# Tutorial 5 Worksheet AY 22/23 Sem 1

DSA2101

## Restaurants Data

In chapter 2, we worked with the JSON version of the restaurants data. The following code converts the file into a tibble within R. Let us use it to practice with `dplyr` data manipulations.

```
library(tidyverse)
library(jsonlite)

rest_json <- stream_in(file("../data/restaurants_dataset.json")) %>%
  unnest(cols=c(address, grades))
```

Recall that in the lecture, there 25359 restaurants. The above procedure has removed the 738 records with NULL in the grades section, but it retains those with "Not Yet Graded".

The output tibble should look like this:

```
## # A tibble: 6 x 11
##   building coord  street zipcode borough cuisine date                grade score
##   <chr>    <list> <chr>  <chr>   <chr>   <chr>   <dttm>              <chr> <int>
## 1 1007     <dbl>  Morri~ 10462   Bronx   Bakery  2014-03-03 08:00:00 A         2
## 2 1007     <dbl>  Morri~ 10462   Bronx   Bakery  2013-09-11 08:00:00 A         6
## 3 1007     <dbl>  Morri~ 10462   Bronx   Bakery  2013-01-24 08:00:00 A        10
## 4 1007     <dbl>  Morri~ 10462   Bronx   Bakery  2011-11-23 08:00:00 A         9
## 5 1007     <dbl>  Morri~ 10462   Bronx   Bakery  2011-03-10 08:00:00 B        14
## 6 469      <dbl>  Flatb~ 11225   Brookl~ Hambur~ 2014-12-30 08:00:00 A         8
## # ... with 2 more variables: name <chr>, restaurant_id <chr>
```

1. As you can see, the `coord` column contains a list instead of numeric elements. Separate `coord` into two columns, lat and long, and drop the original `coord` column. Assign the name `q1_tbl` to the new tibble. It should look like this:

```
## # A tibble: 6 x 12
##   building   lat  long street  zipcode borough cuisine date                grade
##   <chr>    <dbl> <dbl> <chr>   <chr>   <chr>   <chr>   <dttm>              <chr>
## 1 1007      40.8 -73.9 Morris~ 10462   Bronx   Bakery  2014-03-03 08:00:00 A
## 2 1007      40.8 -73.9 Morris~ 10462   Bronx   Bakery  2013-09-11 08:00:00 A
## 3 1007      40.8 -73.9 Morris~ 10462   Bronx   Bakery  2013-01-24 08:00:00 A
## 4 1007      40.8 -73.9 Morris~ 10462   Bronx   Bakery  2011-11-23 08:00:00 A
## 5 1007      40.8 -73.9 Morris~ 10462   Bronx   Bakery  2011-03-10 08:00:00 B
## 6 469       40.7 -74.0 Flatbu~ 11225   Brookl~ Hambur~ 2014-12-30 08:00:00 A
## # ... with 3 more variables: score <int>, name <chr>, restaurant_id <chr>
```

2. Find the range of lat and long for the restaurants in `q1_tbl`. Are there mistakes in the dataset? Identify one restaurant with incorrect coordinates.

3. How many restaurants in Manhattan serve American cuisine?

4. Find the restaurants that have been graded the most number of times. Include the restaurant name in your output. The columns should match this output (only first 2 rows shown):

```
## # A tibble: 2 x 3
##   restaurant_id name       n
##   <chr>         <chr>  <int>
## 1 41177358      S'Mac      9
## 2 41181651      Benton     9
```

5. For each restaurant, compute the shortest duration (in days) between gradings. You may want to take a look at the `difftime` function in R.
6. The 5-number summary of a dataset consists of the minimum, 1st-quartile, median, 3rd-quartile and maximum of the dataset. Compute the five-number summary score for each grade. Include the count for each grade.
7. Use `across` to compute the min. and max. letter grade and score for each restaurant.
8. Count the number of gradings in each calendar month in each borough.
9. Recode the following cuisines as `Asian`, and then find the proportion of Asian restaurants in each borough:
   - Vietnamese/Cambodian/Malaysia
   - Thai
   - Chinese
   - Chinese/Japanese
   - Pakistani
   - Korean
   - Indonesian
   - Indian
   - Asian

   You may want to take a look at `recode()`.
10. List the names of all restaurants that have "C" as their first grade, and "A" as the second grade.
11. List all restaurants with only "A" grades.