

## Objective:

The objective of this assignment is to develop a web scraper to extract specific information from an app/website. This assignment will test your ability to understand the app/web scraping concepts, make use of scraping libraries, and counter challenges such as blocking, authentication, etc., in the data extraction process.

**Problem Statement:** Develop the scraper to scrape Grab Food Delivery.

**Website/App:** Grab food- Singapore

**Location:** You need to enter any of the given locations to scrape the list of restaurants and their details.

1. PT Singapore - Choa Chu Kang North 6, Singapore, 689577
2. Chong Boon Dental Surgery - Block 456 Ang Mo Kio Avenue 10, #01-1574, Singapore, 560456

### Tasks that scraper should perform:

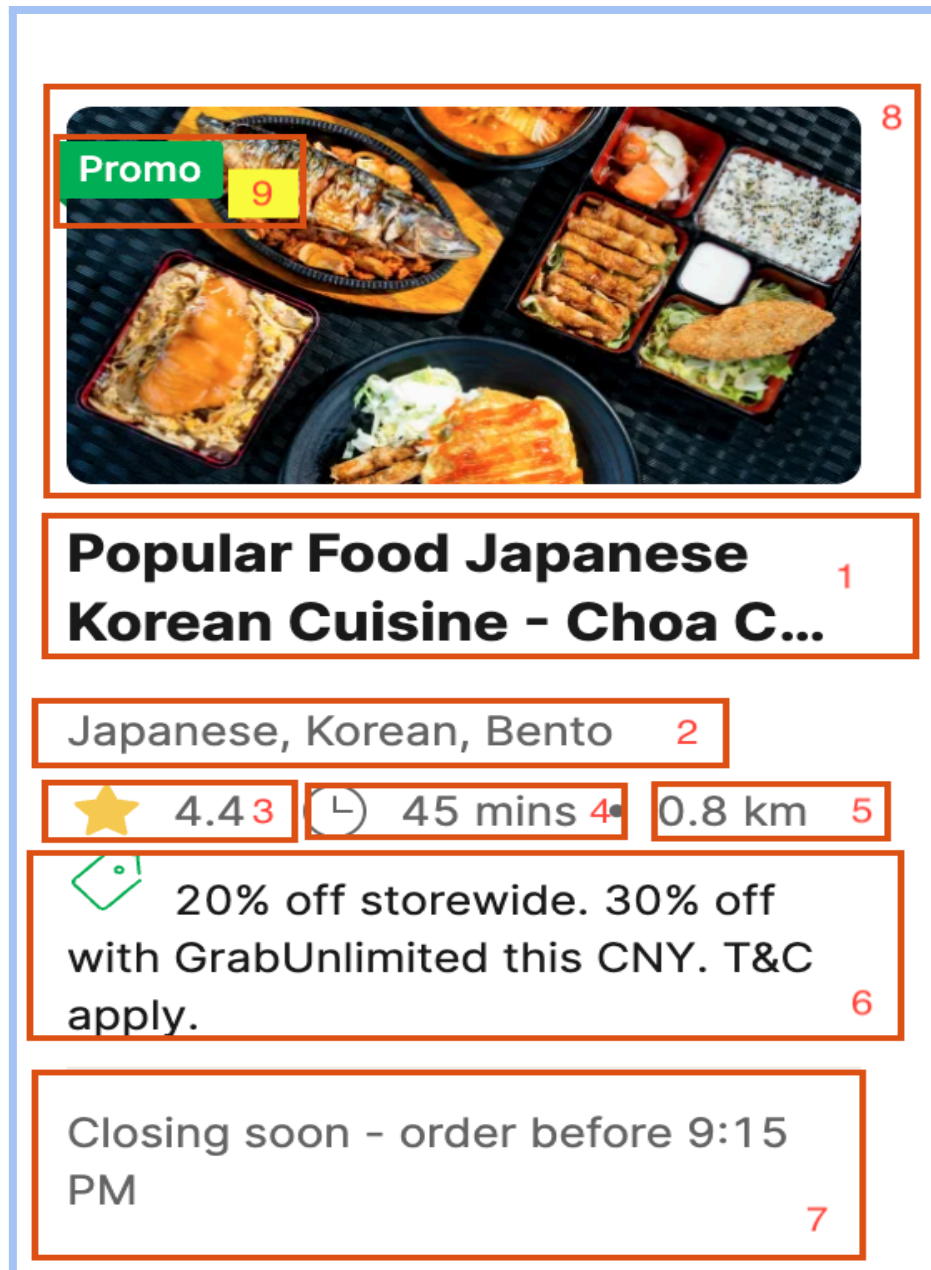
- Extract the restaurant list of the selected location along with restaurant details(Please refer to the schema given below)
- Create a unique restaurant list.
- Extract the delivery fee and estimated delivery time for any one of the selected locations.

**Result format:** gzip of ndjson.

- Write a Web Scraper:
  - Identify the scraping logic as per the given app/website.
  - Develop the scraping service in any language. Please make below programming contracts while coding.
    - Code should follow OOP concepts.
    - Code folder, class, method structure should be as per programming language standards.
    - Programming nomenclature should be followed in code.
    - Code should be highly optimised for higher load. Please make use of multithreading, multi processing, or some other paradigm to make it scalable.
    - Use consistent comment and logging format.
  - Saved the extracted data in gzip of ndjson.

- Choose Data to Extract:

- Extract all the fields/column data that are visible on the website/response



The image shows a restaurant listing with the following elements:

- Image (8):** A collage of food items including a whole fish, a bowl of ramen, and a bento box. A green 'Promo' tag (9) is overlaid on the top left.
- Restaurant Name (1):** Popular Food Japanese Korean Cuisine - Choa C...
- Cuisine (2):** Japanese, Korean, Bento
- Rating (3):** 4.4 (represented by a yellow star icon)
- Delivery Time (4):** 45 mins (represented by a clock icon)
- Distance (5):** 0.8 km
- Promotional Offer (6):** 20% off storewide. 30% off with GrabUnlimited this CNY. T&C apply.
- Notice (7):** Closing soon - order before 9:15 PM

- Explanation of each point displayed in the Image.

- Restaurant list schema.

1. Restaurant Name
2. Restaurant Cuisine
3. Restaurant Rating
4. Estimate time of Delivery
5. Restaurant Distance from Delivery Location
6. Promotional Offers Listed for the Restaurant.
7. Restaurant Notice If Visible.

8. Image Link of the Restaurant
9. Is promo available (True/False)
10. Restaurant ID (Look like this: 4-C2NKCETKWE5WLT)
11. Restaurant latitude and longitude (This is has to be taken from the grab website only)
12. Estimate Delivery Fee

- Ensure that your scraper handles cases where the data may be structured differently on different pages.

## ● Documentation:

- Add comprehensive comments in your code to explain the purpose of each section.
- List down what kind of QC you have performed over the extracted data.
- Stats of data should at least cover the total count as per the selected location, as well as not null and null stats of above mentioned mandatory fields.
- Compose a brief report documenting the following:
  - Your overall approach and methodology.
  - Challenges faced during the scraping process.
  - Any improvements or optimizations you could envision.

## ● Submission:

- Submit your developed code along with the extracted data and the documentation report.

## ● Important Notes:

- Ensure that your web scraping activities comply with the website's terms of service.
- Test your scraper on a small scale before running it on a large dataset.
- Use the right proxy/VPN to access the website.