1. You work at an e-commerce company that sells three goods: widgets, doodads, and fizzbangs. The head of advertising asks you which they should feature in their new advertising campaign. You have data on individual visitors' sessions ([activity on a website](https://en.wikipedia.org/wiki/Session_%28web_analytics%29" \t "_blank), [pageviews](https://en.wikipedia.org/wiki/Page_view" \t "_blank), and purchases), as well as whether or not those users [converted](https://en.wikipedia.org/wiki/Conversion_marketing" \t "_blank) from an advertisement for that session. You also have the cost and price information for the goods.

The goal is to find out whether advertising should use widgets, doodads, or fizzbangs in their new campaign. We can start by using page views of the three goods to compare interest in them. A product with lower page views may turn out to be a good candidate for advertisements.

Next, we can leverage the activity data to find out the conversion rate of a page view to a purchase. A good with a lower page view, but healthy conversion rate can definitely benefit from advertisement and exposure.

We can also use activity data to find out session length and see how long a potential customer spends deciding whether or not to buy. If many customers are on the fence, the good may be a good candidate for advertisement.

Past performance on advertising campaigns is also important. If revenue (# of sales \* cost) greatly increases from an advertisement, it may be a good candidate for additional advertisement.

Additionally, we can use Bayes’ Theory to find:

P(Buy|Ad Seen) = P(Ad Seen|Buy) \* P(Buy) / P(Ad)

Separately, we can use the available data on a user’s activity on a website to track if there is a bottleneck in the purchase process for a customer.

1. You work at a web design company that offers to build websites for clients. Signups have slowed, and you are tasked with finding out why. The [onboarding funnel](https://en.wikipedia.org/wiki/Funnel_analysis" \t "_blank) has three steps: email and password signup, plan choice, and payment. On a user level you have information on what steps they have completed as well as timestamps for all of those events for the past 3 years. You also have information on [marketing spend](https://en.wikipedia.org/wiki/Marketing_spending" \t "_blank) on a weekly level.

We can start by finding out the distribution of steps completed for unfinished sign ups. This way we can find potential bottlenecks. To further confirm this information, we can use time taken for each steps to find out which steps where difficult for customers.

We can definitely also check if marketing spend has decreased in this time period.

1. You work at a hotel website and currently the website ranks search results by price. For simplicity's sake, let's say it's a website for one city with 100 hotels. You are tasked with proposing a better ranking system. You have session information, price information for the hotels, and whether each hotel is currently available.

The first step would be to filter hotels by availability. We can allow the customer to select their price range so that an additional number of hotels can be filtered out.

We can then use session information to find out which hotels customers are most interested in. This information can be found by exploring time spent on a hotel’s page. Based on their browsing activity we can also figure out what price range a customer may be targeting.

1. You work at a social network, and the management is worried about [churn](https://en.wikipedia.org/wiki/Churn_rate" \t "_blank) (users stopping using the product). You are tasked with finding out if their churn is atypical. You have three years of data for users with an entry for every time they've logged in, including the timestamp and length of session.

In this case, we are trying to find anomalies in the data. We can find out if there are any clusters in the churn data outside of what appears to be normal. Maybe there are several categories of churn that happen at the social network.