



The James
Hornsby School

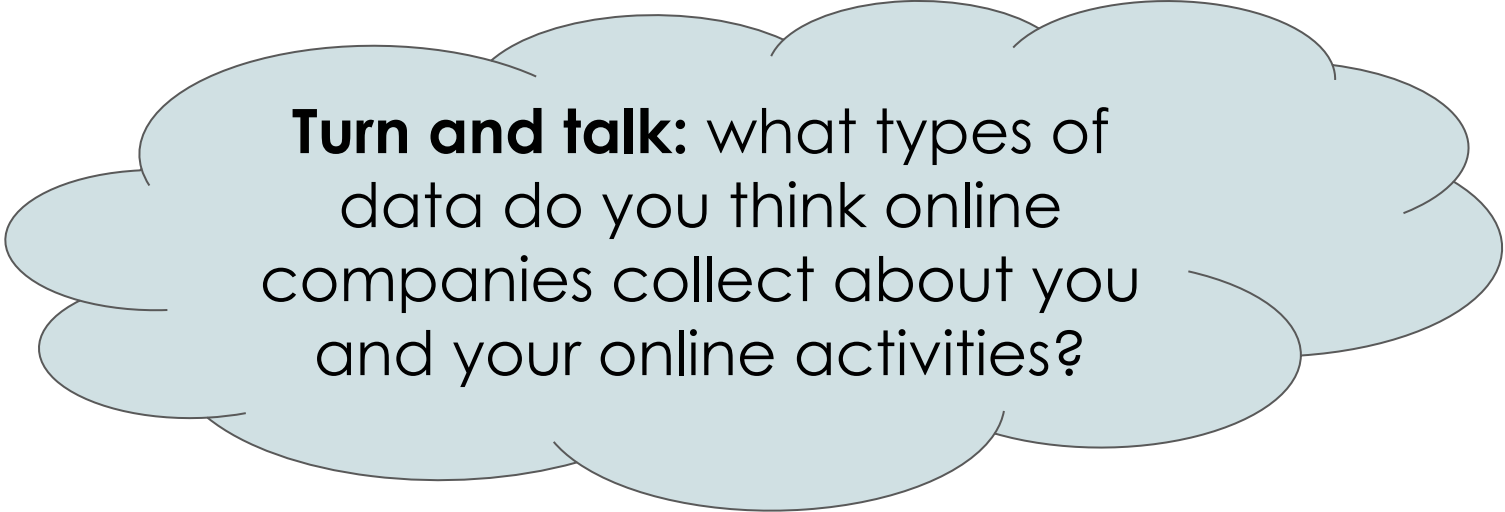
Together we excel

- Everyone has the right to be heard and respected.
- We will use language that won't offend or upset other people.
- We will use the correct terms, and if we don't know them, we'll ask the teacher.
- We will comment on what was said, not the person who said it.
- We won't share our own, or our friends', personal experiences.
- We won't put anyone on the spot, and we have a right to pass.
- We won't judge or make assumptions about anyone.

How is my online data collected and used?

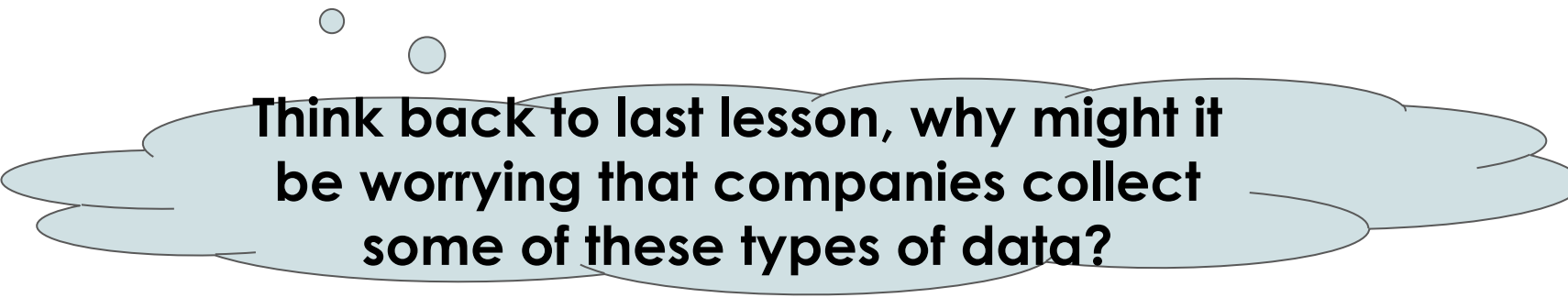
Do now:

1. Describe the relationship between age and fertility
2. Explain one possible route to parenthood for people who struggle to conceive
3. What kind of financial exploitation am I describing if I talk about people willingly allowing their bank accounts to be used to move criminal money?



Turn and talk: what types of data do you think online companies collect about you and your online activities?

- Demographic information (age, gender, job)
- Contact details
- Location
- Search history
- Purchases
- Content you've streamed or downloaded
- Adverts you have watched or clicked (we are going to learn more about this next lesson)
- Mouse movements
- Conversations
- Computer characteristics (browser, battery life, screen size)



Think back to last lesson, why might it be worrying that companies collect some of these types of data?

Big data: Large data sets that are analysed in order to identify patterns and trends, often used by organisations to better understand their customers.

Big data: Large data sets that are analysed in order to identify patterns and trends, often used by organisations to better understand their customers.

Big data is characterised by the three Vs:

- **Volume:** the amount of data in the data set
- **Variety:** the different types of data in the data set
- **Velocity:** the speed at which the data set is produced and how long it remains accurate



Turn and talk: why do you think
companies or organisations
want to collect data from you?



Turn and talk: why do you think companies or organisations want to collect data from you?

- Improve and streamline services
- Target advertisements towards customers
- Optimise project management
- Identify criminal activity
- Predict behaviour

Improve and streamline services

Companies can analyse customer reviews and behaviour to determine what parts of their service are liked and what needs to change. They can also work out when customers are mostly likely to use their service, to ensure they can meet demand.

Turn and talk: Can you think of any companies/organisations that might use data in this way?

Target advertisements towards customers

Companies can analyse the buying behaviour of their customers and then advertise similar products, to improve their adverts' effectiveness.

Optimise project management

When planning a project, companies need to determine how long it will take and how much it will cost. They can use big data to reduce both of these by working out the quickest and cheapest way of carrying out their project.

Identify criminal activity

Big data sets that show customer behaviour can reveal peculiar activity, such as purchases of an unexpected item, or card use in a different country. This can be used to help identify fraudulent activity.

Predict behaviour

Big data sets can also be used to predict what people will do next. Data sets can be created that show how users have behaved in the past or how the people around them have behaved, and can then be used to guess what will happen next. This can be particularly useful when trying to analyse how groups of people are likely to vote in matters of elections for example.

Data analytics

Big data sets are analysed using data analytics. These are the techniques used to combine different types of data and extract meaning from them.

This might involve searching for common patterns in a data set, or trying to characterise the behaviour or certain types of data subject.

Data analytics might involve:

- Machine learning
- Text analysis
- Predictive analysis
- Optimisation problems
- Cleaning and combining data sets

Print out and work through this sheet as a class

Data set 1

Excerpt from a data set collected by an online streaming service, StreamNow. They provide a range of films and TV at a monthly cost.

User ID	Subscription service	Favourite genre	Average watch time	Gender
00011	Premium	Romance	2 hours a day	Female
00012	Basic	Documentary	0.5 hours a day	
00013	Free	Comedy	1 hours a day	Male
00014	Student	Horror	2 hours a day	Female
00015	Premium	Family	3 hours a day	Female
00016	Premium	Romance	2 hours a day	Male
00017	Basic	Comedy	0.25 hours a day	Female
00018	Free	Romance	4 hours a day	

1. You work for StreamNow: What patterns do you notice in user behaviour?

Data set 2

Excerpt from a data set collected by an online shopping site, BuyOnline. BuyOnline is an online marketplace that facilitates online shopping. There are different levels of membership, which provide additional benefits such as discounts and free postage and packing.

User ID	Account	Browsing time	No. of purchases per month	% access via adverts	Amount spent (£)
00011	Guest	5 mins per purchase	1	100	10
00017	Premium	30 mins per purchase	5	20	150
10102	Premium	20 mins per purchase	2	0	17.98
10103	Basic	0 mins per purchase	1	100	14.99
00104	Student	1 hour per purchase	2	50	5
10105	Guest	2 mins per purchase	3	66	40
10106	Student	45 mins per purchase	6	33	100
00016	Premium	1 hour per purchase	8	0	200

2. You work for BuyOnline; which types of users are spending the most money on the site?

1.You work for StreamNow: What patterns do you notice in user behaviour?

- Premium users tend to use the service more than other account holders.
- Users who use the service for several hours a day tend to watch romances.

2.You work for BuyOnline; which types of users are spending the most money on the site?

- Premium users tend to spend more time browsing before making a purchase, make more purchases, and spend more money on the site.
- Basic users are more likely to enter the site via an advert.

3. a) You work for BestAnalytics; what insights (information) can you offer StreamNow to help them improve their services?

- User 00017 doesn't like the current content and may need more comedy films and TV to encourage them to stay a member.
- Romance content appears to be popular; if it were only available to basic or premium members, this might encourage users to pay a subscription fee.
- User 00014 could be targeted with an advert explaining that an account on StreamNow makes a great birthday gift.

3.b) You work for BestAnalytics; what insights (information) can you offer BuyOnline to help them increase their number of premium customers?

- If user 10104 were advertised some bowling shoes, they might purchase them. However, they have not spent much money on the site so the shoes advertised should be cheap.
- Users 00011 and 10105 could be encouraged to create accounts, especially as user 10105 has made several purchases and 00011 has a premium account on StreamNow (and therefore may be interested in the benefits of being a premium member).
- Users 00011 and 00016 like romance films and user 00017 likes comedy. They could be targeted with adverts that appealed to other customers with similar tastes.
- User 00014 could be targeted with adverts for great gifts for mums.



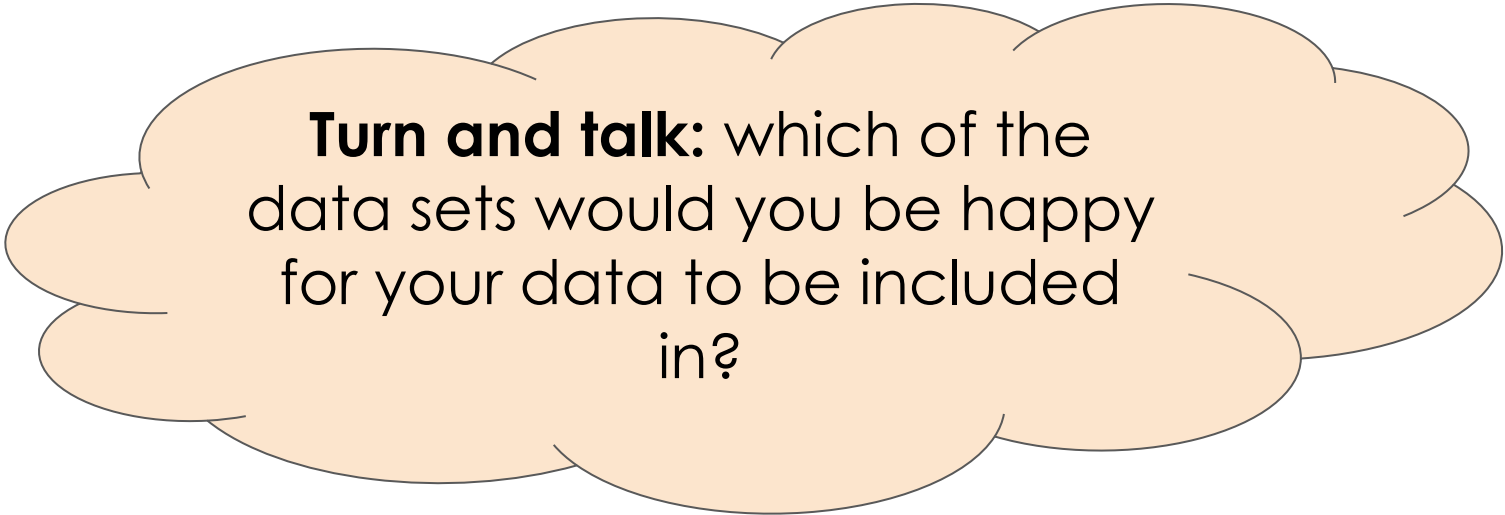
Turn and talk: how do companies
collect your 'big data'?

How is big data collected?

- Asking for it (when you create an account or buy something)
- Cookies (track web pages you visit and when)
- Social media/app activity
- Online marketing/shopping
- Email tracking
- Loyalty programs and cards
- Gameplay
- Satellite imagery

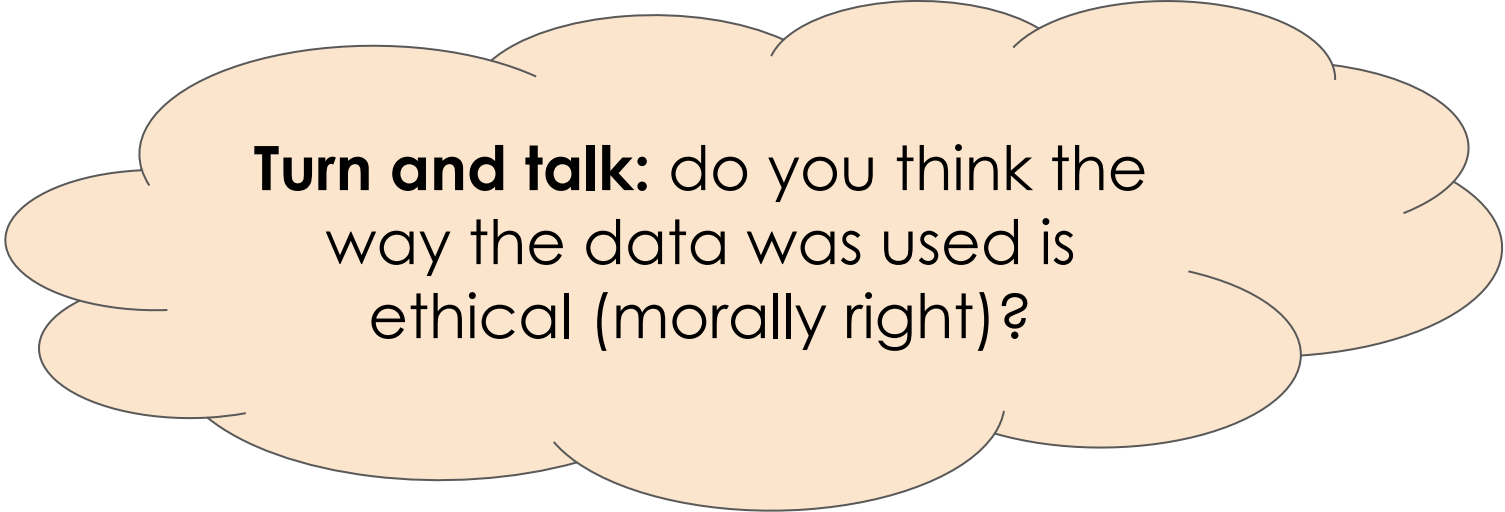


Thinking about the data sets you have just analysed:



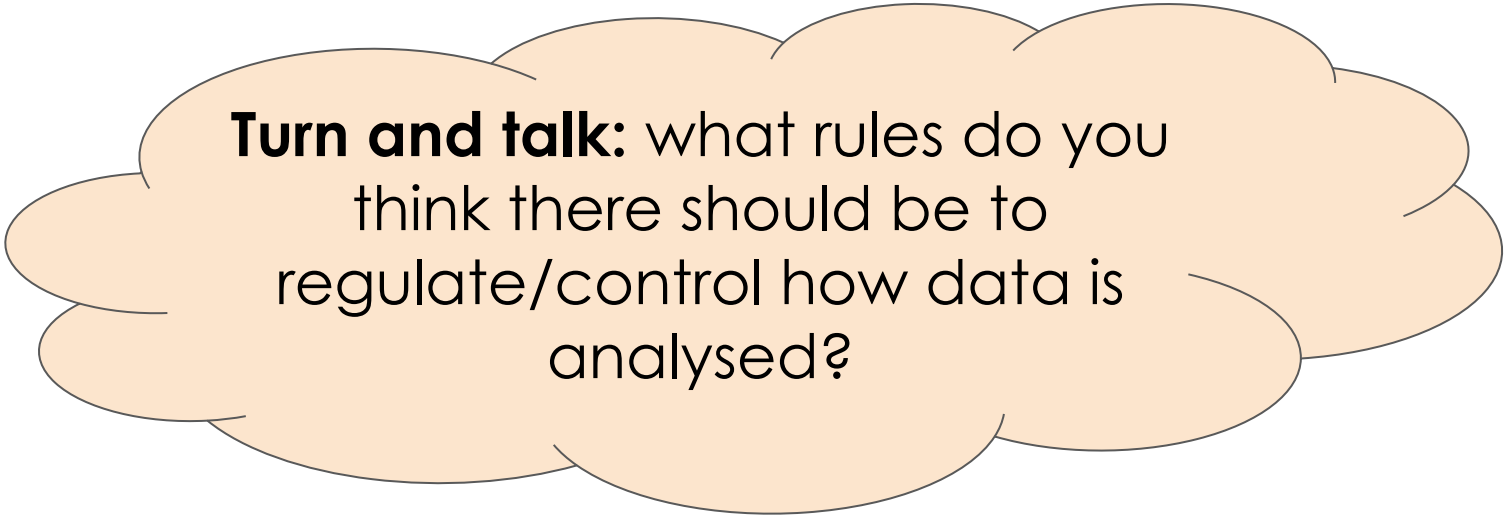
Turn and talk: which of the
data sets would you be happy
for your data to be included
in?

Thinking about the data sets you have just analysed:



Turn and talk: do you think the way the data was used is ethical (morally right)?

Thinking about the data sets you have just analysed:



Turn and talk: what rules do you think there should be to regulate/control how data is analysed?

Big data sets have sometimes been misused...

Privacy violation

Big data sets often contain lots of personal information about their data subjects. For example, Uber collects the location data of passengers. In 2014, it was discovered that an executive had used this data to track the location of someone they were meeting.

Big data sets have sometimes been misused...

Discrimination

Big data is used to make the adverts you see more relevant to you. However, in 2015, it was shown that Google was advertising different jobs to men and women and that the jobs advertised to men were more highly paid.

Big data sets have sometimes been misused...

Inaccuracies

Big data is used to streamline services and improve the decisions made by companies. However, this only works if the data in the big data sets is accurate. In 2017, an Australian debt collecting company was accused of falsely telling people that they owed large sums of money. The company used an automated system that analysed large data sets to identify who owed money, and how much. However, errors in the system led to the wrong people being contacted and the wrong debts calculated.

Big data sets have sometimes been misused...

Inaccuracies

Big data sets can be valuable and contain valuable personal information. They can therefore be the targets of theft. Data subjects rely on the organisation managing the data set to protect their information, but thefts do occur. In 2018, the app MyFitnessPal was hacked and data on 150 million users was stolen. As data sets are often sold, it may also not be clear who is storing your data.

Collecting and using big data is ethical

AGREE ←————→ DISAGREE

Collecting and using big data is ethical

Do you agree with this statement?

Explain your opinion but also try to explain why someone might disagree with you (explain the other side of the argument)

Disagree

Violation of privacy
Targeted advertising
Predict behaviour
Discrimination
Inaccurate
Data theft/fraud

Possible arguments you
could use for either side
in your answer



Agree

Improve services and customer experience
Identify criminal activity