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 whether or not those users converted from an advertisement for that session. You also have the cost and price information for the goods.

Since advertising wants to know which product should be featured on their new campaign, I think two separate questions needs answering through data. First, how each product currently sells independent of advertising? Secondly, how each how effectively can advertisements convert users for each product?

To test the first question, I propose isolating the non-converted user pool, and analyze the purchases rate for each product. Thereby generating an understanding of users not coming from advertisement's preference for the products that are offered.

Secondly, isolating the population that converted from an advertisement, and analyze the percentage of that population's purchasing history for each product.

Then advertising can make the decision based on company needs as well as past performance.

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.

If only the initial email Test if there appears to be a relationship between marketing spending and signups.

If one is concerned about the entire process, then information regarding time between each session can be examined, to see which tier appears to be the biggest drops.

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.

First show all available hotels over any that’s not available. Secondly, assume the longer time that people spend on a hotel’s page to be of more interest, the clicks and length of viewing indicates popularity, then construct some sort of relationship that manages the tradeoff between price and popularity. Assuming

our website has an agenda to promote the more expensive hotels, perhaps show hotels in ascending order of their (price * popularity index)

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.

First define which users are considered to have stopped using the network: (Haven't logged in for over 1 month?)

Then isolate the users that are considered to have stopped using the network, regardless of their reasoning, generate a pattern of the time between their sessions(maybe frequent at first, then less so till gone). Match the potential pattern with active members to see how large a portion is considered to be potential dropouts to determining if it's worth worrying about.