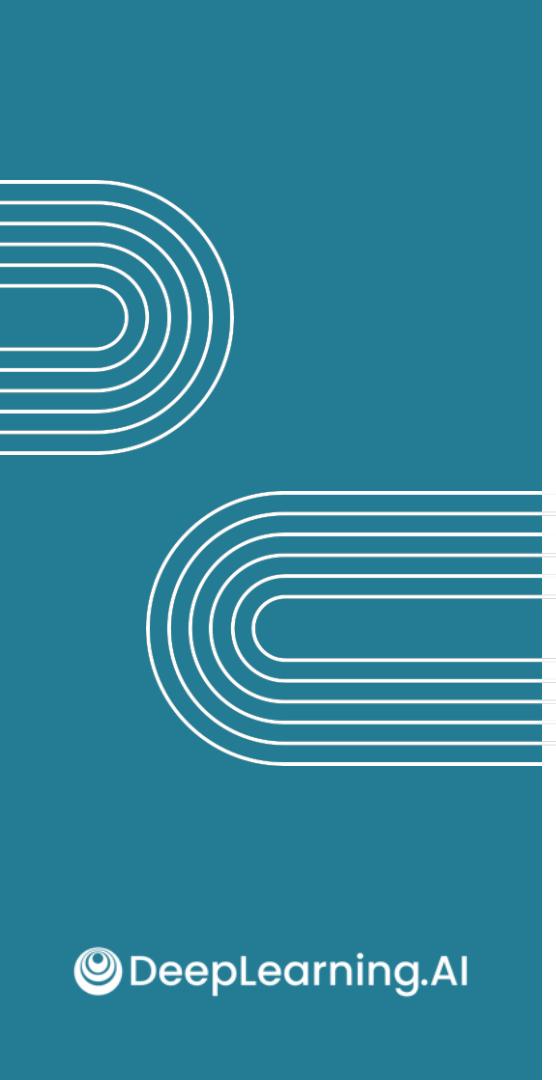


Data Analytics Foundations

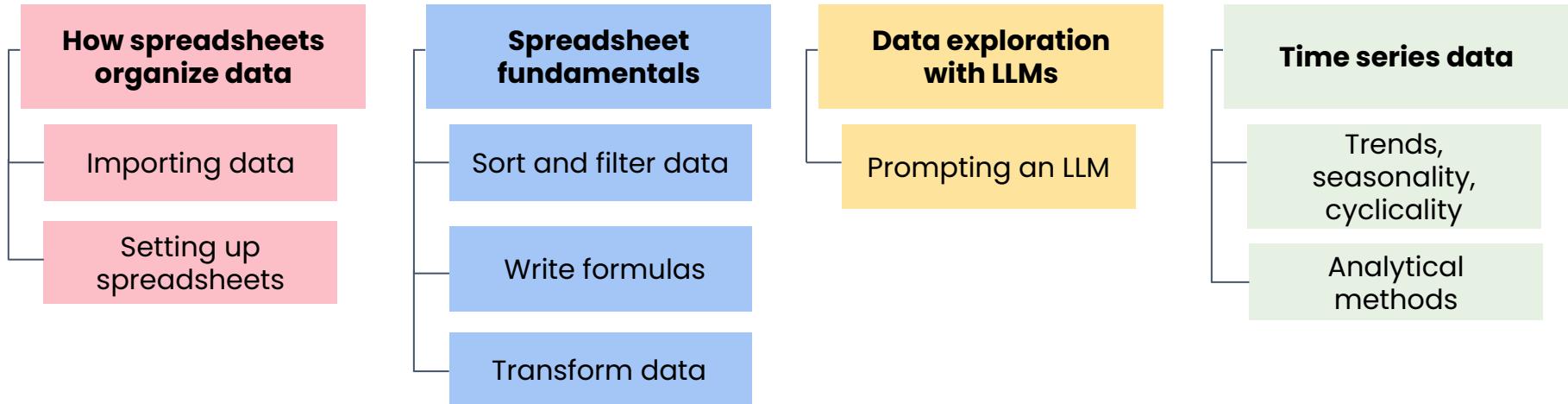
Module 2: Using spreadsheets
for data analytics

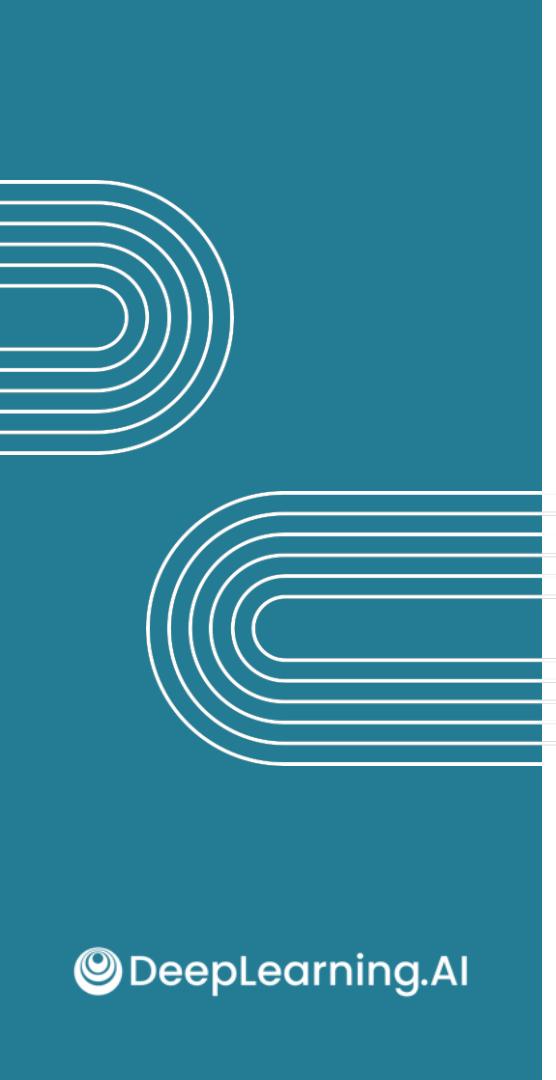


Using spreadsheets for data analytics

Module 2 introduction

Module 2 outline





Using spreadsheets for data analytics

Solving problems with data

What outcomes am I interested in?

Problem	Outcome	Data
 Increase profit	+ Sales	 Sales reports
	- Expenses	 Expense reports
 Improve patient outcomes	+ Patient satisfaction	 Patient surveys
	- Length of hospital stays	 Admission and discharge dates

What data provides context?

Who

What

When

Where

Outcome

Customer ID

Product

Purchase Date

Region

Sale



Increasing revenue for a solar panel company

Who

Customer ID

973

What

Product

Cable set

When

Purchase Date

06/15/2024

Where

Region

North America

Outcome

Sale

\$1,600

Customer ID

111

Product

Solar panel

Purchase Date

09/10/2024

Region

Asia

Sale

\$3,200

- High paying customers?



Increasing revenue for a solar panel company

Who	What	When	Where	Outcome
Customer ID 973	Product Cable set	Purchase Date 06/15/2024	Region North America	Sale \$1,600
Customer ID 111	Product Solar panel	Purchase Date 09/10/2024	Region Asia	Sale \$3,200

- High paying customers?
- Specific products driving sales?



Increasing revenue for a solar panel company

Who	What	When	Where	Outcome
Customer ID 973	Product Cable set	Purchase Date 06/15/2024	Region North America	Sale \$1,600
Customer ID 111	Product Solar panel	Purchase Date 09/10/2024	Region Asia	Sale \$3,200

- High paying customers?
- Specific products driving sales?
- How are sales trending?



Increasing revenue for a solar panel company

Who	What	When	Where	Outcome
Customer ID 973	Product Cable set	Purchase Date 06/15/2024	Region North America	Sale \$1,600
Customer ID 111	Product Solar panel	Purchase Date 09/10/2024	Region Asia	Sale \$3,200

- High paying customers?
- Specific products driving sales?
- How are sales trending?
- Sales vary across regions?



Increasing revenue for a solar panel company

Who

What

When

Where

Outcome

Customer ID

973

Product

Cable set

Purchase Date

06/15/2024

Region

North America

Sale

\$1,600

Customer ID

111

Product

Solar panel

Purchase Date

09/10/2024

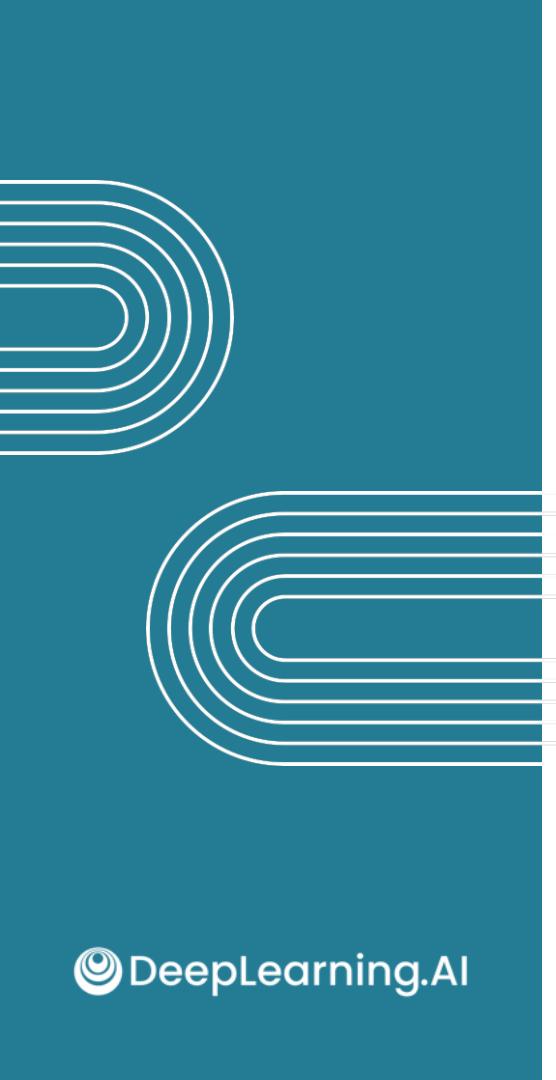
Region

Asia

Sale

\$3,200

- Business or individual purchaser?
 - If business, how large?
- How many orders placed?
- High paying customers?
- Specific products driving sales?
- Sales trends
- Sales vary across regions?



Using spreadsheets for data analytics

Spreadsheets for
business analytics

Why start with spreadsheets?



Industry standard tool

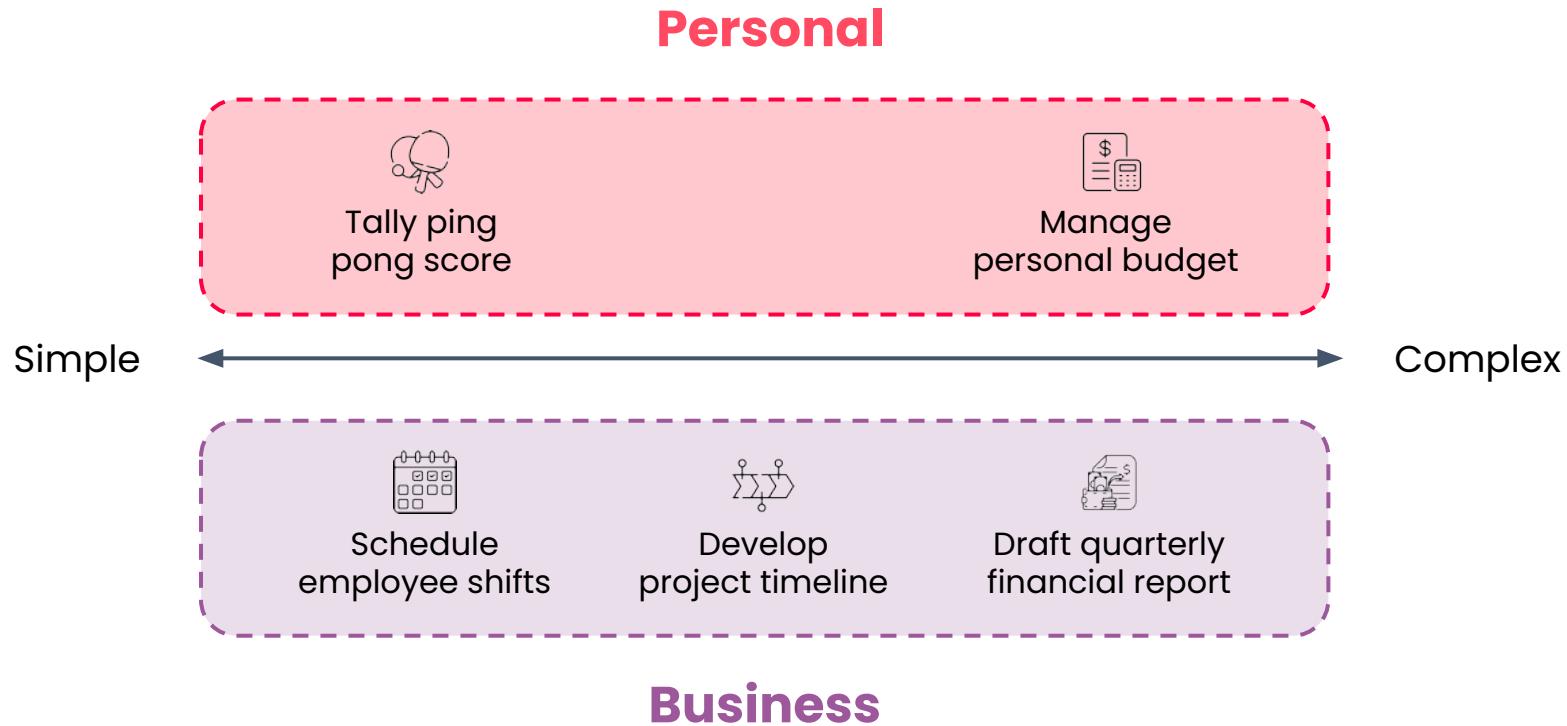


Opens in a few seconds



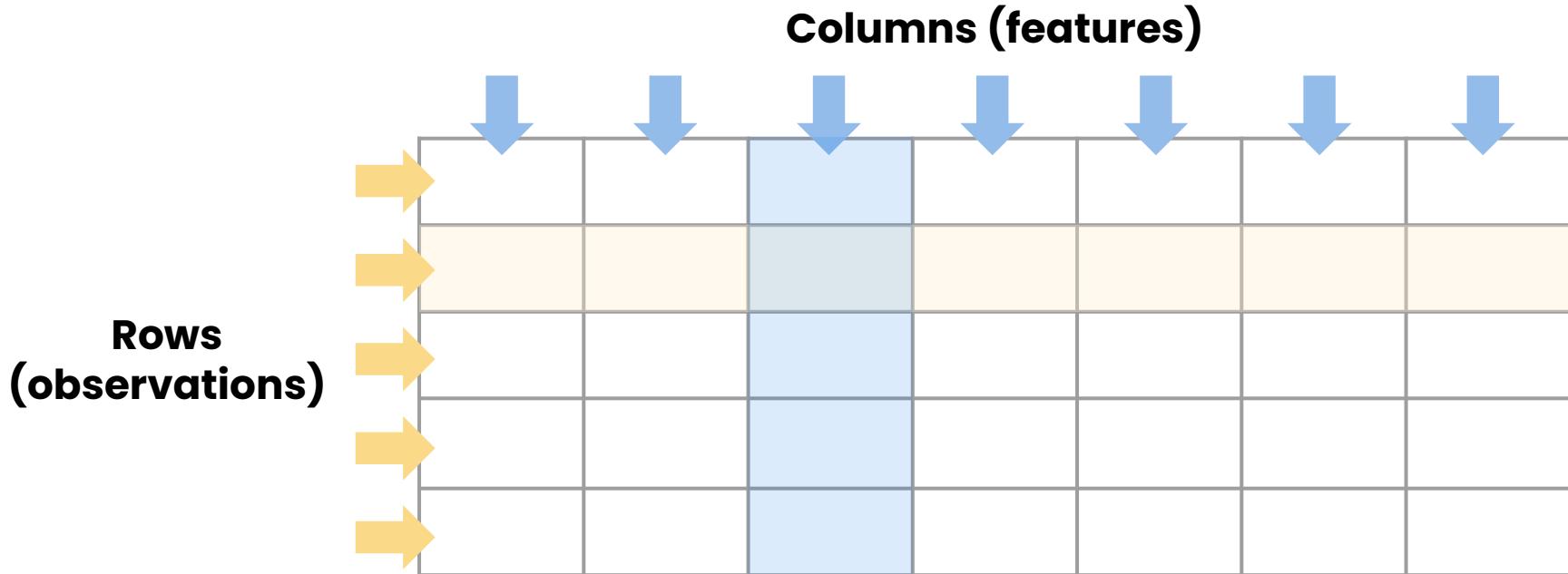
Free

Spreadsheet applications



Spreadsheet applications

- ★ Designed for working with structured data



Unstructured data in spreadsheets

text

images

audio

video

✓ collecting

✓ organizing

✗ analyzing

Essay on African Rhythms											
African rhythms are an integral part of the rich musical heritage of the African continent. These rhythms have evolved over centuries, shaped by the diverse cultures, languages, and traditions of African peoples. From the pulsating beats of West African drums to the intricate polyrhythms of East Africa, these rhythms have had a profound impact on music worldwide, influencing genres such as jazz, blues, rock, and hip-hop.											
One of the most distinctive features of African rhythms is their complexity. Unlike Western music, which often relies on a single, steady beat, African rhythms typically involve multiple layers of interlocking patterns. These patterns are created by various instruments, including drums, rattles, bells, and even hand claps. Each instrument plays a specific role in the overall rhythm, creating a sense of tension and release that propels the music forward and engages the listener.											
Another important aspect of African rhythms is their connection to dance. In many African cultures, music and dance are inseparable, with rhythms serving as the foundation for elaborate dance performances. These dances often have deep cultural and spiritual significance, telling stories, celebrating life events, or honoring ancestors. The rhythms guide the dancers' movements, creating a powerful synergy between music and motion.											
African rhythms also play a crucial role in communication and social bonding. In many African societies, drumming is used as a means of sending messages over long distances, with specific rhythmic patterns conveying different meanings. Rhythms are also used to bring communities together, whether in the context of religious ceremonies, festivals, or everyday gatherings. Participating in group drumming and dancing fosters a sense of unity and shared experience that strengthens social ties.											
The influence of African rhythms extends far beyond the continent itself. During the transatlantic slave trade, enslaved Africans brought their musical traditions with them to the Americas, where they blended with European and indigenous musical styles to create new genres. African rhythms formed the backbone of blues, jazz, and rock and roll, with the syncopated beats and infectious rhythms still evident in these musical roots.											
In the 20th century, African rhythms continued to shape popular music around the world. The rise of Afrobeat in the 1970s, pioneered by Nigerian musician Fela Kuti, fused African rhythms with funk, jazz, and highlife to create a powerful, politically charged sound. In the 1980s and 90s, hip-hop and electronic dance music heavily sampled African rhythms, introducing them to new audiences and sparking a renewed interest in traditional African music.											
Today, African rhythms continue to evolve and inspire musicians worldwide. Contemporary African artists are pushing the boundaries of traditional rhythms, incorporating elements of modern genres like hip-hop, rock, and electronic music. At the same time, there is a growing movement to preserve and promote traditional African rhythms, with initiatives like the African Drum Festival in Nigeria and the Festival on the Niger in Mali showcasing the incredible diversity and richness of African percussion.											
In conclusion, African rhythms are a testament to the creativity, resilience, and spirit of the African people. From their origins in ancient tribal ceremonies to their global influence on popular music, these rhythms have demonstrated an enduring power to move, inspire, and unite people across cultures and generations. As we continue to explore and celebrate the musical treasures of Africa, we can be sure that its rhythms will keep pulsing through the heart of music for generations to come.											
21											
22											
23											
24											

Unstructured data in spreadsheets

text

images

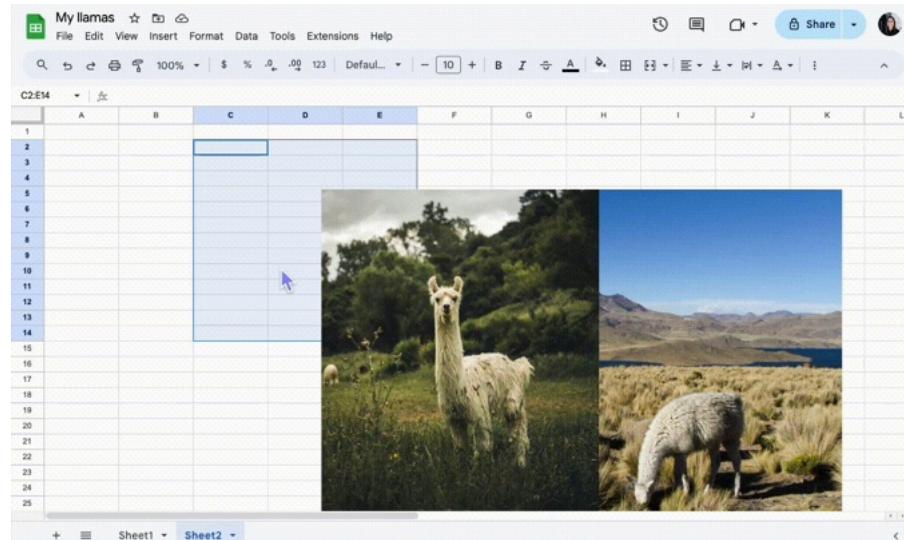
audio

video

✓ collecting

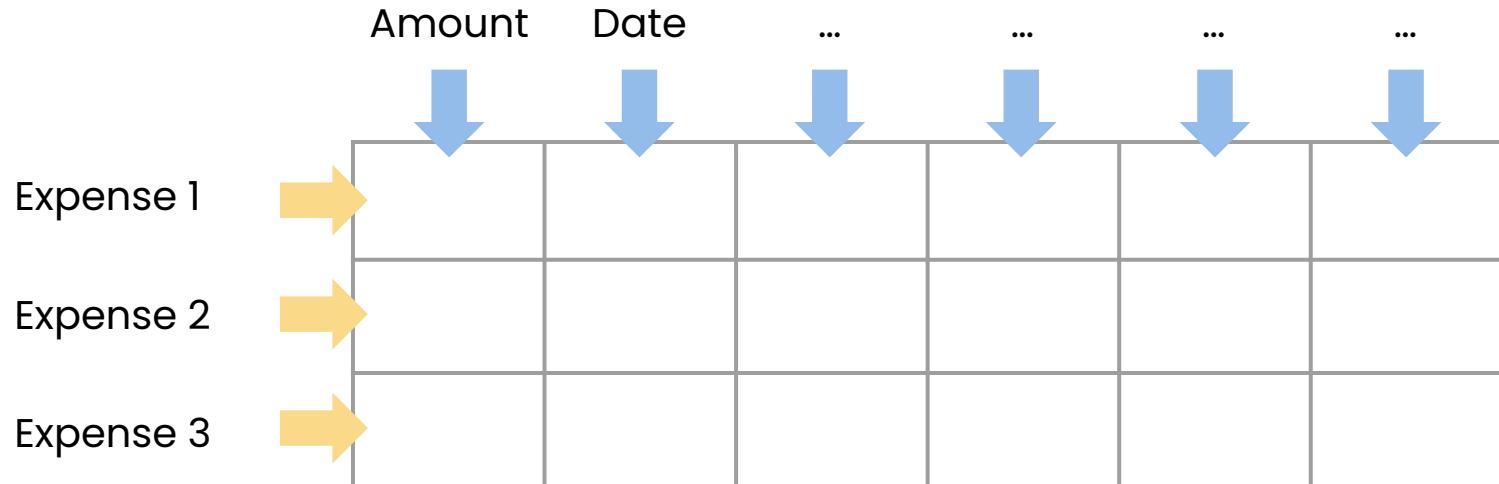
✓ organizing

✗ analyzing



Are spreadsheets right for your use case?

1. Can your data be organized into **rows** and **columns**?



Are spreadsheets right for your use case?

1. Can your data be organized into **rows** and **columns**?

???

Essay on African Rhythms

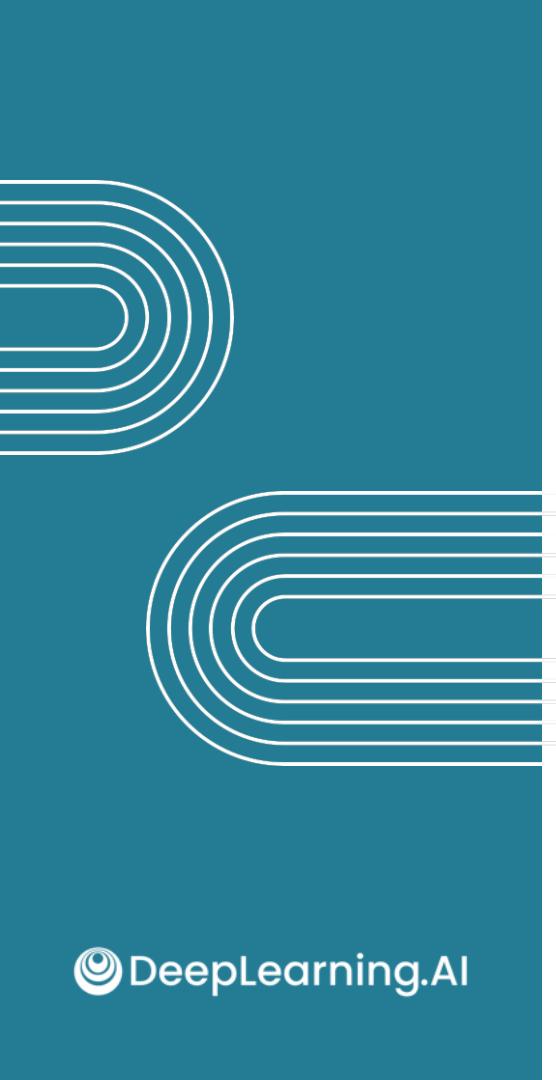
African rhythms are an integral part of the rich musical heritage of the African continent. These rhythms have evolved over centuries, shaped by the diverse cultures, languages, and traditions of African peoples. From the pulsating beats of West African drums to the intricate polyrhythms of East Africa, these rhythms have had a profound impact on music worldwide, influencing genres such as jazz, blues, rock, and hip-hop.

One of the most distinctive features of African rhythms is their complexity. Unlike Western music, which often relies on a single, steady beat, African rhythms typically involve multiple layers of interlocking patterns. These patterns are created by various instruments, including drums, hand claps, and other percussion instruments. Each instrument plays a specific role in the overall rhythmic structure, contributing to the rich tapestry of African musical tradition.

Are spreadsheets right for your use case?

1. Can your data be organized into **rows** and **columns**?

2. Are there **relationships** you want to explore between different aspects of the data?
 - a. Organizing expenses by category
 - b. Finding the month where you spent the most



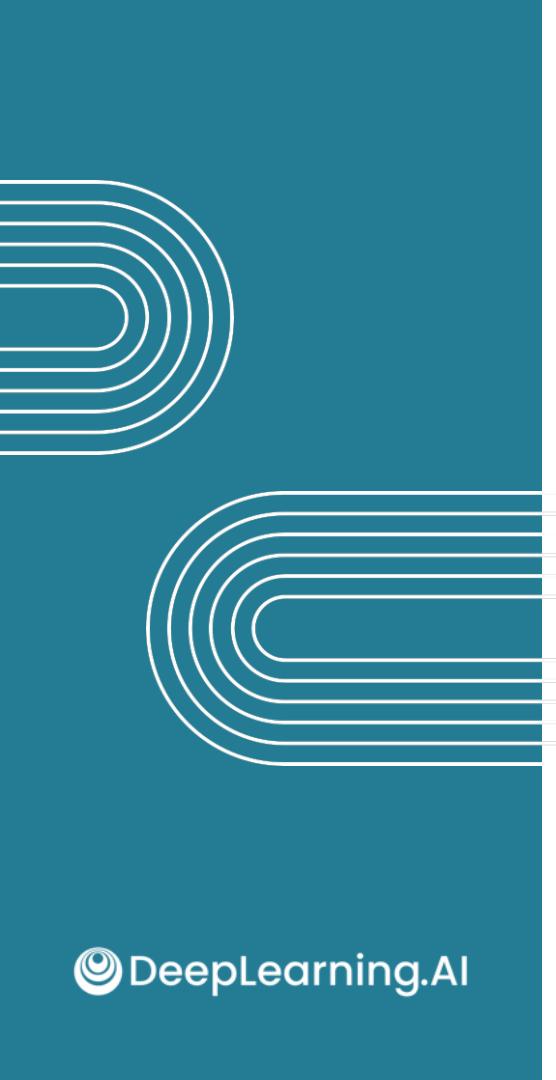
Using spreadsheets for data analytics

Navigating Google Sheets

Google Sheets alternatives



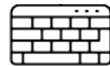
Highly transferable skills



Using spreadsheets for data analytics

Importing data

Loading data into Google Sheets



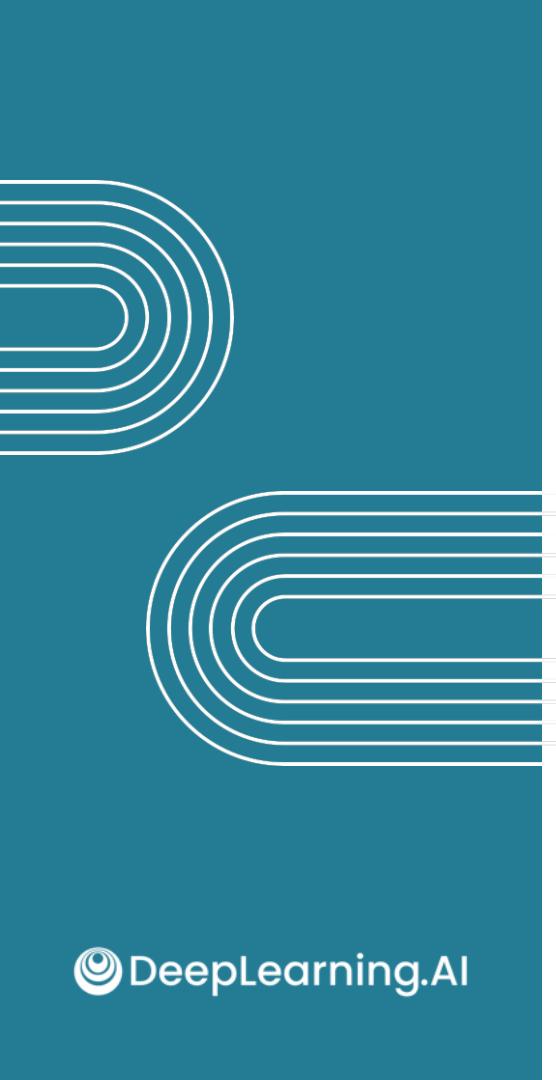
Generate data directly in the spreadsheet



Open an existing file

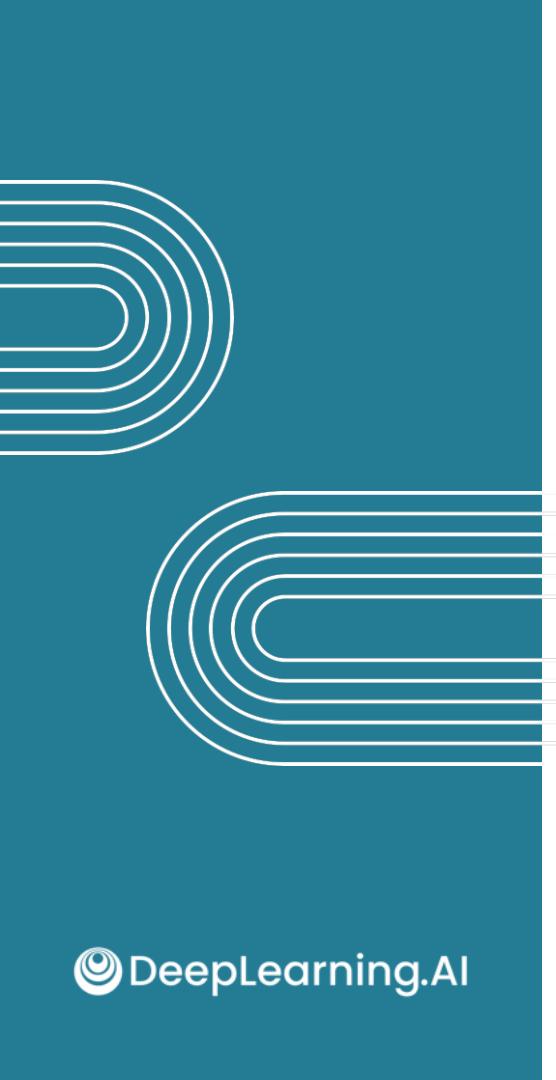


Import a structured data set (.csv, .xlsx)



Using spreadsheets for data analytics

Sorting, filtering,
and formatting



Using spreadsheets for data analytics

Getting to know your data

Techniques for getting to know your data

Structure:



Structured



Unstructured

Dimensions:



of observations



of features

Type:



Numerical



Categorical

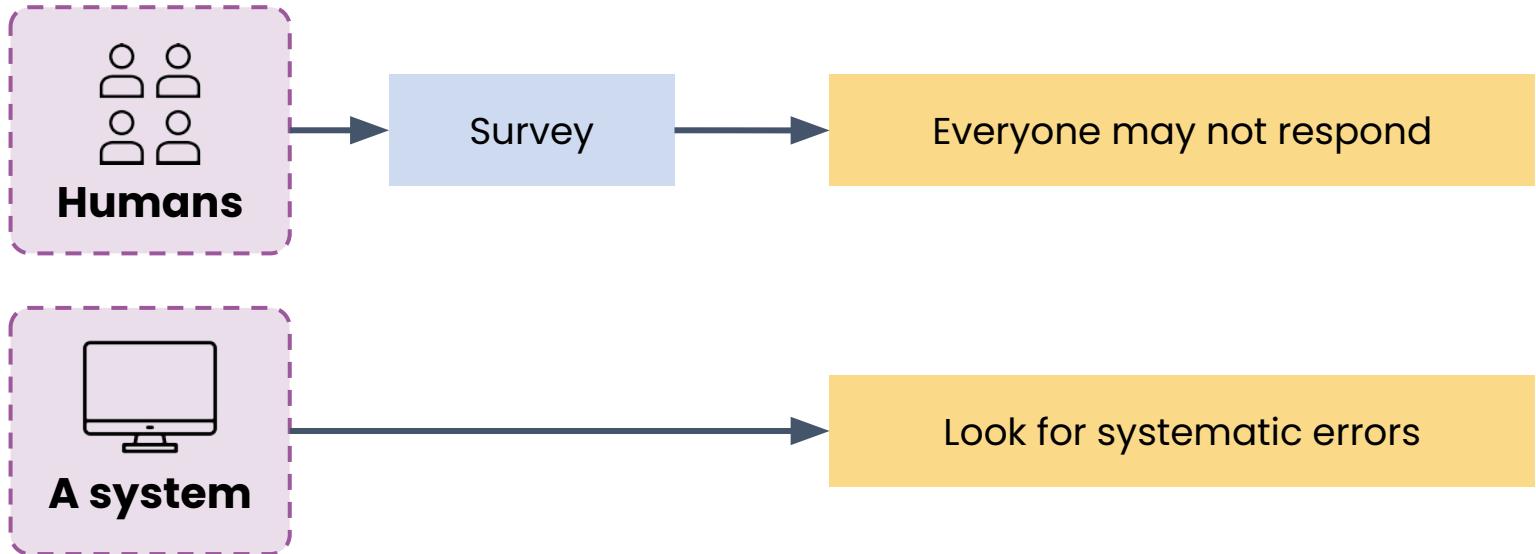
Metadata: data about your data

More metadata

Type of metadata	Example	What it tells you
 File name	Hotel reservations	How you can search for the data
 Author	Charles Xavier	Who produced the data
 Description	About hotel rooms	Context about what information is in data
 File format	.csv, .txt, .json	How you may need to interact with the data
 Created/last updated date	January 20, 2024	How current the data is
 Access controls	View, comment, edit	Who can access the data and how they can interact with it

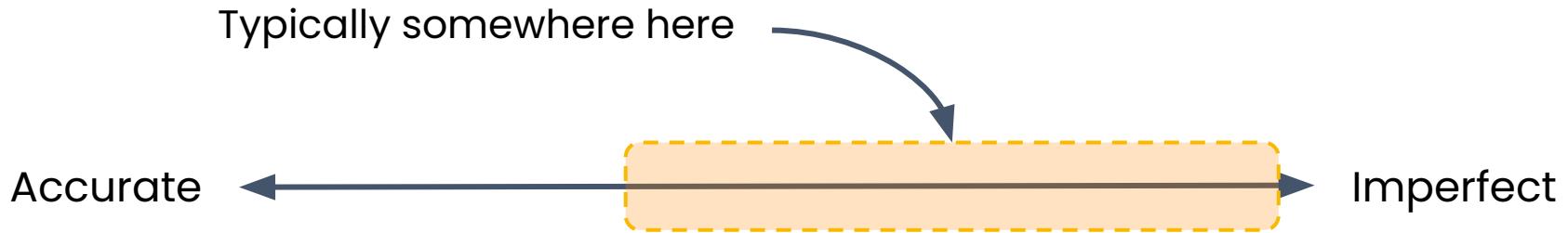
The source of your data

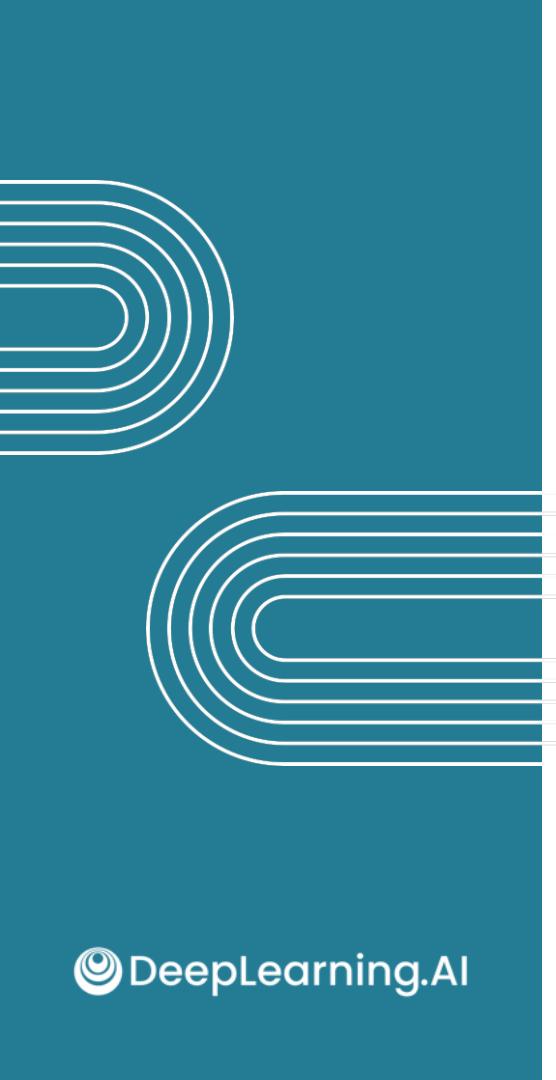
- How was it **collected** or **generated**?



The source of your data

- How was it **collected** or **generated**?
- What is the **quality** of the data?





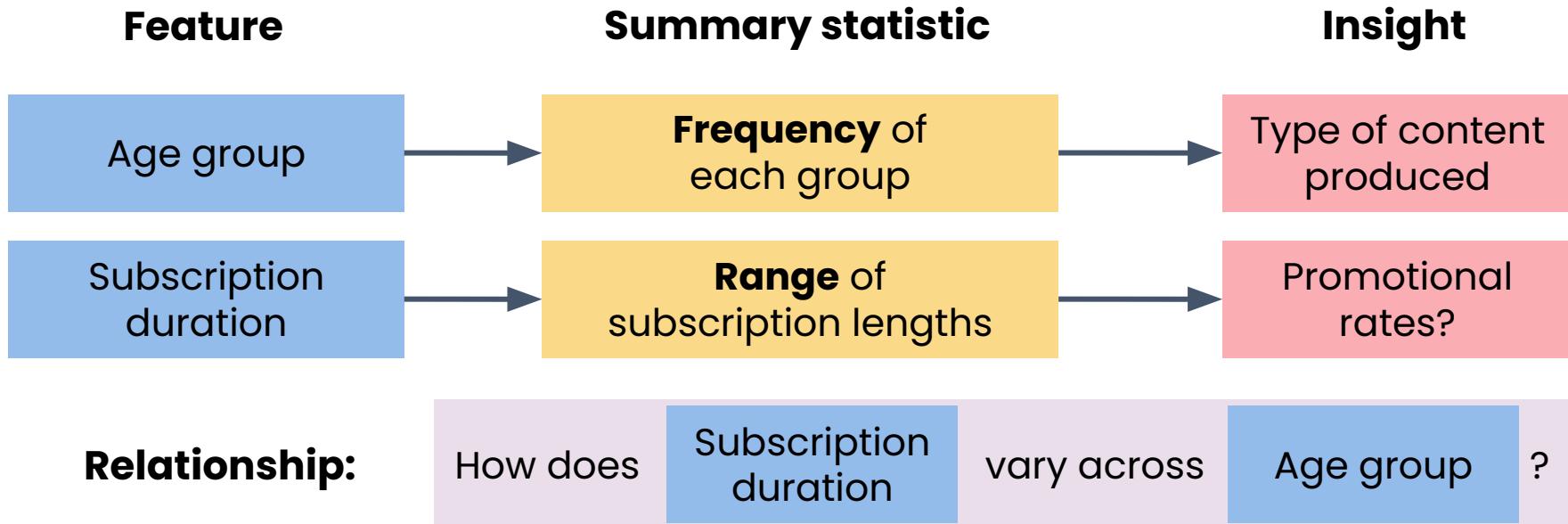
Using spreadsheets for data analytics

**Summary statistics –
MAX, MIN, AVERAGE**

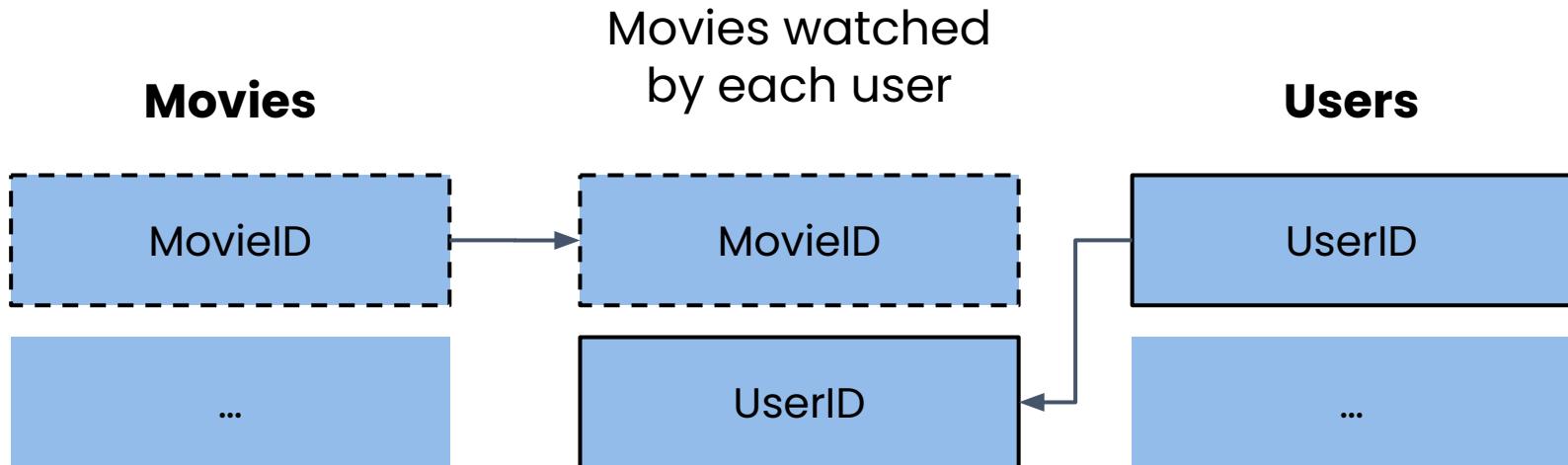
Summary statistics



Survey of active subscribers



Relations to other data sets





Using spreadsheets for data analytics

Conditional formatting

Conditional formatting

Allows you to:

- Apply a particular rule across your data set
 - Numerical data
 - Categorical data
 - Dates



8	8
1	1
3	3
7	7
8	8
9	9
10	10
6	6

Visual signals for:

- Spotting trends across many data points
- Identifying positive or negative changes
- Identifying which values fall above or below a threshold

Two types of conditional formatting

Single color



Highlight cells that satisfy a particular condition

8
1
3
7
8
9
10
6

Rule:

Highlight **green** days
with ≥ 8 sales

Color scale



Applies one of several colors to each cell based on its value

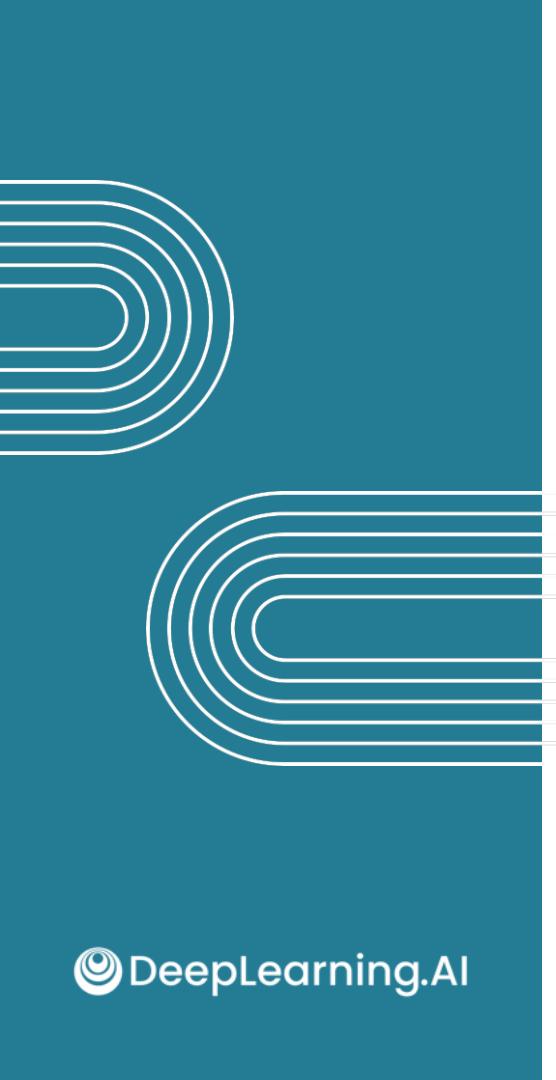
Two common types:

Sequential



Divergent





Using spreadsheets for data analytics

**Summary statistics –
COUNTIF**

Q. What percent of bookings were canceled?



COUNTIF



Conditional



Given a range or group of cells



Tallies them up only if they meet a **condition**



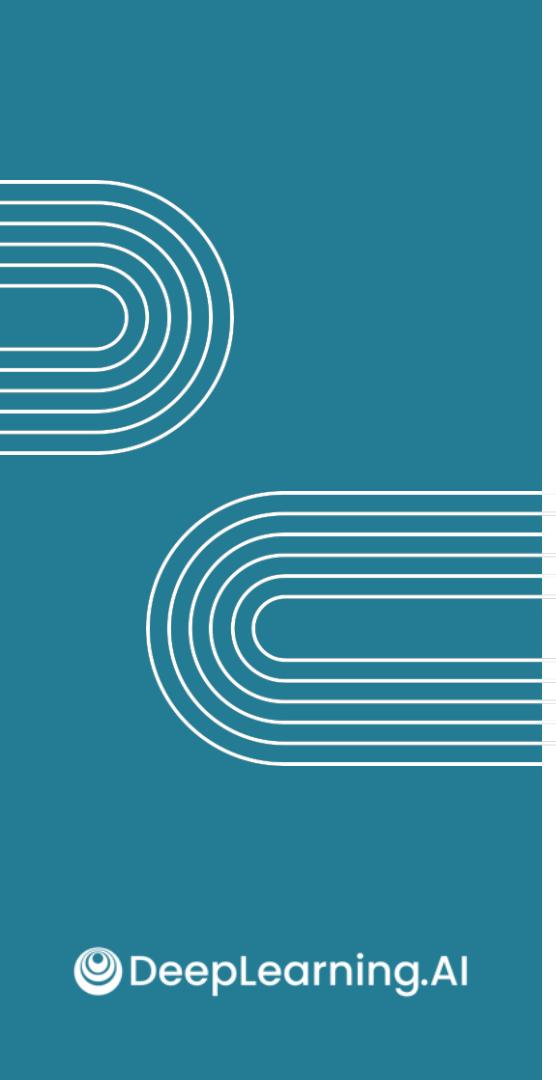
Similar to a filter, but counts rather than displays data

Formula

= COUNTIF (**A1:A9** , “ <=0 ”)

Range

Criteria



Using spreadsheets for data analytics

Summary statistics –
SUMIF, AVERAGEIF

If the **number of adults** equals zero, add the **number of children** up.



SUMIF



Conditional

 **Adds** all the cells in a range if Σ cells meet a condition

 **Check** cells in one range,
add cells in another

Adults = 0?

adults	children
1	2
2	1
0	3
0	1
0	1

Sum = 5

Formula

Range

Condition

Sum range

= **SUMIF** (**A1:A5** , **=0** , **B1:B5**)

Average lead time of bookings that were and were not canceled.



AVERAGEIF



Averages all values that meet the condition

Adults = 0?

Adults	Children
1	2
2	1
0	3
0	1
0	1

Avg = 1.6

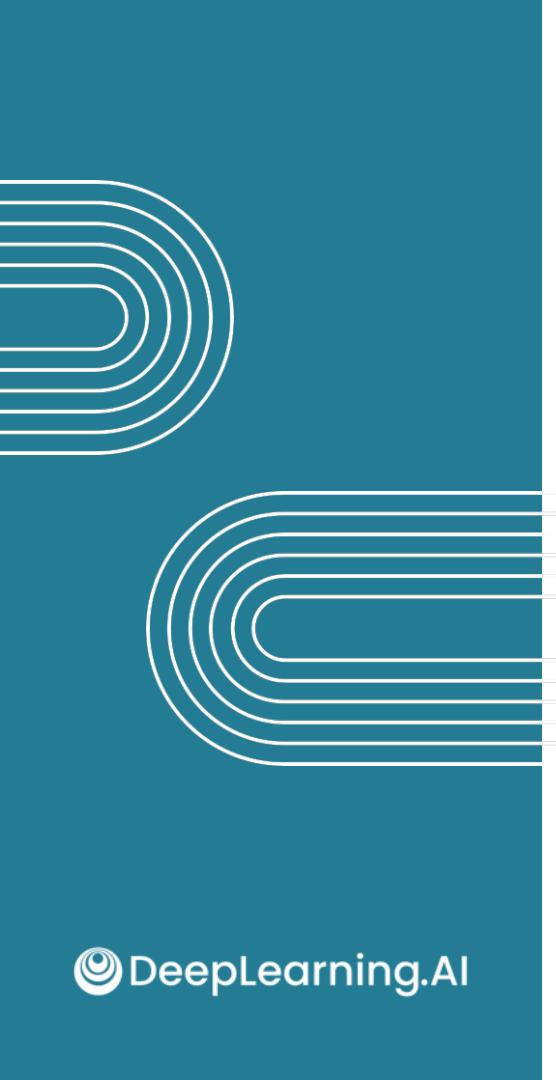
Formula

Range

Condition

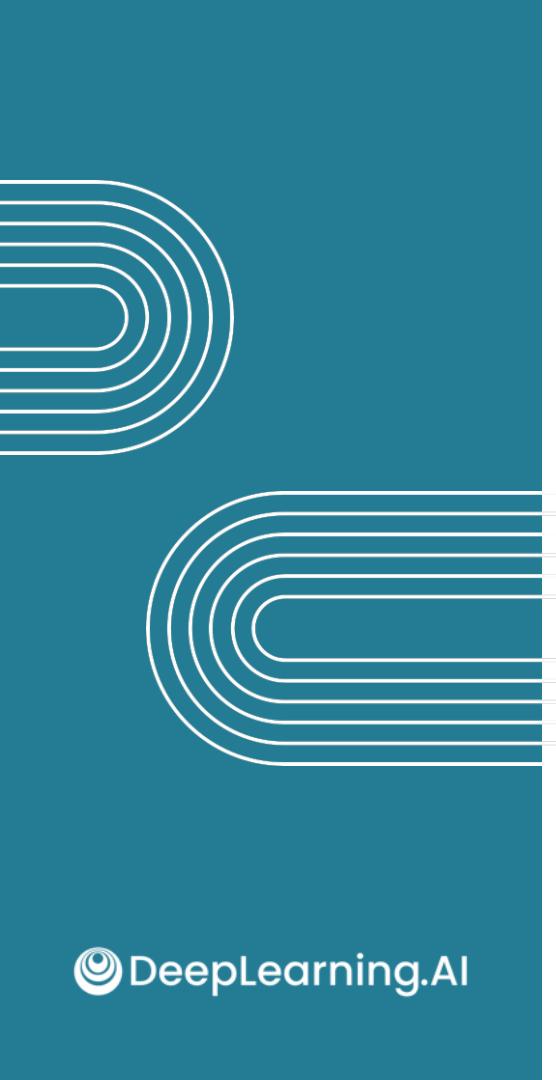
Sum range

AVERAGEIF(, ,)



Using spreadsheets for data analytics

**Summary statistics –
COUNTIFS, SUMIFS**



Using spreadsheets for data analytics

Data processing –
IF, IFS, RIGHT, LEFT

Categorical groupings

Group features with similar relationship to outcome

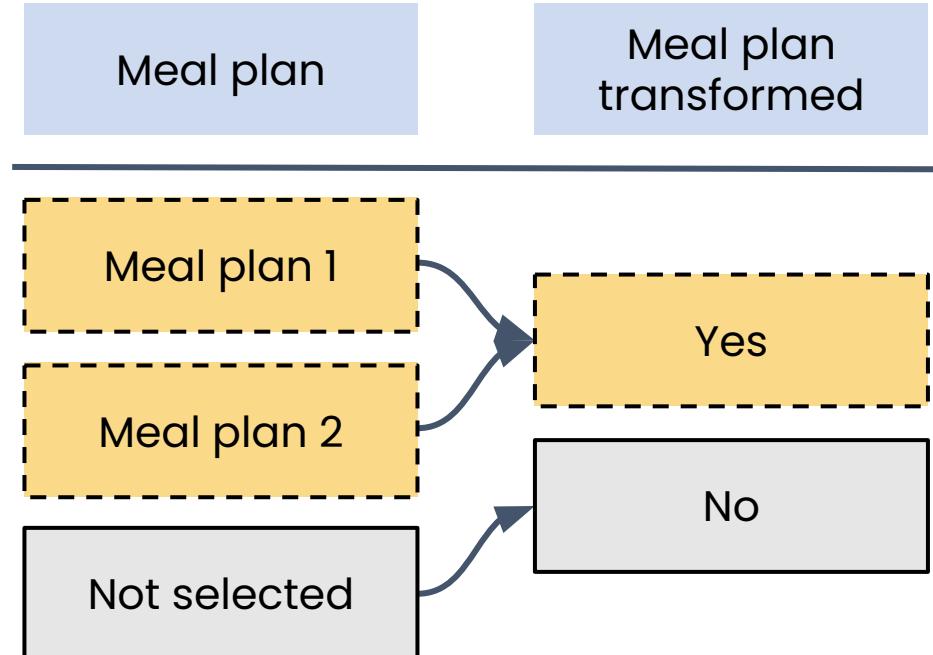


IF

Checks a condition, then returns values depending on whether the condition is true or false

Formula

```
=IF(A2, “=Not selected”, “No”, “Yes” )
```



Text processing

Rename category or extract part of a category

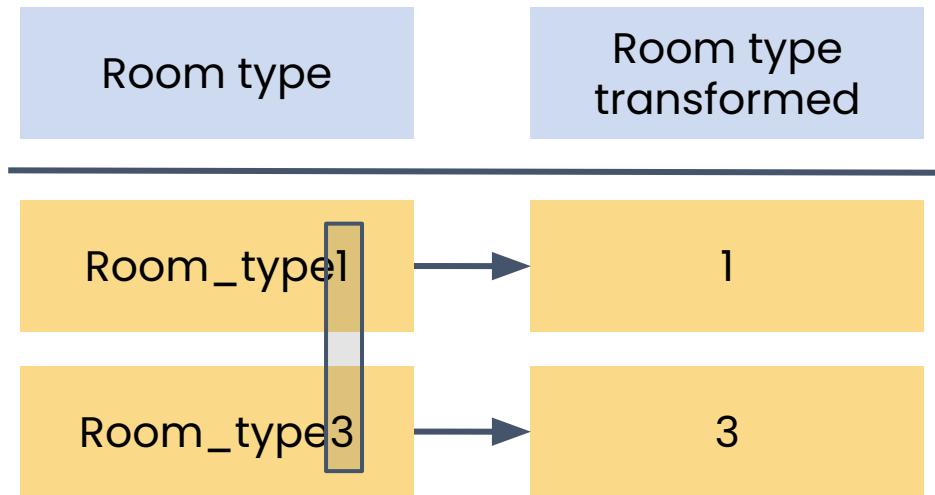


LEFT



RIGHT

Take leftmost or rightmost characters and put them in a new cell



Formula

```
=RIGHT(A2, 1)
```

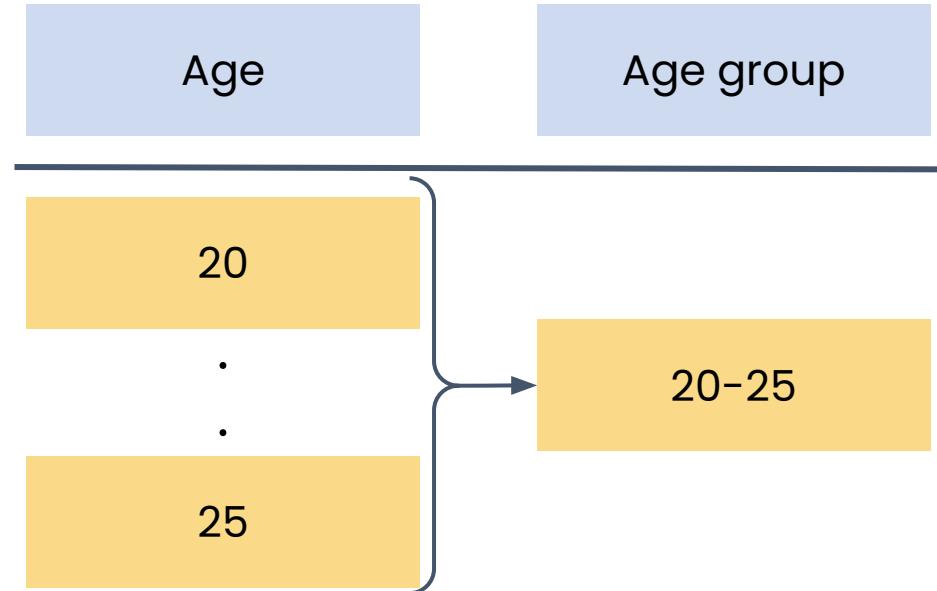
Binning

Group numerical data into categories



IFS

Checks multiple conditions



	Condition1	Value if true1	Condition2	Value if true2
--	------------	----------------	------------	----------------

Formula

=IFS(A2<18, “Under 18”, A2<30, “18-29”)

Spreadsheets functions or data processing



IF



LEFT

Condition 1

Condition 2

=IFS(A2="Not selected", "No", A2="Selected", "Yes")

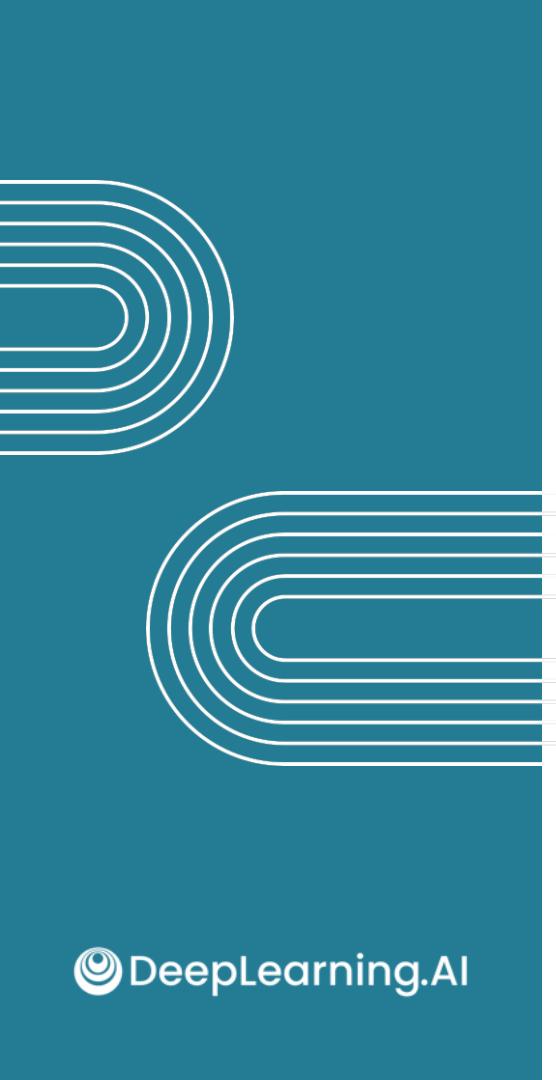


RIGHT



IFS

Checks multiple conditions



Using spreadsheets for data analytics

Where does data
come from?

Data comes from many sources



[keri, Magic 8 ball - MY SOURCES SAY NO,
<https://commons.wikimedia.org/>, CC 2.0, (2014)]



[NASA, NOAA-M, <https://commons.wikimedia.org/wiki/File:NOAA-M.jpg>, (2002)]

Data comes from many sources

1. Direct input

- Explicitly provided through a structured process



Feedback survey



Registration form



2. Behavioral observation

- Passively observing individuals' actions



Website analytics



Mobile app usage



3. Physical sensors

- Continuously monitor some phenomenon



Smart thermostat



Vehicles

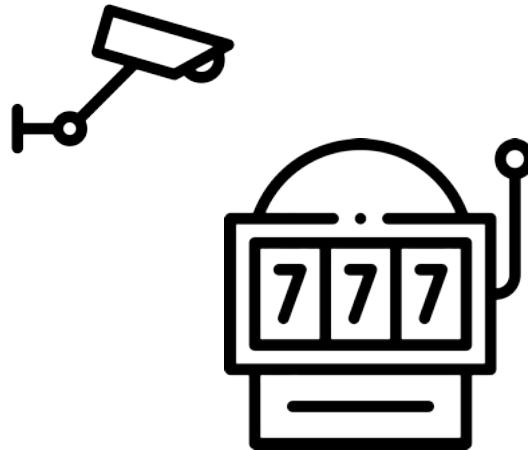


1st party, 2nd party, and 3rd party data

Your company

1st party

Collected by you or
your company directly



1st party, 2nd party, and 3rd party data

Your company
1st party

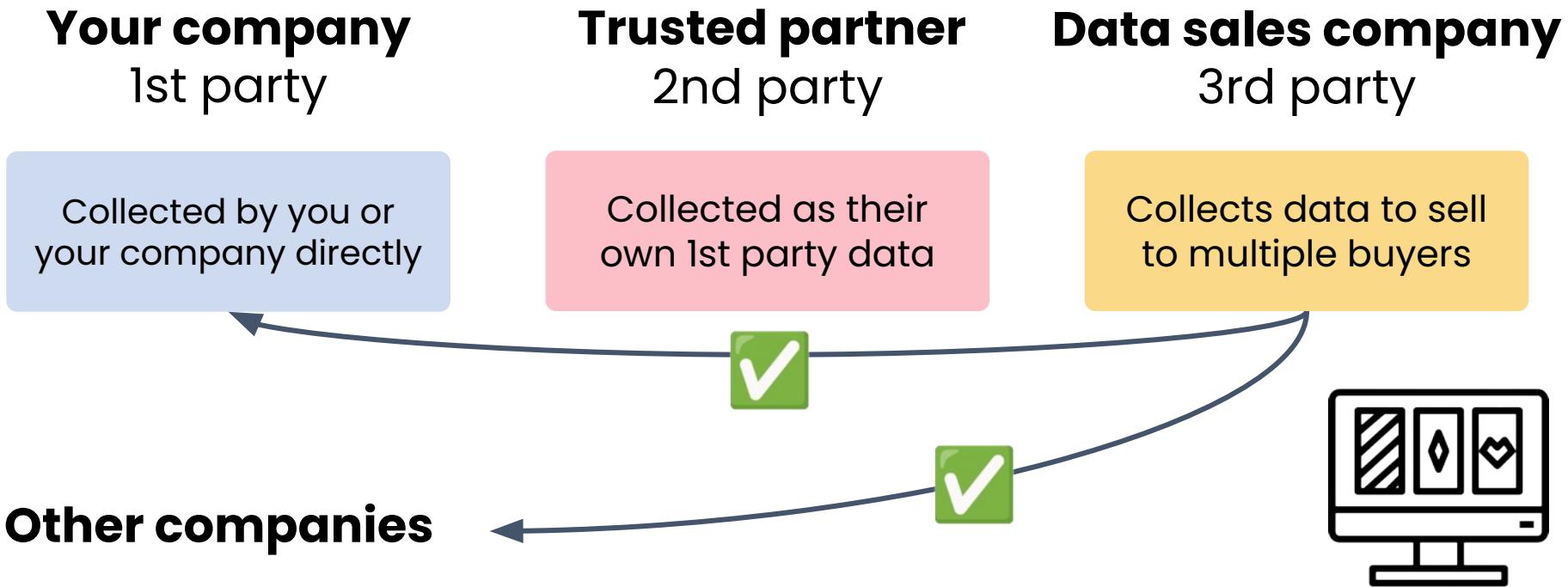
Collected by you or
your company directly

Trusted partner
2nd party

Collected as their
own 1st party data



1st party, 2nd party, and 3rd party data



1st party, 2nd party, and 3rd party data

Your company

1st party

Collected by you or
your company directly

Trusted partner

2nd party

Collected as their
own 1st party data

Data sales company

3rd party

Collects data to sell
to multiple buyers

More
control



Less
control

Publicly available data



Sources

- Government agencies
- Research organizations
- Open-source databases



Usage

- Download directly
- Scrape data with Python



Great resource because

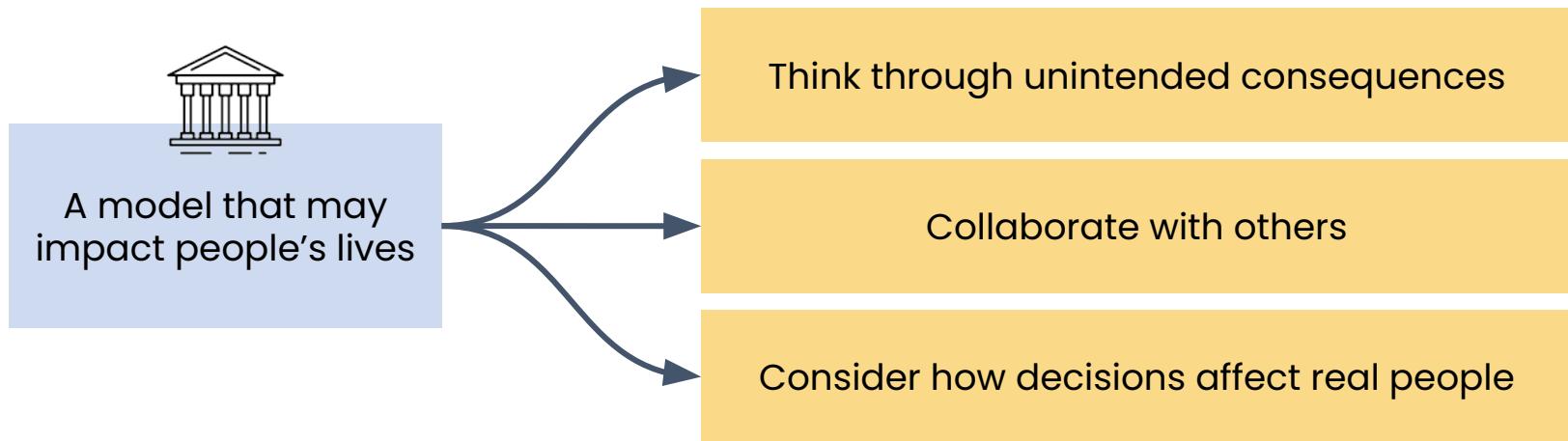
- Free to access
- Real-world data

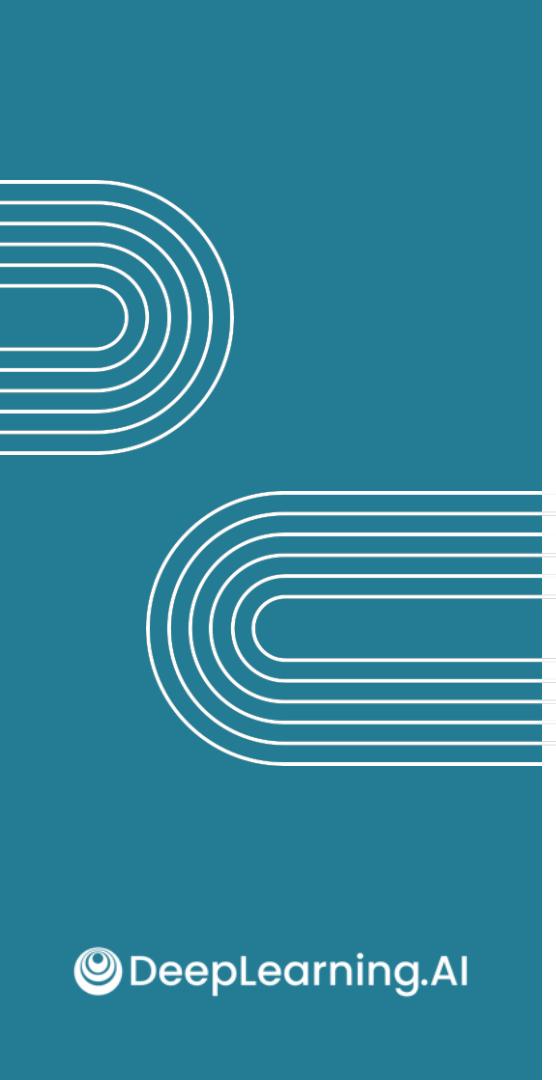
Data privacy

-  Only analyze data you are authorized to access
-  Take training
-  Operate within secure computing environments
-  Strip data of personally identifiable information

Ethical usage

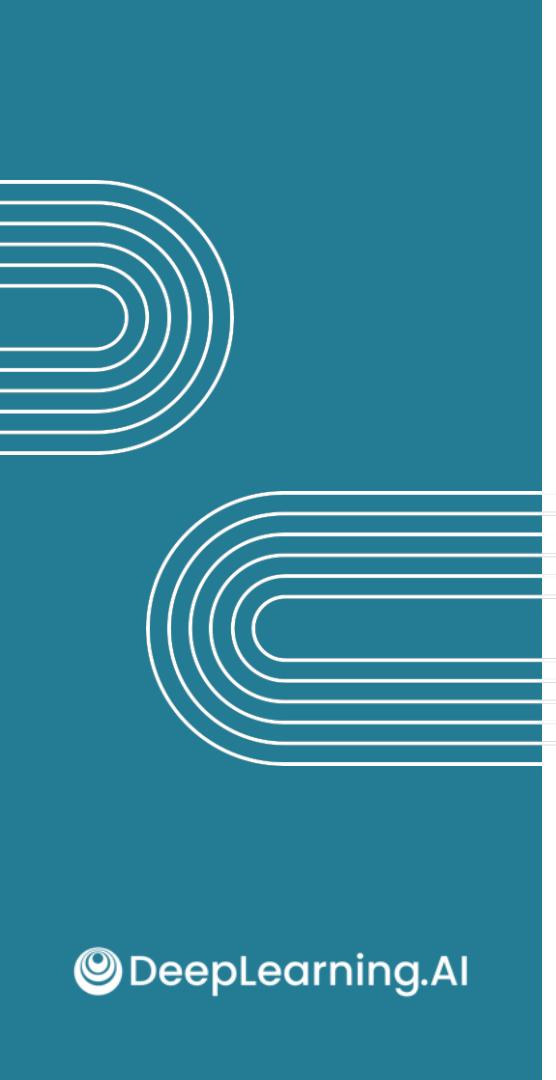
Using the **data** or **insights** you are producing
in a fair and beneficial way





Using spreadsheets for data analytics

Data exploration with LLMs



Using spreadsheets for data analytics

Introduction to time series

Introduction to time series

- Measuring one or more outcomes over fixed periods of time
 - Minutes
 - Hours
 - Days
- Goal: Understand how the outcomes vary over time
 - Identify increasing or decreasing trends
 - Discover unusual events
 - Forecast future outcomes

Common use cases for time series

Measurement	Time interval	Purpose
 Rainfall in the Amazon	Annually	Global warming
 Active users	Weekly	User growth
 Sales	Monthly, quarterly	Revenue forecasting
 Stock market prices	By minute, hourly	Day trading
 Stock market prices	Monthly, quarterly, yearly	Long-term investing

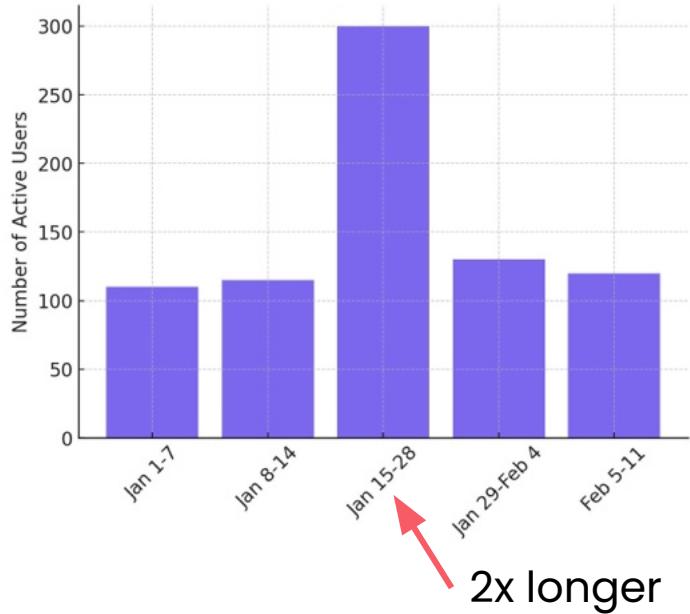
Why have a consistent time interval?

- Interpretability
- Consistent frame of reference
- Focus on comparing the outcome itself



Same data, different interval

Inconsistent interval



Consistent interval



Comparison with cross-sectional data

Type of Data	Measurement	Time Period
Time Series	1	Multiple
Cross-sectional	Multiple	1



- ✓ Focus is typically on each observation as a one-time event
- ✗ Rather than on how those events change over time

Trend

Seasonality

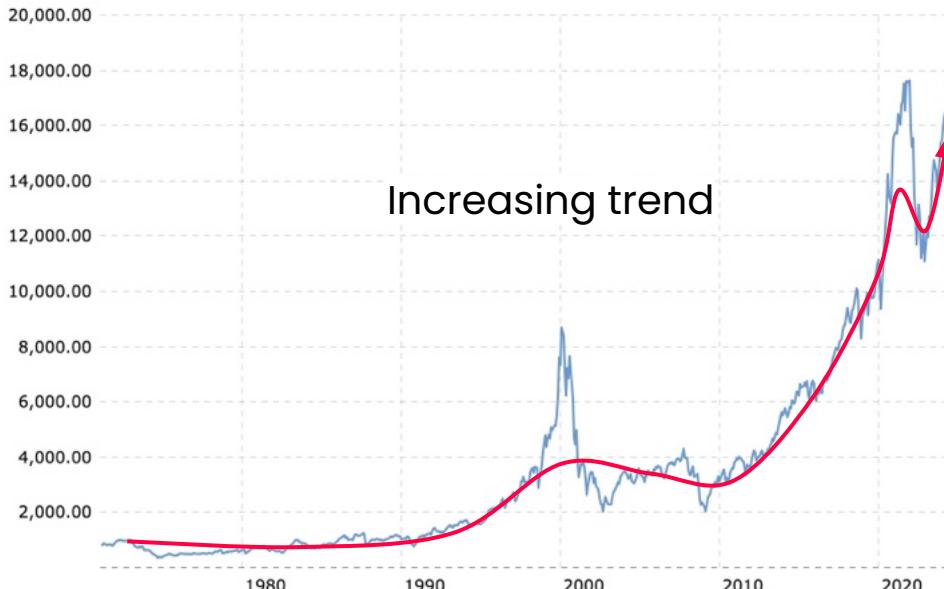
Cyclical fluctuations

Irregularities

Trend: long term direction

- Increasing
- Decreasing
- Stationary

U.S. stock market for past 55 years



[*Nasdaq Composite - 45 Year Historical Chart.* MacroTrends, [www.macrotrends.net/1320/nasdaq-historical-chart.](http://www.macrotrends.net/1320/nasdaq-historical-chart/)]

Trend

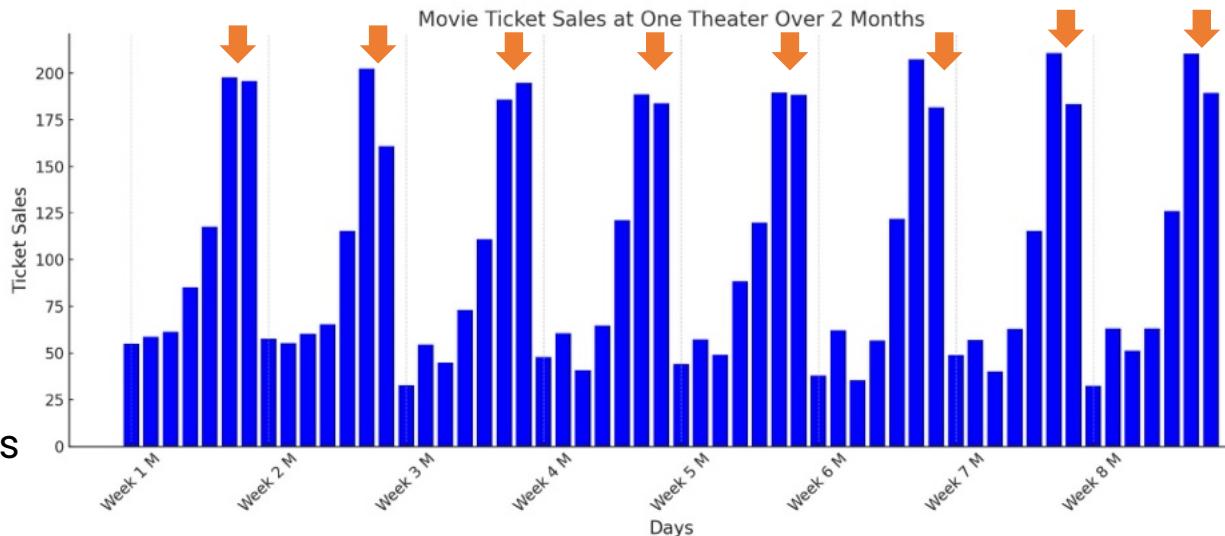
Seasonality

Cyclical fluctuations

Irregularities

Seasonality: a repeating, predictable pattern that occurs at regular intervals

- Daily
- Weekly
- Monthly
- Yearly
- Not just weather seasons
- Can appear at multiple time intervals



Trend

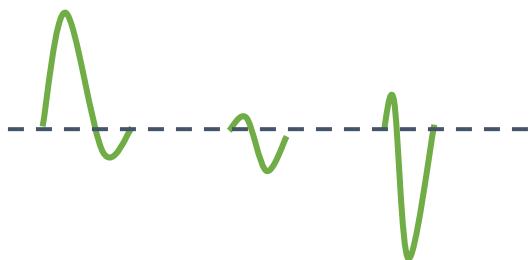
Seasonality

Cyclical fluctuations

Irregularities

Cyclical: repeating ups and downs that don't occur at regular intervals

- Sizes of increases and decreases often not the same
- Harder to predict



U.S. stock market for past 55 years



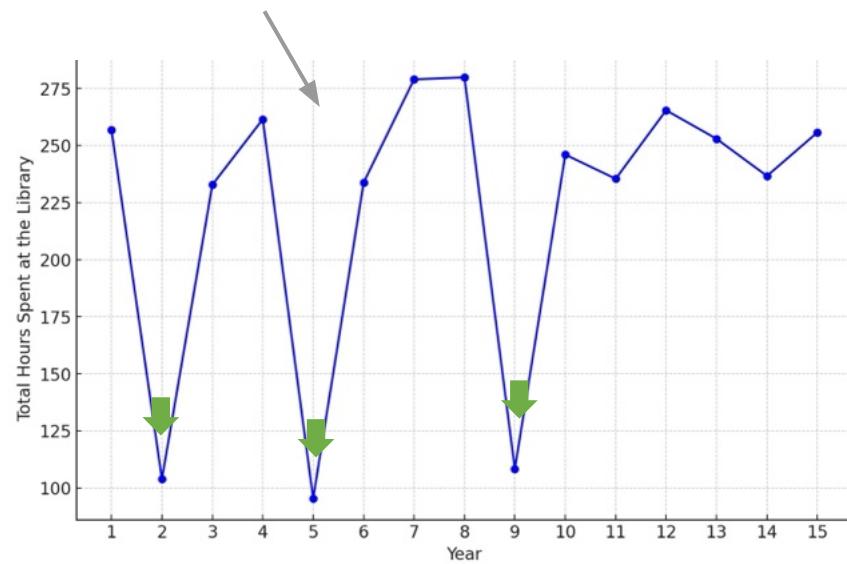
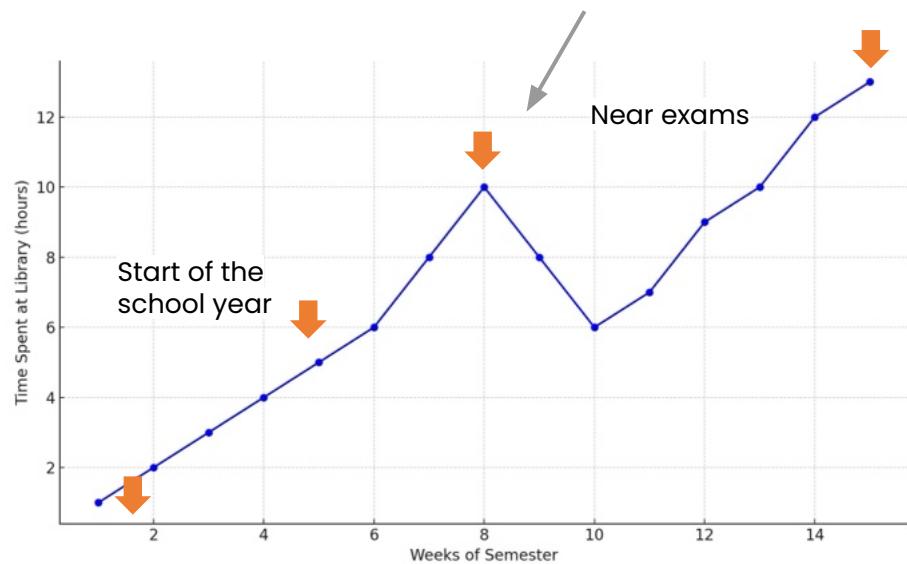
[Nasdaq Composite - 45 Year Historical Chart." MacroTrends, www.macrotrends.net/1320/nasdaq-historical-chart.]

Trend

Seasonality

Cyclical fluctuations

Irregularities



Regular

Unpredictable

Trend

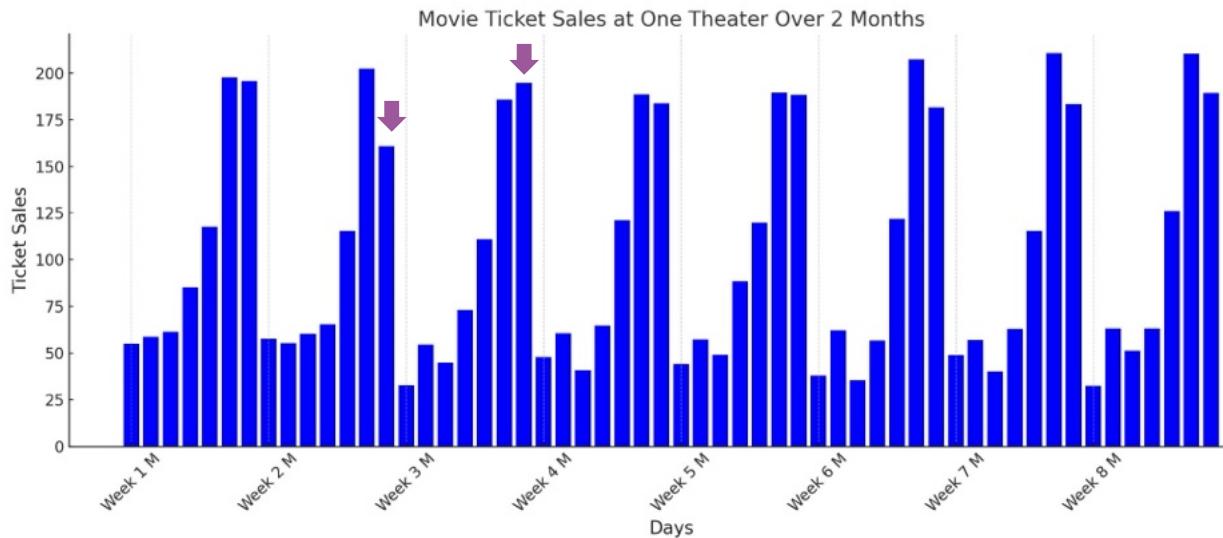
Seasonality

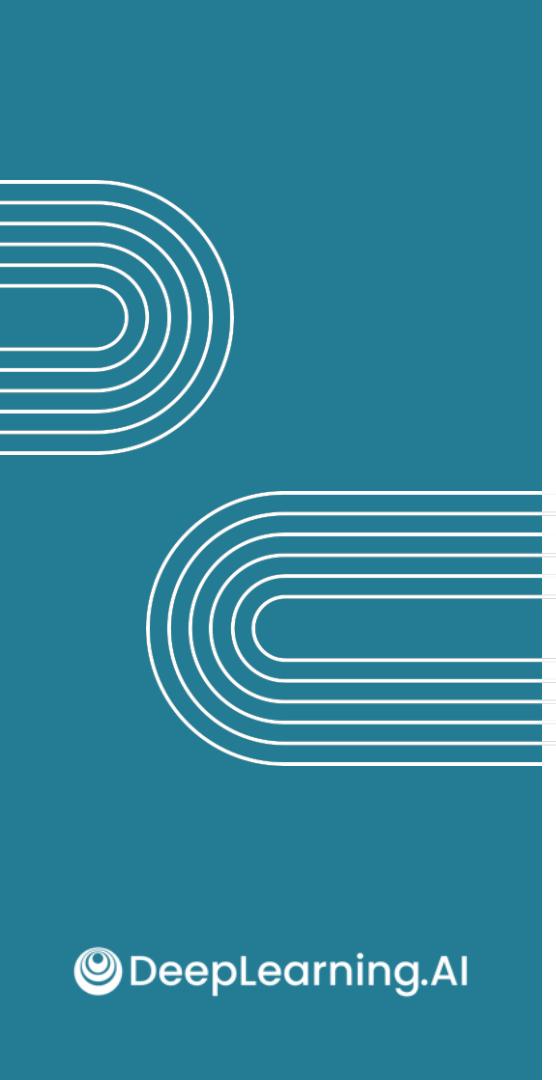
Cyclical fluctuations

Irregularities

Irregularities: random, unpredictable fluctuations in the data

- Also called **noise**
- Makes it harder to spot the trend, seasonality, and cyclicality

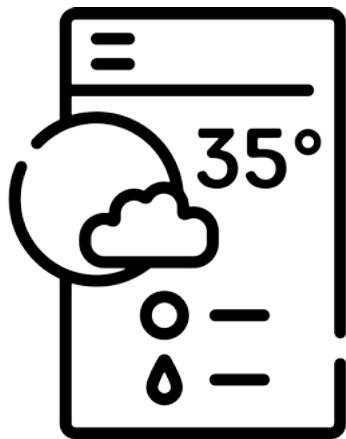




Using spreadsheets for data analytics

Real-world time series

Weather



Time scale

Hours or days



Insights

What should I wear today?

Longer periods



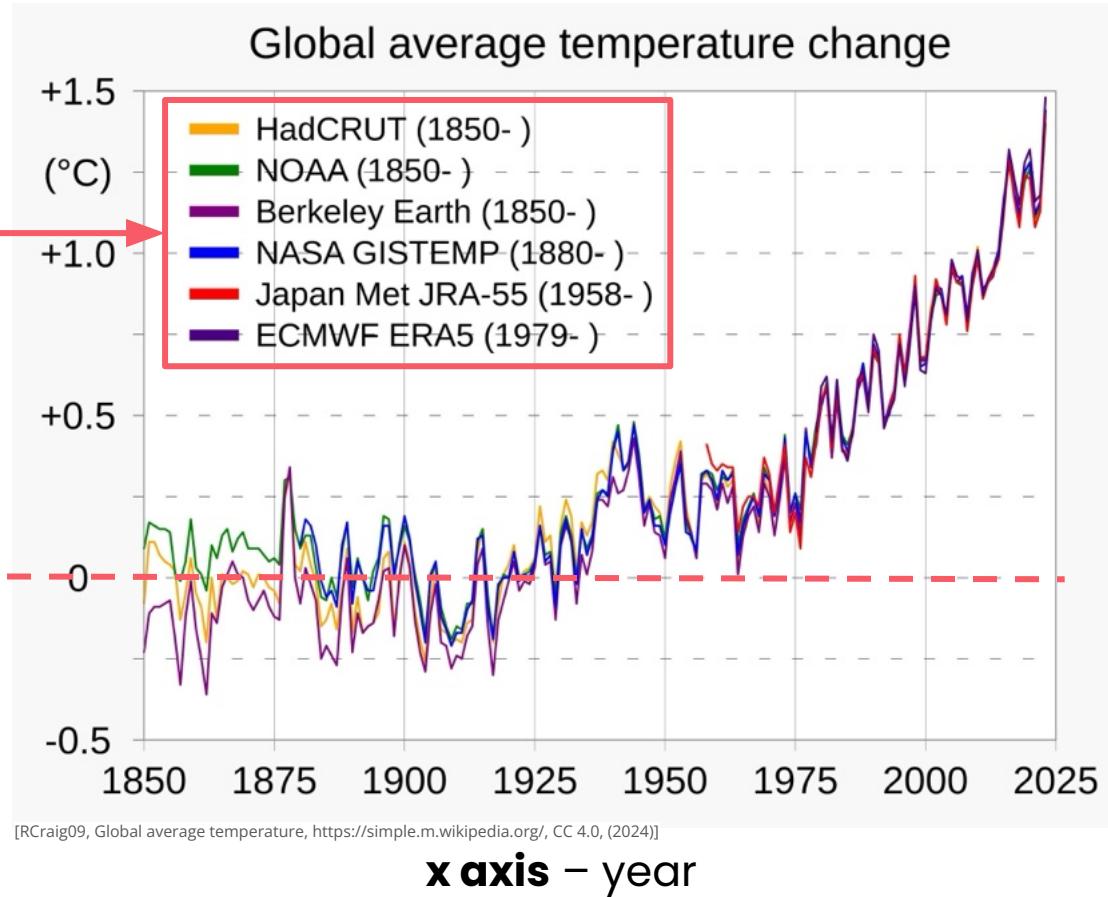
Trends, seasonality,
cyclical

Temperature

Time series

y axis –
temperature

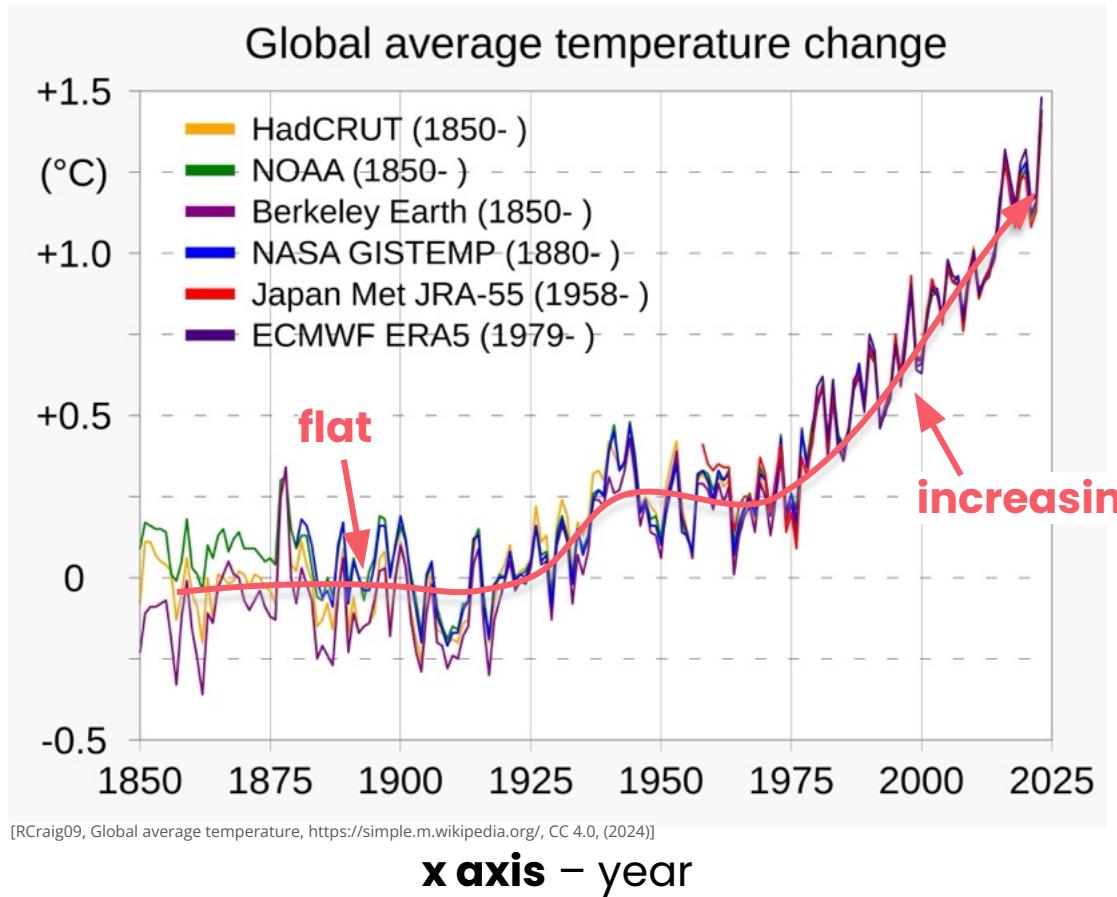
Avg temp
1850 – 1900



Temperature

y axis –
temperature

Trend



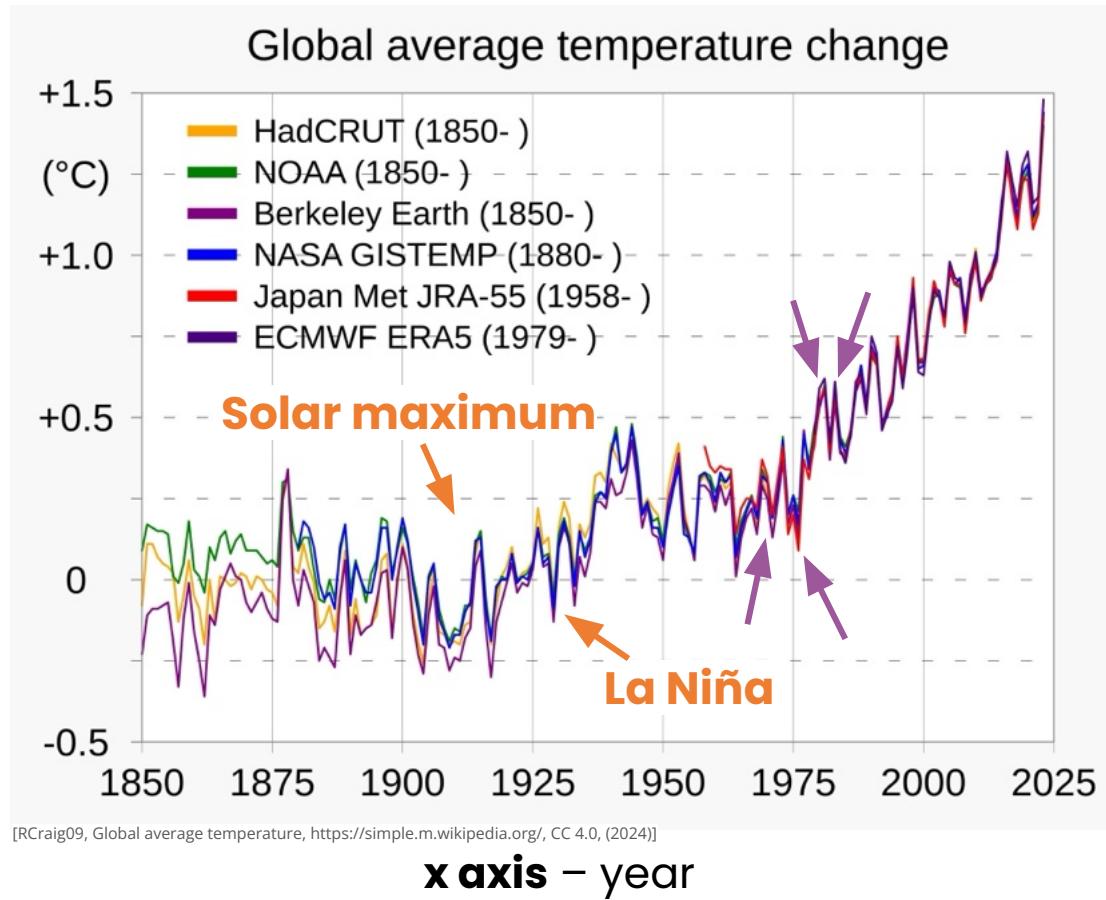
Temperature

y axis –
temperature

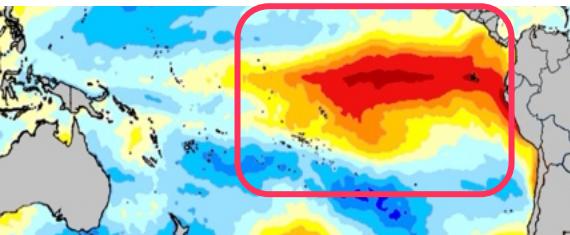
Trend

Seasonality

Noise



El Niño

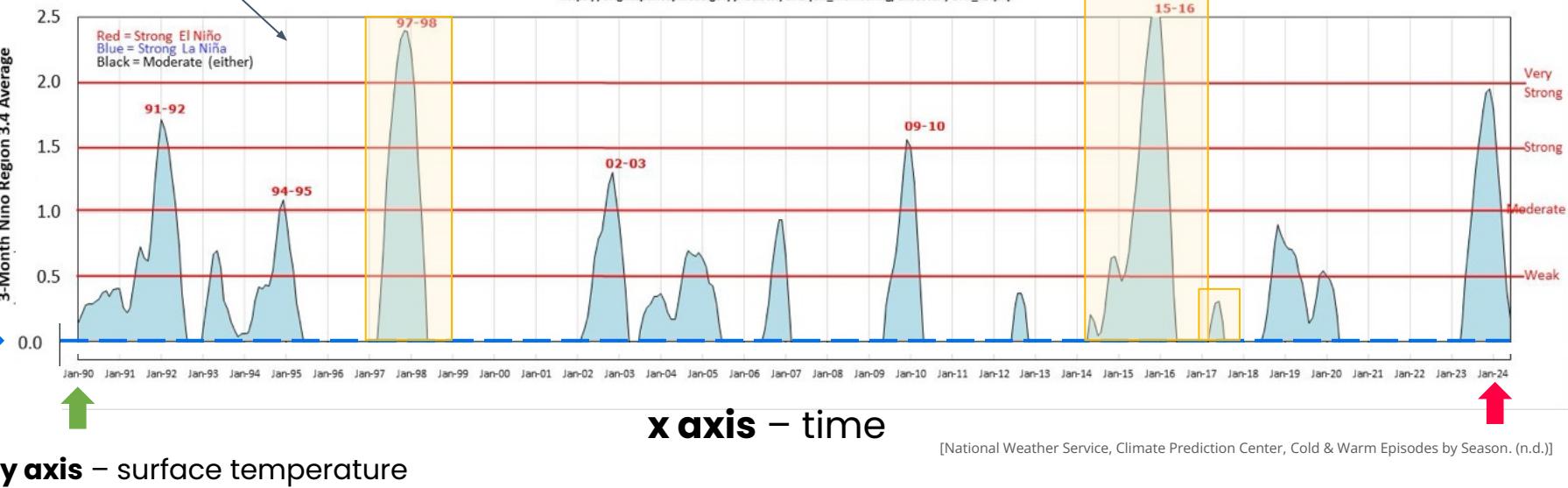


- Caused by wind patterns
- Periodic, but not at fixed intervals
- 9–12 months, but can last for years

vertical line – year

Oceanic Niño Index (ONI) - 1990-present

https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php

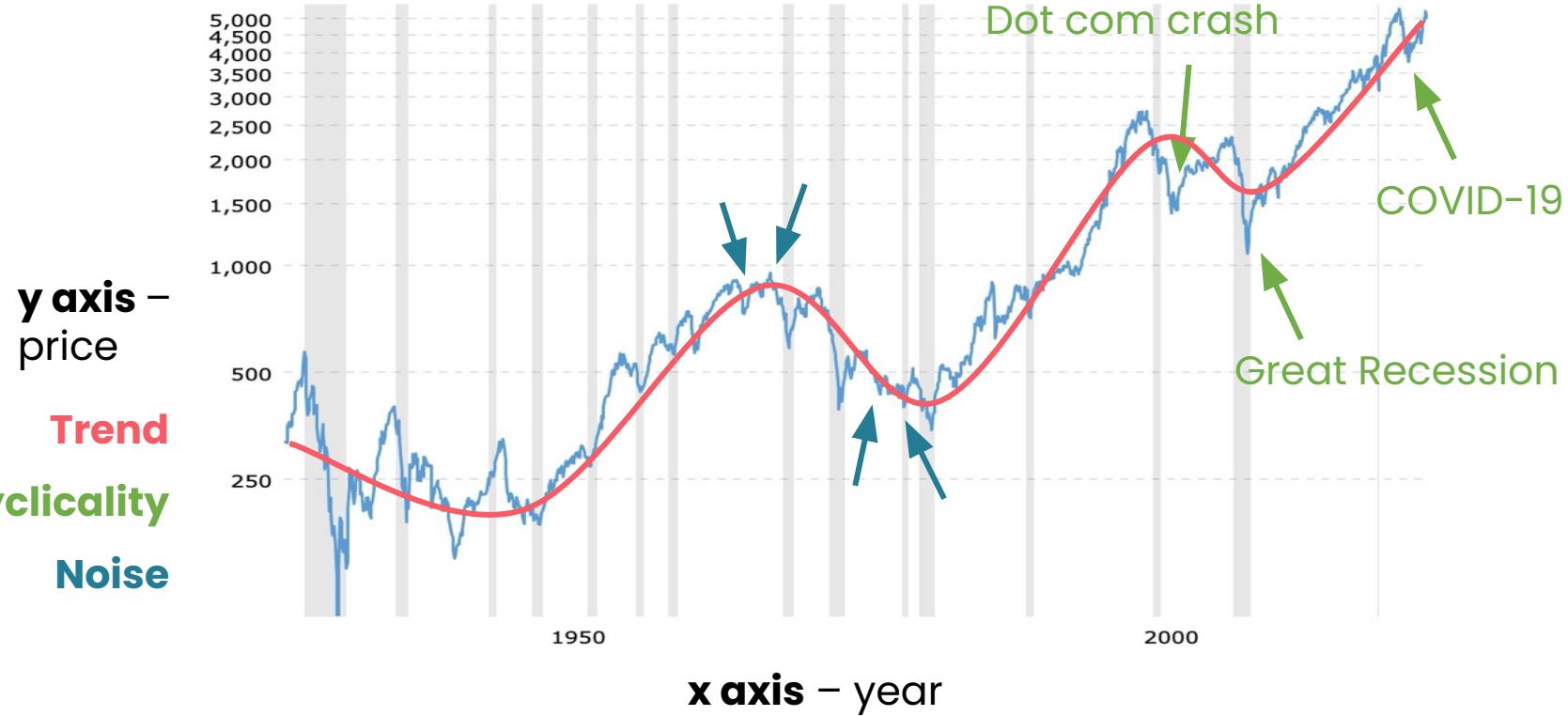


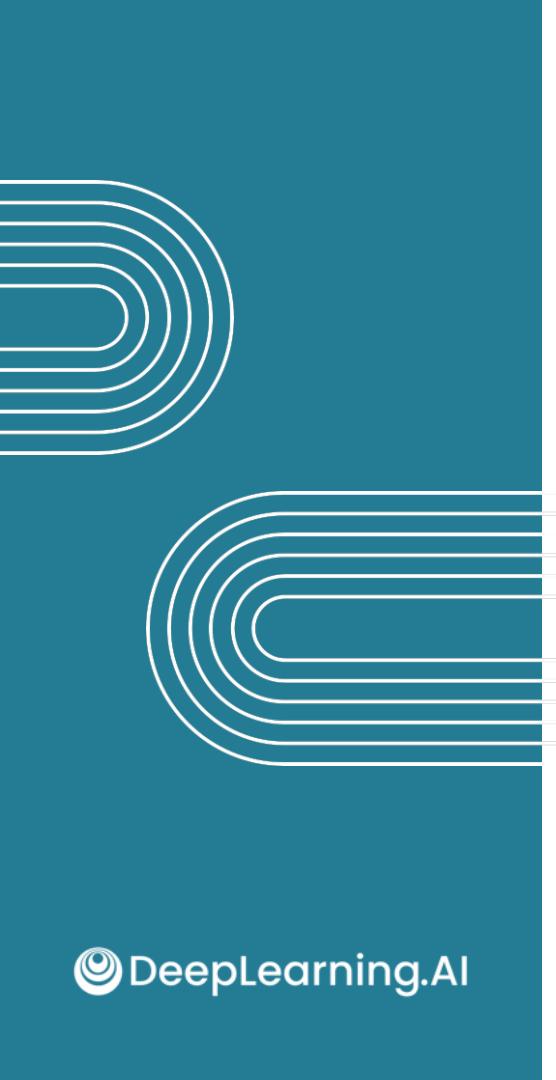
[National Weather Service, Climate Prediction Center, Cold & Warm Episodes by Season. (n.d.)]

Stock market prices - S&P 500



Stock market prices - S&P 500

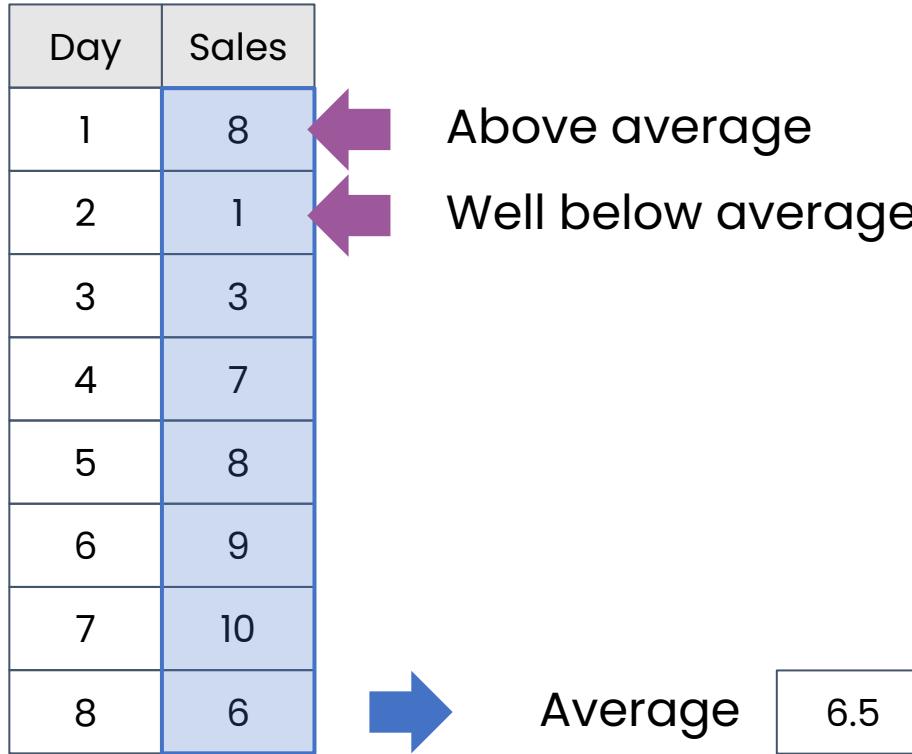




Using spreadsheets for data analytics

Moving averages

Average



Moving average

- Smooth out noisy behavior
- How to calculate a simple moving average:
 - Calculate average of **N** consecutive time periods
 - Calculates a series of values (**N-1** shorter than data)
 - Larger values of **N** are more stable

Day	Sales
1	8
2	1
3	3
4	7
5	8
6	9
7	10
8	6

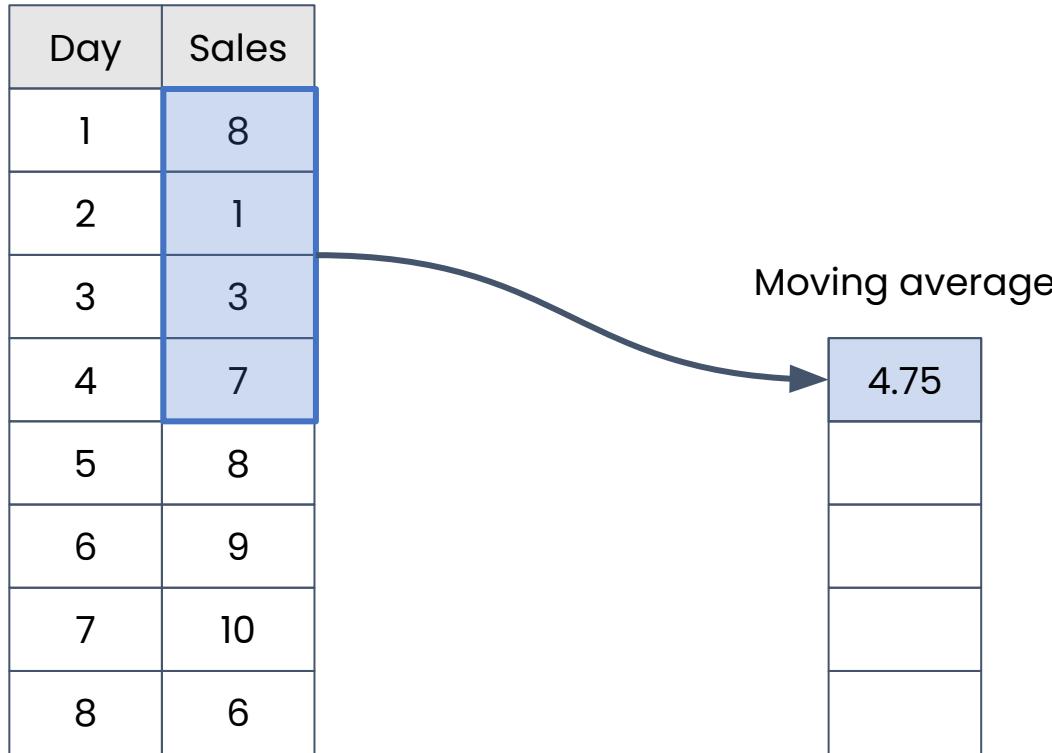
Calculate average for values

Calculate average for values

Calculate average for values

Moving average

N = 4



Moving average

$N = 4$

Day	Sales
1	8
2	1
3	3
4	7
5	8
6	9
7	10
8	6

Moving average

4.75
4.75

Moving average

$N = 4$

Day	Sales
1	8
2	1
3	3
4	7
5	8
6	9
7	10
8	6

Moving average

4.75
4.75
6.75

Moving average

$N = 4$

Day	Sales
1	8
2	1
3	3
4	7
5	8
6	9
7	10
8	6

Moving average

4.75
4.75
6.75
8.50

Moving average

$N = 4$

Day	Sales
1	8
2	1
3	3
4	7
5	8
6	9
7	10
8	6

Moving average

4.75
4.75
6.75
8.50
8.25

Real-world example

- Names tend to follow cycles of popularity

In this exercise:



Work with a dataset of popular US baby names



Predict whether a particular name may see a resurgence



Using spreadsheets for data analytics

Percent change

Percent change

To standardize these differences:

- Convert from original units to percentages:

$$\frac{\text{Sales}_{\text{current}} - \text{Sales}_{\text{previous}}}{\text{Sales}_{\text{previous}}}$$

Day	Sales
1	8
2	1

Difference: -7 sales

$$\% \text{ change} = \frac{-7}{8} \quad -87.5\%$$

Day	Sales
20	108
21	101

Difference: -7 sales

$$\% \text{ change} = \frac{-7}{108} \quad -6\%$$

Percent change

Day before - X_{t-1}

Current day - X_t

$$\frac{X_t - X_{t-1}}{X_{t-1}} \times 100$$

% change → positive

% change → negative

Time series analysis: Percent difference from period to period

$t = 2$

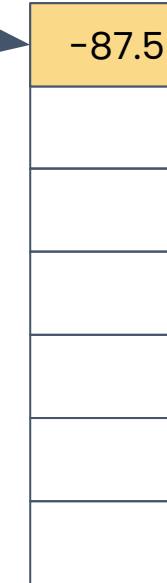
Day	Sales
1	8
2	1
3	3
4	7
5	8
6	9
7	10
8	6

x_{t-1}

x_t

$$\frac{1 - 8}{8} \times 100$$

Percent change



Time series analysis: Percent difference from period to period

$t = 3$

Day	Sales
1	8
2	1
3	3
4	7
5	8
6	9
7	10
8	6

X_{t-1}
 X_t

$$\frac{3 - 1}{1} \times 100$$

Percent change

-87.5
200

Time series analysis: Percent difference from period to period

$t = 4$

Day	Sales
1	8
2	1
3	3
4	7
5	8
6	9
7	10
8	6

X_{t-1}
 X_t

$$\frac{7 - 3}{3} \times 100$$

Percent change

-87.5
200
133.3

Time series analysis: Percent difference from period to period

$t = 5$

Day	Sales
1	8
2	1
3	3
4	7
5	8
6	9
7	10
8	6

X_{t-1}
 X_t

$$\frac{8 - 7}{7} \times 100$$

Percent change

-87.5
200
133.3
14.3

Time series analysis: Percent difference from period to period

$t = 6$

Day	Sales
1	8
2	1
3	3
4	7
5	8
6	9
7	10
8	6

X_{t-1}
 X_t

$$\frac{9 - 8}{8} \times 100$$

Percent change

-87.5
200
133.3
14.3
12.5

Time series analysis: Percent difference from period to period

$t = 7$

Day	Sales
1	8
2	1
3	3
4	7
5	8
6	9
7	10
8	6

X_{t-1}
 X_t

$$\frac{10 - 9}{9} \times 100$$

Percent change

-87.5
200
133.3
14.3
12.5
11.1

Time series analysis: Percent difference from period to period

$t = 8$

Day	Sales
1	8
2	1
3	3
4	7
5	8
6	9
7	10
8	6

X_{t-1}
 X_t

$$\frac{6 - 10}{10} \times 100$$

Percent change

-87.5
200
133.3
14.3
12.5
11.1
-40.0