

# TABLE OF CONTENTS

**Objective:**..... 1

    Levels to be Scrapped:..... 1

    Details to be Extracted:.....2

        For Level 1 (Category Items):..... 2

        For Level 2 (Subcategory Items):.....2

        For Level 3 (Product Items of subcategory):.....3

    Data Format:..... 3

    Note:..... 4

    Deliverables:..... 4

        1. Code File:.....4

        2. Data File:.....4

        3. Images folder:..... 4

## Task: Web Scrapping from Almeera Online Store

*Please do not send us incomplete tasks, we greatly value your and our time. Incomplete tasks will not be entertained.*

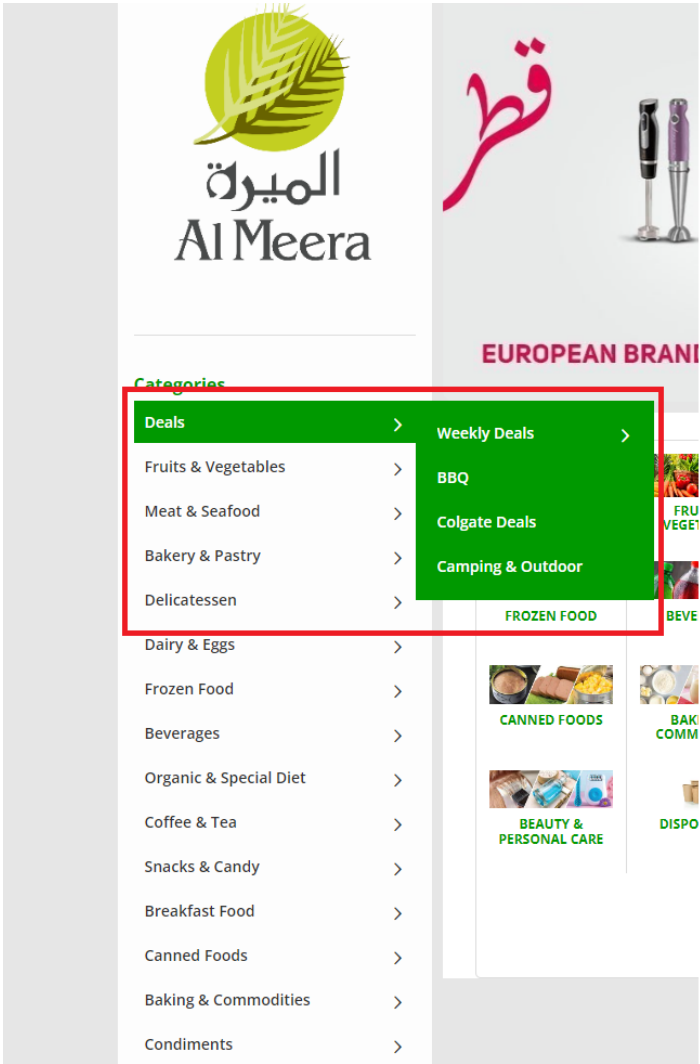
### Objective:

Extract specific data from [Almeera Online](#) involving multi-level category and product details, and store this data in a structured JSON format.

Levels to be Scrapped:

- **Level 1:** Category
- **Level 2:** Subcategories
- **Level 3:** Product Items

**BONUS : FOR THOSE WHO ATTEMPT THIS TASK USING,**  
**“SCRAPY”**



Scrap subcategories up to the highlighted level as show in the above picture

Details to be Extracted:

For Level 1 (Category Items):

- Category title
- Category Image
  - URLs
  - File (download)

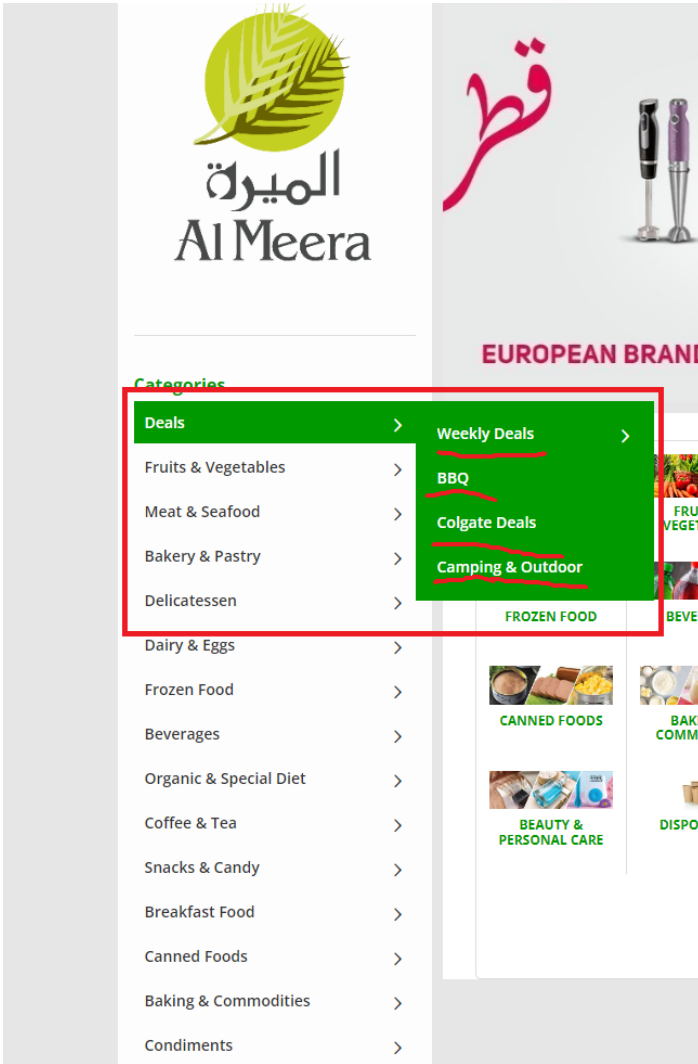
Scrape all the categories.

For Level 2 (Subcategory Items):

- Subcategory title

Scrape all subcategories of a category (do not scrape subcategories of subcategory)

BONUS : FOR THOSE WHO ATTEMPT THIS TASK USING,  
“SCRAPY”



For Level 3 (Product Items of subcategory):

- Product Image
  - URLs
  - File (download)
- Item title
- Item price
- Item Barcode / SKU (if available)

Scrap 2 pages per subcategory  
Scrap only 5 products per page

Data Format:

Store the extracted data in a JSON format structured as follows:

```
{  
  "CategoryTitle": "string",  
  "CategoryImageURL": "string",
```

**BONUS : FOR THOSE WHO ATTEMPT THIS TASK USING,**  
**“SCRAPY”**

```

"Subcategories": [
  {
    "SubcategoryTitle": "string",
    "Products": [
      {
        "ItemTitle": "string",
        "ItemImageURL": "string",
        "ItemPrice": "string",
        "ItemBarcode": "string"
      },
      {
        "ItemTitle": "string",
        "ItemImageURL": "string",
        "ItemPrice": "string",
        "ItemBarcode": "string"
      }
      // Additional products...
    ]
  }
  // Additional subcategories...
]
}

```

#### Note:

1. Ensure the extraction of data adheres to legal and ethical standards.
2. Ensure the web scraping activity complies with the website's robots.txt guidelines.
3. Include appropriate waiting time between requests to avoid burdening the website's server.
4. Validate and clean the extracted data where necessary.
5. Ensure to manage pagination to navigate through multiple pages of products within each category/subcategory.
6. Consider implementing error handling to manage potential issues that might arise during the scraping process.

#### Deliverables:

##### 1. Code File:

Provide the code script used to perform the scraping.

##### 2. Data File:

A JSON file containing the scraped data following the provided structure.

##### 3. Images folder:

A Folder that contains, the extracted images of the products

## How to submit the task

1. Create a **git repository** for the task and send the repository link over the email.
2. It should have
  - a. Code file
  - b. Data file
  - c. Images

**BONUS : FOR THOSE WHO ATTEMPT THIS TASK USING,**  
**“SCRAPY”**