

Briyani



Round 1 – what can you tell me about this plate?



See

What do you notice just by looking at it?



Measure

What could you put a number to?



Ask

What would you have to ask the cook?

Here is what you called out

In a few seconds, one photograph gave us all of this:

Rice

**6 chicken
pieces**

Non-veg

Cooked at 1 pm

Very spicy

₹180 a plate

Hyderabadi

Smells great

**Every one of these is a fact about the same plate —
but they are not all the same kind of fact. Look again.**

Sort them – what kind of fact is each?

What you said

- “6 chicken pieces”
- “Non-veg”, “Hyderabadi”
- “Very spicy”
- “Cooked at 1 pm”
- The photograph itself

What kind of fact it is

- A number you can count or measure
- A name from a fixed set of choices
- A name that also has an order
- A moment on the clock or calendar
- A picture – not a number or a word

Five different kinds of answer from one plate. These are the types of data.

Round 2 – now put a number on it

This time I will only accept numbers. How would you answer these?

How many grams of rice?

How many pieces of chicken?

What did the plate cost?

How many people will it feed?

How hot is it, in °C?

How many calories?

What do all six answers have in common?

Answer — that was quantitative data

COUNT IT

6 pieces of chicken • 2 eggs

Serves 4 people

Discrete — whole numbers only. There is no such thing as $2\frac{1}{2}$ people.

MEASURE IT

480 g of rice • 78 °C when served

₹180 a plate • 620 calories

Continuous — any value in a range. A finer scale just gives more decimal places.

Quantitative = numbers you can do arithmetic with.

Round 3 – answer without using a number

Veg or non-veg?

- Veg
- Non-veg
- Egg only

Which style?

- Hyderabadi
- Ambur
- Lucknowi
- Kolkata

How spicy?

- Mild
- Medium
- Hot
- Extra hot

Not one of these answers is a number. So what are they?

Answer – that was categorical data

Your answer

“Non-veg”



What kind of category it is

Nominal — just a name, no order

“Hyderabadi”



Nominal — one style among many

“Medium spicy”



Ordinal — mild < medium < hot

Categorical = labels that sort data into groups.

Round 4 – when?

- When was this biryani cooked?
- How long did it sit on the dum?
- By when must it be eaten?
- What time does the canteen start serving?
- How old is this recipe?

These are not amounts, and they are not labels. So what are they?

Answer – that was temporal data

Your answer

“Cooked at 1:15 pm”



What kind of time value it is

A point in time — a timestamp

“90 minutes on the dum”



A duration — a length of time

“Eat before 9 pm”



A deadline — the end of an interval

Temporal = data whose value is a moment, or a span, of time.

Round 5 – the canteen's week

BIRYANI PLATES SOLD

Monday — 40
Tuesday — 38
Wednesday — 42
Thursday — 45
Friday — 96
Saturday — 88
Sunday — 35

What do you notice?

- Which two days stand out — and can you guess why?
- Would these numbers still mean the same thing if I shuffled the days into a random order?
- Could you have a sensible guess at next Friday?

These are numbers again — so why do they feel different from “480 g of rice”?

Answer – that was time-series data

What you noticed

Friday and Saturday spike

Shuffling the days ruins it

You could guess next Friday

What we call it

Seasonality — a repeating pattern

Order is part of the data itself

Forecasting — using the past ahead

Time-series = one measure, recorded again and again, in time order.

Round 6 – what kind of data is this?



Not a number. Not a label. Not a time. So what is a photograph?

Answer – that was multi-modal data

Photo

Recipe

Review

Sizzle

78 °C

₹180

Non-veg

1:15 pm

Image

Text · audio · values

One plate, recorded in many different forms at once — a picture, a written recipe, a spoken review, a temperature reading, and every number, label and time you already found.

Multi-modal = several different forms of data describing the same thing.

Your turn – name the type

Five facts about the same plate. Which type is each one? You have 30 seconds.



The five facts

1. 480 g of rice
2. Hyderabadi
3. Cooked at 1:15 pm
4. 40, 38, 42, 45, 96, 88, 35
5. The photograph itself

Choose from

- Quantitative
- Categorical
- Temporal
- Time-series
- Multi-modal

No peeking at the next slide!

Answers – how did you do?

The fact

1. 480 g of rice
2. Hyderabadi
3. Cooked at 1:15 pm
4. 40, 38, 42, 45, 96, 88, 35
5. The photograph itself

The type

1. Quantitative – a measured amount
2. Categorical – a name, no order
3. Temporal – a point in time
4. Time-series – ordered over days
5. Multi-modal – an image, not a value

One plate. Five different types of data.

One plate, five types of data

Everything you worked out today, from a single photograph of biryani.

The five types

- **Quantitative** — 480 g of rice
- **Categorical** — Hyderabadi, non-veg
- **Temporal** — cooked at 1:15 pm
- **Time-series** — plates sold each day
- **Multi-modal** — photo, recipe, review

Why the type matters

- It decides which sums make sense
- It decides which chart to draw
- It decides how to store it
- It decides how a model can use it

Ask the type first — the right summary, chart and method follow.

From one plate to any dataset

You found all five types in a plate of biryani. The same five turn up in every dataset you will ever meet.

- **Quantitative** — sales figures, exam marks, sensor readings
- **Categorical** — regions, product lines, diagnosis codes
- **Temporal** — order timestamps, admission and discharge dates
- **Time-series** — daily demand, hourly traffic, monthly rainfall
- **Multi-modal** — patient records, self-driving cars, social posts

Next: each type in detail — what it is, how to summarise it, how to chart it.

Multi-modal Data

CHALLENGE

MULTI-MODAL DATA HUNT! 🔍

Look at the post on the phone.

Identify the different types of data you can find.


You have
45 seconds!

List as many
types of data
as you can.


Think beyond
the obvious!



What kind of data is present in this single post?



1. Photo

What type of data is this? _____



2. Caption

What type of data is this? _____



3. Number of Likes

What type of data is this? _____



4. Posting Time

What type of data is this? _____



5. Location Tag

What type of data is this? _____



Hint: Think about the nature of each piece of information.
Is it visual, textual, numerical, time-related, categorical, etc.?



Bonus (if you can spot more!)

Can you find any other hidden types of data in this post?

List them:



GOAL: See how one post contains many kinds of data — this is **MULTI-MODAL DATA!**

Multi-modal Data

MULTI-MODAL DATA EXAMPLE

One social media post, many types of data



PHOTO → Image Data

The picture of the mountains, coffee and sunrise.



CAPTION → Text Data

"Morning views, hot coffee and cool breeze. There's no place like this!"



NUMBER OF LIKES → Quantitative

1,248 likes (a numerical value representing engagement).



POSTING TIME → Temporal

May 24, 2025 at 7:15 AM
(date and time of the post).



LOCATION TAG → Categorical

Coorg, Karnataka, India
(a place/category).

WHAT IS THIS?

All these different types of data together describe the same social media post.

This is MULTI-MODAL DATA

Different forms,
one meaning.



Multi-modal data combines multiple types (image, text, numbers, time, categories, audio, video, etc.) to give a complete understanding of the same event or object.

Types of Data

Dr. Panchatcharam M

Associate Professor, Department of Mathematics and Statistics, IIT Tirupati



Types of Data

Five families of data

Why data types matter

- **Storage & structure** — the type decides how data is recorded: numbers, labels, timestamps or files
- **Valid operations** — you can average heights, but the “average postcode” is meaningless
- **Summaries** — means and medians suit quantities; counts and modes suit categories
- **Visualisation** — histograms for quantities, bar charts for categories, line charts for time series
- **Analysis & ML** — models need each type encoded appropriately before they can learn from it

The five families at a glance



Quantitative

Numbers you can measure or count — height, price, marks



Categorical

Labels that sort data into groups — colour, blood type



Temporal

Date and time values — timestamps, durations



Time-series

One measure recorded repeatedly over time, in order



Multi-modal

Several forms combined — text, images, audio, video

Quantitative data

What it is

Data recorded as numbers that represent a measured or counted quantity. The numbers carry arithmetic meaning – you can add, subtract and average them.

Ask yourself

“Does it make sense to do maths with this number?” If yes, it is quantitative.

Everyday examples

- Height of a pupil in centimetres
- Price of a train ticket in pounds
- Number of goals scored in a match
- Time taken to run 100 metres
- Daily rainfall in millimetres

Not quantitative

- Postcode or phone number – digits, but no arithmetic meaning

Discrete vs continuous

Discrete

Countable values with gaps between them — usually whole numbers. You can list the possible answers.

- Number of siblings: 0, 1, 2, 3...
- Cars passing a junction per hour
- Goals in a football match

There is no such thing as 2.4 siblings.

Continuous

Values that can take any point on a scale. Between any two values there is always another — precision is limited only by the instrument.

- Height: 162.4 cm, 162.43 cm...
- Temperature of a room
- Time to finish a race

Continuous data is usually rounded when recorded.

Interval vs ratio scales

Interval

Equal gaps between values, but zero is an arbitrary reference point rather than “nothing”.

- Temperature in $^{\circ}\text{C}$
- Calendar years
- IQ scores

20°C is not “twice as hot” as 10°C — ratios are meaningless.

Ratio

Equal gaps and a true zero that means the complete absence of the quantity.

- Height, mass, length
- Money in an account
- Duration in seconds

6 kg really is twice 3 kg — ratios are meaningful.

Summary statistics: a worked example

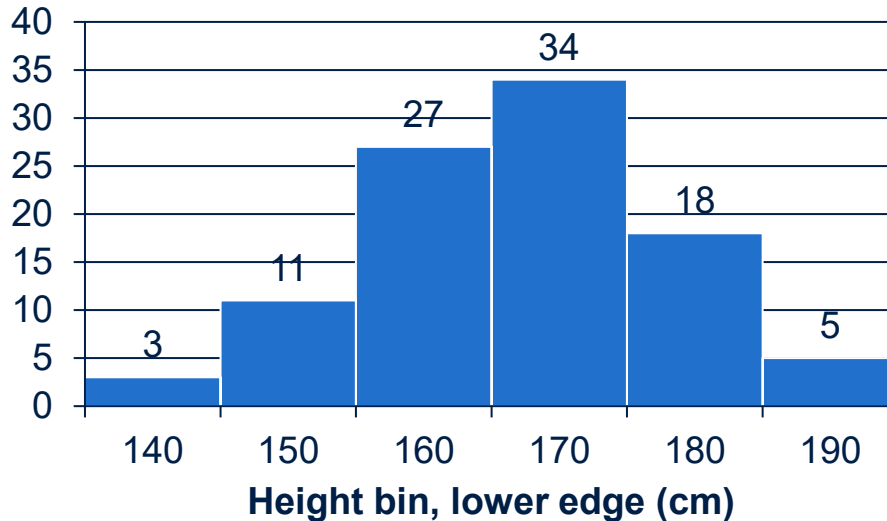
Age of each student:

$n = ?$, total = ?

Statistic	How it is worked out	Value
Mean		
Median		
Mode		
Range		
Interquartile range		
Standard deviation		

Distributions and histograms

Heights of 98 pupils (cm)



Reading a histogram

- Values are grouped into bins; bar height is the count in each bin
- Bars touch, because the scale is continuous
- The shape shows the centre, the spread and any skew
- Bin width matters: too wide hides detail, too narrow looks like noise

Outliers

What is an outlier?

A value that sits far away from the rest of the data. A common rule flags anything more than $1.5 \times \text{IQR}$ beyond the upper or lower quartile.

Where they come from

- Measurement or typing errors
- A different population mixed in
- Genuinely rare but real events

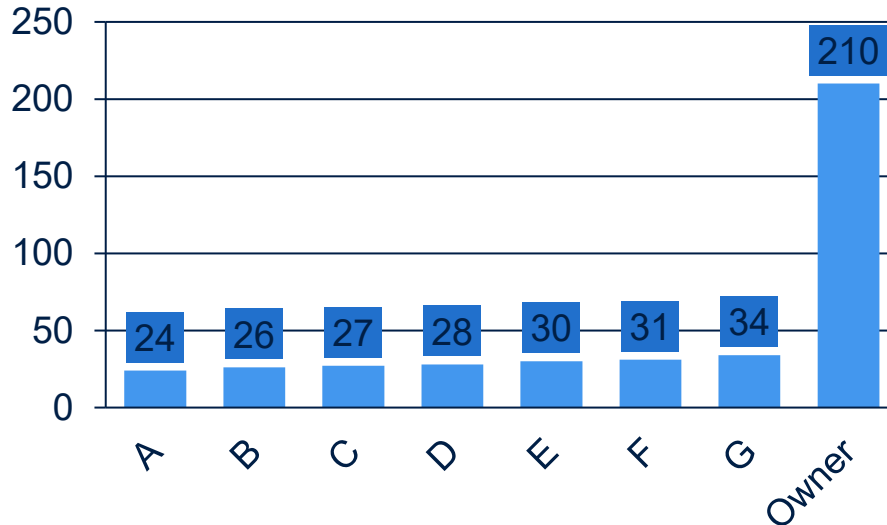
What to do about them

- Investigate before deleting – an outlier is evidence, not litter
- Correct it if it is clearly an error, such as a height of 1,750 cm
- Keep it and report the median instead of the mean if it is genuine
- Say what you did – removing points silently is bad practice

Outliers often turn out to be the most interesting part of the dataset.

Worked example: mean vs median

Salaries at a small firm (£000s)



The same data, two answers

- **Mean = £41,250**
- **Median = £29,000**

Seven of the eight staff earn less than the mean. One very large value has dragged it upwards, so the median describes the typical employee far better.

When data is skewed, prefer the median.

Common mistakes with quantitative data

- **Treating ID numbers as quantities** — the mean of a set of student numbers means nothing.
- **Quoting the mean for skewed data** — one extreme value can make the average unrepresentative.
- **Ignoring units** — mixing metres and centimetres in one column silently corrupts every result.
- **Over-precision** — reporting 23.847619 kg from a scale that reads to the nearest 100 g.
- **Truncating a bar chart axis** — starting the value axis above zero exaggerates small differences.
- **Deleting outliers automatically** — check whether the value is an error before you remove it.

Categorical data

What it is

Data that places each observation into one of a limited set of groups or labels. The value names a category rather than measuring a quantity.

Ask yourself

"Am I sorting things into boxes?" If yes, it is categorical – even if the label happens to be a number.

Everyday examples

- Eye colour: blue, brown, green
- Blood type: A, B, AB, O
- Country of birth
- T-shirt size: S, M, L, XL
- Survey answer: agree, neutral, disagree

The number trap

A "1 = male, 2 = female" coding is still categorical. The digits are shorthand, not amounts.

Nominal vs ordinal

Nominal

Categories with no natural order. No label is “higher” than another — they are simply different.

- Eye colour
- Language spoken at home
- Type of transport to school

You may count them and find the mode, but sorting them is arbitrary.

Ordinal

Categories with a meaningful order, but the gaps between them are not equal or measurable.

- Exam grade: A, B, C, D
- Satisfaction: poor → excellent
- Race position: 1st, 2nd, 3rd

You may rank them and quote a median, but the difference between 1st and 2nd is not the same as 2nd and 3rd.

Binary and other special cases

Binary data

A category with exactly two possible values — the simplest categorical variable there is.

- Yes / no
- Pass / fail
- Present / absent

Often stored as 0 and 1, which makes it convenient for calculations: the mean of a 0/1 column is the proportion of 1s.

Watch out for these

- High cardinality — thousands of distinct categories, such as postcodes, are hard to chart or model
- Multi-label — one record belongs to several categories at once, such as film genres
- Free text — open-ended answers must be grouped before they behave like categories
- Missing or “other” — decide deliberately whether these form their own category

Encoding categories for analysis

Most software needs numbers, so categories have to be converted – carefully.

- **One-hot encoding** – one new 0/1 column per category. Safe for nominal data; use it by default.
- **Ordinal encoding** – map ordered labels to ranks, e.g. poor=1 ... excellent=5. Only for ordinal data.
- **Label encoding** – assign an arbitrary integer per category. Compact, but implies a false order.
- **Grouping rare levels** – collapse infrequent categories into “other” to keep the table manageable.

The golden rule: never invent an order that the data does not have.

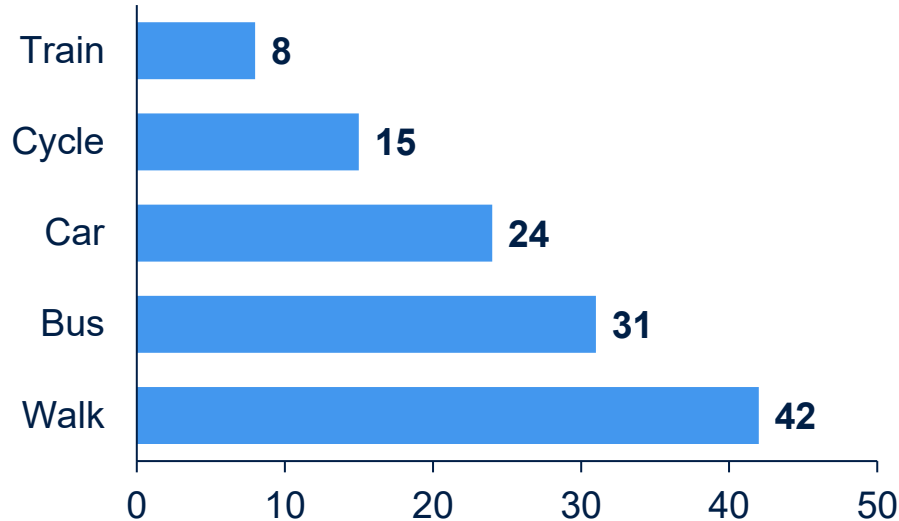
Summarising categorical data

- **Frequency** — how many records fall into each category. The starting point for everything else.
- **Proportion or percentage** — frequency divided by the total. Makes groups of different sizes comparable.
- **Mode** — the most common category. The only average that always works for nominal data.
- **Median** — usable for ordinal data once the categories are placed in order.
- **Cross-tabulation** — counts for two categorical variables at once, to look for association.

Always quote the total alongside a percentage — “60%” means very different things out of 5 and out of 5,000.

Charting categories

How 120 pupils travel to school

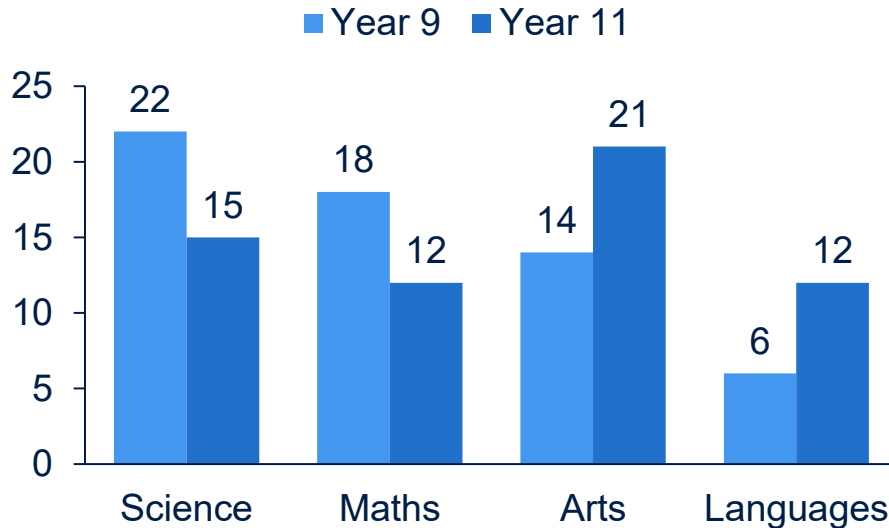


Charting rules of thumb

- Use a bar chart, not a histogram — leave gaps between the bars
- Sort nominal categories by size; keep ordinal ones in their natural order
- Start the value axis at zero
- Horizontal bars suit long category labels
- Avoid pie charts beyond about five slices

Worked example: cross-tabulation

Favourite subject by year group



Two categories at once

Here favourite subject is crossed with year group. Each bar is a count of pupils in one combination.

- Science leads in Year 9; Arts leads in Year 11
- Compare percentages, not raw counts, when the groups differ in size
- A pattern in the table suggests association, not cause

Common mistakes with categorical data

- **Averaging category codes** — if red=1, blue=2 and green=3, a mean of 2.1 is nonsense.
- **Averaging ordinal ranks carelessly** — “mean satisfaction of 3.7” assumes equal gaps that may not exist.
- **Inconsistent labels** — “UK”, “U.K.” and “united kingdom” become three separate categories.
- **Too many slices** — a pie chart with fifteen categories communicates nothing.
- **Hiding the base** — percentages without the underlying counts can mislead badly.
- **Dropping “missing” silently** — non-response is information; record it as its own category.

Temporal data

What it is

Data that records when something happened, or how long it lasted. The value is a point in time or a length of time.

Why it is its own family

Time looks numeric but behaves oddly: 60 seconds roll into a minute, months differ in length, and years occasionally gain a day.

Everyday examples

- Date of birth
- Timestamp of a card payment
- Length of a phone call
- Opening hours of a shop
- Delivery window: 09:00–12:00

Key idea

Temporal data is about individual moments. Once you record the same measure repeatedly over time, you have a time series.

Four kinds of temporal value

- **Date** — a single day with no time attached, such as 14 March 2026. Used for birthdays and deadlines.
- **Timestamp** — a precise moment, such as 14 March 2026 at 09:41:07. Used for transactions and log entries.
- **Duration** — a length of time with no fixed start, such as 90 minutes. Behaves like a quantity and can be averaged.
- **Interval** — a period with a definite start and end, such as 09:00 to 12:00. Two intervals can overlap.

Subtracting two timestamps gives a duration; adding a duration to a timestamp gives another timestamp.

Formats and the ISO 8601 standard

The problem

Is 03/04/2026 the third of April or the fourth of March? It depends who wrote it. Ambiguous date formats are a leading cause of corrupted datasets.

- UK style: day / month / year
- US style: month / day / year
- Spreadsheets often reformat dates on import without warning

The solution: ISO 8601

Write dates largest unit first, with a four-digit year.

- Date: 2026-04-03
- Timestamp: 2026-04-03T09:41:07Z
- Duration: PT90M

Why it works

It is unambiguous worldwide, and text sorting puts the dates in true chronological order.

Time zones and daylight saving

Where it goes wrong

- The same instant has different clock readings in London, New York and Tokyo
- Clocks go forward in spring: one hour of the day simply does not exist
- Clocks go back in autumn: one hour repeats, so 01:30 happens twice
- Comparing events across regions without zones gives wrong orderings

Good practice

- Store timestamps in UTC and convert only for display
- Keep the original time zone as a separate field if local time matters
- Never assume the server clock and the user clock agree
- Test any date calculation across a clock-change weekend

Rule: record the instant unambiguously first, present it conveniently second.

Granularity and rounding

Choosing a grain

Granularity is how finely time is recorded: to the second, the minute, the day, the month.

- Too fine: huge files and noisy charts
- Too coarse: real patterns disappear

Pick the grain that matches the question you are asking, not the grain the sensor happens to offer.

Rolling up safely

- You can always aggregate upwards — seconds into hours, days into months
- You can never recover detail you did not record
- Say how you rounded: does “14:59” belong to the 14:00 hour or the 15:00 hour?
- Beware part-periods: the current month is incomplete and will look artificially low

Store fine, report coarse.

Turning time into useful features

A raw timestamp is rarely useful on its own. Split it into parts that carry meaning.

- **Calendar parts** — year, month, day, hour. Simple and often surprisingly powerful.
- **Day of week** — separates weekday behaviour from weekend behaviour.
- **Is it a holiday?** — a 0/1 flag that explains many otherwise strange spikes and dips.
- **Cyclical encoding** — hour 23 and hour 0 are adjacent, so encode the hour on a circle rather than a straight line.
- **Elapsed time** — days since sign-up, minutes since the last event. Turns a moment into a quantity.

Good time features often matter more than a clever model.

Common mistakes with temporal data

- **Storing dates as text** — “12 March” sorts alphabetically, not chronologically, and cannot be subtracted.
- **Ambiguous formats** — mixing UK and US ordering silently swaps days and months for the first twelve of each month.
- **Ignoring time zones** — events end up in the wrong order, or on the wrong day entirely.
- **Assuming every day has 24 hours** — clock changes break naive arithmetic twice a year.
- **Two-digit years** — “26” could be 1926 or 2026; always store four digits.
- **Comparing an incomplete period** — this month so far will always look smaller than last month in full.

Check your understanding

Name the type

1. Shoe size: 5, 6, 7, 8
2. Postcode of a customer
3. Time a parcel was scanned
4. Rating: poor, fair, good, excellent
5. Litres of fuel in a tank
6. Length of a hospital stay

Answers

1. Ordinal categorical – ordered labels, unequal gaps
2. Nominal categorical – digits with no arithmetic meaning
3. Temporal – a timestamp
4. Ordinal categorical
5. Quantitative, continuous, ratio scale
6. Temporal duration, which behaves like a quantity

Time-series data

What it is

The same measurement taken repeatedly over time, kept in order. Each observation is a pair: a moment and a value.

The defining feature

Order matters. Shuffle the rows of a time series and you destroy the information; shuffle the rows of an ordinary table and nothing is lost.

Everyday examples

- Daily maximum temperature
- Monthly electricity usage
- A share price recorded every minute
- Steps counted each hour by a watch
- Weekly supermarket sales

Key idea

Nearby observations are related. Today's temperature tells you a lot about tomorrow's.

Temporal vs time-series

Temporal

Time is an attribute of each record.
Records are independent of one another.

- A list of customer birthdays
- Timestamps of support tickets
- Booking windows for a clinic

Sorting by time is convenient, but the rows do not depend on each other.

Time-series

Time is the index. Each record is one step in an ordered sequence of the same measurement.

- Temperature at noon each day
- Website visits per hour
- Monthly rainfall totals

Gaps, ordering and spacing all carry meaning – a missing Tuesday is itself information.

Trend, seasonality and noise

Almost every time series can be read as three ingredients layered on top of one another.

- **Trend** — the long-run direction. Is the series drifting up, down or staying flat over months and years?
- **Seasonality** — a pattern that repeats on a fixed cycle: busier at weekends, warmer in summer, higher every December.
- **Cycles** — longer swings with no fixed length, such as an economic boom and slump.
- **Noise** — the irregular leftover once trend and seasonality are removed. Real, but not predictable.

Separating these three is called decomposition, and it is usually the first thing to try.

Sampling, gaps and irregular series

Regular series

Observations arrive at fixed intervals — every hour, every day, every month. Most standard methods assume this.

Irregular series

Observations arrive whenever something happens: a payment, a click, an alarm. The spacing itself is data.

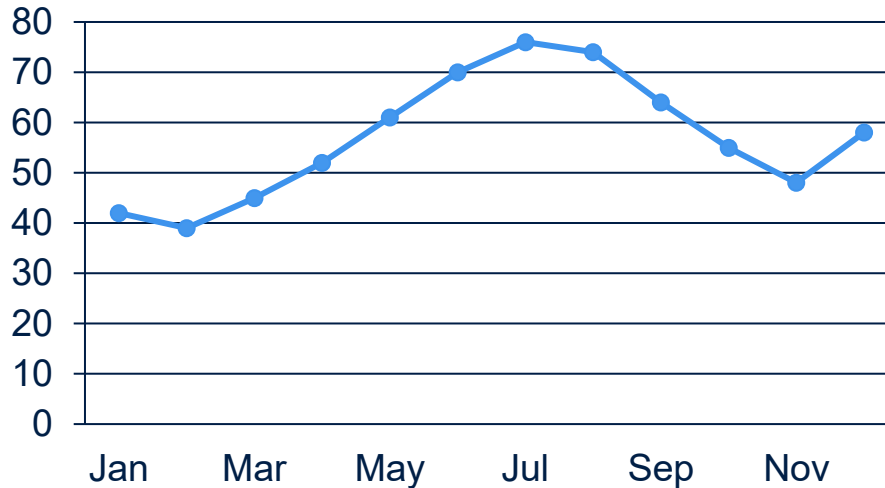
Handling gaps

- Decide whether a missing value means “not recorded” or “genuinely zero” — they are very different
- Resample irregular data onto a regular grid before modelling
- Interpolate cautiously, and flag which points were filled in
- Never simply delete rows — that silently changes the spacing

Sampling too slowly can hide a repeating pattern completely.

Reading a line chart

Ice cream sales per month
(thousands)



What to look for

- **Shape:** sales rise into summer and fall away again — clear seasonality.
- **Peak:** July, as expected for ice cream.
- **Surprise:** December ticks up. Worth asking why before assuming it is an error.
- **Connect the dots:** lines are right here because the order is real. Never join up categories this way.

Lags, smoothing and autocorrelation

Lags

A lag is simply the value from k steps earlier. Lag-1 is yesterday, lag-7 is the same day last week.

Autocorrelation

How strongly the series agrees with its own past. High lag-7 correlation is a fingerprint of a weekly pattern.

This is why time series need their own methods: the observations are not independent.

Smoothing

- A moving average replaces each point with the average of a window around it
- A 7-day window removes the day-of-week wobble and reveals the trend
- Wider windows are smoother but slower to react to real change

Health warning

Smoothing is for seeing the shape, not for reporting the numbers. Always keep the raw series too.

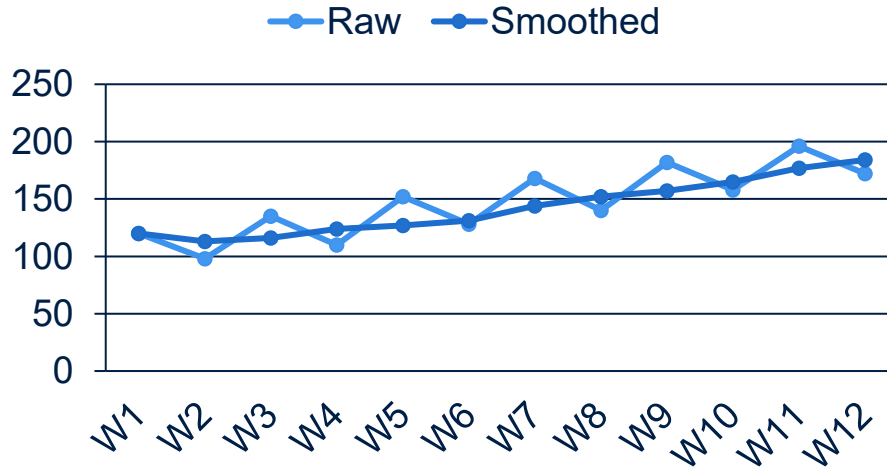
Forecasting in one slide

Forecasting uses the past of a series to estimate its future. Start simple — simple methods are surprisingly hard to beat.

- **Naïve forecast** — tomorrow equals today. Use it as the baseline every other method must beat.
- **Seasonal naïve** — next Monday equals last Monday. Strong when seasonality dominates.
- **Moving average and exponential smoothing** — weight recent observations more heavily.
- **ARIMA and friends** — model the autocorrelation structure explicitly.
- **Always split by time** — train on the earlier part, test on the later part. Random splits leak the future into the past and flatter the model.

Worked example: raw vs smoothed

Weekly orders: raw vs 3-week average



Reading the two lines

- **Raw:** zig-zags every week, which makes the direction hard to judge by eye.
- **Smoothed:** the same data with a 3-week average, showing a steady climb.
- **Decomposed:** trend is upward, the zig-zag is the repeating pattern, and small wobbles are noise.
- **Careful:** quote the raw numbers when reporting. The smoothed line is a reading aid, not the data.

Common mistakes with time series

- **Shuffling the rows** — order is the information. A shuffled time series is no longer a time series.
- **Random train/test splits** — testing on data that came before your training data lets the model see the future.
- **Ignoring seasonality** — “sales fell in January” may just be the annual pattern, not a problem.
- **Comparing unequal periods** — February is shorter; a five-weekend month beats a four-weekend one.
- **Reporting smoothed values as fact** — smoothing is a lens, not a measurement.
- **Confusing correlation with cause** — two series that both trend upward will correlate strongly and may be entirely unrelated.

Multi-modal data

What it is

Data about one thing recorded in several different forms at once. A mode, or modality, is a kind of signal: text, images, audio, video, sensor readings, numbers.

Why bother

Each mode sees something the others miss. Combined, they describe the situation far better than any one alone.

An everyday example

A single social media post can contain:

- A photo – image data
- A caption – text data
- Number of likes – quantitative
- Posting time – temporal
- Location tag – categorical

One record, five different types working together.

The common modalities



Text

Reviews,
notes,
transcripts,
messages



Image

Photographs,
scans, X-rays,
diagrams



Audio

Speech, music,
machine
sounds, alarms



Video

Moving images
plus sound,
over time



Sensor

GPS,
temperature,
accelerometer,
lidar

Why multi-modal data is hard

The challenges

- Alignment — matching the right frame of video to the right line of transcript
- Different sampling rates — audio arrives thousands of times per second, notes arrive once a week
- Different scales — pixels, decibels and words are not comparable numbers
- Missing modes — one patient has a scan, another does not
- Size — video and audio dwarf spreadsheets in storage

How it is handled

- Convert each mode into a numeric representation, called an embedding
- Line the modes up on a shared timeline or shared identifier
- Combine them — either early, as one joint input, or late, by merging separate predictions

Rule of thumb

Get one mode working well before adding a second. Extra modes add cost as well as insight.

Example: a patient record

What the record contains

- **Quantitative** — blood pressure, weight, test results
- **Categorical** — blood group, diagnosis code, ward
- **Temporal** — admission and discharge timestamps
- **Time-series** — heart rate monitored every second
- **Image** — X-rays and MRI scans
- **Text** — the clinician's free-text notes

Why combining helps

A scan alone shows a shadow. The notes explain the symptoms. The vital-signs series shows whether the patient is improving. Together they support a diagnosis that no single mode could justify.

The catch

Notes are unstructured, scans are enormous, and the monitor stops recording when a lead falls off. Real multi-modal data is messy in several ways at once.

Example: a self-driving car

Signals arriving at once

- **Cameras** — read road signs, lane markings and traffic lights
- **Lidar and radar** — measure distance and speed of nearby objects
- **GPS** — gives position on a map
- **Wheel and motion sensors** — report the car's own movement
- **Microphones** — detect sirens the cameras cannot see

Why one mode is not enough

A camera struggles in fog and glare. Radar sees through fog but gives a coarse picture. GPS drifts under bridges. Each mode fails in different conditions, so the combination is far safer than any one alone.

The hard part

Every signal must be stamped with the same clock and merged in a few milliseconds. Alignment is the whole problem.

Multi-modal models at a glance

Modern systems handle several modes at once. The recipe is broadly the same in each case.

- **Encode** — turn each mode into a list of numbers that captures its meaning.
- **Align** — place those lists in a shared space so that a photo of a dog and the word “dog” land close together.
- **Fuse** — combine the aligned representations into a single view of the record.
- **Predict** — use the fused view to answer a question, describe an image or generate a reply.

Everyday examples: image search from a written description, automatic captions and subtitles, and assistants that read a photo and answer questions about it.

Check your understanding

Questions

1. A shop records daily takings for two years. Which type?
2. Why can a line chart mislead when the x-axis holds categories?
3. A podcast episode has audio, a transcript and a play count. How many modes?
4. Give one reason to store timestamps in UTC.
5. Why is a random train/test split wrong for a time series?

Answers

1. Time-series — one measure, repeated, in order
2. Lines imply the gaps between points are meaningful; between categories they are not
3. Three: audio, text and a quantitative count
4. It removes ambiguity from time zones and clock changes, so events sort correctly
5. It lets the model train on data that came after the test data, which cannot happen in reality

Choosing the right chart

Quantitative

- Histogram — shape of one variable
- Scatter plot — two variables together

Categorical

- Bar chart — compare counts
- Stacked bar — parts of a whole

Temporal

- Timeline — events in order
- Heat map — activity by day

Time-series

- Line chart — the default choice
- Seasonal plot — compare cycles

Multi-modal

- Small multiples — one panel per mode
- Dashboard — modes on a shared clock

Always

- Label the axes and give the units
- Start bar charts at zero

Converting between types

Data can be moved from one family to another. Every move trades detail for convenience.

- **Quantitative** → **categorical** — group ages into bands. Easier to summarise, but the exact values are gone.
- **Categorical** → **quantitative** — count how many fall in each group. You now have numbers you can chart.
- **Temporal** → **categorical** — turn a timestamp into weekday or month. Useful for spotting weekly patterns.
- **Temporal** → **time-series** — count events per hour or per day to make an evenly spaced series.
- **Multi-modal** → **quantitative** — measure something from the media: image brightness, word count, speech length.

Conversions are almost always one-way. Keep the original data so you can start again.

Recap: the five families

Type	What it is	Example	Typical chart
Quantitative	Numbers with arithmetic meaning	Height, price, goals scored	Histogram, scatter
Categorical	Labels that sort data into groups	Blood type, county, grade	Bar chart
Temporal	Points and spans of time	Booking timestamp, duration	Timeline, heat map
Time-series	One measure repeated, in order	Daily rainfall, hourly demand	Line chart
Multi-modal	Several forms describing one thing	Patient record, social post	Dashboard, small multiples

Glossary

Discrete — countable values with gaps between them

Continuous — any value within a range, limited only by measuring precision

Nominal — categories with no natural order

Ordinal — categories with a meaningful order but uneven gaps

One-hot encoding — one yes/no column per category

Cardinality — how many distinct categories a variable has

Outlier — a value far from the rest of the data

UTC — Coordinated Universal Time, a single global reference clock

ISO 8601 — the international standard date format, YYYY-MM-DD

Granularity — how finely time is recorded: second, hour, day

Trend — the long-run direction of a series

Seasonality — a pattern that repeats on a fixed cycle

Lag — a value from an earlier point in the series

Modality — one form a signal can take: text, image, audio, sensor

Data Detective Challenge

Look at this railway platform. You have **30 seconds**.

Find one example each of:

- ① Quantitative data
- ② Categorical data
- ③ Temporal data
- ④ Time-series data

गाड़ी संख्या TRAIN NO.	गाड़ी का नाम TRAIN NAME	गंतव्य DESTINATION	समय TIME	प्रस्थान/आगमन A/D	पटरी PF
12603	GRAND TRUNK EXPRESS	NEW DELHI	14:35	DEPARTURE	4
22843	HOWRAH SUPERFAST EXP	HOWRAH	14:50	DEPARTURE	3
17230	SABARI EXPRESS	THIRUVANANTHUPURAM	15:10	ARRIVAL	2
16526	KANYAKUMARI EXPRESS	KANYAKUMARI	15:20	DEPARTURE	1
12617	TAMIL NADU EXPRESS	CHENNAI CENTRAL	15:35	ARRIVAL	4

PLEASE STAND BEHIND YELLOW LINE • Eatables / PLASTIC NOT ALLOWED

4
PLATFORM

14:32:18

HELP DESK
139

CLEAN PLATFORM
GREEN PLATFORM

WAY TO EXIT

STAIRS

DRINKING WATER

सुविधाएं उपलब्ध हैं
FACILITIES AVAILABLE

वेटिंग रूम
WAITING ROOM

भोजनालय
FOOD COURT

रिटायरिंग रूम
RETIRING ROOM

शौचालय
TOILET

बुक स्टॉल
BOOK STALL

TRAIN INFORMATION
12603
G T EXPRESS
NEW DELHI
DEPARTURE
14:35
PF 4

Categorical Data

Groups or labels with categories.

- Train Type: Express
- A/D: Arrival / Departure
- Facilities: Waiting Room, Food Court
- Class: Sleeper, AC, General

Time-Series Data

Data changing over time in a sequence.

- Train schedule through the day (14:35, 14:50, 15:10, 15:20, 15:35...)
- Delay updates over time
- Passenger count over time

गाड़ी संख्या TRAIN NO.	गाड़ी का नाम TRAIN NAME	गंतव्य DESTINATION	संमय TIME	प्रस्थान/आगमन A/D	पटरी PF
12603	GRAND TRUNK EXPRESS	NEW DELHI	14:35	DEPARTURE	4
22843	HOWRAH SUPERFAST EXP	HOWRAH	14:50	DEPARTURE	3
17230	SABARI EXPRESS	THIRUVANANTHAPURAM	15:10	ARRIVAL	2
16526	KANYAKUMARI EXPRESS	KANYAKUMARI	15:20	DEPARTURE	1
12617	TAMIL NADU EXPRESS	CHENNAI CENTRAL	15:35	ARRIVAL	4

PLEASE STAND BEHIND YELLOW LINE • EATABLES / PLASTIC NOT ALLOWED

Quantitative Data

Numbers you can count or measure.

- Train No.: 12603
- Platform No.: 4
- Help Desk No.: 139
- Number of coaches (if displayed)

Temporal Data

Specific date and time.

- 14:35, 14:50, 15:10...
- 14:32:18
- 14 AUG 2026

HELP DESK
139

CLEAN
GREEN

STAIRS

DRINKING
WATER

सुविधाएं उपलब्ध हैं
FACILITIES AVAILABLE

वेटिंग रूम
WAITING ROOM

भोजनालय
FOOD COURT

रिटायरिंग रूम
RETIRING ROOM

शौचालय
TOILET

बुक स्टॉल
BOOK STALL

TRAIN INFORMATION

12603
G T EXPRESS
NEW DELHI
DEPARTURE
14:35
PF 4

14:32:18
14 AUG 2026

4
PLATFORM

Data Detective Challenge

Look at this car dashboard. You have **30 seconds**.

Find one example each of:

- ① Quantitative data
- ② Categorical data
- ③ Temporal data
- ④ Time-series data



Quantitative Data

Numbers you can count or measure.

- Speed: 82 km/h
- RPM: 2.1 x1000
- Fuel Range: 542 km
- Avg. Fuel: 18.6 km/l
- Odometer: 12568 km
- Outside Temp: 28°C
- Tire Pressure: 32 psi

Categorical Data

Groups or labels with categories.

- Drive Mode: D
- Fuel Gauge: E / F
- High Beam: ON
- Lane Assist Icon
- Car Ahead Icon

Temporal Data

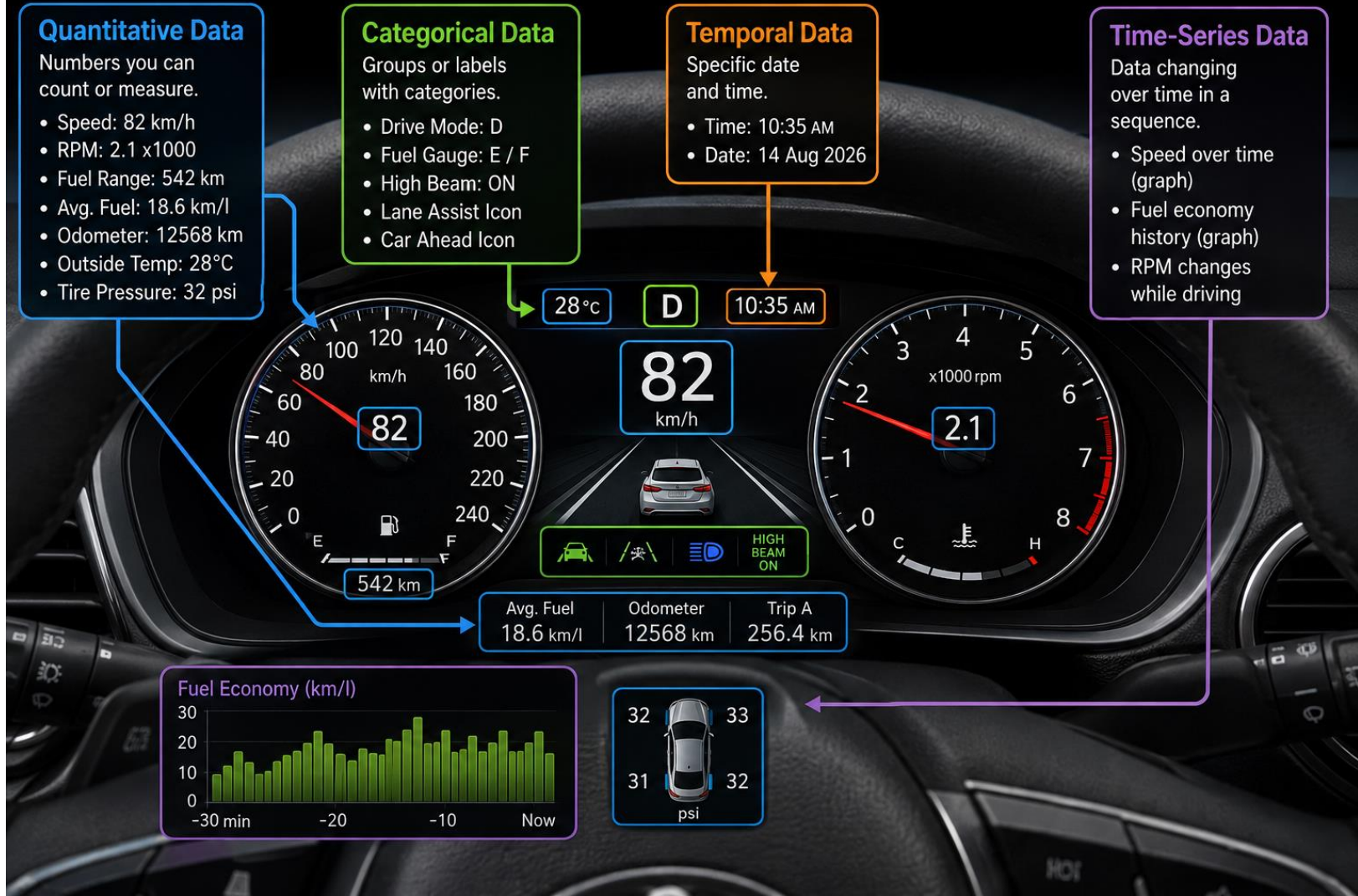
Specific date and time.

- Time: 10:35 AM
- Date: 14 Aug 2026

Time-Series Data

Data changing over time in a sequence.

- Speed over time (graph)
- Fuel economy history (graph)
- RPM changes while driving



Data Detective Challenge

Look at this trading terminal. You have 30 seconds.

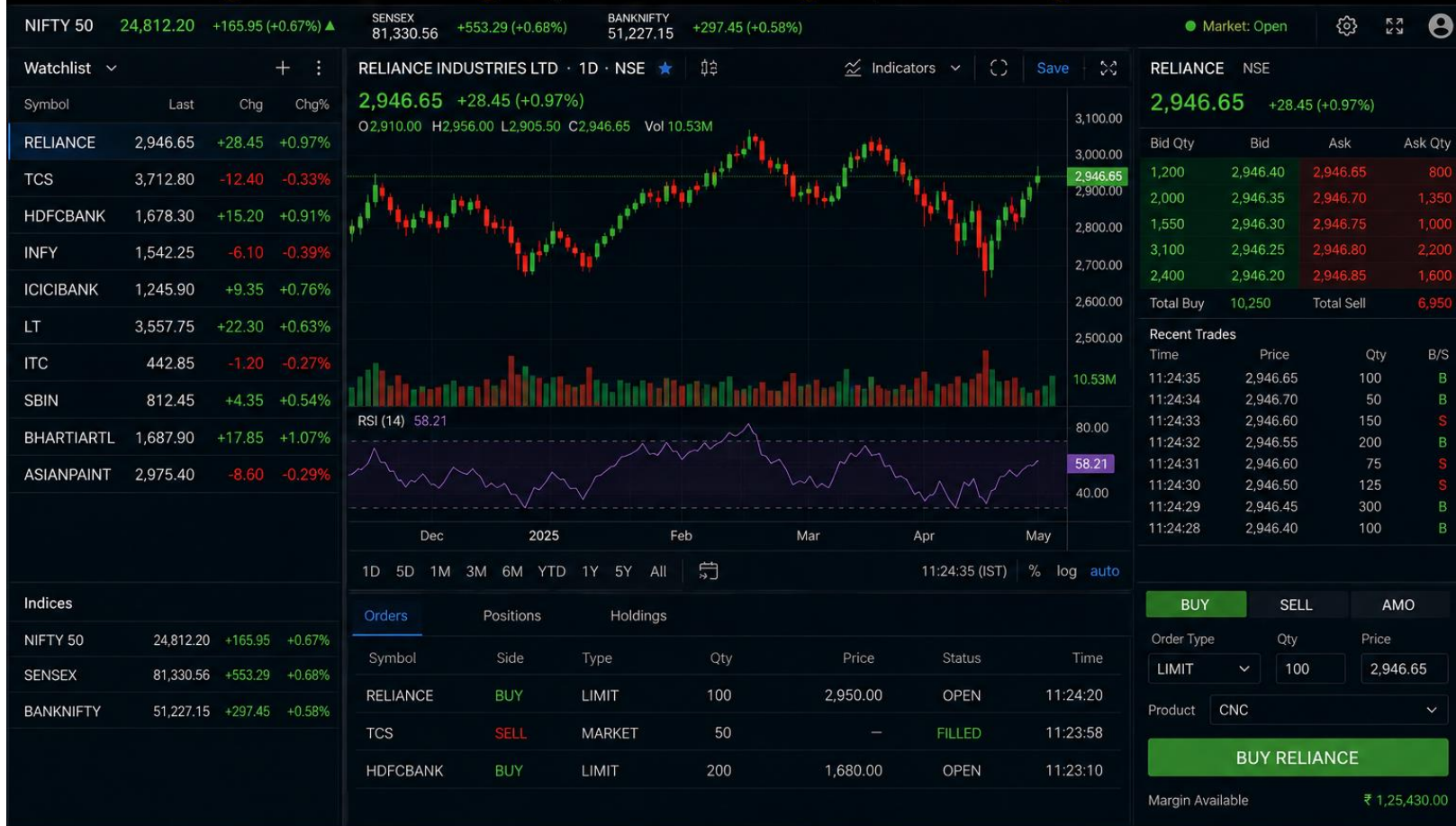
Find one example each of:

① Quantitative data

② Categorical data

③ Temporal data

④ Time-series data



NIFTY 50 24,812.20 +165.95 (+0.67%) ▲

SENSEX
81,330.56 +553.29 (+0.68%)BANKNIFTY
51,227.15 +297.45 (+0.58%)

Watchlist Indices F&O Currency Commodities

Search & add

15/50

Symbol	Last	Chg	Chg%
RELIANCE	2,946.65	+28.45	+0.97%
TCS	3,712.80	-12.40	-0.33%
HDFCBANK	1,678.30	+15.20	+0.91%
INFY	1,542.25	-6.10	-0.39%
ICICIBANK	1,245.90	+9.35	+0.76%
LT	3,557.75	+22.30	+0.63%
ITC	442.85	-1.20	-0.27%
SBIN	812.45	+4.35	+0.54%
BHARTIARTL	1,687.90	+17.85	+1.07%
ASIANPAINT	2,975.40	-8.60	-0.29%

CATEGORICAL DATA

Groups / labels / names

- Symbol (RELIANCE, TCS, ...)
- Index (NIFTY 50, SENSEX)
- Segment (Equity, F&O, ...)

RELIANCE 1m 5m 15m 1H 1D W Indicators

RELIANCE INDUSTRIES LTD · 1D · NSE

O2910.00 H2956.00 L2905.50 C2946.65 +28.45 (+0.97%)



TIME-SERIES DATA

Values recorded over time and shown as a series

- Price candlesticks
- Volume over time
- RSI over time

Volume 20 10.53M 12.81M

RSI 14 58.21

Dec 2025 Feb Mar Apr May

1D 5D 1M 3M 6M YTD 1Y 5Y All

11:24:35 (UTC+5:30) % log auto

QUANTITATIVE DATA

Numbers you can count or measure

- Price: 2,946.65
- Change: +28.45
- Volume: 10.53M
- Index value: 24,812.20

14 May 2025 11:24:35 AM



RELIANCE NSE

2,946.65 +28.45 (+0.97%)

BID QTY	BID	ASK	ASK QTY
1,200	2,946.40	2,946.65	800
2,000	2,946.35	2,946.70	1,350
1,550	2,946.30	2,946.75	1,000
3,100	2,946.25	2,946.80	2,200
2,400	2,946.20	2,946.85	1,600
Total Buy	10,250	Total Sell	6,950

Recent Trades

Time	Price	Qty	B/S
11:24:35	2,946.65	100	B
11:24:34	2,946.70	50	B
11:24:33	2,946.60	150	S
11:24:32	2,946.55	200	B
11:24:31	2,946.60	75	S
11:24:30	2,946.50	125	S
11:24:29	2,946.45	300	B
11:24:28	2,946.40	100	B

TEMPORAL DATA

Date and time stamps

- 14 May 2025
- 11:24:35 AM
- Trade time (11:24:35, 11:24:34, ...)

Order Book Positions Holdings Alerts

RELIANCE NSE 2,946.65 +28.45 (+0.97%)

Buy

Sell

Regular - MIS

Qty.

Price

Product

CNC

Validity

Day

Type Limit

PLACE BUY ORDER

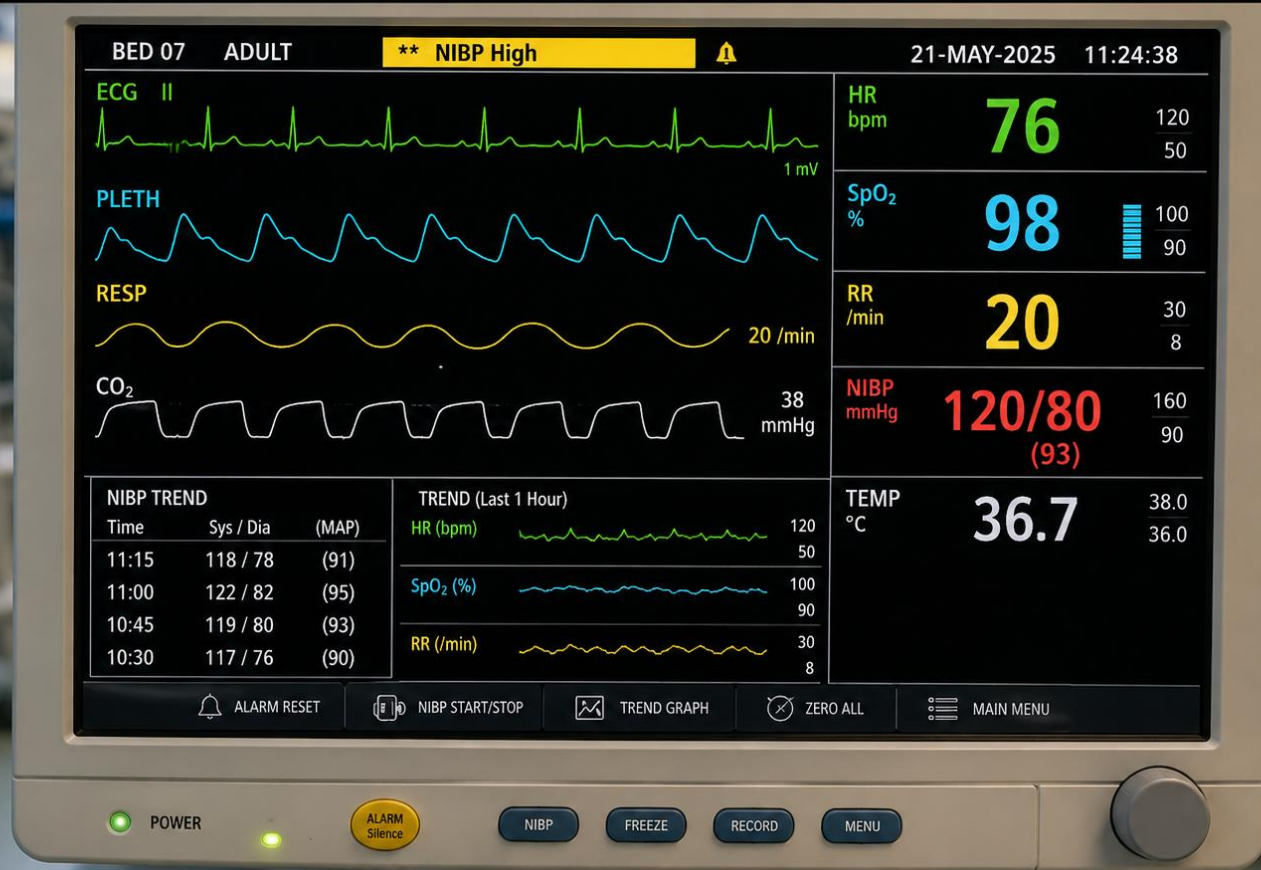
Symbol	Product	Qty	Avg Price	LTP	P&L	P&L %
RELIANCE	CNC	10	2,920.50	2,946.65	+261.50	+0.90%
TCS	CNC	5	3,725.00	3,712.80	-61.00	-0.33%
HDFCBANK	CNC	20	1,671.00	1,678.30	+146.00	+0.44%

Data Detective Challenge

You have 30 seconds.

From this single patient monitor, find:

- ① Quantitative data ② Categorical data ③ Temporal data ④ Time-series data



Data Detective Challenge

You have **30 seconds**.

From this single patient monitor, find:

② Categorical Data

Groups or labels with categories.

Examples here:

- BED 07 (Bed ID)
- ADULT (Patient Type)
- ECG, PLETH, RESP, CO₂ (Parameter types)
- NIBP High (Alarm category)

④ Time-Series Data

Data changing over time in a sequence.

Examples here:

- ECG waveform (over time)
- PLETH waveform (over time)
- RESP waveform (over time)
- CO₂ waveform (over time)
- Trend graph (last 1 hour) for HR, SpO₂, RR

① Quantitative data

② Categorical data

③ Temporal data

④ Time-series data

③ Temporal Data

Specific date and time.

Examples here:

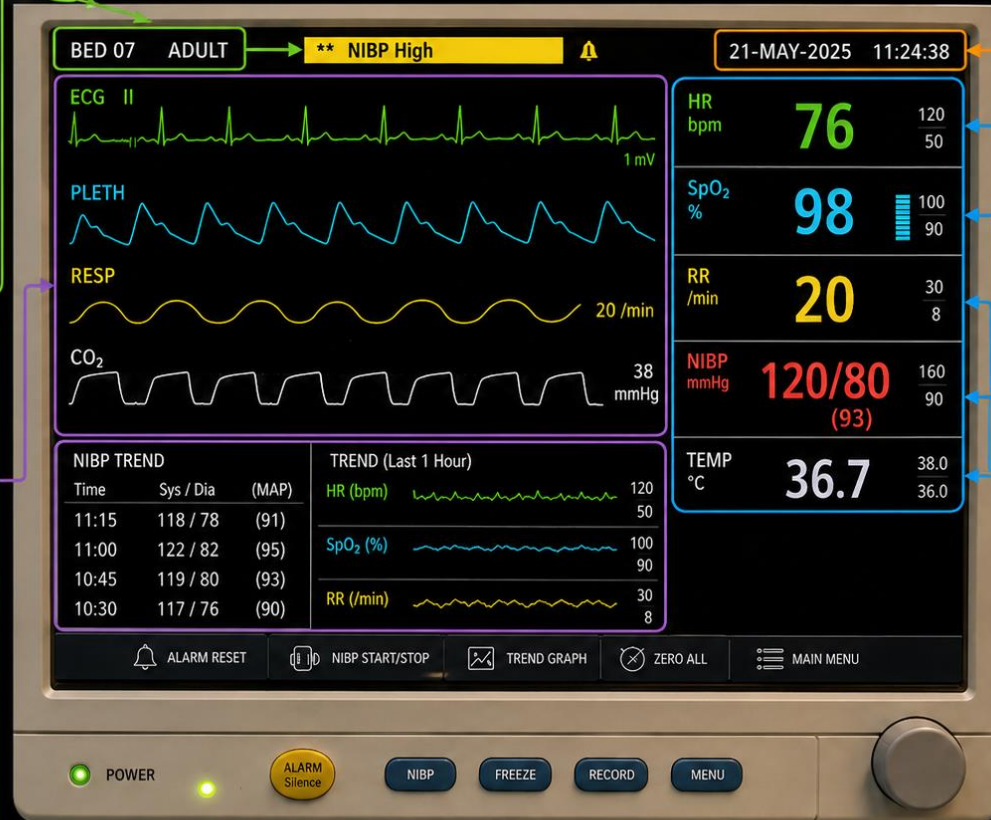
- Date: 21-MAY-2025
- Time: 11:24:38

① Quantitative Data

Numbers you can count or measure.

Examples here:

- HR: 76 bpm
- SpO₂: 98 %
- RR: 20 /min
- NIBP: 120/80 mmHg (93)
- TEMP: 36.7 °C
- CO₂: 38 mmHg
- Waveform scales (e.g., HR scale 50–120)



DATA DETECTIVE CHALLENGE



Look at this device Monitor. You have **30 seconds**.

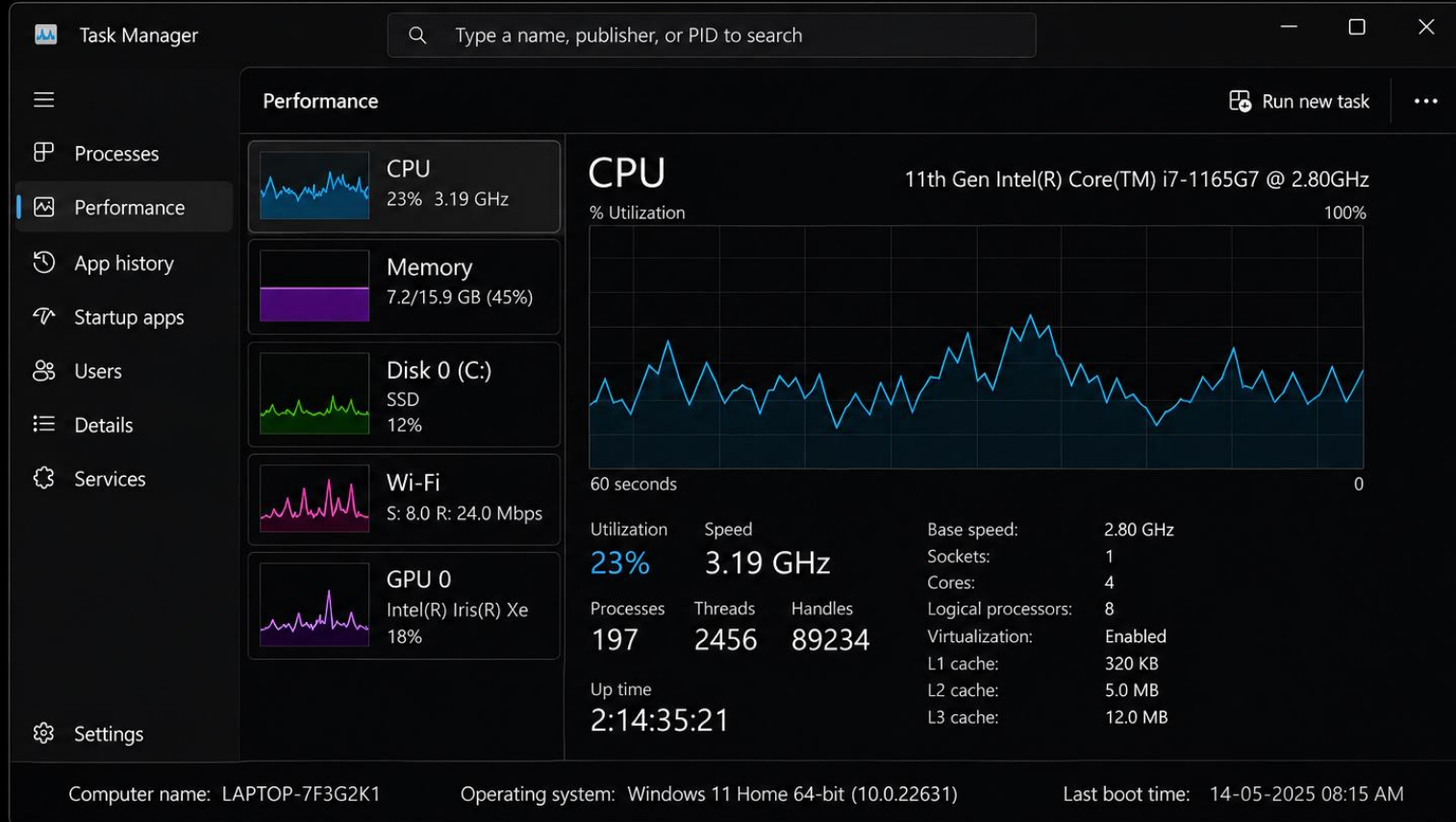
Find one example each of:

① Quantitative data

② Categorical data

③ Temporal data

④ Time-series data



DATA DETECTIVE CHALLENGE

Look at this Device Monitor. You have **30 seconds**.

Find one example each of:

① Quantitative data

② Categorical data

③ Temporal data

④ Time-series data

Quantitative Data

Numbers you can count or measure.

Example here:

- CPU Utilization: 23% (a measurable value)

Categorical Data

Groups or labels with categories.

Example here:

- Process names (CPU, Memory, Disk, Wi-Fi, GPU) (categories of system resources)

Temporal Data

Specific date and time.

Example here:

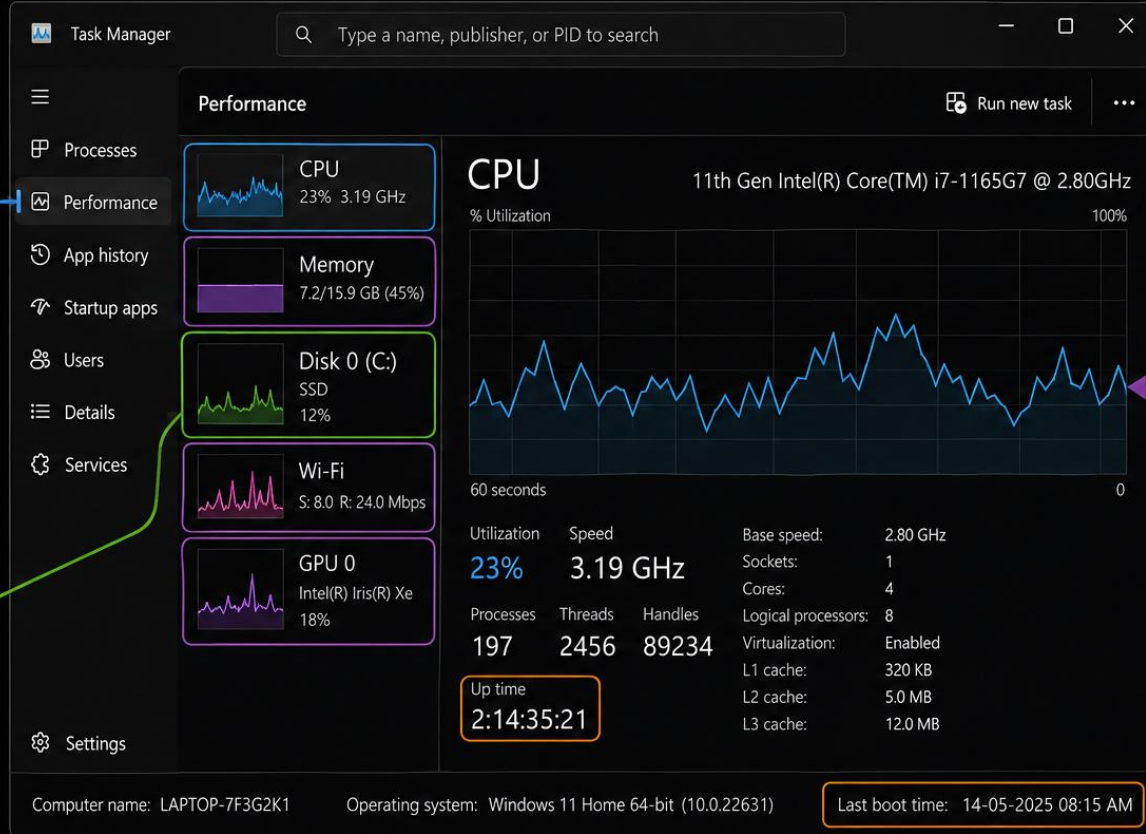
- Last boot time: 14-05-2025 08:15 AM (a specific point in time)

Time-Series Data

Data changing over time in a sequence.

Example here:

- CPU Utilization graph over the last 60 seconds (changes over time)



DATA DETECTIVE CHALLENGE

Look at this Bridge Health Monitoring System. You have **30 seconds**.
Find one example each of:

① Quantitative data

② Categorical data

③ Temporal data

④ Time-series data

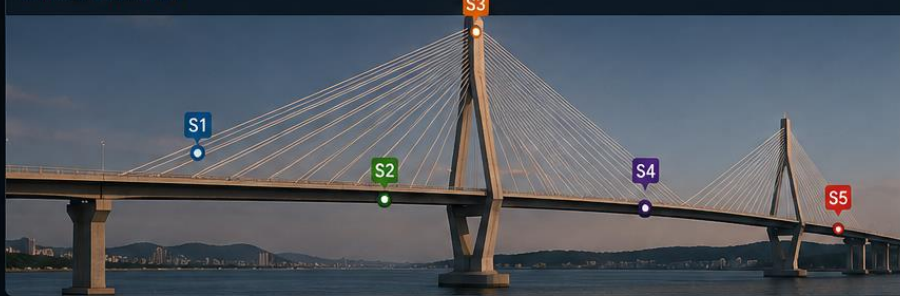


BRIDGE HEALTH MONITORING SYSTEM

14 May 2025

10:24:35 AM

BRIDGE OVERVIEW



SENSOR SUMMARY

Sensor ID	Location	Sensor Type	Status	Battery
S1	Left Abutment	Strain Gauge	OK	92%
S2	Mid Span	Strain Gauge	OK	88%
S3	Tower Top	Accelerometer	OK	76%
S4	Right of Mid Span	Accelerometer	OK	81%
S5	Right Abutment	Strain Gauge	OK	94%

MAX STRAIN

186 $\mu\epsilon$

Sensor: S2 (Mid Span)

MAX ACCELERATION

0.42 g

Sensor: S3 (Tower Top)

TEMPERATURE

28.7 $^{\circ}\text{C}$

Sensor: S4

WIND SPEED

12.6 m/s

Weather Station

ALERTS

✓ All sensors are operating within normal range.
No alerts at this time.

STRAIN ($\mu\epsilon$) - MID SPAN (S2)



ACCELERATION (g) - TOWER TOP (S3)



TEMPERATURE ($^{\circ}\text{C}$) - RIGHT OF MID SPAN (S4)



System Status: ● HEALTHY

Data Update Frequency: 1 Hz

Next Maintenance: 01 Jun 2025

DATA DETECTIVE CHALLENGE ★

Look at this Bridge Health Monitoring System. You have **30 seconds**.

Find one example each of:

① Quantitative data

② Categorical data

③ Temporal data

④ Time-series data



BRIDGE HEALTH MONITORING SYSTEM

14-May-2025 10:35:42 AM

TEMPERATURE

28.4 °C

WIND SPEED

12.6 m/s

HUMIDITY

54 %

SYSTEM STATUS

OK

LATEST SENSOR READINGS

Sensor ID	Sensor Type	Location	Strain ($\mu\epsilon$)	Acceleration (mg)	Timestamp
S1	Strain Gauge	Tower Top	-23.6	1.24	14-May-2025 10:35:40 AM
S2	Strain Gauge	Mid Span	15.2	0.98	14-May-2025 10:35:40 AM
S3	Accelerometer	Quarter Span	-	2.15	14-May-2025 10:35:40 AM
S4	Accelerometer	Pier Base	-	1.68	14-May-2025 10:35:40 AM

STRAIN ($\mu\epsilon$) – MID SPAN (S2)



ACCELERATION (mg) – QUARTER SPAN (S3)



Categorical Data

Groups or labels with categories.

Example here:

- Sensor Type (Strain Gauge, Accelerometer)
- Sensor Location (Tower Top, Mid Span, etc.)

BRIDGE HEALTH MONITORING SYSTEM

14-May-2025 10:35:42 AM

TEMPERATURE

28.4 °C

WIND SPEED

12.6 m/s

HUMIDITY

54 %

SYSTEM STATUS

OK

LATEST SENSOR READINGS

Sensor ID	Sensor Type	Location	Strain ($\mu\epsilon$)	Acceleration (mg)	Timestamp
S1	Strain Gauge	Tower Top	-23.6	1.24	14-May-2025 10:35:40 AM
S2	Strain Gauge	Mid Span	15.2	0.98	14-May-2025 10:35:40 AM
S3	Accelerometer	Quarter Span	-	2.15	14-May-2025 10:35:40 AM
S4	Accelerometer	Pier Base	-	1.68	14-May-2025 10:35:40 AM

STRAIN ($\mu\epsilon$) – MID SPAN (S2)



ACCELERATION (mg) – QUARTER SPAN (S3)



Quantitative Data

Numbers you can count or measure.

Example here:

- Strain: 15.2 $\mu\epsilon$
- Acceleration: 2.15 mg
- Temperature: 28.4 °C
- Wind Speed: 12.6 m/s

Temporal Data

Specific date and time.

Example here:

- 14-May-2025 10:35:42 AM
- 14-May-2025 10:35:40 AM

Time-Series Data

Data changing over time in a sequence.

Example here: Strain and Acceleration signals over time (line charts).

Exercises

1. **Classify a dataset.** Take any spreadsheet you have to hand and label every column with one of the five types. Note any column you find hard to place.
2. **Spot the ordinal.** Find three ordinal variables in everyday life and explain why the gaps between their levels are uneven.
3. **Break a date.** Write the same moment in three different date formats and explain which is safest to store and why.
4. **Build a series.** Count how many messages you sent each day last week. Plot it. Is there a weekly pattern?
5. **Design multi-modal.** Sketch what a supermarket would need to record to detect a spoiled delivery, using at least three modes.

Where to go next

Foundations

- Levels of measurement – the nominal, ordinal, interval and ratio framework behind these categories
- Tidy data principles – one variable per column, one observation per row

Practical skills

- Summary statistics and distributions – mean, median, quartiles, standard deviation
- Date and time handling in your chosen tool – parsing, time zones, resampling
- Encoding categorical variables for models – one-hot, ordinal and target encoding

Going further

- Time-series decomposition and forecasting – trend, seasonality and simple baselines
- Introductory multi-modal learning – embeddings, alignment and fusion

Know your data type

Get the type right and the method follows