PROBLEM STATEMENT 1

The question every advertiser wants to answer is this: How can I spend my ad dollars to get the MOST sales at the LOWEST cost-per-sale? I have a 2017 dataset for television advertising placements, the resulting leads (someone calls the phone number listed on the ad) and sales (the lead converts to a sale - on average, 10 weeks later).

How do the factors of a TV ad placement affect the conversion of a viewer from a lead to a sale? Using my own dataset, I would seek to determine the optimal media placement to hit the cost-per-sale goal. This requires determining the average conversion rate from a lead to a sale, and then using the cost-per-lead combined with the conversion rate to predict the cost-per-sale. Ad dollars are spent based on a cost-per-lead, but cost-per-sale is the ultimate result we would like to predict.

What factors most heavily affect a conversion from lead to sale? Media type (Cable v. Network)? Station? Daypart? Spot length? Creative? Can we predict what the conversion rates and cost-per-sales would be down to such a granular level?

A big challenge may be determining which factors of the ad buy have the most impact. The data is available historically, but multiple years could add a challenge due to varying strategies and budgets cross-year. Media buying happens almost real-time, immediately tracking the leads that come in from a spot airing. The ultimate question I would like to answer is can we better optimize our media to more accurately hit cost-per-sale goals? This project appears to be two-fold: first, determine the average conversion from lead to sale, then calculate what cost-per-lead falls under the threshold to hit your goal cost-per-sale.

PITCH

Television advertisements that show a 1-800 phone number are a dream when it comes to tracking the success of an ad campaign. When someone calls a phone number, they are linked to the exact ad placement. You know how many people responded to a 60-second spot that aired at 11:17 am on a Cable program. This leads to the ability to optimize your media: keep placing dollars behind the spots that drive response. The problem? Leads come in immediately, within hours or days. Sales, however, take an average of 10 weeks to become fully mature. If the end goal of an ad campaign is sales (resulting in a LOW cost-per-sale), then optimization off the cost-per-lead is not truly optimizing effectively. When sales come in ten weeks later, you cannot change how you spent your money nine weeks ago. Therefore, it would be beneficial to find a way to predict the cost-per-sale based on historic data demonstrating the relationship between cost-per-lead and cost-per-sale ,broken down to various levels of granularity such as spot length, creative choice, and media type.

PROBLEM STATEMENT 2

How does geography affect the health insurance marketplace? Using updated 2018 data from the Centers for Medicare & Medicaid Services (CMS), I would like to answer questions such as (1) how many plans are available in each state? (2) how do rates vary between each state? (3) which markets appear saturated with the most competitors, and which are least saturated? (4) is there a concentration of places where there are more out-of-country or out-of-service-area plans?

It would be important to determine which factors of the CMS data are integral to this analysis. There is an abundance of data provided which could slow down the processing and analyzing. Properly joining tables and using the correct fields will be an important step in ensuring a smooth process. There are pieces of the dataset that would not be useful, such as the URL for brochures, which could be removed before performing analysis.

PITCH

The insurance marketplace is overwhelming for many adult Americans. What information could we deduce about the plans available across the country? I am most interested in geographical differences between insurance providers and their plans and rates. In an industry more heavily influenced by the government than most, it would be interesting to see what patterns surface and information could be gleaned from this dataset.