Drill: What Can Data Science Do?

**For each question, describe how you would make it testable and translate it from a general question into something statistically rigorous. Write your answers down in a shareable document and submit the link below.**

**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" \t "_blank), 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.

Ultimately, we want to maximize profit from the selection in advertising. Before we can select a product, we should subtract cost from price to see which of the three goods has the highest profit margin. We should multiply that profit margin the overall amount of purchases that were converted through advertisements. To normalize for the amount of interest with each product, we should divide by the amount of page views for each good. From that number, we can compare to see which products have the highest relative yield. Next, we should compare that to the overall number sold for each to see if there is room for growth (i.e. market share) or to see if the product may not be something that sells as frequently (rare good). Based on those comparisons, we can ultimately decide which product to advertise.

**2)** 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.

We want to maximize new customers as constrained by our marketing spend. Firstly, I would review when the peak dates and times were for each step in the onboarding funnel. I would also compare the average time between each completed step to see if there were any outliers. I would overlap that with the dates of our marketing spend to see if there was any correlation. Additionally, I would compare that to seasonal and industry trends in hiring to account for any external forces. From there, I would try and use the marketing spend at dates or times which the data indicated led to more paid customers.

**3)** 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.

Given the few pieces of information available, we should gather as much insight as we can from the session information. We should see how long the candidates searched results, where they preferred to stay, and the price they paid. First, we can sort all results by availability to keep unavailable hotels at the bottom, thereby not wasting the customer’s time. Second, we can use the insights to see which hotel rooms are sold most frequently and recommend these as the most popular. The remaining results can be sorted by price.

**4)** 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.

To see if the churn is atypical, we need to compare samples using a t-test. We compare a random sample of users churn from recently to a sample from different time frames (i.e. several months ago, a year ago, three years ago). If we see any significant change between them, then we can go into the data for users to try and account for what may have caused the change. If there is no significant change in the users’ habits, but a significant change in the churn, we may consider external factors.