**Google**

Background:

More than any other company, Google are probably responsible for introducing us to the benefits of analysing and interpreting Big Data in our day-to-day lives. When we carry out a Google search, we are manipulating Big Data. The size of Google’s index – its archive of every Web page it can find, which is used to return search results – is estimated to stand at around 100 petabytes (or 100 million gigabytes!) – certainly Big Data, by anyone’s standards.1 But as we’ve seen over the past decade, bringing all the information on the Internet together to make it easier for us to find things was only the start of their plan. Google have gone on to launch Web browsers, email, mobile phone operating systems, map services and the world’s biggest online advertising network – all firmly grounded in the Big Data technology with which they made themselves a household name.

Data/Problem current have:

The Internet is a big place – since we moved online in the 1990s, it’s been growing at a phenomenal rate and is showing no signs of slowing down. This size itself is a problem: when we have access to practically everything that anyone has ever known, how do we find what we need to help us solve our problems? Not only is it big, the Internet is very widespread. Information is uploaded to servers that may be located anywhere in the world, meaning anyone wanting to browse through the data available to them is connecting to computers which are sometimes linked thousands of miles apart from each other. Getting individual bits of specific data through to the user doesn’t take long, with the speed at which information can travel along copper or fibre-optic cables – a matter of seconds. But that supposes the user knows where the data is located in the first place. Searching the entire Internet even for a very simple piece of information, if you didn’t know the precise IP address of the computer on which it was stored would take a very, very long time if you didn’t have an index. With billions of pages of information available online, though, building an index isn’t trivial. It would take an army of humans an eternity to come up with anything approaching a comprehensive database of the Internet’s contents. So it had to be done automatically – by computers. This raised another problem: how would computers know what was good information and what was pointless noise? By default, computers can’t determine this on their own: they have no concept of the difference between useful and useless, unless we teach them and, anyway, what’s useless to one person may be critical to another person in order to solve their problems. Apart from this, Google has a diverse set of data (e.g., email, map, voice, mobile phone data) and it is expanding its company to many new business involving new form data (e.g., driverless car and Google Home). How these data can be used is also a consideration for Google’s future planning.