Unit 1, Lesson 1, Project 4

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.

The function of the advertising campaign is to maximize sales and profit. These are not mutually exclusive as the cost of a product that sells high volume may result in lower profits than high cost products with lower conversion.

First we ask ourselves:

* Which product is getting the highest conversion?
* Which product is getting the most activity?
* Which product generates the most profit (unit profit \* conversion rate \* web activity)

A primary goal of the campaign should be to reduce the activity on the webpage for a product without negatively impacting the conversion (increasing that conversion if possible). So we break down each product the company sells using the same formula. Then we evaluate the expected impact the marketing campaign will have (+10% conversion, +20% pageviews, etc.). Accounting for the profit yield of that increase we should be able to see which product is most profitably impacted by the campaign.

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.

We have 106 weeks worth of data on the marketing spend. We have the numbers of users who have created an account, of those how many have chosen a plan, and of those how many have processed payment. To start, we assess the percentages that funnel from step 1 to step 2 and step 2 to step 3. Using comparable data for similar web-design companies we can see where we are losing more of our potential customers. If they drop off before completing step 1 we should analyze the user-friendliness of the account setup process. If step 2, we should direct some marketing resources to surveying the reason. Perhaps the plans offered are too complex or too many. If step 3, there is likely an issue with cost or financial commitment to continuing services and marketing funds should be redirected to boosting the value perception of the products or advertising a sale.

Similarly, we can compare the user progression through the three steps to shifts in marketing spending, accounting for the natural delay in marketing spending to consumer impact. If a reduction in spending on advertising through digital streaming ads is followed by reduced progression through steps 2 and 3 then we may want to consider increasing spending on that over some other, less effective advertising.

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 only other ranking we can use is by availability. Perhaps if we had other information (proximity of IP to hotel chosen, time between bookings and stays, etc.) we could consider ranking according to that.

However, we can use the session data to assess how much time is being spent looking at hotels that are unavailable for the dates chosen. If we then set thresholds of likelihood of availability and price (>50% availability for dates chosen, lowest 25% in price for example) we can reprioritize the ranking system. This may be wiser than shifting the ranking completely. It could be preferable to create a hybrid price/availability ranking system.

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.

First we create a dataset that shows the time between user logins for all users over the three years. We look for trends that describe the churn and identify the point at which use drops below profitability. For example, we look to see the mean time between new user login and second login, second to third, and so on and create a mathematical model for expected churn over time. Then look for the point at which a statistically significant percentage of users have been churned (between 3 & 6 six months, after 75-125 logins). Then we can analyze the actual sessions that mark the downturn and report our findings for marketing toward customer retention.