



Yelp Dataset Challenge

Yelp Dataset Challenge is doubling up: Now 10 cities across 4 countries! Two years, four highly competitive rounds, over \$35,000 in cash prizes awarded and [several hundred peer-reviewed papers](#) later: the Yelp Dataset Challenge is doubling up. We are proud to announce our latest dataset that includes information about local businesses, reviews and users in 10 cities across 4 countries. The Yelp Challenge dataset is much larger and richer than the [Academic Dataset](#). This treasure trove of local business data is waiting to be mined and we can't wait to see you push the frontiers of data science research with our data.



The Challenge Dataset:

- **1.6M** reviews and **500K** tips by **366K** users for **61K** businesses
- **481K** business attributes, e.g., hours, parking availability, ambience.
- Social network of **366K** users for a total of **2.9M** social edges.
- Aggregated check-ins over time for each of the **61K** businesses

Cities:

- U.K.: Edinburgh
- Germany: Karlsruhe
- Canada: Montreal and Waterloo
- U.S.: Pittsburgh, Charlotte, Urbana-Champaign, Phoenix, Las Vegas, Madison

The Challenge

Not only would we like to give you our data, we'd also like to announce the fifth round of the **Yelp Dataset Challenge**. We challenge you to use this data in an innovative way and break ground in research.

Cultural Trends: By adding a diverse set of cities, we want participants to compare and contrast what makes a particular city different. For example, are people in international cities less concerned about driving in to a business, indicated by their lack of mention about parking? What cuisines are Yelpers raving about in these different countries? Do Americans tend to eat out late compared to the Germans and English? In which countries are Yelpers sticklers for service quality? In international cities such as Montreal, are French speakers reviewing places differently than English speakers?

Location Mining & Urban Planning: How much of a business' success is really just location, location, location? Do you see reviewers' behavior change when they travel?

Seasonal Trends: What about seasonal effects: Are HVAC contractors being reviewed just at onset of winter, and manicure salons at onset of summer? Are there more reviews for sports bars on major game days

The Awards

If you are a student and come up with an appealing project, you'll have the opportunity to win one of ten Yelp Dataset Challenge awards for \$5,000. Yes, that's \$5,000 for showing us how you use our data in insightful, unique, and compelling ways.

Additionally, if you publish a research paper about your winning research in a peer-reviewed academic journal, then you'll be awarded an additional \$1,000 as recognition of your publication. If you are published, Yelp will also contribute up to \$500 to travel expenses to present your research using our data at an academic or industry conference.

and if so, could you predict that?

Infer Categories: Do you see any non-intuitive correlations between business categories e.g., how many karaoke bars also offer Korean food, and vice versa? What businesses deserve their own subcategory (i.e., Szechuan or Hunan versus just "Chinese restaurants"), and can you learn this from the review text?

Natural Language Processing (NLP): How well can you guess a review's rating from its text alone? What are the most common positive and negative words used in our reviews? Are Yelpers a sarcastic bunch? And what kinds of correlations do you see between tips and reviews: could you extract tips from reviews?

Predict Attributes: Can you correlate our reviews with business attributes and predict whether a business is good for kids? Has Wi-Fi? Has Parking?

The deadline for the fifth round of the Yelp Dataset Challenge is **June 30, 2015**. Submit your project to Yelp by visiting yelp.com/challenge/submit. You can submit a research paper, video presentation, slide deck, website, blog, or any other medium that conveys your use of the Yelp Dataset Challenge data.

Round Four Challenge Winners

Coming Soon!

Round Three Challenge Winners

From the completed entries we received, a team of our data mining engineers selected two entries as grand prize winners (in alphabetical order by entry name):

- ["On the Efficiency of Social Recommender Networks."](#) Felix W. Princeton University.
- ["Personalizing Yelp Star Ratings: a Semantic Topic Modeling Approach."](#) Jack Linshi. Yale University.

Round Two Challenge Winner

From the completed entries we received, a team of our data mining engineers selected the following as a grand prize winner:

- ["Valence Constrains the Information Density of Messages."](#) David W. Vinson, Rick Dale. University of California, Merced.

Round One Challenge Winners

From the completed entries we received, a team of our data mining engineers selected four entries as grand prize winners (in alphabetical order by entry name):

- ["Clustered Layout Word Cloud for User Generated Review."](#) Ji Wang, Jian Zhao, Sheng Guo, Chris North. Virginia Tech and University of Toronto. Presented at [Graphics Interface 2014 Montreal](#)
- ["Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text."](#) Julian McAuley, Jure Leskovec. Stanford University. Published in [ACM RecSys '13 Proceedings](#)
- ["Improving Restaurants by Extracting Subtopics from Yelp Reviews."](#) James Huang, Stephanie Rogers, Eunkwang Joo. University of California, Berkeley. Presented at [iConference 2014 Berlin](#)
- ["Inferring Future Business Attention."](#) Bryan Hood, Victor Hwang, Jennifer King. Carnegie Mellon University.

Notes on the Dataset

Each file is composed of a single object type, one json-object per-line.

Take a look at some examples to get you started: <https://github.com/Yelp/dataset-examples>.

business

Find tacos, cheap dinner, Max's

Near Minneapolis, MN

Log In

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

```
{
  'city': (city),
  'state': (state),
  'latitude': latitude,
  'longitude': longitude,
  'stars': (star rating, rounded to half-stars),
  'review_count': review count,
  'categories': [(localized category names)]
  'open': True / False (corresponds to closed, not business hours),
  'hours': {
    (day_of_week): {
      'open': (HH:MM),
      'close': (HH:MM)
    },
    ...
  },
  'attributes': {
    (attribute_name): (attribute_value),
    ...
  },
}
```

review

```
{
  'type': 'review',
  'business_id': (encrypted business id),
  'user_id': (encrypted user id),
  'stars': (star rating, rounded to half-stars),
  'text': (review text),
  'date': (date, formatted like '2012-03-14'),
  'votes': {(vote type): (count)},
}
```

user

```
{
  'type': 'user',
  'user_id': (encrypted user id),
  'name': (first name),
  'review_count': (review count),
  'average_stars': (floating point average, like 4.31),
  'votes': {(vote type): (count)},
  'friends': [(friend user_ids)],
  'elite': [(years_elite)],
  'yelping_since': (date, formatted like '2012-03'),
  'compliments': {
    (compliment_type): (num_compliments_of_this_type),
    ...
  },
  'fans': (num_fans),
}
```

check-in

```
{
  'type': 'checkin',
  'business_id': (encrypted business id),
  'checkin_info': {
    '0-0': (number of checkins from 00:00 to 01:00 on all Sundays),
    '1-0': (number of checkins from 01:00 to 02:00 on all Sundays),
    ...
    '14-4': (number of checkins from 14:00 to 15:00 on all Thursdays),
    ...
    '23-6': (number of checkins from 23:00 to 00:00 on all Saturdays)
  }, # if there was no checkin for a hour-day block it will not be in the dict
}
```

tip

```
{
  'type': 'tip',
  'text': (tip text),
  'business_id': (encrypted business id),
  'user_id': (encrypted user id),
  'date': (date, formatted like '2012-03-14'),
  'likes': (count),
}
```

About

[About Yelp](#)
[Careers](#)
[Press](#)
[Investor Relations](#)
[Content Guidelines](#)
[Terms of Service](#)
[Privacy Policy](#)
[Ad Privacy Info](#)

Discover

[The Weekly Yelp](#)
[Yelp Blog](#)
[Support](#)
[Yelp Mobile](#)
[Developers](#)
[RSS](#)

Yelp for Business Owners

[Claim your Business Page](#)
[Advertise on Yelp](#)
[Yelp SeatMe](#)
[Business Success Stories](#)
[Business Support](#)
[Yelp Blog for Business Owners](#)

Languages

[English](#)

Countries

[United States](#)

[Site Map](#) | [Atlanta](#) | [Austin](#) | [Boston](#) | [Chicago](#) | [Dallas](#) | [Denver](#) | [Detroit](#) | [Honolulu](#) | [Houston](#) | [Los Angeles](#) | [Miami](#) | [Minneapolis](#) | [New York](#) | [Philadelphia](#) | [Portland](#) | [Sacramento](#) | [San Diego](#) | [San Francisco](#) | [San Jose](#) | [Seattle](#) | [Washington, DC](#) | [More Cities](#)

Copyright © 2004–2015 Yelp Inc. Yelp,  and related marks are registered trademarks of Yelp.