Northeastern University

DA5020 17561 Collect/Store/Retrieve Data Fall 2017

PROJECT REPORT

Topic: FOODPANDA

An Analysis of Foodpanda Delivery System

Submitted by:

Raiyan Riyazahmedikkeri

Swagath Varanchi Chandregowda

Submitted to:

Prof. Kathleen Durant

INDEX

Table of contents **Page**

OVERVIEW

The Foodpanda is a global food delivery marketplace headquartered in Belin, Germany, operating in 12 countries and territories. The service allows users to select from local restaurants and place orders via its mobile applications as well as its websites. The company has partnered with over 27,095 restaurants in 193 cities and works with over 15,733 delivery riders. Foodpanda has tie up with 15,000+ restaurants in India with home delivery and take away service also providing discount voucher codes and faster delivery promise. Here in this project we are analyzing food delivery data from 10 cities from Foodpanda India website with information such as restaurants, their ratings, reviews, cuisines etc. The data scraped has 12,000+ rows of data with all the information mentioned above. The data would then be cleaned to get the data that would be used for analysis.

PROBLEMS DRIVING THE PROJECT

- Restaurants currently offering discounts and which city provides the highest.
- Restaurants having no reviews and which city has the most number of 'no reviews' restaurants.
- Top 20 restaurants having highest number of reviews and which city has the highest number of Top 20 restaurants.
- City wise number of restaurants and highest average ratings of the restaurants.
- Obtaining bar graphs for above listed problems to analyze visually.

SCRAPING THE DATA

We used tools such as Instant Data Scraper, import.io and Grepsr to scrap the data from Foodpanda India website. We scraped data for 10 popular Indian cities and transformed the content into analysable form. The unwanted data in the dataset was removed and then after cleaning the data it was stored in a data frame.

PROCESSES IMPLEMENTED

- Used Packages like RSQLite, tidyverse, dplyr, readxl, ggplot2 etc.
- Cleaned the data and stored it in a data frame.
- Performed operations like Average to visualize data.
- Compared the data from the obtained analysis.
- Obtained bar graphs for cities with highest average rating, more number of reviews, and low number of reviews.
- City providing more discounts.
- Stored it in a SQLite database storage system.

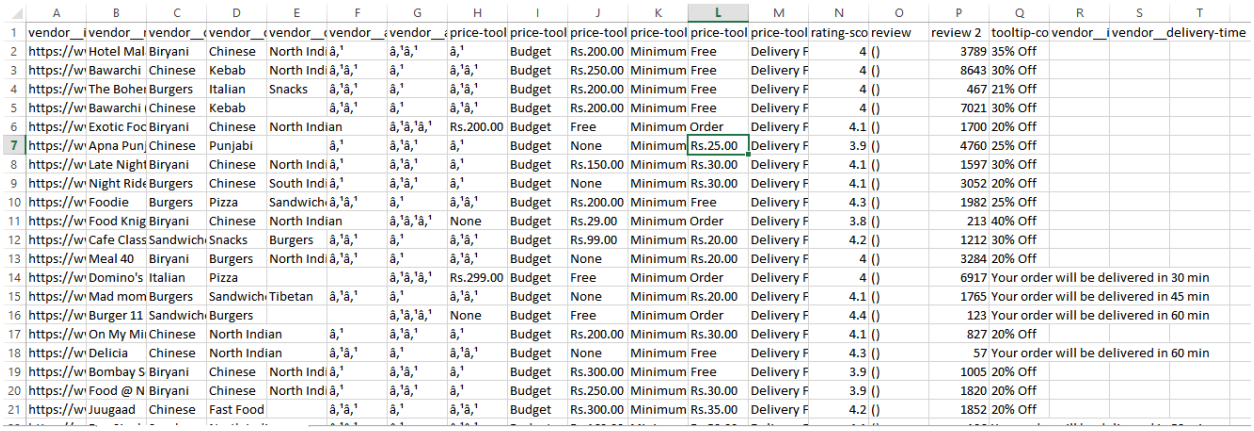
TECHNICAL CHOICES

- **Instant Data Scraper, Grespr and Import.io:**
  Open source tools and extensions were used to scrap the data from the website.
- **RSQLite:** SQLite is an open source, embedded relational database which helps to directly import a data frame into a database connection. Allows us to use queries to communicate with the database.
- **Dplyr:** It provides simple verbs, functions that correspond to the most common data manipulation tasks, to help translate those thoughts into code. It uses efficient data storage back ends.
- **Barplot:** It is used to show visually the results obtained after analysis.

ISSUES

- Issue: Scraping the data from Foodpanda India website was quite difficult as there were thousands of restaurants.
  Solution: We scraped the data on the basis of cities, we scraped the data of 10 cities and generated 10 CSV files and later merged it into one.
- Issue: The data scraped was unclean and we to manipulate it to increase the readability and also to make the retrieving process easier.
  Solution: We cleaned each table individually and stored in separate data frames and finally concatenated into a single data frame.
- Issue: The scraped data was not in required numeric format and had dollar signs and commas along with unwanted rows
  Solution: We made use of gsub function to remove the unwanted characters and converted the mode into numeric.

Uncleaned Data Set:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	vendor__i	vendor__r	vendor__c	vendor__c	vendor__c	vendor__c	vendor__a	price-tool	price-tool	price-tool	price-tool	price-tool	price-tool	rating-sco	review	review 2	tooltip-co	vendor__i	vendor__	delivery-time
2	https://w	Hotel Mal	Biryani	Chinese	North Indi	â,¹	â,¹â,¹	â,¹	Budget	Rs.200.00	Minimum	Free	Delivery F	4	()	3789	35% Off			
3	https://w	Bawarchi	Chinese	Kebab	North Indi	â,¹â,¹	â,¹	â,¹â,¹	Budget	Rs.250.00	Minimum	Free	Delivery F	4	()	8643	30% Off			
4	https://w	The Bohe	Burgers	Italian	Snacks	â,¹â,¹	â,¹	â,¹â,¹	Budget	Rs.200.00	Minimum	Free	Delivery F	4	()	467	21% Off			
5	https://w	Bawarchi	Chinese	Kebab		â,¹â,¹	â,¹	â,¹â,¹	Budget	Rs.200.00	Minimum	Free	Delivery F	4	()	7021	30% Off			
6	https://w	Exotic Foc	Biryani	Chinese	North Indi	an	â,¹â,¹â,¹	Rs.200.00	Budget	Free	Minimum	Order	Delivery F	4.1	()	1700	20% Off			
7	https://w	Apna Pun	Chinese	Punjabi		â,¹	â,¹â,¹	â,¹	Budget	None	Minimum	Rs.25.00	Delivery F	3.9	()	4760	25% Off			
8	https://w	Late Night	Biryani	Chinese	North Indi	â,¹	â,¹â,¹	â,¹	Budget	Rs.150.00	Minimum	Rs.30.00	Delivery F	4.1	()	1597	30% Off			
9	https://w	Night Ride	Burgers	Chinese	South Indi	â,¹	â,¹â,¹	â,¹	Budget	None	Minimum	Rs.30.00	Delivery F	4.1	()	3052	20% Off			
10	https://w	Foodie	Burgers	Pizza	Sandwich	â,¹â,¹	â,¹	â,¹â,¹	Budget	Rs.200.00	Minimum	Free	Delivery F	4.3	()	1982	25% Off			
11	https://w	Food Knig	Biryani	Chinese	North Indi	an	â,¹â,¹â,¹	None	Budget	Rs.29.00	Minimum	Order	Delivery F	3.8	()	213	40% Off			
12	https://w	Cafe Class	Sandwich	Snacks	Burgers	â,¹â,¹	â,¹	â,¹â,¹	Budget	Rs.99.00	Minimum	Rs.20.00	Delivery F	4.2	()	1212	30% Off			
13	https://w	Meal 40	Biryani	Burgers	North Indi	â,¹â,¹	â,¹	â,¹â,¹	Budget	None	Minimum	Rs.20.00	Delivery F	4	()	3284	20% Off			
14	https://w	Domino's	Italian	Pizza			â,¹â,¹â,¹	Rs.299.00	Budget	Free	Minimum	Order	Delivery F	4	()	6917	Your order will be delivered in 30 min			
15	https://w	Mad mom	Burgers	Sandwich	Tibetan	â,¹â,¹	â,¹	â,¹â,¹	Budget	None	Minimum	Rs.20.00	Delivery F	4.1	()	1765	Your order will be delivered in 45 min			
16	https://w	Burger 11	Sandwich	Burgers			â,¹â,¹â,¹	None	Budget	Free	Minimum	Order	Delivery F	4.4	()	123	Your order will be delivered in 60 min			
17	https://w	On My Mi	Chinese	North Indian		â,¹	â,¹â,¹	â,¹	Budget	Rs.200.00	Minimum	Rs.30.00	Delivery F	4.1	()	827	20% Off			
18	https://w	Delicia	Chinese	North Indian		â,¹â,¹	â,¹	â,¹â,¹	Budget	None	Minimum	Free	Delivery F	4.3	()	57	Your order will be delivered in 60 min			
19	https://w	Bombay S	Biryani	Chinese	North Indi	â,¹	â,¹â,¹	â,¹	Budget	Rs.300.00	Minimum	Free	Delivery F	3.9	()	1005	20% Off			
20	https://w	Food @ N	Biryani	Chinese	North Indi	â,¹	â,¹â,¹	â,¹	Budget	Rs.250.00	Minimum	Rs.30.00	Delivery F	3.9	()	1820	20% Off			
21	https://w	Juugaad	Chinese	Fast Food		â,¹â,¹	â,¹	â,¹â,¹	Budget	Rs.300.00	Minimum	Rs.35.00	Delivery F	4.2	()	1852	20% Off			

Cleaned Data Set:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	vendor_id	link	vendor_n	cuisine	cuisine_2	cuisine_3	cuisine_4	affordabil	minimum	delivery_f	rating	review_co	discount	city
2	1	https://w	Night Rest	Chinese	Seafood	North Indian		$	99	50	4.2	792	20% Off	bangalore
3	2	https://w	Eatsomnia	Chinese	Continent	North Indian		$$	99	50	4.1	2927	20% Off	bangalore
4	3	https://w	Hunger Hi	Biryani	Chinese	North Indian		$	150	50	4.2	866	20% Off	bangalore
5	4	https://w	Midnight I	Italian	North Indian			$$$	50		4.2	1060		bangalore
6	5	https://w	Go Italia (I	Burgers	Pizza	Italian		$	200	50	3.9	241		bangalore
7	6	https://w	Night Fork	Biryani	Chinese	North Indian		$$	99	50	4	985	20% Off	bangalore
8	7	https://w	B.M.W (Bl	Chinese	Snacks	North Indian		$		50	4	130	20% Off	bangalore
9	8	https://w	Midnight I	Italian	Pizza			$$$	50		4.2	414		bangalore
10	9	https://w	8 Slice	Snacks	Pizza			$	150	50	0		20% Off	bangalore
11	10	https://w	The Charc	Italian	Pizza			$	150	50	0		20% Off	bangalore
12	11	https://w	Box Of Bir	Kebab	North Indian			$	150	50	0			bangalore
13	12	https://w	Delly Bell	Chinese	North Indian			$$	150	50	0			bangalore
14	13	https://w	Kokomo	Thai	Chinese			$$$			0			bangalore
15	14	https://w	Andhra Ho	Biryani	Chinese	Kebab		$		Free	4	5177	20% Off	bangalore
16	15	https://w	Freshkhila	Biryani	Chinese	North Indian		$		Free	4.1	3364	15% Off	bangalore
17	16	https://w	Maa Redd	Biryani	Chinese	North Indian		$		Free	4.1	8586	20% Off	bangalore
18	17	https://w	Abhiruchi	Chinese	North Indian			$		Free	4.2	1861		bangalore
19	18	https://w	Jaffa's Bir	Biryani	Chinese	North Indian		$$		Free	4	11844	20% Off	bangalore
20	19	https://w	Punjabi Ta	Biryani	Chinese			$		Free	4	5397	20% Off	bangalore
21	20	https://w	Meghana	Chinese	Punjabi	Regional		$$		Free	4	902	20% Off	bangalore

- Cleaned data set with 14 columns and 12058 rows

INPUT

Data Cleaning:

```
vendors <- read.csv("combined.csv")
foodpanda <- read.csv("combine.csv")

#Data cleaning

#affordability rows did not recognize $ function while scraping the data, hence it can be replaced as sho

fp <- select(foodpanda, affordability )
f.v <- as_vector( fp)
df <- data_frame(affordability = character())

for(i in 1:NROW(f.v))
{
  x= f.v[[i]]
  p1 = 'â,''
  p2 = 'â,'â,''
  p3 = 'â,'â,'â,''
  p4 = 'â,'â,'â,'â,''
  p5 = 'â,'â,'â,'â,'â,''

  t1 = if (grepl(p1,x)==TRUE){gsub(p1,"$",x)}
  else if(grepl(p2,x)==TRUE){gsub(p2, "$$", x)}
  else if(grepl(p3,x)==TRUE){gsub(p3, "$$$", x)}
  else if(grepl(p4,x)==TRUE){gsub(p4, "$$$$", x)}
  else if(grepl(p5,x)==TRUE){gsub(p5, "$$$$$", x)}
  else {x}
  df[i, ] = t1
}

df
```

- Data cleaning to remove unwanted characters from Affordability

```
#Some vendors have minimum order and some do not.
#We cleaned the data and make the CSV file more readle to analyze which restaurants have no minimum order

fs <- select(foodpanda, minimum_order )
f.u <- as_vector( fs)
ds <- data_frame(minimum_order = character())

for(i in 1:NROW(f.u))
{
  x= f.u[[i]]
  p1 = 'Rs.'
  p2 = 'Free'
  p3 = 'None'


  t1 = if (grepl(p1,x)==TRUE){gsub(p1,"",x)}
  else if(grepl(p2,x)==TRUE){gsub(p2, "", x)}
  else if(grepl(p3,x)==TRUE){gsub(p3, "", x)}

  else {x}
  ds[i, ] = t1
}

ds

foodpanda$minimum_order <- ds$minimum_order
```

- Data cleaning to make the Minimum order column more readable

Outputs for the questions answered

Number of restaurants in every city:

```
> count <- dbGetQuery(pandadb,"
+             SELECT  city, count(vendor_id) as No_of_restaurants
+              FROM vendors
+              Group by city
+              ORDER BY count(vendor_id) desc; ")
>
>  count
          city No_of_restaurants
1    bangalore               1970
2        delhi               1953
3         pune               1638
4       mumbai               1363
5    hyderabad               1316
6      chennai                825
7      gurgaon                782
8        noida                679
9      Kolkata                575
10 chandigarh                303
11                             0
>
```

- Bangalore has the most number of restaurants

Restaurants having no reviews:

```
> noreviews <- dbGetQuery(pandadb,"
+            SELECT DISTINCT vendor_id, vendor_name, city
+            FROM vendors
+            WHERE review_count is null; ")
>
> noreviews
    vendor_id                                vendor_name       city
1           9                                    8 Slice  bangalore
2          10                        The Charcoal Factory  bangalore
3          11                              Box Of Biryani  bangalore
4          12                   Delly Belly (Koramangala)  bangalore
5          13                                     Kokomo  bangalore
6         766                                Patels Inn  bangalore
7        1136                          Sri Ganesh Darshan  bangalore
8        1189                               Yum Yum South  bangalore
9        1291          Curries & Pickles (A Narayanapura)  bangalore
10       1555                                   Sizzling  bangalore
11       1639                             Glens Bake House  bangalore
12       1677          Curries & Pickles (C.V Raman Nagar)  bangalore
13       1701                              Drunken Monkey  bangalore
14       1702                                Burrito Boys  bangalore
15       1707                                    Wow Momo  bangalore
16       1716                                Xpress Kitchen  bangalore
17       1717                               Nouvelle Garden  bangalore
18       1718            Arya Bhavan Sweets (Millers Road)  bangalore
19       1719                                   Bharanis  bangalore
20       1720                           Darbari Restaurant  bangalore
21       1721                                 Rasoi Ghar  bangalore
22       1723              Firangi Bake (Vinayaka Nagar)  bangalore
23       1724              BOX8- Desi Meals (Whitefield)  bangalore
24       1727             Curries & Pickles (Indiranagar)  bangalore
```

City with highest average rating:

```
> Avg <- dbGetQuery(pandadb,"
+             SELECT  city, AVG(rating) as Rating
+             FROM vendors
+             Group by city
+             ORDER BY AVG(rating) desc; ")
>  Avg
         city   Rating
1   hyderabad 3.931535
2     gurgaon 3.715090
3   bangalore 3.695736
4       noida 3.658910
5        pune 3.598718
6       delhi 3.594521
7  chandigarh 3.454455
8      mumbai 3.427293
9     chennai 3.333333
10    Kolkata 2.970435
11                 NA
> |
```

- Hyderabad tops the list for the highest average ratings

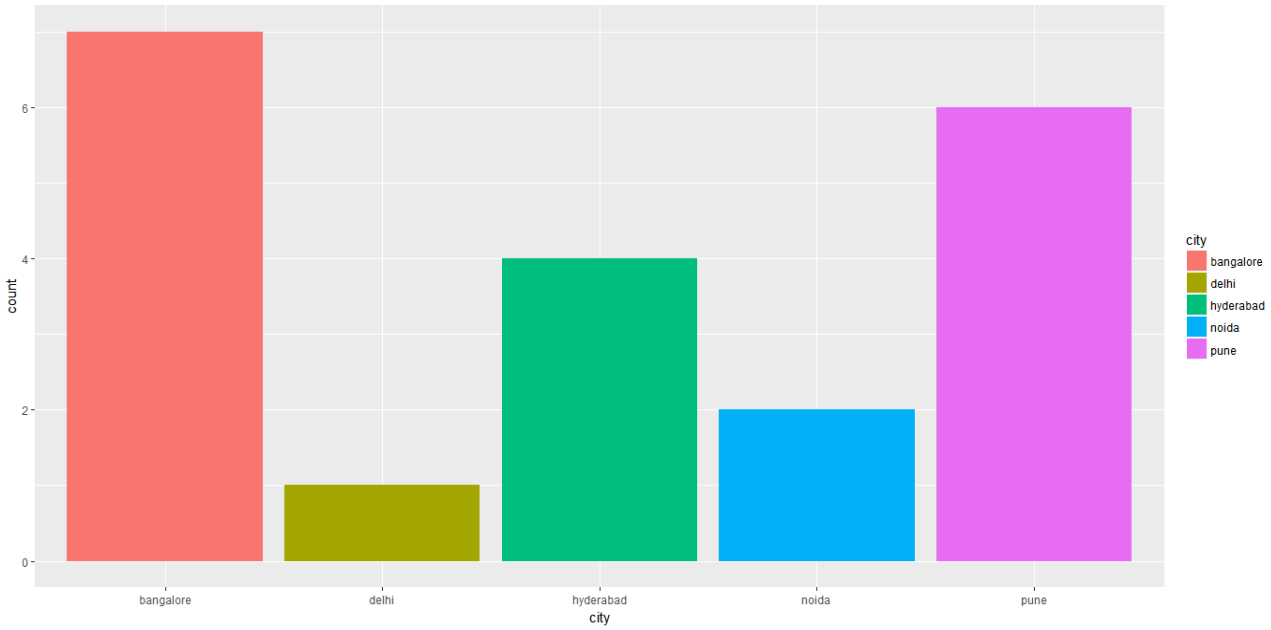
Restaurants currently offering discounts:

```
>  ggplot(data = discount) +
+    geom_bar(mapping = aes(x = city, y = 'No_of_restaurants'), stat = "identity", col=c("brown"))
> discount <- dbGetQuery(pandadb,"
+           SELECT DISTINCT vendor_id, vendor_name, city, discount
+           FROM vendors
+           WHERE discount >0
+           ORDER BY DISCOUNT DESC; ")
> discount
    vendor_id                                          vendor_name       city
1        8454                                             Overload     mumbai
2        3269                                Burger Singh (Rajouri Garden)      delhi
3        4890                                    Burger Singh (Hari Nagar)      delhi
4       10203                                     Burger Singh (Sector 62)      noida
5        5147                                                Big Jack's    gurgaon
6        2532                                    Behrouz Biryani (Kellys)    chennai
7        2566                                Behrouz Biryani (Poonamalle)    chennai
8        2864                                 Behrouz Biryani (Sembakkam)    chennai
9        2876                                   Behrouz Biryani (Navalur)    chennai
10       2882                                       Behrouz Biryani (OMR)    chennai
11       2896                                    Behrouz Biryani (Porur)    chennai
12       2913                               Behrouz Biryani (West Mambalam)    chennai
13       2916                                Behrouz Biryani (Besant Nagar)    chennai
14       2947                                    Behrouz Biryani (Chrompet)    chennai
15       2949                          Behrouz Biryani (Thillai Ganga Nagar)    chennai
16       2951                                Behrouz Biryani (Nungambakkam)    chennai
17       2955                                   Behrouz Biryani (Mogappair)    chennai
18       2995                                   Behrouz Biryani (Velachery)    chennai
19       3415                                    Behrouz Biryani (RK Puram)      delhi
20       3456                                 Behrouz Biryani (Laxmi Nagar)      delhi
21       3478                                 Behrouz Biryani (Patel Nagar)      delhi
22       3479                                     Behrouz Biryani (Kalkaji)      delhi
23       3500                                   Behrouz Biryani (Sector 5)      delhi
24       3503                                  Behrouz Biryani (Janakpuri)      delhi
25       3515                               Behrouz Biryani (Malviya Nagar)      delhi
26       3546                                  Behrouz Biryani (Green Park)      delhi
```

Top 20 restaurants having highest number of reviews:

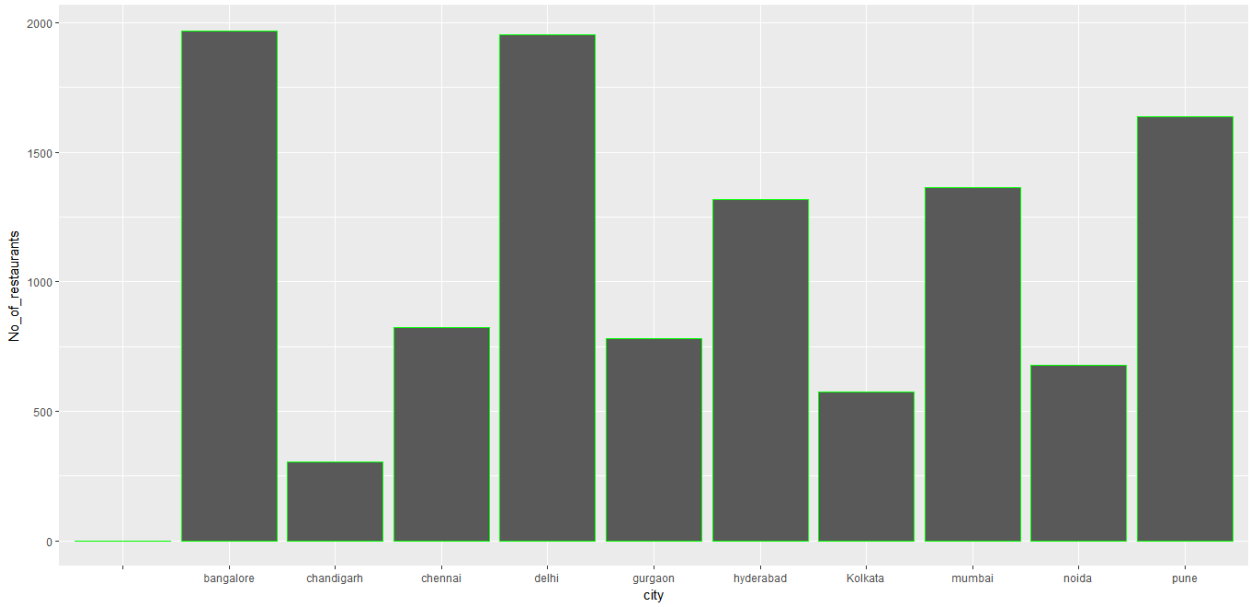
```
> #3.Top 20     Restaurants having highest number of reviews
>
>
> Resthigh <- dbGetQuery(pandadb,"
+             SELECT DISTINCT vendor_id, vendor_name, city, review_count
+             FROM vendors
+             WHERE review_count is not null
+             order by review_count desc limit 20; ")
> Resthigh
   vendor_id                   vendor_name      city review_count
1       1829                      Domino's bangalore        29921
2       3107                      Domino's     delhi        22577
3       9885         Chick Chicken Barbeque     noida        17892
4      10490         Halka Fulka Restaurant      pune        16415
5      10506 Chaitanya Paranthas (Kharadi)      pune        16365
6      10538     Eatsome (Pimple Saudagar)      pune        12653
7         77     Rangla Punjab (Bellandur) bangalore        12580
8       5910     Rangla Punjab (Bellandur) hyderabad        12580
9         63                   Biryani Day bangalore        12354
10      5896                   Biryani Day hyderabad        12354
11      1824                         Moriz bangalore        12267
12     10535   Chaitanya Paratha (Kothrud)      pune        12168
13     10499               Eatsome (Wakad)      pune        12091
14      9849           Vinayak Restaurant     noida        11977
15        18              Jaffa's Biryani bangalore        11844
16      5851              Jaffa's Biryani hyderabad        11844
17     10487   Preets Punjabi Swad (Wakad)      pune        11243
18        75       New Punjabi Tasty Khana bangalore        10972
19      5908       New Punjabi Tasty Khana hyderabad        10972
20        80 Imperio Restaurant (Kadugodi) bangalore        10893
>
```

City with most number of top 20 restaurants.



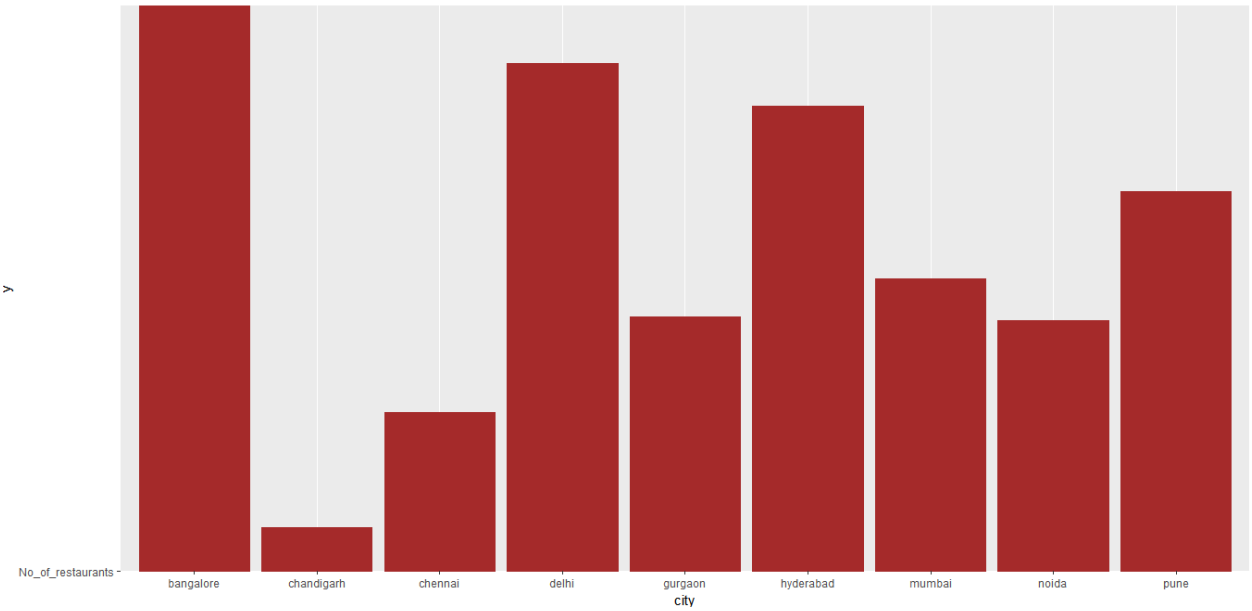
- Bangalore has the most top 20 restaurants

Number of restaurants in every city:



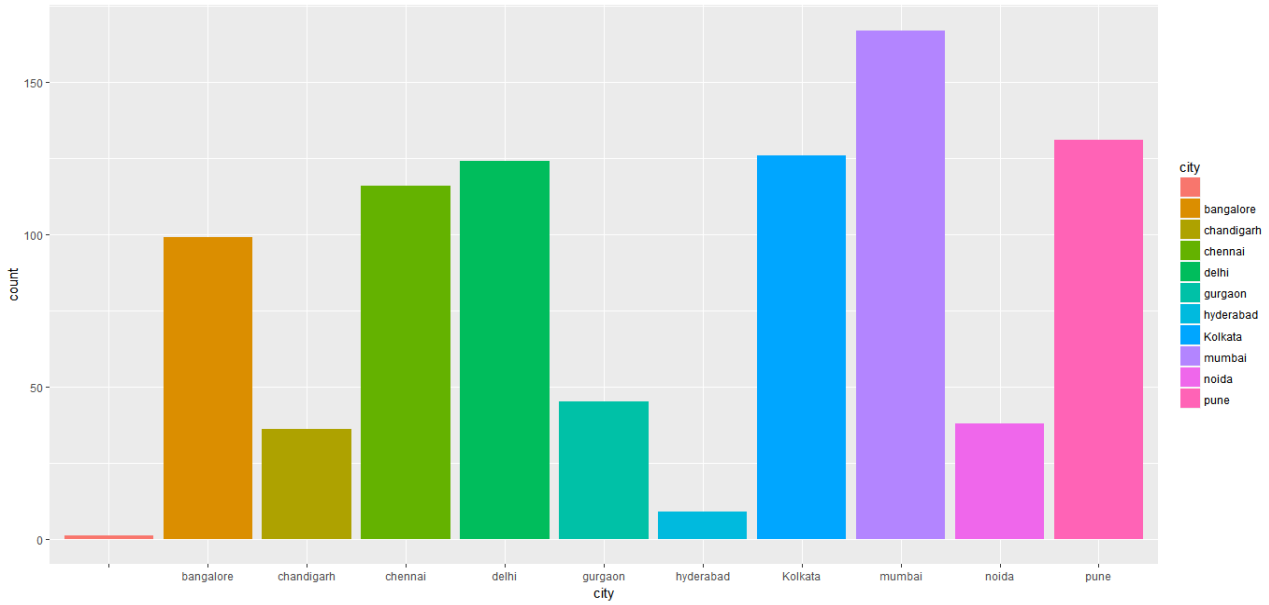
- Bangalore and Delhi has most number of restaurants, Bangalore edges Delhi.

City having most restaurants offering discounts.



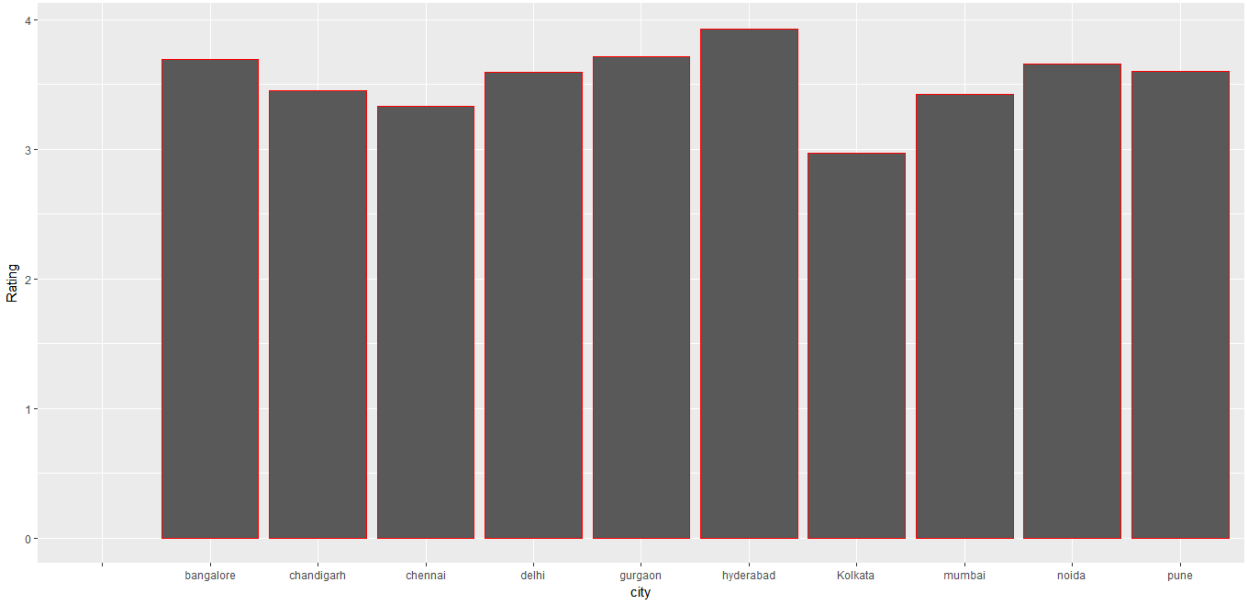
- Bangalore has the most number of restaurants providing discounts.

City with highest number of No reviews:



- Mumbai has the most number of restaurants with no reviews.

City with highest average reviews:



- Hyderabad has the highest average rating among the cities.

CONCLUSION:

- We understood to scrape data from websites using open source tools. We cleaned data to make it more readable and also to retrieve data. We learnt the usage of R packages like dplyr, ggplot, SQLite and performed analysis on the collected data. We obtained bar graphs and designed a SQLite database for storage.
- We analysed Foodpanda to generate insights that can improve the businesses for the company.
- We concluded that among Indian cities, Bangalore has the biggest market for Foodpanda in India. Less number of reviews in Mumbai shows below average consumer satisfaction. The highest average ratings in Hyderabad shows that the city has the best restaurants in India.

REFERENCES:

- www.Foodpanda.in/
- www.stackoverflow.com
- R for Data Science by Hardly Wickham