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, pageviews, and purchases), as well as whether or not those users converted from an advertisement for that session. You also have the cost and price information for the goods.

ANSWER: The primary goal of the company will most likely be to increase profit, so the question is which of the products will garner the most profit. Using the data provided, we can calculate the profit from any given scenario, i.e. what is the return in widgets vs doodas vs fizzbangs. We can do this by the per item-profit (e.g. price-(advertising cost- manufacturing cost - delivery/shipping cost). In the perfect scenario, the data we gathered would produce an accurate estimate for item price that included overall costs. We can also take the data gathered on user activity to determine what people will most likely purchase. Afterall, there is no point in trying to sell a product that has a high profit margin if no one wants to buy it. We can also use data from the visitor sessions to better price our products. If the prices for most of the products currently being purchased at a mid-level tier, viewers might have a less likely chance of buying a similar item prized in a higher tier.

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 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 on a weekly level.

ANSWER: The goal would be to get more people to the third stage of payment. Using timestamp data, we can determine if there is a time when more people are getting to that third step. We could also determine what time of the year the process takes the least amount of time. This would help the web design company what time it might be more beneficial to increase marketing. If, for example, we find that most payments are happening at the beginning of the year around tax season, the company might wanna consider spending more money at this time and advertising on sites that are related to tax preparation. It might be worthwhile to collect other information, suchs as the reason why they want to build a website (personal, business, etc). Using that information will help to create targeted marketing campaigns instead of more broad ones and can also help to determine how to best allocate marketing resources.

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.

ANSWER: While ranking by price is good, I think that most people like myself, are looking for the best deal and that doesn't necessarily mean purchasing the cheapest item, rather it is what will get me the most satisfaction relative to cost. Then we can ask, how do we measure satisfaction? That is something of an individual preference. To some people, amenities are everything. To others it might be distance from the activities that they which to do.

We can take the session information from users to determine what the most looked at properties are and why. However, it is important to know the kind of information the hotel site has collected as session information can mean many things. So what does session information mean? Is this the time spent per page, the number of clicks per person? Also, it what length of time is this in? There are many variables that can affect the quality of any models created.

4. You work at a social network, and the management is worried about churn (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.

ANSWER: We can use the data over the last three years to construct some models to paint a "portrait" of each user. We can do this by taking the timestamp information and creating a time-series plot of their logins. Using this data we can see if there is a trend among users. Maybe there are points in the year where users frequent the site less, thus determining if any churn the company is seeing is just a slow time for the company overall or if indeed is something new and "atypical". To adequately measure churn, we would need to establish a timeframe of no activity in which we can consider the user inactive. Is this three months, 6 months, or more? This can help us see a trend for each user over time. Are they logging in less frequently in year two than year one? Are there a significant amount of users that stop using the product after a year or two years? Using the data to answer questions like these can be beneficial in understand what is happening, but what it fails to explain is why it's happening. The information for that question would come from other data related more at product content and user satisfaction rather than user logins.