Case Study: Online Ads Measurement

At Oracle Advertising, one of our products measures the effectiveness of online advertising. To that end, we look at the differences in sales between households that saw the ad (“exposed”) and those that did not (“unexposed”).

Problem Statement

Using the attached set of data from a recent online campaign for a candy bar, we want to know if the advertising worked by answering the following questions.

- Post campaign sales lift - Did the ads campaign lead to sales lift?
- What assumption should you make?
- What methodology would you use to measure campaign performance?
- Is the sales lift consistent for all demographic and spending groups?
- What other insights could you get from the dataset?

Deliverable

Put together a 15 minutes presentation (Word doc, PPT, PDF, etc) with a compelling argument one way or the other. Be sure to support your argument with data, list any/all assumptions you have made, and include descriptions/reasonings for any models you chose. Your intended audience is a fellow data scientist who can follow along with your methodology but wants your guidance on next steps. Be prepared to walk through your presentation via a screen share with your interviewer during your scheduled time.

P.s. If you summarize your analysis in less than 15 minutes, that’s fine. But please do not go over 20. Better to be concise than drawn-out.

Data

The CSV contains a row per household, with columns containing data related to pre-period sales data, post-period sales data, demographics, online activity, and a flag for whether or not the household was exposed to the candy bar ad.

Pre-period sales data is aggregated over a quarter (13 weeks) and 5 quarters (65 weeks) are included in this data set, where Q1 is the most recent quarter and Q5 is the furthest back in time. Post-period sales data consists of sales during the campaign and in the 4 weeks following the campaign end.

Demographic data includes household income, the number of total individuals who live in the household, the number of children present, and the age of the person who does most of the purchasing.

Online activity data includes how many cookies are active in the household, the number of days a cookie was active, and the total number of times Oracle Advertising saw the cookie.

Column Name	Description
hhid	unique identifier variable
reached_flg	a 0/1 flag indicating if a household was exposed
num_inds	how many total individuals live in the household
children_ind	how many children are present in the home
hh_income_ind	household income
age_ind	age of the person in the household that does most of the purchasing
home_value_ind	value of the household
state	state in which the household lives
num_cookies	how many cookies are active in the household
num_days_online	the number of days a cookie was active
num_events	the total number of times ODC saw the cookie
brand_sales_q5	household-level sales 5 quarters back in time
brand_sales_q4	household-level sales 4 quarters back in time
brand_sales_q3	household-level sales 3 quarters back in time
brand_sales_q2	household-level sales 2 quarters back in time
brand_sales_q1	household-level sales 1 quarter back in time
brand_sales_post	household-level sales during the campaign and in the 4 weeks following the campaign

Definitions

- Cookie: a very small text file placed on an individual's computer that allows the company who placed it to identify the user at a later time.
- Individual: a unique identifier for a single individual person, made up of one or more cookies.
- Household: a unique identifier for a family or cohabitating group, made up of one or more individuals.