

Enhancing Yelp User Experience

Through Text Mining of Yelp User Reviews.

Ian Bassett

Ian.Bassett@student.ncirl.ie

Sapna Gupta

Sapna.Gupta@student.ncirl.ie

Shashi Bhushan

Shashi.Bhushan@student.ncirl.ie

James Matthews

James.Matthews1@student.ncirl.ie

Abstract

In this paper, the authors have taken part in the Yelp Dataset Challenge and discuss the results of text mining 539,535 Yelp user reviews of businesses located in Las Vegas Nevada and Phoenix Arizona.

The authors outline their adopted approach and methodology taken in conducting the analysis activities in identifying the term frequency of words contained within the Yelp user reviews. The results of which show 411 terms that appear 5,000 times in 539,535 reviews. Using a Likert scale the authors evaluate each found term to determine if it is Actionable, Trivial or Inexplicable. The overall aim of the evaluation, is to find terms that are practical and effective in enhancing the Yelp user experience.

Moreover, the authors propose a new feature for Yelp be implemented allowing users to dynamically refine Yelp reviews that are displayed on a business's page through the introduction of new sub-search filters and categories which are based upon the analysis results and evaluation of the 411 terms. The authors have selected ten terms which have been deemed to be effective insights that have a practicality in being assigned as a sub-search filter and category to enhance user experience.

Keywords—yelp dataset challenge; text mining; term-document matrix; enhancing user experience

I. INTRODUCTION

YELP website traffic volumes have seen continual growth year on year, increasing by 33% in 2014 resulting in 135 Million unique monthly visitors to the Yelp throughout Q4 of 2014 [1]. This growth in website traffic has attributed to a significant increase in the number of posted Yelp User Reviews of businesses, which has increased by 82% in comparison to 2013. 71 million cumulative reviews were posted in 2014, written

by users whom wish to share their local business experiences and in turn allowing consumers to voice and convey “word of mouth” online [1].

Providing a seamless and superior user experience where by website users can easily navigate and access information contained within a website is vitally important for any business or organisation. For Yelp, enabling users to easily navigate through and read user reviews is a core and fundamental element of Yelp's business model. The reviews are critical and necessary component that allow Yelp increase its market share and sustain business growth.

The Yelp Dataset Challenge allows academics access to a partial subset of Yelp data to utilise in an innovative way and break ground in conducting research [2]. The authors of this paper have opted to take part in the Yelp Dataset Challenge and propose a new feature for Yelp be implemented which will enable users to dynamically refine reviews that are displayed on a business's page through the introduction of new sub-search filters and categories. The primary objective of this new feature is to optimise and enhance the Yelp user experience.

The formulating of the sub-search filters and categories will be achieved through the application of text analytics against Yelp user reviews resulting in the identification of the most frequently used words associated with user reviews and these words will form the basis of the sub-search filters and categories.

In related work the authors discuss and critically evaluate similar approaches adopted in the Yelp dataset challenge around the domain of text analytics.

Within the methodology section the authors define the Yelp dataset segment and methodology applied to conduct the text analysis activities in identifying the new sub-search filters and categories.

The implementation section specifies the tools, environment and framework utilised to carry out the text analysis activities.

In the results section the authors evaluate and discuss the implications of their findings and how the found insights can be turned these into actionable tasks and recommendations that Yelp can apply to enhance user experience.

Finally in the conclusion section the authors summarise the papers findings and discuss the limitations of the results in relation to using a partial subset of the complete Yelp dataset. Moreover, the authors outline proposed areas of future work that can be conducted in the domain text analytics for the Yelp Dataset Challenge.

II. RELATED WORK

With the success and mainstream interest in the Yelp Dataset Challenge, participants have proposed a large volume of papers discussing related work. The authors agreed to focus on text analytics when brainstorming multiple approaches.

Related work to our data mining approach showed some interesting contributions. McAuley and Leskovec fused latent rating dimensions and latent review topics to create text labels in order to justify text with the assigned yelp ratings [3].

Alternate methods were Chada and Naik try to determine the legitimacy of the reviews by determining on polarity based on positive and negative words [4]. Where Xu et al used numeric ratings to determine legitimacy [5].

With regard to participants using text analytics in the yelp dataset challenge, there is a recurring trend of papers showing the majority focusing on trying to determine the psychology of yelp users and discover does their reviews match their score.

The main problem with this mainstream focus on legitimacy is it is difficult to determine customer's expectations. Where one consumer may be a kind reviewer another one may be strict. There is also unidentifiable variables the reviewer may choose not to disclose that alters their rating because they deem unfair to include in the text review.

The papers contributed as it brought to the authors' attention the challenges using text analytics. Cavnar and Trenkle discuss the challenges in text analytics with double meanings, emoticons, grammar, spelling multilingual text among others [6].

These challenges made the authors understand that based on different cultures, Yelp users have their own style in

writing reviews, which led the authors to discuss on how to gather accurate results.

Other papers that thought of alternative methods in using text analytics were beneficial in brainstorming in order to discover areas not covered in the Yelp Dataset Challenge. Korpoglu and Hall focused on the busiest levels in services to assist consumers based on time constraints [4].

Fan and Khademi alternatively focused on predicting the overall business rating alone using by creating a bag full of words from the top frequent words in the review [5].

Korpoglu and Hall's research brought the authors to focus on improving customer needs and discovering what they truly find important. Where Fan and Khademi's research established a foundation for the authors to start and establish the methodology.

The methodology was built by examining text analytic papers that focused on improving customer reviews. Hu and Liu focused on trying to manage customer opinions from the summarization of customer reviews. The step process involved mining product features that have been commented on by consumers and identifying positive or negative opinions when summarizing the data [7]. Cao et al examined the helpfulness of reviews by extracting semantic characteristics from review texts and discovering these semantic characteristics are more influential than other characteristics [8]. Hu and Liu's contribution on mining features with Cao et al pursuit in trying to discover which characteristics are important allowed the authors to develop the methodology.

The authors decided a geographical parameter would be focused on from Phoenix to Las Vegas in aims to discover the features that were commonly being mentioned in the text. These results would show what consumers find important when going out to a restaurant, allowing Yelp to address these needs that they are unaware of. Cao et al approach created discussion among the authors on how to frame the features that where most important compared to the rest. This led to the discussion of implementing a scale.

III. METHODOLOGY

A. Identifying Yelp Dataset Segment

The 2015 Yelp Academic Dataset is comprised of 1.6 million reviews and 500,000 tips submitted by 366,000 Yelp users for 61,000 businesses that span across ten cities in the North America and Europe.

2015 Yelp Dataset Challenge Cities	
North America	Europe
Charlotte	Edinburgh
Las Vegas	Karlsruhe
Madison	
Montreal	
Phoenix	
Pittsburgh	
Urbana-Champaign	
Waterloo	
Mexican, Restaurants	
Pizza, Restaurants	

Table 1 - Yelp Dataset Challenge Cities

481,000 businesses attributes ranging from business category, opening hours to parking availability are provided within the Academic Dataset.

Summary analysis completed against the Yelp Academic Dataset identified Las Vegas Nevada and Phoenix Arizona as being the most suitable candidate cities for conducting the text mining analysis.

The top ten business categories reviewed in Las Vegas and Phoenix illustrated in Table 2 show that they are of a comparable nature with Restaurant and Food related businesses being the most reviewed business category.

Rank	Comparison of Top 10 Reviewed Business Categories	
	<i>Las Vegas</i>	<i>Phoenix</i>
1	Arts & Entertainment, Performing Arts	Auto Repair, Automotive
2	Auto Repair, Automotive	Beauty & Spas, Nail Salons
3	Beauty & Spas, Nail Salons	Burgers, Fast Food, Restaurants
4	Burgers, Fast Food, Restaurants	Chinese, Restaurants
5	Chinese, Restaurants	Food, Coffee & Tea
6	Food, Coffee & Tea	Hair Salons, Beauty & Spas
7	Hair Salons, Beauty & Spas	Home Services, Real Estate, Apartments
8	Hotels & Travel, Event Planning &	Hotels & Travel, Event Planning &

	Services, Hotels	Services, Hotels
9	Mexican, Restaurants	Mexican, Restaurants
10	Pizza, Restaurants	Pizza, Restaurants

Table 2 – Comparative View of Top 10 Reviewed Business Categories in Las Vegas & Phoenix

The summary analysis results showed that combined both cities represent 36% of all businesses available within the Academic Dataset. In terms the volume of user reviews for Las Vegas and Phoenix this amounted to 539,535 or 34% of all business reviews contained within the Academic Dataset.

The Yelp Dataset segment selected by the authors for the purposes of text mining analysis is 539,535 user reviews from businesses located in Las Vegas Nevada and Phoenix Arizona.

B. Term-document matrices

To enact the text mining analysis of user reviews the authors have elected to build term-document matrices of the user reviews. A term-document matrix denotes the correlation between terms and documents, a row represents a term and column a document. A summation of the total number of occurrences of a term is displayed as an entry [9].

	Docs				
Terms	1	2	3	4	5
Casino	11	20	71	11	1
Cheap	45	54	14	57	74
Checkin	32	21	44	74	14
Clean	74	12	57	41	71
Closed	44	77	41	11	77
Club	54	54	44	10	20

Figure 1 - Sample Term-Document Matrix

The 539,535 selected user reviews are assigned into five corpuses, with each corpus containing 107,907 user reviews respectively.

C. Calculating Frequent Terms

Upon completing the build of the user review term-document matrix the authors next calculate the term frequency. Term Frequency equates how frequently a term occurs within in a document in the matrix. $TF(t)$ quotient is the number of occurrences term t appears in a given document divided by the cumulative number of terms in a document as illustrated in Figure 2 [10].

$$TF(t) = \frac{f(t)}{\sum x(t)}$$

Figure 2 - Term Frequency

Computation of term frequency will be enacted by utilising the R Text Mining (TM) package [11]. Six incremental iterations with minimum defined thresholds of term occurrence will be run to identify the most frequently used words contained within the term-document matrix.

<i>2015 Yelp Dataset Challenge Cities</i>	
<i>Iteration</i>	<i>Minimum Term Occurrence Threshold</i>
1	500
2	1000
3	2000
4	3000
5	4000
6	5000

Table 3 – Iteration minimum defined thresholds

D. Evaluation Criteria

The objective of the analysis is to provide Yelp with actionable tasks and recommendations that can be applied to enhance the Yelp user experience. Returned frequent terms will be evaluated by the authors using a Likert scale to address the following evaluation criteria [12].

- **Actionable** - term generated is deemed a practical and effective insight to enhance user experience.
- **Trivial** - term generated is deemed self-evident and an ineffective insight to enhance user experience.
- **Inexplicable** - term generated is deemed ambiguous and enigmatic insight to enhance user experience.

IV. IMPLEMENTATION

A. Database Environment – MongoDB

Complete Yelp Academic dataset was imported into an instance of MongoDB.

B. MapReduce Category Extraction

A MapReduce program written in Python was used to extract and segment the Yelp Academic dataset by extracting only user reviews for businesses located in Las Vegas and Phoenix.

C. Analysis Tools

The authors predominantly used R and the Text Mining R package TM to conduct their analysis activities.

V. RESULTS

A. Out Putted Term Frequency Results.

Utilising the R text mining package TM the term frequency of words contained within the Yelp user review segment was calculated. The authors ran this phase of the analysis six times in total with each iteration the minimum term occurrence threshold is increased as shown in table 4.

In iteration 1 the minimum term occurrence threshold was set as 500, the outputted results show that there are 2,025 words that occur 500 times within Yelp user review segment. For iteration 2, the minimum term occurrence threshold was increased to 1,000 with the outputted results indicating that there is 1,305 words that occur over 1,000 times.

<i>2015 Yelp Dataset Challenge Cities</i>			
<i>Iteration</i>	<i>Minimum Term Occurrence Threshold</i>	<i>Word Occurrence</i>	<i>Iteration % Change</i>
1	500	2,025	
2	1000	1,305	- 35.6%
3	2000	821	-37.1%
4	3000	591	-28.0%
5	4000	481	-18.6%
6	5000	411	-14.6%

Table 4 – Iteration Results with Word Occurrences

As illustrated in Figure 3 in each subsequent iteration the minimum term occurrence threshold is increased in increments of 1,000, we can see as the threshold increases the volume of word occurrences found decreases.

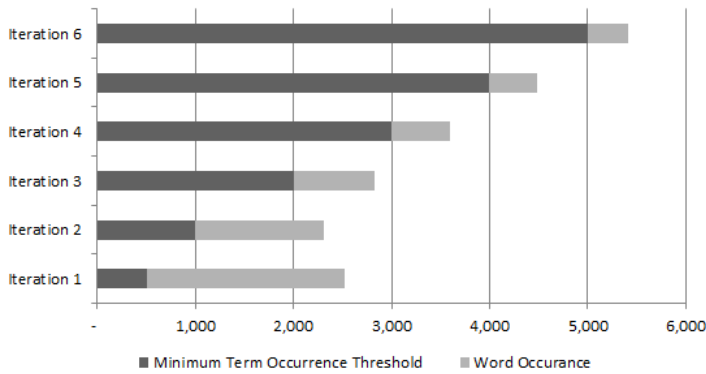


Figure 3 - Threshold increases, volume of word occurrences found decreases.

Examining the last iteration 6, the minimum term occurrence threshold was set at 5,000. The results from this iteration yield that there are 411 words that occur 5,000 times in user review segment. The 411 words from iteration 6 formulate the base data for the next phase in evaluating the results.

B. Applying Evaluation Criteria to Outputted Results.

In evaluating the results set of 411 words from iteration 6, the authors applied their Likert scale to each word in order to gauge its overall validity. The Likert scale entails three possible ratings which are Actionable, Trivial and Inexplicable.

Actionable terms are deemed to be effective insights that have a practicality in being assigned as a sub-search filter and category to enhance user experience. Examining a sample of Actionable words that the authors have selected in Table 5 using the Likert scale, we infer that Yelp users are interested in the atmosphere of a restaurant based upon the frequency of how many times this term appears in reviews.

Likert Sample Results		
Actionable	Trivial	Inexplicable
Atmosphere	Italian	Els
Deals/Offers	Menu	Hous
Events	Pizza	May
Healthy Choice	Italian	Month

Table 5 – Likert Sample Results

The Likert evaluation also shows that Yelp users mention Deals and Offers frequently within their posted reviews. Therefore allowing a user to filter reviews to display those that relate to *Atmosphere* and/or *Deals and Offers* is reputed to be an actionable insight.

Trivial terms scored in Likert scale are deemed to be obvious and apparent therefore these words are ineffective insights and would not enhance user experience. Focusing on a sample of words that have been rated as trivial by the authors as illustrated in Table 5, we can interpret these words as being obvious and apparent.

The authors have already identified in their summary analysis outlined in Table 2 that the majority of user reviews posted on Yelp for Las Vegas and Phoenix are related businesses that are restaurants. Moreover, the word *Pizza* has also be mentioned frequently within user reviews, referring to Table 2 again we can see that *Pizza Restaurants* are the tenth most reviewed business by users on Yelp in Las Vegas and Phoenix.

The authors infer that these trivial words are ineffective insights and hold no basis to be included as a sub-search filter or category to enhance user experience.

Inexplicable terms scored in Likert scale are deemed to be enigmatic and ineffective insights to enhance user experience. Screening the sample of words in Table 5 that have been selected as Inexplicable, we can infer that including words like *Hous* or *May* as a sub-search filters or categories would be counter-intuitive. Therefore the authors have chosen to disregard inexplicable words.

C. Evaluated Results

Upon completion of the evaluation of results phase of the analysis, the authors have identified and selected ten words using the Likert scale. The selected words illustrated in Table 6 and ordered alphabetically are deemed to be effective insights that have a practicality in being assigned as a sub-search filter and category to enhance user experience.

Proposed Sub-Search Filters & Categories			
1	Atmosphere	6	Hygiene
2	Buffet	7	Portion Size
3	Deals/Offers	8	Recommended Dishes
4	Events	9	Vegetarian Dishes
5	Healthy Choice	10	Wait time table

Table 6 – Proposed Sub-Search Filters & Categories

Examining the proposed sub-search filters & categories, the ten words relate to service, environment and food elements of a given restaurant.

For the environment term *Atmosphere* the authors infer that users of Yelp are discussing and therefor interested in gauging the overall ambience of restaurant.

Deals/Offers can be classified as a service term. Yelp alongside companies like Groupon and LivingSocial provide a service whereby said companies will advertise and promote a deal/offer on behalf of a business. For this term the authors interpret that Yelp users discuss and are engaged in sharing their experiences about Deals/Offers for a given restaurant [13],[14],[15], [16],[17].

Events and *Wait time for a table* are service elements for a restaurant. The authors infer here that Yelp users frequently post event based reviews and are observant to duration of time they will wait for a table.

Hygiene being the final environment element, for this term the authors infer that Yelp users are vigilant to hygiene and cleanliness of a restaurant.

Buffet, *Healthy Choice*, *Portion Size*, *Recommend Dishes* and *Vegetarian Dishes* being food elements, the authors infer that Yelp users discuss and review these food elements and there availability in a given restaurant.

D. Proposed Page Mechanic of Sub-search Filters & Categories.

In the current version of Yelp, when a user looks at reviews posted on a business’s page, they can sort reviews by Date, Rating as in the number of stars a review has been assigned along with reviews that been posted by Elite Yelp users as highlighted and illustrated in Figure 4. Users must scroll and paginate through multiple pages of reviews which can be a cumbersome process.



Figure 4 – Yelp Business Page Template Example February 2015

The authors propose the introduction of a new Sub-search Filter and Category menu to the business page template of Yelp based upon the evaluated results. Figure 5 illustrates how the sub search Filter and Category menu will look.



Figure 5 – Illustration of Filter Menu.

Functionally the filter menu will allow users to select one or more terms from the menu. The user’s selection will alter the appearance of reviews displayed on a business’s page, by only displaying reviews relating to those terms selected by the user. The change in appearance of the page will happen dynamically meaning the page will not flicker.

For example if a user selects *Atmosphere* and *Events* for the Soho Japanese Restaurant located in Las Vegas.

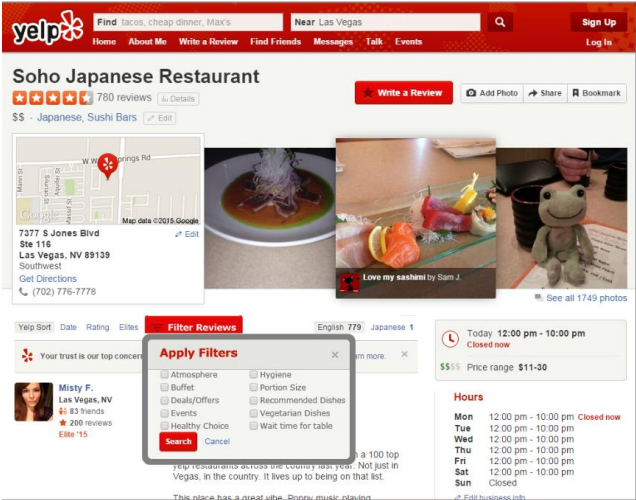


Figure 6 – Proposed Change to Yelp Business Page Template including Filters

Only reviews relating to the terms *Atmosphere* and *Events* will appear on the business’s page. In the event the user removes their selected filters then the page will revert back to its original state and display the standard view of user reviews.

E. Implications of Sub-search Filters & Categories on Yelp User Experience

The change in page mechanic and the introduction of the sub-search filter and category menu will enhance the Yelp user experience. Accessing user reviews is a core and fundamental element of Yelp's business model. The proposed change simplifies and allows users to easily navigate through the user reviews posted on a business's page.

The authors proposed change is eliminating the need for a user to scroll and paginate through multiple pages of reviews and trying to identify those that relate to *Atmosphere* and *Events* for example, for a given restaurant.

VI. CONCLUSION

The authors have taken part in the Yelp Dataset challenge and propose a new feature be implemented which will enable users to dynamically refine reviews that are displayed on a business's page through the introduction of a new sub-search filter and category menu. The overall aim of this new feature is to optimise and enhance the Yelp user experience and allow users to easily access and navigate posted reviews on a business's page relating the most frequently used terms mentioned in reviews.

The authors have conducted text mining analysis against a segment of data from prominently restaurant businesses located in Las Vegas Nevada and Phoenix Arizona as part of the Yelp Dataset Challenge 2015.

The selected dataset segment comprises of 539,535 Yelp user reviews. The adopted approach to conducting the text mining analysis was utilising the TM Text Mining package in R, and the building five corpuses containing 107,907 user reviews which assembled into a term-document matrix.

The authors ran six iterations increasing the minimum term occurrence threshold in increments of 1,000 words to calculate how frequently a term occurs within a document in the term-document matrix. The final outputted results yielded 411 words that occurred 5,000 times in user reviews posted in businesses located in Las Vegas Nevada and Phoenix Arizona.

For evaluating the outputted dataset the authors utilised a Likert scale containing three ratings of Actionable, Trivial and Inexplicable. Each of the 411 words was evaluated using the Likert scale to determine its overall validity and effectiveness in enhancing Yelp user experience in the navigating and access of reviews by users. The authors identified and selected ten words using the Likert evaluation, that are deemed to be effective

insights and have a practicality in being assigned as sub-search filter and category.

The proposed Sub-search filters & categories selected by the authors relate to service, environment and food elements of a restaurant. The proposed Sub-search filters & categories are Atmosphere, Buffet, Deals/Offer, Events, Healthy Choice, Hygiene, Portion Size, Recommended Dishes, Vegetarian Dishes and Wait time for a table.

A limitation in current Yelp Business page template was identified by the authors whereby users must scroll and paginate through multiple pages of reviews which can be a cumbersome process. The authors propose a change to the Yelp Business page template through the introduction of a new Sub-search filter & category menu, based upon the evaluated results. This new feature will allow users to select one or more terms from the menu and in turn the appearance of reviews displayed on a business's page will alter to the only display reviews relating to those terms selected by the user.

The authors highlight that the introduction of the new filter menu simplifies and allows users to easily navigate through the user reviews posted on a business's page, which is a core and fundamental element of Yelp's business model.

A. Limitations of Results

The analysis and findings of this paper have some restrictions and limitations. Firstly the base data utilised is restrictive in terms of size and represents 0.8% of the 71 million cumulative reviews posted by Yelp users in 2014.

The segment utilised in conducting the analysis is based upon reviews of restaurant businesses located exclusively in the cities of Las Vegas, Nevada and Phoenix Arizona which are western states in the United States of America. The authors acknowledge that demographics and location play a pivotal role in the type of business being reviewed and the terms found will differ from region to region.

The inclusion of reviews from cities that have larger population sizes in comparison to Las Vegas and Phoenix, for example New York City and Los Angeles would produce different results. Taking the example of Mexican restaurants and Mexican cuisine would tend to be more densely located in south western states of the USA versus states and cities located further north or on the eastern sea board of the USA.

Therefore the sub-search filters and categories will change based upon business category and location, the authors recommend that the analysis activities carried out in this paper be applied to the full yelp user review dataset and that the sub-search filters and categories be regionalised based up business type, city, state and region.

Furthermore, for Yelp to extend and implement the proposed sub-search filter & category menu to countries and cities outside of the USA, the same analysis methodology would need to be applied to the regionalised datasets.

B. Extensions of Analysis

There is scope to apply extensions to the analysis activities conducted as part of this paper. Firstly, the authors would like to conduct their analysis against a large dataset of Yelp user reviews in the region of 10 to 20 million reviews, that encompasses multiple cities spread across the entire continent of North America.

Secondly, the authors would like utilise topic modelling with focus on the Latent Dirichlet allocation to build more comprehensive corpuses that would allow for the combining of synonym based topics generated from the Yelp user reviews. This approach would allow for the creation of a diverse and versatile sub-search filter and category menu.

Finally the authors would like to factor in a social network element of analysis, by comparing the reviews of Yelp users with reviews of users whom they follow. The authors infer that users whom follow other users could potentially share the same interests; there is scope for Yelp to use these analysis findings in a targeted market campaign by offering exclusive deals or money off coupons to users based upon their review interests.

REFERENCES

- [1] Yelp-ir.com, 'Overview | Investor Relations | Yelp', 2015. [Online]. Available: <http://www.yelp-ir.com/phoenix.zhtml?c=250809&p=irol-irhome>. [Accessed: 29- Apr- 2015]. (references)
- [2] www.yelp.com, 'Yelp', 2015. [Online]. Available: <http://www.yelp.com>. [Accessed: 02- Feb- 2015].
- [3] J. McAuley and J. Leskovec, 'Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text', 2015.
- [4] R. Chada and C. Naik, Data Mining Yelp Data - Predicting rating stars from review text.
- [5] Y. Xu, X. Wu and Q. Wang, 'Sentiment Analysis of Yelp's Ratings Based on Text Reviews', 2015.
- [6] E. Korpeoglu and J. Hall, 'YelpBusy: Prediction of Busyness Levels', 2014.
- [7] M. Fan and M. Khademi, 'Predicting a Business Star in Yelp from Its Reviews Text Alone', 2014.
- [8] W. Cavnar and J. Trenkle, 'N-Gram-Based Text Categorization', Ann Arbor Mi, vol. 48113, no. 2, pp. 161 - 175, 1994.
- [9] G. Chakraborty, M. Pagolu and S. Garla, Text mining and analysis. pp. 93-106.
- [10] R. Baeza-Yates and B. Ribeiro-Neto, Modern information retrieval. New York: Addison Wesley, 2011.
- [11] I. Feinerer and K. Hornik, 'Text Mining Package', Cran.r-project.org, 2014. [Online]. Available: <http://cran.r-project.org/web/packages/tm/index.html>. [Accessed: 28- Mar- 2015].
- [12] G. Fernandez and G. Fernandez, Statistical data mining using SAS applications. Boca Raton: CRC Press, 2010, pp. 360-390
- [13] Yelp-support.com, 'Yelp Deals & Gift Certificates | Support Center | Yelp', 2015. [Online]. Available: http://www.yelp-support.com/Yelp_Deals_Gift_Certificates?l=en_US. [Accessed: 29- Mar- 2015].
- [14] LivingSocial, 'Amazing Phoenix Daily Deals & Discounts | LivingSocial', 2015. [Online]. Available: <https://www.livingsocial.com/cities/8-phoenix>. [Accessed: 29- Mar- 2015].
- [15] LivingSocial, 'Amazing Las Vegas Daily Deals & Discounts | LivingSocial', 2015. [Online]. Available: <https://www.livingsocial.com/cities/30-las-vegas>. [Accessed: 29- Mar- 2015].
- [16] Groupon, 'Deals in Phoenix', 2015. [Online]. Available: <http://www.groupon.com/local/phoenix>. [Accessed: 29- Mar- 2015].
- [17] Groupon, 'Deals in Las Vegas', 2015. [Online]. Available: <https://www.groupon.com/local/las-vegas>. [Accessed: 29- Mar- 2015].

Group C - Contribution of Work to Project.

Our approach to this project was to complete and deliver the final results as team. Our preference was to work face to face as a group at weekends. As team we initially met each Saturday afternoon in NCI, to discuss project ideas and approaches to carrying out our analysis. In conducting the remaining work for the project we met each Sunday in the Cloud Computing room to work on our project.

Scoping out of Project Ideas	
	Ian Bassett, Shashi Bhushan, Sapna Gupta, James Matthews
ETL Phase of Project	
Extraction	James Matthews
Transformation	James Matthews
Loading & Segmenting	Ian Bassett, Shashi Bhushan, James Matthews
Summary & Exploratory Analysis	
	Ian Bassett, Shashi Bhushan, Sapna Gupta, James Matthews
Data Mining Phase	
	Ian Bassett, Shashi Bhushan, Sapna Gupta, James Matthews
Interpreting Results	
Likert Evaluation	Ian Bassett, Shashi Bhushan, Sapna Gupta, James Matthews
Determining Final Results	Ian Bassett, Shashi Bhushan, Sapna Gupta, James Matthews
Poster	
Layout	Ian Bassett, Shashi Bhushan, Sapna Gupta, James Matthews
Format	Ian Bassett, Shashi Bhushan, Sapna Gupta, James Matthews
Design	James Matthews
Paper	
Abstract	James Matthews
Introduction	James Matthews
Related Work	Ian Bassett
Methodology	James Matthews
Results	James Matthews
Conclusion	James Matthews
References	Ian Bassett & James Matthews