Harnessing Restaurant reviews made simple with *RestoScraper*

Moharnab Saikia
North Carolina State University
Raleigh,NC
msaikia@ncsu.edu

Pranjal Deka
North Carolina State University
Raleigh,NC
pdeka@ncsu.edu

Rashmi Sandilya
North Carolina State University
Raleigh,NC
rsandil@ncsu.edu

Sindhu Balakrishnan
North Carolina State University
Raleigh,NC
sbalakr2@ncsu.edu

ABSTRACT

Consumer reviews of restaurants are scattered across various websites in the Internet. A typical user, on an average visits more than two websites to get the reviews of the restaurant to dine in. He also spends a considerable amount of time to get an overall idea of the quality of the restaurant by manually comparing the reviews in those websites. On the other hand, users who visit just one or two sites are not satisfied with the quality of the reviews and are often misled by the ratings. In order to gain insights into the problem, we conducted an online survey with a set of relevant questions and collected the opinion of various users. Based on the survey results, we have implemented three different solutions to ease the process of getting reviews for a restaurant. These solutions aggregates the reviews from the most popular and reliable websites and present them to the users through a web application, a chrome extension and an Android mobile application. The following sections of the paper discuss the above mentioned solutions along with the telemetry analysis of how users perceive them. The results have showed that these solutions are very useful for the users in getting the information fast thus saving their time and effort.

Keywords

Web Scraping, restaurant, online reviews, Aggregated reviews

1. INTRODUCTION

The world of digitalization has given people, the handy option to use online sources for getting instant reviews. Restaurant review websites are the most used online source to exploit the dining alternatives. Often users come across lot of inconsistencies in the information provided in each of these websites. Each of these websites can give a different review for the same restaurant. It becomes very difficult for the user to make an informed decision. How one could leverage information from multiple review websites and make a more informed decision becomes a challenge. In order to find a solution for this problem we wanted to understand the difficulties faced by the users while they try to find reviews for a restaurant. We conducted an extensive study of how users are making use of the existing websites in getting the restaurant reviews.

Data collection on the usage of a certain tool on set of intended user groups helps to identify the problems faced by them better and come up with fix to problem or devise a better tool. We performed two types of data collection to gain a better understanding of the problems faced by the users while collecting restaurant reviews. First we performed an online survey with a specific set of questions and collected the results. We also asked the users to rate the top reviewed sites so that it would be easier for us later to select the websites to aggregate the results while providing fix to the problem. Next we performed participant observation studies where we enquired few users about their experience with the usage of various websites in getting reviews for the restaurants.

Based on the survey results and our analysis, we implemented three different solutions to aggregate the restaurant reviews from the most popular websites [2]. We chose to aggregate the reviews from the three most popular review sites namely yelp, trip advisor and four square. All the three solutions present aggregated ratings and/or reviews of the restaurants from the three websites. We run a crawler periodically that extracts the restaurant reviews from the different websites and stores it in the S3 storage of the AWS. The three solutions namely the web application, the chrome extension and the Android mobile app fetch the stored reviews and displays it to the users through their individual interfaces.

After implementing the three different solutions, we again conducted an extensive user study and analysis. We ran through two iterations of the solutions, adding and removing features based on the user study and feedback. While we conducted online surveys and user observation studies, we also ran some tools like google analytics behind the scene to analyse the user activities while using the three different solutions. As a result of all the observations, we found that the web application is the most popular amongst the users for its broad coverage of the information. Further sections discuss the system architecture, evaluation and analysis.

2. SYSTEM ARCHITECTURE

The system architecture of the three solutions combined is shown below.

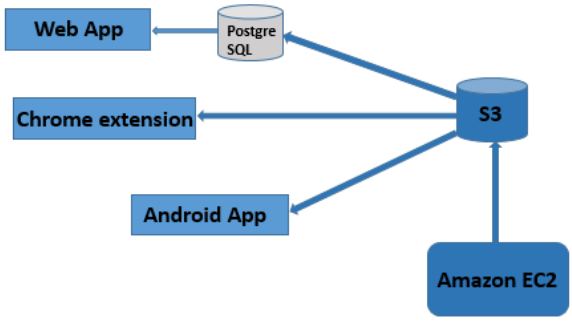

Figure 1: System Architecture

2.1 Crawler & Server

In order to fetch the restaurant reviews from the three websites we have implemented a web crawler that crawls the three websites to get the data related to the restaurant reviews [1]. Information such as overall rating, most popular and recent reviews and individual review rating and location co-ordinates are collected. This data is converted into the JSON format and stored in the s3 bucket of the Amazon Web Services(AWS). We chose to store in AWS to ease the access and maintenance of the data store to be accessible by all the three solutions.

2.2 Web Application

The web application shown in Figure 2 provides the users the option to search for a particular restaurant and view the reviews and ratings from the three popular websites namely Yelp, Foursquare and Trip Advisor. The aggregated (average) rating, the overall rating of the restaurants and the top 3 reviews and individual review ratings from the popular websites are displayed in the web application. The aggregated ratings is calculated by taking a weighted arithmetic mean(weight is the no of users providing the reviews) of the restaurant rating from the top websites. Also the location of the restaurant is displayed in a map view based on the co-ordinates stored in the JSON. By clicking on the hyperlinks, the users can navigate to the restaurant review page of the individual websites.

2.3 Android Mobile App

We have also built an Android mobile application that provides the users the option to search for their favourite restaurant. On clicking the restaurant name, the user can view the aggregated (average) rating, the overall rating of the restaurants and the top 3 reviews and individual review ratings from the popular websites similar to the web application. Again we have calculated the average rating by taking the weighted arithmetic mean of the individual restaurant ratings from the top three websites mentioned earlier. The mobile app also provides hyperlinks clicking on which the users will be navigated to the particular website namely Yelp, Trip Advisor or Foursquare. Figure 4 and Figure 5 show screenshots of the android app.

2.4 Chrome Extension

We have implemented a chrome extension as another solution (Figure 3). The extension can be installed in the chrome browser like any other extension. Whenever a user browses for the restaurant reviews on any of the three websites, the extension will fetch the ratings of that restaurant from the other two websites and display it in the current page as an overlay. This facilitates the user to compare the ratings without navigating to the other websites, thus saving time. This solution does not show the actual user reviews from the other websites as that would make the extension heavy and it is logically unconvincing to load the reviews from the other websites into the current website's page.

3. EVALUATION PLAN

The evaluation plan consists of the different ways by which we analyzed the effectiveness of the three solutions that we implemented. We conducted user surveys and telemetry analysis to understand user needs and improve the solutions based on the results.

3.1 User Survey

For user survey we targeted 20 users who were willing to try the different solutions which we implemented. We observed the users as they were using the three solutions and at the end of the user observation we asked them to fill in an online survey consisting of a set of 5 relevant questions that would help us to understand what the users wanted. Some of the questions which gave us significant results are (i) Which solution do you prefer the most? (ii) How much time you spent to get the reviews using any of the above solution? (iii) Which of the additional feature would you like us to implement? and (iv) Which is the most time taking task in the entire process of dining at a restaurant? The survey results gave us a better idea of the effectiveness of the three solutions.

3.2 Telemetry Analysis

Apart from directly observing the users and asking survey questions, we decided to track user activity behind the scenes to deduce new interesting results and observations that cannot be assessed through direct user surveys. We implemented the logic that keeps track of user clicks and the patterns of usage. As we have hyperlinks to the top three websites in our solutions, we were curious to know the rate at which the users are navigating to the other sites in all the three solutions. This would help us to decide if the users were satisfied with the information provided in our solutions. On the contrary, if the users were not happy with the information, then they would navigate to the original websites to get more information.

3.3 Results & Observations

3.3.1 User Survey

The results of the user survey gave us interesting insights into the effectiveness but more importantly the areas of improvement in all the three solutions.From Figure 13 we can say that 68.18% of the users felt that the web application is the best solution in terms of the amount of information it provides to the users and its availability to all types of users.The chrome extension and the android app had a favoring population percentage of 9.09% and 22.73% respectively.

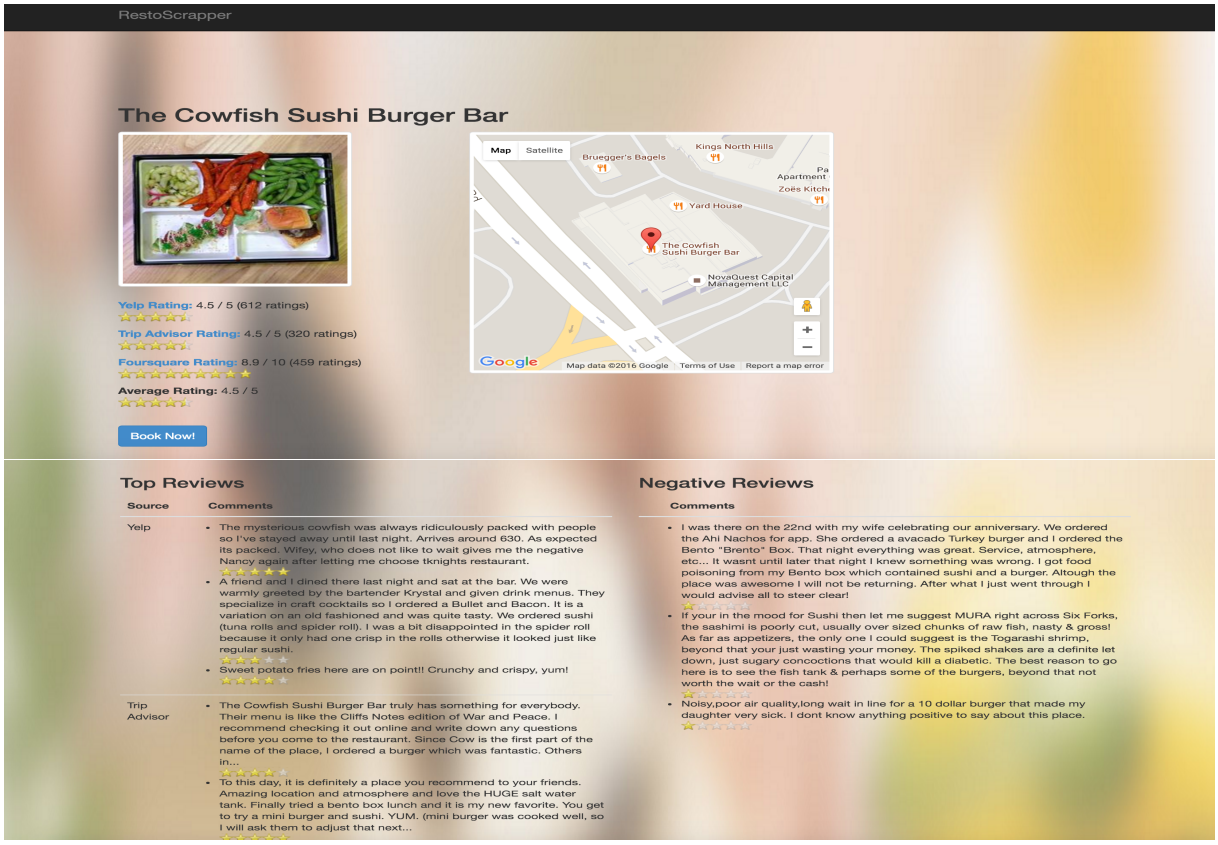

Figure 2: Web Application

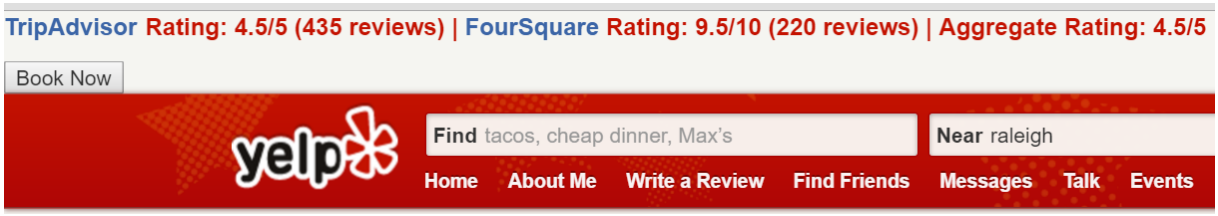

Figure 3: Chrome extension

Even though the Android mobile app is handy and easy to use anywhere, it targets only the android users whereas the chrome extension needs the users to have the chrome browser installed. Even if we argue that installing the chrome is not a big deal, 7 of the 20 users commented on the amount of information available in the chrome extension.

In our initial survey we observed that 37% of the users spent more than 5 minutes in checking the reviews for a restaurant by visiting multiple websites.This can be referred from Figure 6. Using our solutions the time spent by the users in getting reviews for a restaurant has reduced and from Figure 12 we can see that now at an average 63.64% of users take 0-3 minutes and 90.91% take 5 minutes or less.This is a significant improvement as now 27% more users compared to January are taking 5 minutes or less for getting their reviews.Thus we can say that we have achieved our goal of saving the effort and time spent by the users in looking for restaurant reviews.

3.3.2 Telemetry Analysis

The telemetry logs shown in Figure 7 is the actual log collected by the Web Application. We have implemented the code that tracks user activities such as hit count, Yelp redirect count, Foursquare redirect count, Trip Advisor redirect count and booking count. These values will give us many interesting observations which are discussed as follows.

Though initially we did not expect to gain much of insights through telemetry analysis, the logs that we collected gave us very interesting results. We could refer from Figure 9 that 67.4% of the users clicked the hyperlinks to the individual websites(Yelp, Trip Advisor, Foursquare) in case of the chrome extension and the rate was 48.6% in the android application. But in the case of the web application only 36.5% users navigated through the hyperlinks to the other

Figure 4: Android App

Figure 5: Android App

websites. This indicates the fact that most of the users were satisfied with the quality of aggregated information provided in the web application. The results for the chrome extension are self-explanatory as it provides the users only with the aggregated rating information.

We also observed that about 35.3% of users clicked the 'Book now' button in the web application, 15.5% of the users who used the android app clicked the same and the numbers were 9.6% in the case of chrome extension.This can be referred from Figure 11. Even though booking a table is not a primary feature of the solutions, a considerable amount of people clicked the option while some were very curious about its functionality.

3.4 Solution Iterations

Based on the user feedback and our evaluation results, we went through two iterations of the solution development where we incorporated extensive changes to our code base. The changes made at each iteration and the analysis results that motivated the changes are discussed below. At the end of each iteration we also made sure that we had not negatively impacted the already existing features in the three solutions. This was achieved by adding questions about changes in the user experience in the subsequent surveys.

3.4.1 Iteration 1: Negative reviews

In order to add additional features to our product we conducted a user survey as mentioned in the previous section. We asked the users the features that they would be mostly interested in. Through this we came to know that 59.09 % of the users supported that negative reviews be displayed along with the popular reviews.Also the 2nd and 3rd popular features were hospitality reviews and cleanliness ratings respectively with percentages 22.73% and 18.18%. Hence we re-wrote the crawler that pulled the top 5 negative reviews along with the top 5 popular reviews. The web application and the android mobile app were also modified to reflect the change by displaying the negative reviews in their individual user interfaces.

3.4.2 Iteration 2: Book Now

After the first iteration of implementing the negative review display in the applications, we again conducted a user survey with the same population mentioned earlier and also telemetry analysis to make sure that we had not negatively impacted any existing feature of the applications. Though the users felt that there was not much of a change in the overall user experience of the solutions, we also wanted to know the most time consuming process in the entire process of dining in a restaurant.As shown in Figure 10 we came to know that after reading reviews(53.33%) the most time a user spent was on booking a table(24.17%). To mitigate this we implemented a 'Book Now' button in all three of our solutions clicking on which the users will be redirected to the actual restaurant website where they can book a table as per their convenience.

Apart from the above changes, we had also consistently made some improvements to the user interfaces both in terms of usability and ease of use. These changes had been made in both the iterations and at the end of them we also got user feedback to ensure that we have not negatively impacted any usability feature that was working properly prior to

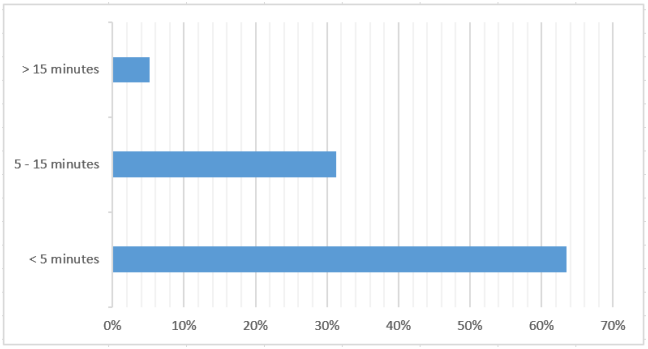

Figure 6: Time Spent(January Results)

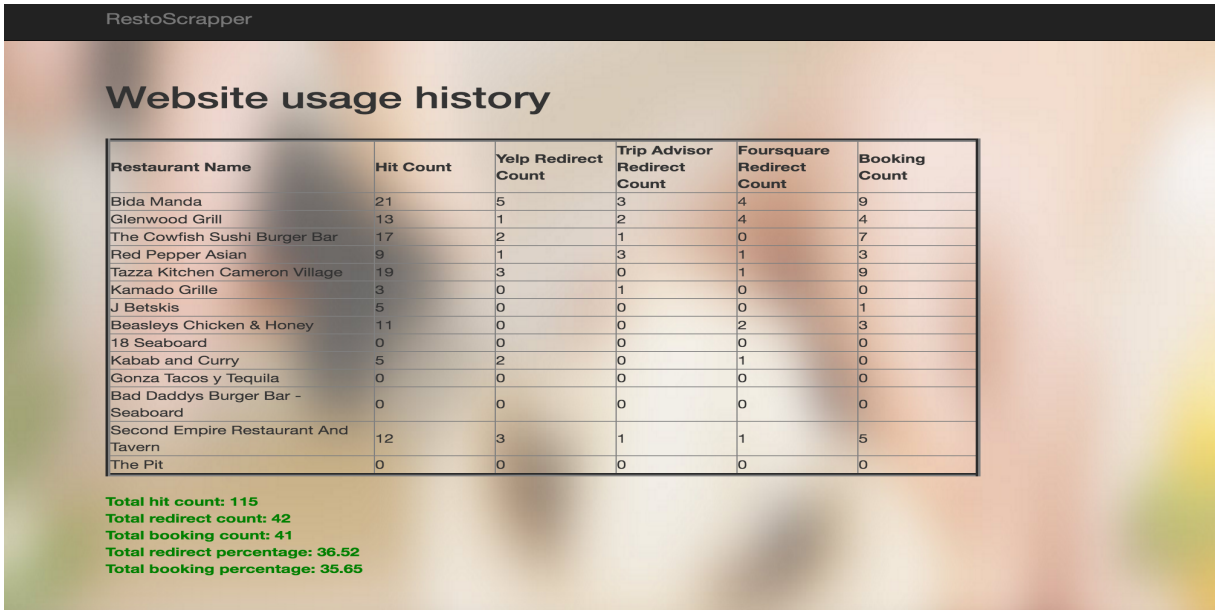

Figure 7: Web Application - Telemetry Reports

the iterations. During this survey we came to know from 4 users(total 20) that the negative reviews being displayed below the top reviews in the web app made it a little difficult to view all the information in a compact manner. So, we changed the UI and displayed both reviews side by side using two columns.

4. RESULTS

In this section we will discuss the final survey results and telemetry reports to show the effectiveness of the three solutions and how the changes we made improved the solutions. We also discuss the reasons that made us conclude that the web application is the most popular of the three solutions among the users.

4.1 Hyperlink Click Rate

We can see in the Figure 9 that the number of users who are navigating away from our applications is relatively lower in the case of the web application. Only 36.5% of the users have navigated to the original websites from our web application whereas in the case of the chrome extension the rate at which the users are clicking the hyperlinks is observed to be 67.4%. This is understandable as the chrome extension does not provide the users with much data and it merely gives them only the ratings. This was the feedback we got from most of the users.The android application stayed in a middle zone with a outbound click rate of 48.60%.

4.2 Book Now

Based on the user feedback from the second survey about the most time consuming part of dining in a restaurant, we came to know that booking a table was only next to reading the actual reviews(Figure 10). The booking feature was implemented as a result of our second iteration and it got very good user reviews when we carried out the user satisfaction survey at the end of the iterations. We observed the number of users who are clicking the 'Book now' button to book a table in a restaurant while using our solutions , as shown in the Figure **??** was 35.30% in case of the web app and 25.5% and 14.6% in case of the android app and chrome extension respectively. This was more that the population that told us about booking taking up their time(24.17%). Users had used the feature which convinced us that we have

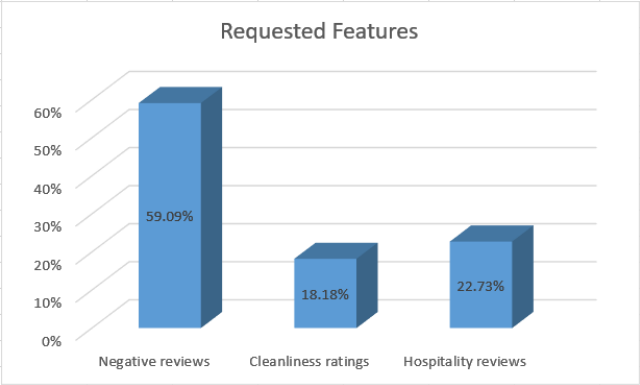

Figure 8: Requested Features(Iteration 1)

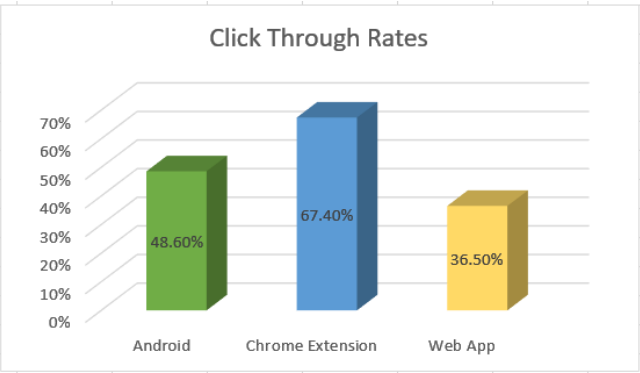

Figure 9: User outbound click rate

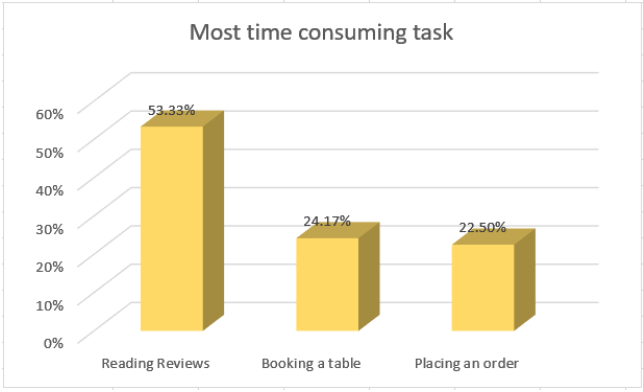

Figure 10: Most time consuming task

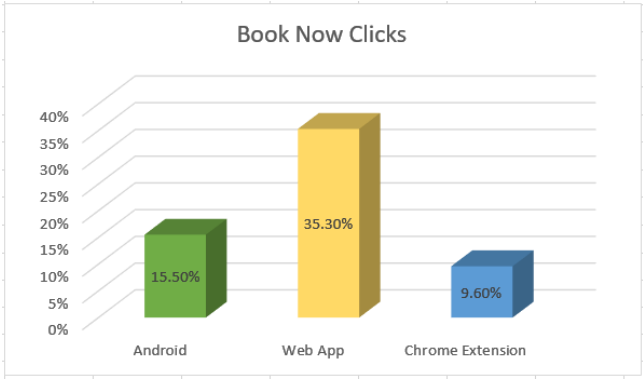

Figure 11: Book Now

provided them with the feature they wanted.

4.3 Time Spent

In our January survey we found out that users spent more time in looking for the restaurant reviews by visiting multiple web pages and manually aggregating the information to get an overall idea of the restaurant. Hence we took it as a challenge to help the users save the time and effort they spent on searching for the restaurant reviews. Our solutions that aggregate the information from the top websites are very helpful to the users even though they had some drawbacks which we constantly tried to improve during the subsequent development iterations in the month of March. Initially as seen in Figure 6 37% the users spent more than 5 minutes when they were searching for the reviews. After using the solutions that we developed they time spent has dramatically reduced such that 90.91% take 5 minutes or less time as shown in the Figure 12.

4.4 User ratings

We have also collected the user satisfaction ratings for all the three solutions and the responses have been plotted as shown in the Figure 13. It appeared that the users are very satisfied with the web application(68.18%) and they gave a variety of reasons comparing it with the other solutions. They felt that the web application gave them more information compared to the other solutions and that it is available for everyone to use. On the other hand many people were also happy with the mobile app but the only concern was it is not available for non-android users. The general feedback about the chrome extension was that it did not provide the users much of information they wanted.

5. FUTURE WORK

The goal of the project may not be limited only to aggregating the reviews from the top 3 websites that we surveyed for but also to identify fraudulent reviews and eliminate those kind of reviews in our calculation of average ratings. Some analysis has already been done in the paper [3]. Also more emphasis can be given to those reviews that are highly useful for the users. Also the data collected may can be distributed as an API such that it will be easy for all applications to use. This will also help by making it easy for other developers to use the data we collect.

6. CONCLUSION

Our extensive real time research with online survey and participant study during the month of January helped us having a closer look at the problem the users are facing in making an informed decision based on website reviews. Based on the January results ans analysis we implemented three solutions that aggregate the restaurant reviews from the most popular websites and present it to the users by means of three different applications namely the web application, chrome extension and an android mobile application. In order to test the effectiveness of the solutions we performed a series

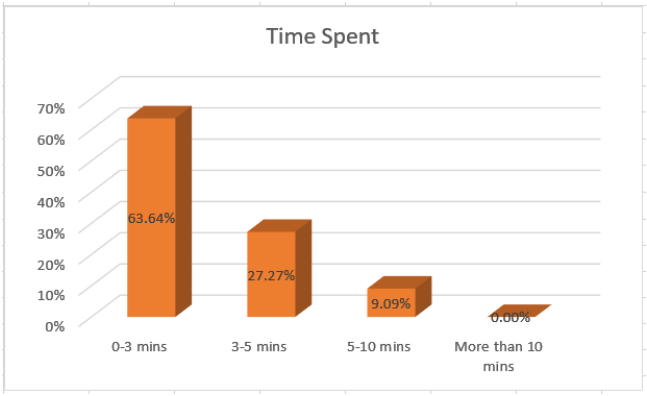

Figure 12: Time Spent

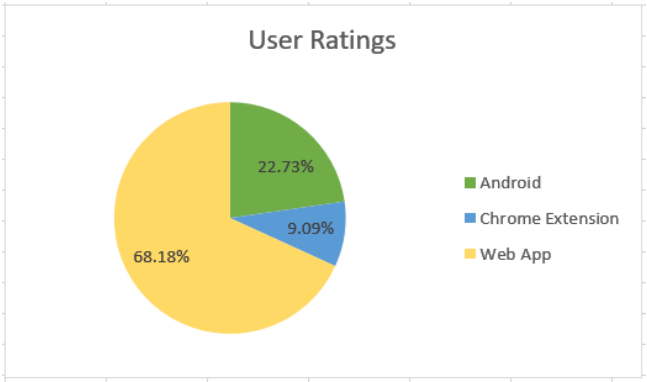

Figure 13: User ratings

of user studies and telemetry analysis and ran through two iterations of the development phase where we made extensive modifications to the code base to incorporate the new changes and features. After the iterations and further study and analysis helped us to decide that the most popular solution among the users is the web application. Sufficient statistical data and reports are presented in the paper to prove the same.

7. REFERENCES

[1] J. I. Fernández-Villamor, J. Blasco-García, C. A. Iglesias, and M. Garijo. A semantic scraping model for web resources - applying linked data to web page screen scraping. In *ICAART 2011 - Proceedings of the 3rd International Conference on Agents and Artificial Intelligence, Volume 2 - Agents, Rome, Italy, January 28-30, 2011*, pages 451–456, 2011.

[2] A. J. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos. Trueview: Harnessing the power of multiple review sites. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 787–797, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.

[3] A. Mukherjee, B. Liu, and N. Glance. Spotting fake reviewer groups in consumer reviews. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 191–200, New York, NY, USA, 2012. ACM.