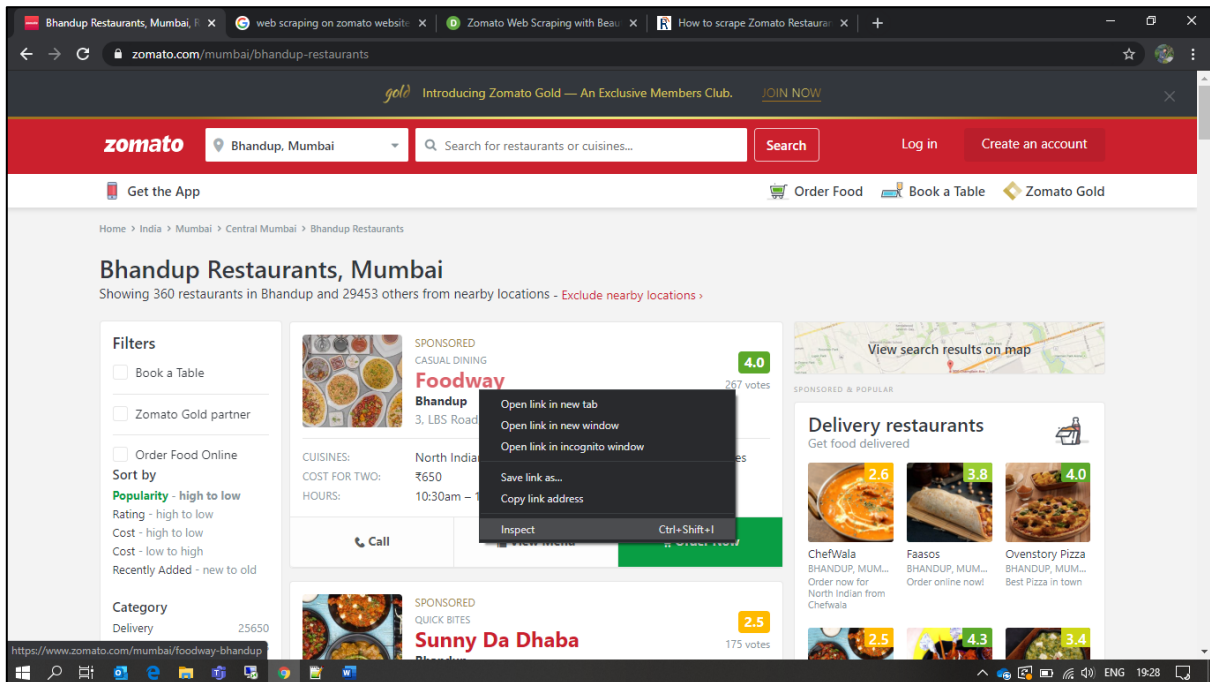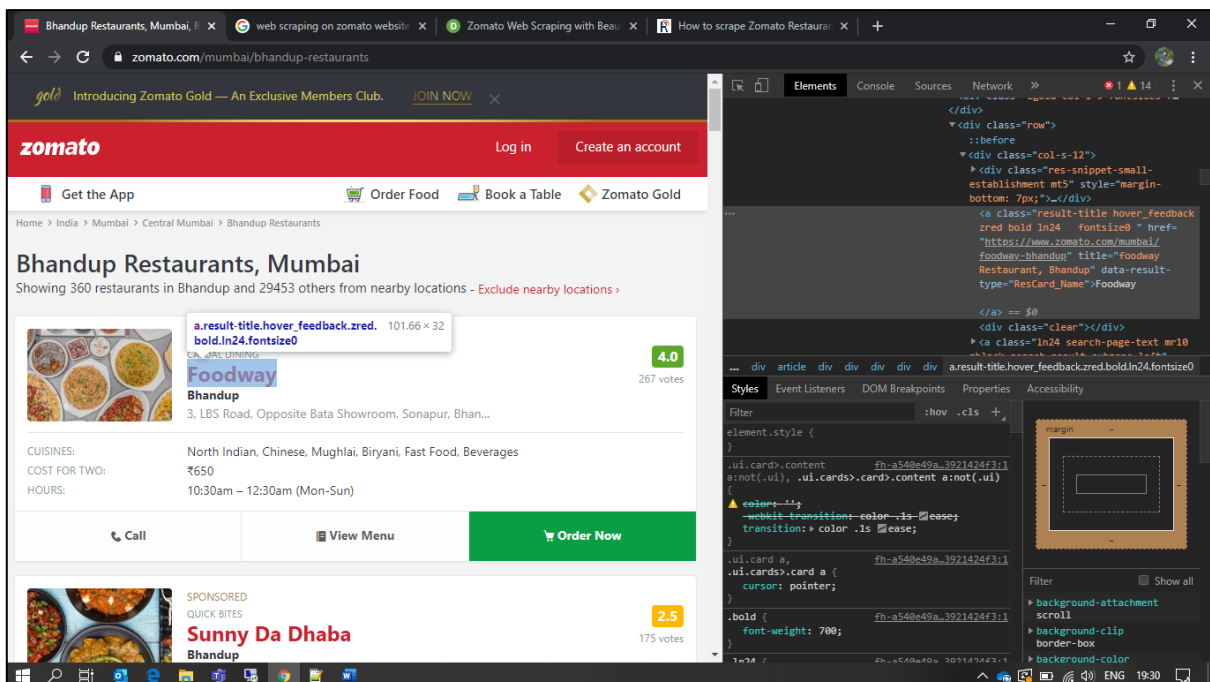1. Go to Link

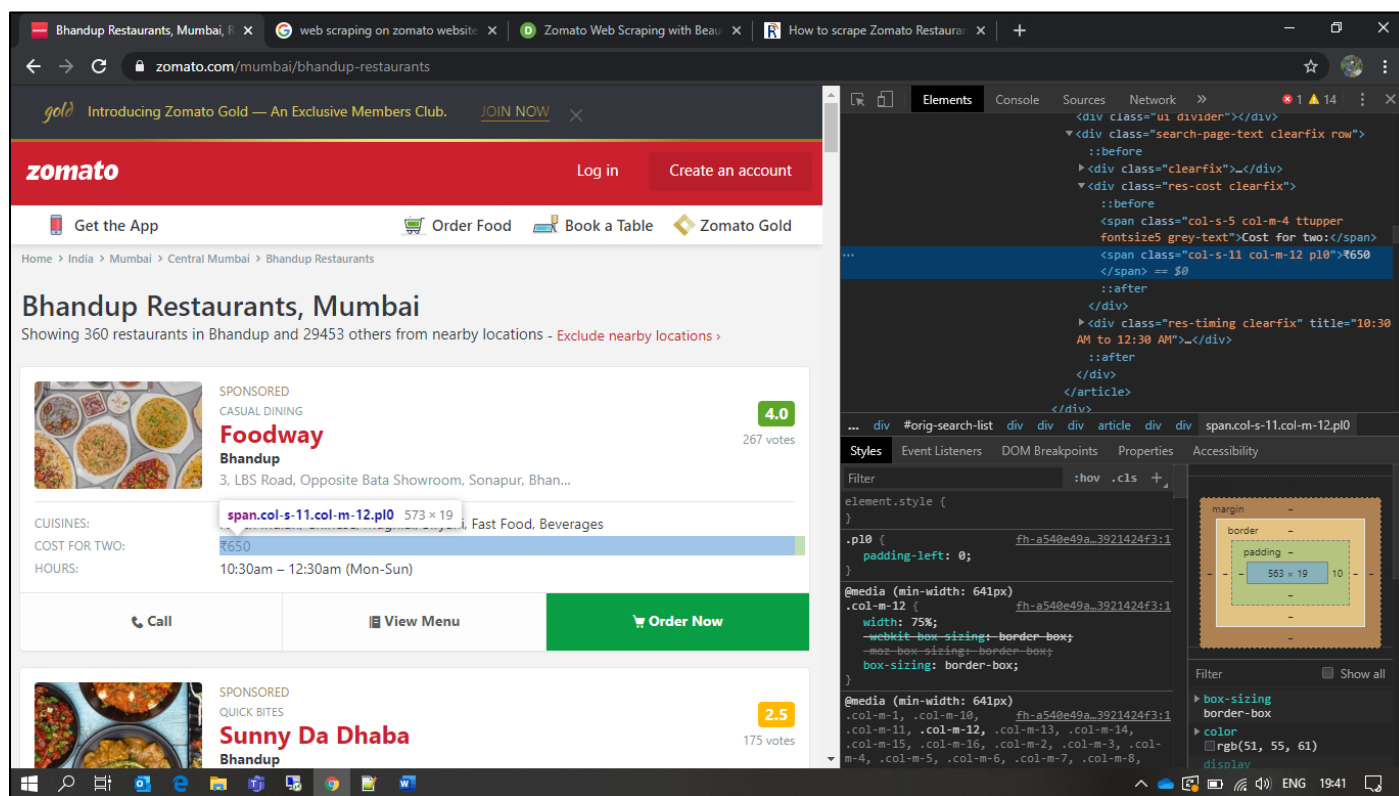*https://www.zomato.com/mumbai/bhandup-restaurants*

2. Right Click on any restaurant name and select inspect element

3. Hover over the elements until you find only the restaurant name is selected. Check for the HTML element and the class ID used here. In this case HTML element is **a** and class id is **result-title** thus, put **a.result-title** as html node and then extract the title attrinute using **html-attr** method. If we look at the title attribute closely we see the restaurant name and place both are mentioned separated by a **","**. Hence use **stsplit** to separate the data. (See line 13 in the script)

4. Similar we get the Cost for two and timings of the restaurant. In this case, we see the data is under division (kind of a box which can contain things) **res-cost** and **res-timing.** Within them, **span.p10** is the element.class combination for cost. Likewise, for timing it is **div. search-grid-right-text**. (See line 22 and 23 in the script)

5. Now, we have the data and we must perform some post -processing on the data like removing white space, if any (line 29) and replace character with alternative text which excel won't understand.

Limitations:

Currently, the script only scrapes details from 1 page out of the multiple possible pages. One page shows only 15 restaurants at a time hence, in the data there are only 15 entries. More records can be parsed but currently it's limited.

References:

1. https://www.datacamp.com/community/tutorials/r-web-scraping-rvest
2. https://www.analyticsvidhya.com/blog/2017/03/beginners-guide-on-web-scraping-in-r-using-rvest-with-hands-on-knowledge/
3. https://medium.com/free-code-camp/an-introduction-to-web-scraping-using-r-40284110c848
4. https://www.r-bloggers.com/how-to-scrape-zomato-restaurants-data-in-r/