5 mo	~12 mo	~26 mo	~39 mo

8. MODEL LIMITATIONS

- KPI independence: min() operator may understate joint probability — duration & effort are correlated.
- Double-counting bug (corrected): original HTML had Pluto in both LOAD_ONGOING and LOAD_PLUTO layers. Fixed to exclusive A+B+C / Pluto / D+E layers matching Excel exactly.
- Stationarity: historical distribution assumed stable — organisational or market shifts not captured.
- Right-skew: Follower CV=58% — Log-Normal/Weibull likely outperform Gaussian; confirmed by Excel errors.
- Poisson/Exponential: poor fits for R&D; duration (errors 0.35-2.53 vs Gaussian 0.06-0.27). Included for completeness.
- Sample size: 1st in Family n=20 gives ±19% error margin. DB reference uses full Excel dataset (n=27/70/79) but HTML RAW array has 27 fewer records — to be reconciled.

PLUTO — Predictive Project Management | ThinknSolv | Full Methodology Report

Copyright 2026, ThinknSolv.net

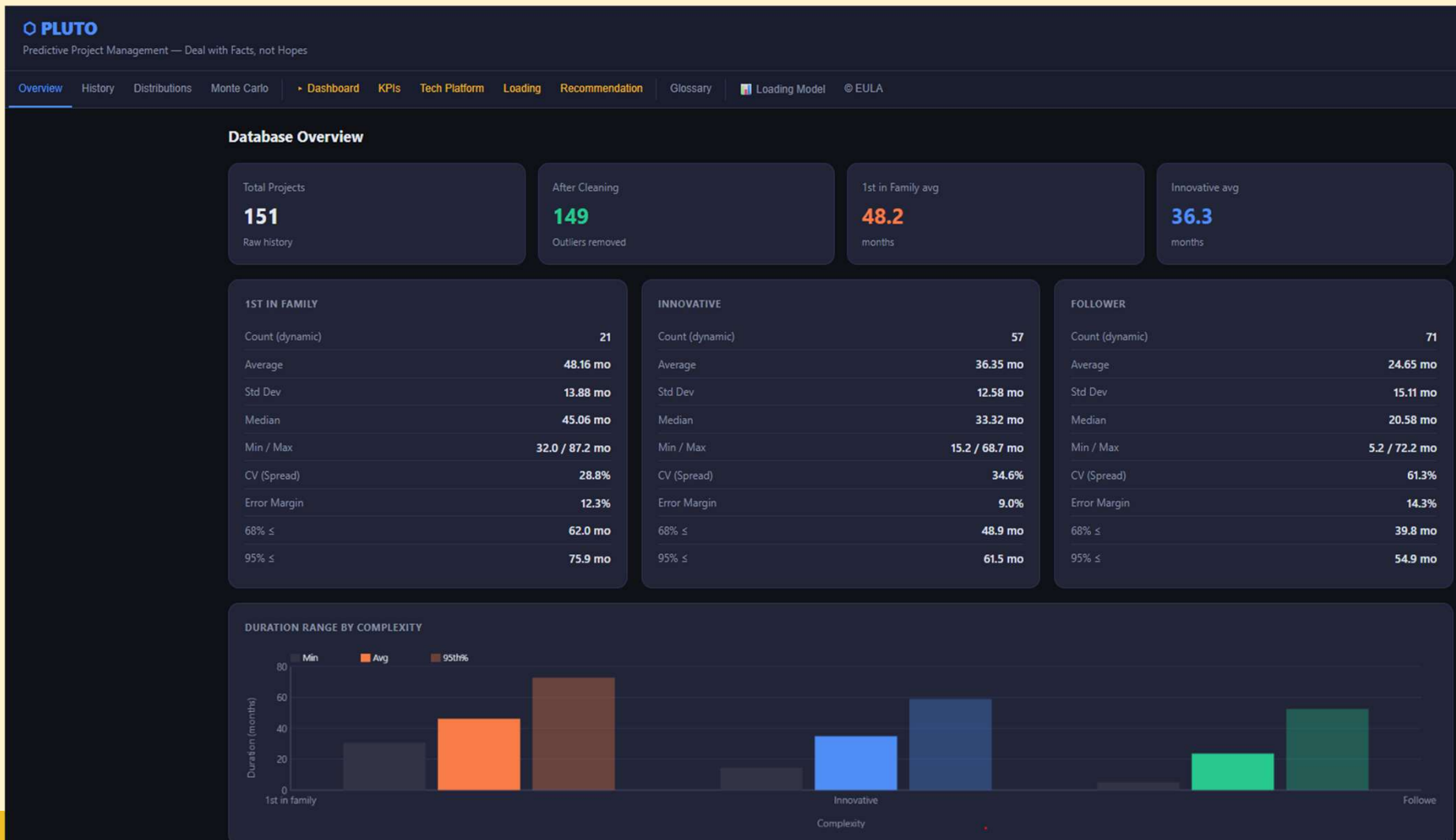
All Rights Reserved- laurentperier@thinknsolv.com

Data only indicative

 Pluto startdate
Code · HTML

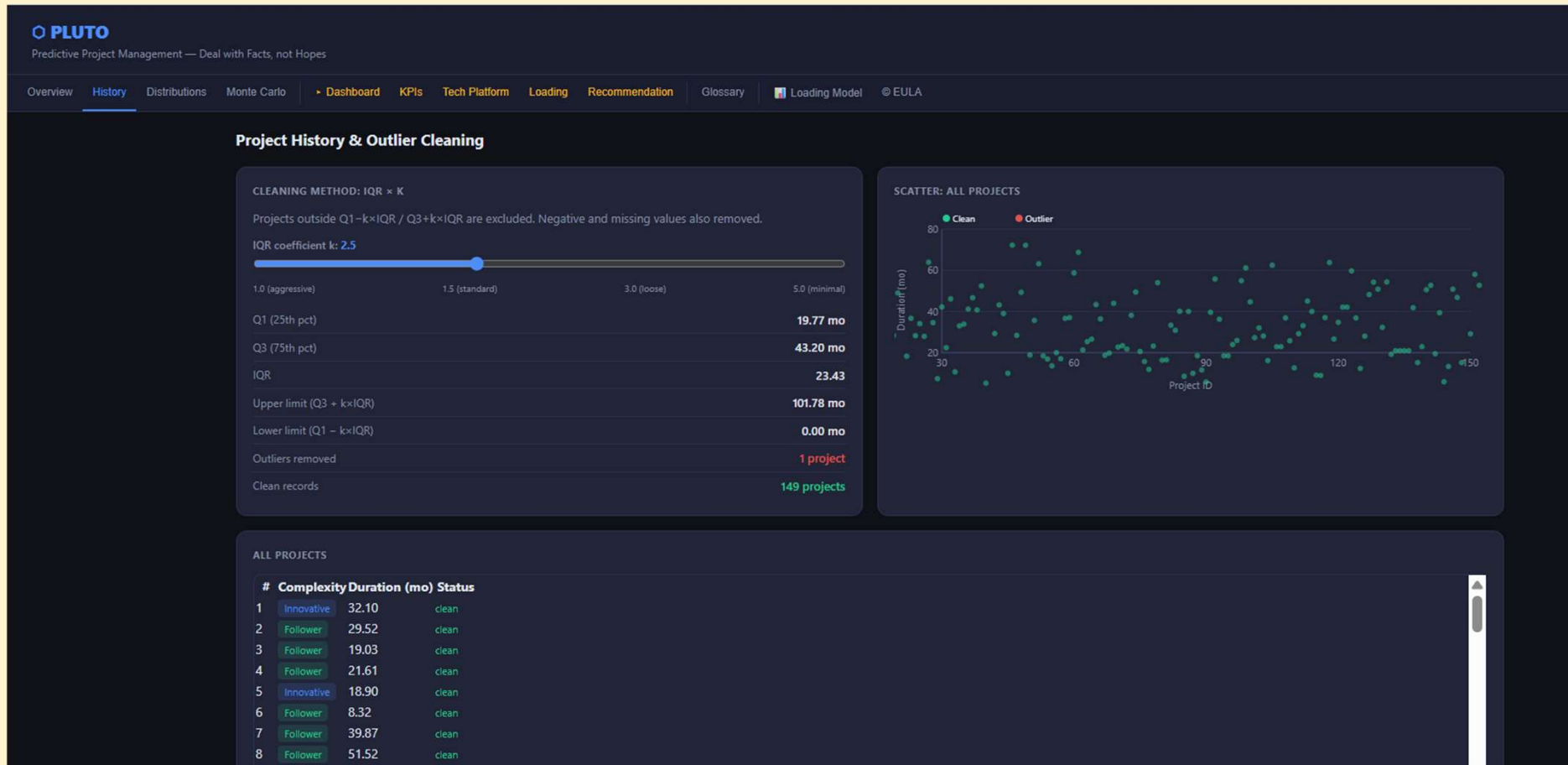
 Open in Google Chrome

Action 16: AI Web Page Overview



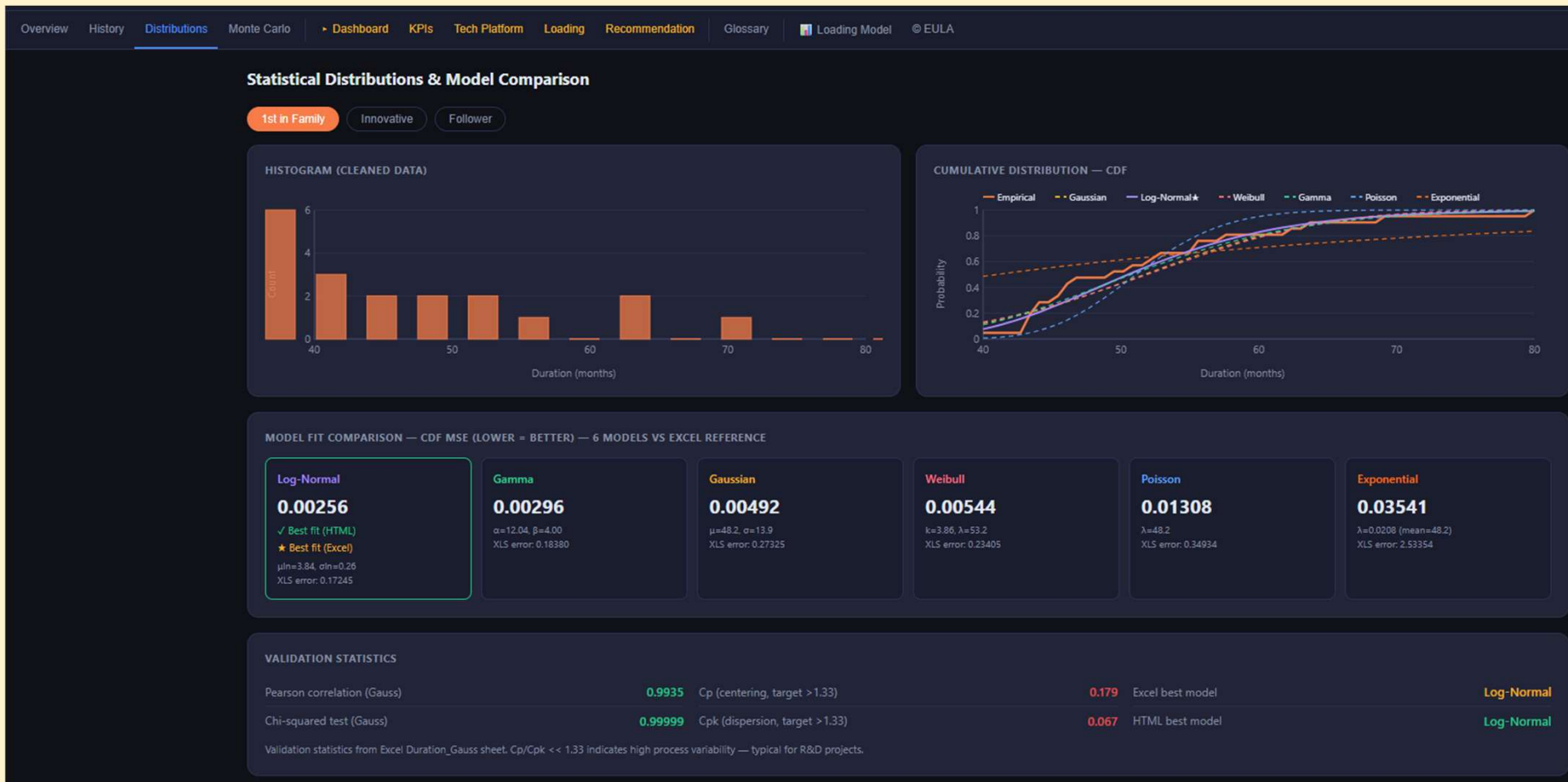
Data only indicative

Action 16: AI Web Page History



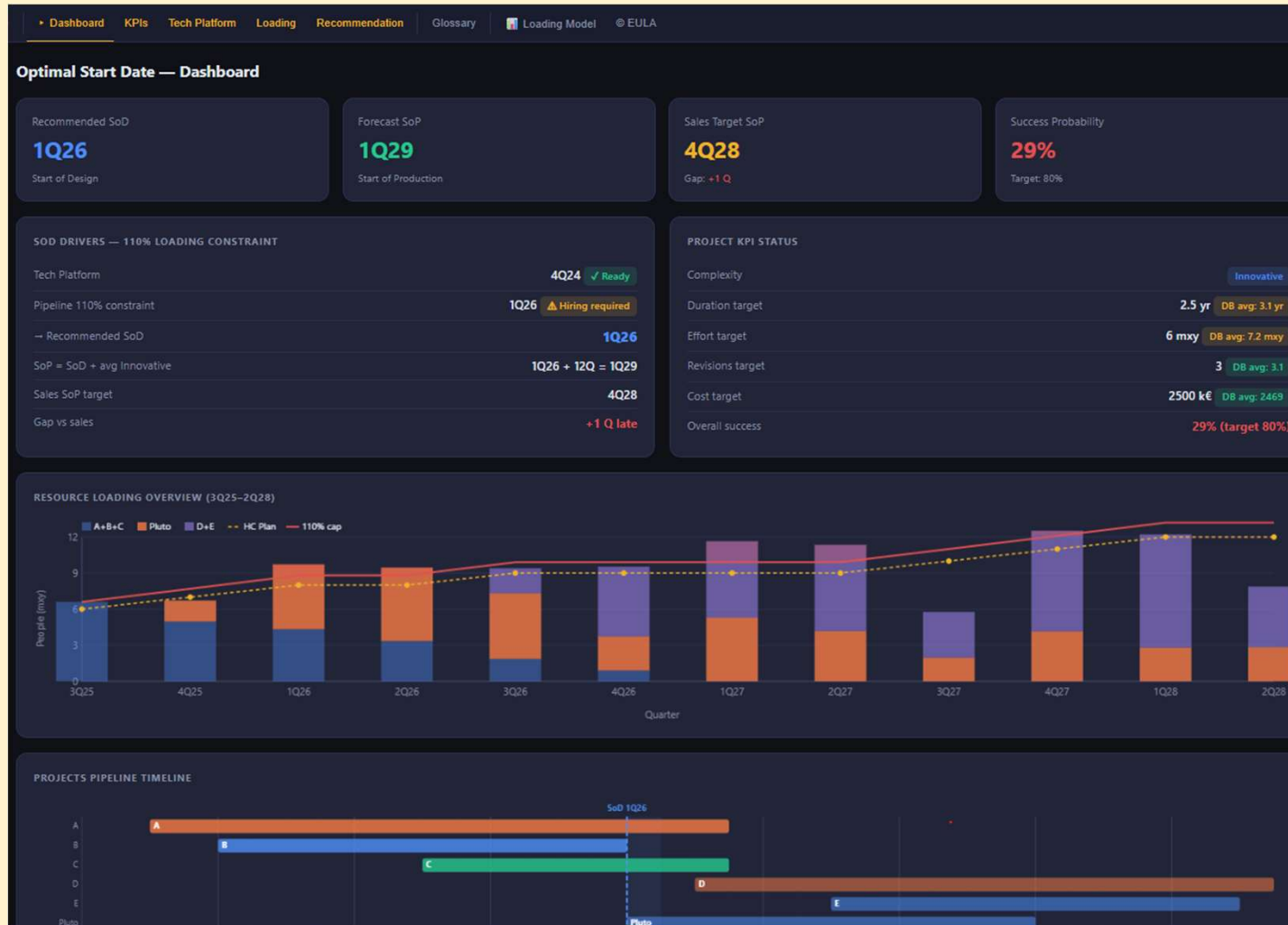
Data only indicative

Action 15: AI Web Page Distributions



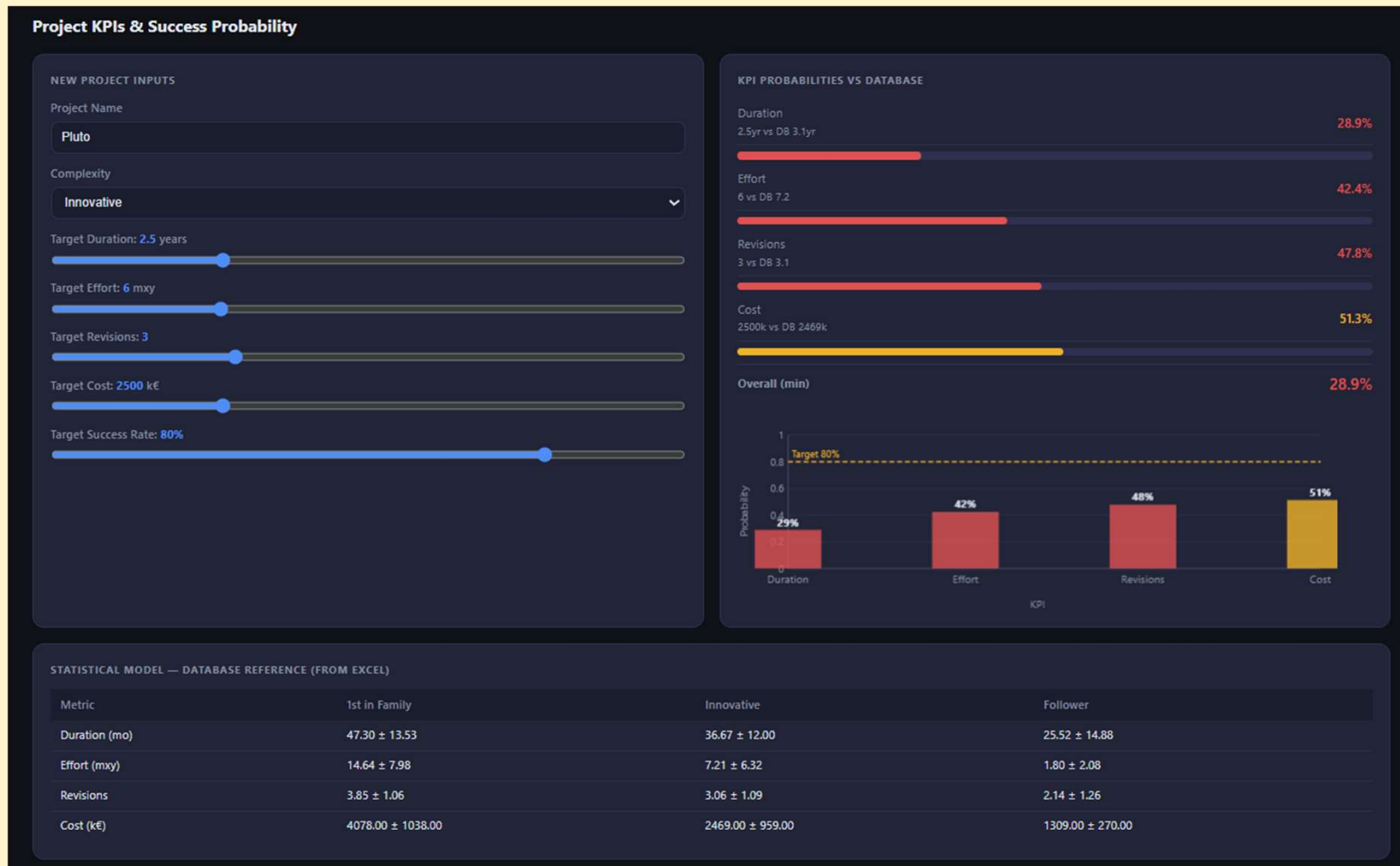
Data only indicative

Action 15: AI Web Page Dashboard



Data only indicative

Action 15: AI Web Page KPIs



Data only indicative

Action 15: AI Web Page Loading

Resource Loading & Pipeline

R&D HEADCOUNT PLAN (BASE)

Year	Total HC	Design	Indus
2025	6→7	4→5	2
2026	8→9	6	2
2027	9→11	7	2
2028	12	9	3

🔒 Hard cap: utilization ≤ 110% — extra hires triggered automatically.

110% CONSTRAINT ANALYSIS

Recommended SoD	1Q26
Design max util	108% ✓
Indus max util	286% ⬆ 3Q25,4Q25,1Q26,2Q26,3Q26,4Q26,1Q27,2Q27,3Q27,4Q27,1Q28,2Q28
Indus 3Q25	+2 → HC 4
Indus 4Q25	+2 → HC 4
Indus 1Q26	+4 → HC 6
Indus 2Q26	+2 → HC 4
Indus 3Q26	+1 → HC 3
Indus 4Q26	+1 → HC 3
Indus 1Q27	+2 → HC 4
Indus 2Q27	+2 → HC 4
Indus 3Q27	+1 → HC 3
Indus 4Q27	+3 → HC 6
Indus 1Q28	+1 → HC 4
Indus 2Q28	+1 → HC 4

★ DESIGN — RESOURCE LOADING (HC START: 6 • 75% PLACEHOLDER)

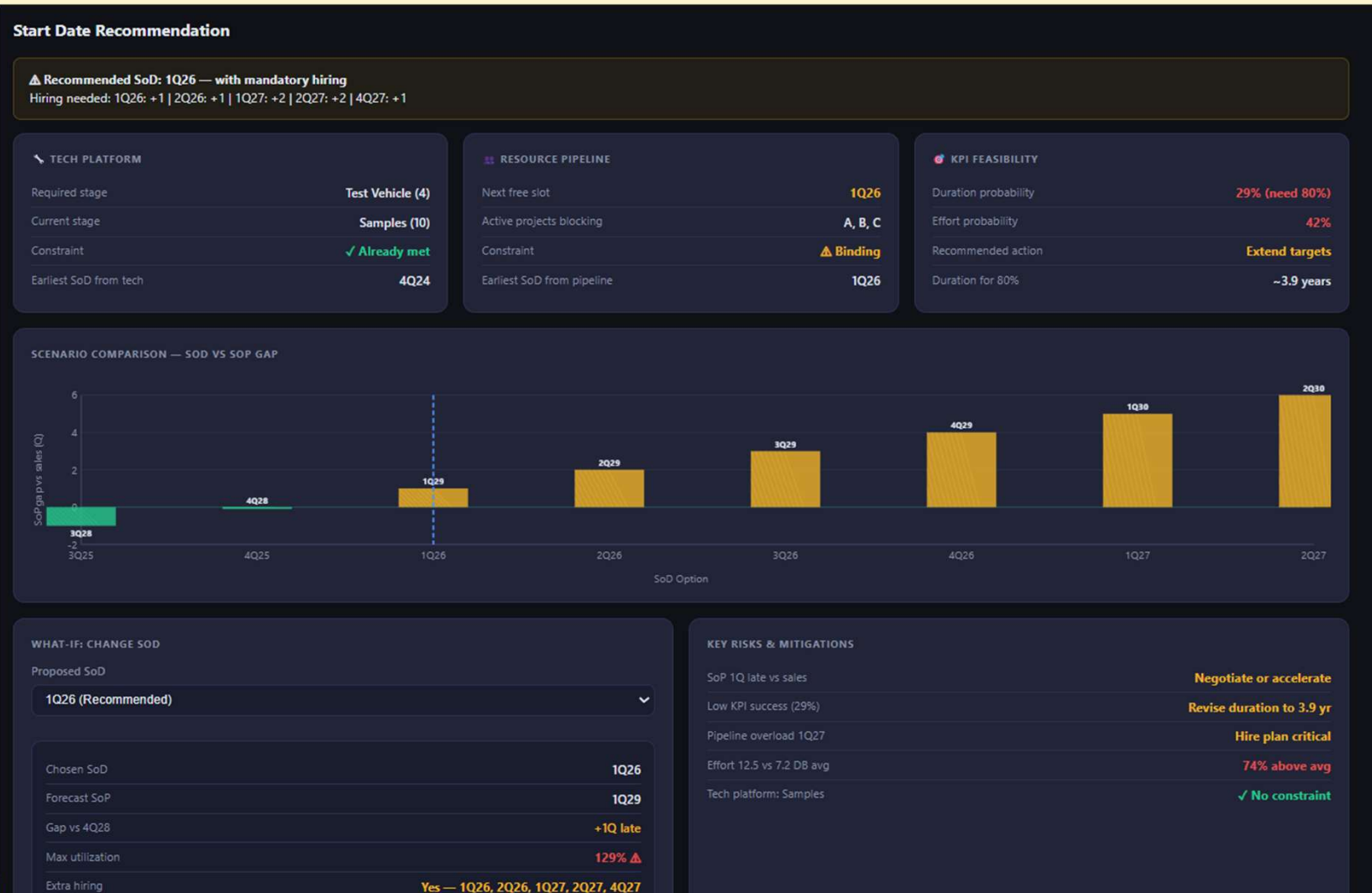


🏭 INDUSTRIALISATION — RESOURCE LOADING (HC START: 2 • 25% PLACEHOLDER)



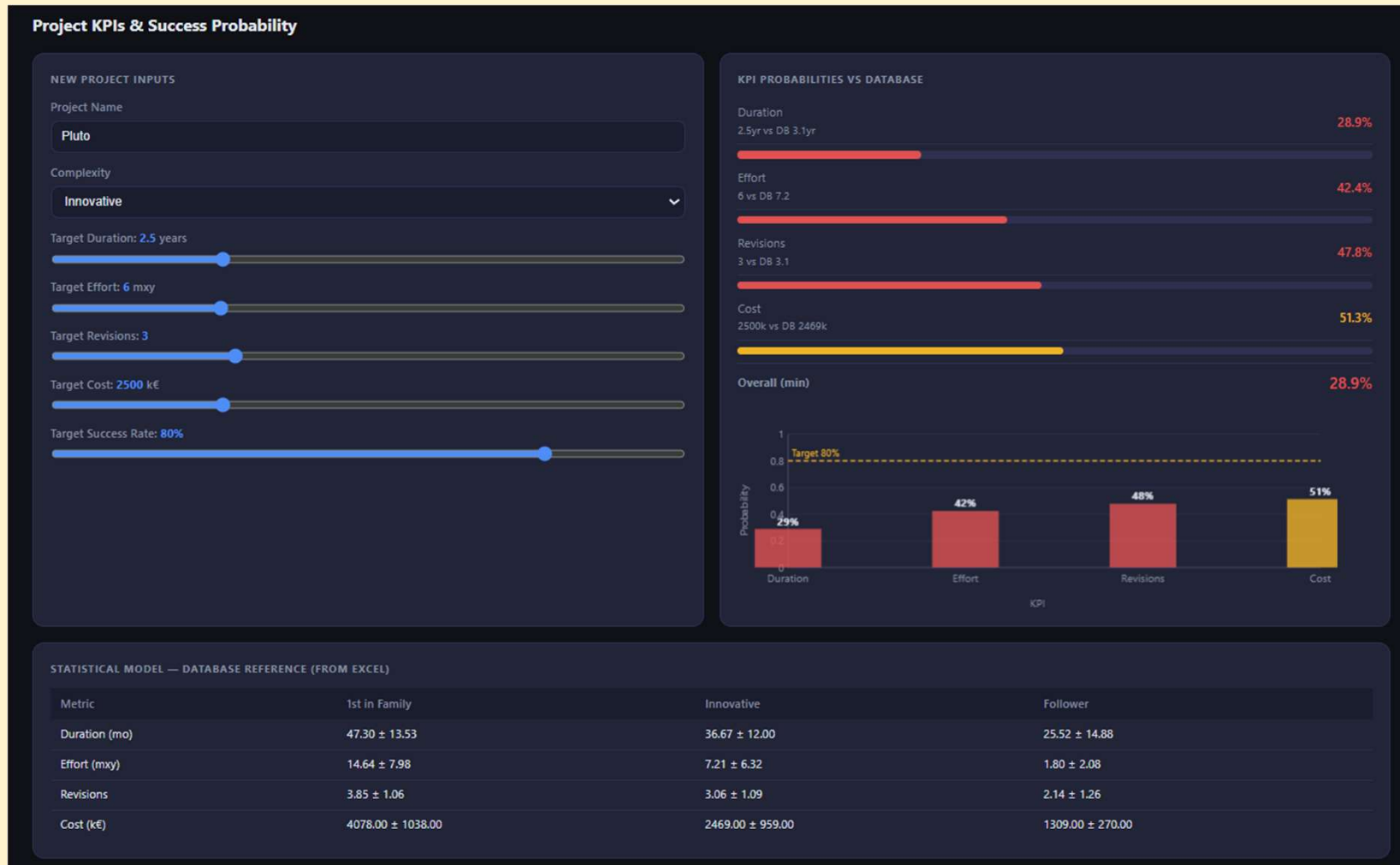
Data only indicative

Action 15: AI Web Page Recommendations



Data only indicative

Action 15: AI Web Page KPIs



Processus DA+AI

Detailed

1. Representation (historical database)
2. Selection (removal of outliers)
3. Conversion (based on criteria)
4. Distribution (statistical)
5. Modeling (initial statistical model)
6. Optimization (error minimization)
7. Exploration (testing other models)
8. Prediction (forecasting based on the best model)
9. DA vs. AI comparison (dedicated prompt)
10. Verification (prediction testing against distribution)

Data-Driven Decisions with Excel & AI

Summary

- Structure the data
- Formalize an explicit analysis process
- Derive a predictive model for data-driven decision
- Deploy the process using a structured AI prompt

Artificial Intelligence or Human Stupidity?



- AI: Naive, positive and tireless personal collaborator
- Communication: Precise question => Good answer
- Best Use case: Clean data, measurable gains, explicit steps

Grazie mille

Danke

Muchas gracias

obrigado

Dziekuje bardzo

谢谢

Thanks

СПАСИБО

Thank you

dhanyavad



Glossary

Term	Definition
Predictive planning	Use of historical data and models to estimate future project outcomes and support project selection.
Probability of success	Estimated likelihood that a project will meet its objectives under defined conditions.
Delay risk	The probability that a project will finish later than planned.
Budget deviation	Difference between planned budget and actual or expected cost outcome.
Revision intensity	Frequency or magnitude of changes made to a project during execution.
Complexity	Project class used in the case to segment the history database into 1st in family, innovative, and follower projects.
1st in family	A project type with the highest novelty and typically the greatest uncertainty.
Innovative	A project type with moderate novelty and moderate uncertainty.
Follower	A project type that reuses known approaches and usually has lower novelty.
Outlier	An observation that is unusually high or low and may distort statistical modeling.
IQR	Interquartile range; used here as a robust method for defining outlier limits.
Gaussian model	Normal distribution model used as a transparent baseline for forecasting duration.
Log-Normal model	Distribution model that fits right-skewed duration data better than a normal distribution in this case.
Poisson model	Distribution sometimes used for count or event data; one of the candidate alternative models.
Exponential model	Candidate statistical model often used for waiting times or decay processes.
Weibull model	Flexible distribution often used for reliability and duration modeling.
Gamma model	Flexible right-skewed distribution often used for time and cost data.
Pareto / Power Law	Heavy-tailed distribution sometimes used when extreme values dominate the shape of the data.
Histogram fit error	Measure of how far a proposed model is from the observed history distribution.
Margin of error	Statistical range around an estimate showing the likely precision of the average.
Scenario analysis	Comparison of multiple possible project outcomes under different assumptions.
Risk-adjusted attractiveness	Project desirability after accounting for uncertainty, delay, cost, and other risks.