Using Social Media Advertising Data to Monitor Global Migration

Ingmar Weber
March 28, 2019
FACEBOOK ADVERTISING DATA
Audience
Define who you want to see your ads. Learn more.

Locations
Qatar
(25.2014, 51.4473) + 2 km

Age
13 → 65+

Gender
All Men Women

INCLUDE people who match at least ONE of the following

Behaviours > Ex-pats
Lived in India (formerly Expats – India)
Lived in Nepal (formerly Ex-pats – Nepal)

Detailed targeting
and MUST ALSO match at least ONE of the following

Behaviours > Mobile Device User > All Mobile Devices by Operating System
Facebook access (mobile): Android devices

Audience size
Your audience selection is broad. This requires a large budget.
Potential reach: 42,000 people

Estimated daily results
Reach
2.7K-17K
Post Engagement
82-510

Removing Instagram and Audience Network may result in 45% fewer Post Engagement, based on your past campaign performance. We recommend choosing automatic placements for the best results.

The accuracy of estimates is based on factors such as past campaign data, the budget you've entered and market data. Numbers are provided to give you an idea of performance for your budget, but are only estimates and don't guarantee results.

Were these estimates helpful?
http://fb-doha.qcri.org
Expats Across US States

Fraction of expats on Facebook 2017 vs. Fraction of foreign-born in the ACS 2014

- FB underestimates
- FB overestimates

- Mexicans in CA
- Mexicans in NM
- Filipinos in HI
Age-Specific Selection Biases

Mexican men in California

Mexican women in California

Mexican men in Texas

Mexican women in Texas

Fraction of state population

Fraction of state population

Fraction of state population

Fraction of state population

SOURCES: American Community Survey (ACS 2014); Facebook Adverts Manager.
Bias Reduction via Model-Fitting

\[
\log(\text{ACS foreign-born population})_{zij}
\]

- \( z \) = age-gender group
- \( i \) = country of birth
- \( j \) = US state of residence

Mean out-of-sample absolute percentage error 37%, down from 56% without origin-age bias correction

Adjusted \( R^2 = 0.70 \)

Does not use GDP, language, internet penetration...
Expats Across Countries

log(Fraction of immigrants from World Bank data) 2015

log(Fraction of immigrants on Facebook) 2017

Continent
- Africa
- Asia
- Europe
- Latin America
- North America
- Oceania

regression line
https://tinyurl.com/FB-Georgians-abroad
LINKEDIN ADVERTISING DATA
Studied in X, Lives in Y

- Compile a list of all universities for European countries
- Query number of LinkedIn users who studied in country X who now live in country Y
- Disaggregate by gender, age, industry, ...
Advertising Audience Estimates

+ Global reach with over 2 billion users
+ FB, LinkedIn, Google, Snapchat, IG, ...
+ Real-time estimates
+ Uses anonymous and aggregate data
+ Good for relative comparisons (densities/trends)
+ Proxies for education and income
+ Non-traditional attributes such as interests
Advertising Audience Estimates

- Black box on how attributes are inferred
- Needs modeling for bias correction
- Hard to obtain absolute numbers
- Usage patterns change over time
- Black box changes over time
- Only includes people who are online

Useful to augment, not replace, traditional data sources. Comes with uncertainty – beware point estimates!
Other Data Sources

• Yahoo data with geo-located IP addresses
  • 10-100’s of millions of users – but hard to get
• Google+ and “places lived”
  • Cute, but no longer around
• Geo-tagged tweets
  • More biased but more fine-grained
Thanks!