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), [pageviews](https://en.wikipedia.org/wiki/Page_view), and purchases), as well as whether or not those users [converted](https://en.wikipedia.org/wiki/Conversion_marketing) from an advertisement for that session. You also have the cost and price information for the goods.

Speaking of statistics, this is actually a comparison between three groups of products. Considering that the final goal is to maximize the total benefit of the company, for each of the goods a target benefit function could be formed as:

Weekly Benefit due to advertising = number of sales\* rate of conversion from an ad \* (Price-Cost) – advertising expenditure

The benefit could then be compared among the three groups and statistical inferences could be made. However, to avoid the simpson’s paradox, it is reasonable to have an idea of the market including customer demographics, time of the year, events that might cause bias in the sales/purchases, etc. and do the comparison across certain segments of customers or timelines to avoid biased inferences.

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) 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) on a weekly level.

The very first step is probably draw a graph indicating sign-up frequencies along time to see when exactly the slow down started. When the start-down point is detected, a comparison could be made on a before-after basis to see if there is a significant difference between several metrics. One important metric could be the amount of time a user accomplishes two consecutive steps. That’s is from step 1 to step 2 and then from 2 to three. Comparison of these two metrics provide insights on where exactly the issue rises, if any. For instance, if a person passed the first step but did not go through the second step after a considerable amount of time, the reason could be sth. other than payment plans or prices. On the other hand, if a person had finished step 2 but did not make a payment, that means he probably found the plans inappropriate or the pricing was too high. For the first issue, probably a demographic analysis of customers is required as well as their intentions for having a website. This could shed light on what types of customers are not willing to signup even before seeing the prices. For the second issue, the company could potentially consider other plans and payment options to attract more people into their system.

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.

Surely there are many other parameters that need to be taken into account, including:  
Availability dates, distance to specific locations/attractions (accessibility), previous reviews, amenities, etc.

1. You work at a social network, and the management is worried about [churn](https://en.wikipedia.org/wiki/Churn_rate) (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.

This is very similar to question number 1. The slow down process could be defined by two different metrics: either a user did not log in to the system as frequently as he did before, or he is spending shorter sessions on the network. A time-series analysis needs to be done comparing the values at period t continuously with their previous values at period t-1. Comparison of the values could be done per individual or at aggregate level. Considering the Simpson’s paradox, it is reasonable to do the analysis within certain market segments based on demographic or socio-economic attributes. Also the trend needs to be compared with other social-networks if their data is available too.