Yelp Dataset Challenge
New Yorker Exercise

Toavina Andriamanerasoa

January 2017

THE YELP DATASET IS EXTENSIVE

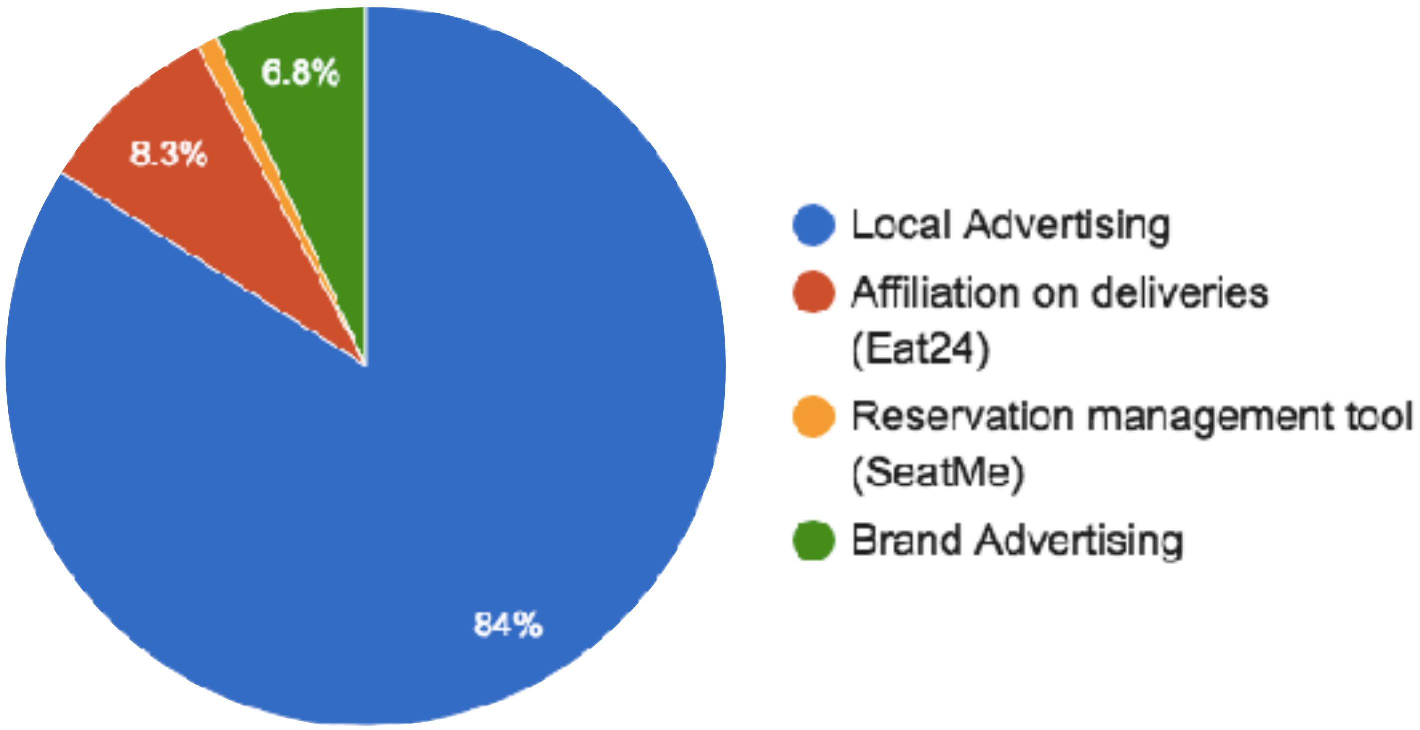
The dataset when combined provides a wealth of opportunities for interesting analyses

- The Yelp challenge consists of the following datasets:
    - Business data - Information about c.86k businesses in 10 cities across the UK, Germany, Canada and the US
    - Check-in data - Information about users “checking in” to locations
    - Reviews - Over 2.6m reviews for close to 100% of the businesses in the dataset
    - Tips - Almost 650k user tips for 57k businesses
    - User data - Data from c.687k users, with information about friends, number of fans and linkable to reviews
    - Photos - 200,000 photos for c.42k businesses
- This provides significant scope for interesting questions about the dataset, but after careful analysis, it is possible to isolate key topics

BY UNDERSTANDING YELP'S BUSINESS MODEL, THE SCOPE OF INTERESTING QUESTIONS NARROWS SIGNIFICANTLY

Yelp's revenues are derived from user views driven by reviews. Given the importance of significant network effects, **any measure that increases user traffic, engagement (through better user experience) and especially reviews,** or provides valuable monetisable information for businesses **will have positive business impact**

Yelp 2015 Revenue Sources(1)



Key Drivers

- # and quality / trustworthiness of reviews as impacts both businesses and users, and confirms value of Yelp to businesses
- Coverage of local businesses (one-stop shop increases lock-in) and their engagement
- **Traffic and repeat traffic (over 150m unique visitors per month as of Q3 2016)(2) through greater user and business engagement as network effects take place**

Sources: (1) http://hackingrevenue.com/revenue/how-does-yelp-make-money/ (2) https://www.yelp.co.uk/factsheet

I THUS FOCUSED ON QUESTIONS THAT COULD INCREASE THE NUMBER OF REVIEWS OVER THE LONG-TERM

- While many interesting questions that increase Yelp's business could also addressed, the following constraints were born in mind:

  - Lack of data - Some questions (e.g. Do new photos increase the number of reviews over time?) cannot be answered satisfactorily because timestamps are not provided

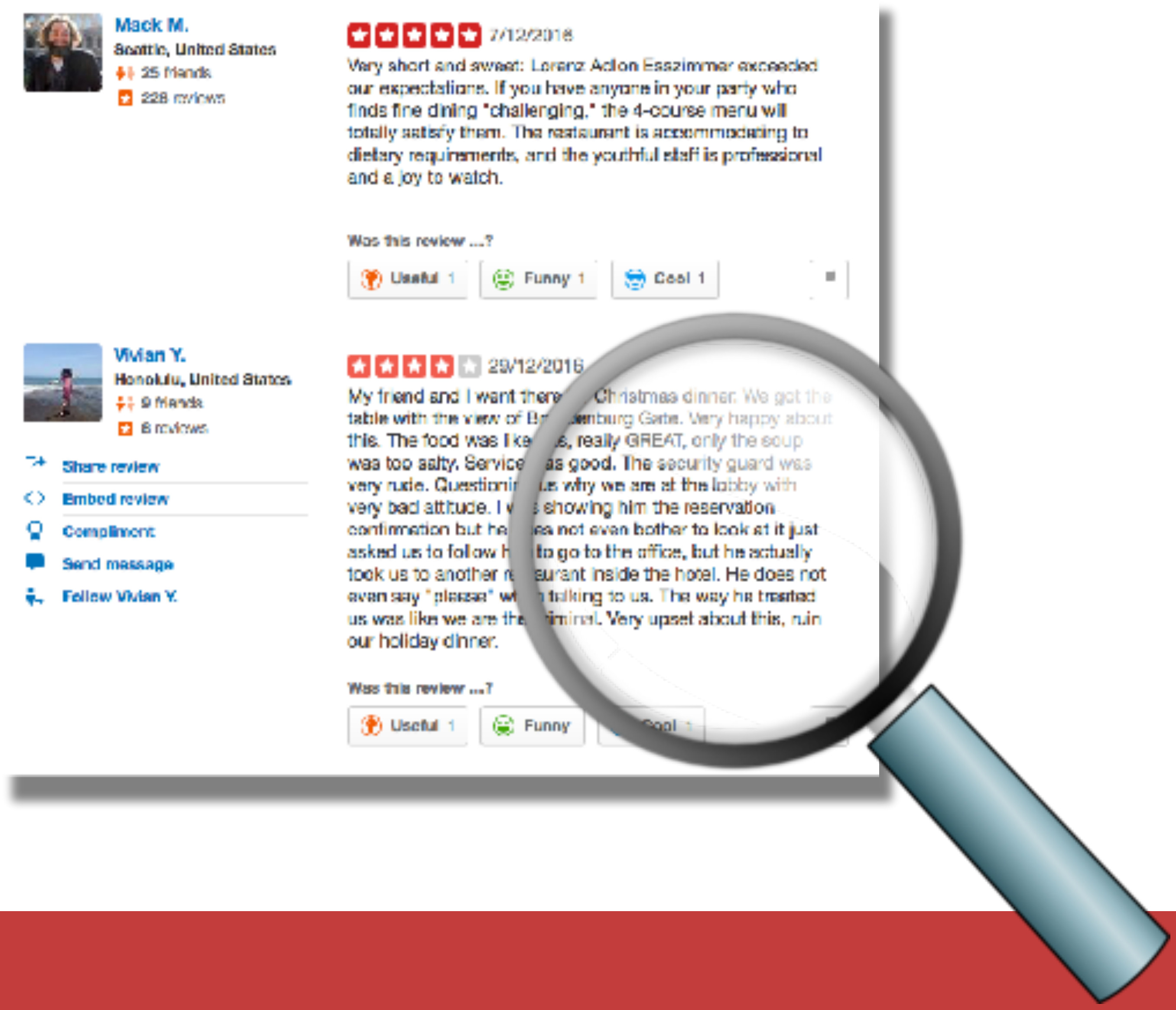
  - Resource constraints / scope - Given the size of the dataset and computing resources, I had to focus on certain subsets of the data and problems which could be addressed in a reasonable timeframe

  - Value-add from machine learning - Some interesting questions can be solved without complicated models and with simple explorations - despite their potential impact those were left out

- In light of this, I chose the following questions:

  - **Can we predict the rating a user will apply to a restaurant review from the review text alone?**

  - **Can we use collaborative filtering to suggest restaurants in Las Vegas to users?**



Mack M.
Seattle, United States
25 friends
228 reviews

7/12/2016

Very short and sweet: Lorenz Adlon Esszimmer exceeded our expectations. If you have anyone in your party who finds fine dining "challenging," the 4-course menu will totally satisfy them. The restaurant is accommodating to dietary requirements, and the youthful staff is professional and a joy to watch.

Was this review ...?

Vivian Y.
Honolulu, United States
9 friends
8 reviews

29/12/2016

My friend and I want there [illegible] Christmas dinner. We got the table with the view of Br[illegible]enburg Gate. Very happy about this. The food was l ke[illegible]s, really GREAT, only the soup was too salty. Service [illegible]as good. The security guard was very rude. Questioni[illegible] us why we are at the lobby with very bad attitude. I w[illegible]s showing him the reservation confirmation but he [illegible]es not even bother to look at it just asked us to follow h[illegible] to go to the office, but he actually took us to another re[illegible]aurant inside the hotel. He does not even say "please" w[illegible]n talking to us. The way he treated us was like we are th[illegible]riminal. Very upset about this, ruin our holiday dinner.

Was this review ...?

1. Can we predict the rating a user will apply to a restaurant review through the review text alone?

1. CAN WE PREDICT THE RATING A USER WILL APPLY TO A REVIEW THROUGH THE REVIEW TEXT ALONE? - MOTIVATION & PROCESS

Motivation

- By predicting user ratings from text alone, Yelp could benefit from:
    - Decreased user friction = More reviews - if users type text first and the rating is suggested as they type, this saves time for users
    - Additional reviews (perhaps tagged as other sources) - useful for rating a new business to drive traffic
    - Fake review or user error detection - If there is a discrepancy between the rating and the text, this could be flagged as an potential anomaly for further analysis

Modelling Process

1. Load and merge review and business data
2. **Filter for restaurants in English speaking states** for dimensionality and consistency of vocabulary
3. **Split ratings into 3 classes (above average (4-5 stars), average (3 stars), below average (1-2 stars) - sufficient to provide value**
4. After splitting into training and test sets, use bag of words (5,000 features) to vectorise most common words (NB: currently implementing Doc2Vec - will append once results are in)
5. Evaluate model performance and thoughts

1. CAN WE PREDICT THE RATING A USER WILL APPLY TO A REVIEW THROUGH THE REVIEW TEXT ALONE? - BAG OF WORDS RESULTS

The results are promising save for 3 star reviews, but should be sufficient to be tested in production

Description

- Classifiers tried with BoW features
  - Logistic regression (LR)
  - Linear SVM
  - ExtraTrees
- Best Accuracy: LR with c.84% on test set (30% of data)

Confusion Matrix - LR Classifier

	Below Avg	Avg	Above Avg
Below Avg	74,896	7,232	11,039
Avg	13,925	19,949	30,439
Above Avg	4,317	8349	299,049

Key Classification Metrics with LR Classifier

	Precision	Recall	F1-Score	Support
Below Avg	0.80	0.80	0.80	93,167
Avg	0.56	0.31	0.40	64,313
Above Avg	0.88	0.96	0.92	31,1715
Avg/Total	0.82	0.84	0.82	469,195

1. CAN WE PREDICT THE RATING A USER WILL APPLY TO A REVIEW THROUGH THE REVIEW TEXT ALONE? - CONCLUSION & THOUGHTS

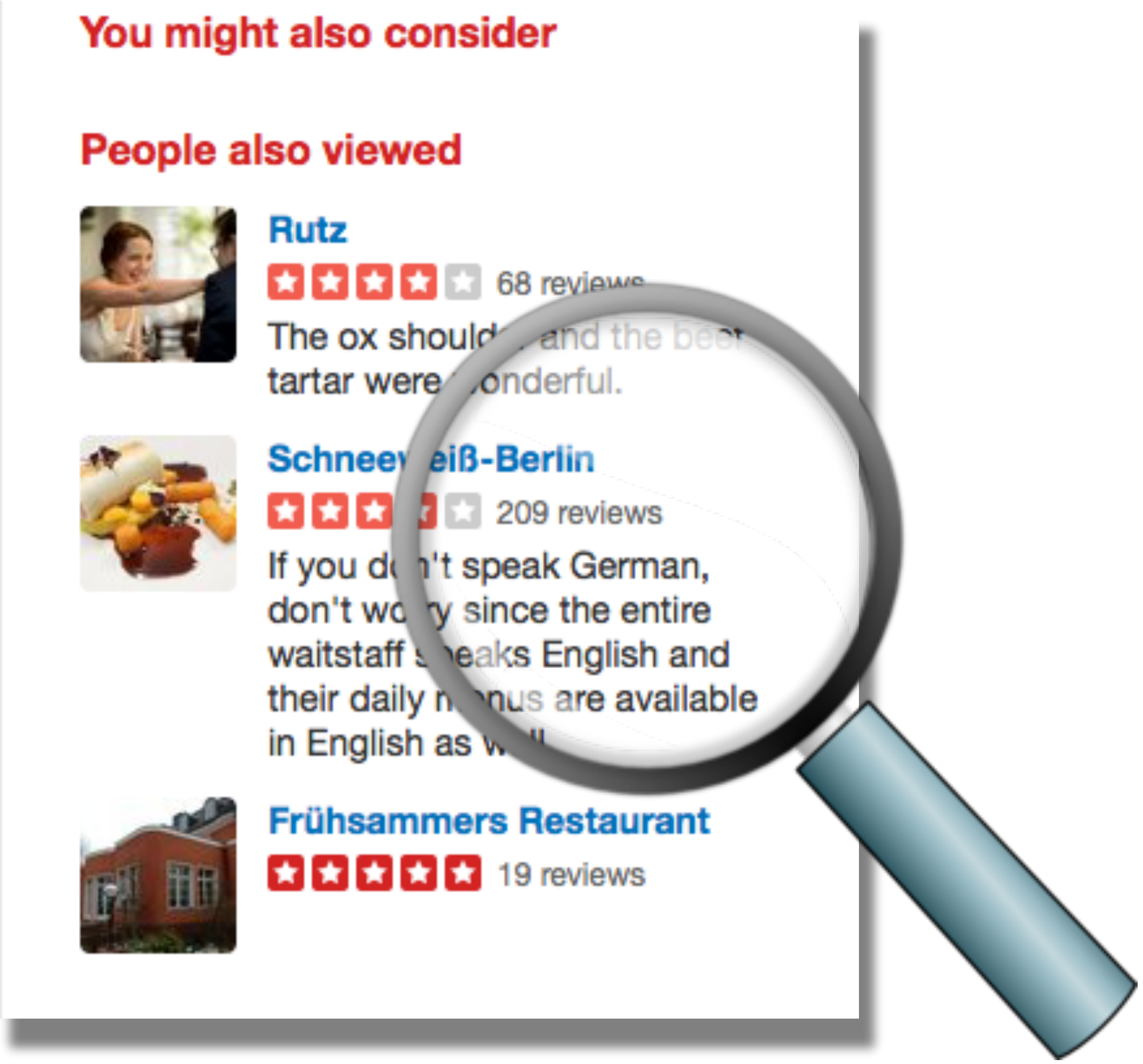
The results seem good enough to do some A/B testing with users - it may be worth looking at other algorithms to see if results can be improved

Conclusion

- Good results for a first pass, would benefit from being rolled out further
- Unsurprising that "average class" results are sub-par, but may not impact performance too much - to be tested in production

Improvements & Next Steps

- Resample minority classes to see whether this improves results
- Filter for other languages
- Increase number of examples and maximum number of words
- Try TF-IDF & Doc2Vec to see if results improve
- Test other classifiers and perform GridSearch to optimise hyperparameters and use cross-validation to estimate metrics
- Do some A/B testing to see impact on user friction
- Develop concept of "Ghost Reviews"



You might also consider

People also viewed

Rutz

★★★★☆ 68 reviews

The ox should and the beef tartar were wonderful.

Schneeweiß-Berlin

★★★★☆ 209 reviews

If you don't speak German, don't worry since the entire waitstaff speaks English and their daily menus are available in English as well.

Frühsammers Restaurant

★★★★★ 19 reviews

2. Can we use collaborative filtering to suggest restaurants in Las Vegas to users?

2. CAN WE USE COLLABORATIVE FILTERING TO SUGGEST RESTAURANTS IN LAS VEGAS TO USERS? - MOTIVATION & PROCESS

Motivation

- By using collaborative filtering (unsure what the mix is between content and collaborative filtering currently), Yelp could benefit from:
  - Potentially better recommendations through personalisation - drives user traffic and reviews
  - With some filtering, more variance of content per user - users can see restaurants with less reviews which would help drive traffic to those businesses and engage them with Yelp

Modelling Process

1. Load and merge review and user data
2. **Filter for restaurants in Las Vegas for 1,000 users with at least 5 reviews** for dimensionality to make model relevant (geography being key)
3. After splitting into training and test sets, use Alternating Least Squares to compute user-review matrix
4. Evaluate model performance and thoughts

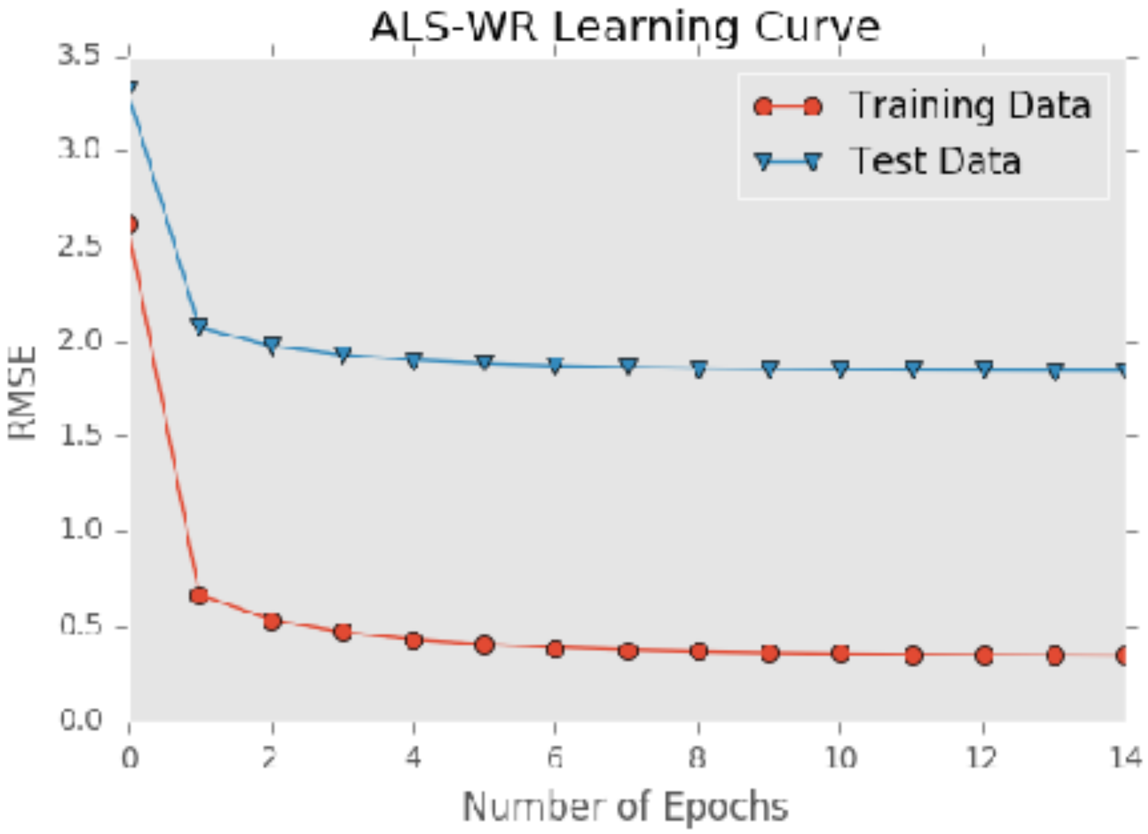
2. CAN WE USE COLLABORATIVE FILTERING TO SUGGEST RESTAURANTS IN LAS VEGAS TO USERS? - ALS RESULTS

The results are passable, with RMSE of c.1.8 in the test set

Description

- Test set: 30% of the data
- RMSE Training set: c.0.34
- RMSE Test set: c.1.8
- Passable results in test set, but may work well in real-life

ALS - RMSE over Epochs - Training & Test Data



2. CAN WE USE COLLABORATIVE FILTERING TO SUGGEST RESTAURANTS IN LAS VEGAS TO USERS? - CONCLUSION & THOUGHTS

Conclusion

- Results with respect to RMSE are passable and could stand from being optimised and tested
- Nevertheless, the recommended results may end up being more interesting than simple content-based filters - would benefit from testing

Improvements & Next Steps

- Test on larger sample
- Use optimised algorithms that can be computed much faster (e.g Spark MLlib implementation
- Backtest result further once implemented in real life
- Perform GridSearch to optimise hyperparameters to minimise the loss function and attempt to improve predictions
- Do some A/B testing to see impact on user engagement and reviews
- Test various combinations of content-based filtering and collaborative filtering to see what works best

yelp

DATASET CHALLENGE

Conclusion and Next Steps

CONCLUSIONS

1. There is much scope for a data-intensive business such as Yelp to benefit from machine learning

2. With more thorough analysis of each question and use of more sophisticated algorithm, there remains scope for improvement of the preliminary models shown

3. As the business develops and new features appear (check-in, ordering…), new opportunities will continue to arise for further analysis and value creation, creating a virtuous circle

4. As Yelp's users are increasingly using mobile devices, speed of delivery becomes even more paramount, while the amount and relevance of information provided is constrained by screen size. Nevertheless, features such as checking-in, calling and ordering food from the mobile present new opportunities

5. Businesses that can benefit from Yelp-like network effects and information will have an advantage going forward

APPENDIX - OTHER QUESTIONS TO CONSIDER WITHIN DATASET SCOPE

1. Is it possible to detect anomalies (up and down) in restaurant average ratings over time?
    1. Why: Could be used to suggest trending restaurants if good to test as a recommendation strategy, or to alert businesses if negative change
    2. How: Time series analysis and anomaly detection algorithms
2. Do reviews from superusers impact the number of subsequent reviews?
    1. Why: Could be used as paid product for businesses to increase visibility to superusers if there is an impact
    2. How: Time series analysis - regression (change in parameters)
3. Is there a geographic pattern and by type of business for check-ins at various points in time?
    1. Why: Could be used as information to sell to potential physical advertisers for good product placement or to advise new businesses as to where to locate
    2. How: Geodata analysis
4. Can you recommend users to follow / friends?
    1. Why: Could increase user engagement as they meet new people
    2. How: Clustering (kNN / SVM)