**Data Exploration & Preparation**

1. **Data Exploration**

**Table:**

| **Table Name** | **Number of Columns** | **Number of Rows** |
| --- | --- | --- |
| hca\_users | 78 | 100 |
| hca\_food\_items | 19 | 100 |
| hca\_recipes | 21 | 100 |
| hca\_food\_logs | 15 | 100 |
| hca\_google\_fence | 9 | 100 |

**1.1 HCA Merged Table**

Check for null values:

item\_name has 63 null values

substitute\_itemsids has 63 null values

substitute\_recipeids has 36 null values

food\_category has 63 null values

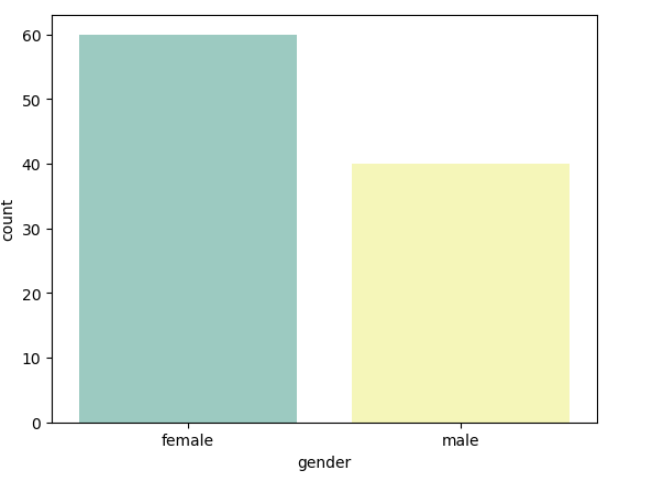
recipe\_name has 36 null values

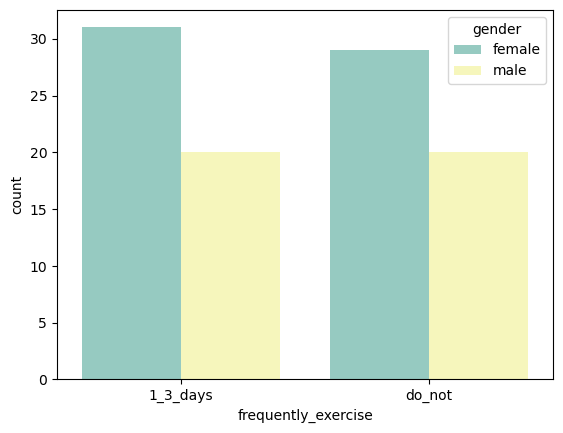
Eqv has 36 null values

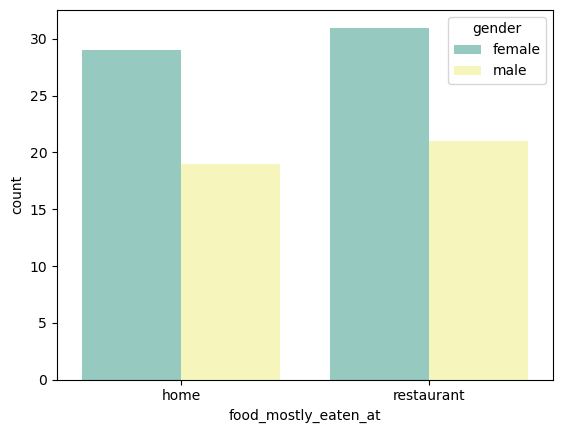
Replaces all null values with an empty string.

Categorical values:

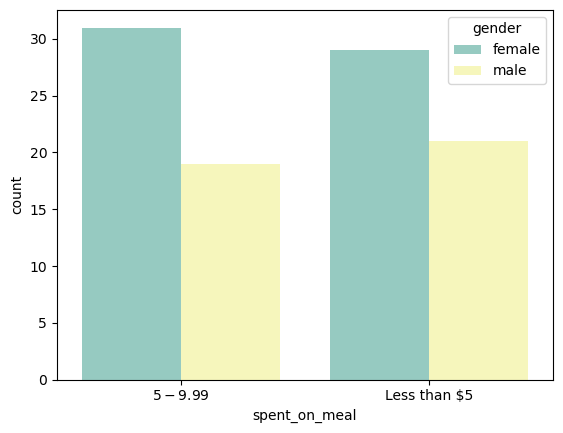
| **Variable Name** | **Values** |
| --- | --- |
| gender | [‘female’, ‘male’] |
| frequently\_exercise | [‘1\_3\_days’, ‘do\_not’] |
| food\_mostly\_eaten\_at | [‘home’, ‘restaurant’] |
| spent\_on\_meal | [‘$5-$9.99’, ‘Less than $5’] |
| diet\_plan | [1, 2, 3, 4] |

The ratio of female to male users is 60% to 40% respectively.

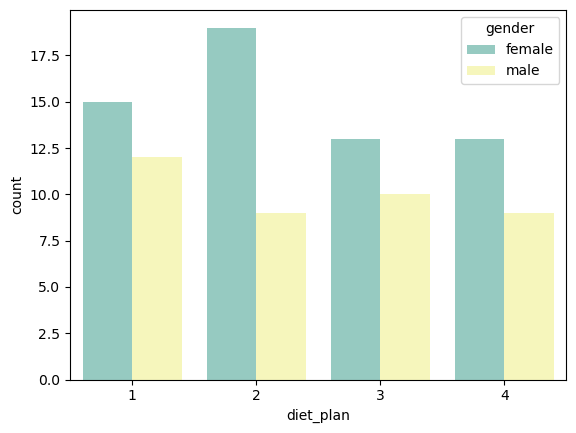
Half of the users frequently exercise and 50% of the male users do not frequently exercise whereas a higher percentage of females exercise.



A higher percentage of users, both male and female, eat at restaurants.

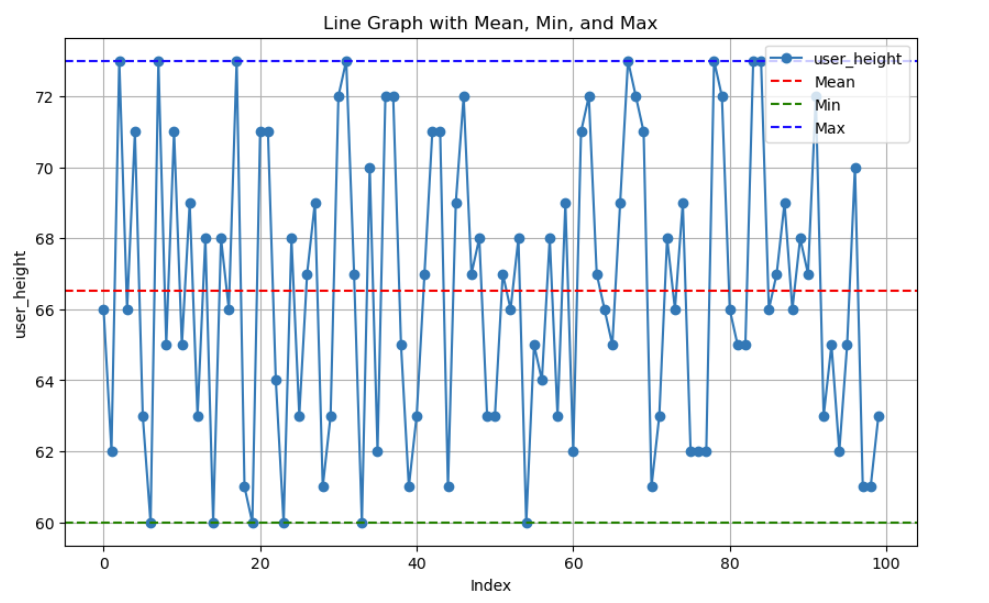


A higher percentage of females spend 5-9.99$ on food whereas higer percentage of males spend less than 5$.



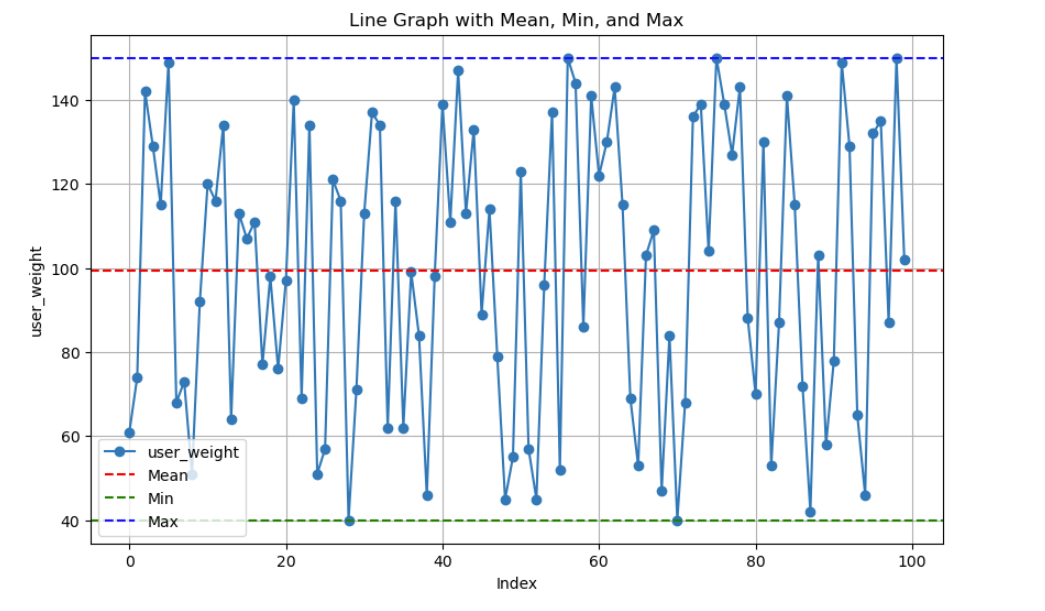
Females largely prefer the 2nd diet plan while males mostly prefer the 1st diet plan.

Numerical values:



The mean height of the users is 66.52. Minimum is 60 and maximum is 71.

User\_weight- 99.26



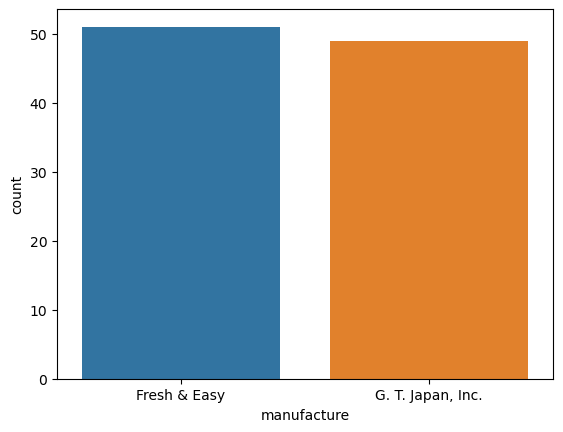
The mean weight of the users is 99.26. Minimum is 40 and maximum is 150.

**1.2 HCA food items Table**

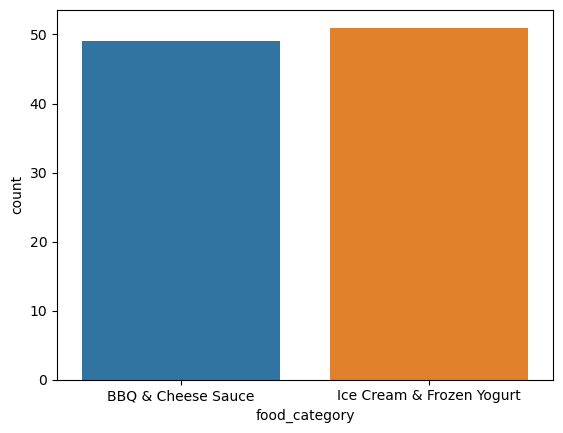
Check for null values: *No null values were found in this table.*

Categorical values:

| **Variable Name** | **Values** |
| --- | --- |
| manufacturer | ['Fresh & Easy' 'G. T. Japan, Inc.'] |
| food\_category | ['BBQ & Cheese Sauce' 'Ice Cream & Frozen Yogurt'] |

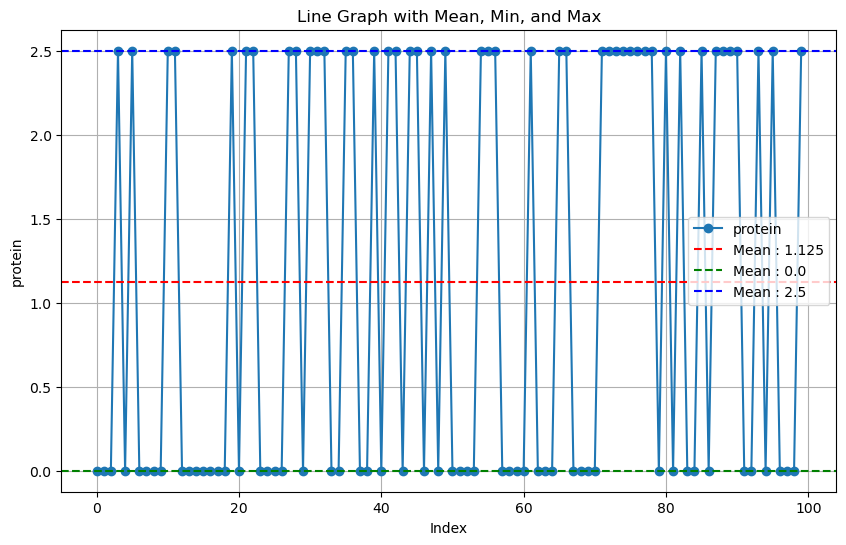


Fresh & Easy and G.T.Japan, Inc. manufacturers have an almost even spread of 52% and 48% respectively

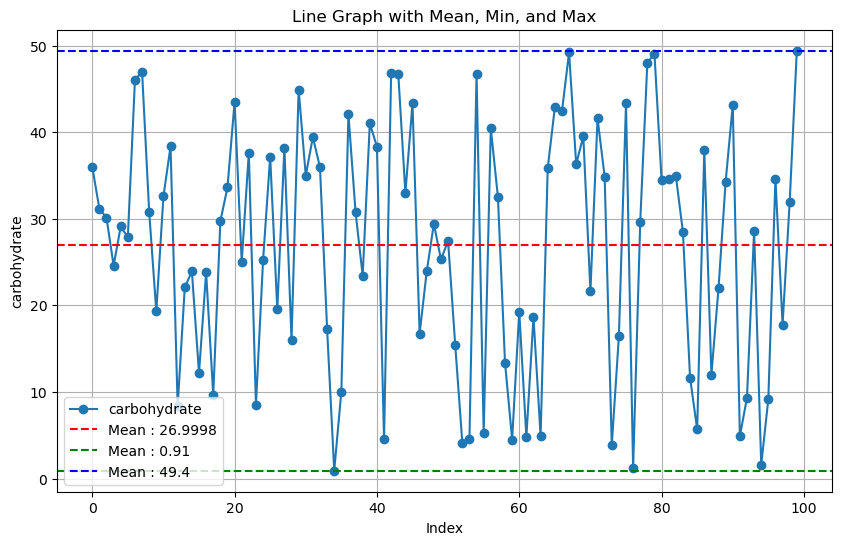


Ice cream & frozen yogurt, 52% is consumed slightly more than BBQ & cheese sauce 48%.

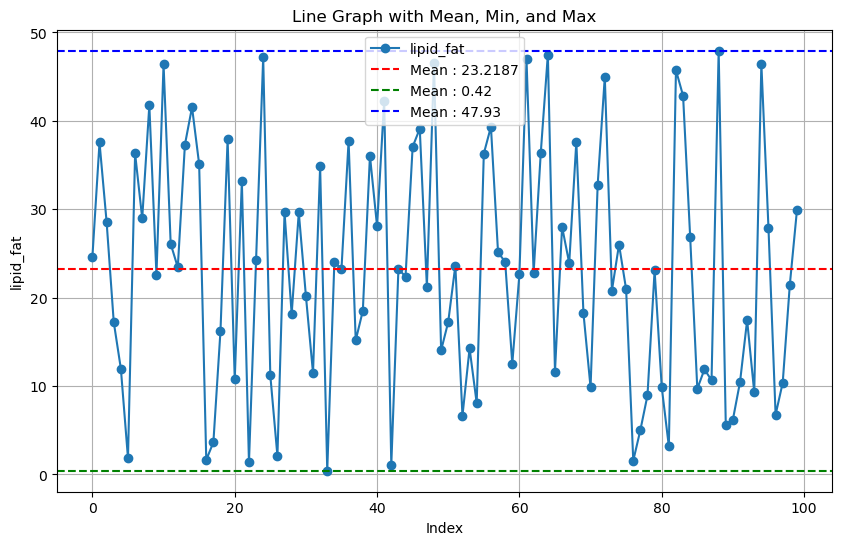
Numerical values:



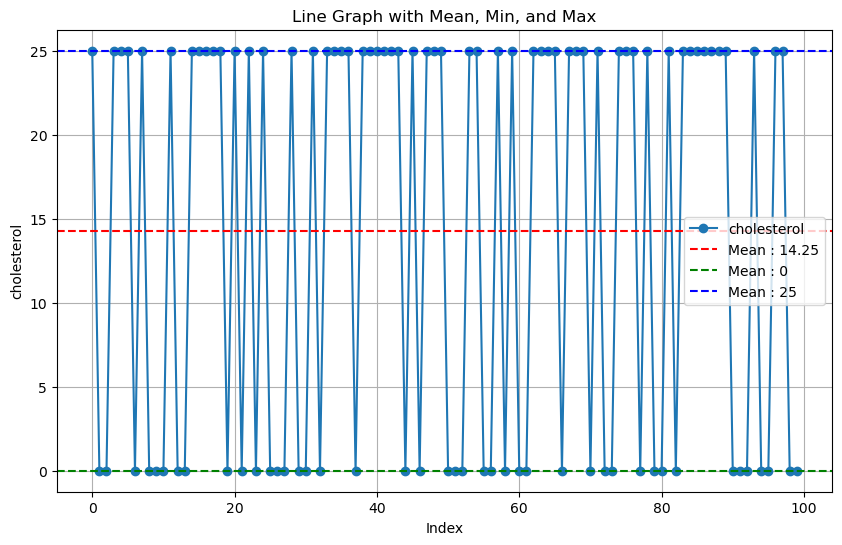
The mean protein of the food items is 1.124. Minimum is 0.0 and maximum is 2.5.



The mean carbohydrate of the foor items is 26.99. Minimum is 0.91 and maximum is 49.4.



The mean lipid\_fat of the food items is 23.2187. Minimum is 0.42 and maximum is 47.93.

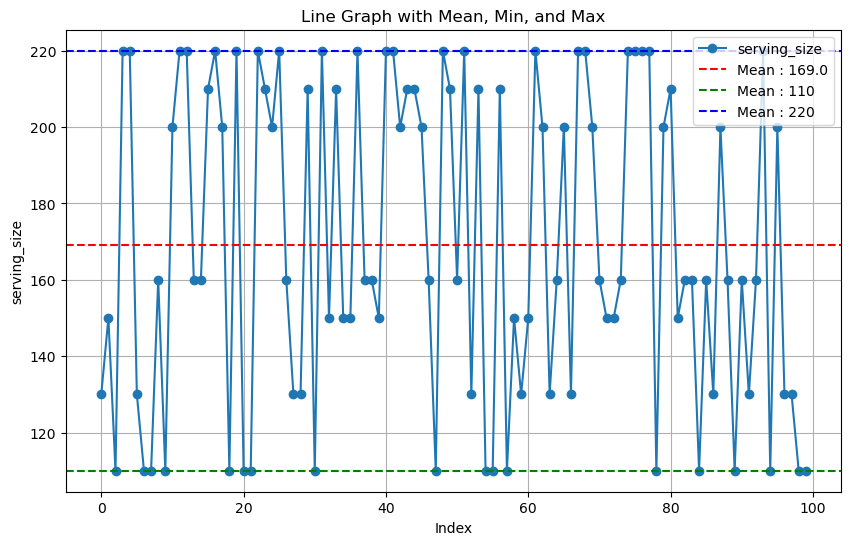


The mean cholesterol of the food items is 14.25. Minimum is 0 and maximum is 25.

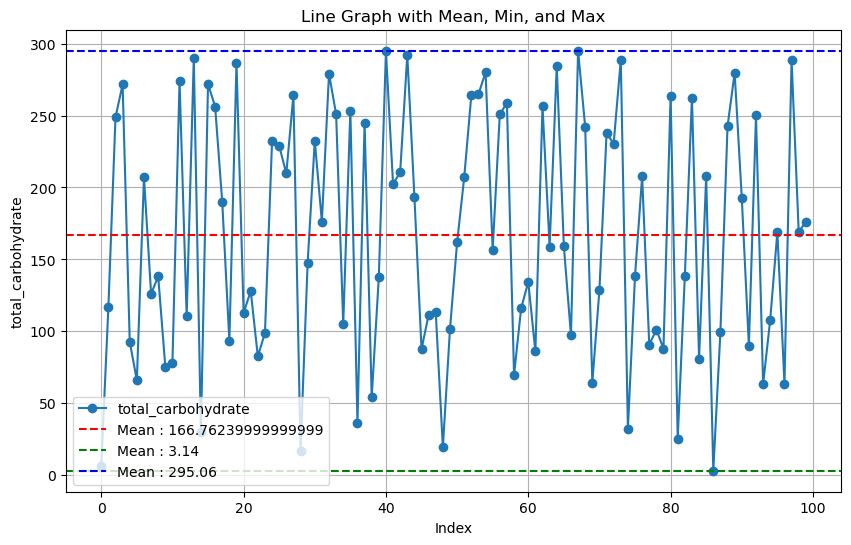
**1.3 HCA food recipes Table**

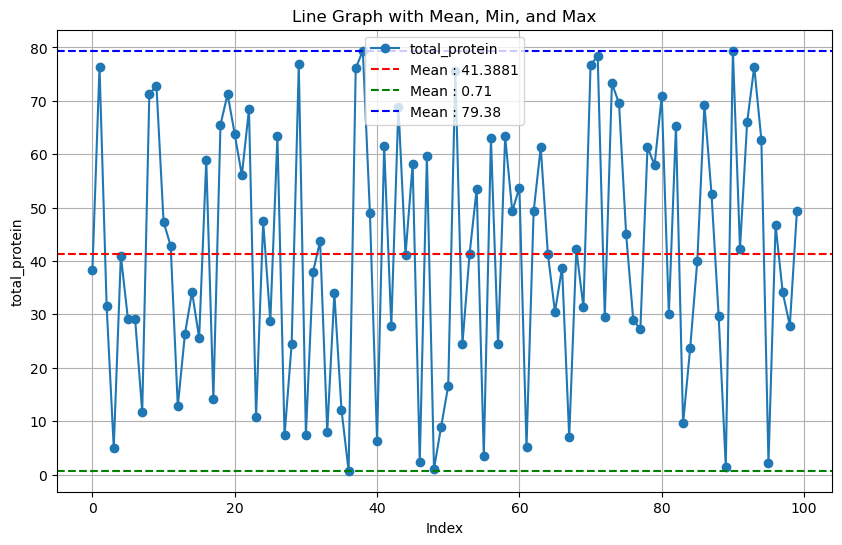
Check for null values: *null values were found in the column.* ***Micronutrient***

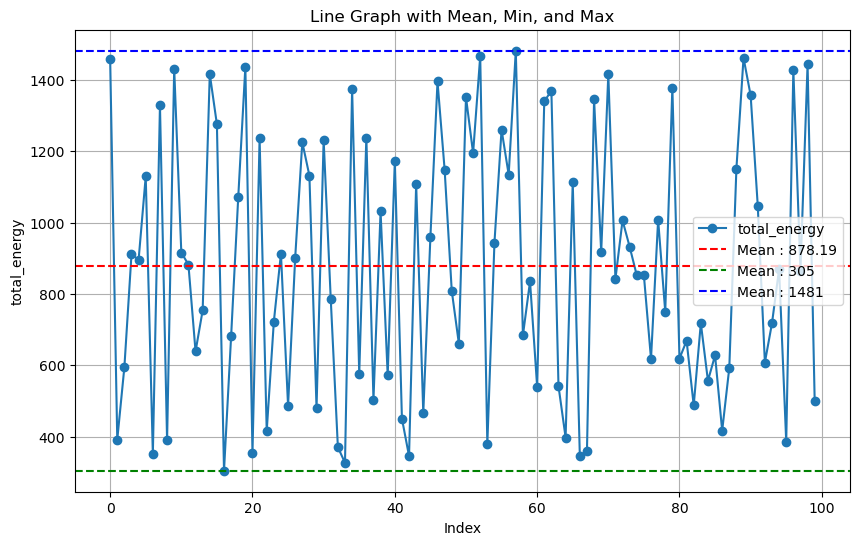
Numerical values:

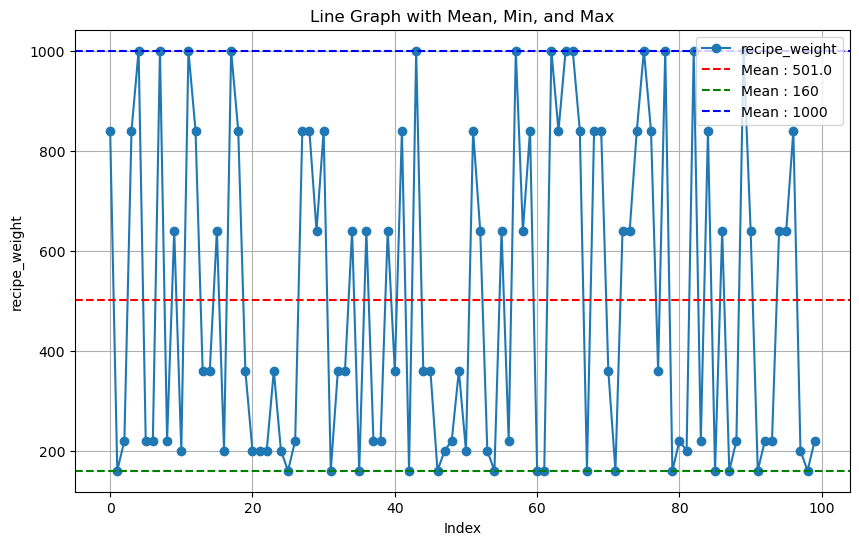


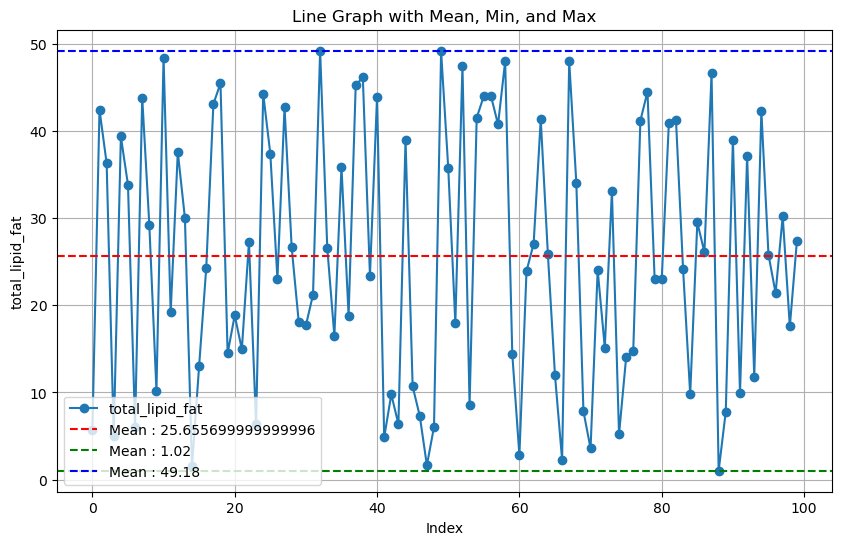
The mean serving\_size of the recipes is 169.0. Minimum is 110 and maximum is 220.











**1.4 HCA food log Table**

Check for null values: *No null values were found in this table.*

Categorical values:

| **Variable Name** | **Values** |
| --- | --- |
| food\_type | ['food\_item' 'recipe'] |
| item\_unit | ['1 medium' '1 slice' 'gm' 'ml'] |

