Question #1. You work at an e-commerce company that sells three goods: widgets, doodads, and fizz bangs. 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.

My Response:

Considering activity on a website, pageviews, purchase:

Say, 60% of visitors spend less than a minute on your site. If people don’t engage with your website, you won’t be able to sell them your product. Also, when people browse more pages from your site, they’re more likely to buy something. If people spend more time exploring more about a particular product, that product might be popular.

Considering number of items sold:

If our primary goal is including the product which one is generating maximum profit:

Since we have the Cost and price data for each product we can calculate per- product profit.

Say profit of widgets = number of widgets sold X (price of each widget – cost of each widget)

And select the product which gave the maximum profit.

So, in the new advertising campaign we can feature the product which gives either more profit or which is more popular among the people.

Question #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.

My Response:

In the onboarding funnel process, I will look at following questions:

* What step we lost our people?
* How much time did they spend on each step?
* Which step was the final one to be completed?

I will be familiar with the data first. Then, I can divide the data into two groups: people who went through the three-step funnel, AND a random sample of people, who for some reason, left our website in between and do some analysis on both groups looking for any trend / pattern in past 3 years.

It could be that there's nothing wrong with the website, and our slowdown in sales is reflective of a general slowdown.

Perhaps many of user got to the plan choice page but could not find a plan perfectly suitable for them, which made them leave half way.

Perhaps many of user got to the pricing page but figured that they just can't afford our web services right now.

It's also possible that they found signup forms confusing, which made them spend more time on a step, imposing an obstacle to finalize the deal.

It may be possible that for some reasons, the marketing was not proper for a week which reduced the signups for that week.

We will analyze all these factors and try to figure it out.

Question #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.

My Response:

Currently the website ranks the search results by price. But price may not be the only factor. There can be some other preferences as well. We can tailor the ranking in accordance with the user's preferences as much as possible. Factors like room sizes, parking availability, restaurant quality, accessibility, Internet access and other amenities can help to determine the ratings. The location of the hotel, can also be another preference for some visitors.

We can rank the website on how closely they conform to personal preferences and availability for the hotel. Perhaps it would be good to allow users to rate hotels on five-star scale, with five indicating the best, and one (or zero) representing the worst and assign weights to the preferences while rating. The final rating determining the rank, would be a weighted average of how well the rooms meet the user’s preferences and the availability.

Question #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.

My Response:

At user level we will count number of log-ins per week and session length. It would be useful to know about the distribution of typical user data for these two variables separately. The distribution is important, because means and standard deviations could be used to detect deviation from what's typical.

We can count number of hours per week by user, and then find any deviations from the pattern for user churn.

At website-level, we need to find weekly proportion of users for these years and try to find any pattern or deviation.