Advertising Dynamics in Medicare Advantage

Da-Young Kim and Zhirui Li
David Meyers, PhD
Andrew Ryan, PhD

School of Public Health

Project Overview

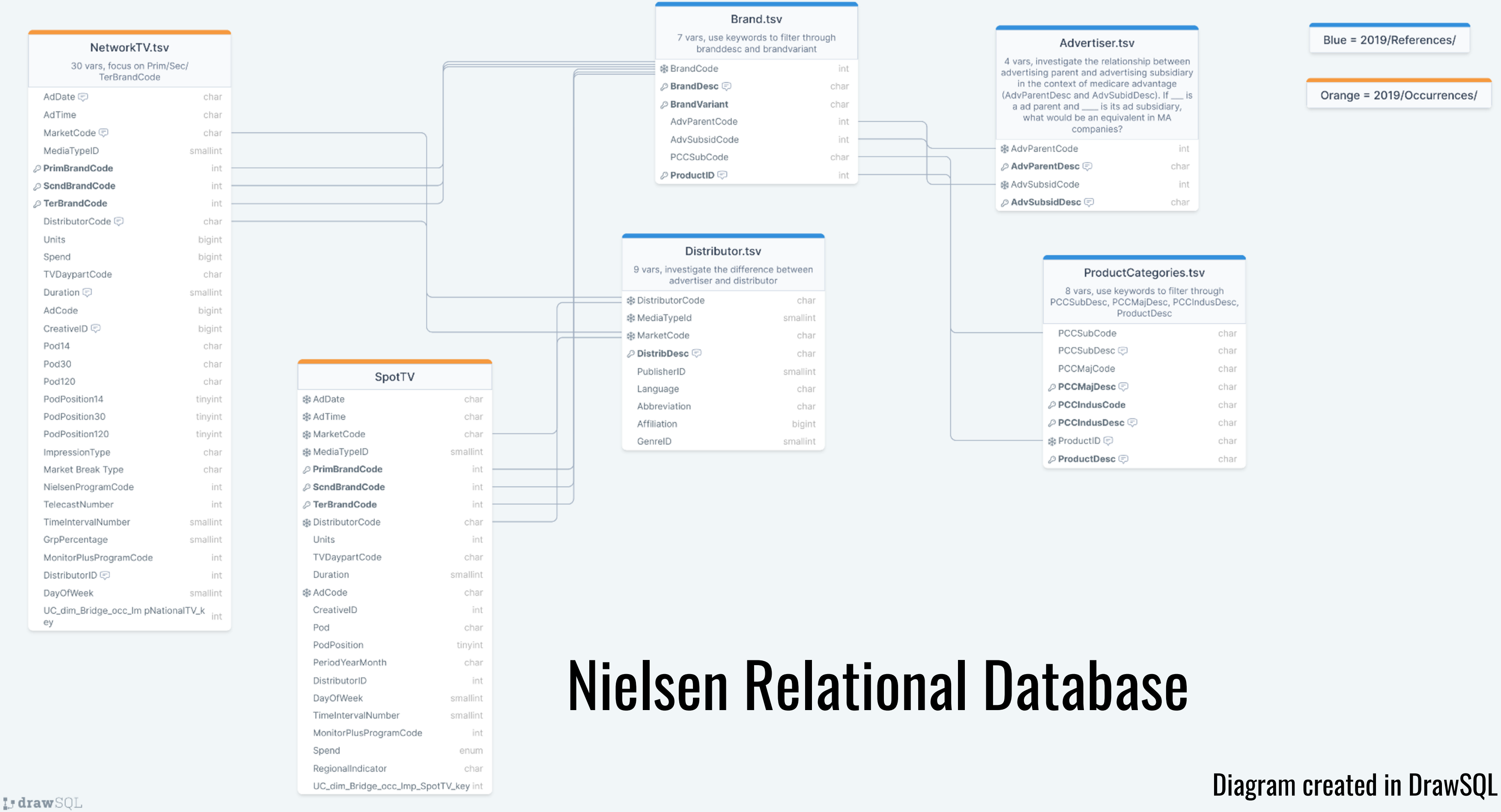
Aim: To investigate advertising dynamics (size, distribution, and impact of advertiser spending) in the Medicare Advantage space

Research Question: What is the extent of TV advertising in MA? What is the distribution of spending on advertising in MA and how does it vary across markets?

Four-Week Focus: Given the structure of the Nielsen Ad Intel Database, how do we *identify* "Medicare Advantage TV ads"?

Motivation:

Medicare Advantage (MA) is steadily growing in popularity amongst seniors, and Medicare Advantage is expensive compared to traditional Medicare. Spending on advertising by health insurance companies may contribute to federal overspending on Medicare... where else could that money go?



Nielsen Relational Database

Diagram created in DrawSQL

Methodology

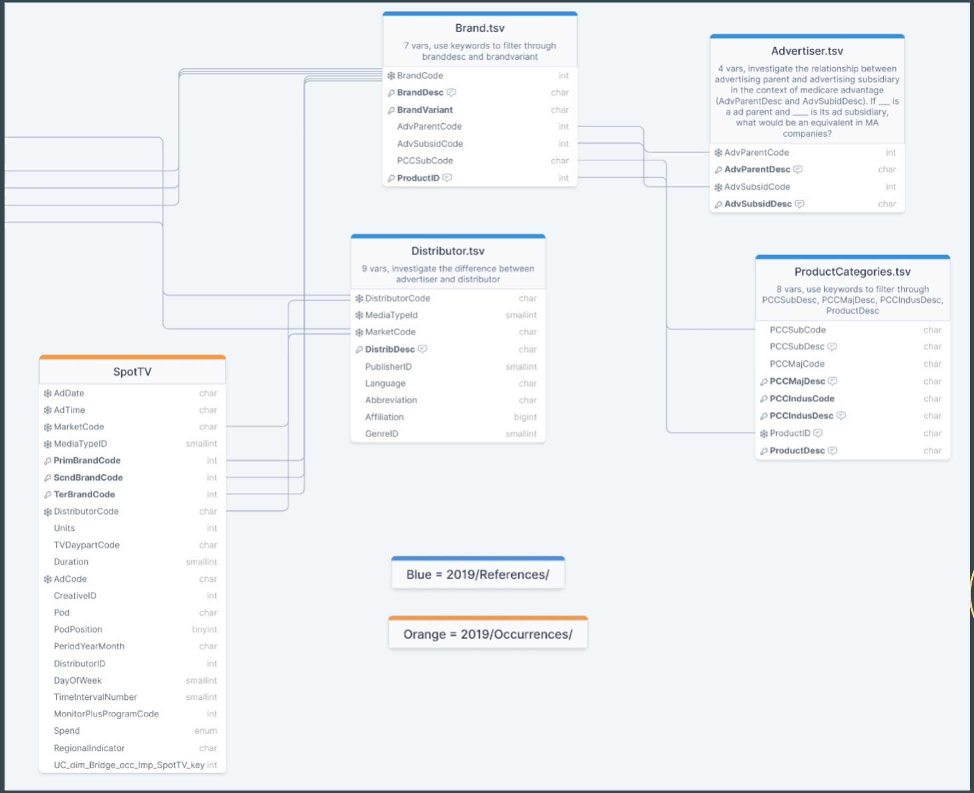
First Approach:

Filter Advertisers by Known Companies → Link Advertisers to Brand Codes → Filter Advertiser-linked Brands by MA-related Keywords → Filter Occurrences by MA-related Brand Codes

Second Approach:

Filter Brands by MA-related Keywords → Link Brand Codes to Advertisers → Identify All Parent Advertisers that are Associated with MA-related Brands

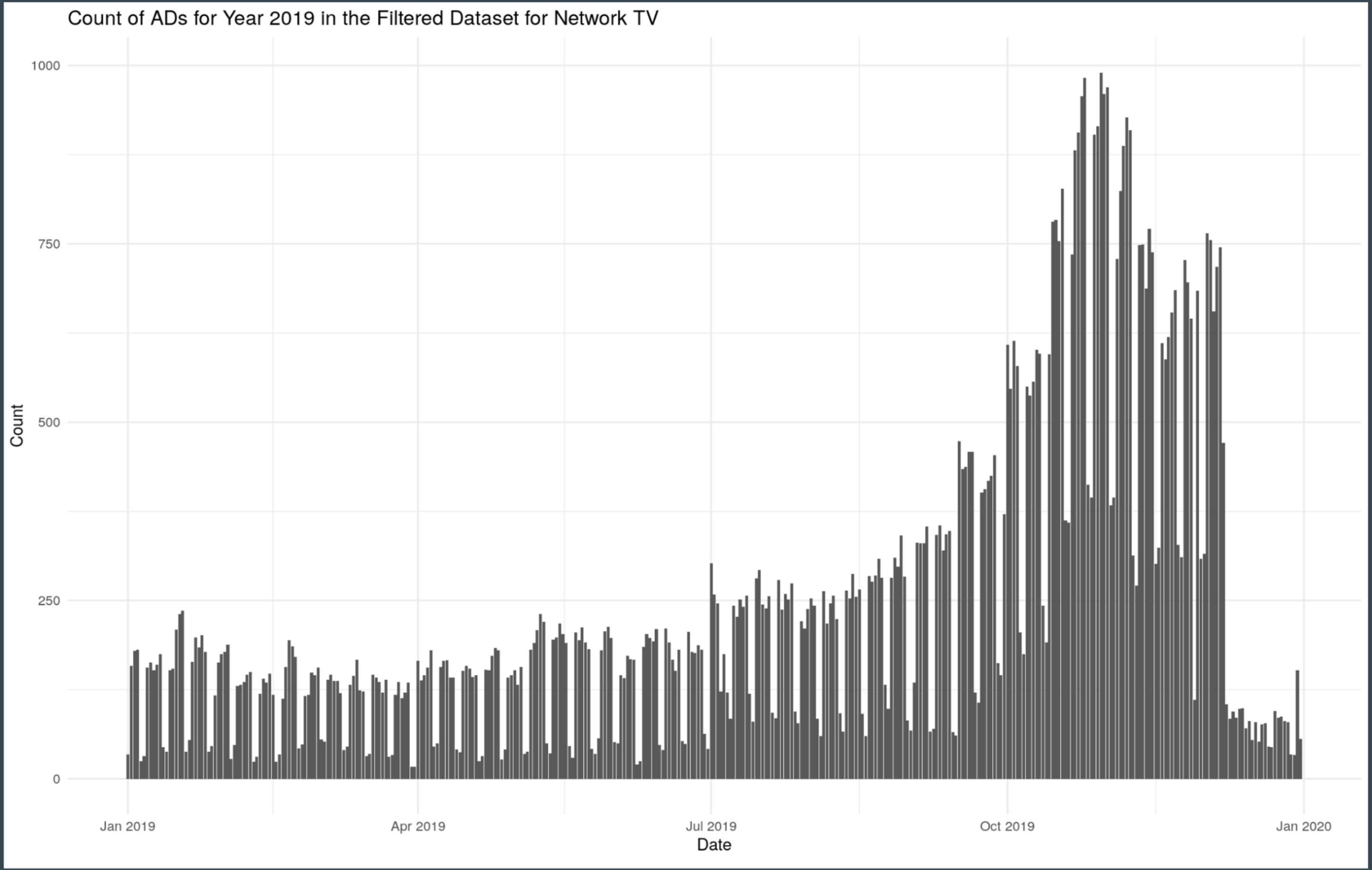
Our Filtering Strategy



<u>Challenge</u>: designing a strategy for identifying MA TV ads that balanced **sensitivity** and **specificity**

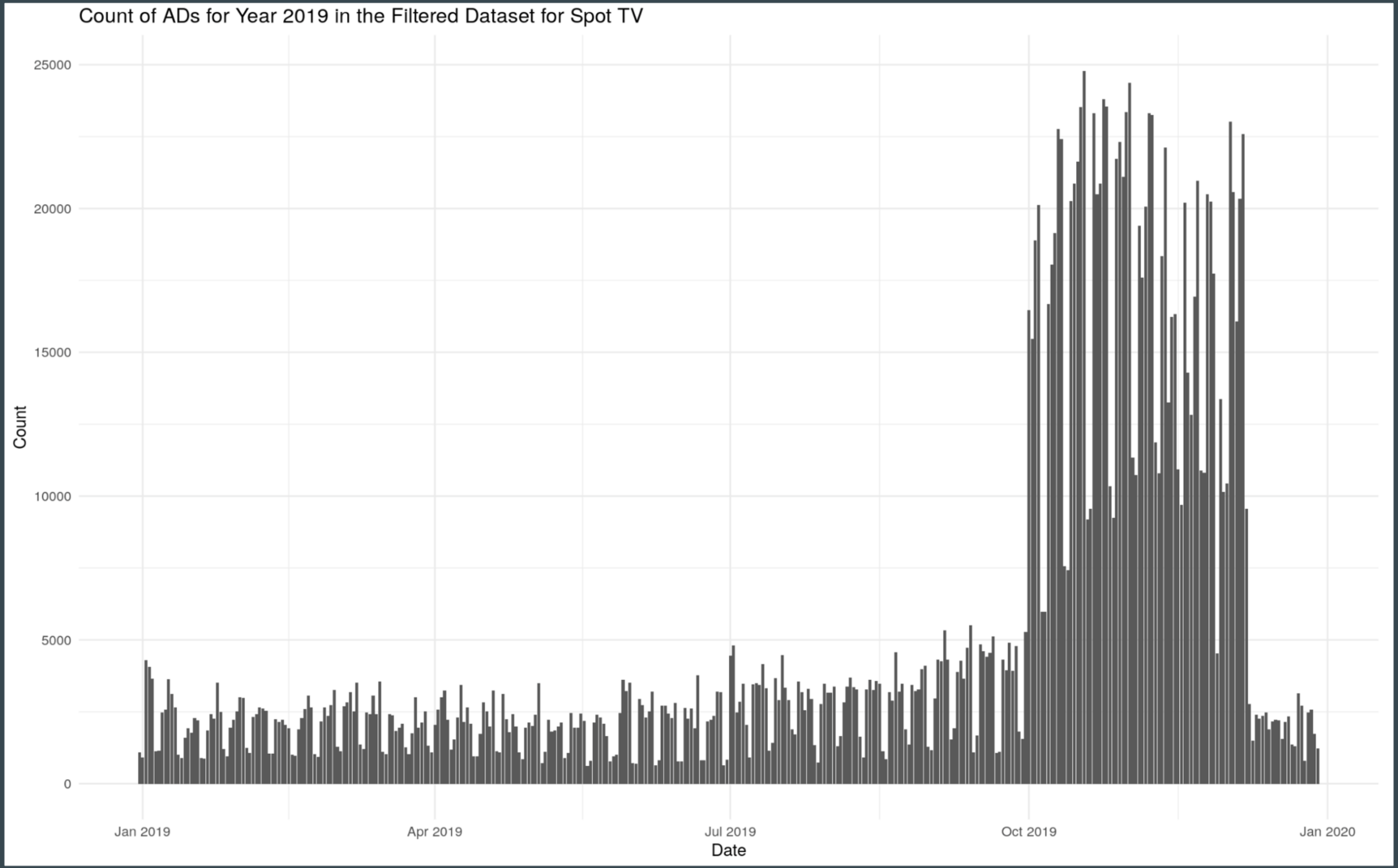
1. Filtering brands and advertisers for "MEDICARE" and "MEDICARE ADV"
   a. Too specific, sensitive enough
2. Using known MA companies as advertisers to identify MA-related health insurance brands
   a. Specific, but not sensitive enough
3. Filtering brands by MA-related keywords to identify MA-related advertisers and ultimately all MA-related brands
   a. Keywords: "MEDICARE", "MEDICARE ADV", "SENIOR", "65"

Results

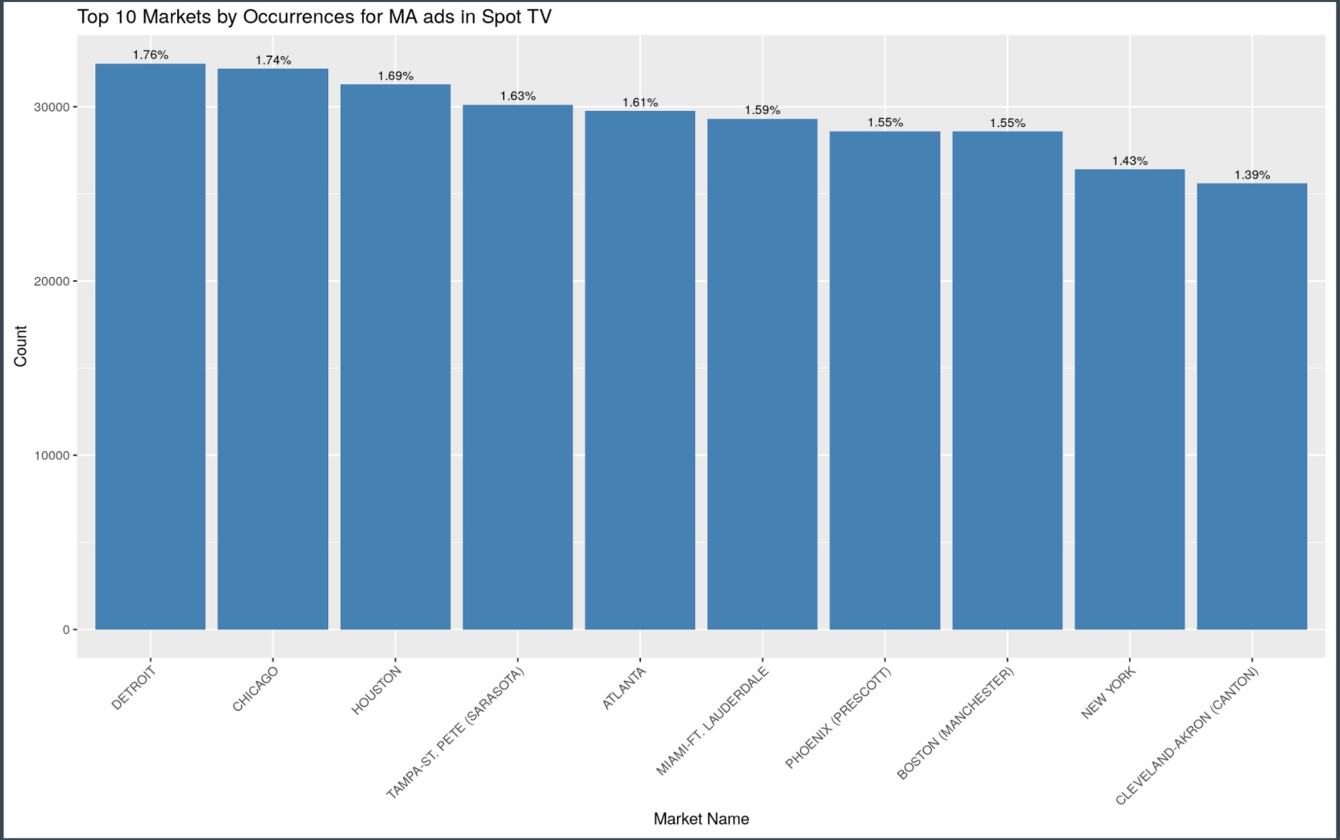


Count of ADs for Year 2019 in the Filtered Dataset for Network TV
Count
1000
750
500
250
0
Jan 2019
Apr 2019
Jul 2019
Oct 2019
Jan 2020
Date

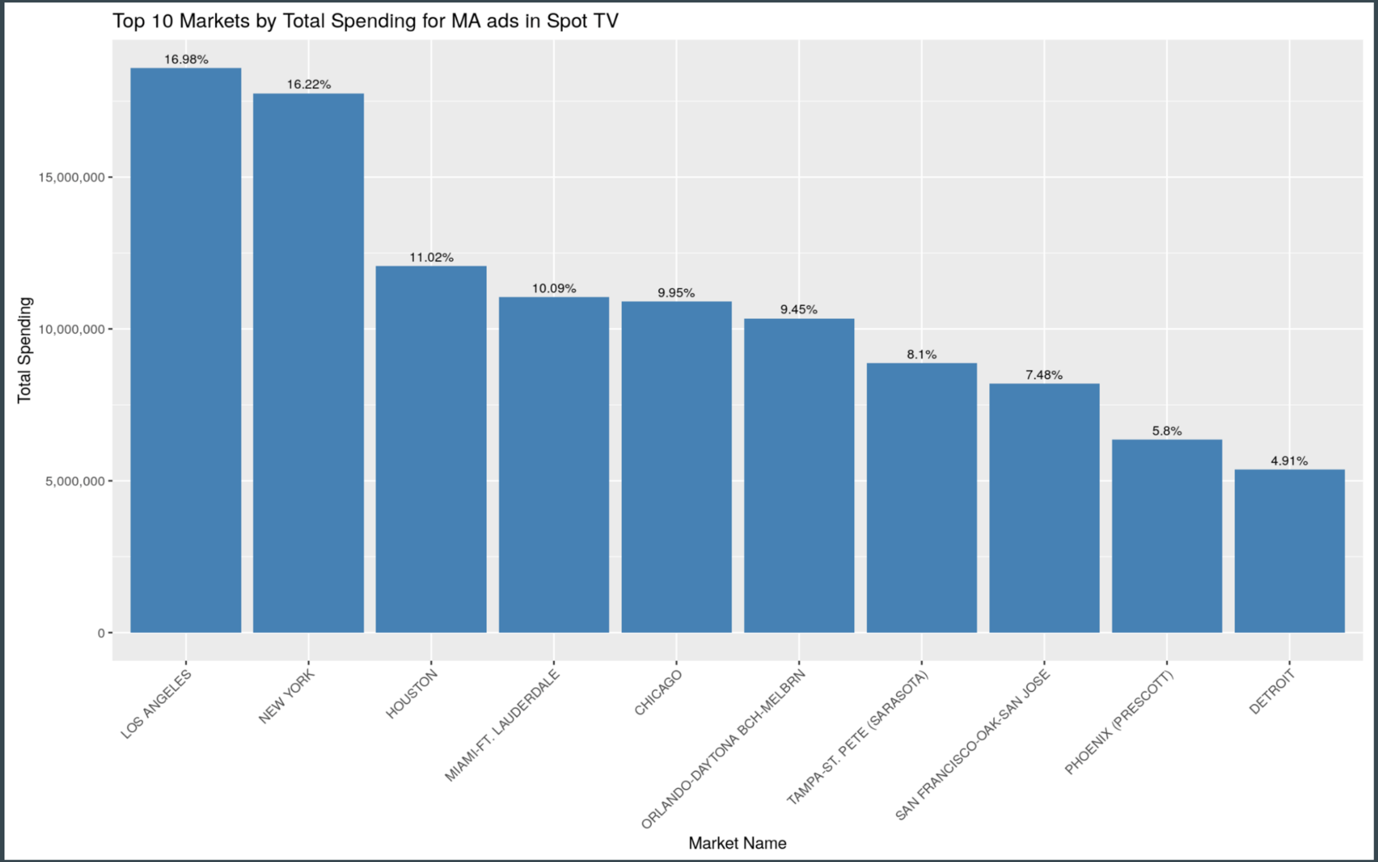
Results


Count of ADs for Year 2019 in the Filtered Dataset for Spot TV
25000
20000
15000
10000
5000
0
Count
Jan 2019
Apr 2019
Jul 2019
Oct 2019
Jan 2020
Date

Results


Top 10 Markets by Occurrences for MA ads in Spot TV
Count
30000
20000
10000
0
1.76%
1.74%
1.69%
1.63%
1.61%
1.59%
1.55%
1.55%
1.43%
1.39%
DETROIT
CHICAGO
HOUSTON
TAMPA-ST. PETE (SARASOTA)
ATLANTA
MIAMI-FT. LAUDERDALE
PHOENIX (PRESCOTT)
BOSTON (MANCHESTER)
NEW YORK
CLEVELAND-AKRON (CANTON)
Market Name

Results



Top 10 Markets by Total Spending for MA ads in Spot TV
Total Spending
15,000,000
10,000,000
5,000,000
0
16.98%
16.22%
11.02%
10.09%
9.95%
9.45%
8.1%
7.48%
5.8%
4.91%
LOS ANGELES
NEW YORK
HOUSTON
MIAMI-FT. LAUDERDALE
CHICAGO
ORLANDO-DAYTONA BCH-MELBRN
TAMPA-ST. PETE (SARASOTA)
SAN FRANCISCO-OAK-SAN JOSE
PHOENIX (PRESCOTT)
DETROIT
Market Name

Discussion

- Future steps:
  - Perform filtering strategies on Nielson datasets for the years 2017 and 2018
  - Conduct filtering analyses on Nielsen datasets for a broader range of ad mediums, beyond just Network TV and Spot TV
  - Using parallel processing to speed up computing
- Limitations:
  - We cannot perform machine learning analyses due to the structure of our datasets, no meaningful variables to predict or model
  - Identifying criteria requires further validation

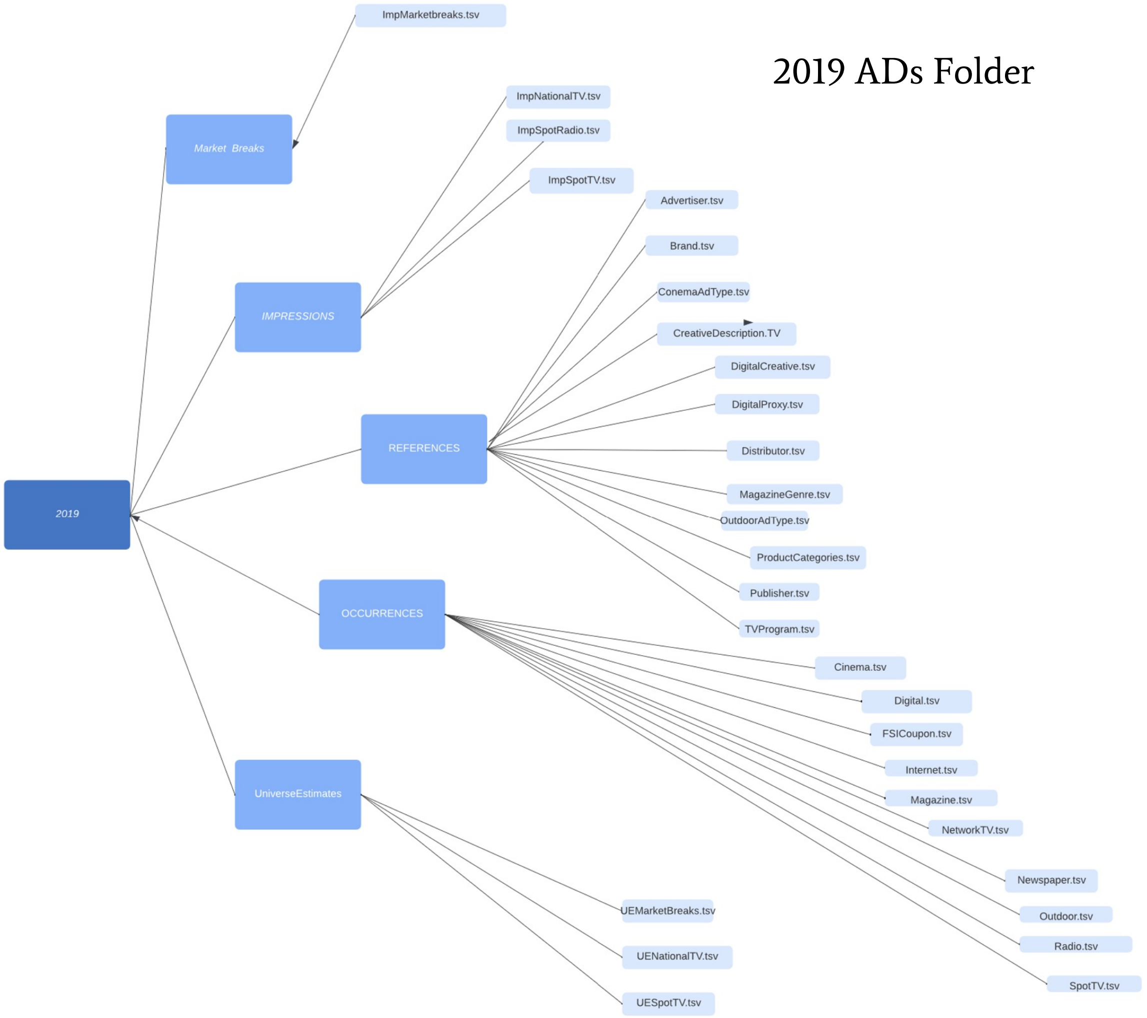
Discussion: Big Data

The Nielsen database is big and complex.

To manage the size and breadth of the database, we...

- Focused on 2019 data and TV ads: Network TV (~4GB) and Spot TV (~20GB) ad occurrences
- Used OSCAR to run R code and explore the files
    - High processing power, speed, and RAM
- Used the Vroom package to load in ~318mil entries in ~3 mins
- Optimized code: select columns, computing calculations

2019 ADs Folder



Machine Learning Approaches

- Initial thought: using k-modes clustering to help identify MA TV ads
    - Attaching factor labels by frequency of MA-related terms to cluster brand and/or advertiser data
    - Limitations: descriptions are short, having to pursue natural language processing or regex processing to create factors

→ Conclusion: resources could be better used elsewhere!

- Future ML applications: predicting which TV ad occurrences are actually MA-related based on a validated training dataset, using ML-assisted Multiple Correspondence Analysis (MCA) or k-modes clustering during MA utilization analysis

Special thanks to Dr. Andrew Ryan, Dr. David Meyers, and Haroon Janjua for expert domain knowledge and technical mentoring. Our work was only made possible because of Arlene Correa, Ashley Gomez, and Dinka Morillo at the Center for Advancing Health Policy through Research (CAHPR), Brown School of Public Health.

Thank You for Listening! Questions?