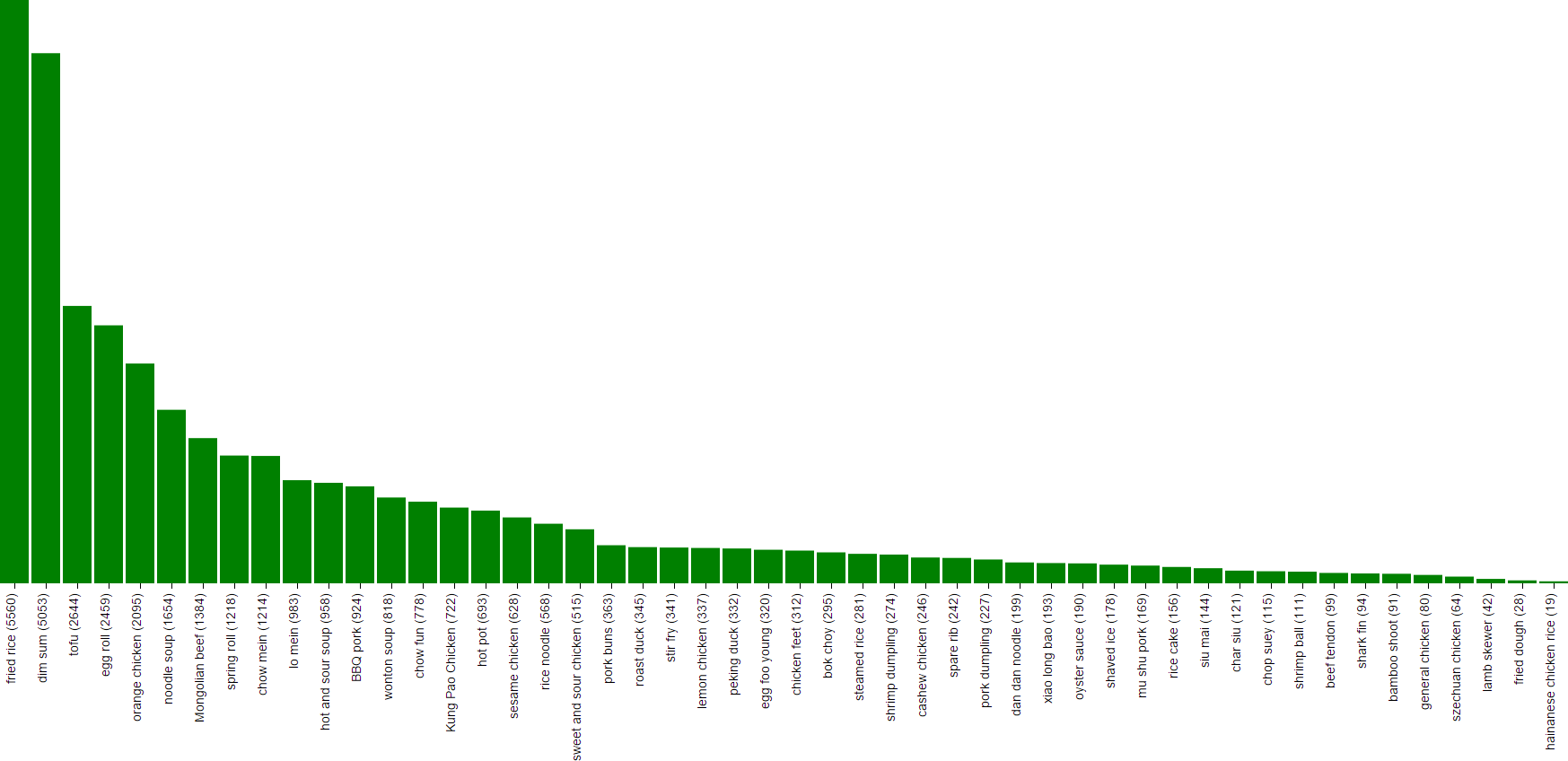
Task 4: Popular Dishes

# Task 4.1 Count Occurrence



## What

In this task, popular dish is the dish that people talk most, may not exactly the dish people love most (ratings).

## Dataset

Dish names: 50 manually tagged Chinese dish names from Task 3

Chinese restaurants reviews: categories/Chinese.txt provided

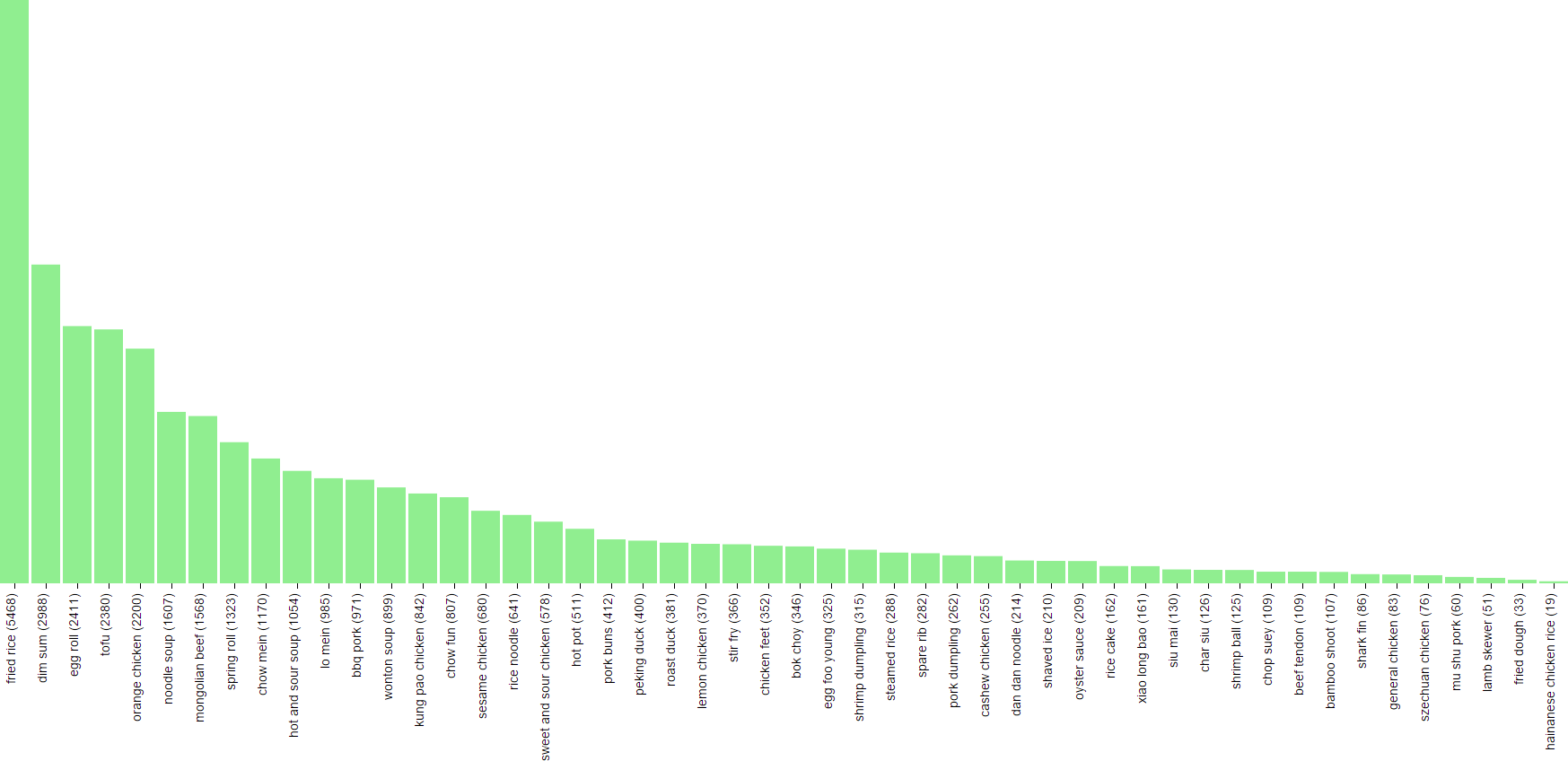
## How

1. Count the dish name in the Chinese cuisine review text file
2. Plot the data on D3

## Where

D3 visualization source code: <https://github.com/pauldeng/MOOC/blob/master/Data%20Mining%20Capstone/Task%204%20%26%20Task%205/Task%204/Task%204.1/task41.html>

# Task 4.2 Count Occurrence + Star Rating



## What and Result Analysis

In this task, popular dish is the dish that people talk and love most.

Compare with the result of Task 4.1, the changes are:

* Dishes rank up: egg roll, hot and sour soup, Kung Pao Chicken , szechuan chicken…
* Dishes rank down: toufu, lo mein, mu shu pork, chow fun…

## Dataset

Dish names: 50 manually tagged Chinese dish names from Task 3

Raw Yelp reviews: yelp\_academic\_dataset\_review.json

Raw Yelp business: yelp\_academic\_dataset\_business.json

## How

### Star rating weights

If dish name is found in the review text, depends on the rating, a weighted popular index will be calculated.

|  |  |  |
| --- | --- | --- |
| Star Ratings | Weights | Design decision |
| 5 | 2 |  |
| 4 | 1.5 |  |
| 3 | 1 | base, neutral weight score |
| 2 | 0.5 |  |
| 1 | 0.1 | At least customer purchased it. It is a sign of popularity. |

### Calculate dish popularity score

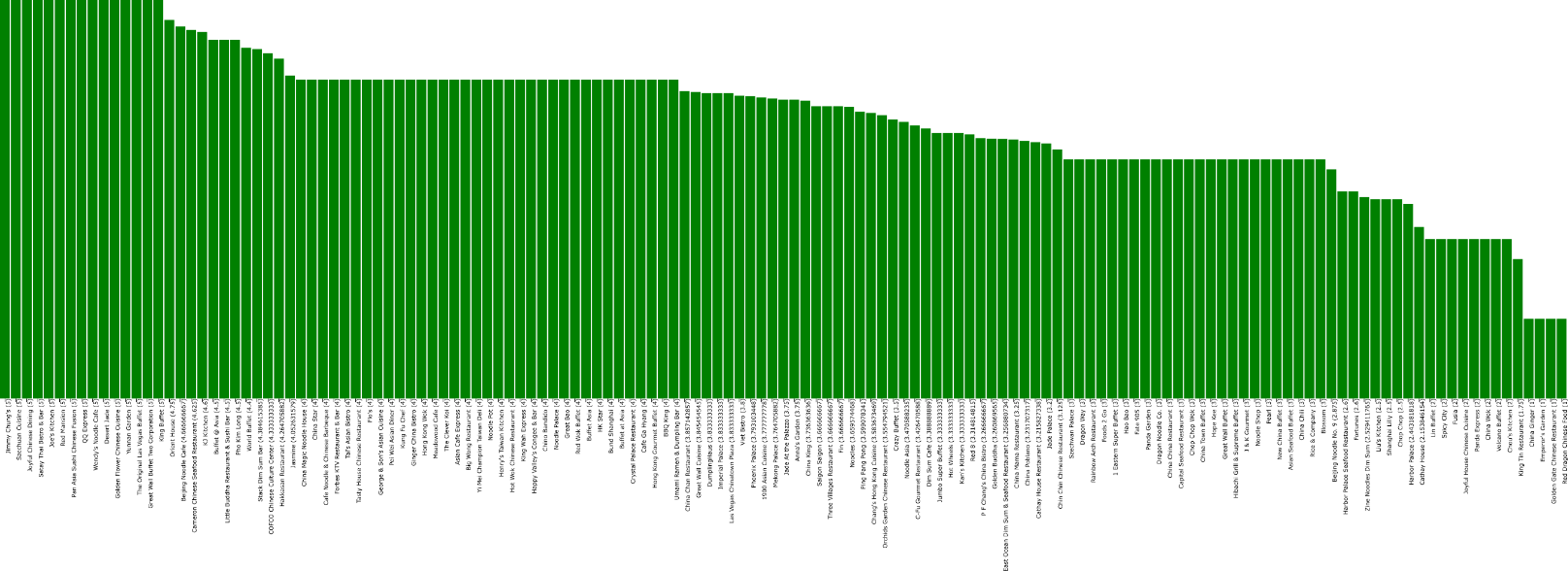
Dish\_Popularity\_Score = Sum\_Of\_All\_Occourance(1 \* Star\_Rating\_Weight)

## Where

D3 visualization source code: <https://github.com/pauldeng/MOOC/blob/master/Data%20Mining%20Capstone/Task%204%20%26%20Task%205/Task%204/Task%204.2/task42.html>

Task 5: Restaurant Recommendation

# Task 5.1 Average Restaurant Rating



## What

In this task, I rank the dim sum restaurants based on their review and ratings.

## Dataset

Dish: dim sum

Raw Yelp reviews: yelp\_academic\_dataset\_review.json

Raw Yelp business: yelp\_academic\_dataset\_business.json

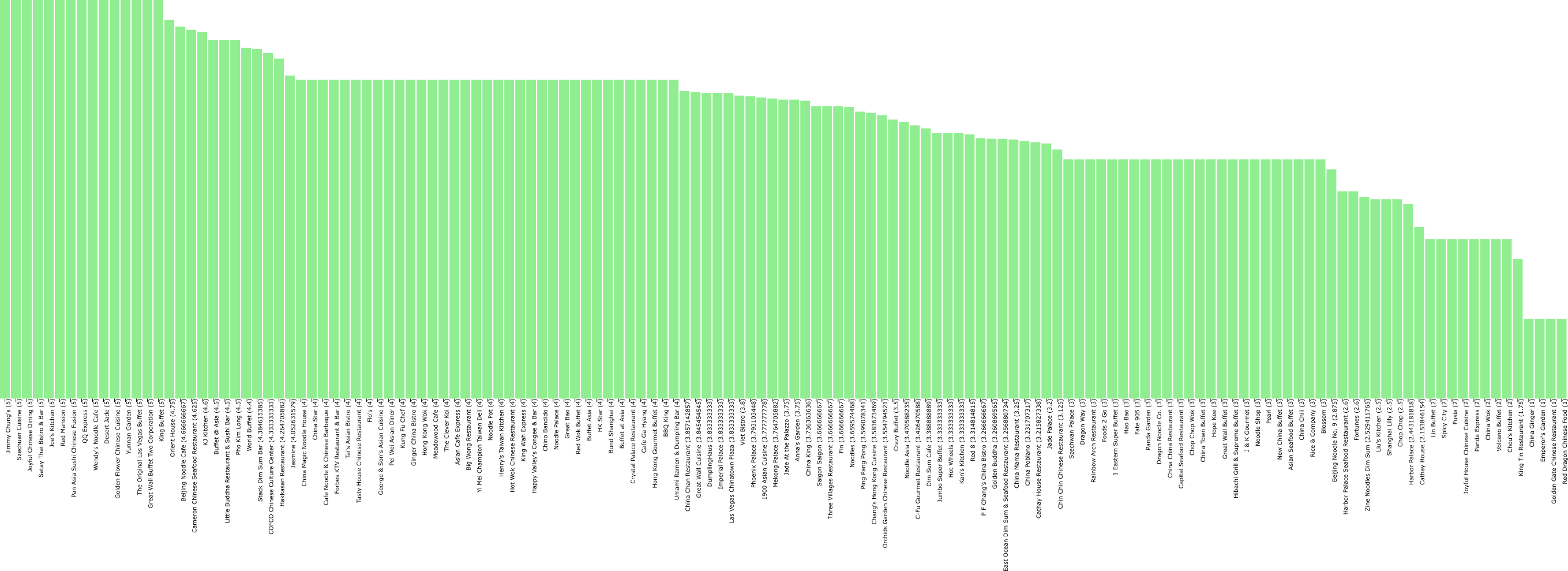
## How

1. Find review text contains “dim sim”
2. Find the review’s restaurant name and review rating
3. Calculate average rating

## Where

D3 visualization source code: <https://github.com/pauldeng/MOOC/blob/master/Data%20Mining%20Capstone/Task%204%20%26%20Task%205/Task%205/Task%205.1/task51.html>

# Task 5.2 Average Restaurant Rating + Review Useful Votes



## What and Result Analysis

In this task, I rank the dim sum restaurants based on their review opinion and ratings.

Compare the result of Task 5.1, there is no restaurant rank change. It tells me the customer star ratings are highly related to review comment text.

## Dataset

Dish: dim sum

Raw Yelp reviews: yelp\_academic\_dataset\_review.json

Raw Yelp business: yelp\_academic\_dataset\_business.json

## How

### Sentimental analysis

|  |  |  |
| --- | --- | --- |
| Opinion | Probability | Example |
| Positive | x | Sentimental\_Analysis\_Result = 2 \* x |
| Neutral | 0 | Sentimental\_Analysis\_Result = 0 |
| Negetive | y | Sentimental\_Analysis\_Result = -2 \* y |

### Calculate dish popularity score

Restaurant\_Score = Average ( 3/5 \* Star\_Rating + 2/5 Sentimental\_Analysis\_Result)

## Where

D3 visualization source code: <https://github.com/pauldeng/MOOC/blob/master/Data%20Mining%20Capstone/Task%204%20%26%20Task%205/Task%205/Task%205.2/task52.html>