Northeastern University
College *of* Engineering

# Project:

# Data Analysis of Trip Advisor Website

Collect, Store, Retrieve Data

Data Analytics DA5020

## Master of Science in Data Analytics

**Hemanth Lakshman Raju**

Under the guidance of Prof. Kathleen Durant

## Introduction:

Travelling is an integral part of life, especially for people who have travelling as a passion. Not only they provide the experience, but they also give a quick view on the culture and the food they consume. Travelling also provide a break from the rat race that everyone runs. The most important player in this field is tripadvisor.com - the website which helps you make plan from the flight to the restaurant that you want to stay in.

Tripadvisor.com is one of the most reliable website that provides services like finding restaurants, booking hotels and flight booking. The website gives us suggestions while we search for the hotels, restaurants and flights which have the best deals in each of the service. It is the largest travel site in the world, with more than 315 million members and over 500 million reviews and opinions of hotels, restaurants, attractions and other travel-related businesses. The website services are free to users, who provide most of the content, and the website is supported by a hotel booking facility and an advertising business model.

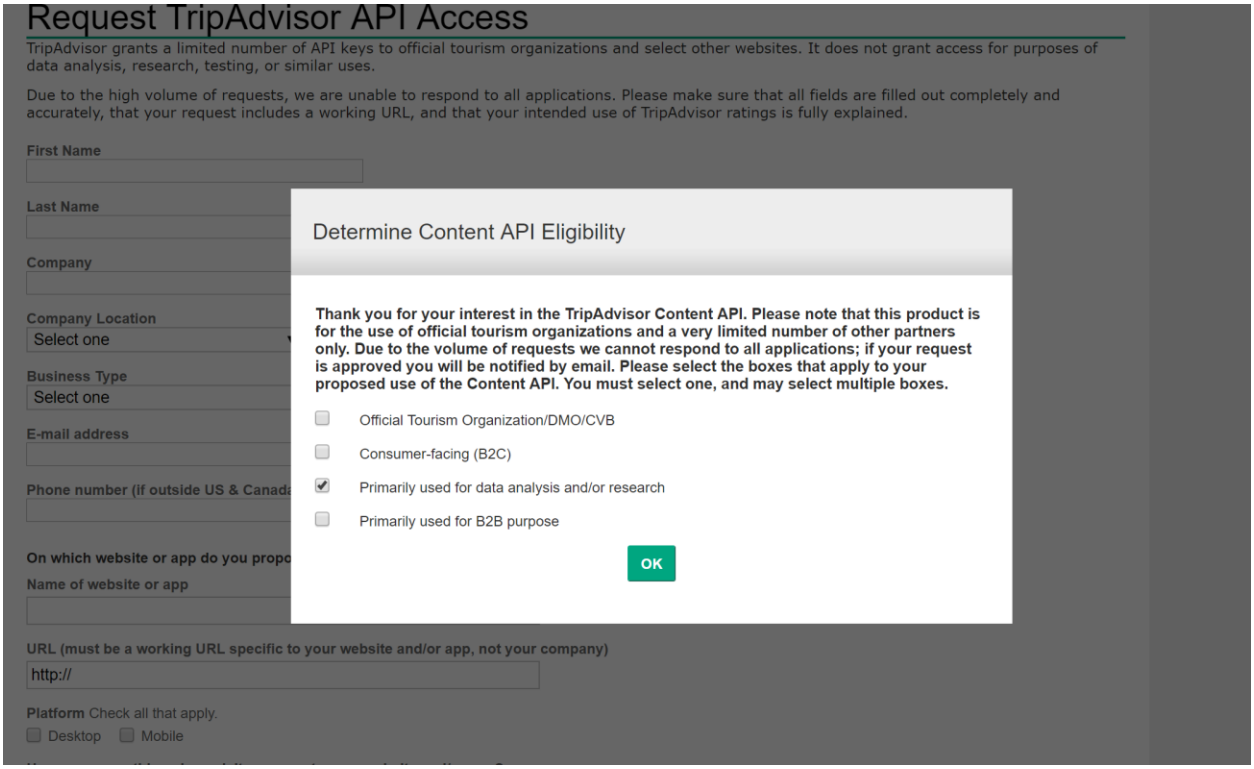
**The idea of the project is to create a bird's eye view of a locality showing Hotels, Restaurants and Tourist Spots in the locality. Trip advisor does not have a consolidated view of all the services it offers. We plan to achieve that with this project. For this Project, we have selected one of the most popular tourist spots, Paris.**

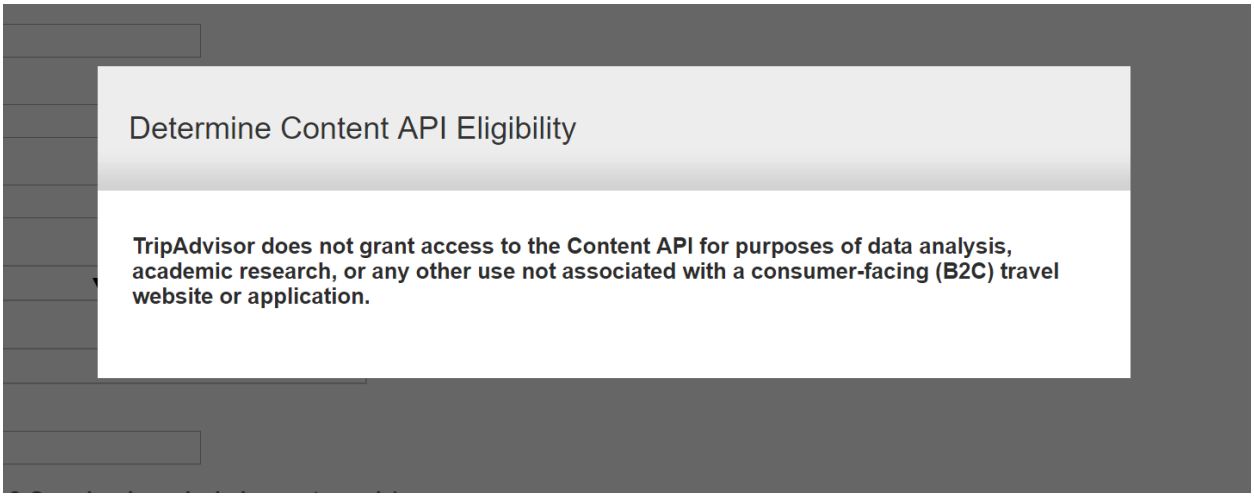
The objective of the proposed project is to scrape data from tripadvisor.com about the restaurants, hotels and popular tourist spots and visualize the data using R packages such as ggmap() and ggplot(). For this project we will consider one particular city and perform Analysis on the data which we get from TripAdvisor. The database will be created in SQLite.

Web Scraping will be done on the TripAdvisor Website to gather the information. After the data scraping we perform multiple analysis of data in various forms such as sentimental analysis and analysis based on location of the best services provided. These analyses are done to ease the selection of restaurants, hotels and tourist spots for the tourists.

## Step 1: Data Collection

We planned on using the TripAdvisor API for our project but when we tried to sign-up as a developer, we were faced with this issue.



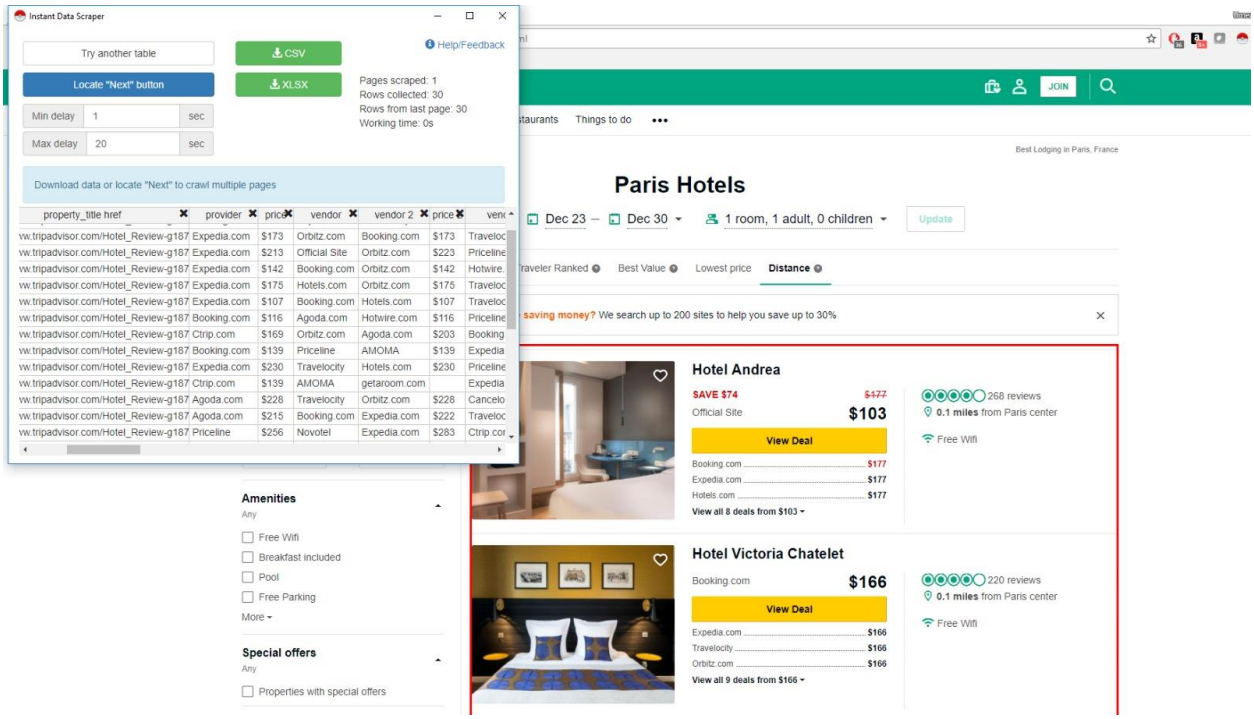




So, after discussing the issue with the professor, we decided to scrape the data using Google Chrome Extensions namely **'Instant Data Scraper'** and **'Web Scraper'**

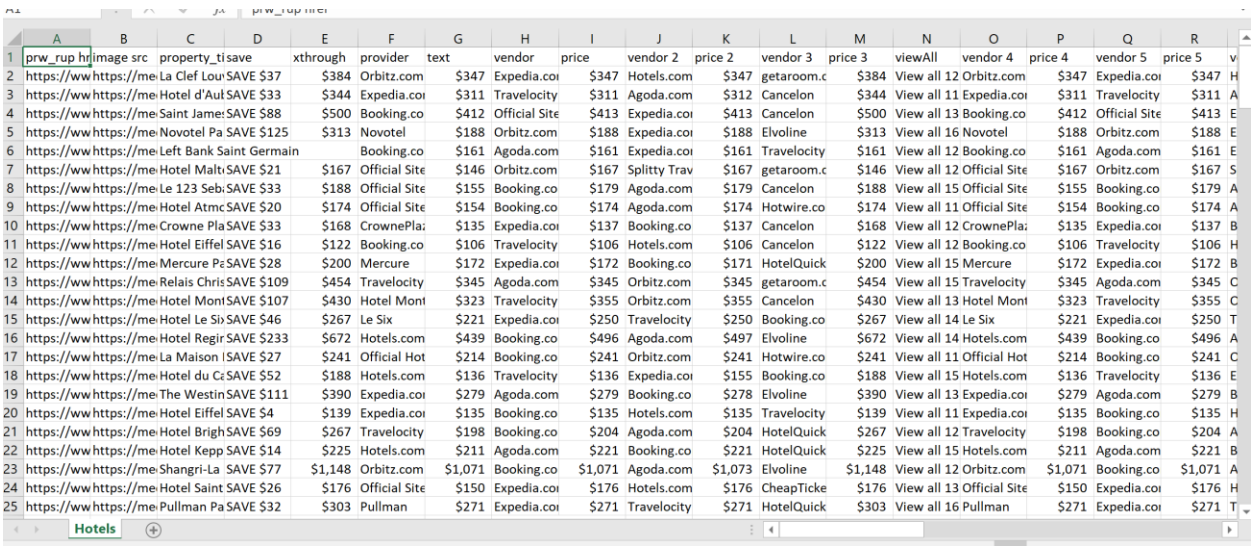
Collection of data from the tripadvisor.com website on the **Hotels** in Paris. The data collected from the website are,

- Hotel Name
- Hotel web link
- Cheapest provider
- Price
- Reviews
- Review count
- Distance
- Amenities

Screenshot showing the data scraping




Scraped Data



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | prw_rup hr | image src | property_ti | save | xthrough | provider | text | vendor | price | vendor 2 | price 2 | vendor 3 | price 3 | viewAll | vendor 4 | price 4 | vendor 5 | price 5 |
| 2 | https://ww | https://me | La Clef Lou | SAVE $37 | $384 | Orbitz.com | $347 | Expedia.co | $347 | Hotels.com | $347 | getaroom.c | $384 | View all 12 | Orbitz.com | $347 | Expedia.co | $347 |
| 3 | https://ww | https://me | Hotel d'Aul | SAVE $33 | $344 | Expedia.co | $311 | Travelocity | $311 | Agoda.com | $312 | Cancelon | $344 | View all 11 | Expedia.co | $311 | Travelocity | $311 |
| 4 | https://ww | https://me | Saint James | SAVE $88 | $500 | Booking.co | $412 | Official Site | $413 | Expedia.co | $413 | Cancelon | $500 | View all 13 | Booking.co | $412 | Official Site | $413 |
| 5 | https://ww | https://me | Novotel Pa | SAVE $125 | $313 | Novotel | $188 | Orbitz.com | $188 | Expedia.co | $188 | Elvoline | $313 | View all 16 | Novotel | $188 | Orbitz.com | $188 |
| 6 | https://ww | https://me | Left Bank Saint Germain | | | Booking.co | $161 | Agoda.com | $161 | Expedia.co | $161 | Travelocity | $161 | View all 12 | Booking.co | $161 | Agoda.com | $161 |
| 7 | https://ww | https://me | Hotel Malt | SAVE $21 | $167 | Official Site | $146 | Orbitz.com | $167 | Splitty Trav | $167 | getaroom.c | $146 | View all 12 | Official Site | $167 | Orbitz.com | $167 |
| 8 | https://ww | https://me | Le 123 Seb | SAVE $33 | $188 | Official Site | $155 | Booking.co | $179 | Agoda.com | $179 | Cancelon | $188 | View all 15 | Official Site | $155 | Booking.co | $179 |
| 9 | https://ww | https://me | Hotel Atmc | SAVE $20 | $174 | Official Site | $154 | Booking.co | $174 | Agoda.com | $174 | Hotwire.co | $174 | View all 11 | Official Site | $154 | Booking.co | $174 |
| 10 | https://ww | https://me | Crowne Pla | SAVE $33 | $168 | CrownePla | $135 | Expedia.co | $137 | Booking.co | $137 | Cancelon | $168 | View all 12 | CrownePla | $135 | Expedia.co | $137 |
| 11 | https://ww | https://me | Hotel Eiffel | SAVE $16 | $122 | Booking.co | $106 | Travelocity | $106 | Hotels.com | $106 | Cancelon | $122 | View all 12 | Booking.co | $106 | Travelocity | $106 |
| 12 | https://ww | https://me | Mercure Pa | SAVE $28 | $200 | Mercure | $172 | Expedia.co | $172 | Booking.co | $171 | HotelQuick | $200 | View all 15 | Mercure | $172 | Expedia.co | $172 |
| 13 | https://ww | https://me | Relais Chris | SAVE $109 | $454 | Travelocity | $345 | Agoda.com | $345 | Orbitz.com | $345 | getaroom.c | $454 | View all 15 | Travelocity | $345 | Agoda.com | $345 |
| 14 | https://ww | https://me | Hotel Mont | SAVE $107 | $430 | Hotel Mont | $323 | Travelocity | $355 | Orbitz.com | $355 | Cancelon | $430 | View all 13 | Hotel Mont | $323 | Travelocity | $355 |
| 15 | https://ww | https://me | Hotel Le Si | SAVE $46 | $267 | Le Six | $221 | Expedia.co | $250 | Travelocity | $250 | Booking.co | $267 | View all 14 | Le Six | $221 | Expedia.co | $250 |
| 16 | https://ww | https://me | Hotel Regir | SAVE $233 | $672 | Hotels.com | $439 | Booking.co | $496 | Agoda.com | $497 | Elvoline | $672 | View all 14 | Hotels.com | $439 | Booking.co | $496 |
| 17 | https://ww | https://me | La Maison I | SAVE $27 | $241 | Official Hot | $214 | Booking.co | $241 | Orbitz.com | $241 | Hotwire.co | $241 | View all 11 | Official Hot | $214 | Booking.co | $241 |
| 18 | https://ww | https://me | Hotel du Ca | SAVE $52 | $188 | Hotels.com | $136 | Travelocity | $136 | Expedia.co | $155 | Booking.co | $188 | View all 15 | Hotels.com | $136 | Travelocity | $136 |
| 19 | https://ww | https://me | The Westin | SAVE $111 | $390 | Expedia.co | $279 | Agoda.com | $279 | Booking.co | $278 | Elvoline | $390 | View all 13 | Expedia.co | $279 | Agoda.com | $279 |
| 20 | https://ww | https://me | Hotel Eiffel | SAVE $4 | $139 | Expedia.co | $135 | Booking.co | $135 | Hotels.com | $135 | Travelocity | $139 | View all 11 | Expedia.co | $135 | Booking.co | $135 |
| 21 | https://ww | https://me | Hotel Brigh | SAVE $69 | $267 | Travelocity | $198 | Booking.co | $204 | Agoda.com | $204 | HotelQuick | $267 | View all 12 | Travelocity | $198 | Booking.co | $204 |
| 22 | https://ww | https://me | Hotel Kepp | SAVE $14 | $225 | Hotels.com | $211 | Agoda.com | $221 | Booking.co | $221 | HotelQuick | $225 | View all 15 | Hotels.com | $211 | Agoda.com | $221 |
| 23 | https://ww | https://me | Shangri-La | SAVE $77 | $1,148 | Orbitz.com | $1,071 | Booking.co | $1,071 | Agoda.com | $1,073 | Elvoline | $1,148 | View all 12 | Orbitz.com | $1,071 | Booking.co | $1,071 |
| 24 | https://ww | https://me | Hotel Saint | SAVE $26 | $176 | Official Site | $150 | Expedia.co | $176 | Hotels.com | $176 | CheapTicke | $176 | View all 13 | Official Site | $150 | Expedia.co | $176 |
| 25 | https://ww | https://me | Pullman Pa | SAVE $32 | $303 | Pullman | $271 | Expedia.co | $271 | Travelocity | $271 | HotelQuick | $303 | View all 16 | Pullman | $271 | Expedia.co | $271 |

Hotels

However, these data scrapers have their limitations. It was unable to scrape the address of the hotels because it was not readily available on the front page. (seen in screenshot)

We had to get the link of each Hotel and fetch the address data from that hyperlink. We wrote the code in R to do that.

We got the web link from the scraped data. Using that web link for each hotel, we ran a 'for' loop to fetch the address of the Hotels using html nodes.

Here is the code.

```{r}
hotel$Hotel_Web_Link <- as.character(hotel$Hotel_Web_Link)
class(hotel$Hotel_Web_Link)
hotel$Hotel_address <- rep(NA,nrow(hotel))
for(i in 1:nrow(hotel)){
  htm <- read_html(hotel$Hotel_Web_Link[[i]])
  add1 <- htm %>% html_node(css="span.street-address") %>% html_text()
  add2 <- htm %>% html_node(css="span.extended-address") %>% html_text()
  add3 <- htm %>% html_node(css="span.locality") %>% html_text()
  add4 <- htm %>% html_node(css="span.country-name") %>% html_text()
  hotel$Hotel_address[[i]] <- paste(add1,add2,add3,add4,collapse=',')
}

print(hotel)
```

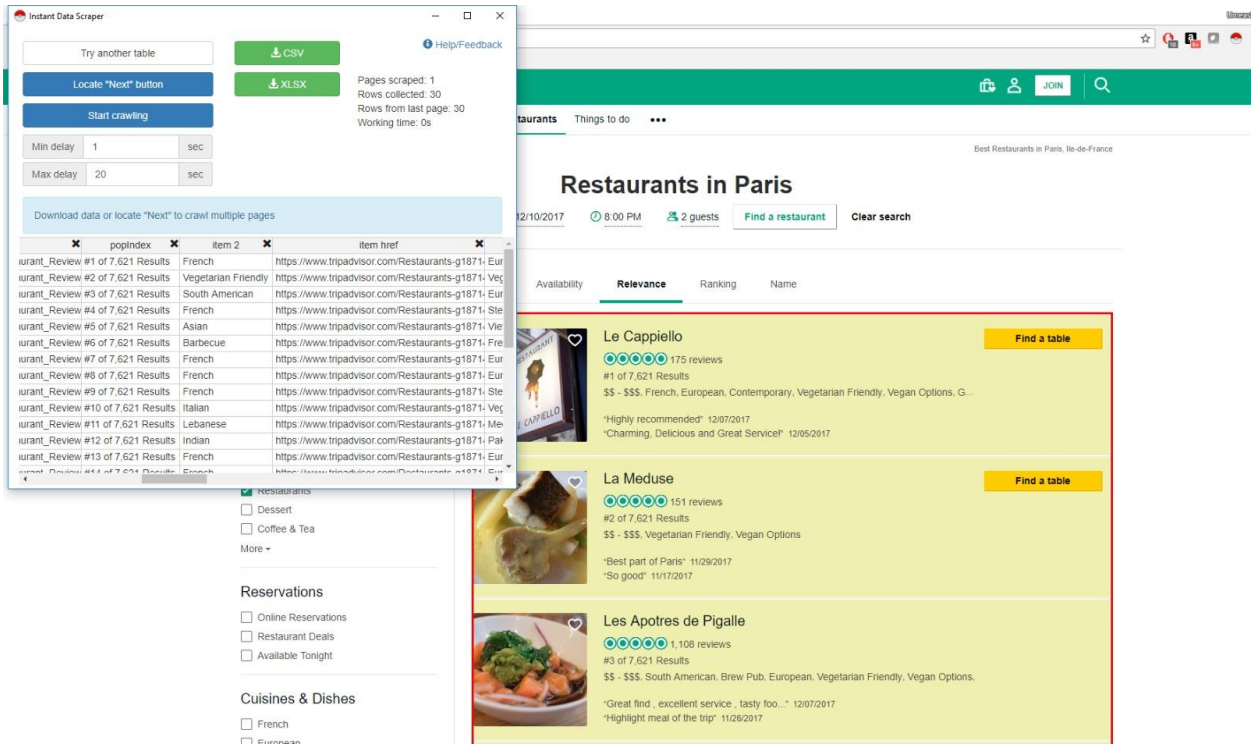
Filter

| | Cheapest_Provider | Price | Review_Count | Distance | Amenities_1 | Amenities_2 | Hotel_address | hotel_id |
|---|---|---|---|---|---|---|---|---|
| el_Review-g187147-d472... | Expedia.com | 146 | 205 | 0.2 | Free Wifi | Room Service | 1 Place du Parvis Notre Dame 04 Arr. 75004 Paris, France | 472020 |
| el_Review-g187147-d302... | Official Site | 161 | 934 | 0.3 | Free Wifi | Bar/Lounge | 1 quai Saint-Michel 5th Arr. 75005 Paris, France | 302993 |
| el_Review-g187147-d219... | Orbitz.com | 145 | 358 | 0.4 | Free Wifi | | 65 rue Saint Louis en L Ile 75004 Paris, France | 219994 |
| el_Review-g187147-d319... | Booking.com | 201 | 1,230 | 0.4 | Free Wifi | Breakfast included | 18 rue de la Harpe 75005 Paris, France | 319955 |
| el_Review-g187147-d194... | Agoda.com | 151 | 210 | 0.4 | Free Wifi | Room Service | 59 rue Saint Louis en L Ile 75004 Paris, France | 194270 |
| el_Review-g187147-d194... | Expedia.com | 225 | 484 | 0.5 | Free Wifi | Bar/Lounge | 54 rue Saint Louis en L Ile 75004 Paris, France | 194304 |
| el_Review-g187147-d774... | Booking.com | 291 | 263 | 0.5 | Free Wifi | Pool | 7 rue du Bourg L Abbe 75003 Paris, France | 7740508 |
| el_Review-g187147-d568... | Official Site | 116 | 257 | 0.5 | Free Wifi | | 2 rue Malher 75004 Paris, France | 568921 |
| el_Review-g187147-d506... | Booking.com | 97 | 178 | 0.6 | Free Wifi | | 39 rue de Turbigo 03 Arr. 75003 Paris, France | 506359 |
| el_Review-g187147-d243... | Travelocity | 199 | 367 | 0.6 | Free Wifi | Room Service | 11 rue des Gravilliers between Place des Vosges and Cent... | 2432527 |
| el_Review-g187147-d506... | Booking.com | 75 | 649 | 0.6 | Free Wifi | | 6 rue Greneta 75003 Paris, France | 506893 |
| el_Review-g187147-d565... | Booking.com | 89 | 264 | 0.6 | Free Wifi | | 9 rue d Ormesson 75004 Paris, France | 565827 |
| el_Review-g187147-d218... | Booking.com | 174 | 51 | 0.6 | Free Wifi | | 7 rue des Vertus 75003 Paris, France | 2186989 |
| el_Review-g187147-d124... | Hotels.com | 226 | 61 | 0.6 | | | 243 rue saint Martin 75003 Paris, France | 12456055 |
| el_Review-g187147-d548... | Orbitz.com | 228 | 453 | 0.7 | Free Wifi | Room Service | 29 rue de Poitou 75003 Paris, France | 548458 |
| el_Review-g187147-d248... | Booking.com | 161 | 186 | 0.7 | Free Wifi | | 16 rue de Saintonge 75003 Paris, France | 248398 |
| el_Review-g187147-d399... | Orbitz.com | 118 | 156 | 0.7 | Free Wifi | Restaurant | 31 rue d Alexandrie 75002 Paris, France | 3998722 |
| el_Review-g187147-d275... | Hotels.com | 131 | 130 | 0.7 | Free Wifi | Room Service | 94 rue des Archives 75003 Paris, France | 275021 |
| el_Review-g187147-d219... | Hotwire.com | 156 | 245 | 0.7 | Free Wifi | Bar/Lounge | 30 rue de Turenne 75003 Paris, France | 219989 |
| el_Review-g187147-d296... | Orbitz.com | 126 | 377 | 0.7 | Free Wifi | | 6 rue Montgolfier 3rd Arr. 75003 Paris, France | 296785 |
| el_Review-g187147-d103... | Official Site | 157 | 225 | 0.7 | Free Wifi | Room Service | 87 rue des Archives 75003 Paris, France | 10319445 |
| el_Review-g187147-d228... | Agoda.com | 134 | 248 | 0.7 | Free Wifi | Bar/Lounge | 4 rue Salomon de Caus 75003 Paris, France | 228719 |
| el_Review-g187147-d188... | Expedia.com | 470 | 832 | 0.8 | Free Wifi | Room Service | 28 Place des Vosges 75003 Paris, France | 188738 |
| el_Review-g187147-d620... | Booking.com | 73 | 50 | 0.8 | | | 183 rue du Temple 75003 Paris, France | 620131 |
| el_Review-g187147-d616... | Expedia.com | 67 | 136 | 0.8 | Free Wifi | | 26 rue de Picardie 3rd Arr. 75003 Paris, France | 616466 |
| el_Review-g187147-d276... | Expedia.com | 102 | 236 | 0.8 | Free Wifi | | 76 rue de Turbigo 75003 Paris, France | 276938 |
| el_Review-g187147-d248... | Hotels.com | 74 | 403 | 0.8 | Free Wifi | | 69 rue Meslay 3rd Arr. 75003 Paris, France | 248396 |
| el_Review-g187147-d233... | Official Site | 92 | 291 | 0.9 | Free Wifi | Room Service | 3 rue Meslay 75003 Paris, France | 233760 |
| el_Review-g187147-d233... | Expedia.com | 91 | 117 | 0.9 | Free Wifi | | 2 B boulevard Saint Martin 75010 Paris, France | 233761 |

Collection of data from the tripadvisor.com website on **Restaurants** in Paris. The data collected from the website are,

- Restaurant name
- Restaurant web link
- Price range
- Cuisines offered
- Reviews
- Review date
-

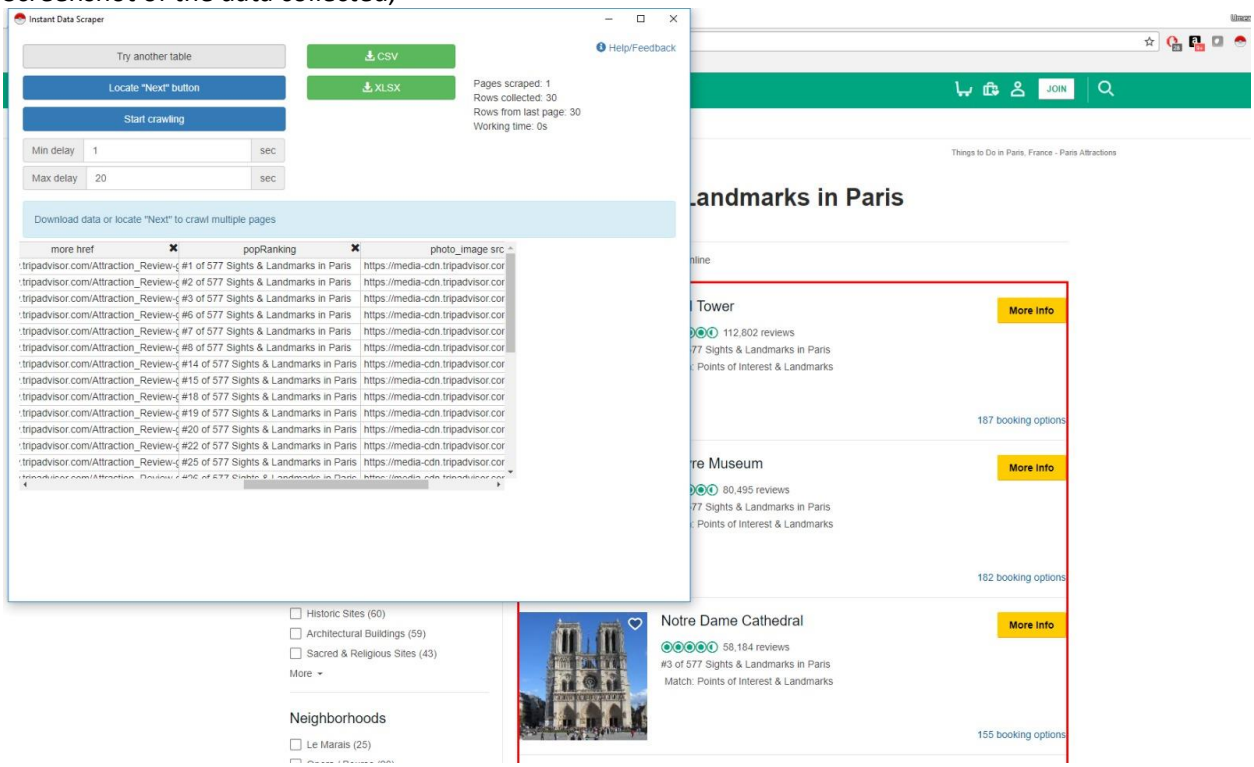
Screenshot of the data collected,



Collection of data from the tripadvisor.com website on **Tourist Spots** in Paris. The data collected from the website are,

- Name
- Web link
- Number of reviews
- Ranking
- Number of booking options

Screenshot of the data collected,




## Step 2: Cleaning the Data

We got a lot of irrelevant and unnecessary data which was not needed for our Analysis. Hence we cleaned up the data as per our requirements.

- We removed the unwanted columns.
- We changed the Column Names
- Combined the cuisines offered into one column.
- Retrieved ID and Address using the Web Link.
- Removed unnecessary commas, words and Null Values from the dataset

## Cleaned Data – Restaurants

Filter

| ırant_Name | Restaurant_Web_Link | Review_Count | Price.Range | Cuisine | Review_1 |
|---|---|---|---|---|---|
| eduse | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 148 | $$ - $$$ | Vegetarian Friendly,Vegan Options | Best part of Paris |
| appiello | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 167 | $$ - $$$ | French,European,Vegetarian Friendly,Vegan Options,Glut... | Absolutely Amazi |
| ıpotres de Pigalle | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 1,098 | $$ - $$$ | South American,European,Vegetarian Friendly,Vegan Opt... | Highlight meal of |
| it Un Square | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 1,547 | $$ - $$$ | French,Steakhouse,Vegetarian Friendly | Great Burgers |
| strot d'Indochine | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 301 | $$ - $$$ | Asian,Vietnamese,Vegetarian Friendly,Vegan Options | Just fantastic! |
| Restaurant et Canal Saint Martin | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 176 | $$ - $$$ | Barbecue,French,Steakhouse,European,Gluten Free Options | Perfect Steak, Sid |
| neur' Affine | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 738 | $$ - $$$ | French,European,Vegetarian Friendly | AMAZING CHEES |
| strot d'Yves | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 270 | $$ - $$$ | French,European,Vegetarian Friendly | Popular neighbor |
| talia Brasserie | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 898 | $$ - $$$ | Italian,Vegetarian Friendly,Vegan Options | Glad we found th |
| e fleur | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 2,410 | $$ - $$$ | French,Steakhouse,European | NOT TO BE MISSE |
| ram | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 528 | $$ - $$$ | Lebanese,Mediterranean,Middle Eastern,Vegetarian Frien... | Best Lebanese fo |
| Jawad Longchamp | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 235 | $$ - $$$ | Indian,Pakistani,Vegetarian Friendly,Vegan Options | Awesome is the w |
| ıagerie Danard | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 548 | $$ - $$$ | French,European,Vegetarian Friendly | Dinner at Paris |
| Picnic | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 451 | $$ - $$$ | French,European,Street Food,Vegetarian Friendly,Gluten F... | Perfect for a large |
| rante Lo Spaghettino | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 449 | $$ - $$$ | Italian,Mediterranean,European,Vegetarian Friendly,Vega... | Very good |
| ssue | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 244 | $$ - $$$ | French,Cafe,European,Vegetarian Friendly,Vegan Options | Nice environmen |
| ie du jour | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 359 | $$ - $$$ | French,European,Vegetarian Friendly,Vegan Options | Totally enjoyable |
| | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 243 | $$ - $$$ | French,European,Vegetarian Friendly,Vegan Options,Glut... | Excellent dinner i |
| ouille | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 282 | $$ - $$$ | French,European | Great food / Orig |
| | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 206 | $$ - $$$ | Italian,French,Pizza,European,Vegetarian Friendly | Pavilla |
| attoria Dell'isola | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 447 | $$ - $$$ | Italian,Pizza,Seafood,Vegetarian Friendly | Nice place |
| on | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 928 | $$ - $$$ | French,European | Dinner |
| oite aux lettres | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 487 | $$ - $$$ | French,European,Vegetarian Friendly,Vegan Options | Excellent food an |
| Imogene | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 400 | $$ - $$$ | French,Vegetarian Friendly,Vegan Options,Gluten Free Op... | Wonderful! |
| olio e pomodoro | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 295 | $$ - $$$ | Italian,Mediterranean,European,Vegetarian Friendly,Vega... | delicious food,frie |
| York a Paris | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 461 | $$ - $$$ | French,American,European,Vegetarian Friendly,Gluten Fre... | Would defiantly r |
| cini | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 493 | $$ - $$$ | Italian,Pizza,European,Vegetarian Friendly | Cute, small, Italia |
| tus | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 458 | $$ - $$$ | French,European,Vegetarian Friendly,Gluten Free Options | Delicious |
| ıche et Le Cuisinier | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 751 | $$ - $$$ | French,European,Vegetarian Friendly | My kind of French |
| canailles | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 480 | $$ - $$$ | French,European | Fantastic place |
| pacionu | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 419 | $$ - $$$ | Italian,Pizza,Vegetarian Friendly,Vegan Options | Very popular pizz |
| cora Cebicheria | https://www.tripadvisor.com/Restaurant_Review-g187147-... | 316 | $$ - $$$ | Peruvian,Seafood,South American,Vegetarian Friendly,Ve... | Excellent food |

We performed the same steps for Hotels and Tourist spots.

## Step 3: Storing the Data

For storing the data, we used the RSQLite package to use SQLite as our database.

We created a database called **TripAdvisorDB**.

After the collection of data from the website, we created three Tables namely

- **RestaurantTable**
- **HotelTable**
- **POI_Table**

While creating the tables we needed a primary key for each Table, so, we used the weblink to get a unique ID for each entity and assigned it as the primary key for each table.

```{r}
db<- dbConnect(SQLite(), dbname = 'TripAdvisorDB')
```

```{r}

str(restaurant)
dbSendQuery(conn = db,  "CREATE TABLE RestaurantTable (
            Restaurant_Name TEXT,
            Restaurant_Web_Link TEXT,
            Review_Count TEXT,
            `Price.Range` TEXT,
            Cuisine TEXT,
            Review_1 TEXT,
            Review_1_date TEXT,
            Review_2 TEXT,
            Review_2_date TEXT,
            Category TEXT,
            Restaurant_address TEXT,
            Restaurant_ID numeric,
            PRIMARY KEY(Restaurant_ID)
            )
            WITHOUT ROWID")
dbWriteTable(conn = db, name = "restaurant", value = restaurant, row.names=FALSE, append = TRUE)

```

```{r}
str(hotel)
dbSendQuery(conn = db,  "CREATE TABLE HotelTable (
            Hotel_Name TEXT,
            Hotel_Web_Link text,
            Cheapest_Provider TEXT,
            Price num,
            Review_Count text,
            Distance num,
            Amenities_1 TEXT,
            Amenities_2 TEXT,
            Hotel_address TEXT,
            hotel_id num,
            PRIMARY KEY(hotel_id)
            )
            WITHOUT ROWID")

dbWriteTable(conn = db, name = "hotel", value = hotel, row.names=FALSE, append = TRUE)
```

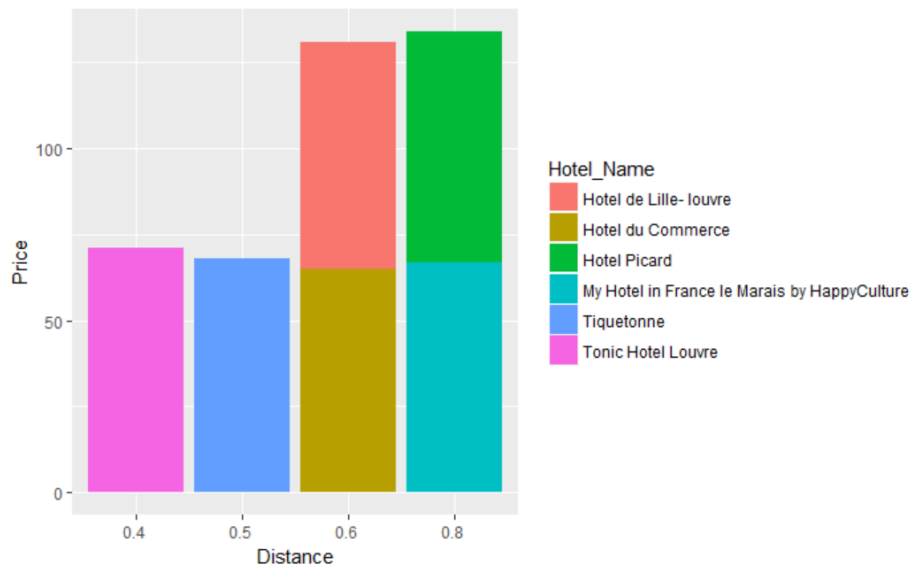
59:1 (Top Level) R Markdown

## Step 4: Data Analysis

We did various types of analysis with the data that we collected.

1. Sentimental Analysis of the reviews.
2. Analysis of Hotels – How Price varies with distance from Paris center.
3. Highest Reviews, Highest bookings, Cheapest Hotels etc
4. Plotting location of hotel, restaurant and tourist spots on a map.
5. A bird's eye view of the location showing the hotels, restaurants and tourist spots in the area for ease of planning for the tourist.

```{r}
hotel$Price = hotel$Price%>% as.integer()
pricedata <- hotel[order(hotel$Price),]
bb<-head(pricedata)
ggplot(bb,aes(x=Distance,y=Price))+geom_bar(stat="identity",aes(fill=Hotel_Name))
```




Price in dollars vs Distance in miles

Hotels Ranked according to highest reviews.

```{r}
hotel$Review_Count = hotel$Review_Count%>% as.numeric()
rwdata <- hotel[order(hotel$Review_Count, decreasing=TRUE),]
print(rwdata)
```

| Cheapest_Provider <fctr> | Price <int> | Review_Count <dbl> | Distance <chr> | Amenities_1 <fctr> | Amenities_2 <fctr> |
|---|---|---|---|---|---|
| Agoda.com | 177 | 967 | 0.7 | Free Wifi | Room Service |
| Booking.com | 106 | 964 | 0.5 | Free Wifi | |
| Travelocity | 108 | 959 | 0.6 | Free Wifi | |
| Melia.com | 155 | 945 | 0.4 | Free Wifi | Room Service |
| Official Site | 161 | 934 | 0.3 | Free Wifi | Bar/Lounge |
| Agoda.com | 336 | 900 | 0.5 | Free Wifi | Room Service |
| Official Site | 112 | 900 | 0.5 | Free Wifi | Bar/Lounge |
| Booking.com | 107 | 888 | 0.6 | Free Wifi | Room Service |
| Agoda.com | 146 | 865 | 0.6 | Free Wifi | |
| Booking.com | 135 | 852 | 0.5 | Free Wifi | Bar/Lounge |

1-10 of 180 rows | 4-9 of 10 columns

Previous 1 2 3 4 5 6 ... 18 Next

We performed sentimental analysis on the reviews by the travelers for each of the three entities.

For example, here is the sentimental analysis of the reviews of each restaurant.

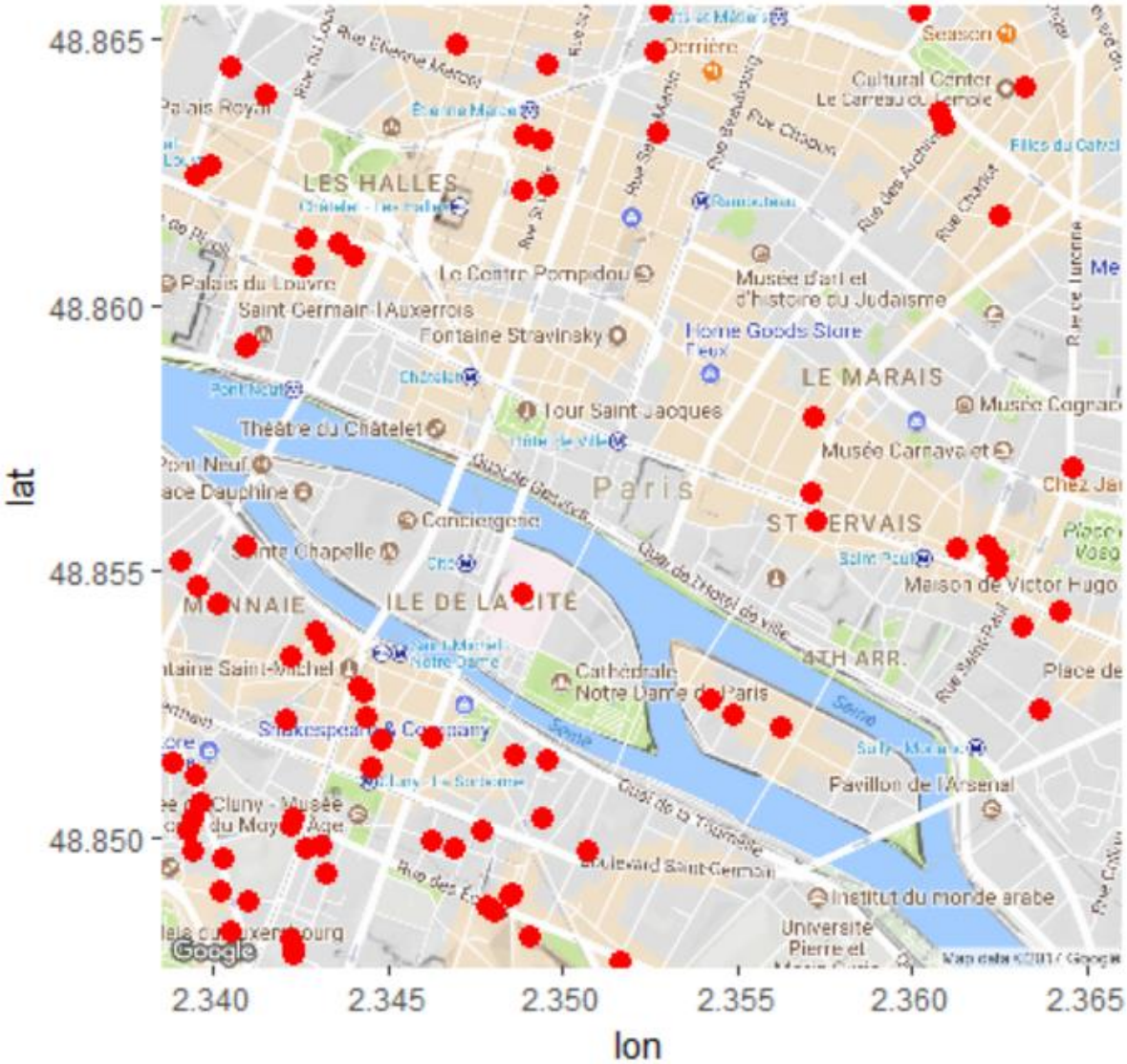
We decided to create a wordcloud of the words picked up during sentimental analysis:

```{r}
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
text <- restaurant$Review_1
docs <- Corpus(VectorSource(text))


toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
docs <- tm_map(docs, toSpace, "/")
docs <- tm_map(docs, toSpace, "@")
docs <- tm_map(docs, toSpace, "\\|")


dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 10)


set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
          max.words=200, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
```

**WordCloud**


- Plotting the Location of the Hotels on the map.

```{r}
# GEOCODE
geo.paris_locations <- geocode(as.character(hotel$Hotel_Name))#gets all lat and long
# COMBINE DATA
hotel.df <- cbind(hotel,geo.paris_locations )


# USE WITH GGPLOT  pointing each n every hotel
get_map("Paris", zoom = 15) %>% ggmap() +
  geom_point(data = hotel.df, aes(x = lon, y = lat), color = "red", size = 3)
```

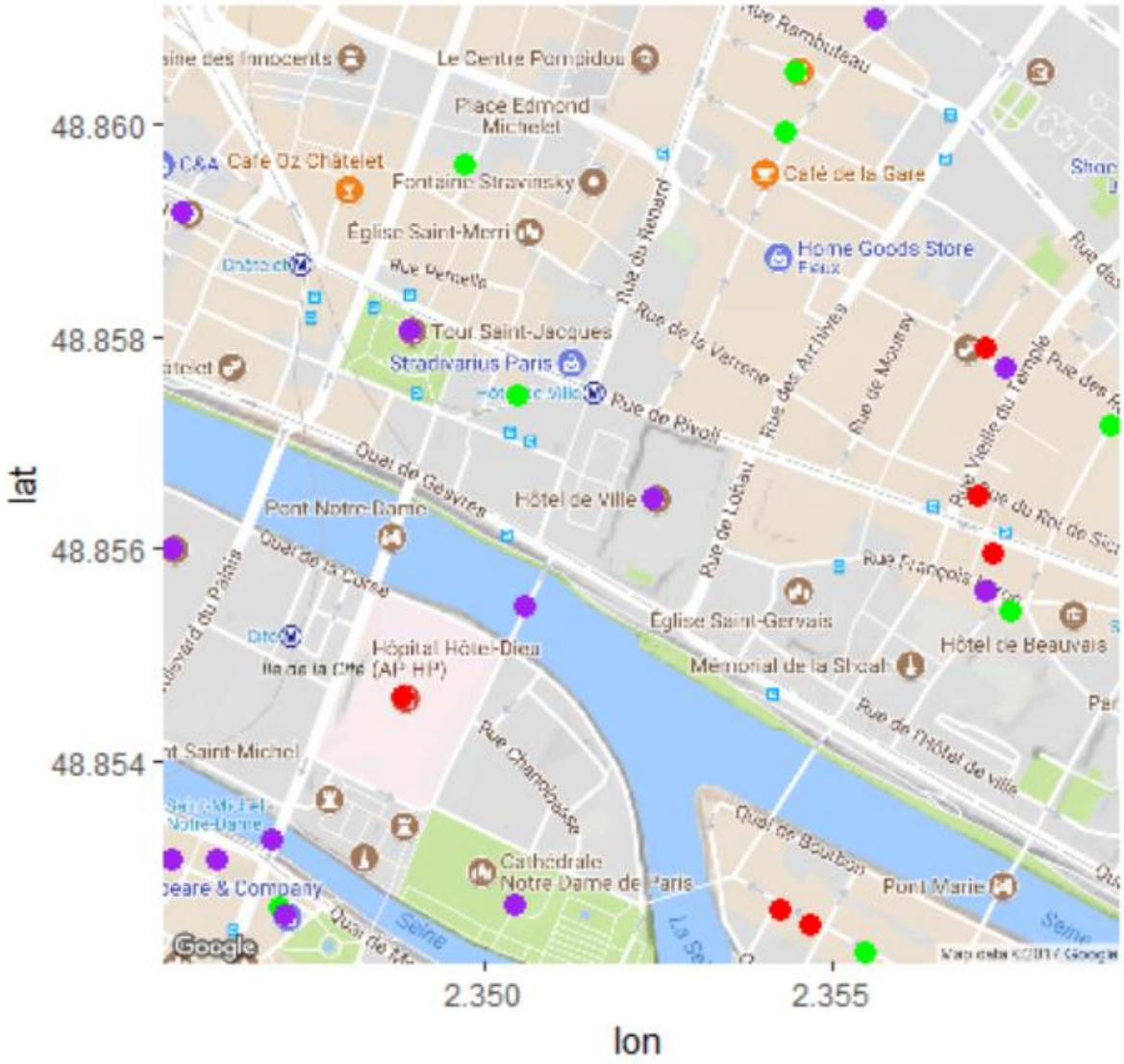



- Bird's Eye View of all the Restaurants, Hotels and Tourist Spots in the area.

```{r}
# geocode
geo.hotel_locations <- geocode(as.character(hotel$Hotel_Name))#gets all lat and long of hotels
geo.restaurant_locations <- geocode(as.character(restaurant$Restaurant_Name))#gets all lat and long of restaurants
geo.poi_locations <- geocode(as.character(Points_of_interest$Name))#gets all lat and long of POI

```

```{r}
# location coordinates
hotel.df <- cbind(hotel,geo.hotel_locations)
restaurant.df <- cbind(restaurant,geo.restaurant_locations )
poi.df <- cbind(Points_of_interest,geo.poi_locations )


# Plotting on the map
get_map("Paris", zoom = 16, maptype = "terrain") %>% ggmap() +
  geom_point(data = hotel.df, aes(x = lon, y = lat), color = "red", size = 3)+
  geom_point(data = restaurant.df, aes(x = lon, y = lat), color = "green", size = 3)+
  geom_point(data = poi.df, aes(x = lon, y = lat), color = "purple", size = 3)
```

**Red- Hotels** **Green – Restaurants** **Purple – Tourist Spots**



# Insights

**Learning Outcomes:**

We chose to do Analysis on TripAdvisor.com because the data offered for a search is very elaborate and often very confusing for the user. With this project, we aimed to give a very concise and straightforward response to a user's query.

During the course of this project, we learnt a lot of new techniques for analysis and visualization.

- We learnt to effectively use ggmap function.
- How to use HTML nodes to our advantage while scraping the data.
- How to assign primary keys to a database.
- Text analysis – Sentiment Analysis, Wordcloud.

We had difficulty to get the address of the hotel as it was inside a follow-up link. To overcome this problem, we tried using a variety of web scrapers and Chrome Extensions. We even tried to build our own web scraper using BeautifulSoup in Python.

Finally, we figured out that we could use the hyperlink extracted during our initial web scraping to get to the web page of each hotel and individually scrape the address.

**Future Possibilities:**

- We could further enhance the project by building a 'Shiny' app.
- We could use the location specified by the user to get the Hotels, Restaurants and Tourist spots in the locality.
- We can extend our project to include flight details and form a tour package for the specified budget.

**References:**


1. TripAdvisor.com
2. Wikipedia.com
3. R for Data Science – Hadley Wickham