Title: Professional and crowd-sourced critics: tracking influencers' impact on restaurant reviews and outcomes

Problem

The digital space is crucial to bringing in business for restaurants, and many platforms exist for broadcasting food opinions that affect this bottom line. These platforms are so diverse and numerous that it may seem impossible to disentangle whether a single mention by a large news organization, such as the New York Times, or positive early reviews by customers on Yelp, are responsible for kicking off a restaurant's success (as measured by restaurant reviews/ratings and popularity vs. closure). Profit margins are already notoriously slim and closures a regular occurrence in the competitive and stressful restaurant industry, and making headway towards this goal could allow restaurants to better adapt to the changing ways with which people consume, share, and decide where to spend their dollars in the food industry.

Client & Pitch

New York Times and other local/national news outlets with a dining reviews section. Such outlets would be interested in determining the impact of their restaurant critics' reviews on Yelp reviews, as well as restaurant outcomes (success or closure). Authorship and consumption of restaurant reviews have increasingly transferred to the digital space, and "traditional" news outlets that are widening their focus and looking to adapt to changing consumer habits will benefit from metrics that can quantify the impact of their dining reviews.

Restaurants and food businesses. Understanding how different platforms bring business to restaurants, and how these platforms interact with each other, provide valuable marketing insights that can be turned into strategies that better leverage the growing importance of social and digital media to restaurants/food businesses.

Dataset

This project focuses on restaurant reviews in New York City from 2012 to 2018. Although Yelp was started in 2005, there were several turning points in its timeline that may confound its impact on local businesses. In its early years, the number of users and reviews experienced rapid growth. In 2006 Yelp only had ~1 million unique monthly users and 100,000 reviews. In 2008, when the Yelp mobile app was released, this statistic was up to ~15.8 million unique monthly users (and from 12 to 24 cities, compared to 2007). Finally, in 2012 Yelp's stock began publicly trading - this is the year we will start with for analysis.

Our data is gathered from two sources: 1) Yelp API, and 2) New York Times restaurant reviews.

We obtain NYC Yelp reviews from Yelp's API, rather than the well-known academic dataset that Yelp has released on its website. This is due to the fact that the latest academic dataset is not complete and does not actually contain many NYC Yelp reviews.

To assemble a database of NYT restaurant reviews, we use a combination of automatic and manual methods. We copy HTML from NYT's dining section, use BeautifulSoup to extract basic information, and manually input review data since NYT's layout of its restaurant reviews makes it difficult to scrape this information.

Approach

Determine patterns in the timeline of restaurant reviews (Yelp and NYT). Compile reviews, mentions, and shares across two platforms: 1) a crowd-sourced, social media platform: Yelp, and 2) a more traditional, professional critic news outlet: New York Times, to determine whether patterns within and between these platforms help popularize local restaurants.

Within-platform patterns include a 4/5 or 5/5 by a well-known, professional NYT restaurant critic, or an initial wave of >4 star reviews on Yelp during a restaurant's opening weeks. Between-platform patterns include whether the aforementioned early success on Yelp increases the likelihood that NYT will pick up on the establishment and write a positive review, or vice versa.

If possible, cluster Yelpers and food critics by different features, such as whether they are Yelp Elite or number of NYT food reviews, and how these features interact with their ability to influence local restaurants' ratings.

Apply NLP/sentiment analysis. Determine whether NYT and Yelp reviews influence each other, and whether certain word frequencies tend to correlate with positive/negative reviews. Influence between platforms (between NYT and Yelp) is determined by tracking frequently used words in Yelp reviews and whether this pattern changes after the onset of a NYT review (or vice versa).

Deliverables

1. Code:
    - Data acquisition: 1) a database of New York City Yelp reviews from 2012-2018, and 2) a corresponding database of New York Times restaurant reviews
    - Data wrangling to set up NLP, integrate Yelp and NYT data, and perform time-course analyses
    - Word frequencies over time and correlations with positive/negative reviews

2. Report on capstone project

3. Presentation on capstone project (slides and final report)