# *Eat, Rate, Love

A proposal for a sub-rating within Yelp to help choose good ethnic restaurants

Robert Chen,  1/25/2016

# The problem

- Yelp is a great tool for finding restaurants in the area of a specific cuisine.

- However, what happens when the results yield a lot of restaurants with similar ratings?

- For example, a search of "Indian Restaurants" in the Schaumburg metropolitan area yields 14 different restaurants all rated between 3 and 4 stars and all within a 3 mile radius (shown on next slide)

# The problem (2)
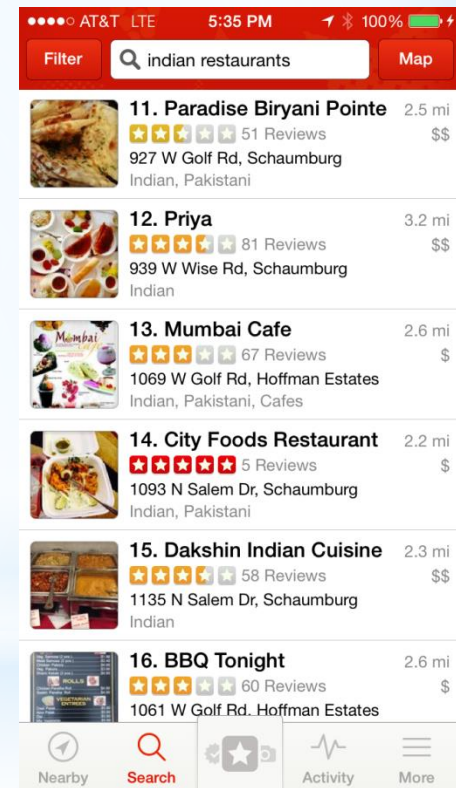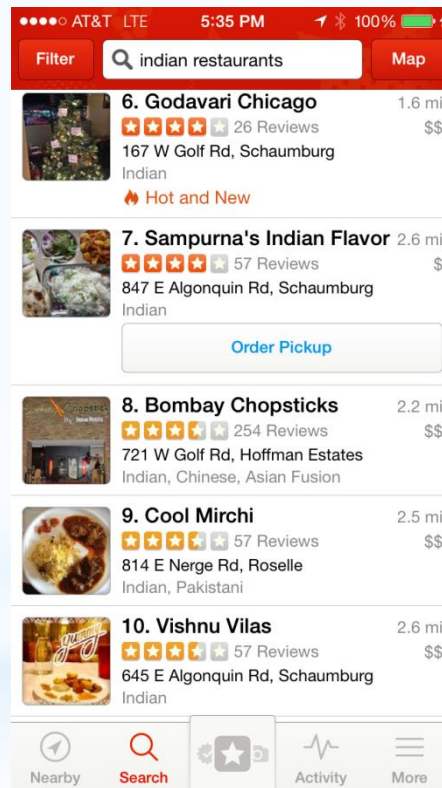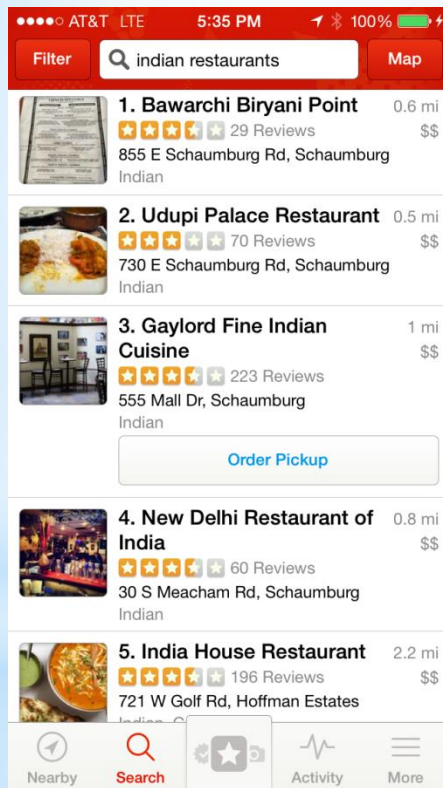
# Proposed solution(s)

- Analyze whether a "sub-rating" can be added to the primary star rating to help a user "cut through the clutter" and select between many different choices.

- Two methods are examined:

  (a) A rating which gives more weight to reviewers who have reviewed the same cuisine.

  (b) An "Immigrant Rating" based on user names that match the ethnicity of the cuisine

# Data / Method used

- Data from the Yelp Dataset Challenge was used.
  - 1.6 million reviews
  - 366,000 users
  - 61,000 business in 10 cities

- This data was first preprocessed using Python scripts to filter out only the fields needed and to convert from JSON to CSV. The resulting CSV files were then analyzed using R.

# Method 1: Description

- Proposed algorithm: If a user has reviewed 2 restaurants of the same cuisine, double the weight of the review. If 3, triple it etc.

- The first question is does the usage pattern justify trying this – are there enough users of the main ethnic cuisines who have reviewed more than 1 restaurant of that cuisine?

# Method 1:  Cuisine review pattern

| Cuisine | Total Reviewers | # > 1 Review | % > 1 Review | Avg Reviews per person | Max Reviews |
|---|---|---|---|---|---|
| (1)   Chinese | 33,359 | 7,733 | 23% | 1.56 | 212 |
| (2)   Mexican | 54,138 | 14,828 | 27% | 1.77 | 145 |
| (3)   Italian | 51,245 | 12,606 | 25% | 1.63 | 90 |
| (4)   Japanese | 44,849 | 11,688 | 26% | 1.67 | 66 |
| (5)   Greek | 10,820 | 1,799 | 17% | 1.29 | 19 |
| (6)   French | 19,127 | 3,413 | 18% | 1.32 | 40 |
| (7)   Thai | 20,699 | 4,032 | 20% | 1.40 | 74 |
| (8)   Spanish | 12,736 | 1,804 | 14% | 1.25 | 27 |
| (9)   Indian | 9,549 | 1,735 | 18% | 1.38 | 93 |
| (10) Mediterranean | 19,400 | 3,534 | 18% | 1.35 | 30 |

Source of Ranking:  "Top 10 Ethnic Cuisines Americans Crave Most",
Parade Magazine, 5/15/2015

# Method 1:  Cuisine review pattern (2)

- Each cuisine has at least 10% of reviewers who have reviewed more than 1 restaurant of that cuisine.

- The number is as high as 27% for Mexican cuisine and drops to as low as 14% for Spanish cuisine.

- In general, the higher the # of reviewers, the higher the percentage of users who have given multiple reviews.

- **At least 10% -> Enough to give Method 1 a try**

# Method 1: Results

- Range = -1.14 stars to + 0.77 stars, mean = -0.09 stars

- Sample results for Tempe, Arizona:

|  | Official | Old | New |
|---|---|---|---|
| Bombay Palace | 4 | 3.83 | 3.63 |
| Curry Corner | 4 | 3.75 | 3.46 |
| Delhi Palace | 4 | 3.75 | 3.64 |
| India Grill | 4 | 3.87 | 3.84 |
| **Little India** | **4** | **4.06** | **4.26** |
| Tasty Kabob | 4 | 4.00 | 3.75 |
| The Dhaba | 4 | 4.06 | 3.97 |
| Udupi Indian | 4 | 3.96 | 3.34 |

Rounds up to 4.5 ->
Clear choice identified

- It works well for one city, but how about for others …

# Method 1:  Results for all cities

| City | # Restaurants w/ 4 stars | Result | Recommended Restaurant(s) |
|------|--------------------------|--------|---------------------------|
| Champaign | 1 | No additional rating needed | ----- |
| Chandler | 3 | 3 reduced to 2 | Woodlands / Indus |
| Charlotte | 8 | 8 reduced to 1 | Aroma |
| Edinburgh | 12 (*) | 12 reduced to 1 | Noor |
| Karlsruhe | 1 | No additional rating needed | ----- |
| Las Vegas | 9 | 9 reduced to 2 | Mint / Saffron |
| Madison | 9 | 9 reduced to 1 | Maharaja |
| Mesa | 2 | 2 reduced to 1 | Guru Palace |
| Montreal | 10 (*) | 10 reduced to 1 | Restaurant Tibetan |
| Phoenix | 7 | 7 reduced to 4 | Star/ Khyber / Garden / Saffron |
| Pittsburgh | 7 | 7 reduced to 2 | Tamarind / India on Wheels |
| Scottsdale | 3 | 3 reduced to 2 | Indian Paradise / Jewel of Crown |
| Tempe | 6 | 7 reduced to 1 | Little India |
| Waterloo | 1 | No additional rating needed | ----- |
| (*) = 4.5 stars used instead of 4 because there were many 4.5 star restaurants | | | |

Results:  3 cities had just one choice
5 cities had > 1 choice, reduced some
6 cities had > 1 choice, reduced to 1 choice

# Method 2:  Description

- Proposed algorithm:  Generate a separate rating for users if they have a user name that is unique to the ethnicity of that cuisine.

- In this example, Indian cuisine is examined.

- The first question is what is the percentage of "Indian names" in the reviews of Indian restaurants?

# Method 2: List of Indian names

- The list of user names of reviewers of Indian cuisine was analyzed. If they had a name that was uniquely Indian it was added to a file which was later read in as a list in R.

- The sites www.indiaparenting.com, www.modernindiababynames.com, www.indiachildnames.come were used.

- A total of 608 names were found.

# Method 2:  Examples of Indian names used

* Aayush
* Abhi
* Abhijeet
* Abhijit
* Abhilash
* Abhinandan
* Abhinav
* Abhinay
* Abhisek
* Abhishek
* Abilash
* Achyuthan
* Aditi
* Aditya
* Adnan

* Ahana
* Aisha
* Aishaa
* Ajay
* Ajinkya
* Akash
* Akhil
* Akhilesh
* Akshada
* Akshay
* Akshaya
* Alina
* Alok
* Amal
* Amalee

* Amalia
* Aman
* Amandeep
* Amar
* Amaris
* Amarnath
* Ambika
* Ami
* Amil
* Amir
* Amirah
* Amisha
* Amit
* Amogha
* Amrita

* Anagha
* anand
* Anand Kumar
* Ananth
* Ananya
* Aneesh
* Ani
* Aniket
* Anil
* Anindya
* Anirudh
* Anish
* Anju
* Ankan
* Ankit

* * *

* Vineet
* Vineeth
* Vinit
* Vinod
* Vinoth
* Vinuth
* Vipin
* Vipin Das
* Vipul
* Vishal
* Vishnu
* Vishwa
* Vivek
* Yogesh
* Yuvaraj

# Method 2:  Initial analysis

- Out of 13,146 reviews, 1,274 of them, or **9.7%**, were done by someone with a clearly Indian name.

- This was enough to warrant giving this a try. What were the results?

# Method 2: Results

- Range = -1.92 stars to + 0.31 stars, mean = -0.44 stars

- Sample results for Tempe, Arizona:

|  | Official | Old | New |
|---|---|---|---|
| Curry Corner | 4 | 3.75 | 3.57 |
| Delhi Palace | 4 | 3.75 | 2.33 |
| India Grill | 4 | 3.87 | 3.00 |
| **Little India** | **4** | **4.06** | **4.13** |
| The Dhaba | 4 | 4.06 | 3.66 |
| Udupi Indian | 4 | 3.96 | 3.21 |

Rounds to 4.0 ->
Clear choice identified

- It works well for one city, but how about for others ...

# Method 2: Results for all cities

| City | # Restaurants w/ 4 stars | Result | Recommended Restaurant |
|------|---------------------------|--------|------------------------|
| Champaign | 1 | No need for additional rating | ----- |
| Chandler | 2 | 2 reduced to 1 | NASHA |
| Charlotte | 6 | 6 reduced to 1 | Aroma |
| Las Vegas | 6 | 6 reduced to 1 | Taj Palace |
| Madison | 2 | 2 reduced to 1 | Maharani |
| Mesa | 1 | No need for additional rating | ----- |
| Montreal | 1 | No need for additional rating | ----- |
| Phoenix | 2 | 2 reduced to 1 | Star of India |
| Pittsburgh | 3 | 3 reduced to 1 | Tamarind Flavor |
| Scottsdale | 2 | 2 reduced to 1 | Jewel of the Crown |
| Tempe | 6 | 6 reduced to 1 | Little India |

Results:  3 cities had no choices
3 cities had just 1 choice
8 cities had > 1 choice, reduced to 1 choice

# Comparison of the 2 methods

|  | Method 1 | Method 2 |
|---|---|---|
| Range | -1.14 to + 0.77 | -1.92 to + 0.31 |
| Mean | -0.09 | -0.44 |
| Cities left out | 0 | 3 |
| # reduced to 1 choice | 6 out of 11 (55%) | 8 out of 8 (100%) |

- Method 2 has more effect on ratings (more negative)
- Method 2 is better at reducing to 1 choice (100% vs 55%)
- Method 1 is more inclusive (0 left out vs 3)

# Comparison of the 2 methods (2)

- Since Method 2 seems better at "cutting through the clutter", recommendation = use **Method 2** as the sub-rating

# Proposed use of sub-rating

- If a restaurant gets a rating of 4 or more using Method 2 (or 3.5 if there are no 4 star restaurants in that city), add an "Authenticity Badge" on the Summary and Restaurant page:
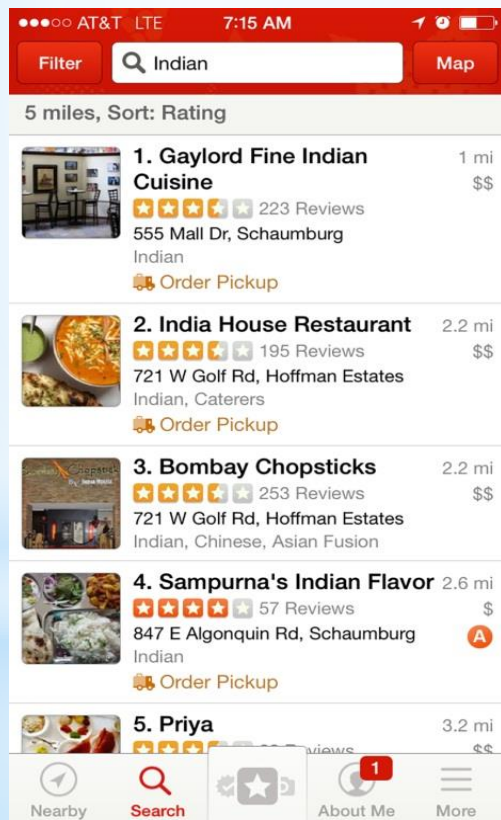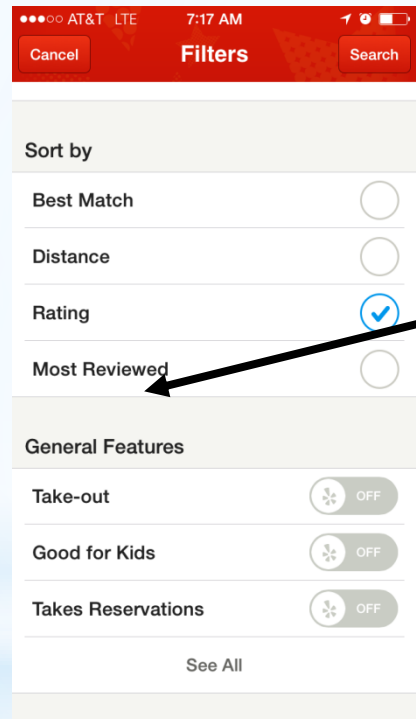


Authenticity Badge



Authenticity Badge

# Proposed use of sub-rating (2)

- Also add an option for filtering for "Authenticity" after "Most Reviewed":



Add "Authenticity" here

# Advantages of sub-rating

- Potential engine for increasing ad revenue

- Help daily deals program

- Improve accuracy of recommendations

- Increase user engagement

# Proposed next steps

## (1) Independently confirm results
- Confirm w/ human panels, solicit opinion of local restaurant critics

## (2) Improve "authenticity rating"
- Analyze text reviews for "authenticity" information
- Can percentage of reviewers who have Indian names for a restaurant, reviewers who also review ethnic grocery stores, restaurants marked as "touristy" be used to add to the "authenticity rating"?

## (3) Extend further
- Automate name generation by scraping Indian baby name web site, then scrubbing common American names from it
- Try Chinese, Vietnamese, Persian names
- Consider a roll-out in areas with large concentrations of ethnic places (Chinese restaurants in the Bay Area, Indian restaurants in Chicago, Pho restaurants in the San Diego Area …)

# Thank you to:

## Andi Bandyopadhyay

## Springboard

# Thank you and Bon appétit!!