**Abstract:**

The report presents deep analysis of restaurants in yelp dataset. To analyse reviews (Statements) and ratings (1 to 5 stars) of the restaurants in the dataset on a city and state level; To find how popular a restaurant is in a city. To understand how much restaurants are reviewed in different cities and state. Finally, analysis of number of reviews by a user to understand how much reviews usually a user puts (user behaviour). To find restaurants with the highest number of check-ins. To use the reviews by user and find how much does reviews does help the restaurants to improve.

**Introduction:**

Restaurants has played an important role in business, social, intellectual and artistic life. Besides great food, restaurants are important for meeting friends, relatives; spending some ‘me’ time; office meetings. Public opinion and expert reviews play an important factor for a restaurant’s success or failure. Reviews sites like Zomato, Yelp are platforms where such reviews are shared. The main objective of this report is to use the data of yelp and provide insights that can be useful for different segment of job roles and business.

Statement of the project objectives:

* To find which city has the most restaurants
* To visualize the popular restaurants in a city
* To find which restaurant has the highest number of check-ins
* To discover which restaurants users reviews the most
* To find how much reviews does a user give
* To analyse which state has the most restaurants
* Understand how much reviews are useful for restaurants
* Find how much restaurants are there in a state that were rated 5 stars by a user

Motivation of the problem:

Analysing the yelp data will help the people from various departments and business with insights which can be used for future.  
By analysing the number of restaurants in different city; the yelp sales and marketing team can target the area with less numbers to register with them.

By analysing the popularity of a restaurant; a restaurant franchise can decide where to put up a new outlet

By analysing the user review number; users can be given incentives in order to increase the review count and encourage them to visit the restaurant more often thus helping yelp to get more search traffic

Relevance of chosen topic

Considering the fact that how much importance restaurants plays in part of todays world for an individual and how reviews and ratings given by a user effects the business and how a new business can use this to open a outlet; we decided to analyse the data of restaurants to get those insights

Elicitation of appropriately formed research question(s):

The principal contributions presented in this work can be summarized as follow

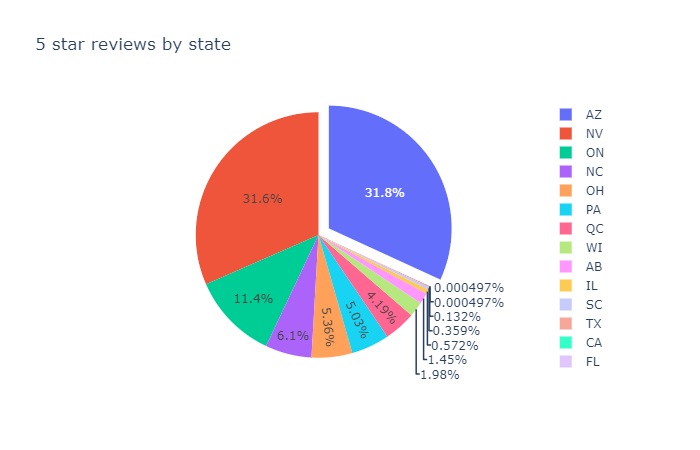
* To analyse restaurants data of yelp
* Analyse reviews, check-ins and stars of restaurants to derive insights to understand user behaviour, restaurant popularity and restaurants registered on yelp

Related Work:

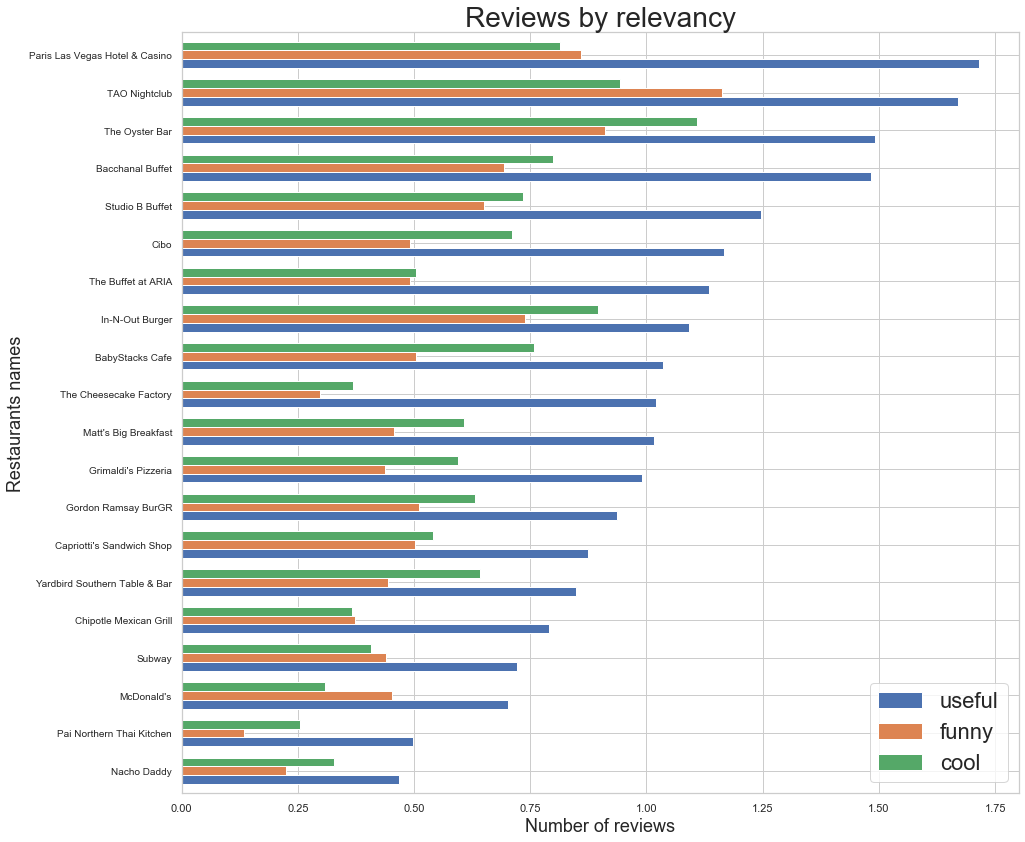
The paper has addressed the problem of setting up a new restaurant. How location can be used to boost profit for the restaurant. Location suggestion could have been done based of the cruise of the restaurant which would be more helpful [1]. Study shows how geographical proximity is important in popularity of a business in different cities [2]. To predict whether a restaurant(business) belongs to positive reviewed class or below it [3].

Results:

* States with most 5-star restaurants

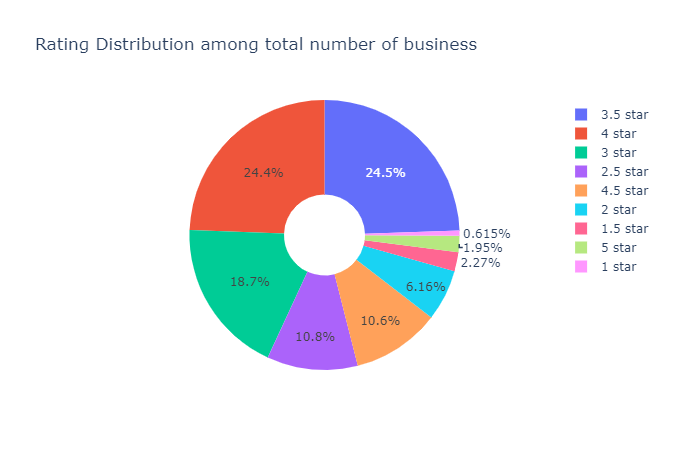
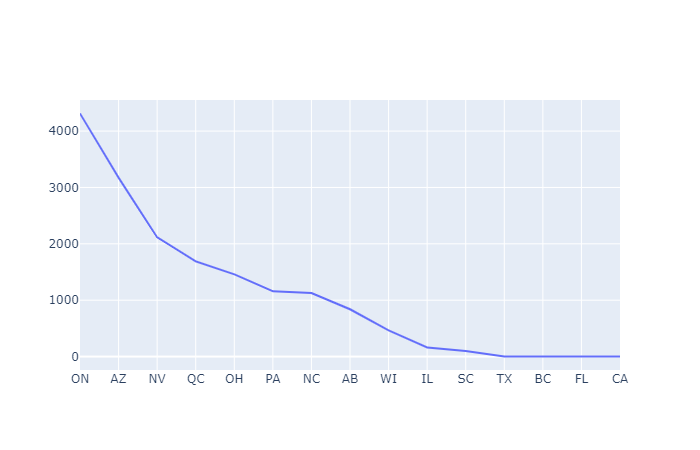
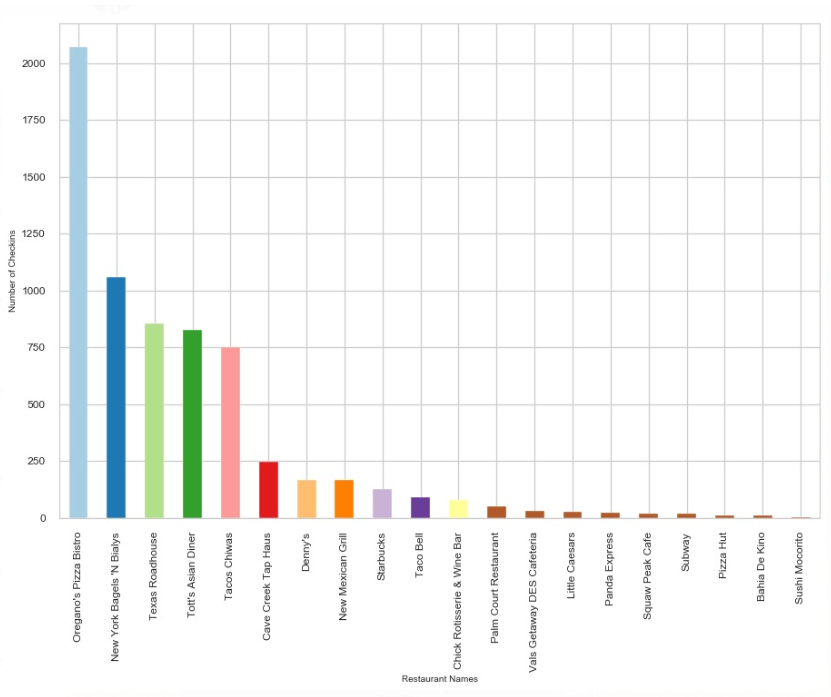
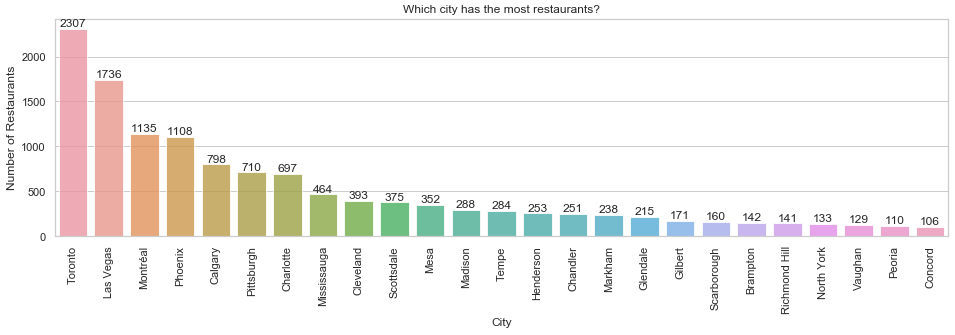
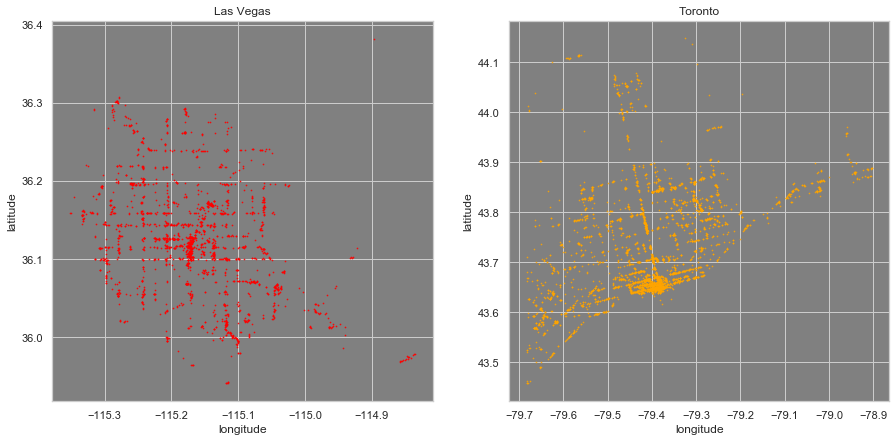


The above pull out chart gives details the distribution of 5-star restaurants among the different states. As seen Arizona has the Highest number of restaurants with 5-star ratings followed by Nevada and Ontario. This graph is particularly helpful to understand that the reviewers in other states are not completely satisfied with the restaurants and this can be also be used for a franchise which has always maintained good reviews to target the states with less 5-star restaurants to open a new outlet.

* Reviews by relevancy  
    
  The above bar graph gives a brief about the restaurants with the breakdown of the reviews. The reviews are classified as useful (good or bad), funny and cool. Of the 3 type of reviews the restaurants use the useful reviews as they are a proper feedback of the food and service and can be used to enhance in any area they are lagging. The Graph also shows that of all the 3 type of reviews; the maximum number of reviews are useful.
* Number of reviews given by a user



This graph is used to understand the user behaviour. For this graph we have kept the count of reviews to 40 i.e. no user who gave more than 40 reviews have been considered. It is seen that the average number of reviews that a user give is less than 5. This can be used by the yelp team to promote users to give more reviews as more the number of reviews the more the search traffic visits their page.

* Rating distribution  
    
  This is one of the most important insight. This pie chart shows the distribution of ratings of restaurants of the entire dataset. As per most of the restaurants are rated 3.5 or 4 stars (48.9%). Only 0.615% restaurants have the least rating i.e. 1 star and only 1.95% restaurants are rated 5 stars.
* Restaurants by State  
    
  The above chart shows the number of restaurants in different states. Ontario has the highest number of restaurants followed by Arizona. States like Texas, Florida and California have very few restaurants registered. This data is helpful for the sales team and marketing team of yelp to target these states and get more restaurants.
* Restaurants with highest number of check-ins  
    
  The above histogram shows the number of check-ins in the state of Arizona (state which has the maximum 5-star restaurants). Oregano’s Pizza Bistro has the highest number of check-ins as per the dataset. Famous Restaurant chains such as Subway and Pizza hut has few numbers of check-ins. This data can be used by Restaurants to encourage user to check-in in yelp and review them.
* Restaurants by City  
    
  The histogram shows the number of restaurants in a city. Toronto has the highest number of restaurants followed by Las Vegas. Such data can be used by business for setting up new restaurants and by yelp sales and marketing team to target cities with fewer number or restaurants.
* Popularity of Restaurants  
    
  This is a heatmap of restaurants based by their popularity. The popularity was calculated by considering the Review count \* Number of stars. The heatmap is of Toronto and Las Vegas which are the top 2 cities by number of restaurants. This chart can be used for business before setting up a new restaurant by opening a restaurant in the more popular area as this would help their business as those are the areas with the high review count areas.

Conclusions and Future Work

In this project, we analyse the data for restaurants of yelp. We found that how much usually user reviews; how much restaurants are there in a city and state; how popular the restaurant is and how much of the user reviews are useful. Finally, which restaurants user check-ins the most in a city and how the data can be useful for Yelp sales and marketing team as well as restaurant owners.

For the future, more analysis can be done on a user to find how does he/she reviews restaurants by a specific cuisine and if the reviews of user have a pattern. Also, how opening of a new restaurant in the same locality has affected the restaurant reviews and popularity. Finally, how factors such as amenities, easy of finding restaurant etc affect the restaurant performance.

References:

1. Restaurant Setup Business Analysis Using Yelp Dataset
2. <https://ieeexplore.ieee.org/document/7962431>
3. <https://ieeexplore.ieee.org/document/8169897>