Impact of Elite User Reviews on Businesses (Yelp Data)

Vishwajeet Khadilkar, Lalit Pathak, Ricardo Slamao Da Silva Jr., Shikhar Srivastava

17169984, 18110088, 18147607, 18106960

MSc in Data Analytics

8th February 2018

1. Problem Background and Research Question

User reviews nowadays are very important for a business to run successfully to make profit. Some websites where users post restaurant and hotel reviews do not publicly provide their data online for analysis purpose like Trip Advisor but some websites do like Yelp. It has already been stated by Trip Advisor that user reviews posted online have a great impact on whether an individual wants to visit a restaurant or not (Perkins, 2014). There are social media influencers who are like famous people and their reviews might prove to be an important driving force for a business (Total Food Service, 2018). Unlike old days, restaurants now depend a lot on reviews given by users online for their restaurants on several online platforms like Yelp. Yelp, majorly operates in the US and is known for its restaurant reviews posted by millions of its users. Restaurant businesses also trust Yelp and monitor user reviews to serve better and improve. Yelp provides rich datasets on their website for academic purpose; all 5 publicly available datasets are listed below:

1. Business
2. Review
3. User
4. Check in
5. Tip

Among these 5 datasets provided by YELP, we'll be using Business, Review and User. For our data mining project, we'll be analysing these 3 datasets to measure the impact of reviews given by elite users on a particular business, where elite users can be regarded as influencers. To achieve this, we we'll be doing sentiment analysis on reviews given by elite users for respective restaurants and will then compare these sentiments with those calculated for users other than elite users. We will further check if there has been a change in star rating of restaurants based on elite user reviews. This analysis leads to our research question:

"Do the reviews given by elite users have any impact on reviews given by other users and restaurant star rating?"

And subsequently we will analyse which business types and geographies show highest impact.

2. Specific Items to be Addressed

- Project revolves around impact of reviews posted by elite users on businesses.
- We'll first filter all elite user IDs from user dataset.
- Sentiment analysis will be done on all reviews posted by elite users from review dataset where user ID is present.
- Sentiment analysis will further be done on reviews posted by users other than elite users from review dataset.

- We are further planning to see on which business type do the reviews posted by elite users have a greater impact, also whether geographies play a role in it or not.
- Star rating of business will also be compared by elite user reviews.

3. Data Description of Datasets used in the Project (Yelp)

- Business

{

// string, 22 character unique string business id

"business_id": "tnhfDv5Il8EaGSXZGiuQGg",

// string, the business's name

"name": "Garaje",

// string, the full address of the business

"address": "475 3rd St",

// string, the city

"city": "San Francisco",

// string, 2 character state code, if applicable

"state": "CA",

// string, the postal code

"postal code": "94107",

// float, latitude

"latitude": 37.7817529521,

// float, longitude

"longitude": -122.39612197,

// float, star rating, rounded to half-stars

"stars": 4.5,

// integer, number of reviews

"review_count": 1198,

```
  // integer, 0 or 1 for closed or open, respectively
  "is_open": 1,

  // object, business attributes to values. note: some attribute values might be objects
  "attributes": {
    "RestaurantsTakeOut": true,
    "BusinessParking": {
      "garage": false,
      "street": true,
      "validated": false,
      "lot": false,
      "valet": false
    },
  },

  // an array of strings of business categories
  "categories": [
    "Mexican",
    "Burgers",
    "Gastropubs"
  ],

  // an object of key day to value hours, hours are using a 24hr clock
  "hours": {
    "Monday": "10:00-21:00",
    "Tuesday": "10:00-21:00",
    "Friday": "10:00-21:00",
    "Wednesday": "10:00-21:00",
    "Thursday": "10:00-21:00",
    "Sunday": "11:00-18:00",
    "Saturday": "10:00-21:00"
  }
}
```

- User

```
{
    // string, 22 character unique user id, maps to the user in user.json
    "user_id": "Ha3iJu77CxlrFm-vQRs_8g",

    // string, the user's first name
    "name": "Sebastien",

    // integer, the number of reviews they've written
    "review_count": 56,

    // string, when the user joined Yelp, formatted like YYYY-MM-DD
    "yelping_since": "2011-01-01",

    // array of strings, an array of the user's friend as user_ids
    "friends": [
        "wqoXYLWmpkEH0YvTmHBsJQ",
        "KUXLLiJGrjtSsapmxmpvTA",
        "6e9rJKQC3n0RSKyHLViL-Q"
    ],

    // integer, number of useful votes sent by the user
    "useful": 21,

    // integer, number of funny votes sent by the user
    "funny": 88,

    // integer, number of cool votes sent by the user
    "cool": 15,

    // integer, number of fans the user has
    "fans": 1032,

    // array of integers, the years the user was elite
    "elite": [
        2012,
        2013
    ],

    // float, average rating of all reviews
    "average_stars": 4.31,

    // integer, number of hot compliments received by the user
    "compliment_hot": 339,

    // integer, number of more compliments received by the user
    "compliment_more": 668,

    // integer, number of profile compliments received by the user
    "compliment_profile": 42,
```

```
    // integer, number of cute compliments received by the user
    "compliment_cute": 62,

    // integer, number of list compliments received by the user
    "compliment_list": 37,

    // integer, number of note compliments received by the user
    "compliment_note": 356,

    // integer, number of plain compliments received by the user
    "compliment_plain": 68,

    // integer, number of cool compliments received by the user
    "compliment_cool": 91,

    // integer, number of funny compliments received by the user
    "compliment_funny": 99,

    // integer, number of writer compliments received by the user
    "compliment_writer": 95,

    // integer, number of photo compliments received by the user
    "compliment_photos": 50
}
```

- Review

```
{
    // string, 22 character unique review id
    "review_id": "zdSx_SD6obEhz9VrW9uAWA",

    // string, 22 character unique user id, maps to the user in user.json
    "user_id": "Ha3iJu77CxlrFm-vQRs_8g",

    // string, 22 character business id, maps to business in business.json
    "business_id": "tnhfDv5Il8EaGSXZGiuQGg",

    // integer, star rating
    "stars": 4,

    // string, date formatted YYYY-MM-DD
    "date": "2016-03-09",

    // string, the review itself
    "text": "Great place to hang out after work: the prices are decent, and the ambience is fun. It's a bit loud, but very lively. The staff is friendly, and the food is good. They have a good selection of drinks.",

    // integer, number of useful votes received
```

```
    "useful": 0,

    // integer, number of funny votes received
    "funny": 0,

    // integer, number of cool votes received
    "cool": 0
}
```

Bibliography

Perkins, C. (2014, February 12). *TripAdvisor: Study gives insight into impact of reviews*. Retrieved from bighospitality.co.uk: https://www.bighospitality.co.uk/Article/2014/02/12/TripAdvisor-study-gives-insight-into-impact-of-reviews

Total Food Service. (2018, August 24). *How Social Influencers Can Impact the Restaurant Industry*. Retrieved from totalfood.com: https://totalfood.com/social-influencers-instagram-impact-restaurant-industry/

Yelp. (n.d.). *Yelp Dataset JSON*. Retrieved from https://www.yelp.com/dataset/documentation/main