Survival of Gentrification / Depreciation in Restaurants

F. M. Marsh and J. A. Clithero

September 4, 2016

1 Abstract

2 Introduction

Yelp Dataset Papers:

[Alghunaim, 2015, Byers et al., 2012, Cawkwell et al., 2015, Chepurna and Makrehchi, 2015, Feng and Qian, 2013, Gutierrez, 2014, Hajas et al., 2014, Hu et al., 2014, Liu et al., 2015, Mashhadi et al., 2012, Quattrone et al., 2015]

Zillow Dataset Papers:

Has previously been combined with the yelp dataset Bonnar et al..

Our goal is to use the correlation of two time-series:

1. The monthly median rent, as tracked by Zillow Rental Data. 2. The median restaurant review rating (stars) for each restaurant in a neighborhood.

Zillow rental data can be used to detect appreciating, and depreciating neighborhoods.

As rents rise in a given neighborhood, which types of businesses fare / worse better in the reviews? As rents fall in a given neighborhood, which types of business fare / worse better in the reviews?

We hope to present concrete suggestions to restaurant owners to improve the survivability of their businesses in times of strong appreciation / depreciation in the housing market.

3 Data

3.1 Yelp Academic Dataset

The Yelp Academic Dataset contains five files.

3.2 Zillow Public Dataset

The Zillow Public Dataset (hereafter Zillow dataset) contains many files.

Zillow divides homes into geographic “neighborhoods” with well defined boundaries. The ZHVI, and rental data is reported for every neighborhood, on a monthly basis.

4 Methods

4.1 Combination of Datasets

Each Yelp business is tagged with a geographic (latitude, longitude) coordinate. Each Zillow neighborhood has geographic boundaries. We sort each Yelp business into its appropriate Zillow neighborhood.

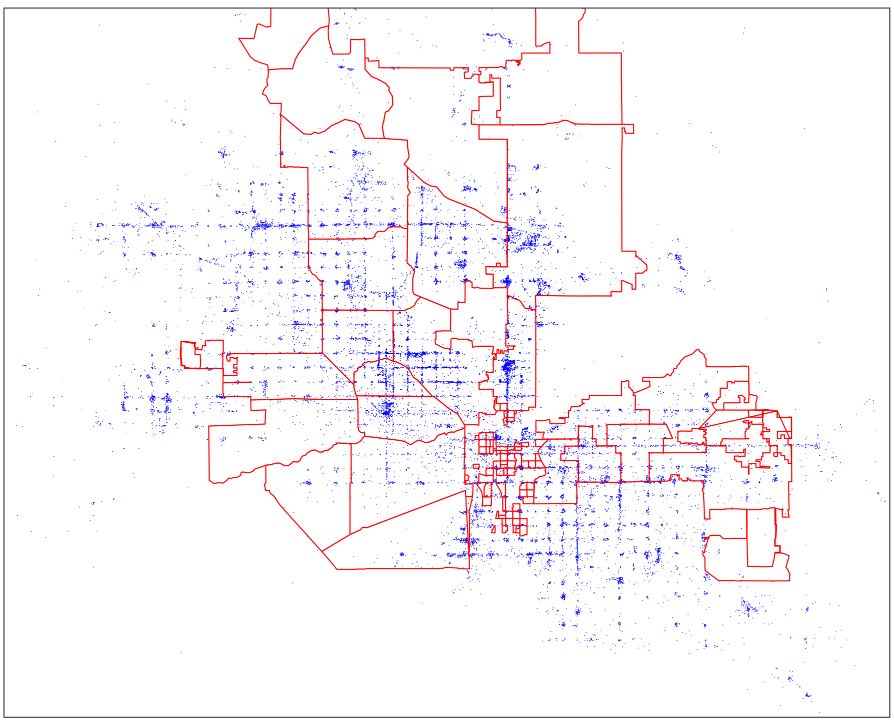

Figure 1: default

4.2 Determination of Most Common Yelp Tags

Each Yelp business has user-generated tags, that allow other users to determine what genre the business is. For restaurants, common tags are "Mexican", "Chinese", etc.

4.3 Computation of Price Trend

References

A. Alghunaim. *A Vector Space Approach for Aspect-Based Sentiment Analysis*. PhD thesis, Massachusetts Institute of Technology, 2015.

C. Bonnar, F. Cordeiro, and J. Michelman. With a little help from yelp.

J. W. Byers, M. Mitzenmacher, and G. Zervas. The groupon effect on yelp ratings: a root cause analysis. In *Proceedings of the 13th ACM conference on electronic commerce*, pages 248–265. ACM, 2012.

P. B. Cawkwell, L. Lee, M. Weitzman, and S. E. Sherman. Tracking hookah bars in new york: Utilizing yelp as a powerful public health tool. *JMIR public health and surveillance*, 1(2), 2015.

I. Chepurna and M. Makrehchi. Exploiting class bias for discovery of topical experts in social media. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 64–71. IEEE, 2015.

H. Feng and X. Qian. Recommendation via user's personality and social contextual. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1521–1524. ACM, 2013.

L. A. Gutierrez. *Noise reduction in user generated datasets*. PhD thesis, RENSSELAER POLYTECHNIC INSTITUTE, 2014.

P. Hajas, L. Gutierrez, and M. S. Krishnamoorthy. Analysis of yelp reviews. *arXiv preprint arXiv:1407.1443*, 2014.

L. Hu, A. Sun, and Y. Liu. Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 345–354. ACM, 2014.

J. Liu, J. Shang, C. Wang, X. Ren, and J. Han. Mining quality phrases from massive text corpora. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1729–1744. ACM, 2015.

A. Mashhadi, G. Quattrone, L. Capra, and P. Mooney. On the accuracy of urban crowd-sourcing for maintaining large-scale geospatial databases. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, page 15. ACM, 2012.

G. Quattrone, L. Capra, and P. De Meo. There's no such thing as the perfect map: Quantifying bias in spatial crowd-sourcing datasets. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1021–1032. ACM, 2015.