1. What is the 1994 rate of juvenile delinquency in the US?

This seems like a good question and could be answered with a census of a small, randomly selected, population of delinquency centers in the US and the whole value extrapolated to the population.

1. What can we do to reduce juvenile delinquency in the US?

This is a bad question, because it is very nonspecific. It’s hard to imagine how data science can be used to answer this. I would change it to some specific techniques that could be tested.

1. Does education play a role in reducing juvenile delinquents’ return to crime?

This is a good question and can be answered with an experiment (like an AB test) that could compare the rates of return to delinquency in juveniles who had extra education during their corrective time compared to those who did not have extra educational opportunity.

1. How many customers does AT&T currently serve in Washington, DC?

This is a good question and can be answered by taking a random sample of various counties or neighborhoods in DC and identifying AT&T customers through a survey.

1. What factors lead consumers to choose AT&T over other service providers?

This is not a very good questions, because it is nonspecific. There could be literally hundreds of different factors that may influence which provider someone selects. I would suggest narrowing this down to a few factors and then administering a survey to a group of customers selected at random to figure out which factors drove their decision making.

1. How can AT&T attract more customers?

The way this question is phrased right now makes it more of a marketing strategy question. It’s not very good in terms of data science. I would change it to a couple of factors or strategies that then could be tested with an AB test to see which one is the best at attracting customers.

1. Why did the Challenger Shuttle explode?

This is not a good question, because it’s very nonspecific. I would suggest selecting a few characteristics of the Challenger Shuttle to model.

1. Which genes are associated with increased risk of breast cancer?

This is a good question. You could mine the genetic data that is usually publicly available from the NIH and compare the genetic profiles of individuals with and without cancer.

1. Is it better to read to children at night or in the morning?

This is not a very good question, because it’s not clear what ‘better’ means. This needs to be reworked so that some more specific outcome can actually be measured. For example, does reading to children in the morning vs at night improve their test scores?

1. How does Google’s search algorithm work?

This does not really seem like a data science question, it’s more of a research question on the algorithm behind Google.