Steps for Zomato data wrangling:

Data extraction:

- The data is extracted from Zomato website using Python package '**Beautiful Soup'**.
- This is data is for all restaurants of Bangalore city which is around 12k-13k of records, pulled as of January 2020.
- Following are the fields
    1. **restaurant_link**: Link for the restaurant
    2. **restaurant_ID**: Unique restaurant id
    3. **restaurant_name**: Name of the restaurant
    4. **locality**: neighbourhood of the restaurant
    5. **restaurant_category**: Category of restaurant based on what food they serve, like dining or quick bites, etc.
    6. **zomato_gold**: Whether the restaurant provides zomato gold benefits
    7. **discounts**: Discounts offered by the restaurant
    8. **photos_taken**: Number of photos taken at the restaurant
    9. **rating**: Zomato rating
    10. **votes**: Votes for the ratings or reviews
    11. **cuisines**: Type of cuisines served
    12. **approx._cost_for_2**: Approx cost for 2 people
    13. **opening timings**: Opening and closing timings of the restaurant
    14. **address**: Detailed address of the restaurant
    15. **latitude**: Latitude of restaurant
    16. **longitude** :Longitude of the restaurant
    17. **more_info**: main features or services provided by the restaurant like delivery, outside seating, etc
    18. **featured_in**: Featured in which categories of Zomato collections
    19. **most_liked_food**: Most liked or famous for in food items and rating

20. **most_liked_service**: Most liked service of the restaurant and rating
21. **most_liked_look&field**: Most liked, look and feel of the restaurant and rating
22. **reviews**: Reviews available on first page of the restaurant along with time of review posted and sentiments.

Data cleaning:

- Most of the data is cleaned/formatted while scraping.
- Some columns are manipulated to tuples.
- Opening and closing timings are transformed to datetime formats.
- Missing values are transformed to np.NAN
- Duplicates rows, if any, are removed based on the restaurant_id.
- No outliers.

For Data cleaning and EDA refer:
https://github.com/Anandpatil412/DSC/blob/master/CapstoneProject1/Zomato/ZomatoDataExtractingAndCleaning/zomatoDataCleaning.ipynb

For Web scarping refer:
https://github.com/Anandpatil412/DSC/blob/master/CapstoneProject1/Zomato/ZomatoDataExtractingAndCleaning/zomatoScrapper.ipynb