

Data Analytics on a Yelp Data Set

by

Maitreyi Tata

B.Tech., Gitam University, India, 2015

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Computer Science
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2017

Approved by:

Major Professor
Dr. William H. Hsu

Copyright

© Maitreyi Tata 2017.

Abstract

In this report, I describe a query-driven system which helps in deciding which restaurant to invest in or which area is good to open a new restaurant in a specific place. Analysis is performed on already existing businesses in every state. This is based on certain factors such as the average star rating, the total number of reviews associated with a specific restaurant, the price range of the restaurant etc.

The results will give an idea of successful restaurants in a city, which helps you decide where to invest and what are the things to be kept in mind while starting a new business.

The main scope of the project is to concentrate on Analytics and Data Visualization.

Table of Contents

List of Tables	vii
Acknowledgements.....	viii
Chapter 1 Introduction	1
1.1 Problem Statement.....	1
1.2 Goals and Scope.....	2
1.3 Solution and Overview	2
1.4 End Users in the application.....	3
Chapter 2 Background and Related Work	4
2.1 MongoDB	4
2.2 Data Set and Analysis.....	6
2.2.1 Yelp Dataset:.....	6
2.2.2 Collections	7
2.2.3 Pipeline of the Project.....	8
2.3 NodeJS	9
2.3.1 Special features of NodeJS	9
2.3.2 Components of NodeJS.....	9
Chapter 3 System Design and Requirements.....	11
3.1 System Design	11
3.1.1 Use Case Diagram.....	11
The above use case diagram gives the use cases that are used in the project. These use cases represent the factors that are responsible for success of a restaurant.....	13
3.1.2 Use case Description.....	13
3.2 System Diagram.....	15
3.3 Functional Requirements:.....	16
3.4 Sequence Diagram	17
Chapter 4 Results	18
4.1 Analytics	18
4.1.1 Star Rating:	18
4.1.2 Wi-Fi Facilities:	19

4.1.3 Valet Parking:	21
4.1.4 Credit Card Factor:	23
4.1.5 Map Reduce function	24
4.2 User Interface.....	26
Chapter 5 Conclusions and Future Work.....	28
5.1 Conclusions:.....	28
5.2 Future Work:.....	29
5.3 Contribution of CIS 798 (Big Data Analytics) and CIS 732 (Machine Learning and Pattern Recognition) courses in this project.....	29
Chapter 6 References	30

Figure 2-1 MongoDB Architecture ^[3]	5
Figure 2-2 Components of NodeJS ^[6]	10
Figure 3-1 Use Case Diagram.....	12
Figure 3-2 System Diagram.....	15
Figure 3-3 Sequence Diagram.....	17
Figure 4-1 Star Rating Data.....	18
Figure 4-2 Star Rating Graph.....	19
Figure 4-3 Wi-Fi Data.....	20
Figure 4-4 Restaurants with Wi-Fi facilities.....	21
Figure 4-5 Parking Data.....	22
Figure 4-6 Parking Graph.....	22
Figure 4-7 Restaurants which accept Credit Card.....	23
Figure 4-8 Restaurants which do not accept Credit Card.....	24
Figure 4-9 MapReduce Function.....	25
Figure 4-10 MapReduce output on a portion of yelp data set.....	26
Figure 4-11 User Interface.....	27

List of Tables

Table 1 Use Case Description 1	13
Table 2 Use Case Description 5	14
Table 3 Functional Requirements	16

Acknowledgements

I would like to thank my major professor Dr. William H. Hsu of Computer Science Department for his constant academic support and feedback on this project. I would also like to express my gratitude to my graduate academic adviser, Dr. Mitchell Neilsen who had helped me throughout my Master's degree and Dr. Carlos Castellanos for taking time to be in my committee.

I would like to thank my family for their tremendous support in every way possible and for having faith in me.

Chapter 1 Introduction

1.1 Problem Statement

Due to the vast number of restaurants spread worldwide, it is very difficult to get to know the quality of a restaurant and the services they provide. Yelp is one such website that provides all the information regarding a restaurant. Yelp provides us with reviews and ratings provided by the users and other services offered by the restaurants such as free Wi-Fi, parking etc.

In-order-to invest in a business or start a new one, one should have an idea of the reasons behind the success of the restaurant. The original yelp data set has the general details of restaurants but there are no analytics to infer the customer satisfaction and the reasons behind the success of a restaurant.

If the factors contributing to the success of a restaurant are known, it would be easy for the future investors to make a wise financial decision easily. The solution to this problem consists of some analytics that give an idea to the future owner about the success factors of the restaurants.

Data visualization is representing the data in a pictorial or graphical format. It helps the decision makers to see the analytics visually, so that it would become easy for people to understand the concepts or patterns.

Data visualization helps in the following ways ^[1]

- Identify areas that need attention or improvement
- Clarify which factors influence customer behavior
- Help us understand where and what type of products to be placed
- Predict sales volumes

1.2 Goals and Scope

The main goal of this project is to provide an insight to the future investor/owner, on where and what business to invest in or where to start a new business based on the customer satisfaction.

The scope of this project is mainly Data Analytics and Data visualization.

1.3 Solution and Overview

The proposed solution can help in the following ways.

1. It gives information about the factors that are responsible for the success of a restaurant. In this way it gives the top most successful businesses and the factors responsible for the success based on customer satisfaction so that future investor/owner gets an idea of where to invest or start a new business.
2. The analytics also help the owners to concentrate on areas that need improvement or invest in an already successful business which help in the success of a restaurant.

This solution consists of some functional requirements in which MapReduce and Mongo queries are used to perform the analysis. The results thus obtained are tabulated and are also represented in graphical forms for a better understanding.

The obtained results help in identifying the factors contributing to the success and the factors to be kept in mind while starting a new business.

If someone wanted to invest or start a new business in a state and city, the analytics help in grouping the top restaurants in that area. The results are ordered by state, which makes it easy to search for and can decide based on it.

1.4 End Users in the application

Current Business Owners – They are the people who own businesses and are registered on yelp and update their business information.

Future owners and/or Investors – They are the people who would like to invest or start a new business in future

Chapter 2 Background and Related Work

This is a type of full stack application which uses data from the database, performs the analytics and visualization and it is projected in the front end for the users to use as a web application. MongoDB can analyze any type of data in real time and it does it faster and better.

2.1 MongoDB

Advantages of using MongoDB over other databases ^[2]

1. MongoDB can incorporate any type of data no matter what size, format and source.
2. MongoDB is built to scale out very large amount of data present in the cloud or data center.
3. MongoDB analyzes any type of data within its own database and gives us real-time results.

MongoDB has a flexible data model and dynamic schema which allows the users to store data in multi-dimensional documents to quickly build and evolve.

MongoDB has a great flexibility, scalability and high performance when compared to all other databases. It enables to store and process any kind of data with little or no difficulty.

Architecture:

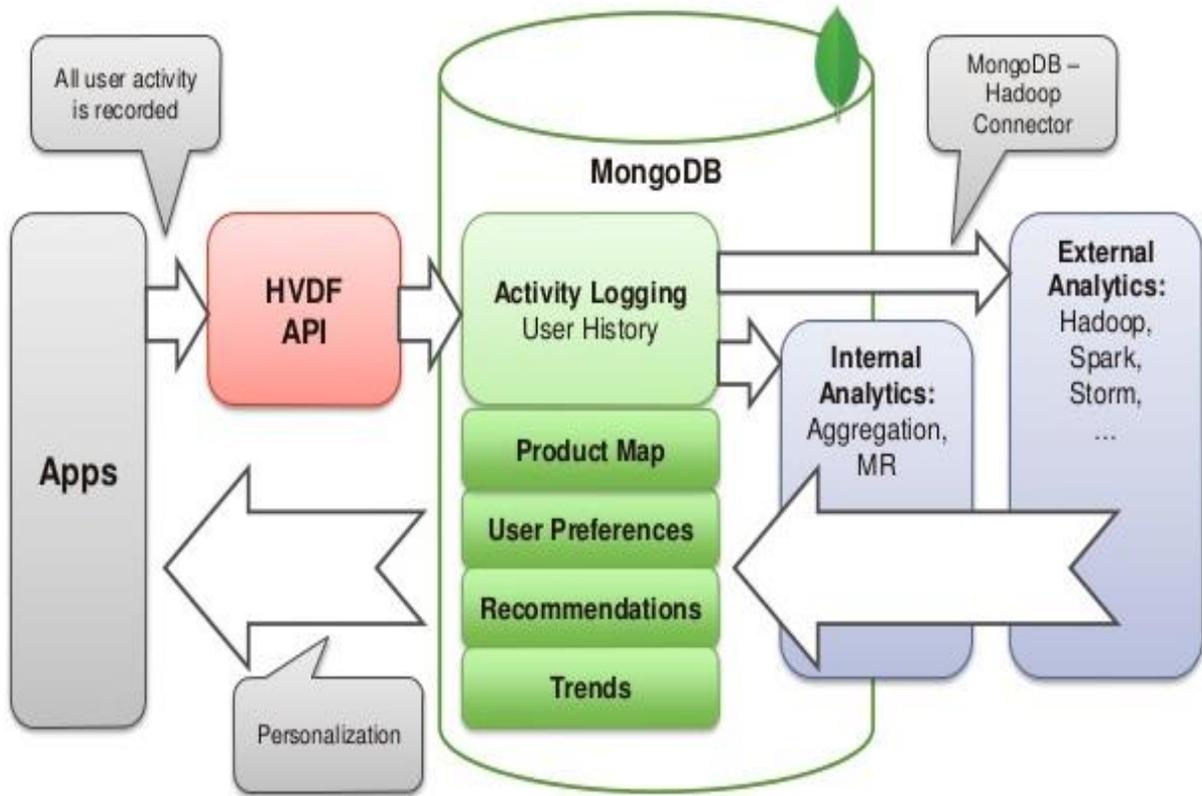


Figure 2-1 MongoDB Architecture ^[3]

The data set is downloaded and the collections which are JSON files are imported into the MongoDB. It records all the user activities and history. The queries written to perform analytics are stored in its own database. The visualization is performed and the results can be set to any type of chart. The results obtained are then connected to the front-end web application using NodeJS which is accessed by the users.

MongoDB helps in the following ways.

- MongoDB lets you store the data in JSON file formats which are highly flexible. It means that the data structure can be changed over time.
- The MongoDB document model maps to the objects in the application code making it easy to work with the data.
- MongoDB is a distributed database and its built in high availability, geographic distribution makes it easy to use.
- Indexing, Real time aggregation provide the most powerful ways to access and analyze the data.
- MongoDB has indexes, which are similar to the indexes in any other database. It supports indexes on any field of a document. ^[4]

2.2 Data Set and Analysis

2.2.1 Yelp Dataset:

The dataset downloaded from yelp was of 9.2 GB. There were 5 collections in total. They are

- business.json
- checkin.json
- review.json
- tip.json
- user.json

In this project, the analysis was mainly concentrated on business table as the information required to the future owners is mainly located in the business collection.

2.2.2 Collections

Business

This collection consists of several details related to every business such as location, business id, review count, type of food it provides, parking factors, peak hours and what type of environment it has etc.

It contains almost every detail related to the business.

Check in

This collection gives the day of check in and the check in time of customers. It also displays the business id corresponding to it.

Review

The review collection of this dataset contains columns such as the user id, their rating and the message that they post as review. Natural Language Processing can be used to analyze this collection to determine what exactly the customer thinks.

Tip

The tip collection consists of a useful message from the customer to other customers.

It consists of the user id of the person who posted the message and the date on which the tip was posted.

User

This collection contains several details of the yelp users such as the user id, user name, since when are they using yelp, review count, friends of the users etc.

All the above collections are given in the dataset out of which the important details that we need for this project are present in the business table.

These collections are JSON format files. They are imported to the MongoDB and then the analysis is done on the data and the results are visualized for better understanding.

2.2.3 Pipeline of the Project

- Clean the irrelevant data present in all the files and store them in a new and cleaned dataset.
- Identify the relevant details that can be analyzed with the help of data that is already given in the data set.
- Identify the overall quality of the business by considering the average star rating.
- Perform the data analysis and compare the factors responsible for success with the average star rating of that business.

2.3 NodeJS

NodeJS is a JavaScript runtime that runs on Chrome's v8 engine. It uses a non-blocking and event driven I/O model.

2.3.1 Special features of NodeJS

NodeJS has many special features that make it powerful and efficient. ^[5]

Asynchronous and Event Driven – All the API of NodeJS are asynchronous, that is, non-blocking. It means that a NodeJS server never waits for an API to return the data. After calling it, the server then moves to the next API.

Very Fast – As it was built on Google Chrome's V8 engine, the NodeJS library is super-fast in code execution. This helps in short execution time of the data even if the size is huge.

High Scalability – NodeJS uses a single threaded program. It can provide service to a much larger number of requests than any other server.

No Buffering – The NodeJS applications never buffer any kind of data and they output the data in small chunks.

2.3.2 Components of NodeJS

The NodeJS modules, NPM, Development tools and Frameworks, Node CLI altogether constitute the NodeJS components package. The JSON files hold various meta data

related to the project. This is used to give information to the NPM which will allow to identify the project. The NPM is the package manager to NodeJS packages. It hosts thousands of free packages to download and use them. [6]

NodeJS component diagram

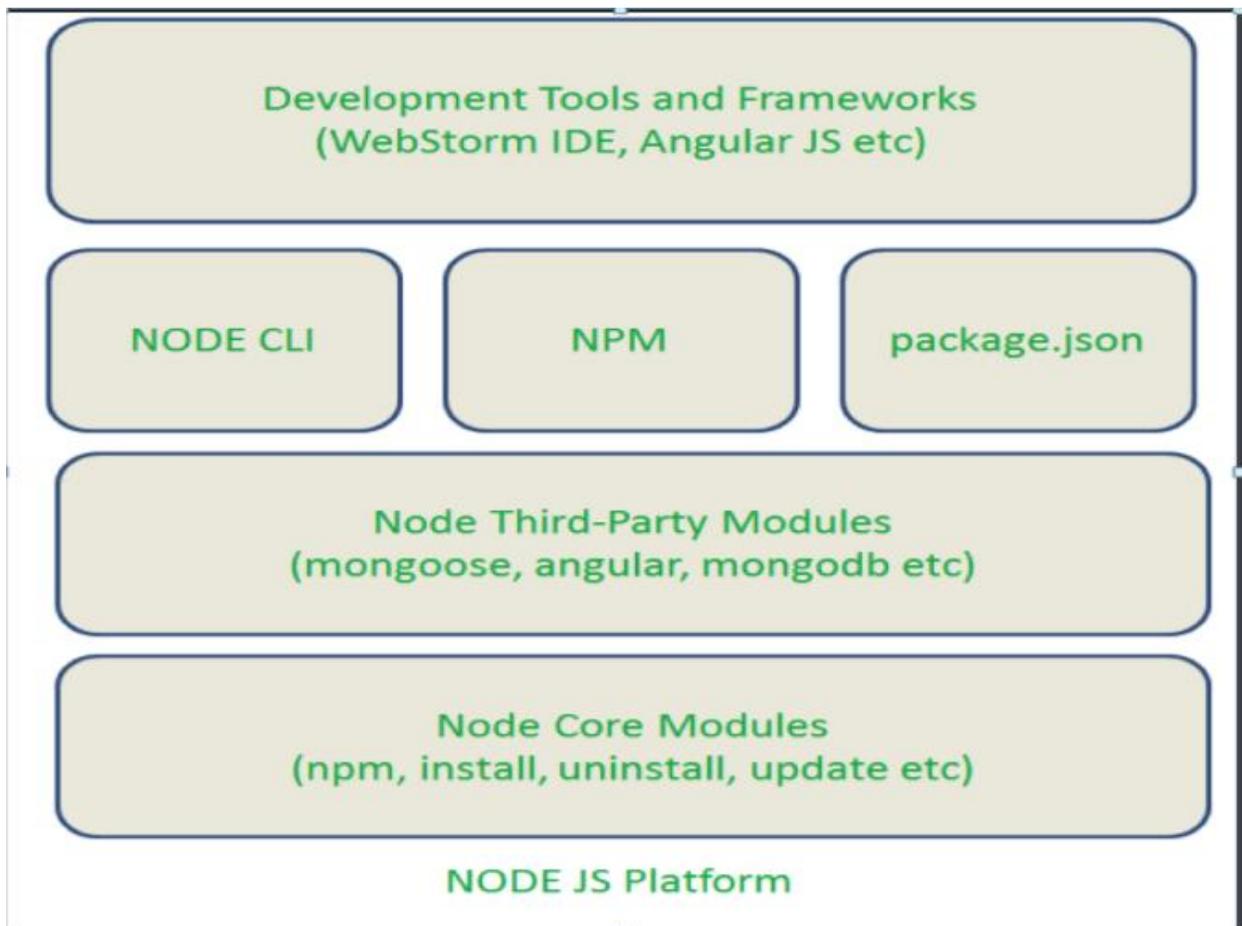


Figure 2-2 Components of NodeJS [6]

Chapter 3 System Design and Requirements

3.1 System Design

The files downloaded from the yelp website are imported to MongoDB. They are then queried and analyzed for required specifications. The results thus obtained are visualized and plotted on various charts for easy understanding. The pictorial representation helps the users to understand the underlying trends and patterns very easily.

The system design is a process in which the architecture, modules, and the data for a system are specified to satisfy the requirements. It is the total process of systems theory to product development.

3.1.1 Use Case Diagram

The Use Case Diagram are referred to as behavior diagrams which are used to describe the series of actions that some of the systems perform in collaboration with the external users.

The Use Case Diagram here represents all the actions that are carried out through the process. ^[7]

Use Case Diagram

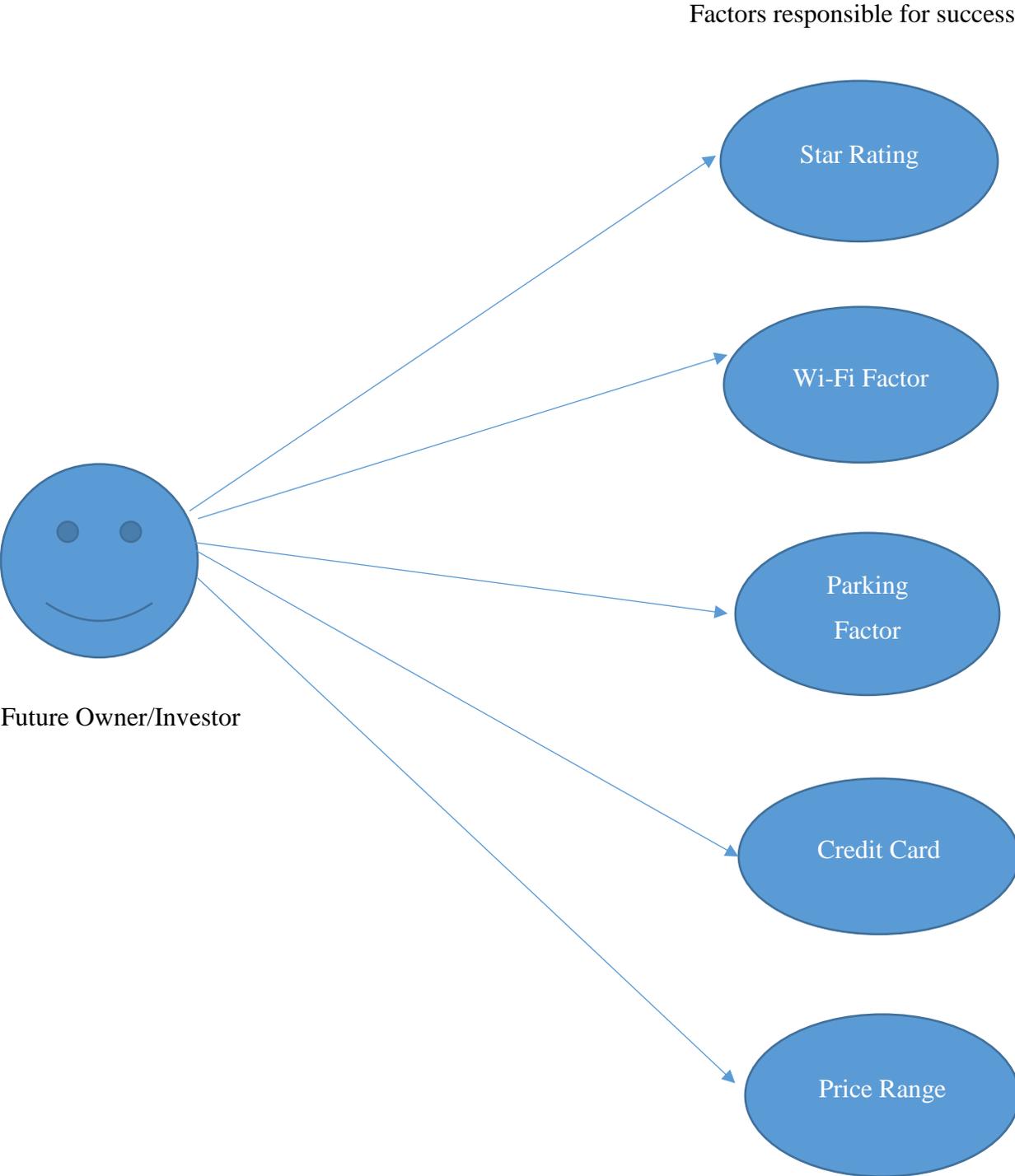


Figure 3-1 Use Case Diagram

The above use case diagram gives the use cases that are used in the project. These use cases represent the factors that are responsible for success of a restaurant

3.1.2 Use case Description

Use case 1	Identify the restaurants with star 3.5 or higher
Initiating Actors	Future investor/owner
Actor’s Goal	To list out all the restaurants with average star rating equal to or higher than 3.5 in the area requested by the user.
Pre-Condition	The current business owners should be registered on yelp so that they can provide all the information regarding their business and future investors who have access to the data.
Post-Condition	System lists out all the restaurant businesses with average star rating 3.5 or higher.

Table 1 Use Case Description 1

Use Case 5	Price Range
Initiating Actors	Future investor/owner
Actor's goal	To list out the highest and lowest price ranges of the restaurants and compare them based on the star rating they have.
Pre-Condition	Current business owners should provide the relevant data and future investors must have access to it to determine the required information.
Post-Condition	System determines the highest and lowest priced restaurants and compares them against the star rating.

Table 2 Use Case Description 5

The above two use cases give the higher rated restaurants and compares the restaurants with higher price range and lower price range efficiently.

In the first use case, the future owner is the actor and he would like to know the list of restaurants that have star rating of 3.5 or higher.

3.2 System Diagram

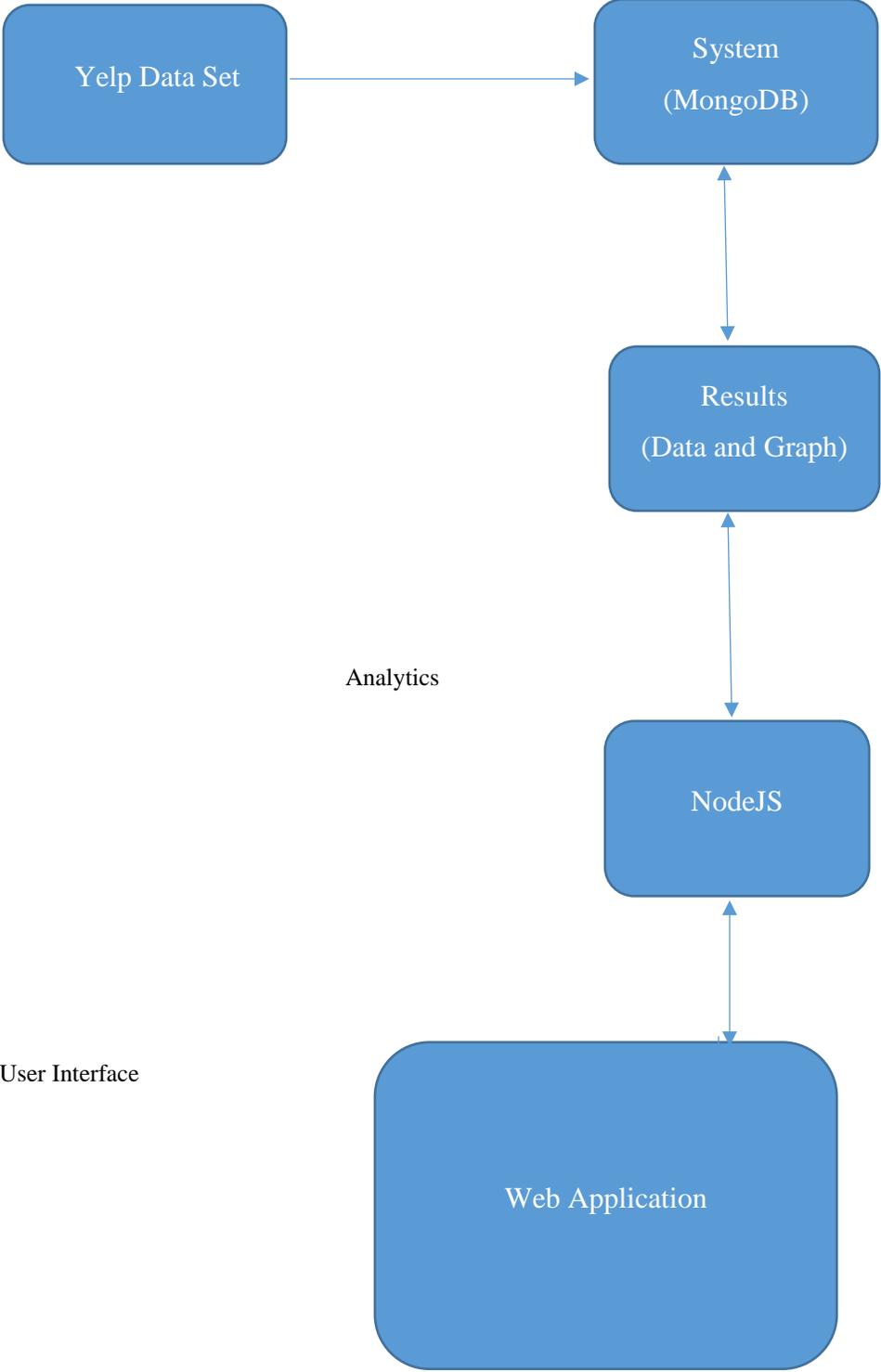


Figure 3-2 System Diagram

3.3 Functional Requirements:

Query	Requirement
1. Find out top rated restaurants in a state.	To display the list of all restaurants with start rating equal to 3.5 or higher
2. Find out the Wi-Fi factor	To find out the restaurants that have free wi-fi and their corresponding star rating.
3. Valet Parking	To find out the restaurants that have valet parking and compare the star rating to check if plays a role in the success.
4. Credit Card	Get the list of restaurants that accept credit cards and check their star rating to identify how much effect it has on the overall success of a business.
5. Price Range	Find out the restaurants that have higher and lower price range and compare the results based on star rating

Table 3 Functional Requirements

3.4 Sequence Diagram

Every time the user uses the web application to see the analytics, it hits the MongoDB and gets updated. It then retrieves the data and gives out the results.

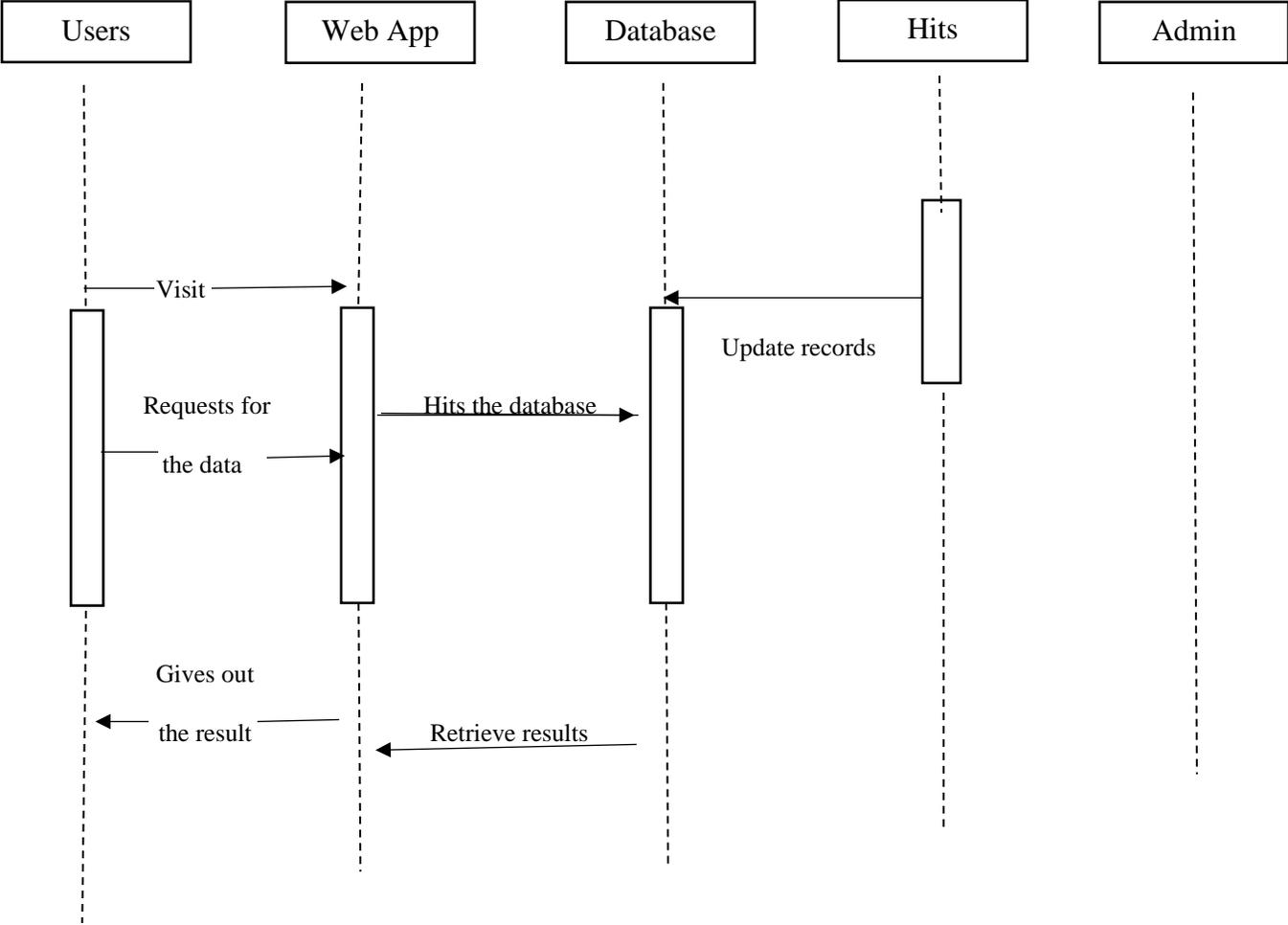


Figure 3-3 Sequence Diagram

Chapter 4 Results

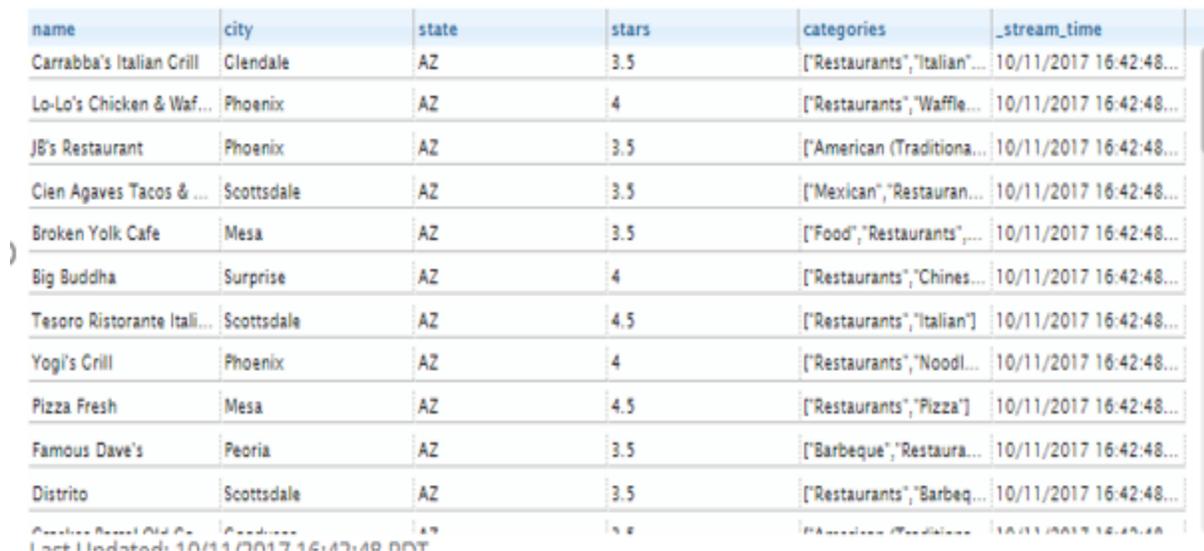
4.1 Analytics

The data is analyzed for various use cases that are mentioned above. The results obtained through analytics are explained below.

4.1.1 Star Rating:

The star rating of the restaurants ranges from 1 to 5. The restaurants with rating 3.5 or higher are most likely the most successful restaurants in a state. The result is arranged by state so that it's easy to check for all the restaurants in a state where the future owner would like to start the business.

List of restaurants with star rating 3.5 or higher:



name	city	state	stars	categories	_stream_time
Carrabba's Italian Grill	Glendale	AZ	3.5	["Restaurants", "Italian"]	10/11/2017 16:42:48...
Lo-Lo's Chicken & Waf...	Phoenix	AZ	4	["Restaurants", "Waffle"]	10/11/2017 16:42:48...
JB's Restaurant	Phoenix	AZ	3.5	["American (Traditiona..."]	10/11/2017 16:42:48...
Cien Agaves Tacos & ...	Scottsdale	AZ	3.5	["Mexican", "Restauran..."]	10/11/2017 16:42:48...
Broken Yolk Cafe	Mesa	AZ	3.5	["Food", "Restaurants", "..."]	10/11/2017 16:42:48...
Big Buddha	Surprise	AZ	4	["Restaurants", "Chines..."]	10/11/2017 16:42:48...
Tesoro Ristorante Itali...	Scottsdale	AZ	4.5	["Restaurants", "Italian"]	10/11/2017 16:42:48...
Yogi's Grill	Phoenix	AZ	4	["Restaurants", "Noodl..."]	10/11/2017 16:42:48...
Pizza Fresh	Mesa	AZ	4.5	["Restaurants", "Pizza"]	10/11/2017 16:42:48...
Famous Dave's	Peoria	AZ	3.5	["Barbeque", "Restaura..."]	10/11/2017 16:42:48...
Distrito	Scottsdale	AZ	3.5	["Restaurants", "Barbeq..."]	10/11/2017 16:42:48...
...

Figure 4-1 Star Rating Data

This above table shows all the restaurants with average rating equal to or higher than 3.5. Later, many other factors are compared with the star rating they have got, to understand the effect of that facility on the rating.

Graph:

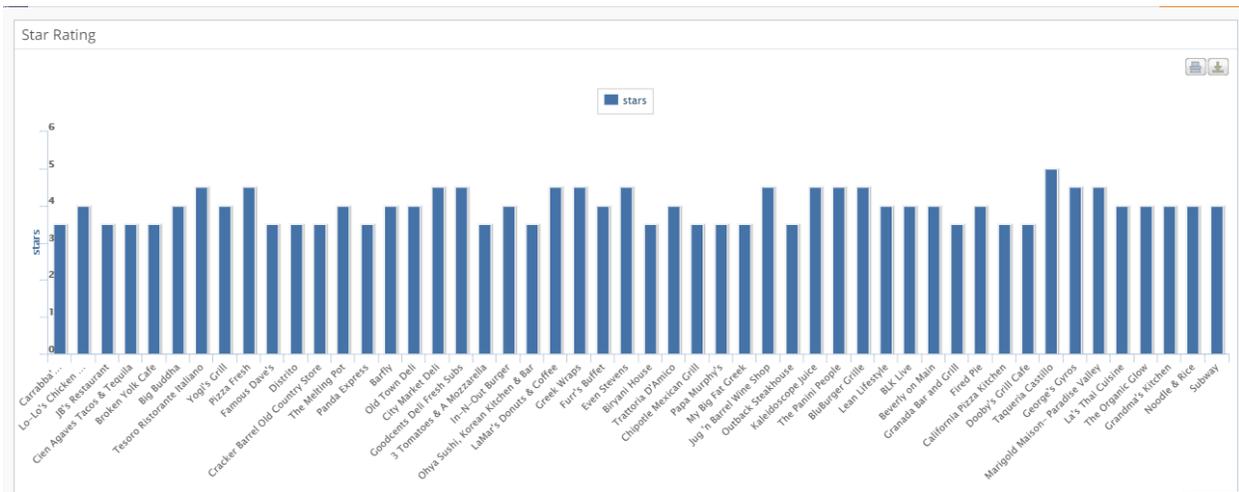


Figure 4-2 Star Rating Graph

4.1.2 Wi-Fi Facilities:

If the future owner wants to start a restaurant in the downtown of a city, Wi-Fi plays an important role. As most of the offices are located in the downtown of a city, people who visit a restaurant for lunch might have to finish up important office work. And for this,

having Wi-Fi is very helpful. In this project, the restaurants with Wi-Fi are compared against the star rating they got.

Restaurants with Wi-Fi facilities

Data:

name	state	stars	attributes	categories	_stream_time
Lo-Lo's Chicken & Waf...	AZ	4	["Alcohol: full_bar", "A...	["Restaurants", "Waffle...	10/27/2017 23:30:06...
Cien Agaves Tacos & ...	AZ	3.5	["Alcohol: full_bar", "A...	["Mexican", "Restauran...	10/27/2017 23:30:06...
Broken Yolk Cafe	AZ	3.5	["Alcohol: full_bar", "A...	["Food", "Restaurants", ...	10/27/2017 23:30:06...
Coco's Family Restaur...	AZ	2.5	["Alcohol: none", "Amb...	["American (Traditiona...	10/27/2017 23:30:06...
Big Buddha	AZ	4	["Alcohol: beer_and_w...	["Restaurants", "Chines...	10/27/2017 23:30:06...
Distrito	AZ	3.5	["Alcohol: full_bar", "A...	["Restaurants", "Barbeq...	10/27/2017 23:30:06...
Pecan Grove Restaura...	AZ	2.5	["Alcohol: none", "Amb...	["Burgers", "American (...	10/27/2017 23:30:06...
The Melting Pot	AZ	4	["Alcohol: full_bar", "A...	["American (New)", "Ste...	10/27/2017 23:30:06...
LaMar's Donuts & Coff...	AZ	4.5	["Alcohol: none", "Bike...	["Donuts", "Breakfast &...	10/27/2017 23:30:06...
Creek Wraps	AZ	4.5	["Alcohol: none", "Amb...	["Creek", "Restaurants"...	10/27/2017 23:30:06...
McDonald's	AZ	2.5	["Alcohol: none", "Amb...	["Fast Food", "Burgers", ...	10/27/2017 23:30:06...
Burger King	AZ	2.5	["Alcohol: none", "Amb...	["Fast Food", "Burgers", ...	10/27/2017 23:30:06...

Figure 4-3 Wi-Fi Data

Graph

The names of restaurants are plotted on X-axis and Star rating was plotted on Y-axis

Restaurants with Wi-Fi facilities

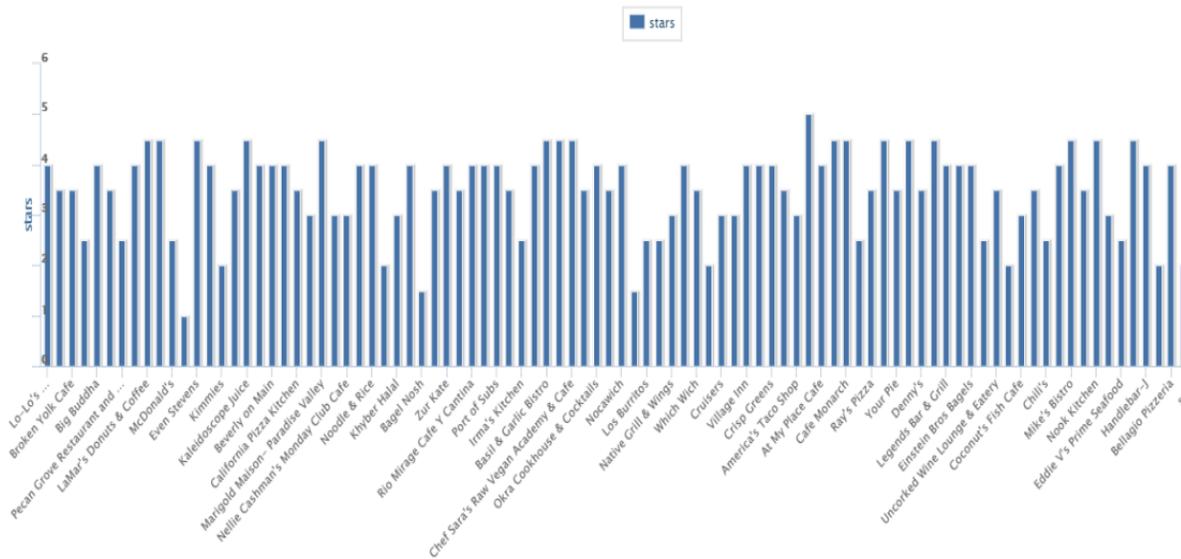


Figure 4-4 Restaurants with Wi-Fi facilities

4.1.3 Valet Parking:

It would be so frustrating when we do not find a parking and just keep looking for one wasting a lot of time. To check out whether the customers may find it as a plus point, I ran a query that checks the star rating against valet parking facility.

It was found that most of the restaurants with valet parking have the rating 3.5 or higher. This makes the future owner consider the valet parking to be an important factor while establishing a new business

Restaurants with Valet Parking and their star rating:

name	state	stars
Lo-Lo's Chicken & Waffles	AZ	4
Distrito	AZ	3.5
Nellie Cashman's Monday Club Cafe	AZ	3
Marcella's Italian Kitchen	AZ	3
Eddie V's Prime Seafood	AZ	4
Sutra Sushi at Barcelona	AZ	3.5
Cibo	AZ	4.5
SWB, a southwest bistro	AZ	3.5
Arcadia Farms Cafe	AZ	4
J. Alexander's	AZ	3.5
Kai	AZ	4.5

Figure 4-5 Parking Data

Graph:

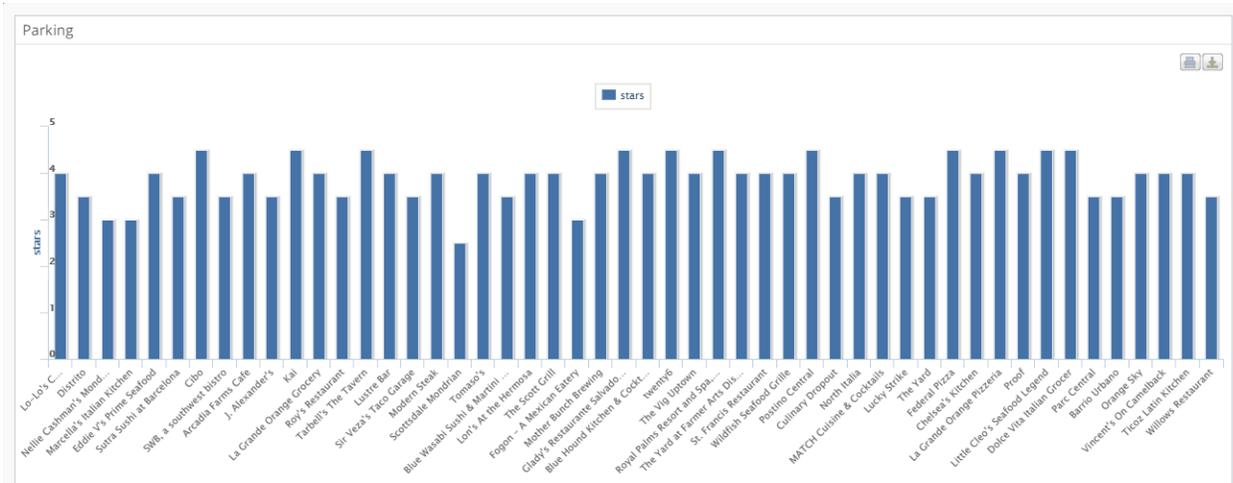


Figure 4-6 Parking Graph

4.1.4 Credit Card Factor:

Some of the restaurants do not accept Credit Card and as a customer, I can say that it is kind of a drawback that affects the overall rating of the restaurant. To find out whether this has effect on overall rating of the restaurant, I ran a query and compared the ratings of restaurants which accept Credit Card and which do not accept.

Surprisingly, it's found out that the ratings fluctuate with in both the results. There are restaurants that accept Credit Card with lower rating and vice versa indicating no strong correlation between accepting the Credit Card and the star rating.

Graph:

Restaurants which accept Credit Card

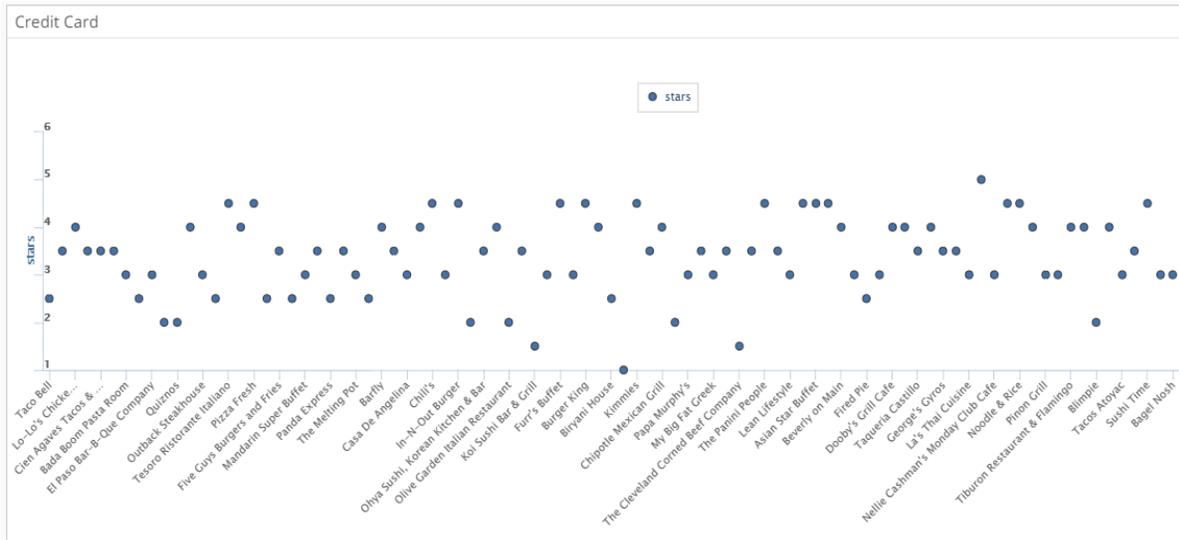


Figure 4-7 Restaurants which accept Credit Card

The above figure gives the plot of restaurants which accept Credit Card and we can see a lot of restaurants that have lower rating which is below 3.

The graph for the restaurants which do not accept Credit Cards is shown below.

Restaurants which do not accept Credit Card

It can be clearly seen that there are lot of restaurants with higher ratings compared to the restaurants that accept Credit Card.

Graph:

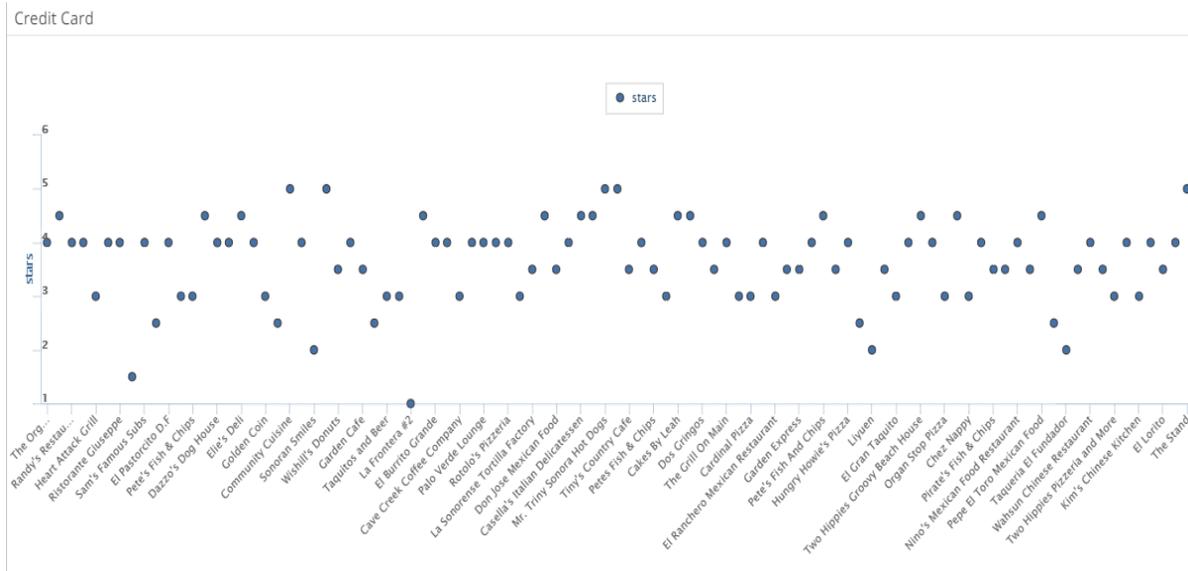


Figure 4-8 Restaurants which do not accept Credit Card

4.1.5 Map Reduce function

I have written a MapReduce operation to combine business and review collections. This in turn gives the details of business and reviews details such as the message and review id.

The function below gives us the business id and average star rating given by all users. It combines with the review collection and gets the combined result.

```
> business_map = function() { emit(this.business_id, {"review_text": this.text, "business_name": this.name, "business_address": this.full_address}); }
function () { emit(this.business_id, {"review_text": this.text, "business_name": this.name, "business_address": this.full_address}); }
> review_map = function() { emit(this.business_id, {"review_text": this.text, "business_name": this.name, "business_address": this.full_address }); }
function () { emit(this.business_id, {"review_text": this.text, "business_name": this.name, "business_address": this.full_address }); }
> r = function(key, values) {
... var result = {"review_text": "", "business_name": "", "business_address":"" };
... values.forEach(function(value) { if( value.review_text!=null){ result.review_text= value.review_text; } if(value.business_name !=null){result.business_name= value.
business_name; } if( value.business_address!= null) { result.business_address= value.business_address; } });
...
... return result;
... }
function (key, values) {
var result = {"review_text": "", "business_name": "", "business_address":"" };
values.forEach(function(value) { if( value.review_text!=null){ result.review_text= value.review_text; } if(value.business_name !=null){result.business_name= value.busi
ness_name; } if( value.business_address!= null) { result.business_address= value.business_address; } });
return result;
}
}
> res = db.business.mapReduce(business_map, r, { out: { reduce:'joined' } });
{
  "result" : "joined",
  "timeMillis" : 241354,
  "counts" : {
    "input" : 144072,
    "emit" : 144072,
    "reduce" : 0,
    "output" : 144072
  },
  "ok" : 1
}
}
> res = db.review.mapReduce(review_map, r, { out: { reduce:'joined' } });
{
  "result" : "joined",
  "timeMillis" : 1193491,
  "counts" : {
    "input" : 4153150,
    "emit" : 4153150,
    "reduce" : 208260,
    "output" : 144072
  },
  "ok" : 1
}
}
```

Figure 4-9 MapReduce Function

```

fFMolF6v1hdEgV2rxIg", "Y50kARmLdzCvL6CnN8Ww", "sYqYXDJGaj7wfaqz5u8KQ", "rSfZv6qsE0XJxdNYWdtPDA", "x1uoFAXRPMJryzW7HiGU6g", "R4eLJKwG9yQ44vfwaUH53g", "fH
_spVpnDn-K5dqEGEDuTA", "5uiCy8vSIknuDEHC7yHRk", "kSkpoS0dr8RQW50aFfV6Hw", "gTQntiB7ec-HCOYYWLSGg", "q5zFC4Bw1xnIILAZNwbk1g", "-chWR5gKglj38yZka9PU1w", "e
avB03KRzBf8sM6KGA_-NQ", "5V8eXkTJb6IejJKMDaj_Bw", "6z26YTb8Dz_ljsqj7q0PQ", "w_PyyaK-7Knpl_pIQfum3A", "KeuwC8E12385P6xyKpBM6w", "7YvMoKqLVXQ_DciRg5UDaQ", "
SrF972IxsELyQ3eaWB5X4A" ] } ] }
{ "_id": ObjectId("58c63bf0eb818f5d8882bbd1"), "business_id": "6TKJDgTrUoD3zu77A4JHnQ", "name": "Coco's Family Restaurant", "neighborhood": "", "address":
"4700 N Scottsdale Rd", "city": "Scottsdale", "state": "AZ", "postal_code": "85251", "latitude": 33.5047782, "longitude": -111.9261807, "stars":
2.5, "review_count": 13, "is_open": 0, "attributes": [ "Alcohol: none", "Ambience: {romantic: False, intimate: False, class: False, hipster: F
alse, divey: False, touristy: False, trendy: False, upscale: False, casual: False}", "BusinessAcceptsCreditCards: True", "BusinessParking: {gара
ge: False, street: False, validated: False, lot: True, valet: False}", "Caters: False", "GoodForKids: True", "GoodForMeal: {dessert: False, lat
enight: False, lunch: False, dinner: False, breakfast: False, brunch: False}", "HasTV: False", "NoiseLevel: average", "OutdoorSeating: False", "Re
staurantsAttire: casual", "RestaurantsDelivery: False", "RestaurantsGoodForGroups: True", "RestaurantsPriceRange2: 2", "RestaurantsReservations: False", "R
estaurantsTableService: True", "RestaurantsTakeOut: True", "WheelchairAccessible: True", "WiFi: free" ], "categories": [ "American (Traditional)", "Bakeri
es", "Food", "Restaurants" ], "hours": null, "type": "business", "starsData": [ { "_id": "6TKJDgTrUoD3zu77A4JHnQ", "avgStars": 2.6923076923076925, "us
ers": [ "KL-JE4VkgW02LzeS1W3e6Q", "pajL4WZeymHvrxP5eG5-lQ", "Xj002l0bp633ebmG468aZw", "-gJJzPghhC_gwHT-D8wLKQ", "Rzf35jKV8e8EJdssrae_AA", "nIR40CydnpipHM
1hd3U9Q", "sYqYXDJGaj7wfaqz5u8KQ", "oQnDZhyJlxypwq_RG3PTbw", "DFc1QxqlUjBJlGeyhbJNA", "6aSVrKFko8GyfmU6VmGyTg", "dXDNggVa8q1Gsja2Vc5ug", "GIL4XSLZguuqk
ZFRxHq_g", "A0j21z2Q1HGic7jW6e9h7A" ] } ] }
{ "_id": ObjectId("58c63bf0eb818f5d8882bbea"), "business_id": "LHueU80xZ60btISJ18niEw", "name": "El Paso Bar-Que Company", "neighborhood": "", "addre
ss": "13823 W Bell Road", "city": "Surprise", "state": "AZ", "postal_code": "85374", "latitude": 33.6383422, "longitude": -112.3579458, "stars": 3,
"review_count": 3, "is_open": 0, "attributes": [ "BusinessAcceptsCreditCards: True", "BusinessParking: {garage: False, street: False, validated: F
alse, lot: False, valet: False}", "GoodForKids: True", "GoodForMeal: {dessert: False, latenight: False, lunch: False, dinner: False, breakfast
: False, brunch: False}", "OutdoorSeating: True", "RestaurantsAttire: casual", "RestaurantsDelivery: False", "RestaurantsGoodForGroups: True", "Restaura
ntsPriceRange2: 2", "RestaurantsTableService: True", "WheelchairAccessible: True" ], "categories": [ "Restaurants" ], "hours": null, "type": "business",
"starsData": [ { "_id": "LHueU80xZ60btISJ18niEw", "avgStars": 3, "users": [ "p_azadin_uWFOXAhK3ag", "1s0Q1KwGpJIKvD-SRSpwjw", "7hfJxm8Te09lJp-Zw9Io
w" ] } ] }
> db.business.aggregate([{$lookup:{from:"result",localField:"business_id",foreignField:"_id",as:"starsData"}},{ $match :{$and: [{ categories:"Restaurants" }
,{state:"AZ"}]},{ $limit:10}]]

```

Figure 4-10 MapReduce output on a portion of yelp data set

4.2 User Interface

The User Interface consists of a login page and a registration page. The future investors/owners should register into the application before using it to retrieve the information. After they register, they should login using the credentials. It gives the list of all states that are present in the database. And they should select a state in-order-to view the information of restaurants in that area. After selecting the state, it displays a list of restaurants and their average star ratings given by all users combined.

A screenshot of the web application page is given below.

States in USA	
<input checked="" type="checkbox"/> AZ	
<input type="checkbox"/> NV	
<input type="checkbox"/> ON	
<input type="checkbox"/> PA	
<input type="checkbox"/> BW	
<input type="checkbox"/> QC	
<input type="checkbox"/> NC	
<input type="checkbox"/> WI	
<input type="checkbox"/> OH	
<input type="checkbox"/> EDH	
<input type="checkbox"/> SC	
<input type="checkbox"/> WLN	
<input type="checkbox"/> IL	
<input type="checkbox"/> NI	
<input type="checkbox"/> MLN	
<input type="checkbox"/> STG	
<input type="checkbox"/> HLD	
<input type="checkbox"/> NY	
<input type="checkbox"/> FIF	

Business Name	Average Rating
Taco Bell	2.3333333333333335
Carrabba's Italian Grill	3.453781512605042
Lo-Lo's Chicken & Waffles	3.946708463949843
Pho Viet	2.6666666666666665
JB's Restaurant	3.4583333333333335
Broken Yolk Cafe	3.7445255474452557
Cien Agaves Tacos & Tequila	3.7083333333333335
Bada Boom Pasta Room	3.130434782608696
Coco's Family Restaurant	2.6923076923076925
El Paso Bar-B-Que Company	3

Figure 4-11 User Interface

Chapter 5 Conclusions and Future Work

5.1 Conclusions:

The trends and patterns obtained from analytics, vary for various businesses. The factors responsible for success are unpredictable in some cases like the restaurants which accept Credit Cards and which do not accept. The data is analyzed for all the use cases that are mentioned at the starting of the project. The analytics used for customer satisfaction gives an insight to the future investor/owner on where to invest or start a new business and the factors that are to be considered for a successful business apart from just the quality of food. The main concentration is done on Analytics platform and Data Visualization. The project as a whole implements as a full stack as it has database, Analytics and User Interface. The results of analytics queries are projected in the User Interface for the users of web application.

This project has helped me in deep understanding of MongoDB and NodeJS which is a complete new topic for me. Major difficulty was faced while learning NodeJS from scratch and implementing it along with MongoDB.

5.2 Future Work:

In the continuation of this project, I would like to implement it in Scikit Learn. I would like to explore the hidden patterns and use Natural Language Processing and Sentimental Analysis which can give more accurate results. I want to use Sentimental Analysis on the review collection, to find out how the customers really feel about the restaurant. It helps to understand and analyze the review that they have written.

As I started learning R, which is very much easy to implement the predictive Analysis, I would like to try out the analytics in R platform.

5.3 Contribution of CIS 798 (Big Data Analytics) and CIS 732 (Machine Learning and Pattern Recognition) courses in this project

My strong interest towards Analytics started with the course CIS 798. The assignments consisted of wide range of technologies and it helped me to understand how each technology/language differ from the other in performing analytics.

CIS 732 introduced me to the concepts of machine learning and I would like to use it in my future work.

Chapter 6 References

1. SAS Institute Inc., Data Visualization-what it is and why it matters
Retrieved from
https://www.sas.com/en_us/insights/big-data/data-visualization.html
2. MongoDB Inc., Real Time Analytics
Retrieved from
<https://www.mongodb.com/use-cases/real-time-analytics>
3. SlideShare., MongoDB., Oct 16, 2014
Retrieved from
<https://www.slideshare.net/mongodb/prepare-for-peak-holiday-season-with-mongodb>
4. MongoDB Inc., MongoDB Architecture
Retrieved from
<https://www.mongodb.com/mongodb-architecture>
5. Tutorials Point., Node.js Introduction
Retrieved from
https://www.tutorialspoint.com/nodejs/nodejs_introduction.htm
6. Rambabu Posa., Node.js Components., Nov 08, 2017
Retrieved from
<https://www.journaldev.com/7423/node-js-components-modules-npm-install-update-uninstall-example>

7. UML Use Case diagrams

Retrieved from

<http://www.uml-diagrams.org/use-case-diagrams.html>