**Bootcamp Unit 1**

**Question 1 widgets, doodads, and fizzbangs.**

**Have:**

**Individual visitors’ sessions**

**Activity on a website**

**Pageviews**

**Purchases**

**Whether or not converted from an advertisement**

**Cost and Price for the goods.**

**Head of Advertising asks which they should feature in the new advertising campaign?**

**Convert Information to testable format and then again from general discussion to statistically rigorous.**

Goal: sell more products/create more conversions

Thoughts:

Figure out potential reasons why some sales convert and others do not:

If we break this down into a grid analysis:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Widgets | Doodads | Fizzbangs |
| Time spent viewing |  |  |  |
| Page views |  |  |  |
| Cost |  |  |  |
| Price |  |  |  |
| Purchases |  |  |  |
| Converted from ad |  |  |  |
| Not Converted |  |  |  |

Calculate the average value for each of the conditions and see if any one factor stands out as a positive or a negative correlation.

**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**](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.**

First put numbers to the onboard funneling steps:

Maybe count the number of drops at a certain step to find out which step seems to be the most averse.

|  |  |  |  |
| --- | --- | --- | --- |
|  | email | Password signup | Plan choice |
| # of drops at this step |  |  |  |
| Average amount of time spent on step before drop |  |  |  |

Create a graph showing weekly numbers:

X-axis = weeks

Y-axis =

For email, password signup, and plan choice = # of drop instances at this category as well as average amount of time spent before drop plus the average amount of time spent at each point when not dropped.

For marketing spend = total amount spent per week on marketing.

# of drop instances at a particular point would tell us where the bottleneck in the process lies.

Comparing time before drop to average overall time if not dropped might tell us if time was spent making the decision to drop at a particular point. For example if the more time is spent in password signup when this step is dropped than when it is not, there may a glitch in the system that occasionally prevents users from successfully completing the step, it could be certain browsers giving issues, or perhaps something happening at a certain time of the day (like system maintenance) that is making this step difficult.

And finally looking at marketing costs in parallel with user activity will tell you if there is any correlation. If user activity decreases when marketing costs decrease then you know you might need to continue marketing efforts. If user activity seems immune to marketing costs then shifts might need to be made to target more effective methods of advertising or customer support.

**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.**

While price is important in deciding whether to rent a room or not, perhaps the first concern is whether the room is available or not. Can’t rent an unavailable room regardless of price. This would likely be reflected in the session information for users – Do users who encounter many unavailable options drop the session before converting more often than those who don’t?

Next would be to look at what users’ who actually book a room activity looks like – did they book the first room that was available? Or did they look at many price options before making their decisions?

Based on the answers to these two questions I would propose ranking by availability first and then lowest to highest or vice versa depending on which seems a more popular choice amongst the converted browsers.

**Question 4: 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.**

Question: By “their churn” – do they mean the users who stop using the product (vs users who remain users)? Or is “their churn” the churn rate of the social network as compared to other networks?

Since I don’t have information for competitors I am going to assume the question wants me to compare churned users to continuing users.

Since churned users fall into two categories – those who leave and come back vs those who leave for good – I will treat them separately.

So we need to define 3 groups: continuing users, churned but returned, and permanently churned.

First we would need to find out what a typical user’s pattern of use is in order to define a period of time a user would be considered churned.

As this is a social network typical usage would be multiple times a day. Considering normal circumstances when a user might be expected to be away from whatever device they use to access the social network it is not inconceivable that a user might be inactive for several weeks if on vacation, camping/backpacking, sailing/boating, or otherwise taking a break from technology. It is unusual for a typical working individual to be in that circumstance for greatly extended periods of time and 6 weeks is a reasonable amount of time of disuse that allows for extenuating circumstances beyond a typical vacation plus a cushion of time to ensure that it is not simply a mismeasurement.

So we define a user as churned once they reach 6 weeks of inactivity.

Any user defined as churned who begins activity again is churned but returned.

Any user with continuous use or no more than a 6 week gap in activity is a continuing user.

Next create a graph where weeks of use are plotted on the x-axis (so all users’s first week of use is calculated as week 1 regardless of when they began using the service) and the average length of time a user spends represented on the y-axis. Plot the 3 groups simultaneously on the same graph. In the churn group where users return to service – the latter part of their service will be handled in a later graph – for this particular graph it will be ignored. The purpose of this graph is to see if there is any indication in the time spent using the service that people are likely to drop the service.

If necessary a second graph can be made with the number of users at each week as well to indicate attrition rate.

For users who were churned but returned we can further analyze the length of time they refrained from use and if there is any change in the length of time they used the service before churning and after churning. Or if they have a pattern of multiple churning. Depending on the patterns that emerge further analysis could be done to explain behavior.