



PORTOFOLIO - INTENSIVE BOOTCAMP

INTRODUCTION TO DATA ANALYTICS, STATISTICS, AND PROBLEM SOLVING

Felicia Emma Angeline

INTRODUCTION TO DATA ANALYTICS

Data Analytics vs Data Analyst vs Data Analysis

Requirement of a Data Analyst

Analytical Thinking

Analytical Thinking Workflow



COURSE SUMMARY

Data Analytics

Data analytics converts raw data into actionable insights.

Data Analysis

The process or skill of breaking a complex topic or substance into smaller parts to gain better understanding of the data

Data Analyst

The person who examines, cleans, transform, and models data to discover useful information, draw conclusions, and support decision-making.

Requirement of a Data Analyst

Output

Data Analysis/
Investigation



Skill Requirement

Analytical thinking,
SQL, Sheets

Dashboard



Looker / Redash /
Tableau / Other Data
Visualization Tool

Dataset



SQL, Google Sheets

Pipeline and Data
Quality Maintenance



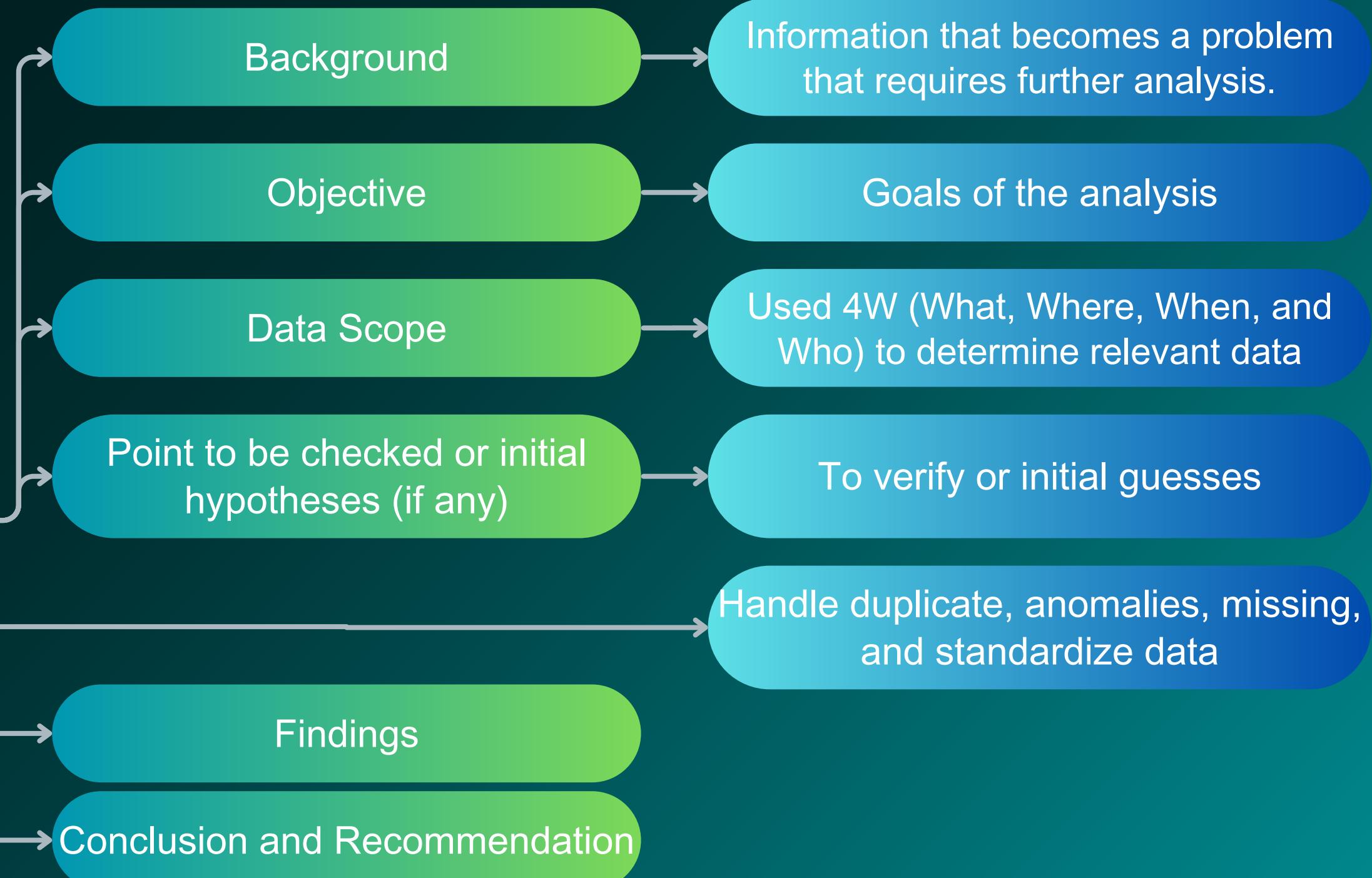
DBT, Python

COURSE SUMMARY

Analytical Thinking

Analytical thinking is a tool to :

1. Identify and define problems,
2. Extract key information from data, and
3. Develop workable solutions for the problems identified.



TASK – PROBLEM 1

Context :

- In one of the Online Travel Agent Company, you act as a data analyst for the Flight product.
- An anomaly was found that there was a spike in sales on November 3, 2023.
- The spike was 50% compared to the previous day and 25% compared to the same date last month.
- On the other hand, 2 promo marketing campaigns are running:
 - Promo Flight 50% with max 200k for new customers.
 - Promo Flight 10% with max 500k for existing customers.
- On the other hand, there is a referral promo running this month where
- Every user who joins/transacts for the first time will get a gift of IDR 50k for the person who is invited and invites.
- Users come to you to find out what caused the above incident.

Create an Analytical Thinking Framework for the above case that includes:

- Background
- Objective
- Data
- Point to be checked / Initial Hypotheses

TASK - ANSWER FORM 1

Context : Data Analyst at OTA

Background

- On November 3, 2023, a significant spike in flight sales was observed:
 - +50% compared to November 2, 2023
 - +25% compared to October 3, 2023

Objective

- Identify the root cause of the sales spike on November 3, 2023.

Data

- What : Flight transaction data (number of transactions and total value), Promo usage details
- Where : Platform where transactions occurred
- When : Transaction timestamps (at least one month before and after November 3, 2023)
- Who : New users vs. existing users, users who used promos vs. those who didn't, users involved in referral program

Point to be checked/ Initial Hypotheses

- Was the spike mainly driven by promotional campaigns (new/existing users)
- Were there external triggers (e.g., long weekend, airline sale, news)

TASK – PROBLEM 2

Context :

- At one of the Money Transfer companies.
- It was found that the average transaction waiting time had spiked.
- This caused many users to complain.
- The incident occurred during the period Nov 1-5, 2023.
- The cause of the above incident is not yet known.

Create an Analytical Thinking Framework for the above case that includes:

- Background
- Objective
- Data
- Point to be checked / Initial Hypotheses

TASK - ANSWER FORM 2

Context : Data Analyst at Money Transfer

Background

- Between November 1–5, 2023, the average transaction waiting time increased significantly, resulting in a surge of user complaints.
- The exact cause of the delay is unknown.

Objective

- Identify the root cause of the increased transaction waiting time during Nov 1–5, 2023.

Data

- What : Transaction logs (timestamps: initiated, processed, completed)
- Where : Affected platforms (mobile app, website, API)
- When : Focus on Nov 1–5, 2023
- Who : User segments affected

Point to be checked/ Initial Hypotheses

- Was there any system degradation (e.g., high server load, slow processing)
- Was there any recent product deployment or system update before Nov 1

TASK – PROBLEM 3

Context :

- You are one of the Marketing Data Analysts at a Tech Company.
- It was found that there was a decrease in the number of customer acquisitions*, even though the advertising budget spent tended to be stagnant.
- This causes the cost of customer acquisition (CAC, customer acquisition cost) to be more expensive.
- The incident occurred during the 3rd and 4th week of the end of 2023 yesterday.

*customer acquisition: the number of customers who sign up for an app/platform

Create an Analytical Thinking Framework for the above case that includes:

- Background
- Objective
- Data
- Point to be checked / Initial Hypotheses

TASK – ANSWER FORM 3

Context : Marketing Data Analyst at Tech Company

Background

- During Week 3 and Week 4 of late 2023, there was a noticeable decrease in customer acquisition (i.e., fewer users signing up for the platform).
- However, the advertising budget remained relatively stable, causing the Customer Acquisition Cost (CAC) to increase.

Objective

- Identify the factors leading to the drop in customer acquisition during the specified period.

Data

- What : Daily/weekly new customer acquisition numbers
- Where : Platforms used, user acquisition sources
- When : Week 3–4 of late 2023
- Who : Target audience segments (age, gender, device, interest group), segmentation by campaign targeting (new vs. returning users)

Point to be checked/ Initial Hypotheses

- Were there technical issues (slow load, errors, broken links)
- Were there seasonal trends, holidays, or market events affecting behavior
- Did competitors launch aggressive campaigns during the same period

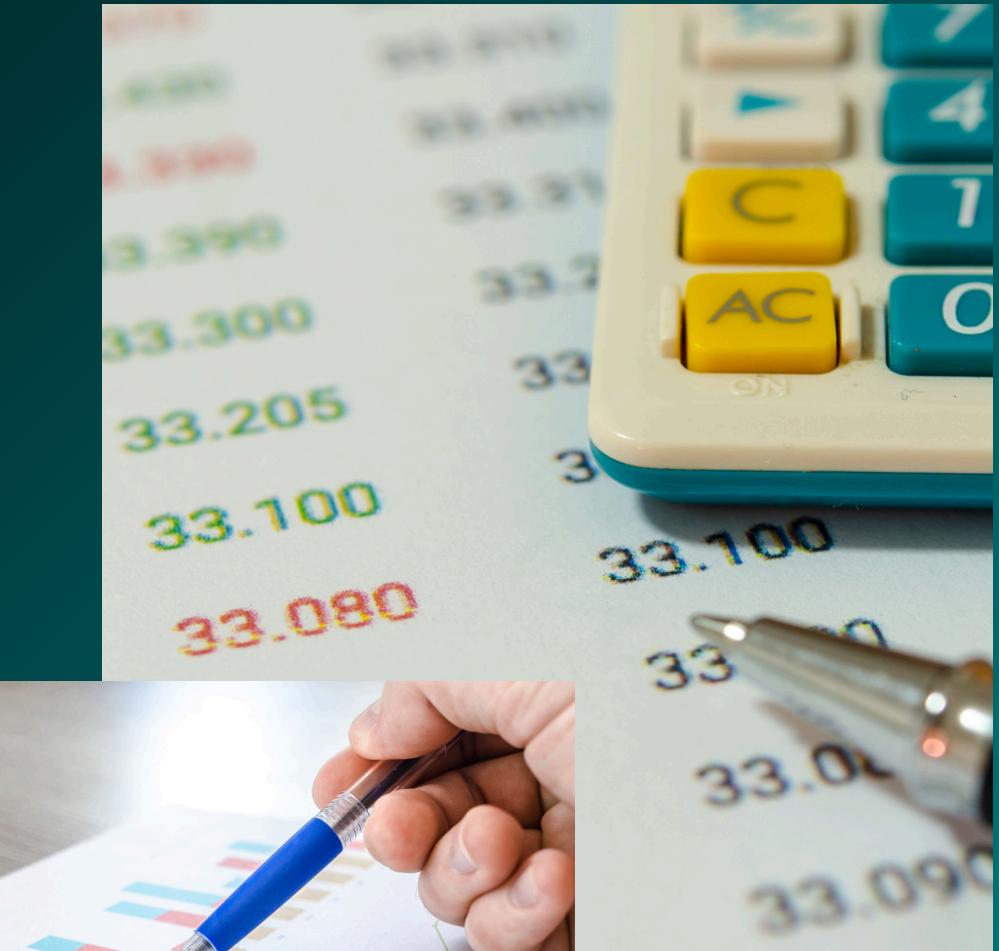
INTRODUCTION TO BASIC STATISTICS

Statistics, Statistical

Type of Data

Measure of Centric

Measure of Distribution



COURSE SUMMARY

Statistics

Data (or fact) measure values.

Statistical

The study of collecting, processing, analyzing, and making conclusions about data.

Continuous (Measured)

Airplane speed, time spent waiting in line

Discrete (Counted)

Number of pets, number of packages shipped

Type of Data

Observation/Data

Quantitative / Numeric

Qualitative / Categorical

Ordinal / Rank (Ordered)

Train passenger class (economy, business, executive)

Nominal (Unordered)

Married/unmarried, country of residence

COURSE SUMMARY

Measure of Centric

Information where data is collected with a certain size/number.

→ **Mean**

$$x_{mean} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

→ **Median**

$$\text{Med}(X) = \begin{cases} X[\frac{n+1}{2}] & \text{if } n \text{ is odd} \\ \frac{X[\frac{n}{2}] + X[\frac{n}{2}+1]}{2} & \text{if } n \text{ is even} \end{cases}$$

→ **Mode**

→ **Range**

Measure of Distribution

Information about the distribution of data around the data center.

→ **Variance**

→ **Standard Deviation**

→ **Z-Score**

→ **Quartile, Interquartile Range, and Percentile**

→ **Outlier**

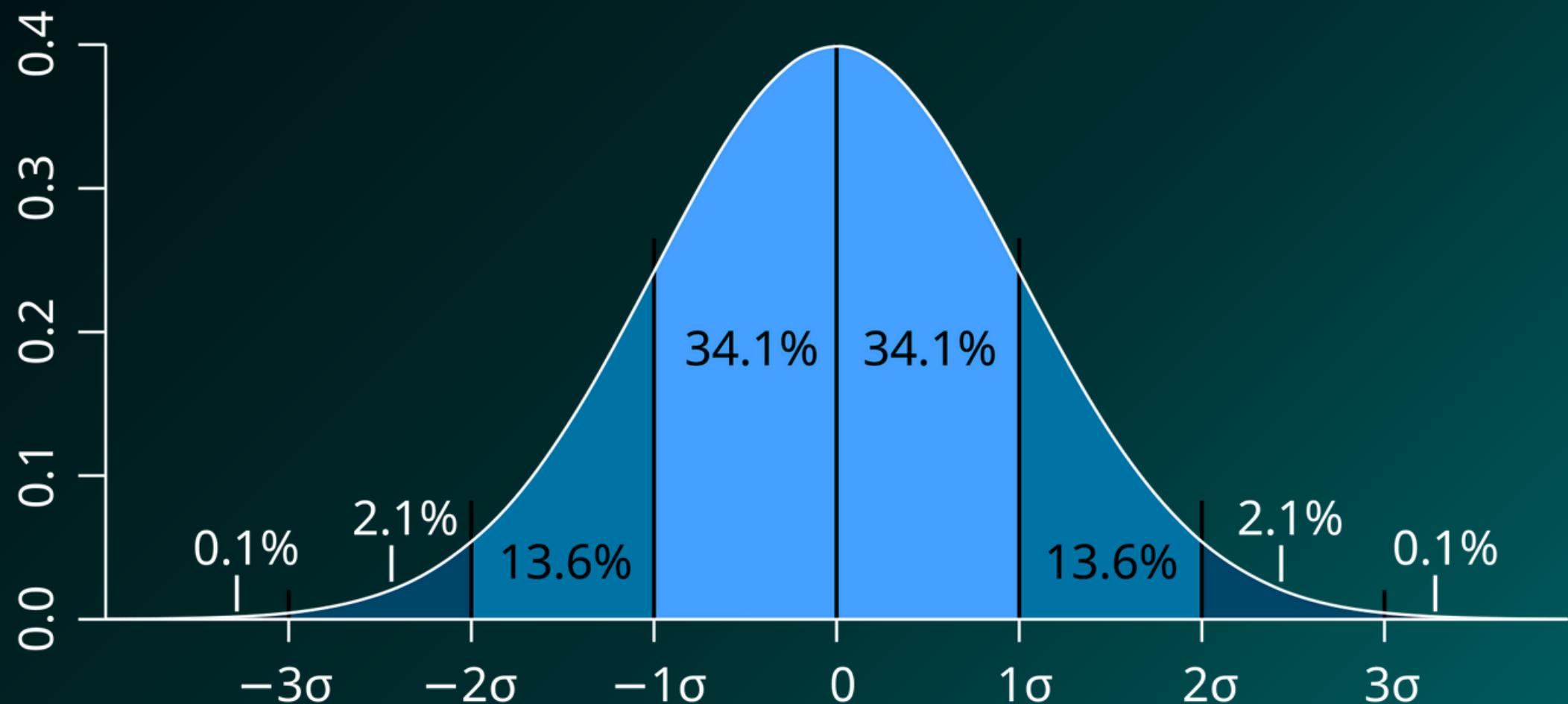
COURSE SUMMARY

Variance

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

Standard Deviation (Sample)

$$s = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}}$$



The higher the curve, the closer the standard deviation value is to the mean, the lower the standard deviation.

The shorter the curve, the further away the standard deviation value is from the mean, the higher the standard deviation.

Normal Distribution

± 1 std.dev (left and right of the mean) has covered 68.2% of the data

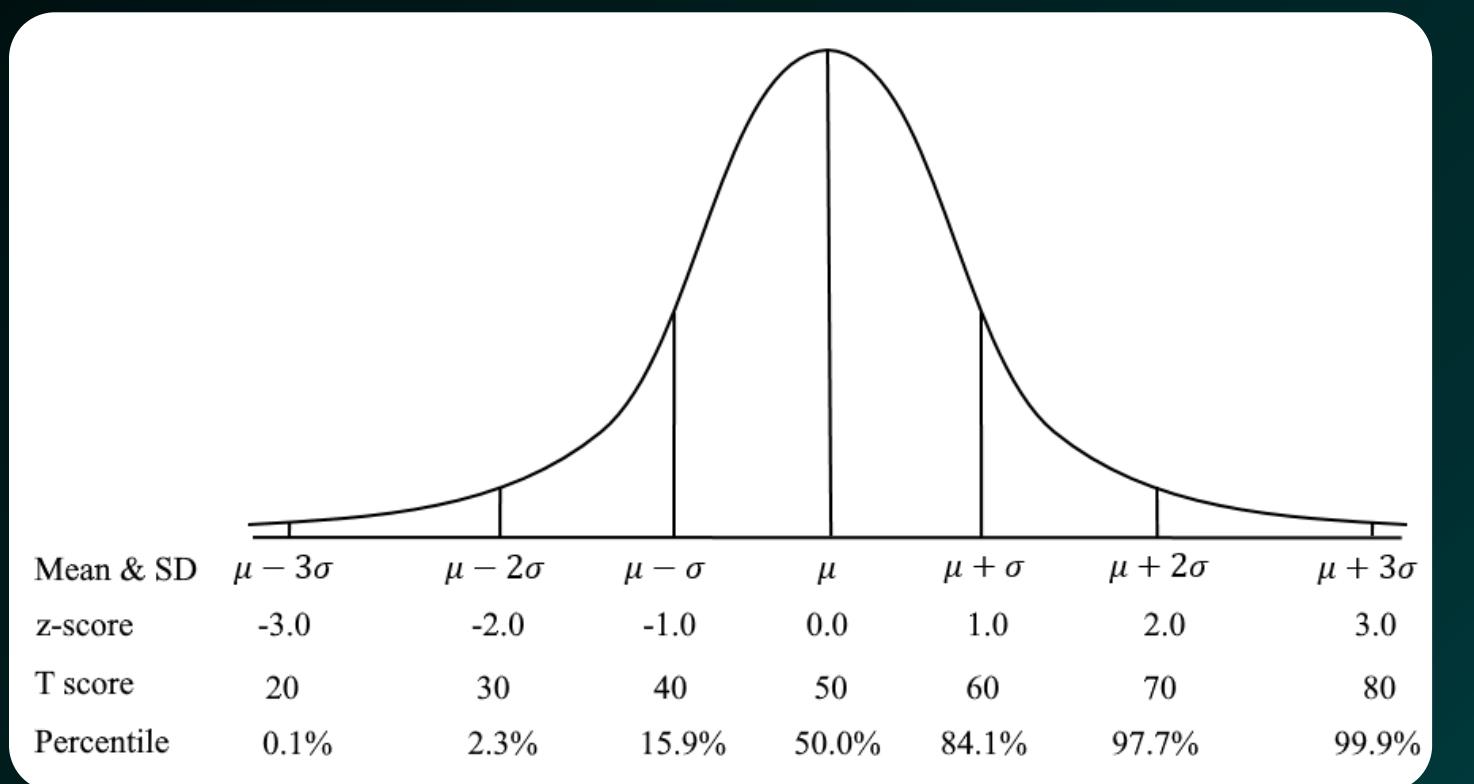
± 2 std.dev has covered 95.4% of the data

± 3 std.dev has covered 99.8% of the data

COURSE SUMMARY

Z-Score

Statistical measure that quantifies how far individual values in a data set are from the mean in units of standard deviations.



$$Z = \frac{(X - \mu)}{\sigma}$$

- If the Z-score is positive, the value is above average.
- If the Z-score is negative, the value is below average.
- The larger the Z-score (positive or negative), the further the value is from the average.

**Standard Normal Distribution
Z-Score Table**

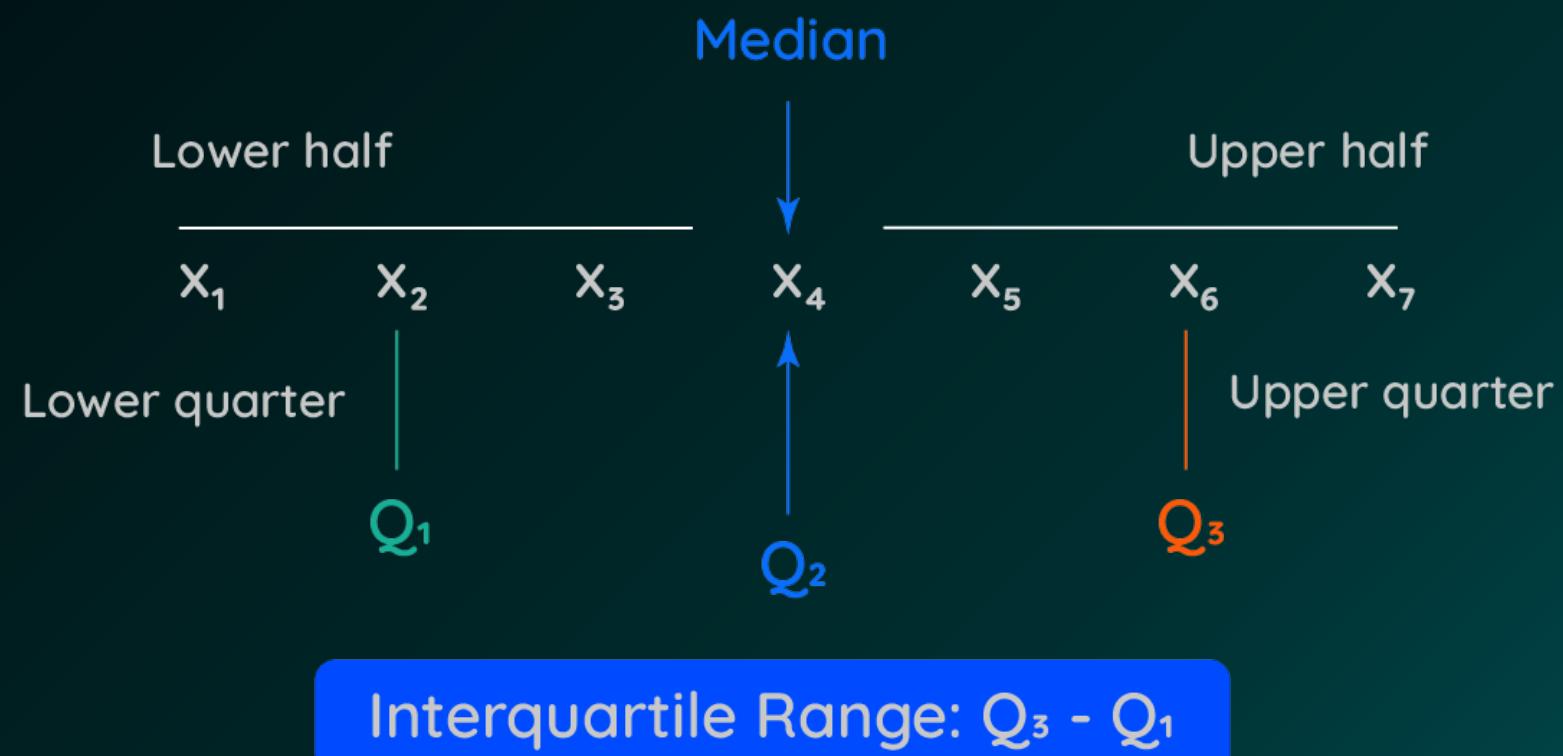
Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-1.4	0.08076	0.07927	0.07780	0.07636	0.07493	0.07353	0.07215	0.07078	0.06944	0.06811
-1.3	0.09680	0.09510	0.09342	0.09176	0.09012	0.08851	0.08691	0.08534	0.08379	0.08226
-1.2	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10383	0.10202	0.10027	0.09853
-1.1	0.13567	0.13350	0.13136	0.12924	0.12714	0.12507	0.12302	0.12100	0.11900	0.11702
-1.0	0.15866	0.15625	0.15386	0.15151	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
-0.9	0.18406	0.18141	0.17879	0.17619	0.17361	0.17106	0.16853	0.16602	0.16354	0.16109
-0.8	0.21186	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
-0.7	0.24196	0.23885	0.23576	0.23270	0.22965	0.22663	0.22362	0.22062	0.21770	0.21476
-0.6	0.27425	0.27093	0.26763	0.26435	0.26109	0.25785	0.25463	0.25143	0.24825	0.24510
-0.5	0.30854	0.30503	0.30153	0.29806	0.29460	0.29116	0.28774	0.28434	0.28096	0.27760
-0.4	0.34458	0.34090	0.33724	0.33360	0.32997	0.32636	0.32276	0.31918	0.31561	0.31207
-0.3	0.38209	0.37828	0.37448	0.37070	0.36693	0.36317	0.35942	0.35569	0.35197	0.34827
-0.2	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	0.39743	0.39358	0.38974	0.38591
-0.1	0.46017	0.45620	0.45224	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42465
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189

COURSE SUMMARY

Quartile, Interquartile Range, and Percentile

Quartiles are values that divide a set of data into four equal parts.

Percentile it's the ratio of data points below a given value to the total number of data points, expressed as a percentage.



Quartile Formula



Lower Quartile (Q1) = $(N+1) \times \frac{1}{4}$

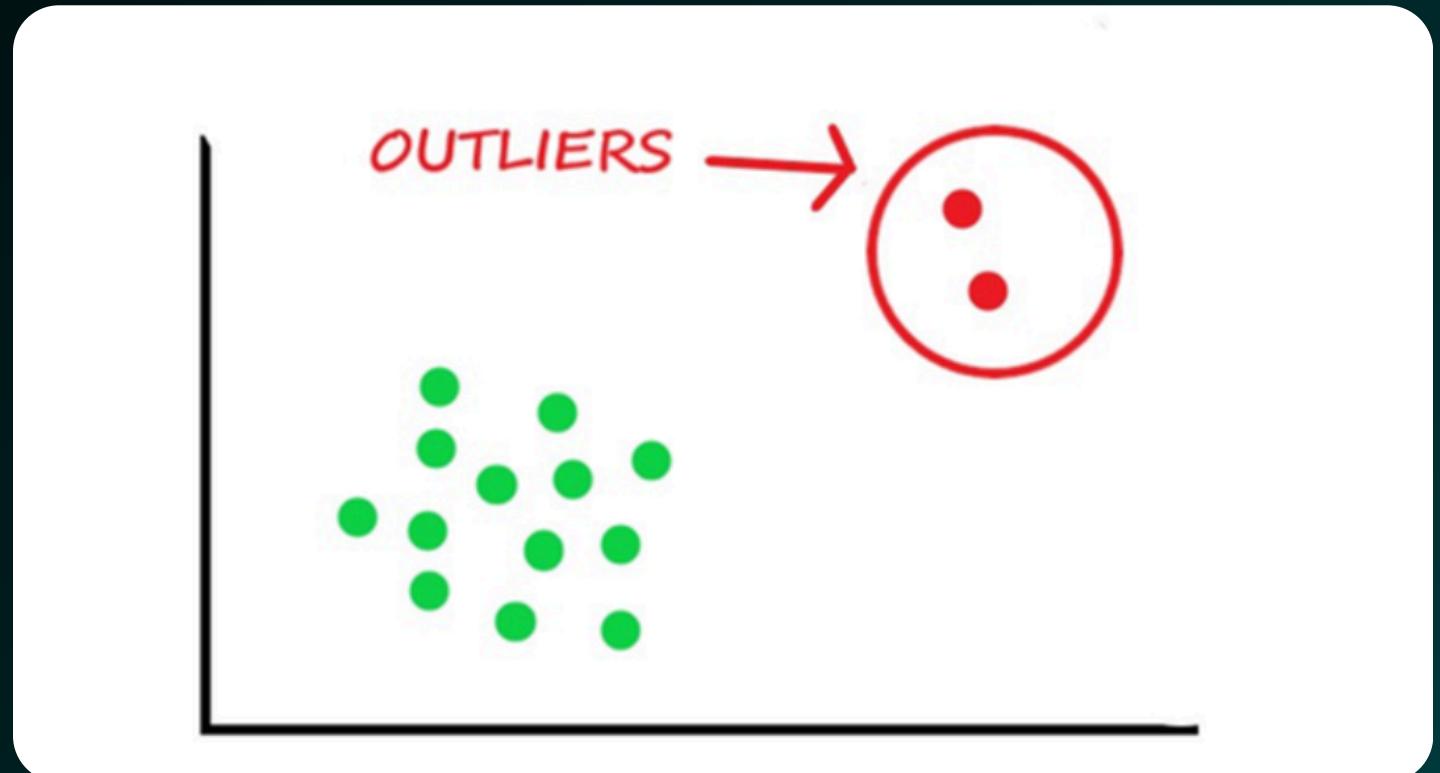
Middle Quartile (Q2) = $(N+1) \times \frac{2}{4}$

Upper Quartile (Q3) = $(N+1) \times \frac{3}{4}$

COURSE SUMMARY

Outlier

Single data point that goes far outside the average value of a group of statistics



Steps for determining outliers with IQR

1. Calculate the first Quartile (Q1) and the third Quartile (Q3)
2. Calculate the Interquartile Range (IQR)
 - $IQR = Q3 - Q1$
3. Determine the outlier boundary
 - Lower bound = $Q1 - 1.5 * IQR$
 - Upper bound = $Q3 + 1.5 * IQR$
4. Outlier Identification

Data that is smaller than the lower limit or larger than the upper limit is considered an outlier.

TASK – STANDARD DEVIATION

Date	Visitor of the Public Service
2022-12-01	4
2022-12-02	2
2022-12-03	5
2022-12-04	10
2022-12-05	7
2022-12-06	8
2022-12-07	10
2022-12-08	5
2022-12-09	8
2022-12-10	4
2022-12-11	9
2022-12-12	6
2022-12-13	10
2022-12-14	8
2022-12-15	4
2022-12-16	6
2022-12-17	8
2022-12-18	5
2022-12-19	7
2022-12-20	3
2022-12-21	9
2022-12-22	7
2022-12-23	5
2022-12-24	4
2022-12-25	7
2022-12-26	7
2022-12-27	7
2022-12-28	5
2022-12-29	8
2022-12-30	5

Case

If you are the leader of one institution in Surabaya, how many chair that you need to prepared to cover 68% of all visitior at least will get the seat?

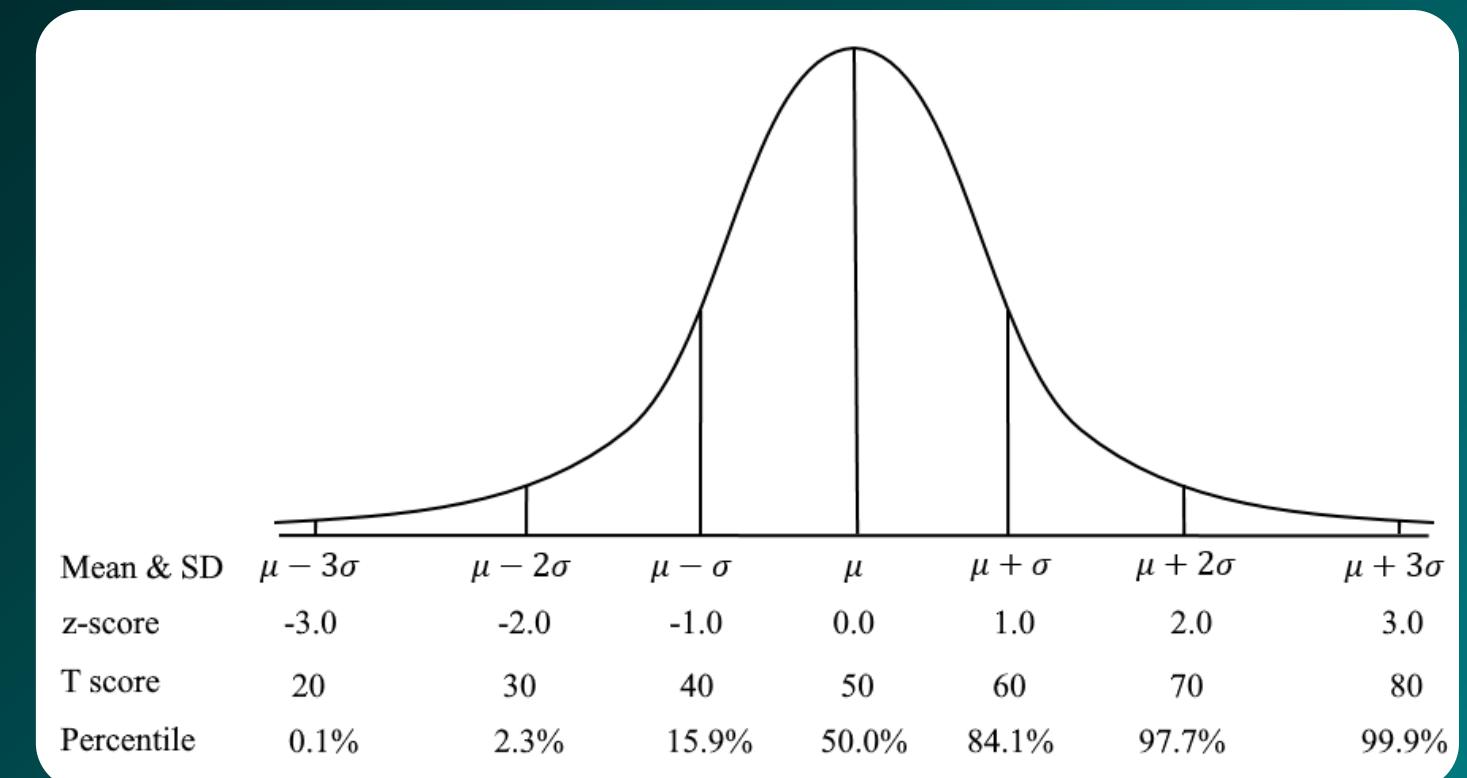
Answer

MEAN	6 . 43	=AVERAGE(B2 : B31)
STDEV	2 . 14	=STDEV . S(B2 : B31)

With assuming the data is normally distributed
We can prepare the chair to cover 68% of the seat at least for:

$$\mu + (1 \times \sigma) = 6.43 + (1 \times 2.14) = 8.57 \approx 9$$

Need 9 seats to be prepared



TASK - Z-SCORE

Udin's Friend Data	Monthly Living Costs	
	Bekasi	Tuban
1	4,800,000	2,500,000
2	5,000,000	2,100,000
3	4,200,000	3,000,000
4	4,600,000	3,000,000
5	4,500,000	2,500,000
6	5,200,000	2,400,000
Job Offer with Salary to the Udin	Bekasi	Tuban
	5,000,000	3,500,000

Case

Which one should Udin take when considering the cost of living based on his neighborhood?

Answer

	BEKASI	TUBAN	
MEAN SAMPLE	4,716,667	2,583,333	=AVERAGE()
STD. SAMPLE	360,092.58	354,494.95	=STDEV.S()
Z-SCORE	0.79	2.59	$=\frac{(X - \mu)}{\sigma}$
Z-SCORE TABLE	78.52%	99.52%	

Based on statistical calculations, the salary offer in Tuban is more favorable considering the average cost of living in Tuban is much cheaper with the salary that will be obtained.

TASK - PERCENTILE

Transaction ID	Service Duration	Transaction ID	Service Duration
1	1	26	2
2	4	27	7
3	8	28	10
4	5	29	7
5	5	30	4
6	1	31	3
7	10	32	10
8	5	33	8
9	6	34	10
10	1	35	4
11	1	36	10
12	10	37	10
13	2	38	5
14	1	39	10
15	4	40	9
16	6	41	3
17	8	42	7
18	4	43	9
19	9	44	9
20	5	45	6
21	10	46	3
22	7	47	5
23	10	48	7
24	2	49	6
25	10	50	8

Case

In one money transfer company, the expected of **Percentile 90 SLA** (Service Level Agreement) is **under 5 mins** to ensure the customer satisfaction of the service provided.

There are **50 transactions** occurred with each of SLA is attached.

Is the company was achieved P90 Satisfaction Level Condition?

What is your recommendation to product team to solve this condition?

Answer

P90	10	=PERCENTILE.INC(\$B\$2:\$B\$51;0.9)
-----	----	-------------------------------------

The company did not achieve the P90 SLA target (≤ 5 minutes)

It is recommended to **identify the causes** of transaction delays, **improve system efficiency**, and implement real-time monitoring to ensure transaction processing runs faster and consistently under the 5-minute limit.

TASK - STATS

uid	Voucher Redeem
1	98
2	112
3	1
4	41
5	106
6	111
7	86
8	142
9	143
10	88
11	32
12	66
13	92
14	94
15	118
16	53
17	119
18	500
19	37
20	90
21	128
22	79
23	86
24	45
25	33
26	54
27	115
28	129
29	56
30	83

Case

You are a part of CRM Team and asked to evaluate the redeemed voucher of the day.

1. What are the mean, median, mode, std.dev, and upper & lower threshold
2. Is there any outlier from the last month performance?
3. If yes? how many?

Answer

MEAN	97.9	=AVERAGE(\$B\$2 : \$B\$31)
MEDIAN	89	=MEDIAN(\$B\$2 : \$B\$31)
MODE	86	=MODE(\$B\$2 : \$B\$31)
VARIANCE	7054.92069	=VAR(\$B\$2 : \$B\$31)
STD DEV S	83.99357529	=STDEV.S(\$B\$2 : \$B\$31)
PERCENTILE 25 , Q1	54.5	=PERCENTILE.INC(\$B\$2 : \$B\$31 ; 0.25)
PERCENTILE 75 , Q3	114.25	=PERCENTILE.INC(\$B\$2 : \$B\$31 ; 0.75)
IQR	59.75	=E21-E20
LOWER THRESHOLD	-35.125	=E20-(1.5*E22)
UPPER THRESHOLD	203.875	=E21+(1.5*E22)
COUNTIF	1	=COUNTIF(\$B\$2 : \$B\$31 ; ">203.875")

There is one outlier with a value of 500 in uid 18.

INTRODUCTION TO PROBLEM SOLVING

What is Problem?

Problem Solving Framework



COURSE SUMMARY

What is Problem?

Problem = Gap

Gap between ideal and existing condition

Problem Solving Framework

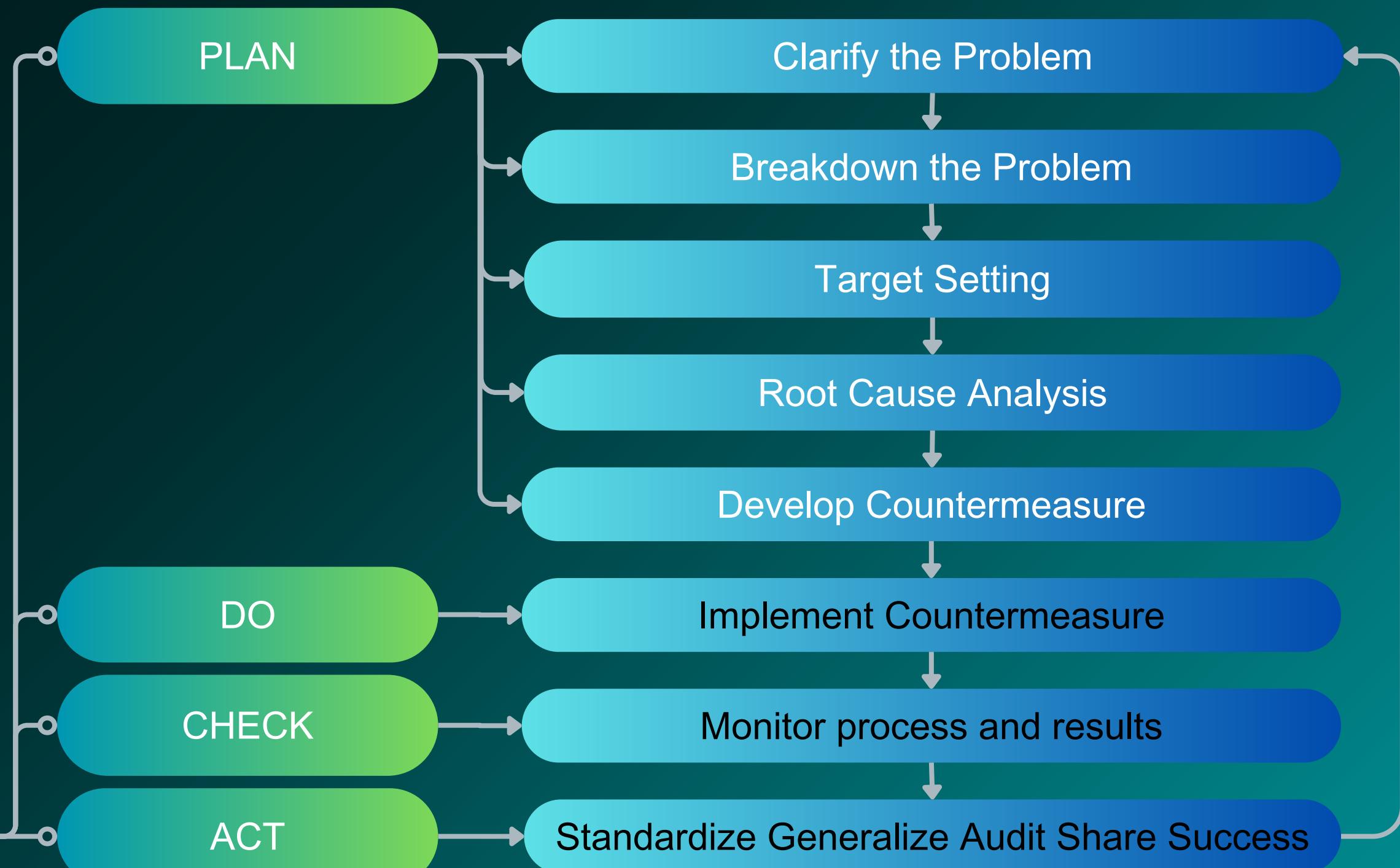
There is many alternative tool in problem-solving process:

1 A3 Problem Solving

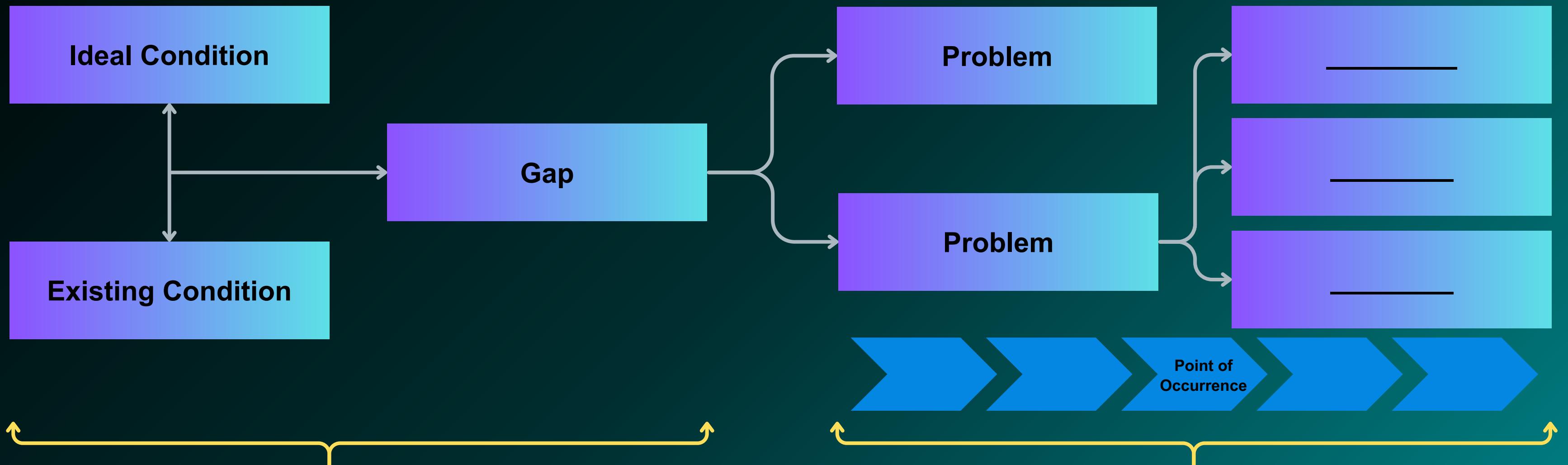
2 DMAIC (Six Sigma)

3 Issue Tree

4 PDCA



COURSE SUMMARY



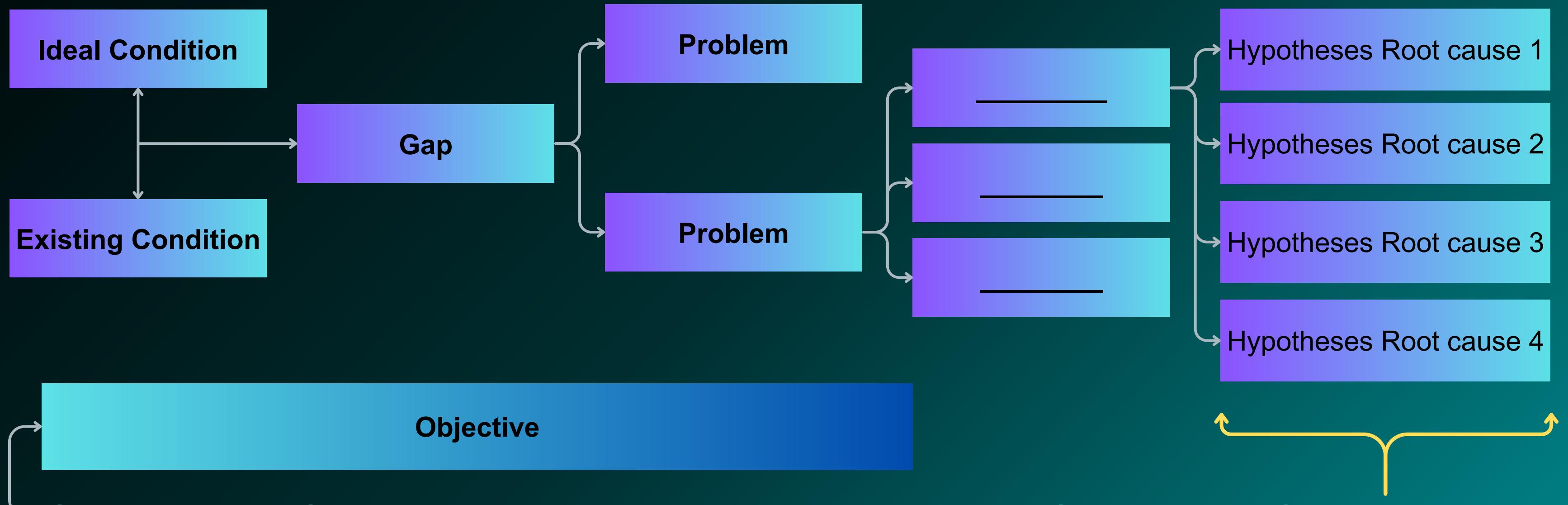
Step 1: Clarify the Problem

- Clarify the Ultimate Goal
- Clarify the Ideal Situation and Current Situation
- Visualize the gap

Step 2: Breakdown the Problem

- Prioritize the broken down problems
 - Specify the Point of Occurrence
- Consider 4W :What? Where? When? Who?

COURSE SUMMARY



Step 3: Target Setting

- Make a commitment
- Set measurable, concrete, and challenging targets

Step 4: Root Cause Analysis

- Examine the Point of Occurrence
- Specify the root cause

COURSE SUMMARY

Step 5: Develop Countermeasure

- Select the highest value-added countermeasures
- Build consensus with others
- Create a clear and concrete action-plan

There are 2 kind of countermeasures:

Temporary Countermeasure

- Solve existing undesirable situation, or bring a situation back to its original state for a set period of time
- Can't fix the **root cause**
- May **not last for long**

Countermeasure

- Will **eliminate** the root cause
- **Solve** the prioritized problem at the point of occurrence
- Prevent the problem from **reoccurrence**

TASK

Halojek is a company with 2 main product:

- Halofood : a service to provide food delivery
- Haloride : a service to provide mobility using motor cycles

The revenue on this quarter (Q3, 23) is decreased for 50% compared to same quarter last year (Q3, 22)

The cost for this quarter (Q3, 23) is as follow:

- Marketing cost : IDR 800 Bio,
- Labor cost: IDR 750 Bio
- Infrastructure cost: IDR 3,500 Bio

The cost for last quarter (Q3, 22) is as follow:

- Marketing cost : IDR 600 Bio,
- Labor cost: IDR 600 Bio
- Infrastructure cost: IDR 2000 Bio

Objective

Develop problem solving framework for the case above, you are a data analyst to deep dive the root cause of cost increasing on this quarter and also provide the alternative

TASK - STEP 1: CLARIFY THE PROBLEM

Clue: Quantified the gap

Ideal Condition:

The cost for last quarter (Q3, 22) is:

- Marketing cost : IDR 600 Bio,
- Labor cost: IDR 600 Bio
- Infrastructure cost: IDR 2000 Bio

Total: IDR 3200 Bio



Gap:
IDR 1850 Bio

Existing Condition:

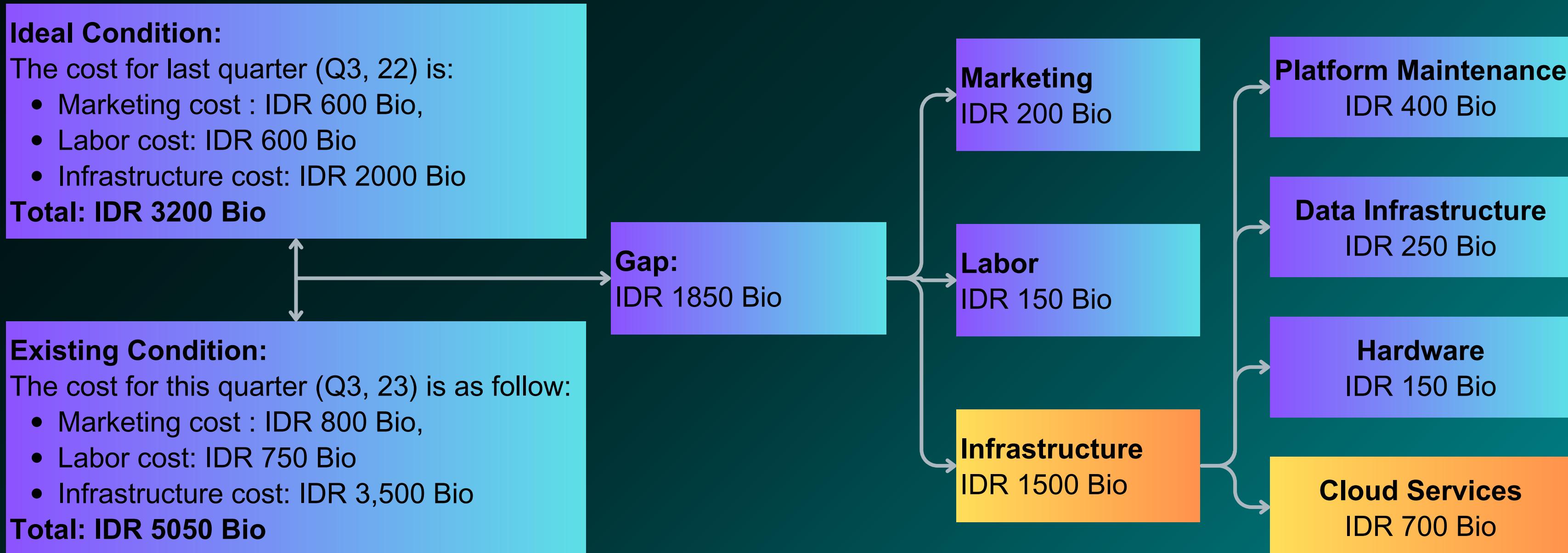
The cost for this quarter (Q3, 23) is as follow:

- Marketing cost : IDR 800 Bio,
- Labor cost: IDR 750 Bio
- Infrastructure cost: IDR 3,500 Bio

Total: IDR 5050 Bio

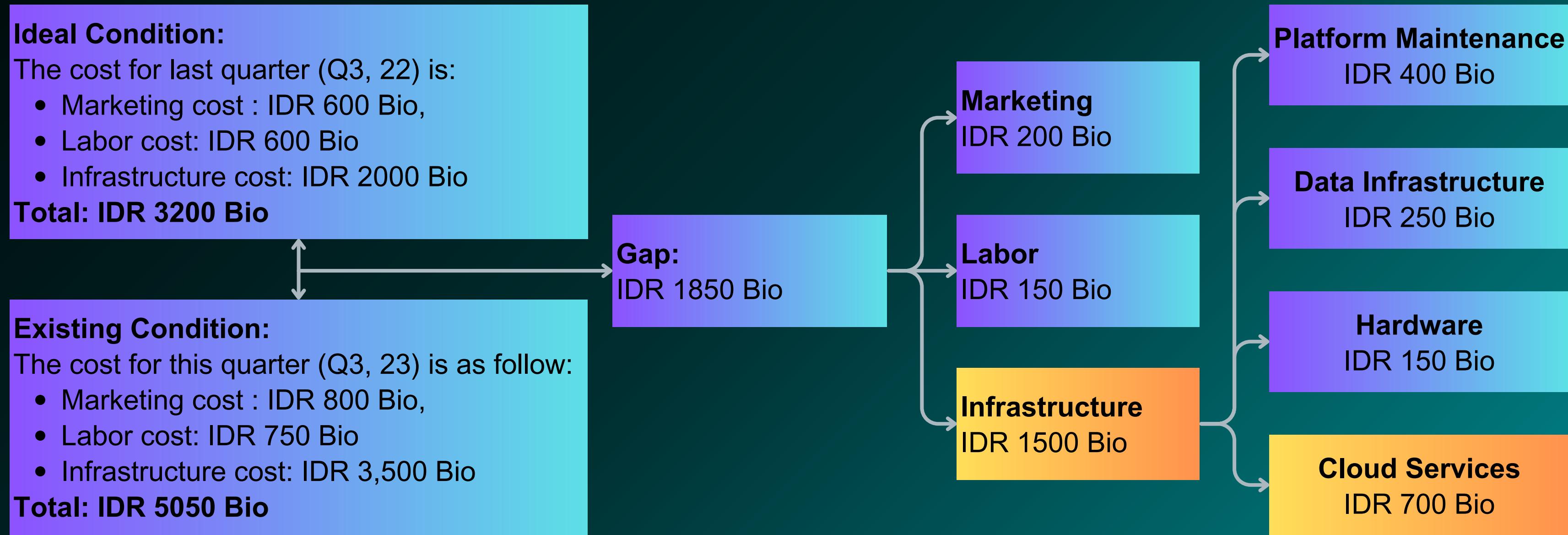
TASK - STEP 2: BREAKDOWN THE PROBLEM

Clue: Should be data driven | 4W - What, Where, Who, When



TASK - STEP 3: SET THE OBJECTIVE

Clue: Objective should be quantified and time-bound



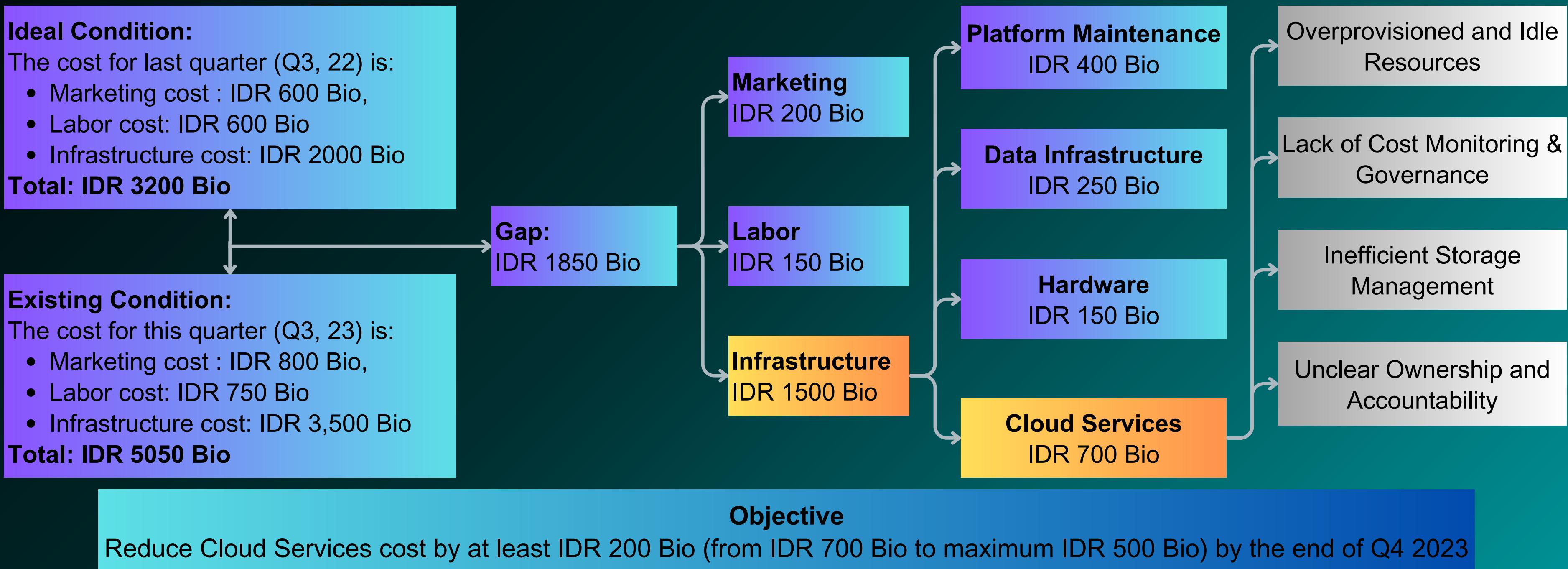
Objective

Reduce Cloud Services cost by at least IDR 200 Bio (from IDR 700 Bio to maximum IDR 500 Bio) by the end of Q4 2023

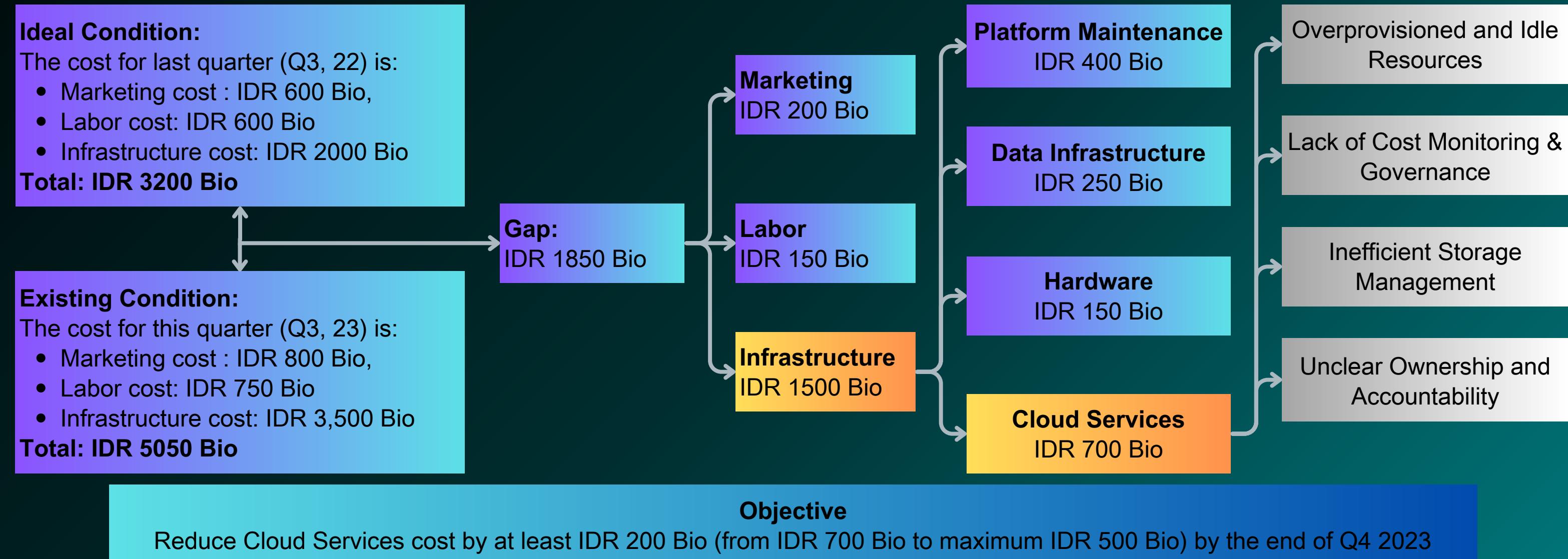
TASK - STEP 4: DEFINE THE ROOT CAUSE

Clue:

- Why, why, why
- Brainstorm is allowed



TASK - STEP 5: DEVELOP COUNTERMEASURES



Problem	Alternative Countermeasure	Feasibility	Cost	Security	Conclusion
Idle and overprovisioned cloud resources	Implement auto-scaling and schedule auto-shutdown for non-production instances	✓	✓	✓	●
No visibility on cloud spending	Build real-time cloud cost dashboard with alert system	✓	✓	✓	●
Inefficient storage usage	Migrate cold data to low-cost storage and set auto-deletion rule	✓	✓	✓	●
No ownership of cloud cost	Assign cloud cost owner per product team and implement cost accountability report	✓	✓	✓	●

THANK YOU



FELICIA EMMA ANGELINE



FELICE.SITORUS@GMAIL.COM



[HTTPS://GITHUB.COM/FELICIAEMMAA/FELICIA-S-PORTOFOLIO.GIT](https://github.com/FeliciaEmmaa/Felicia-S-Portofolio.git)