Valentina Rizzati

The launch of Spotify KIDS

21ˢᵗ April 2021

QUESTION

I am a Data Scientist for the Product team at Spotify and I am working with the Product Marketing team on the launch of a new product: Spotify KIDS. This new platform will include both audio and video content for children and teenagers (age 3-19 years old). Considering the immense challenges faced by the education system during covid, the initial focus of the platform will be on educational content.

As part of the multi-channel campaign for the product launch, the Marketing team plans on leveraging the great power of outdoor advertising. Also, the product will first be launched in New York City (five boroughs). It's useful to mention that in Marketing, most of the time, targeting children means in fact targeting their parents.

Due to the global pandemic, the team is faced with a couple of challenges: a reduced marketing spend and a reduced foot traffic in major cities including New York. Therefore, they asked for my support in prioritizing the areas where to run the Out-Of-Home (OOH) campaign so as to ensure we run a profitable (i.e. positive ROI defined as customer campaign value / marketing cost) marketing campaign. Because their foot traffic is relatively greater than the one of other areas, subway stations will be the primary target of our geo targeting strategy.

DATA DESCRIPTION

To define the targeting component of the OOH campaign I am planning to use the following datasets:

1. Census data to target the most relevant zip codes based on the following parameters:
   a. Median Age > 30
   b. Household income > $100k
   c. Median Persons per Household > 2
2. MTA turnstile data to prioritize subway stations by foot traffic:
   a. 12 weeks of data, from the week ending Jan 9th 2021 to the week ending Mar 27th 2021

My station targeting strategy is two-pronged. First, I will narrow the list of zip codes based on the socio-demographic data provided by the Census. Concurrently, I will define the variable of *foot traffic*, based on which I will prioritize station selection (i.e. stations with greater foot traffic will be prioritized at the expense of stations with lower foot traffic).

As a second step, I will examine the distribution of specific stations by *foot traffic* and I will rank the stations in descending order based on their specific *foot traffic*. This last step will be instrumental for me to provide the Marketing team with a list of subway stations to prioritize for the OOH campaign.

It's useful to mention that, although it would generally be a part of such a project, marketing spend allocation by station is out of scope for the purpose of this project.

TOOLS

I am planning to use SQLAlchemy on a Jupyter Notebook in order to conduct an initial exploration of the data from both sources. For a deeper exploratory data analysis, I will use pandas. Finally, I am planning to use seaborn for visualizations.

If needed, I will use Tableau for more advanced and interactive visualizations.

MVP GOAL

The MVP will consist of the most relevant visualizations of the distributions of subway stations by *foot traffic* as well as socio-demographic characteristics that are part of our target by zip code.

These visualizations will be accompanied by short paragraphs presenting the main insights and conclusions from the analysis.