

## PREDICTING USEFUL YELP REVIEWS

On March 27, 2013, Yelp launched a competition on the public data science competition website Kaggle.<sup>1</sup> The challenge was to design a new algorithm for predicting how many “useful” votes a Yelp review would receive. To the winner would go a coveted prize: an interview for a position on a Yelp data mining team.

Yelp’s website and mobile application offered people a way to search for local businesses and read reviews about those businesses. In order to show reviews in a meaningful order, Yelp allowed users to vote on a review’s quality as useful, funny, or cool. Because a review’s freshness was also extremely important to users, Yelp wanted to estimate the number of useful votes a review would receive before waiting to see the user community’s votes.

### Yelp and its Review Ordering

Founded in 2004, Yelp helped people find local businesses such as restaurants, dentists, and mechanics. Yelp was heavily trafficked. In the first quarter of 2013, Yelp averaged 102 million monthly unique visitors. Not only could a user search for a type of business near their current location and find a listing of those businesses in their area, users could also write reviews about the businesses they experienced. Since inception, users had written over 39 million reviews of businesses.<sup>2</sup>

Although Yelp made money by selling ads to local businesses, it did not allow paying advertisers to change or reorder its reviews. Yelp used an algorithm to determine the order in which reviews were shown. Its goal was “to present reviews in a meaningful order.”<sup>3</sup> Yelp also allowed users to sort reviews based on date, rating, voting, etc.

Showing users useful reviews was a critical part of Yelp’s value proposition. Because reviews were anonymous and unverifiable, their quality varied. Some reviews were thoughtful

---

<sup>1</sup> Kaggle website: <http://www.kaggle.com/c/yelp-recruiting>. Accessed July 14, 2013.

<sup>2</sup> Yelp website: <http://www.yelp.com/about>. Accessed July 14, 2013.

<sup>3</sup> Yelp website: [http://www.yelp.com/faq#review\\_ordering](http://www.yelp.com/faq#review_ordering). Accessed July 14, 2013.

commentaries that detailed a user's experience at a business, while others were anger-filled rants or fake reviews written by friends of the business owner. To help determine which reviews were the best, Yelp added three buttons at the bottom of each review to let users vote on the review's quality. A user could vote on whether the review was useful, funny, or cool. (See **Exhibit 1** for a sample listing from Yelp's website.)

## Yelp's Challenge to the Crowd

Yelp's ability to identify and prioritize high quality reviews had big implications for improving the experience of its users. To that end, Yelp posted a challenge to Kaggle's crowd of data scientists. The goal was to predict the number of useful votes a review would receive.

For the competition, Yelp provided a training set of 229,907 reviews of businesses near Phoenix, Arizona, written between March 7, 2007 and January 5, 2013. The training set included useful votes, funny votes, and cool votes, all measured as of January 19, 2013. Also included in the training set were data about the review writers and the businesses reviewed. The testing set contained 22,956 reviews without votes. The useful votes to be predicted in the testing set were measured as of March 12, 2013.

Immediately after an individual submitted his or her test-set predictions, Kaggle's system automatically scored 30% of the predictions and posted the individual's name and rank on its *public* leaderboard. An individual could submit as many predictions as he or she wanted. At the end of the competition, Kaggle would score the other 70% of each competitor's most recent predictions and announce the winner on its *private* leaderboard. The data scientist with the lowest average root mean squared logarithmic error (RMSLE) would be the winner.<sup>4</sup>

In the end, 352 individuals competed, and Yannis Sismanis (aka Outis) was declared the winner. His RMSLE was 0.44049. And even though the competition ended on June 30, 2013, Kaggle left its submission box open, which gave latecomers the opportunity to hypothetically compete against some of the best data scientists in the world.

Yelp allowed its data to be used for academic purposes under special terms. (See **Exhibit 2** for Phoenix Academic Dataset Terms of Use.) Before working with company data, a student had to agree to its terms.


---

<sup>4</sup> Kaggle used root mean squared logarithmic error (RMSLE) to measure the accuracy of a competitor's useful vote predictions:  $\sqrt{\frac{1}{n} \sum_{i=1}^n (\log(p_i+1) - \log(a_i+1))^2}$  where  $n$  is the number of reviews in the testing set,  $p_i$  is the predicted number of useful votes for review  $i$ ,  $a_i$  is the actual number of useful votes for review  $i$ , and  $\log(x)$  is the natural logarithm of  $x$ .

## Exhibit 1

## PREDICTING USEFUL YELP REVIEWS

## Yelp Listing for Bodo's Bagels in Charlottesville, VA



[Home](#)
[About Me](#)
[Write a Review](#)
[Find Friends](#)
[Messages](#)
[Talk](#)

### Bodo's Bagels

78 reviews [Rating Details](#)

Categories: Bagels, Breakfast & Brunch

1418 Emmet St N  
Charlottesville, VA 22903

(434) 977-9598  
<http://www.bodosbagels.com>

[Explore the Menu](#)

Hours:  
Mon-Fri 6:30 am - 8 pm  
Sat 7 am - 8 pm  
Sun 8 am - 4 pm

Good for Kids: Yes  
Accepts Credit Cards: Yes  
Parking: Private Lot  
Attire: Casual  
Good for Groups: Yes

Price Range: \$  
Takes Reservations: No  
Delivery: No  
Take-out: Yes  
Waiter Service: No  
Outdoor Seating: No  
Wi-Fi: No  
Good For: Breakfast, Lunch

Alcohol: No  
Noise Level: Loud  
Ambience: Casual  
Has TV: No  
Caters: No  
Wheelchair Accessible: Yes

[Add Photos](#)

[Edit Business Info](#)

[First to Review](#) Erik F.

[Send to Friend](#) [Bookmark](#) [Send to Phone](#) [Write a Review](#)

[Increase Website Ranking](#)  
 SEO Specialists. Get Top Rankings. Get Free Site Analysis & SEO Report  
[www.FindMyCompany.com](http://www.FindMyCompany.com)

[Cville's Best Meals](#)  
 True Gourmet-To-Go, Amazing Bakery, Reasonably-Priced Gourmet Cafe  
[www.Hotcakes.biz](http://www.Hotcakes.biz)

Ads by Google

### 78 reviews for Bodo's Bagels

#### Review Highlights

[What's this?](#)

"Definitely look forward to my Saturday morning Bodo Bagel."  
In 47 reviews

"The bagels were always fresh and delicious."  
In 5 reviews

"Whenever I am in the C-Ville area, this will always be my first stop."  
In 4 reviews

Rating Distribution | Trend

5 stars: [Bar chart showing 5 stars is the most common rating]

4 stars: [Bar chart showing 4 stars is the second most common rating]


3 stars: [Bar chart showing 3 stars is the third most common rating]

2 stars: [Bar chart showing 2 stars is the fourth most common rating]

1 star: [Bar chart showing 1 star is the least common rating]

Sort by: [Yelp Sort](#) | [Date](#) | [Rating](#) | [Elites](#) | [Facebook Friends](#)

78 reviews in English

 **Enoch C. Raleigh, NC** 28 helpful votes 27 stars 5/21/2013

I had been hearing how amazing Bodo's was and finally had the opportunity to try it.

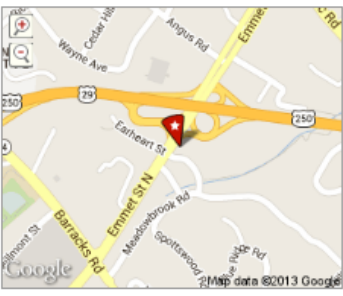
Upon scooping out then menu, I was pleasantly surprised by how inexpensive everything was, in addition to the variety of sandwiches offered.

I was super hungry so I ordered two bagel sandwiches: One bacon-egg-cheddar on an everything bagel, and one chicken salad on a poppy seed bagel. Both were simple, yet so delicious! The bagels have the perfect paradoxical texture of crunchy-softness. Bagel umami-ness!!!

The service was kind and very quick to take and produce the order. I completely understand why Bodo's is a staple in Charlottesville and why every location I drive by seems to be packed, regardless of the time of day.


Was this review ...? [Useful](#) [Funny](#) [Cool](#)

[Bookmark](#) [Send to a Friend](#) [Link to This Review](#) [Add owner comment](#)



[View Larger Map/Directions](#)

**Browse Nearby:**  
[Restaurants](#) | [Nightlife](#) | [Shopping](#) | [Movies](#) | [All](#)



**CANTINA DOUBLE STEAK QUESADILLA**  
 #CANTINAQUESADILLA  
 GET YOURS

At participating locations. ©2013 Taco Bell Corp. Limited time only. Double the steak, compared to 1 soft, portofino steak taco.

#### People Who Viewed This Also Viewed...

**Bodo's Bagels**  
 78 reviews  
 "The bagel sandwich was delicious and the staff was friendly."

**Bluegrass Grill & Bakery**  
 105 reviews  
 "Pig candy is intriguing if you like bacon with a kick."

**Blue Moon Diner**  
 80 reviews  
 "with a side of homemade sweet potato fries."

**Bodo's Bagels**  
 85 reviews  
 "This is definitely the home of the perfect bagel."

**Shenandoah Joe Coffee Roasters**  
 41 reviews  
 "Try an iced Americano or a traditional cappuccino."

#### Related Lists

[Curbing my Coffee Fix](#)  
 Here are my reviews of coffee shops, both local and in my travels

[More »](#)

Exhibit 2

**PREDICTING USEFUL YELP REVIEWS**

Phoenix Academic Dataset Terms of Use

**PHOENIX ACADEMIC DATASET TERMS OF USE**

*Last Updated: March 11, 2013*

This document governs the terms under which you may access and use the data that Yelp makes available for download through this website for academic purposes (the “Data”). This document incorporates the terms of the following additional document, including all future amendments or modifications thereto (collectively, and together with this document, the “Data Agreement”):

- [Yelp Terms of Service](#)

By accessing or using the Data, you agree to be bound by the Data Agreement and represent that the contact information you provide to Yelp is correct. If you access or use the Data on behalf of a university, school, or other entity, you represent that you have authority to bind such entity and its affiliates to the Data Agreement and that it is fully binding upon them. In such case, the term “You” and “Your” will refer to such entity and its affiliates. If you do not have authority, or if you do not agree with the terms of the Data Agreement, you may not access or use the Data. You should read and keep a copy of each component of the Data Agreement for your records. In the event of a conflict among them, the terms of this document will control.

**1. Purpose**

The Data is made available by Yelp Inc. (“Yelp”) to enable you to access valuable local information to develop an academic project as part of an ongoing course of study. With this in mind, Yelp reserves the right to continually review and evaluate all uses of the Data provided under the Data Agreement.

**2. Changes**

Yelp reserves the right to modify or revise the Data Agreement at any time. You should visit this website periodically to review the documents comprising the Data Agreement and check for updates. Your continued use of the Data after the effective date of any such changes will constitute your acceptance of and agreement to such changes. **IF YOU DO NOT WISH TO BE BOUND TO ANY NEW TERMS, YOU MUST TERMINATE THE DATA AGREEMENT BY IMMEDIATELY CEASING USE OF THE DATA AND DELETING IT FROM ANY SYSTEMS OR MEDIA.**

**3. License**

Subject to the terms set forth in the Data Agreement, Yelp grants you a royalty-free, non-exclusive, revocable, non-sublicensable, non-transferable, fully paid-up right and license to use, access, and create derivative works of the Data in electronic form for academic purposes only.

## Exhibit 2 (continued)

You may not use the Data for any other purpose without Yelp's prior written consent. You acknowledge and agree that Yelp may request information about, review, audit, and/or monitor your use of the Data at any time in order to confirm compliance with the Data Agreement. Nothing herein shall be construed as a license to use Yelp's registered trademarks or service marks, or any other Yelp branding.

### **4. Restrictions**

You agree that you will not, and will not encourage, assist, or enable others to:

- A. use the Data in connection with any commercial purpose;
- B. display, perform, or distribute any of the Data, or use the Data to update or create Your own business listing information;
- C. use the Data in any manner or for any purpose that may violate any law or regulation, or any right of any person including, but not limited to, intellectual property rights, rights of privacy and/or rights of personality, or which otherwise may be harmful (in Yelp's sole discretion) to Yelp, its providers, its suppliers, end users of this website, or Your end users;
- D. use the Data on behalf of any third party without Yelp's consent;
- E. create, redistribute or disclose any summary of, or metrics related to, the Data (e.g., the number of reviewed business included in the Data and other statistical analysis) to any third party or on any website or other electronic media not expressly covered by this Agreement, this provision however, excludes any disclosures necessary for academic purposes, including without limitation the publication of academic articles concerning your use of the Data;
- F. use the Data in a manner that is competitive in nature with Yelp;
- G. display Data in a manner that could reasonably imply an endorsement, relationship or affiliation with or sponsorship between you or a third party and Yelp, other than your permitted use of the Data under the terms of the Data Agreement;
- H. rent, lease, sell, transfer, assign, or sublicense, any part of the Data;
- I. modify, rate, rank, review, vote or comment on, or otherwise respond to the content contained in the Data;
- J. display the Data in any way, or on any site, that disparages Yelp or its products or services, or infringes any Yelp intellectual property or other rights;
- K. use the Data in a manner that could reasonably be interpreted to suggest that Yelp is the author or entity that is responsible, in whole or in part, for the creation or development of any Data or that such Data represents the views of Yelp; or
- L. use the Data for any purpose prohibited by law;

### **5. Ownership**

As between you and Yelp, the Data and any derivative works You create from the Data, and all intellectual property rights contained in the foregoing, are and will at all times remain the sole and exclusive property of Yelp and are protected by applicable intellectual property laws and

Exhibit 2 (continued)

treaties (whether those rights happen to be registered or not, and wherever in the world those rights may exist.).

**6. Indemnity**

You agree that your use of the Data is at your own risk and you agree to hold harmless, defend (subject to Yelp's right to participate with counsel it selects) and indemnify Yelp and its subsidiaries, affiliates, officers, agents, employees and suppliers from and against any and all claims, damages, liabilities, costs and fees (including reasonable attorneys' fee) arising from, or in any way related to your or your end users' use or implementation of the Data. You will not agree to any settlement that imposes any obligation on Yelp without Yelp's prior consent.

**7. No Warranties by Yelp; No Entitlement to Support from Yelp**

THE DATA IS PROVIDED "AS IS", "WITH ALL FAULTS" AND "AS AVAILABLE" WITHOUT WARRANTY, OF ANY KIND AND AT YOUR SOLE RISK. EXCEPT TO THE MAXIMUM EXTENT REQUIRED BY APPLICABLE LAW, YELP DISCLAIMS ALL WARRANTIES, REPRESENTATIONS, CONDITIONS, AND DUTIES, WHETHER EXPRESS, IMPLIED OR STATUTORY, REGARDING THE DATA, INCLUDING, WITHOUT LIMITATION, ANY AND ALL IMPLIED WARRANTIES OF MERCHANTABILITY, ACCURACY, RESULTS OF USE, RELIABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE, INTERFERENCE WITH QUIET ENJOYMENT AND NON-INFRINGEMENT OF THIRD-PARTY RIGHTS. FURTHER, YELP DISCLAIMS ANY WARRANTY THAT YOUR USE OF THE DATA WILL BE UNINTERRUPTED, SECURE, TIMELY OR ERROR FREE. FOR THE AVOIDANCE OF DOUBT, YOU ACKNOWLEDGE AND AGREE THAT THE DATA AGREEMENT DOES NOT ENTITLE YOU TO ANY SUPPORT FOR THE DATA. NO ADVICE OR INFORMATION, WHETHER ORAL OR IN WRITING, OBTAINED BY YOU FROM YELP WILL CREATE ANY WARRANTY NOT EXPRESSLY STATED IN THE DATA AGREEMENT.

**8. Limitation of Liability**

THE DATA IS BEING PROVIDED FREE OF CHARGE. ACCORDINGLY, YOU AGREE THAT YELP SHALL HAVE NO LIABILITY ARISING FROM OR BASED ON YOUR USE OF THE DATA. REGARDLESS OF WHETHER ANY REMEDY SET FORTH HEREIN FAILS OF ITS ESSENTIAL PURPOSE OR OTHERWISE, AND EXCEPT FOR BODILY INJURY, IN NO EVENT SHALL YELP OR ITS SUBSIDIARIES, AFFILIATES, OFFICERS, AGENTS, EMPLOYEES AND SUPPLIERS BE LIABLE TO YOU OR TO ANY THIRD PARTY UNDER ANY TORT, CONTRACT, NEGLIGENCE, STRICT LIABILITY OR OTHER LEGAL OR EQUITABLE THEORY FOR ANY LOST PROFITS, LOST OR CORRUPTED DATA, COMPUTER FAILURE OR MALFUNCTION, INTERRUPTION OF BUSINESS, OR OTHER SPECIAL, INDIRECT, INCIDENTAL OR CONSEQUENTIAL DAMAGES OF ANY KIND ARISING OUT OF THE USE OR INABILITY TO USE THE DATA, EVEN IF YELP HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH LOSS OR DAMAGES AND WHETHER OR NOT SUCH LOSS OR DAMAGES ARE FORESEEABLE. ANY CLAIM ARISING OUT OF OR RELATING TO THE DATA AGREEMENT MUST BE BROUGHT WITHIN (1) YEAR AFTER THE OCCURRENCE OF THE EVENT GIVING RISE TO SUCH CLAIM. IF SUCH CLAIM IS NOT FILED, THEN THAT CLAIM IS PERMANENTLY BARRED. THIS APPLIES TO YOU AND YOUR SUCCESSORS, AND TO

## Exhibit 2 (continued)

YELP AND ITS SUCCESSORS. NOTWITHSTANDING THE FOREGOING, YELP'S MAXIMUM LIABILITY UNDER THIS DATA AGREEMENT SHALL NOT, IN ANY EVENT, EXCEED US\$50.00.

### **9. Limited Relationship**

Yelp and you are, and will remain, independent contractors, and nothing in the Data Agreement will be construed as creating an employer-employee relationship, partnership or joint venture. Although you are permitted to publicize your use of the Data, you agree not to make any other statements, without the prior written consent of Yelp, implying a different kind of relationship between you and Yelp, including any implied endorsement by Yelp. You do not have any authority of any kind to bind Yelp in any respect whatsoever.

### **10. Term and Termination**

This Data Agreement is effective as of the date You download or otherwise access the Data ("**Effective Date**") and shall continue in full force and effect for an initial term of one (1) year, which shall renew automatically thereafter for two (2) consecutive renewal terms of one (1) year each, unless earlier terminated by the Parties or expires in accordance with this Section 13. Such initial term, together with any renewal terms, is referred to herein as the "**Term**."

You may terminate this Data Agreement, for any reason or for no reason, by providing at least thirty (30) days' written notice to Yelp. Yelp reserves the right, in its sole discretion (for any reason or for no reason) and at any time without notice to you, to change, suspend or discontinue the Data and/or suspend or terminate your rights under the Data Agreement to access, use, create derivative works from, and/or display the Data.

Any termination of the Data Agreement will also immediately terminate the licenses granted to you hereunder. Upon any termination of the Data Agreement, you will promptly: (i) delete and remove all Data from any location, including any web pages, scripts, widgets, applications and any other software in Your possession or under Your control; (ii) destroy and remove from all computers, hard drives, networks and other storage media in Your possession or under Your control all copies of any Data; and (iii) certify in writing to Yelp that such actions have been taken.

### **11. Miscellaneous**

The Data Agreement encompasses the entire agreement between you and Yelp regarding the subject matter discussed therein. The Data Agreement, and any disputes arising from or relating to the interpretation thereof, will be governed by and construed under the laws of the State of California without regard to its conflict of law provisions. You agree to personal jurisdiction by and venue in the state and federal courts of the State of California, City of San Francisco. The failure of Yelp to exercise or enforce any right or provision of the Data Agreement will not constitute a waiver of such right or provision. The failure of either party to exercise in any respect any right provided for herein will not be deemed a waiver of any further

Exhibit 2 (continued)

rights hereunder. If any provision of the Data Agreement is found to be unenforceable or invalid, that provision will be replaced with terms that most closely match the intent of the provision that is not enforceable to the minimum extent necessary so that the remaining Data Agreement will otherwise remain in full force and effect and enforceable. The Data Agreement is not assignable, transferable or sublicensable, in whole or in part, by you except with Yelp's prior written consent. Any attempt to do so is void. Yelp may assign the Data Agreement, in whole or in part, at any time with or without notice to you. The section titles in the Data Agreement are for convenience only and have no legal or contractual effect.

**12. Survival**

Sections 5 through 13 will survive any expiration or termination of this Data Agreement for any reason.

**13. Contact and Violations**

Please contact Yelp with any questions regarding the Data Agreement. Please report any violations of the Data Agreement [here](#).