Drill: What can Data Science do?

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.

Notes:

* Time-oriented – continuity of user activity
* Navigation-based – continuity in chain of related pages
* Purchases
  1. There are a number of approaches. We may want to motivate sales of one of the 3 goods, depending on the profitability, or relationship to another type of product or inventory level.
  2. A testable technical plan would include an analysis of correlation between both activity and pageviews with purchasing behavior.
  3. I would also analyze how the investment in current advertising was related to the number of conversions from advertisements.
  4. I would determine if there was a difference in purchasing behavior, based on whether or not users came from advertisements or not.

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.
   1. I would determine the rate at which customers dropped off at each of the 3 steps, based on historic information. There may be some operational improvements required.
   2. I would determine what type of correlation existed between marketing spend and engagement with our site, since I have time stamps and weekly marketing spend.
      1. Hypothesis: If more money is spent on marketing, there will be more engagement – define engagement level
         1. If there is a strong correlation, accept hypothesis
         2. Important to look at strengths, low(0.1) may still hold useful information
      2. Consider lag in time-series analysis for causality, could go from positive to negative correlation. Experiment with time-period lag. Use regression.
      3. High-frequency data, lags are more important. Lags might disappear on monthly or quarterly data
   3. I would analyze whether specific plan choice was determined by marketing spend.
2. 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.
   1. I would analyze whether time on page or navigation behavior in each session was related to current price ranges or hotel availability.
   2. I would measure the rate at which availability or price influenced the click-through rate.
   3. I would weight the ranking system based on the level of price, availability and session behavior.
3. 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.
   1. I would determine the average rate of churn, then use a t-test to tell us the likelihood of the current churn rate being statistically different from the historic churn rate.