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.

Organize the data into a form where it is all accessible by a single programming language/interface. Secondly we must expand the main variables activity on a website, pageviews and purchases. Activity on a website includes session length and user actions per session, pageviews is self explanatory, and purchases I assume is the number of purchases per user session. I assume that the objective of the company is to increase revenue so the company cost minus the price per product would give the margin of profit which would be the products with the most profit yielding the highest differences. This too would narrow our search space to determine which products to increase marketing on and perhaps a few prospects that are similar in their outcomes. A PCA could be used to determine the correlations between individual variables and outcomes in the form of maximized profit. An A/B test can also be run to determine if the treatment conversion/purchase relies on a specific variable. A ttest between the non-purchase/conversion mean and conversion/purchase mean for a specific factor can be run.

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.

Convert the timestamp data into difference data for each user. Identify whether there is a pattern of increased time difference between those that do not sign up. Is there a step in the process where users consistently drop out. Construct some conditional probabilities to determine the likelihood of signup based on the steps taken. Join the marketing data with the dates from the onboard funnel dates. Is there a pattern that can be established between market spending and user signup for that week.

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.

Based on the users activity on the website classify whether there was conversion and non-conversion based on whether the user clicked on initial search. Then create another group which clicks on or converts based on other searches or clicks on the website. Perform an A/B test to see whether the two group conversions are significantly different. If they are this could suggest that other variables are more important in determining conversion. Within this space do a PCA to reduce the feature space. The new feature space would suggest ways to change the search based on the user session information.

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.

Identify the important questions that could be asked about atypicality. I believe the question is ill defined since the company only has data on its customers. If other data is present or if there is an industry standard to compare against then it is possible to move forward. If not then im now sure how this company could proceed. Questions to ask

* 1. What is the frequency of user login per month per year etc
  2. Is there a correlation between user login frequency and churn rate
  3. Perform A/B test for those who login above a certain frequency and those that login below that frequency. Is there a significant difference statistically
  4. Examine length of session per login
  5. Perform A/B test for session length per login above a threshold (which would be determined by taking the median of session length per login per user) and those below the threshold for session length per login
  6. Compare A/B test to that of industry standard