Mining Google Trends for Entertainment Analytics

Qing Wu,
Principal Economist, Google
Los Angeles, May 23, 2019
Google Trends: Vodka and Hangover
1. Consumer Insights
2. Predictions
3. Causal inferences
Google and Youtube Search Trends for Movies

Mad Max: Fury Road

Danish Girl
Geo Variances: Madea Family Funeral
Spread of Trend: K-Pop Girl Band Blackpink
1. Consumer insights
2. Predictions
3. Causal inferences
Can we use Google trends to help with predictions in the entertainment industry?

- Opening weekend boxoffice
- TV viewership
- Video game sales
- Concert ticket sales
Machine Learning: Opening Weekend Boxoffice Prediction

**Features**
- Google/Youtube trends
- Youtube trailer stats
- Number of theatres
- MPAA rating
- Genre
- Studio type
- Sequel & franchise
- Themes: kids, african.american, musical
- Inferred features:
  - Holiday, special.day, year
  - Number of big movies
  - Big blockbuster last week

**ML Methods**
- Regressions
  - Elastic net
  - Spike-Slab
- Random forest
- Xgboost
- Rulefit
- Neuralnet
- Ensemble
- Challenges:
  - Small training set
  - Large vs small films

Iterative
Ensemble Did the Best Again!

Accuracy increases significantly in the final week before release.

Accuracy flat between 1.5 months and 3 months.
Insights: Variable Importance from Random Forest

importance chart for trend model

importance chart for trailer model
Impact of Rotten Tomato Score
1. Consumer insights
2. Predictions
3. Causal inferences
Randomized control and experiment is the ultimate approach to draw causal inference and estimate the advertising effectiveness.
Geo Experiment Process

Test Design
Optimize parameters (duration, accuracy, geos) to maximise statistical power.

Experiment
Geo-target ad-spend to randomized sub-regions and expose changes.

Analysis
Measure incremental revenue and visits and draw effectiveness conclusion.
Designing an Experiment

- Non-overlapping sub-regions
- Consistent with ad serving
- Similarly sized in terms of population / queries
- Minimize contamination between GMAs
- Matching pre-treatment trends
Time Based Methodology (CausalImpact)

- Pre test
- Test

- Observed response metric in the test regions
- Counterfactual prediction based on control regions

Impact X%
But experiments are expensive!
Looking for an “natural experiment”
An event where ad exposures are randomly boosted in some cities but not others!
Using Super Bowl Fans as the Instrument

- 10-20% more viewers in home cities of competing teams
- Advertisers purchase ad slots long before it is known who will compete
- In 2 cities “randomly chosen” out of 32 cities with NFL teams we see a 10-20% elevation in viewership

<table>
<thead>
<tr>
<th></th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston Final TV Rating:</td>
<td>56.7</td>
<td>Baltimore Final TV Rating:</td>
</tr>
<tr>
<td>Baltimore TV Rating:</td>
<td>47.9</td>
<td>Baltimore TV Rating:</td>
</tr>
</tbody>
</table>
Procedure from “Super Returns”

- Outcome: Google trends, opening weekend boxoffice
- 10 Super Bowl games from 2004-2013.
- 72 movies that advertised during games.
- 72 placebo movies that did not advertise during game.
- Compare outcomes in home cities of competing teams to outcome in other cities.
Bump in Home City Queries by Week
Back-of-Envelope Impact Estimate

- 21% ticket sales increase
- Average opening weekend revenue: $40M
- Super Bowl slot: $3M
- 40-50% goes to studio
- ROI: 1.4 to 1, or 40%

Robustness Test

- Placebo movies: similar movies that weren’t advertised in the Super Bowl
- Runners up: teams that just missed qualifying for Super Bowl.
- In both cases: no effect.
Weather Determines Whether You Stay Indoors or Not
Difference-in-Difference Method: Traffic vs. Temperature
Weather Impact Example: Heatwave in Netherlands

Heat Wave Impact for Netherlands Computer-Tablet Queries: May 2014

- Movies
- Home furnishings
- Weddings
- Luxury Goods
- Air Travel
- Hotels
- Pharmacy
- Pest Control
- Water Sports

-15.0% -10.0% -5.0% 0.0% 5.0%
But Is that the Case in US?

From February to May warm weather in Washington DC has negative impact on movie searches. But from July to August, it is much more muted. The impact is slightly positive over the weekends as people seek ACs in the theaters.
Using data to get insights and make decisions!

THANKS!