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

Below we have a series of questions for you to translate into a technical plan. For each question, describe how you would make it testable and translate it from a general question into something statistically rigorous. Write your answers down in a shareable document and submit the link below.

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 if those users converted from an advertisement for that session. You also have the cost and price information for the goods.

* Calculate the unit margin for each of the products (MARGIN)
* Distribution of each of the products across the year (AVERAGE PER MONTH/WEEK)
* Distribution of conversion rate across the year (AVERAGE + VARIANCE PER PRODUCT)
* Time between advertisements and conversion rates (PATTERN)

This will answer not only which one but also when depending on the time of the year it asks. Additionally it will depend on the marketing effort required related to the margin it brings (% of marketing expenditure of the margin per item/product)

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 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.

* Average time between steps and variance
* Conversion ratio on average
* Frequency for each step timestamp and distribution of marketing spend (is there any correlation)
* Type of distribution at each step and correlation between them (on a per day basis too?)

No correlation between marketing spend in a weekly basis and an increase in conversion from sign up to payment will indicate that either the target is wrong or there is something not working in the pipeline. Analysis of the different steps to see where we most of the clients are lost/time will unveil additional problems if any.

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.
   * Average time spent per visit viewing each hotel
   * Reservations made in real time and price paid
   * Average price per hotel room with similar entry/exit dates
   * Average number of visitors landing in each hoter
   * Variance in prices
   * Availability of the hotel (average and variance) how much time it needs to be unavailable

The rank I´ll use would be weighted average of prices. The weights will be the average availability and users looking for a specific hotel.

1. 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.

* Average time per session (timeline)
* Churn per user and correlation with time
* Regression based on timestamp and length of session to see if there is any relationship
* Average churn per year/month compared to previous months/years

Build a regression model based on timestamp and length of session to see if there is any relationship between variables and with churn. See to which extent both explain separately or together churn. Clean outliers and see the trend in terms of churn to see the trend.