**Drill: What can data science do?**

Black: General Question

Red: Technical Question

Blue: Plan

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.

Which products are frequently purchased together?

Can we find association rules that show sets of frequently bought items?

Based on the association rules found, we can do advertisement on those items that users are not buying. Users are not purchasing an item may be due to lack of understanding about a product or insufficient advertising efforts by the company.

Are those users who stay more on the website make a lot of purchases?

What is the relationship between amount of time spent on the website total price of purchases?

The obvious assumption would be to think that as users spend more time they select more products and pay more. If the reverse is true, we might think to make a long, interesting advertisement about a product that a user is not buying that much.

Did the advertisement affect users to change their purchase?

What is the likelihood of people to change their product purchasing pattern or change selected product on a specific session, because of advertisement?

A good ad will fulfil its goal, for example an advertisement about Udemy led me to learn some courses there rather than using Coursera for all learning needs. The goal of this e-commerce company is to make as much sales as possible, thus, an advertisement aim shall be to make users purchase more products, even products they never tried before.

Which products are mostly purchased by older people and younger ones?

What is the distribution of, ages of people, who purchased widgets or doodads or fizz bangs separately?

Are more expensive products selling well?

Is there a significant difference between purchase history of cheap products and expensive ones?

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.

Which one takes longest: signup, plan choice or payment?

What is proportion of users who spend longest time on signup, plan choice, or payment?

This will notify the company to make the UI appealing and easily understandable, so users can quickly go through. If considerably large number of users are spending longest time on plan choice, it implies something is wrong with the instructions provided on plan choice screen.

Is our marketing spending paying off?

What is the relationship between company’s weekly marketing cost and number of sign ups?

A line graph can show whether company’s weekly spending is increasing or fluctuating and compare it with another line graph which shows number of signups on every week. If the spending and signups are both declining, it may be a sign of insufficient spending on marketing and if spending is same or increasing while signups decline, it implies the company is not achieving its goal, so a new marketing strategy is required or there is a problem with the website itself.

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.

Did users find top 3 results helpful?

How many users choose a hotel from top 3 results?

The session information will contain which hotel users selected, so if most users found a desired hotel from the top three results, it may imply users like to pay more money or high-priced hotels are available. If the reverse is true, users don’t afford top ranked hotels or unavailable houses are shown on the top; so, users are not satisfied with the results and a new ranking system shall be employed.

A better ranking is to include other attributes in addition to price, like the location of the person, **food choice**, **number of reservations**, **type of room**, **price range**, **availability**

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.

Are more users stopping using the product on the last year?

How the length of session is varying through three years for those who churn and for those who didn’t?

This will help the management decide to do something to prevent users from stopping. A time series analysis can reveal the pattern of variation in length of session for those who stopped and we can compare that pattern with current users and if it is similar.