

Data Literacy Fundamentals for Learning

Zsolt Olah



Images used in the deck were created by MidJourney.

Learning circa 1023

Zsolt's data classification profile



Quantitative
(numerical)

Continuous

Ratio

Height: 5' 10" or 177.80 cm

Interval

Favorite outdoor temperature: 72 F or 22.222 C

Discrete

Number of publications: 2

Qualitative
(categorical)

Ordinal

Level: 6

Experience level: High

Nominal

Name: Zsolt Olah

Title: Sr. Learning Technologist at Amazon

What is data literacy?

“Gartner defines data literacy as the ability to read, write, and communicate data in context[...]

salesid	manage	comple	training	training	performance	deals_c	deals_k	perform	manage	region	performance_t	Duration	score	feedback	feedback	manage	pre_score	learn_gain	perform	c
189	184	1	34	43.23096	19.93168037	0	0	30	-1	S	-10.06831963	9.230955	88	5			69	61	-16	
196	195	1	38	45.91352	19.93819311	0	0	32.41512	-1	N	-12.47692613	7.91352	89	5			40	82	-19	
242	240	1	34	48.65533	20.51441205	0	0	30	-1	N	-9.485587946	14.65533	82	4			70	40	-14	
83	76	1	32	44.68476	20.64086913	0	0	30	-1	E	-9.359130871	12.68476	85	4	Integer aliquet, massa		71	48	-14	
260	257	1	36	48.83551	21.5450002	0	0	30	-1	N	-8.454999802	12.83551	84	5			76	33	-13	
274	272	1	36	45.94415	21.57091358	0	0	30	-1	N	-8.42908642	9.944149	85	3			70	50	-13	
258	257	1	36	41.58324	21.59225548	0	0	30	-1	N	-8.40774452	5.583238	82	4	Vivamus metus arcu,		52	63	-13	
284	283	1	37	48.73652	21.63515921	0	0	30.16318	-1	N	-8.528019618	11.73652	93	5			73	74	-13	
263	262	1	38	50.44731	21.6355028	0	0	33.61544	-1	N	-11.97993735	12.44731	87	5			61	67	-18	
110	107	1	38	48.20194	21.90080643	0	0	32.80963	-1	E	-10.90882785	10.20194	88	5			64	67	-16	
173	168	1	33	45.43844	21.96457034	0	0	30	-1	S	-8.035429659	12.43844	89				60	73	-13	
218	215	1	39	45.89072	22.06048939	0	0	34.54484	-1	N	-12.48435174	6.890721	85	3			88	-25	-18	
254	251	1	35	43.73429	22.18425672	0	0	30	-1	N	-7.815743275	8.734292	84	3	Duis consequat dui ne		60	60	-11	
102	99	1	32	43.61393	22.244257	0	0	30	-1	E	-7.755743005	11.61393	83	5			52	65	-11	
160	157	1	33	44.86677	23.24093659	0	0	30	-1	S	-6.759063406	11.86677	86	5			68	56	-10	
214	213	1	34	42.6395	23.96313376	0	0	30	-1	N	-6.036866238	8.639501	82	5			72	36	-10	
136	127	1	32	37.3062	24.26743172	0	0	30	-1	E	-5.732568277	5.306198	83				59	59	-9	
149	147	1	33	44.98557	24.28283627	0	0	30	-1	E	-5.717163731	11.98557	87	5			53	72	-9	
129	127	1	40	49.98602	24.46754531	0	0	35.60112	-1	E	-11.13357127	9.986022	82	4	Maecenas ut massa q		43	68	-17	
194	191	1	40	47.83634	25.00248716	0	0	35.99107	-1	S	-10.98858191	7.836341	87	5			55	71	-15	
134	127	1	41	47.50272	25.35535006	0	0	37.38935	-1	E	-12.03400379	6.50272	84				66	53	-19	
183	182	1	34	43.71404	25.98801905	0	0	30	-1	S	-4.011980949	9.714041	83				66	50	-7	
63	62	1	31	45.46906	26.0049267	0	0	30	-1	E	-3.995073301	14.46906	88	5			63	68	-6	
244	240	1	35	42.1212	26.68480323	0	0	30	-1	N	-3.315196775	7.1212	87	5			70	57	-6	
278	277	1	37	43.00757	26.71094877	0	0	31.3428	-1	N	-4.631851098	6.007566	81	4	Etiam pretium iaculis		68	41	-7	
282	277	1	36	48.42217	26.98834409	0	0	30	-1	N	-3.011655906	12.42217	85	5			59	63	-6	
108	107	1	32	41.7772	27.29571179	0	0	30	-1	F	-2.704288214	9.777197	84	3	Vivamus metus arcu,		61	59	-4	

Scope for today?

**Data
Literacy**

The diagram features a central white circle labeled 'Data Literacy' on the left. To its right, three purple circles containing the numbers 1, 2, and 3 are arranged vertically. Each number is connected by a line to a black rounded rectangle on the right. The background is a light gray world map. The three components are: 1. Data Classification (bottom), 2. Thinking: 4C's (middle), and 3. Tools, stats skills, and processes (top).

1

Data Classification

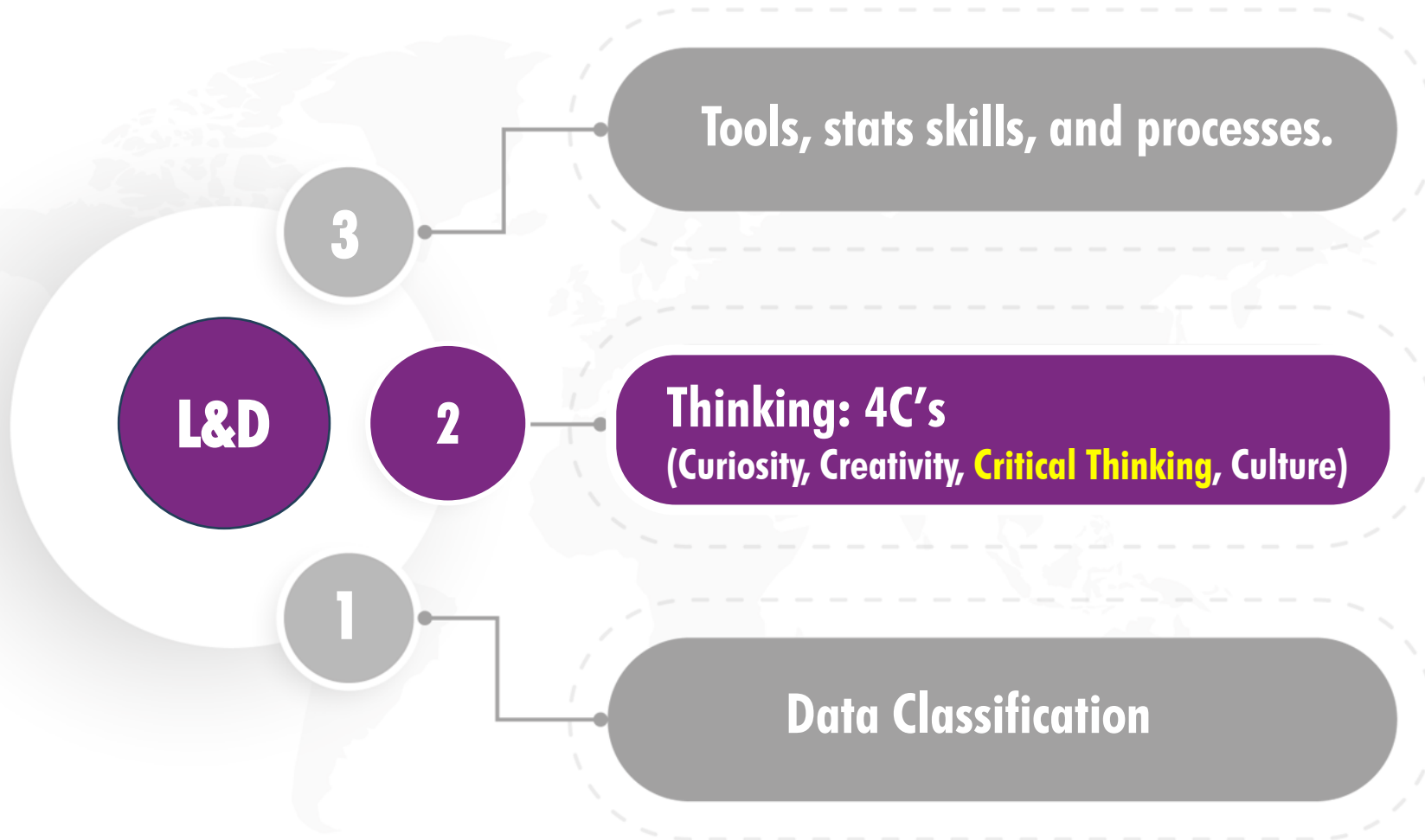
2

Thinking: 4C's
(Curiosity, Creativity, Critical Thinking, Culture)

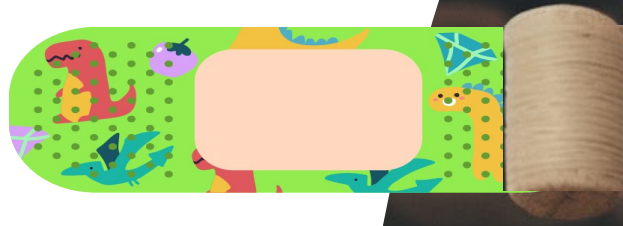
3

Tools, stats skills, and processes.

Scope for today?



M e d i c a l A d v i c e :



“You need to calm
down!”

Dr. Dash Swift





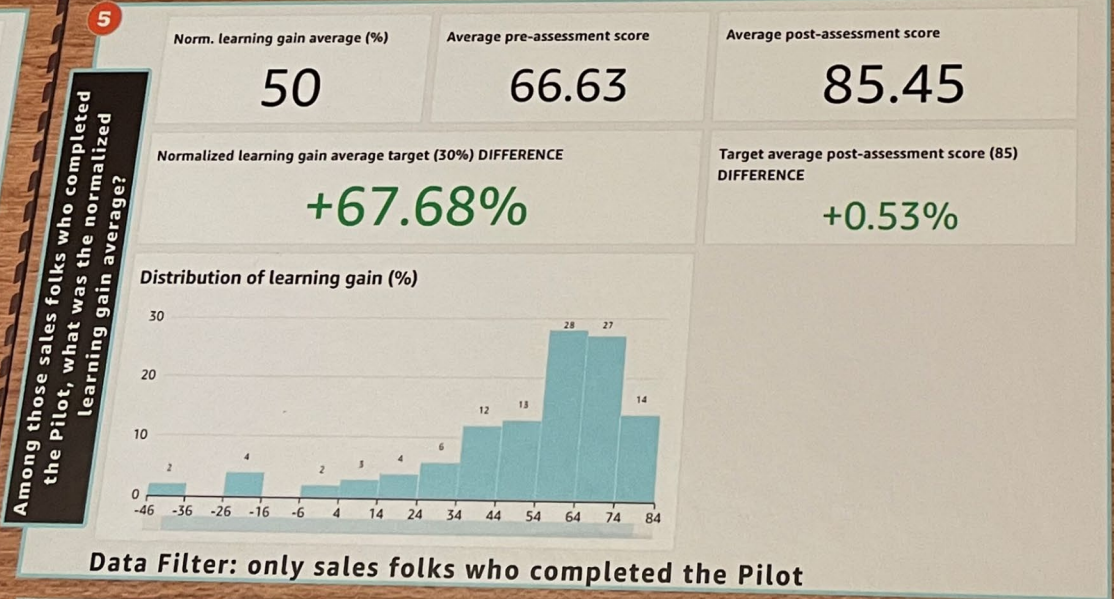
Data, the language of impact

- Don't have to be perfect!
- Data literacy is the ability to use the language of impact **well enough** that you're not just right but able **to make a difference**.
- People can speak and still say nothing...

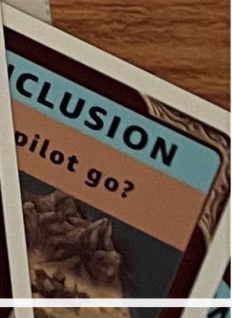
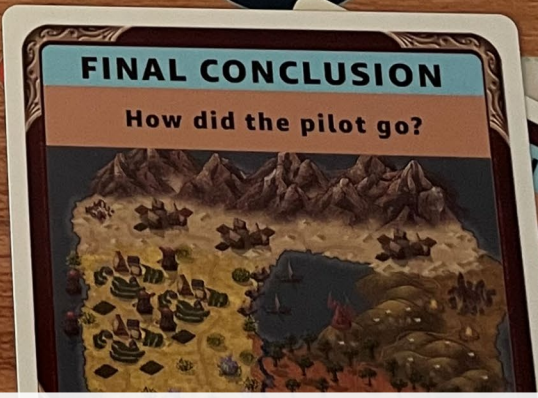


Critical
Thinking:
Calm Down

Average post-assessment score
85.45
Target average post-assessment score (85)
DIFFERENCE
+0.53%



Was the Pilot successful? What's your final conclusion?



What's On The Table (OTT)?

Context: Sales Pilot in a Fictitious Land

“Improve the skills of low-tenure sales folks to be able to address objections and close the sale.”

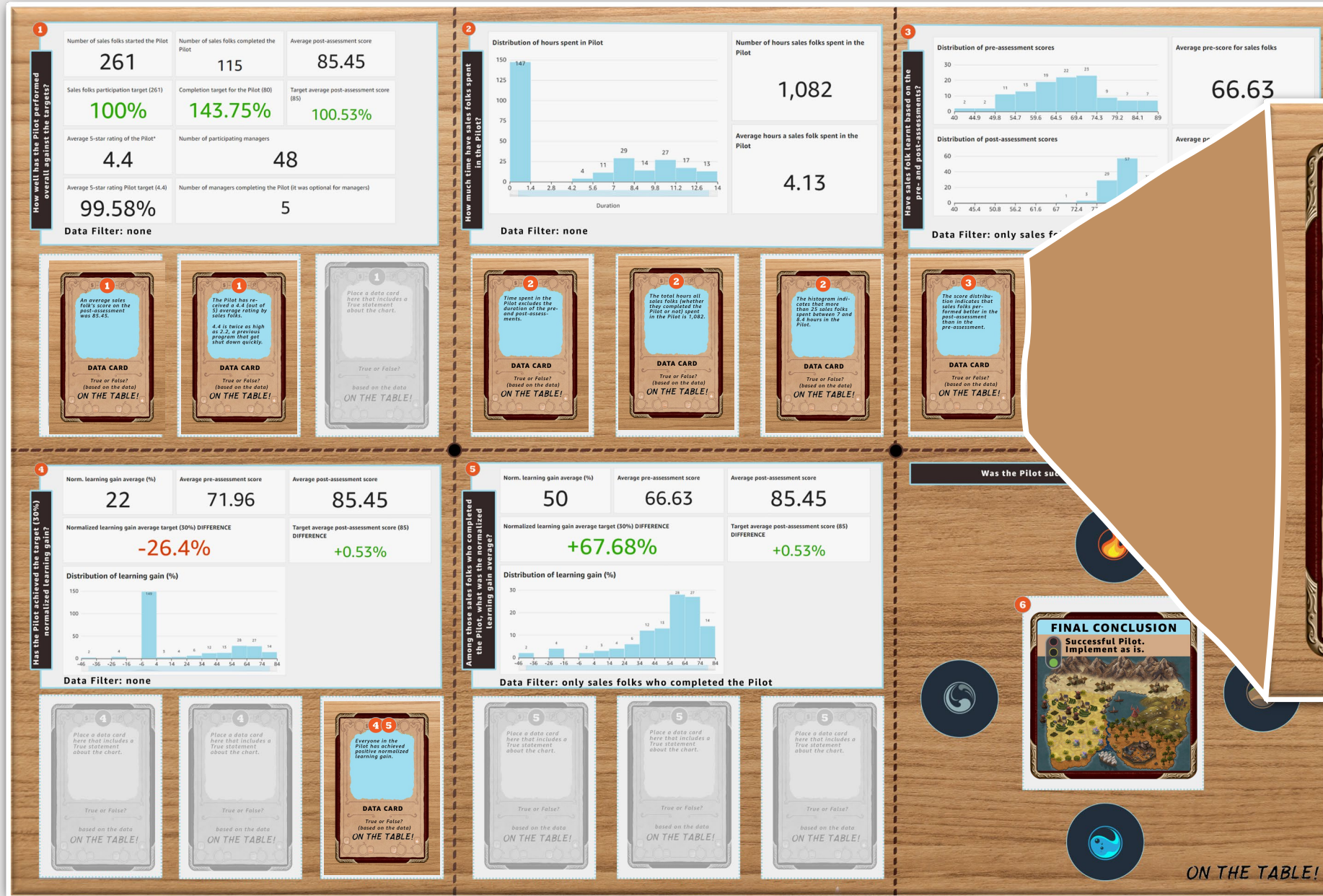


Pilot results

“Was the pilot successful? Partially successful? Or not successful?”

salesid	manage	comple	training	training	performance	deals_c	deals_k	perform	manage	region	performance_t	Duratio	score	feedbac	feedbac	manage	pre_score	learn_gain	perform	c
189	184	1	34	43.23096	19.93168037	0	0	30	-1	S	-10.06831963	9.230955	88	5			69	61	-16	
196	195	1	38	45.91352	19.93819311	0	0	32.41512	-1	N	-12.47692613	7.91352	89	5			40	82	-19	
242	240	1	34	48.65533	20.51441205	0	0	30	-1	N	-9.485587946	14.65533	82	4			70	40	-14	
83	76	1	32	44.68476	20.64086913	0	0	30	-1	E	-9.359130871	12.68476	85	4	Integer aliquet, massa		71	48	-14	
260	257	1	36	48.83551	21.5450002	0	0	30	-1	N	-8.454999802	12.83551	84	5			76	33	-13	
274	272	1	36	45.94415	21.57091358	0	0	30	-1	N	-8.42908642	9.944149	85	3			70	50	-13	
258	257	1	36	41.58324	21.59225548	0	0	30	-1	N	-8.40774452	5.583238	82	4	Vivamus metus arcu,		52	63	-13	
284	283	1	37	48.73652	21.63515921	0	0	30.16318	-1	N	-8.528019618	11.73652	93	5			73	74	-13	
263	262	1	38	50.44731	21.6355028	0	0	33.61544	-1	N	-11.97993735	12.44731	87	5			61	67	-18	
110	107	1	38	48.20194	21.90080643	0	0	32.80963	-1	E	-10.90882785	10.20194	88	5			64	67	-16	
173	168	1	33	45.43844	21.96457034	0	0	30	-1	S	-8.035429659	12.43844	89				60	73	-13	
218	215	1	39	45.89072	22.06048939	0	0	34.54484	-1	N	-12.48435174	6.890721	85	3			88	-25	-18	
254	251	1	35	43.73429	22.18425672	0	0	30	-1	N	-7.815743275	8.734292	84	3	Duis consequat dui ne		60	60	-11	
102	99	1	32	43.61393	22.244257	0	0	30	-1	E	-7.755743005	11.61393	83	5			52	65	-11	
160	157	1	33	44.86677	23.24093659	0	0	30	-1	S	-6.759063406	11.86677	86	5			68	56	-10	
214	213	1	34	42.6395	23.96313376	0	0	30	-1	N	-6.036866238	8.639501	82	5			72	36	-10	
136	127	1	32	37.3062	24.26743172	0	0	30	-1	E	-5.732568277	5.306198	83				59	59	-9	
149	147	1	33	44.98557	24.28283627	0	0	30	-1	E	-5.717163731	11.98557	87	5			53	72	-9	
129	127	1	40	49.98602	24.46754531	0	0	35.60112	-1	E	-11.13357127	9.986022	82	4	Maecenas ut massa q		43	68	-17	
194	191	1	40	47.83634	25.00248716	0	0	35.99107	-1	S	-10.98858191	7.836341	87	5			55	71	-15	
134	127	1	41	47.50272	25.35535006	0	0	37.38935	-1	E	-12.03400379	6.50272	84				66	53	-19	
183	182	1	34	43.71404	25.98801905	0	0	30	-1	S	-4.011980949	9.714041	83				66	50	-7	
63	62	1	31	45.46906	26.0049267	0	0	30	-1	E	-3.995073301	14.46906	88	5			63	68	-6	
244	240	1	35	42.1212	26.68480323	0	0	30	-1	N	-3.315196775	7.1212	87	5			70	57	-6	
278	277	1	37	43.00757	26.71094877	0	0	31.3428	-1	N	-4.631851098	6.007566	81	4	Etiam pretium iaculis		68	41	-7	
282	277	1	36	48.42217	26.98834409	0	0	30	-1	N	-3.011655906	12.42217	85	5			59	63	-6	
108	107	1	32	41.7772	27.29571179	0	0	30	-1	F	-2.704288214	9.777197	84	3	Vivamus metus arcu,		61	59	-4	

Learning Dashboard (completed)



The score distribution indicates that sales folks performed better in the post-assessment than in the pre-assessment.

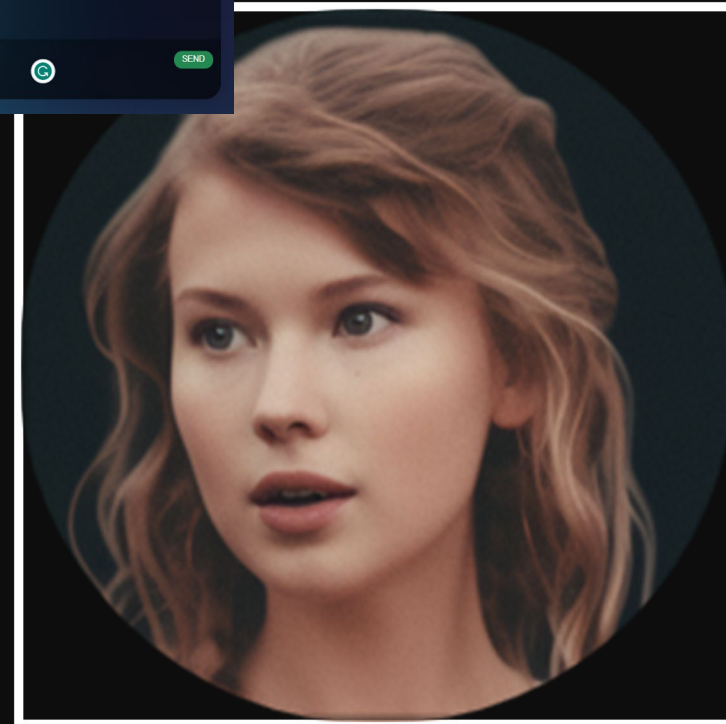
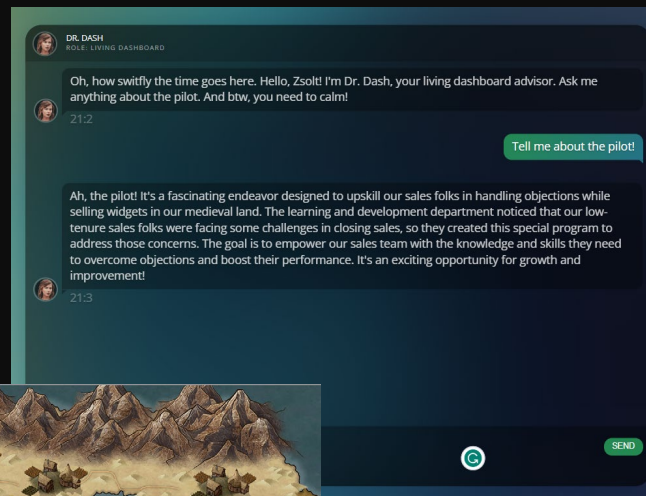
DATA CARD

True or False?
(based on the data)
ON THE TABLE!

Need some help from Dr. Dash?



Page 7



Data Literacy

“Just one more
thing...”

Columbo



Region cards

INSTRUCTIONS: DO NOT show them to anyone!
You can share insights but not the card!

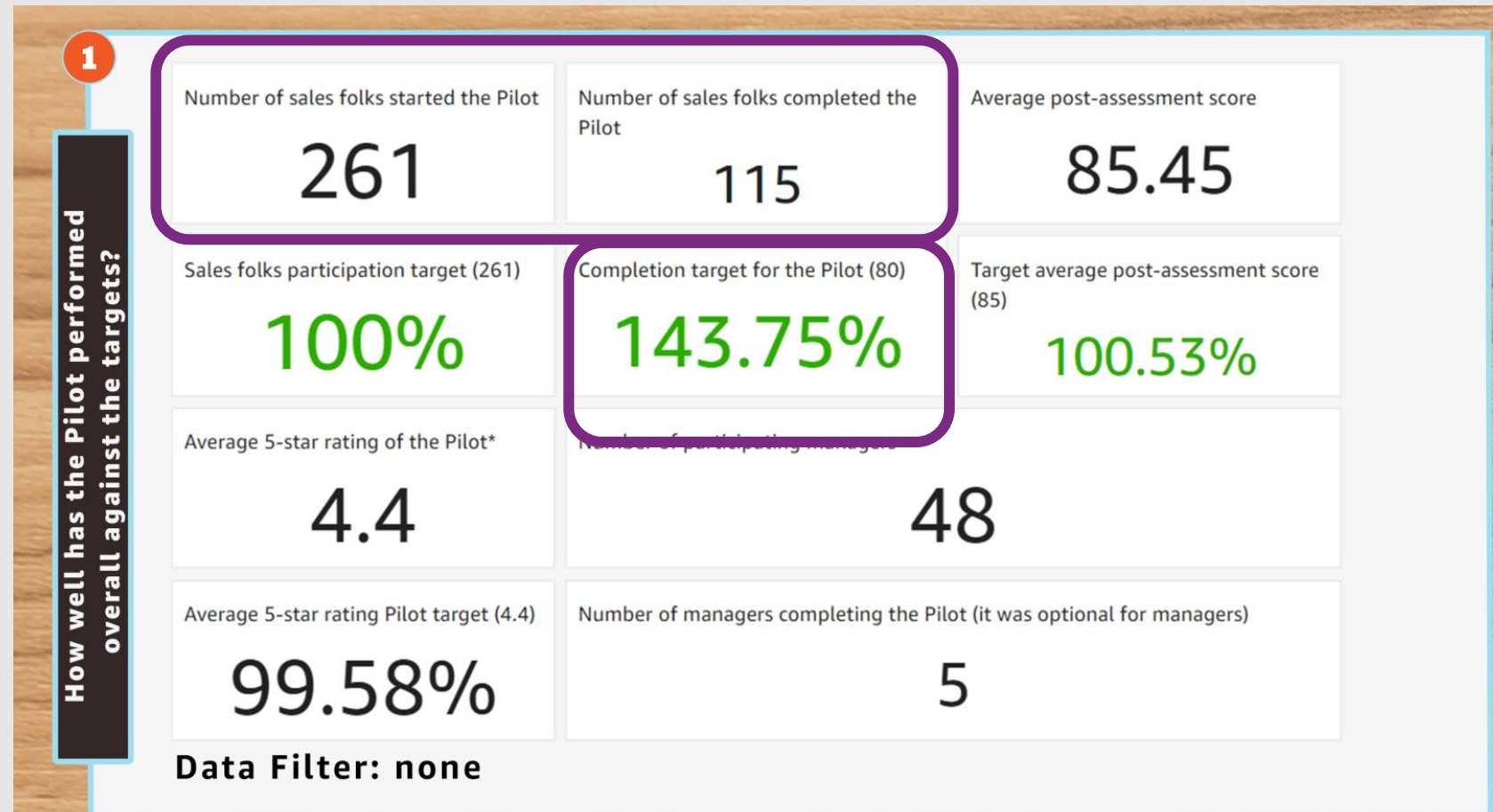
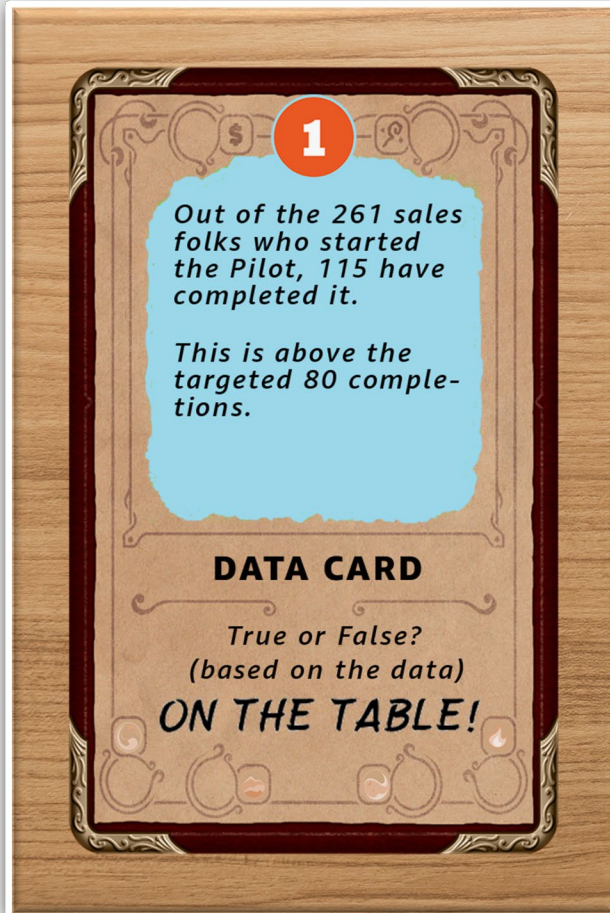


"Perp Card" card

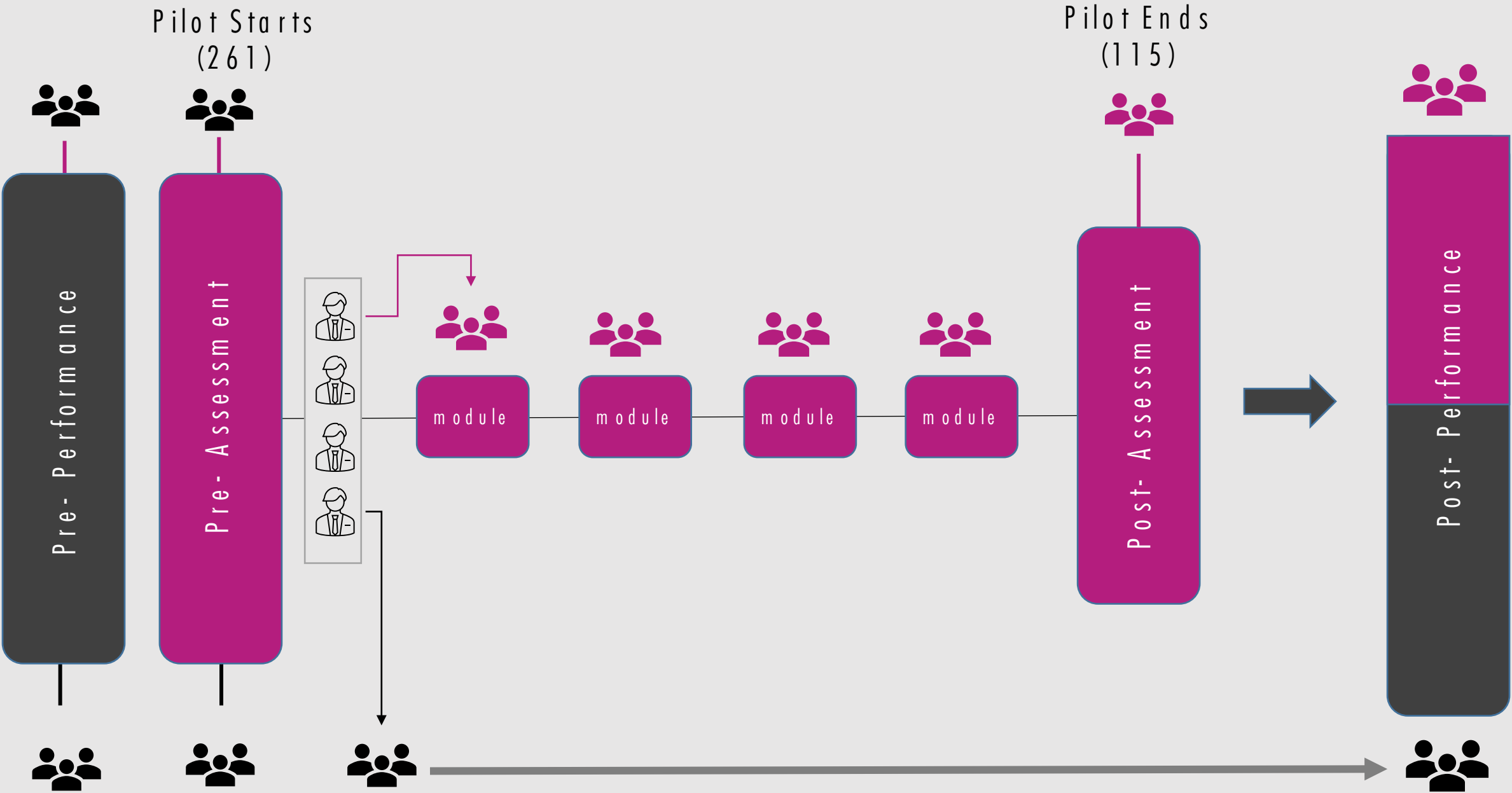
INSTRUCTIONS: DO NOT reveal your agenda!



Practice Round: Learning Dashboard – Chart 1



Pilot Execution



Successful or not?

Pilot success criteria:

1. 100% (261) sales folks (under 2 years of tenure) to take the pre-assessment
2. 30% (80) sales folks to complete the pilot
3. 85 (or above) average post-assessment score
4. 4.4 satisfaction (1 - very unsatisfied, 2 - unsatisfied, 3 - neutral, 4 - satisfied, 5 - very satisfied)
5. 30% average learning gain (calculated as normalized learning change with matched data*: Marx and Cummings 2007).

$$c = \begin{cases} \frac{\text{post-pre}}{100-\text{pre}} & \text{post} > \text{pre} \\ \text{drop} & \text{post} = \text{pre} = 100 \text{ or } 0 \\ 0 & \text{post} = \text{pre} \\ \frac{\text{post-pre}}{\text{pre}} & \text{post} < \text{pre} \end{cases}$$

*matched data calculation requires both pre- and post-assessment



M Y T H # 1

“Data, like hips,
don’t lie!”

Shakira





What do
you
remember?

Learning Dashboard – Chart 1 (true)

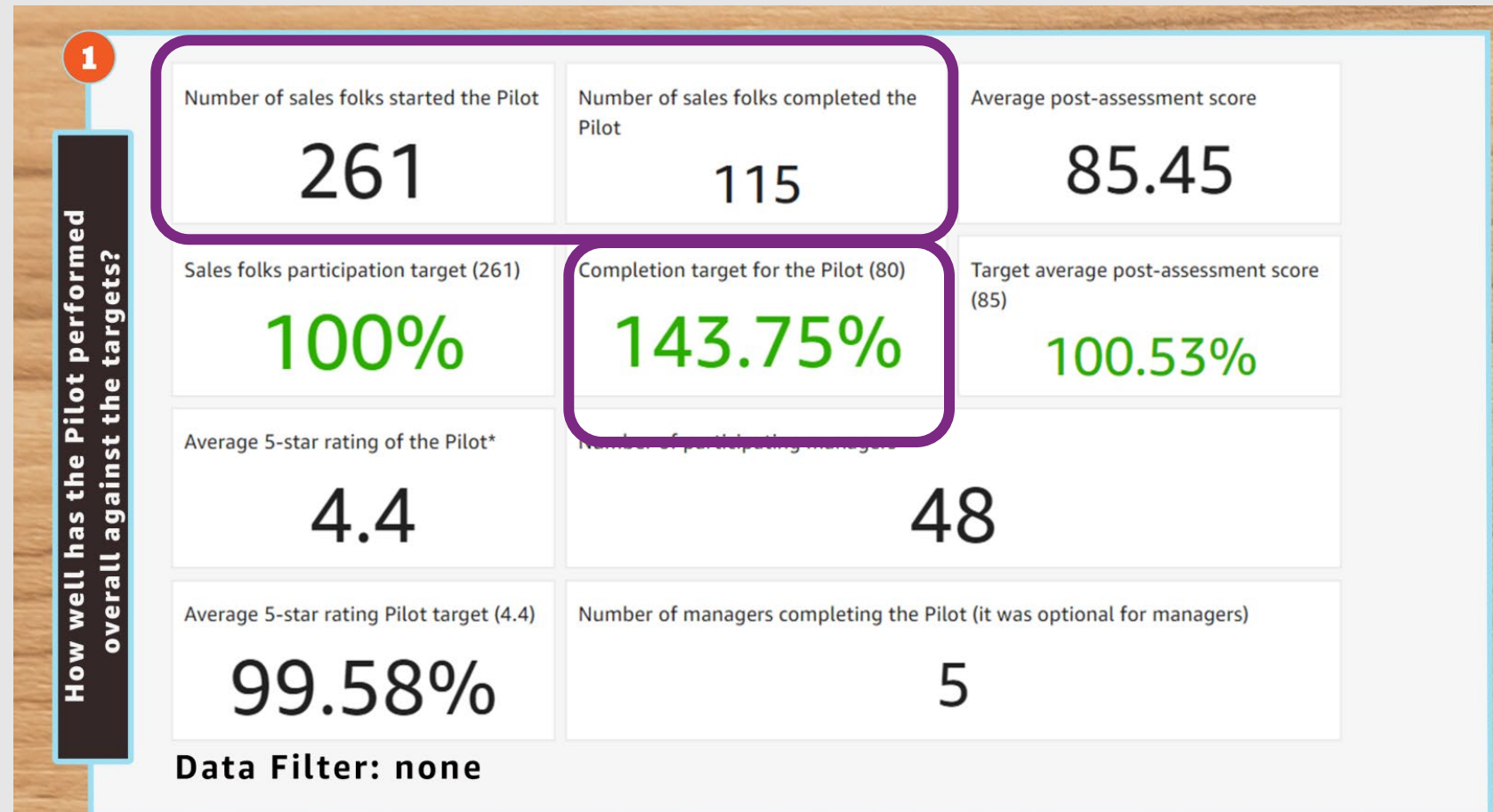
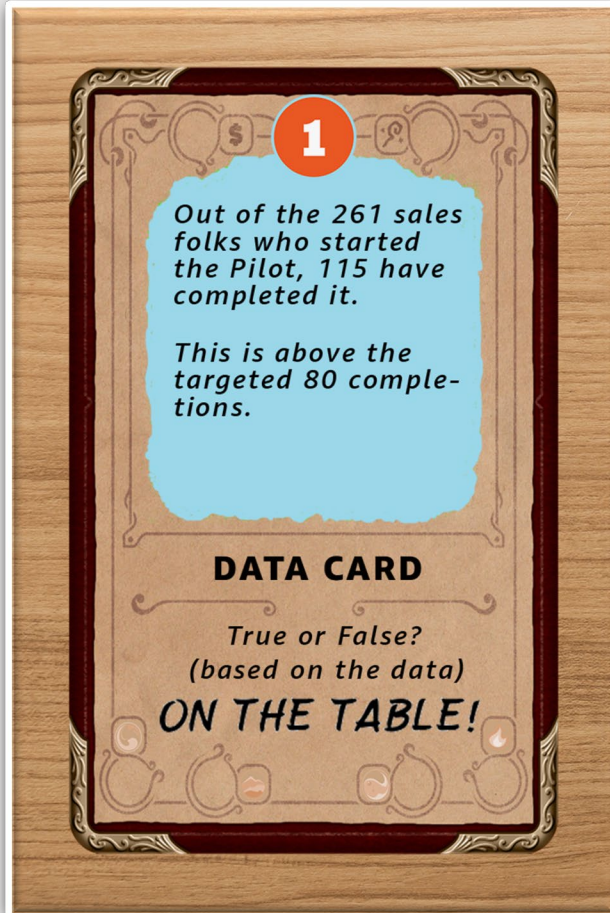
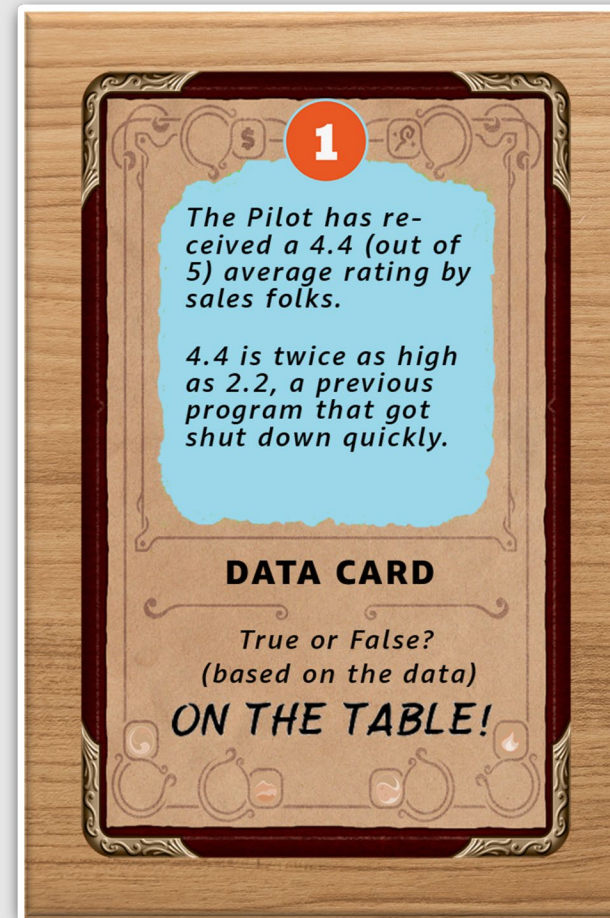
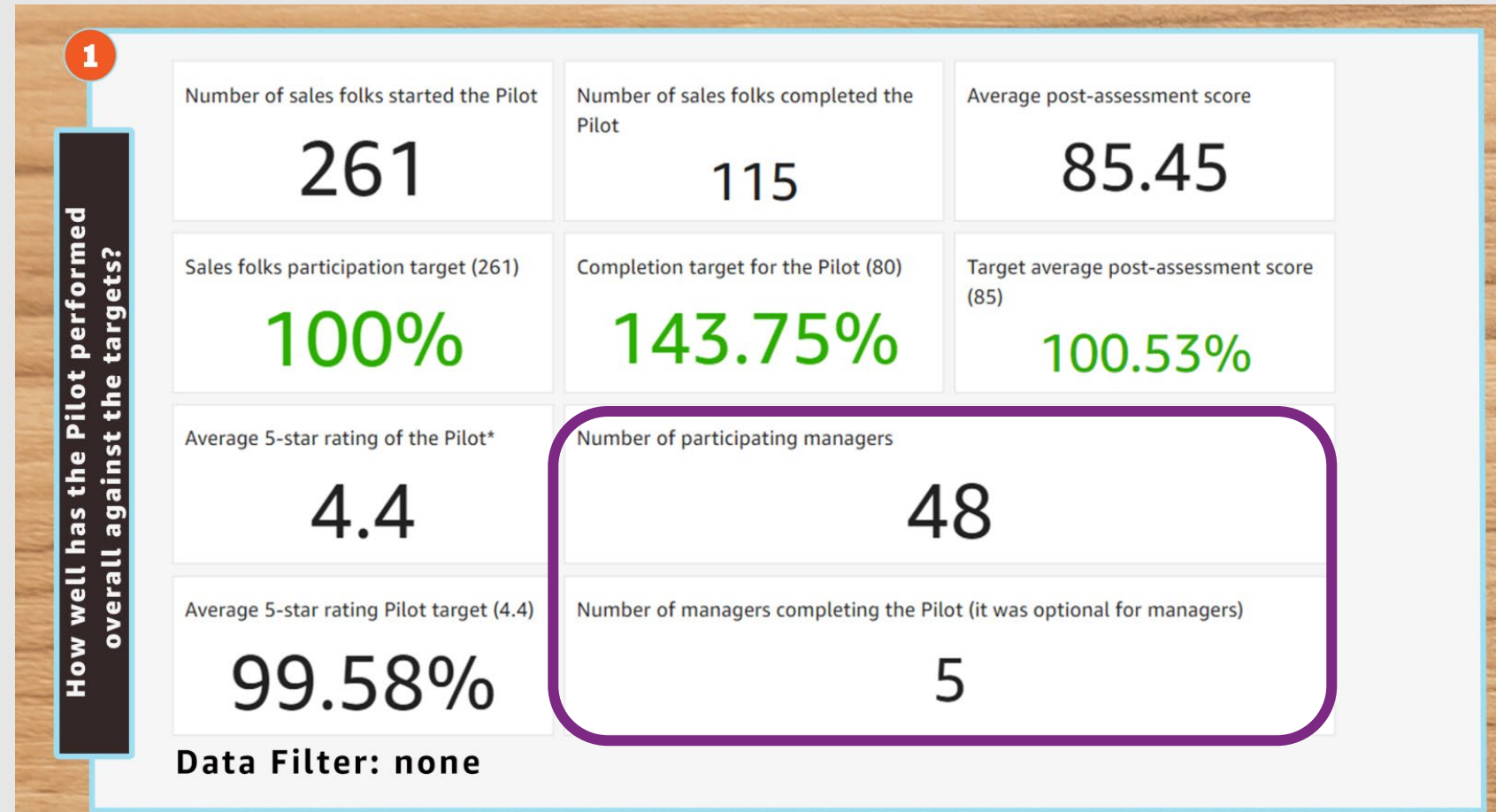


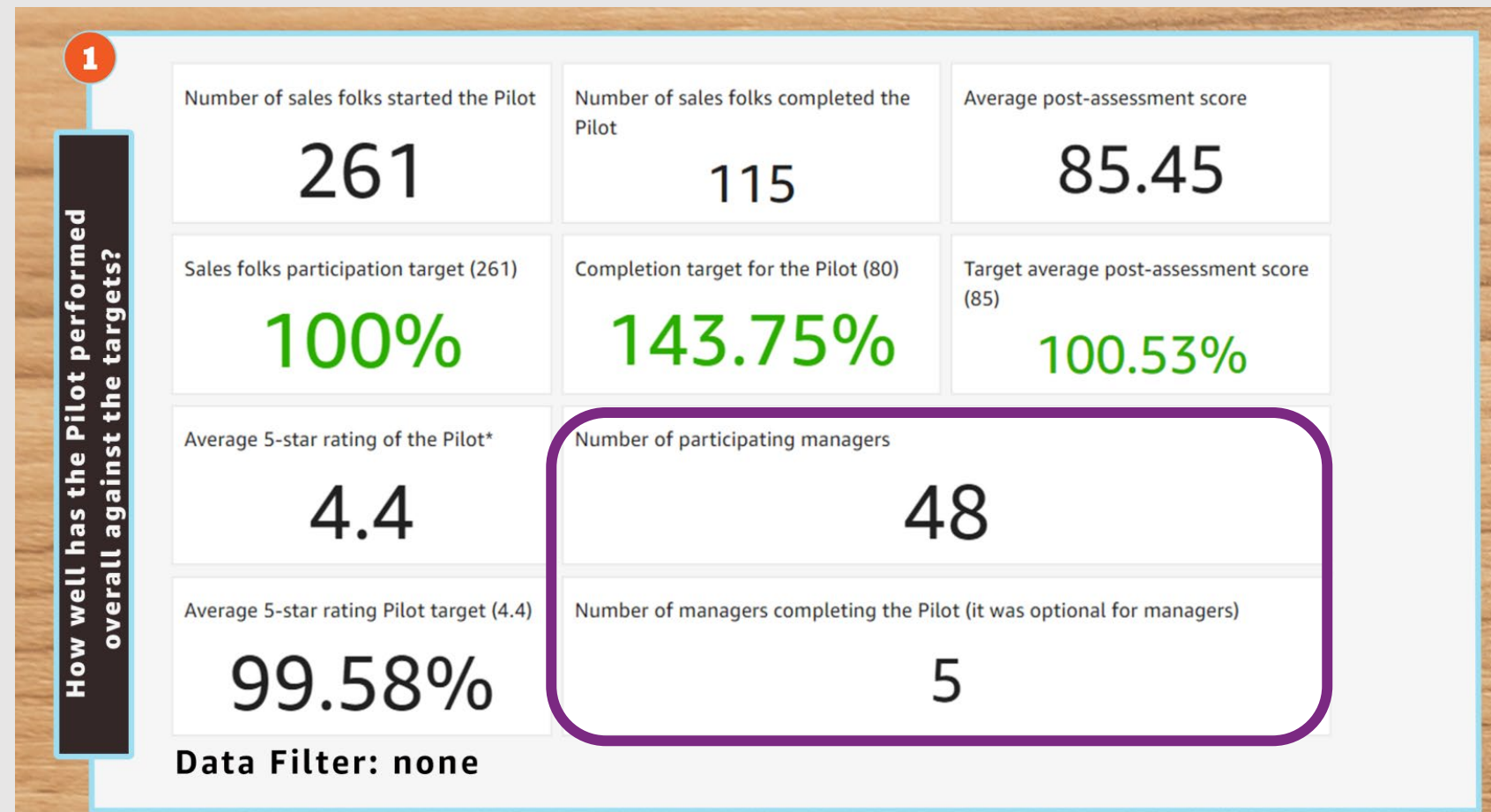
Chart 1 Data Cards (10 min)



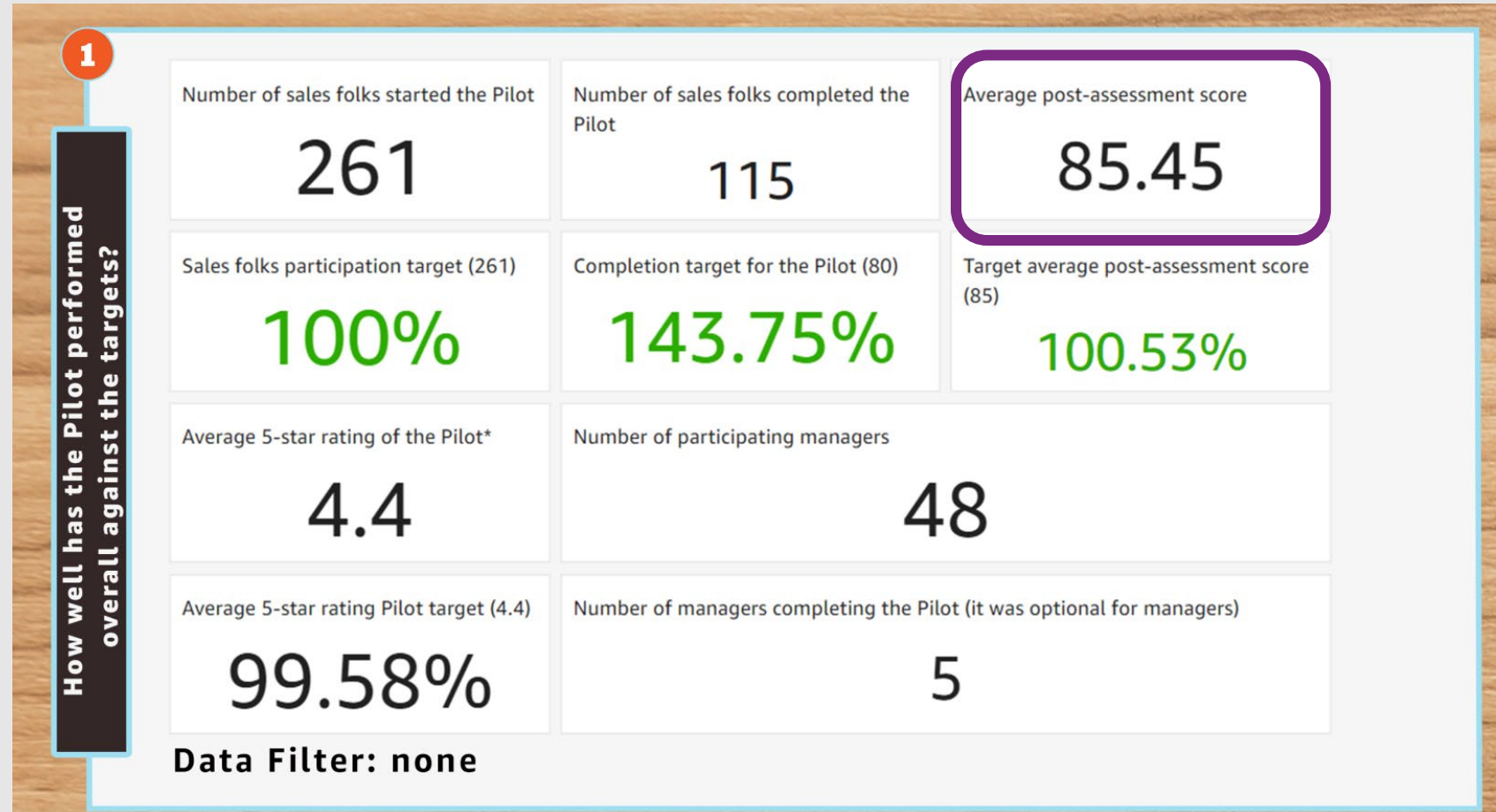
Learning Dashboard – Chart 1



Learning Dashboard – Chart 1 (true but...)



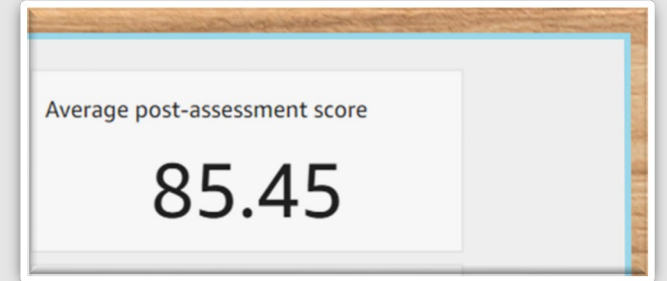
Learning Dashboard – Chart 1



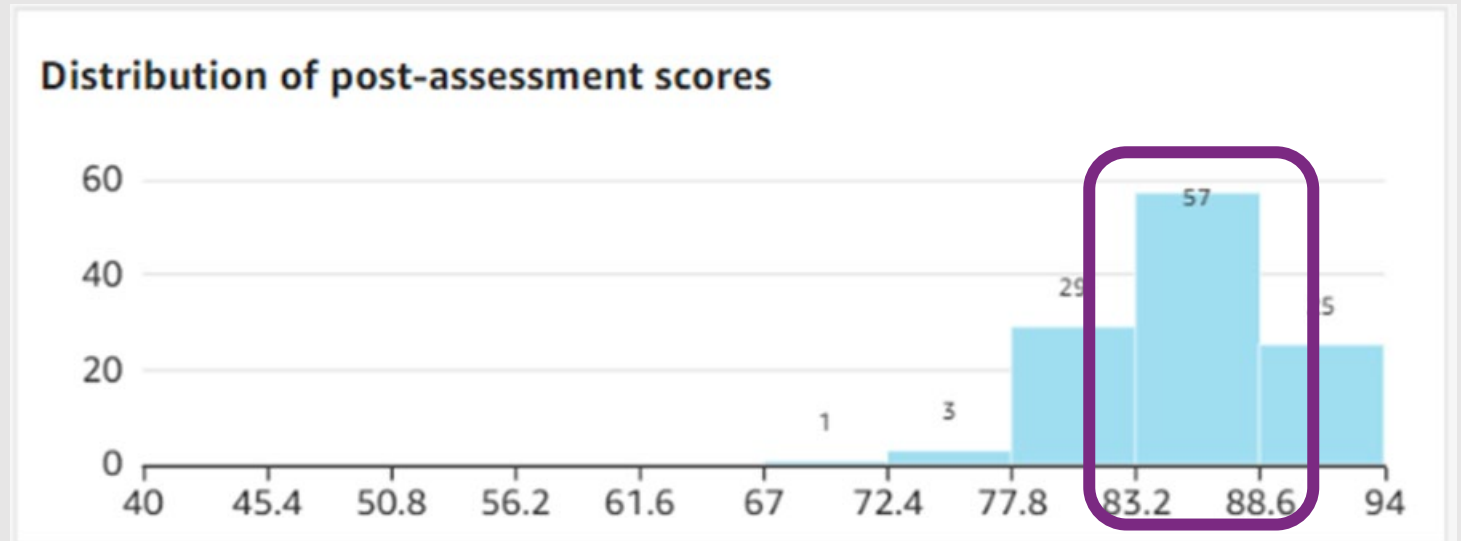
Learning Dashboard – Chart 1 (exactly 85.45 is unknown)



Based on Chart 1: it is unknown



Based on Chart 3: it is unknown



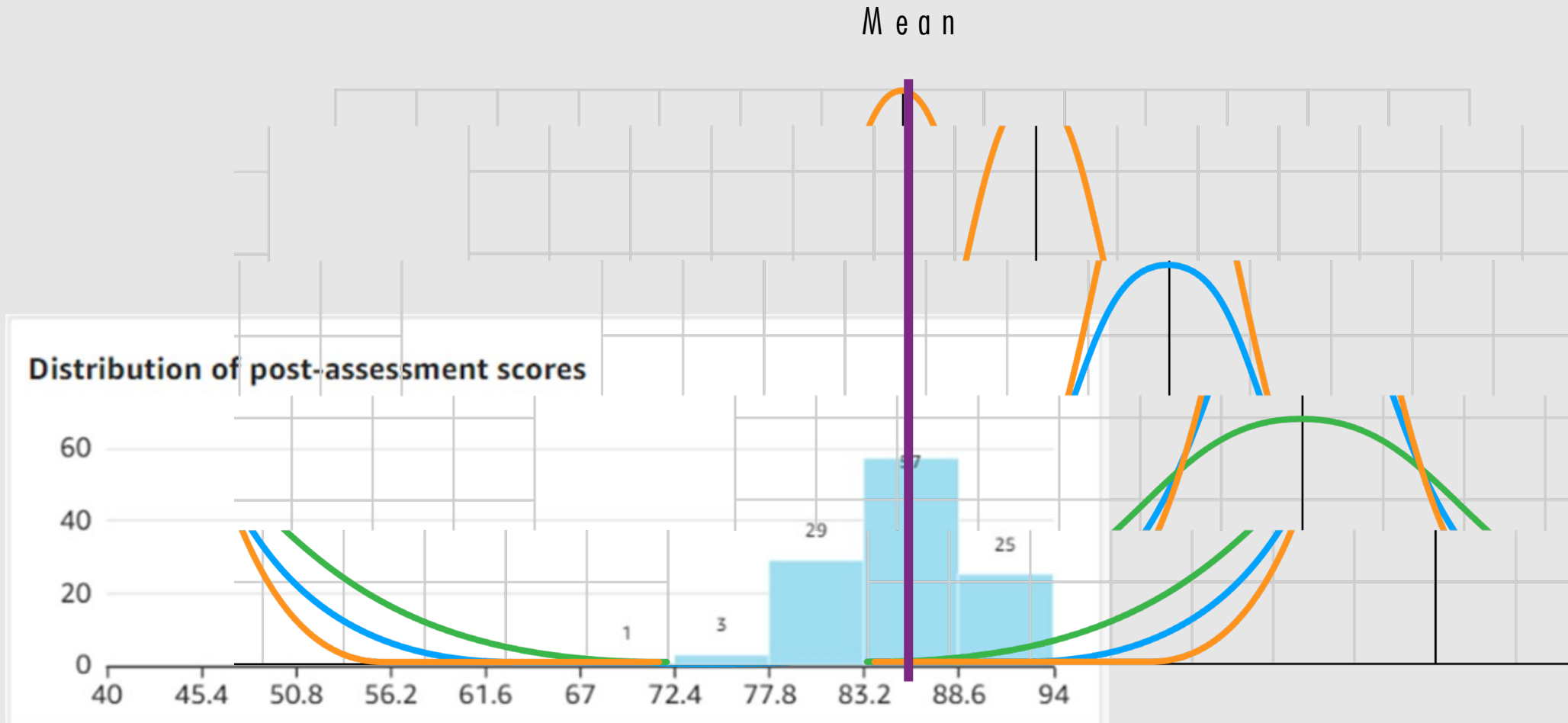
M Y T H # 2

“Mean is always meaningful because all that we see or seem is but a mean within a mean...”

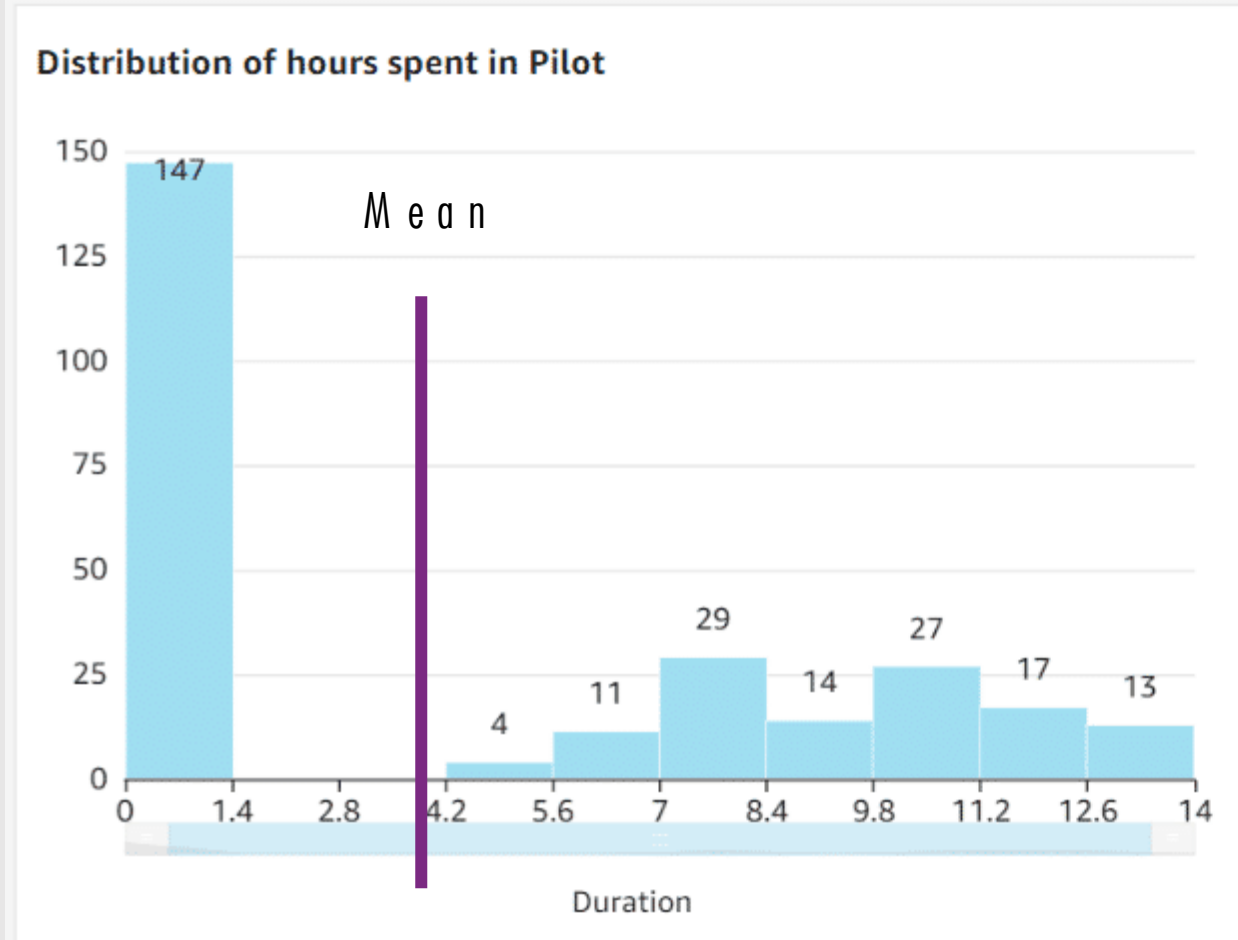
Poe Nevers'more



Mean in normal distributions:



Mean can be deceiving (aka meaningless)



Number of hours sales folks spent in the Pilot

1,082

Average hours a sales folk spent in the Pilot

➡ 4.13

“Normal
Distribution: Shape
of U (upside-
down)”

Ed Sheeran

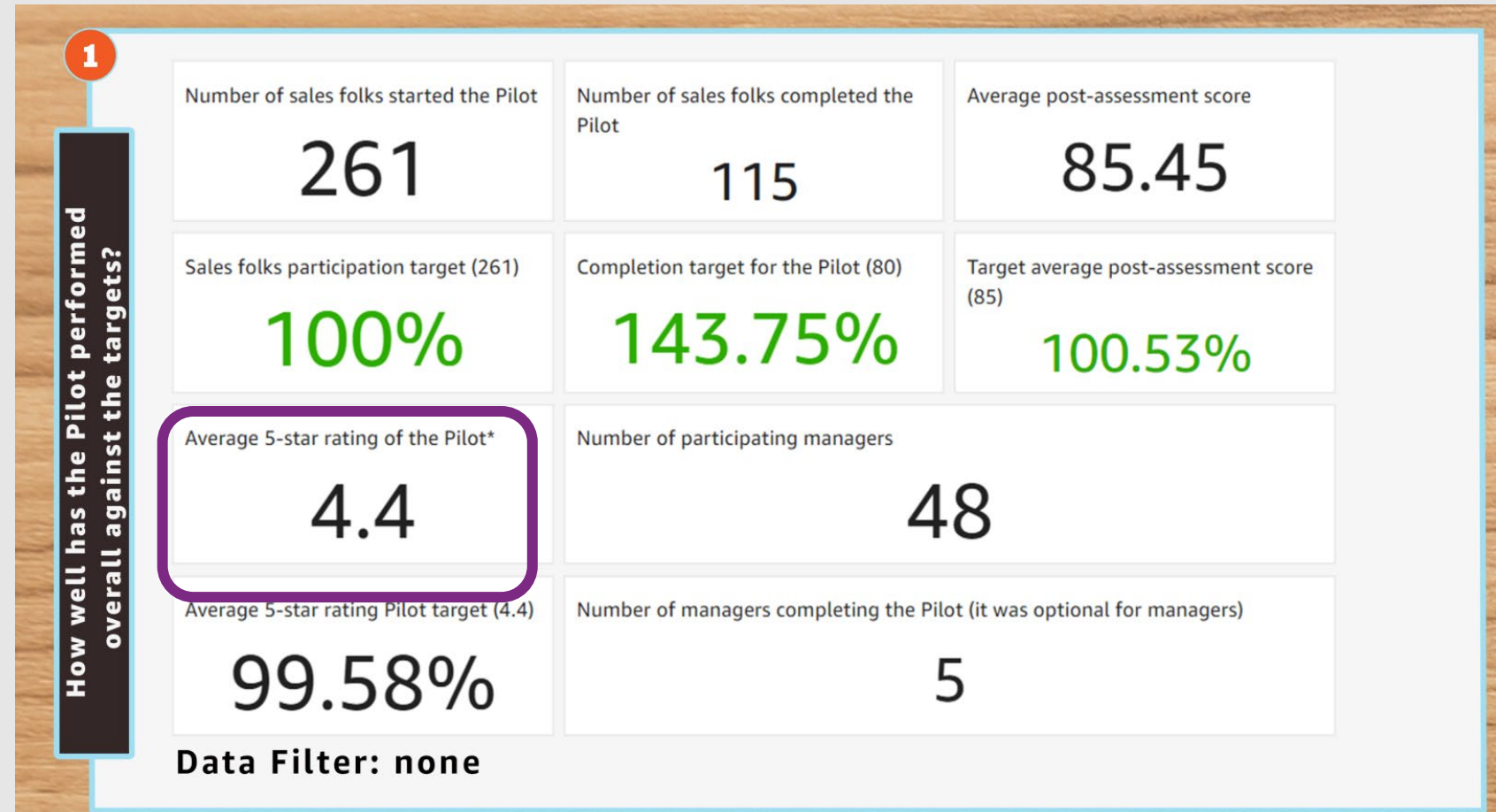
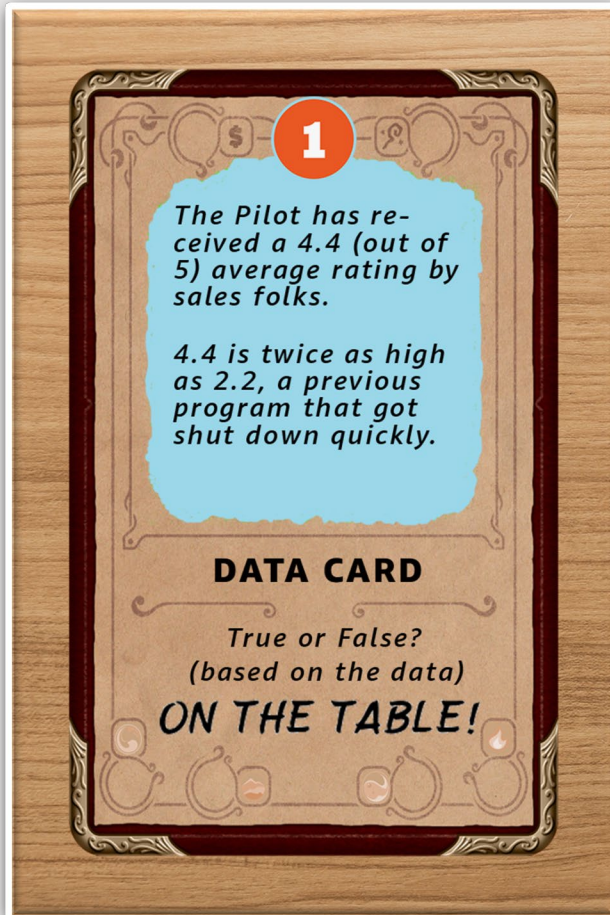
Early Draft from the 1200's





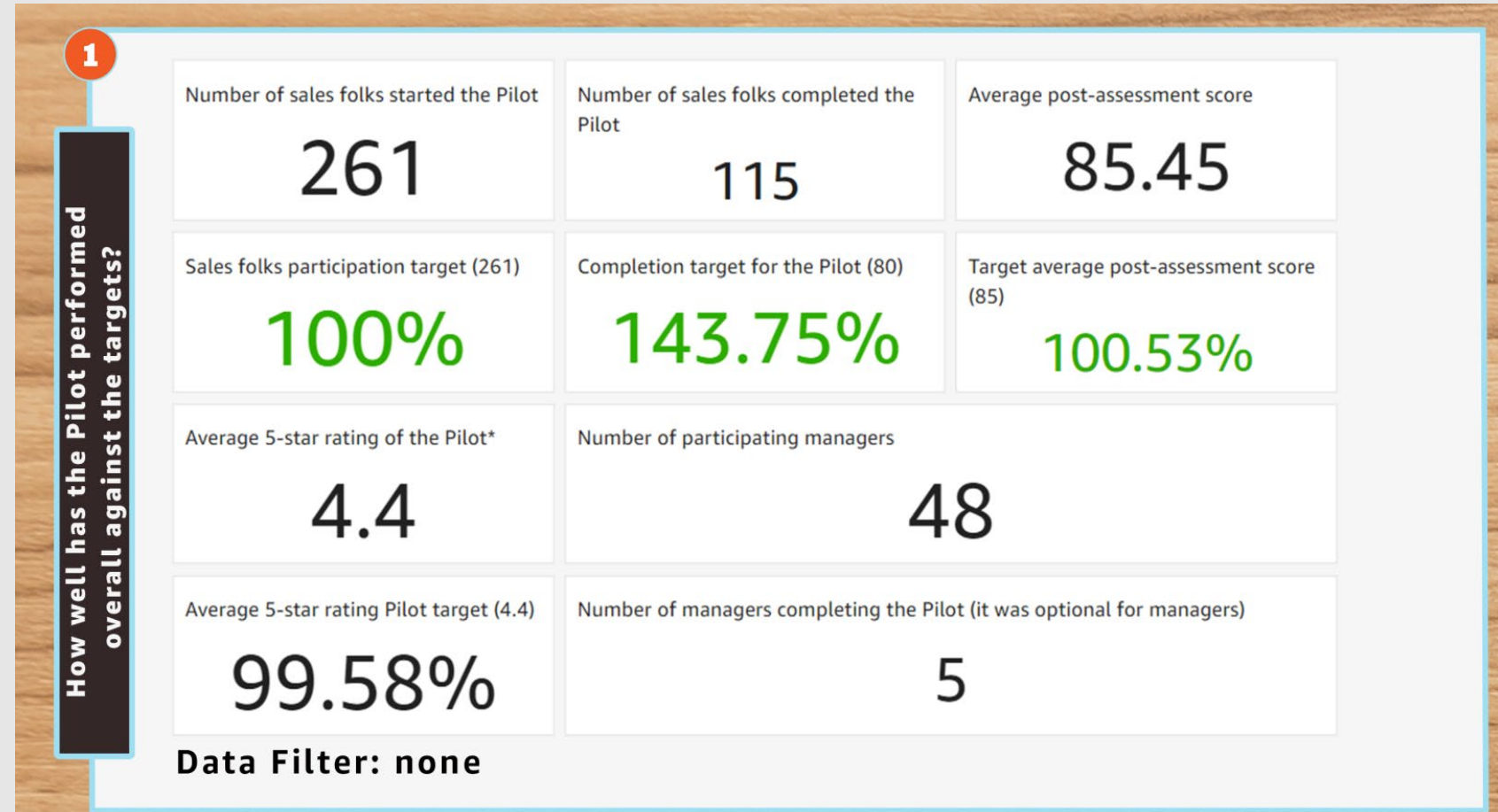
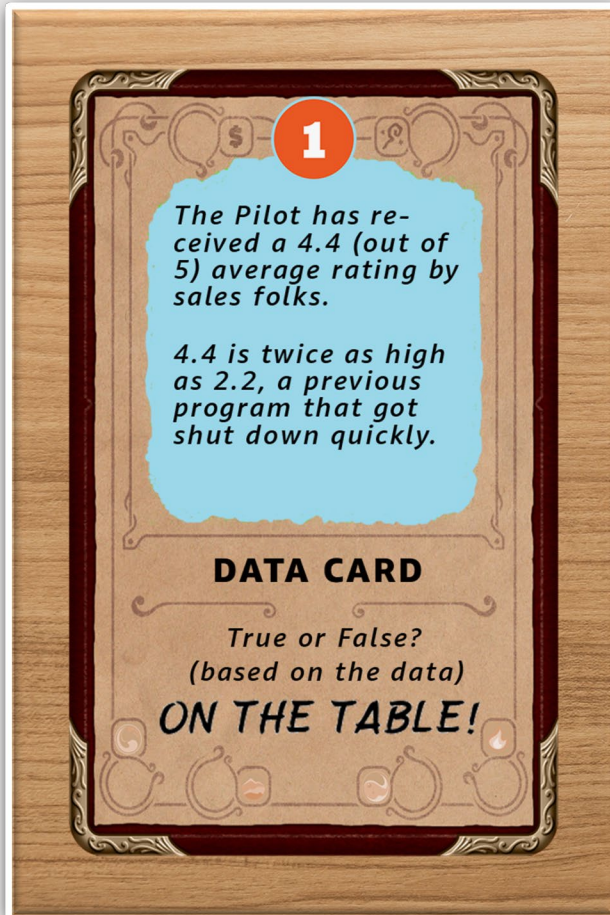
What do
you
remember?

Learning Dashboard – Chart 1



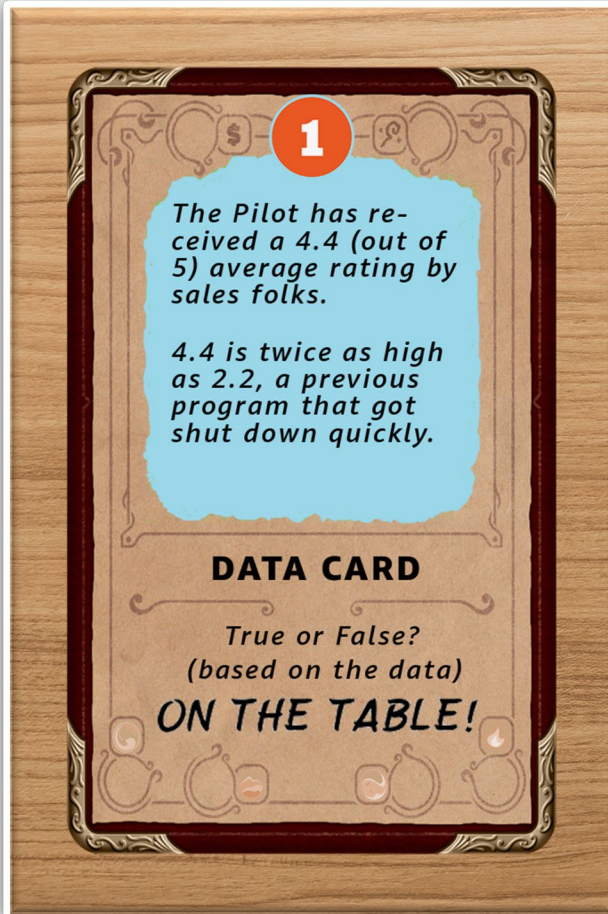
* A single outlier was removed from the calculation because the very low score was due to internet connection not the course content. The 99.58% includes the outlier as well.

Learning Dashboard – Chart 1 (first part: *true)



* A single outlier was removed from the calculation because the very low score was due to internet connection not the course content. The 99.58% includes the outlier as well.

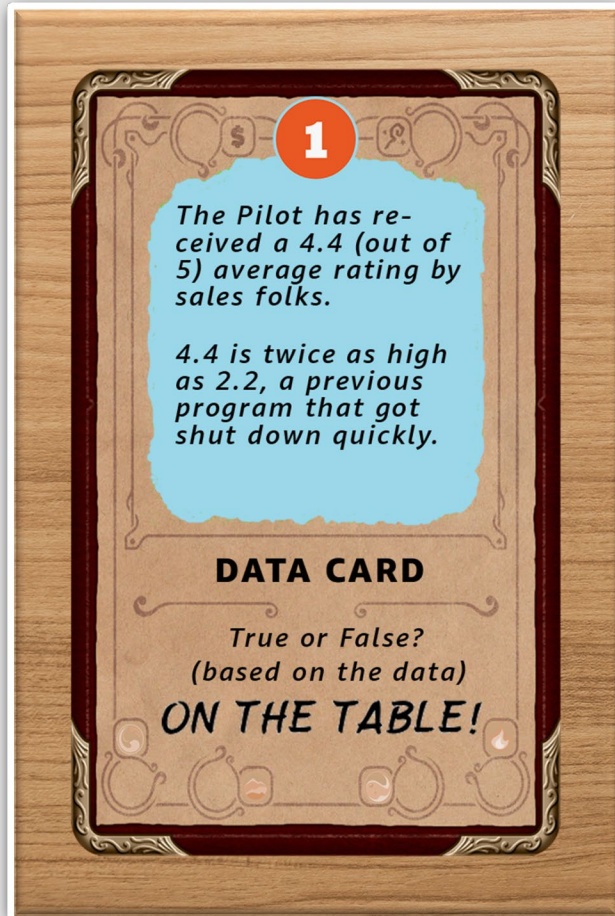
Learning Dashboard – Chart 1 (false or unknown)



First data point: true (4.4 average)

Second data point: 4.4 average
satisfaction is twice as high as 2.2
average satisfaction.

Round 2: Learning Dashboard – Chart 1 (false or unknown)



1 – Very Dissatisfied, 2 – Dissatisfied, 3 – Neutral, 4 – Satisfied, 5 – Very Satisfied

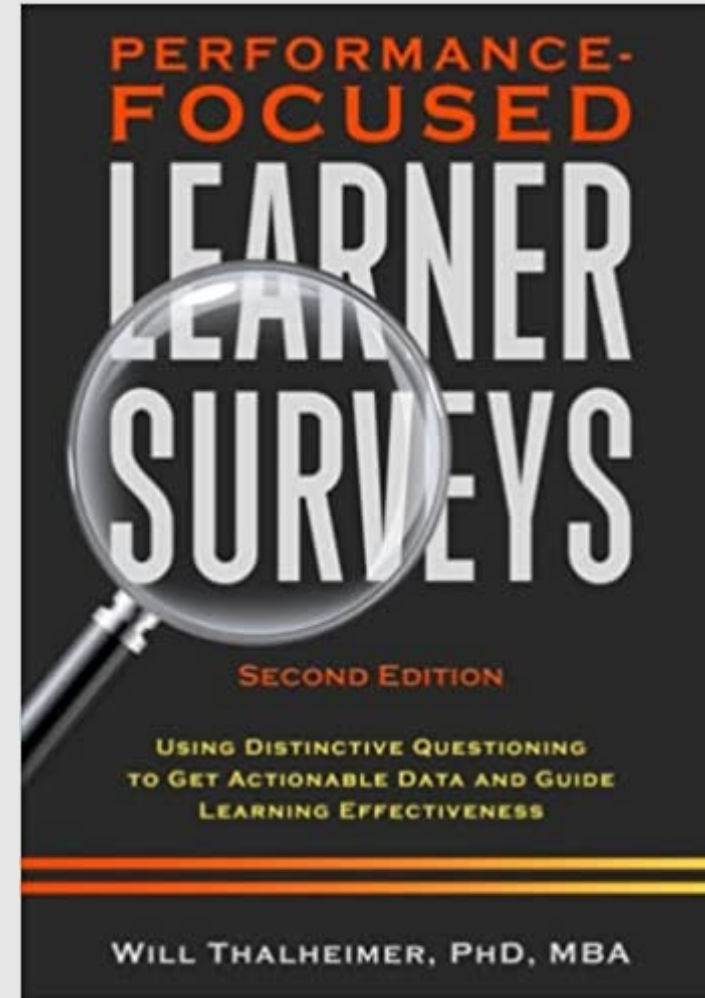
- Can you treat it as a **continuous number** between 1 - 5?
- Are the **distances** between 1,2,3,4, and 5 **equal**? Would it take the **same effort** to move from "Dissatisfied" to "Neutral" as from "Satisfied" to "Very Satisfied?"
- Is there an **absolute 0 (true zero)** representing the lack of satisfaction?

	Nominal	Ordinal	Interval	Ratio
Categories	✓	✓	✓	✓
Ranking		✓	✓	✓
Equal spacing			✓	✓
True zero				✓

Level 1 Likert, NPS, etc.

What's more practical?
Actionable insights:

Performance-focused, forward-looking to support the transfer of learning.



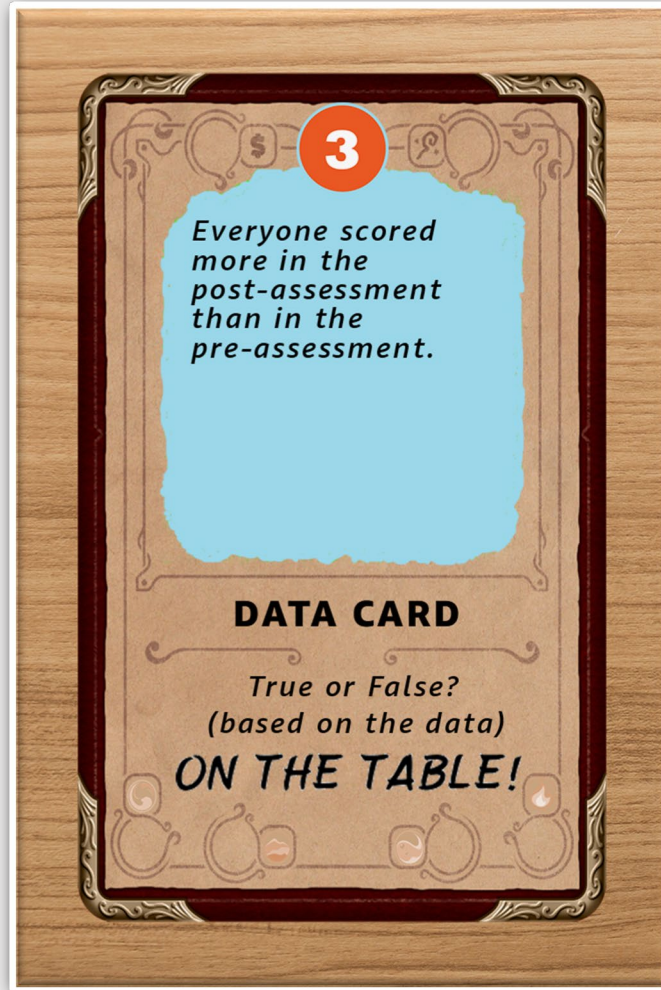
Data Literacy Advice:

“Band-aids don’t fix bullet holes! Counting band-aids is not the same as chasing the cure.”

Dr. Dash Swift



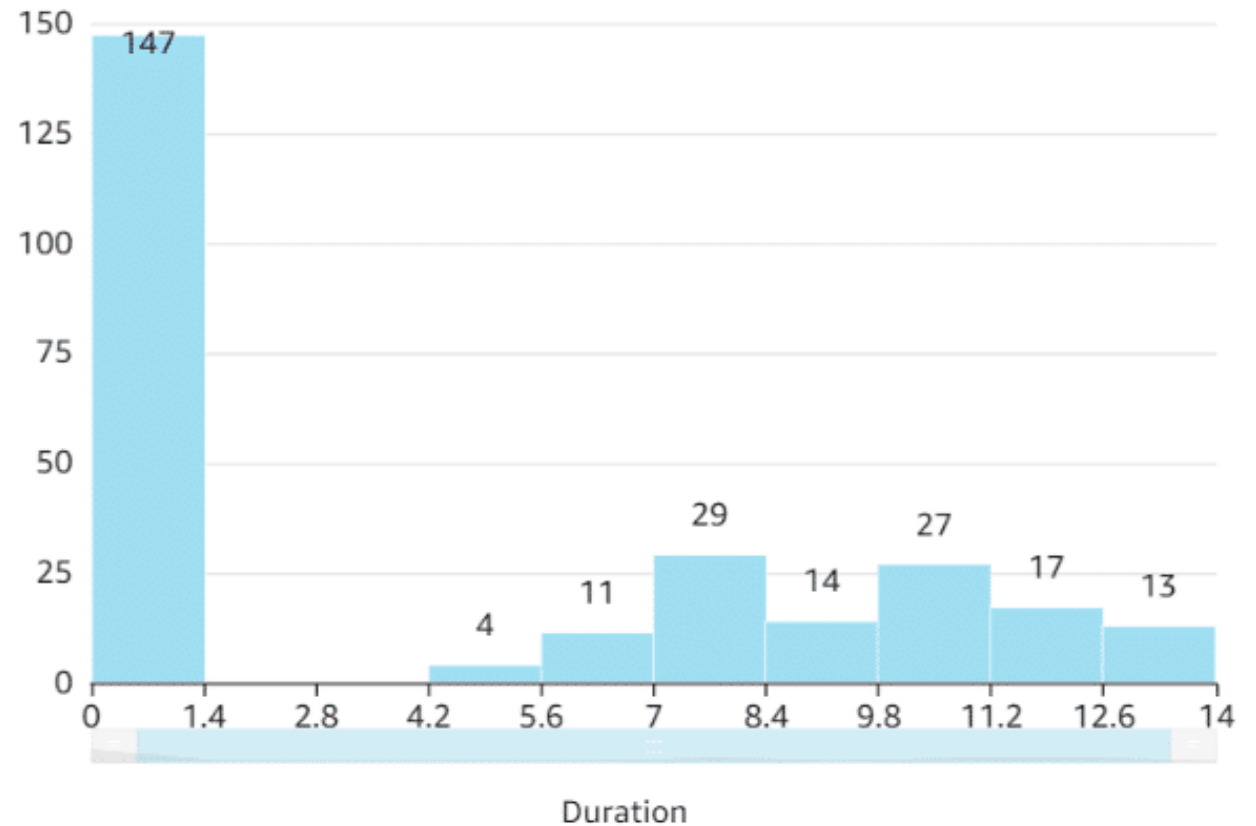
Chart 2 and Chart 3 Data Cards



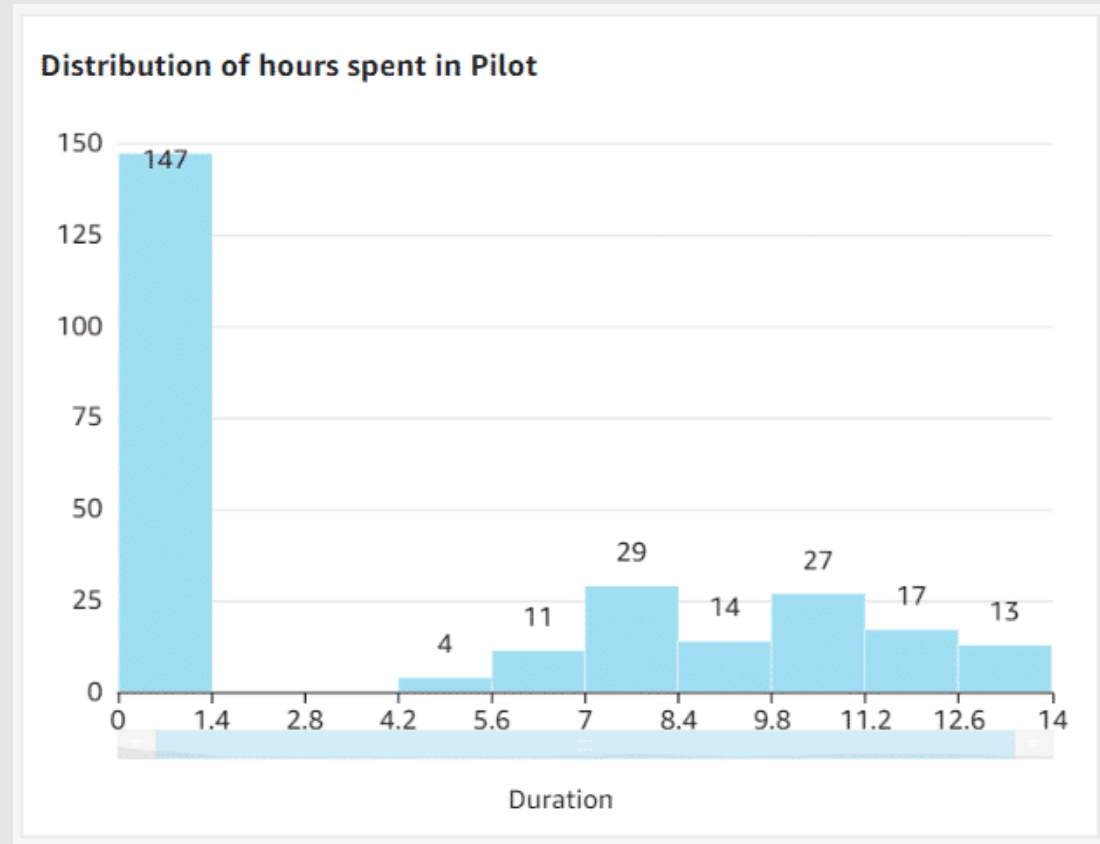
Learning Dashboard – Chart 2



Distribution of hours spent in Pilot



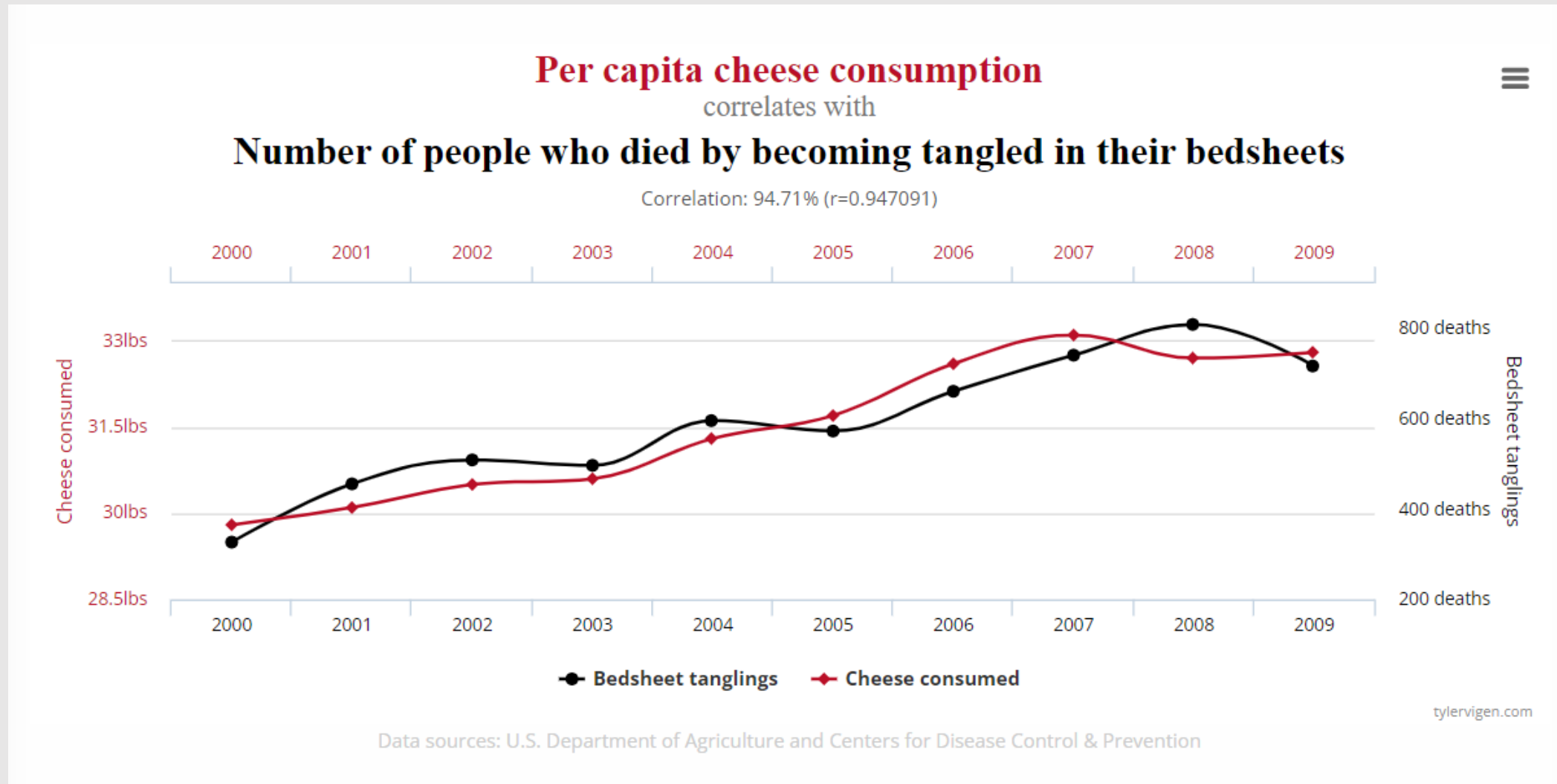
Learning Dashboard – Chart 2 (false or unknown)



Based on Chart 2: unknown.

IMPORTANT: Even if there was a correlation. It does not mean causation.

Correlation is relationship (+, -, 0)



<https://www.tylervigen.com/spurious-correlations>

MYTH # 4

“Put a ring on it!
Correlation =
causation!”

Beyonce





What do
you
remember?

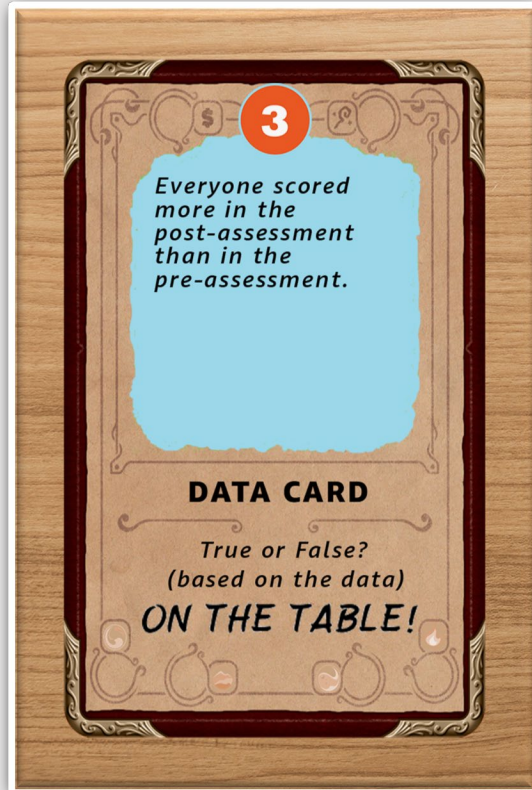


- Mean within a mean: careful relying on a single (meaningless) number!
- You need to calm down: use critical thinking and start chasing the cure!
- Shape of U: know the shape of your data distribution! Is it normal?
- Put a ring on it: correlation may not be causation!
- Data, like hips, don't lie: may not be intentional. Careful with biases!

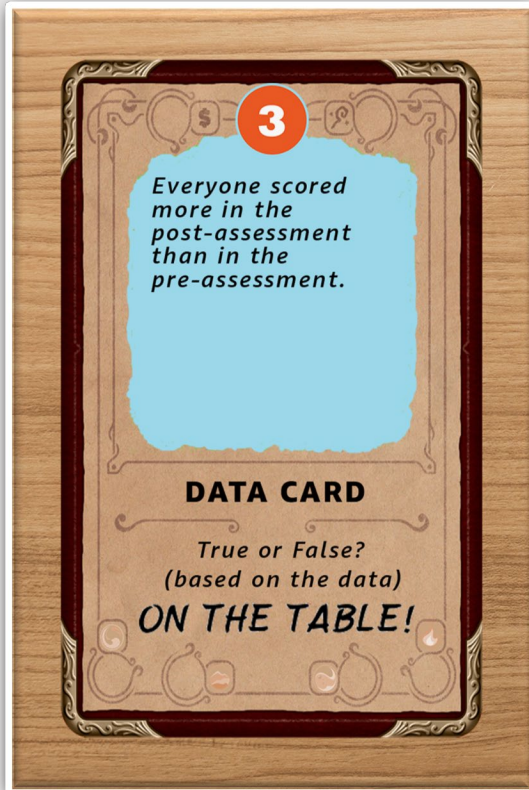
Performance Dashboard: Did the Pilot make a difference?



Round 2: Learning Dashboard – Chart 3



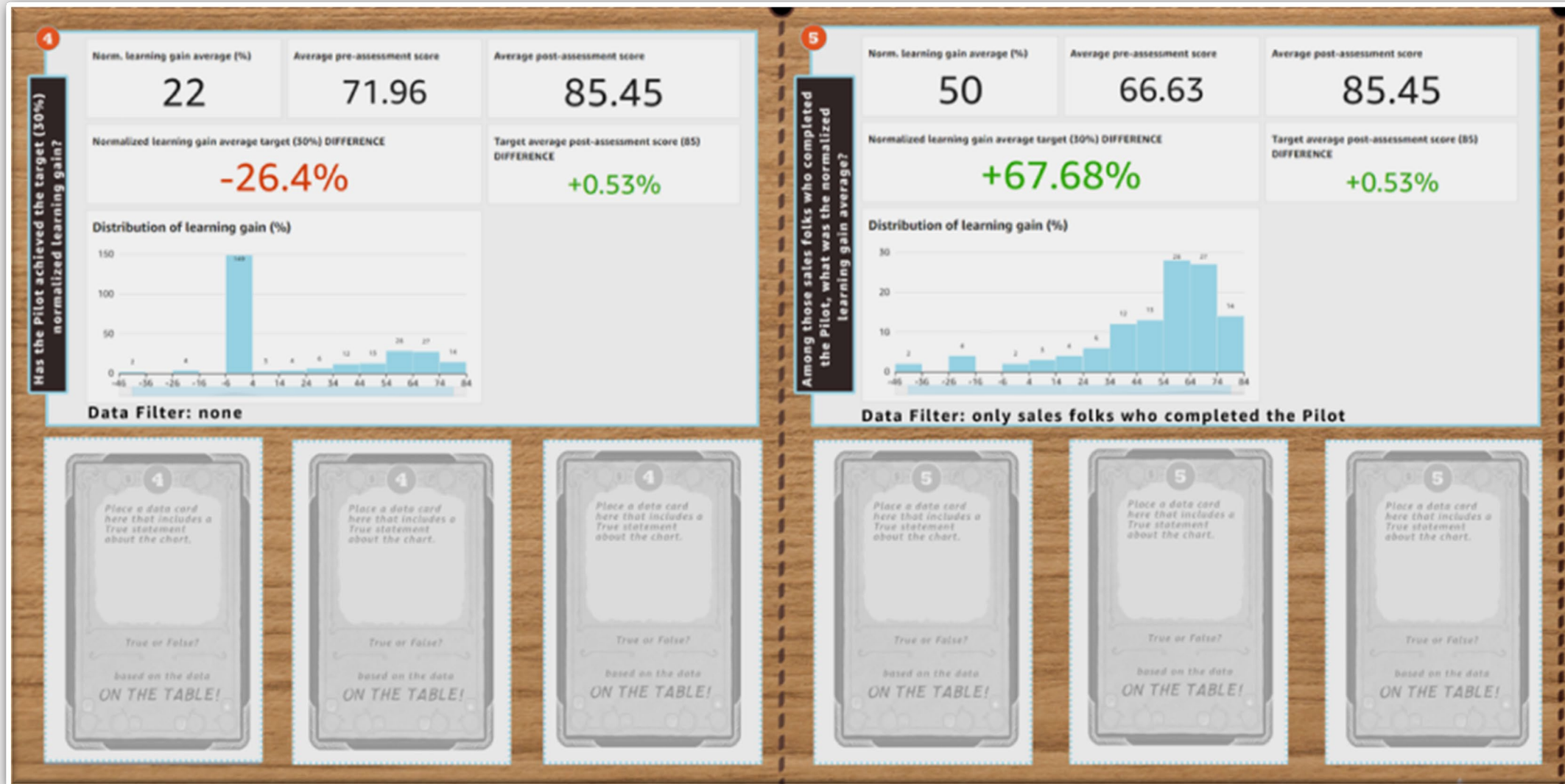
Round 2: Learning Dashboard – Chart 3 (false or unknown)



Based on Chart 3: unknown.
Based on Chart 4 or 5: false

Learning Dashboard – Chart 4 or Chart 5?

Overall, has the Pilot achieved the desired learning impact?



Evaluation Criteria

Pilot success criteria:

1. 100% (261) sales folks (under 2 years of tenure) for pre-assessment
2. 30% (80) sales folks to complete the pilot
3. 85 average post-assessment score
4. 4.4 satisfaction (1 - very unsatisfied, 2 - unsatisfied, 3 - neutral, 4 - satisfied, 5 - very satisfied)
5. 30% average learning gain (calculated as normalized learning change with matched data*: Marx and Cummings 2007).

$$c = \begin{cases} \frac{\text{post} - \text{pre}}{100 - \text{pre}} & \text{post} > \text{pre} \\ \text{drop} & \text{post} = \text{pre} = 100 \text{ or } 0 \\ 0 & \text{post} = \text{pre} \\ \frac{\text{post} - \text{pre}}{\text{pre}} & \text{post} < \text{pre} \end{cases}$$

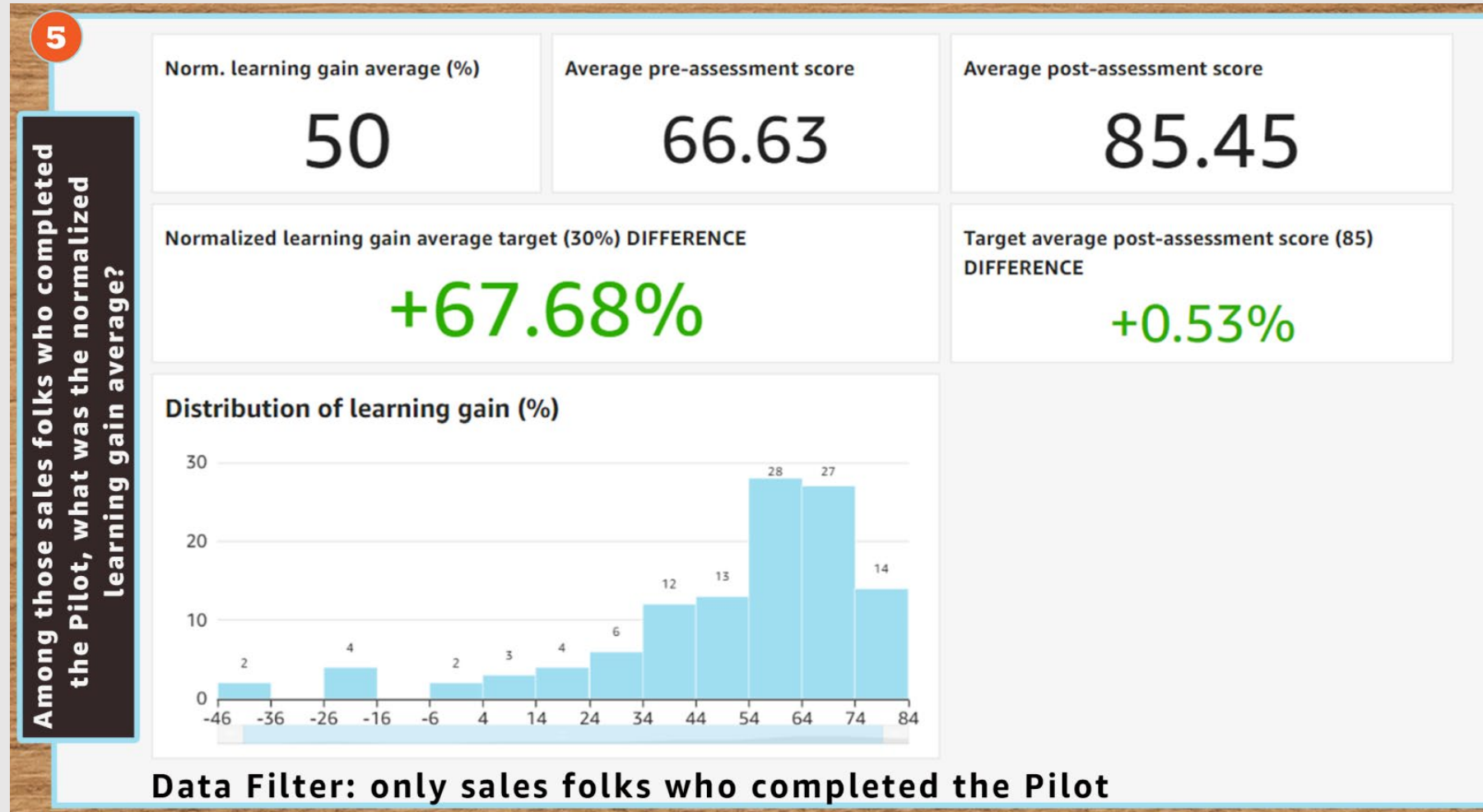
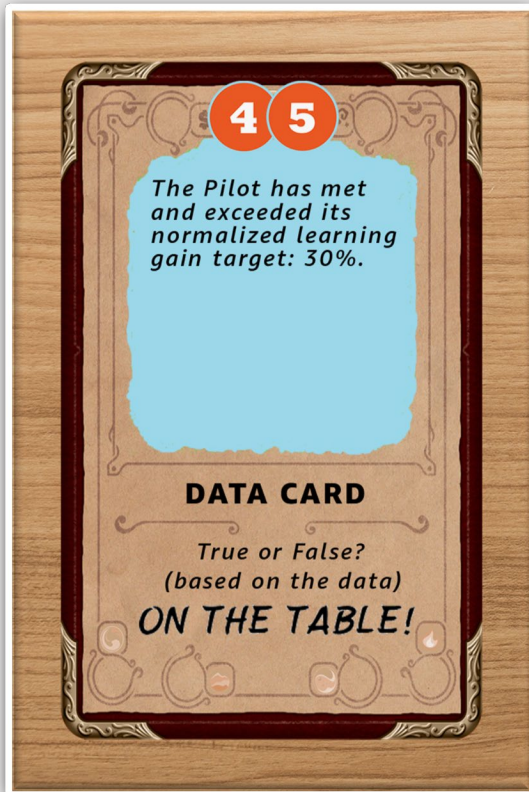
*matched data calculation requires both pre- and post-assessment

Round 4: Learning Dashboard – Chart 5

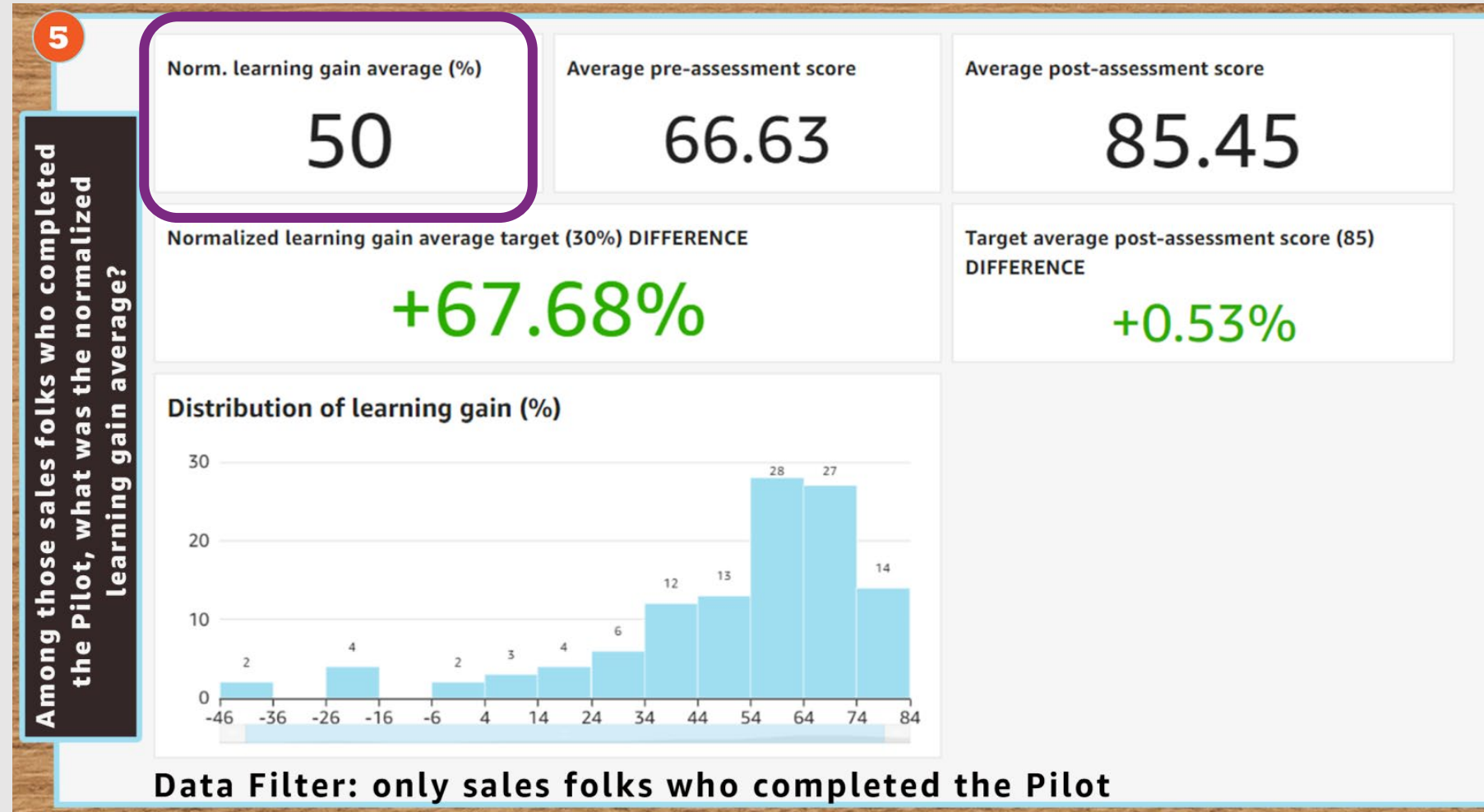
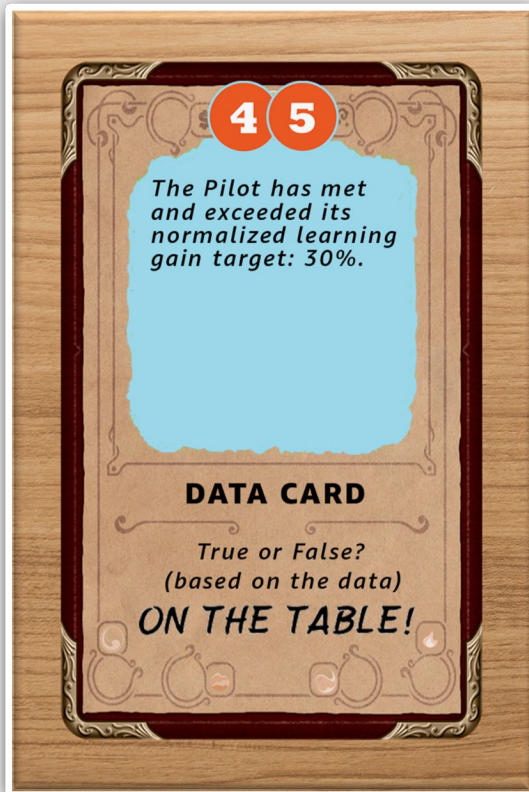
Correct answer: chart 5. Based on the evaluation methodology and criteria agreed on.



Round 2: Learning Dashboard – Chart 5



Round 2: Learning Dashboard – Chart 5 (true)



Performance Dashboard: Did the Pilot cause a performance drop?

2

How well have sales folk who completed the Pilot performed post-Pilot?

Normalized average performance gain (%)

-1.21

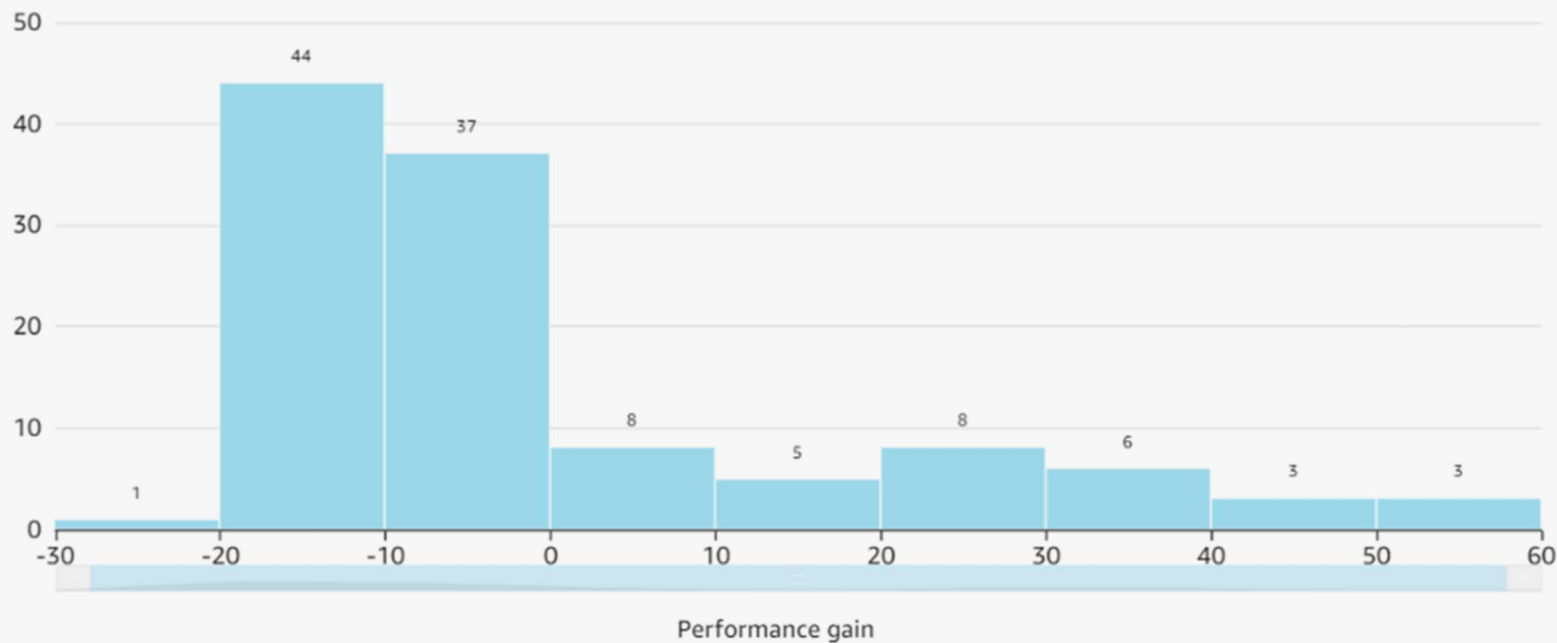
Average performance before Pilot (%)

36.02

Average performance after a Quarter (%)

36.63

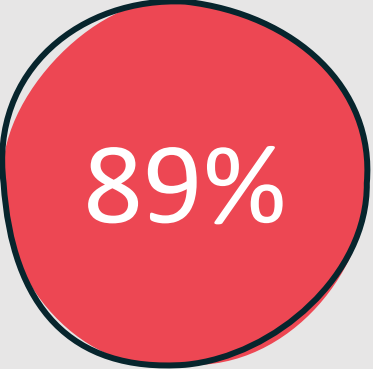
Distribution of performance gain (%)



Data Filter: only sales folks who completed the Pilot

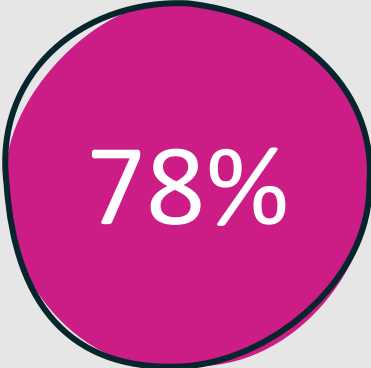
APPENDIX

Do we need data literacy?



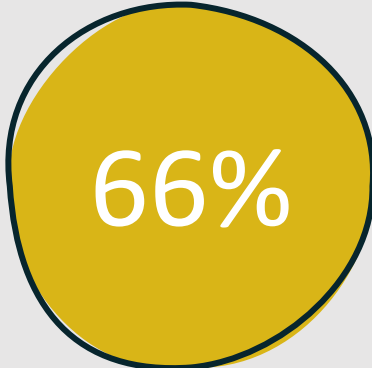
89%

of UK leaders: **data literacy is important** for their teams' daily tasks.



78%

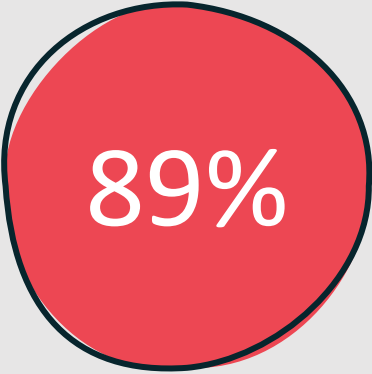
of US leaders: **data literacy is important** for their teams' daily tasks.



66%

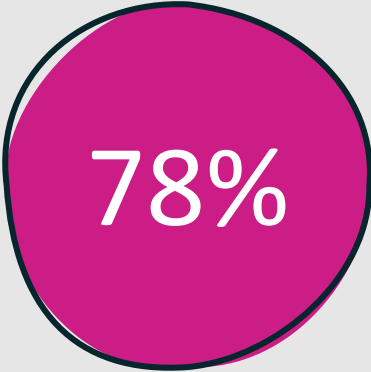
of all leaders: **willing to pay more** for the data literacy skills

Do we need data literacy (n=558)?



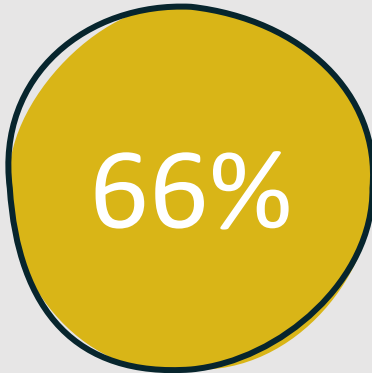
89%

of UK leaders: **data literacy is important** for their teams' daily tasks.



78%

of US leaders: **data literacy is important** for their teams' daily tasks.



66%

of all leaders: **willing to pay more** for the data literacy skills

4C's: Calm down and think!

- Curiosity
- Creativity
- Critical Thinking
- Culture

Curiosity, Creativity, and Critical Thinking are from Jordan Morrow's book: *Be Data Literate*.



Power of the 4C's: Critical Thinking



- Does it make sense?
- What does this mean? (Not how it's calculated!!)
- Is it true? True everywhere?
- Sample size (n)? Methodology?
- How did they define X?
- How was the data collected? Who collected it? What incentives might have been influencing the data collection?
- Is this a snapshot or a trend? Seasonal or average?
- What does the segmentation or data distribution reveal?
- What assumptions were made that we should be aware of?