Detecting Trending Venues Using Foursquare's Data

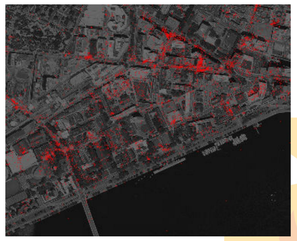
Stephanie Yang, Max Sklar

The Problem

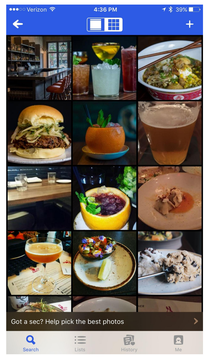
Foursquare is a search and discovery tool which helps users discover venues around the world. In order to keep recommendations fresh, we developed a **trendiness score** to promote venues that are new or have had a recent surge in activity.

The Data

We have several different signals that measure user interactions with venues. The raw signals are the daily counts of each type of interaction.

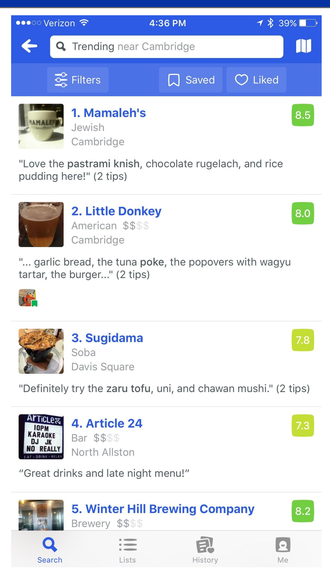


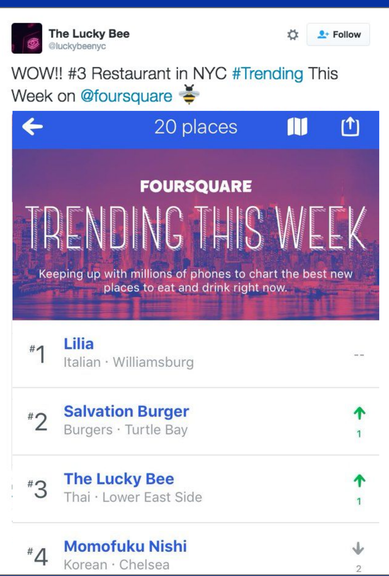
User **visits** come from explicit checkins in our location sharing app Swarm as well as our passive location technology (above).



Users often upload **photos** of their favorite venues (above). This is a great signal to indicate increasing buzz around a place.

Reviews come in the form of explicit ratings (right) and tips which are short snippets of text.





TRENDING THIS WEEK

New York City

Keeping up with millions of phones to chart the best new places to eat and drink right now.
For the week of August 29, 2016. Updated every Tuesday at 9:01am ET.

#	Last week		Neighborhood
1.	–	City Vineyard at Pier 26 Wine Bar	Tribeca
2.	▲ 2	Emmy Squared Pizza Place	Williamsburg
3.	▼ 1	Pondicheri Indian Restaurant	NoMad
4.	▲ 9	The PokéSpot Hawaiian Restaurant	Greenwich Village
5.	▲ New	Blacktail Cocktail Bar	Battery Park City
6.	▼ 1	Merriweather Coffee Shop	West Village
7.	▲ 5	Atoboy Korean Restaurant	NoMad
8.	–	Paowalla Indian Restaurant	SoHo
9.	▼ 3	The Tang Noodle House	East Village
10.	–	Casa Apicii Italian Restaurant	Greenwich Village

The Signals

For each source of data, we calculate two types of signals.

The first type of signal is the **slope of the regression line** fit to the timeseries, divided by the standard error of the slope.

$$S = \frac{\hat{\beta}}{\sigma_{\hat{\beta}}}$$

The second type of signal is a **weighted decayed sum** of the aggregate activity over a time window in the past.

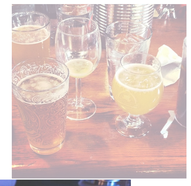
$$D = \sum_d c_d e^{\lambda d}$$

Both types of signals are **normalized to a Gaussian distribution** before being combined into a final sum.

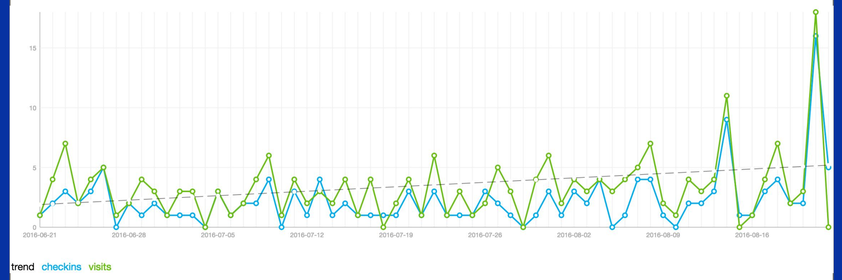
$$N = \Phi^{-1}(r)$$

The Hottest Spot in Boston

The hottest spot in Boston the summer of 2016 was Picco, a pizza joint in the South End which offers a fine selection of draft beers.



Summer weather has led to a noticeable increase in foot traffic to many places which offer outdoor seating, including Picco.



Fine Tuning

Noisy results ⟷ Stale results

Causes:
- λ is too negative
- Too little data

Causes:
- λ is too positive
- Too much data

The signals are balanced to ensure both quality and freshness.

If we consume too little data, or set the half-life of the signals to be too small, then low-quality venues appear to be trendy.

The opposite problem arises from consuming too much data. In this case, newer venues rarely appear and the weekly lists become stale since they display the same results for extended periods of time.