AP CSPTopic 2.3: Extracting Information from Data

Big Idea 2 — Data

Programs can process data to help us discover information and create new knowledge.

Part A — Understanding Data and Information (10 pts)

1- What is the difference between **data** and **information**?

Data is raw facts or numbers that haven't been interpreted yet, while information is what we learn after analyzing that data to find its meaning or patterns.

2- Give one example of **data** and the **information** you could find from it.

Example: Data → "3, 5, 7." Information → "The numbers are odd."

Data → "60F, 62F, 65F" Information → "The temperature is rising over time."

3- How can computers help us find **patterns or trends** in data?

Computers can process large datasets quickly, create graphs, and use algorithms to detect patterns.

4- What does **correlation** mean?

Correlation means that two things change in a related way, when one changes, the other changes too.

Give one example of two things that are correlated but not caused by each other.

Ice cream sales and the number of people swimming are correlated because both increase in summer, but one doesn't cause the other.

5- Why is it important to remember that "correlation does not mean causation"?

Because just because two things happen together doesn't mean one causes the other, they might be linked by coincidence or another factor.

Part B — Metadata (10 pts)

6- What is **metadata**?

Metadata is data that describes other data.

7- Give two examples of metadata that an image might include.

File size and date the photo was taken.

Look at two images:

- Image A = 3.2 MB
- Image B = 400 KB
  Which one is more compressed? How can you tell?

Image B is more compressed because its file size is smaller, meaning more data was removed to make it smaller.

9- Does changing metadata (like file name or date) change the image itself? Explain.

No, changing metadata doesnt affect the image content, it just changes information about the image.

10- How can metadata help a photo app organize your pictures automatically?

A photo app can use metadata like date, time or location to automatically sort and group pictures into albums or by events.

Part C — Processing and Cleaning Data (10 pts)

1- What problems might happen if your data includes missing or repeated values?

It can make results inaccurate, cause errors in analysis, or lead to false conclusions.

2- What does it mean to **clean data**?

Cleaning data means fixing or removing errors, duplicates or missing information so its ready for analysis.

3- Why is cleaning data important before analyzing it?

Because clean data ensures accurate results and helps avoid misleading patterns or mistakes in interpretation.

Part D — Bias, Scale, and System Limits (10 pts)

4- Why can big data sets be hard for one computer to process?

Because large datasets take up lots of memory and processing power, one computer might not have enough storage or speed to handle it all efficiently.

5- Give one example of how data collection could create **bias**.

If a survey only includes people from one city or age group, it could lead to biased results that don't represent everyone.

Part E — Reflection (10 pts)

6- What was the most interesting thing you learned about data or metadata?

I found it interesting that metadata can tell so much about a file without even opening it, like when or where a picture was taken.

7- Why is data cleaning important for getting accurate results?

Because errors, duplicates, or missing data can change the outcome and lead to incorrect conclusions.

8- How can metadata help us organize digital information?

Metadata helps categorize files by type, size or date making it easier to search, sort and manage information automatically.

9- What's one example of bias that could happen in real life?

A facial recognition system might work better for some skin tones than others if the training data wasnt diverse enough.

10-How does processing data help humans make better decisions?

Processing data turns raw numbers into useful insights, helping people spot trends, solve problems, and make informed choices in business, science and everyday life.