



# Identifying eNewsletter Audience Characteristics Driving Advertising Revenue

04.01.2021

---

Mila Khalitova  
Bulletin Healthcare

## Business Problem

Bulletin Healthcare produces eNewsLetters for various medical professionals which are distributed to specialized medical associations. The NewsLetters are free for recipients, and their production costs are covered by money from advertisers who would like to target and reach physicians. Bulletin Healthcare is looking to expand its network of healthcare professionals and increase the reach of their briefings but is unsure what kind of physicians they should focus on first.

Advertisers can target BH recipients based on the following criteria or any combination of them:

- **Professional characteristics**
  - Area of specialization (e.g. OBGYNs, oncologists, podiatrists, etc)
  - Medical Associations in which a recipient is a member (some of them are based on the areas of specialization such as ACC - American College of Cardiology, others are based on location such as Texas Medical Association, etc)
- **Professional lifecycle** (eg. Student, Nurse Practitioner, Resident, physician, etc)
- **Location:** State, city, zip code
- **Hospital characteristics:**
  - Location
  - Size
- **Prescription behavior**
  - Specific drugs prescribed in the past
  - Prescription frequency/Prescription volume

One option for Bulletin Healthcare is to focus their efforts in trying to expand its network among doctors with rare specialties because advertisers tend to pay higher to reach recipients with rare specialties such as hematologic oncologists. On the other hand, expanding reach among more general specialties such as family medicine, internal medicine, etc might also make sense because there may be fewer advertisers trying to reach highly specialized audiences than those advertising their products to general customers, and because there are more doctors in general practice than there are in specialized areas.

## Goals

1. Determine recipients' characteristics that drive advertising revenue

2. Identify the optimal path of expanding Bulletin Healthcare reach and advise Bulletin Healthcare management on their expansion strategy.

## Datasets

As the primary data, the project will use proprietary BH data registering briefing delivery and associated recipients with information about them.

The data on prescription volume for each doctor will be obtained on the CMS Medicare website. The two datasets will be joined using recipients' NPI number, a number that uniquely identifies every single healthcare provider in the U.S.

## Approach

### I. Collect data on 200,000 eNewsletter sends

Randomly selected data from the BH database; count total number of Impressions per recipient, cost of impression to advertiser, and the total revenue driven by each recipient in 2020. This will be the dependent variable.

### II. Collect prescription data

Using CMS prescription data and physician's NPI number, identify how many drugs each physician prescribed over the past year.

### III. Visually inspect the data to identify the most relevant characteristics associated with higher revenue.

The initial BH dataset will contain data on all targeting characteristics such as location, specialty, medical associations, etc.

### IV. Train a regression model predicting revenue from recipients' characteristics and identify the most impactful ones by comparing their b-coefficients.

## Deliverables

1. Python notebooks executing each of the steps:
  - Data Wrangling
  - EDA
  - Data Preprocessing and Model Training
  - Modeling
2. The original dataset and the cleaned dataset
3. Slide deck presentation describing the project
4. A report describing the project, including EDA and data story