# Exercise 1

Step 1. Go to https://www.kaggle.com/openfoodfacts/world-food-facts/data

Step 2. Download the dataset to your computer and unzip it.

Step 3. Use the tsv file and assign it to a dataframe called food

```
In [2]:  import pandas as pd
         import numpy as np

         df = pd.read_table("D:/Courses/Projects/en.openfoodfacts.org.products.tsv")
         df
```

```
C:\Users\Hua\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3444: DtypeWarning: Columns (0,3,5,19,20,24,25,26,27,28,36,37,38,39,48) have mixed types.Specify dtype option on import or set low_memory=False.
  exec(code_obj, self.user_global_ns, self.user_ns)
```

Out[2]:

| | code | url | creator | created_t | created_datetime | last_modified_t | last_modified_datetime | product_name | generic_name | quantity | ... | fruits-vegetables-nuts_100g | fruits-vegetables-nuts-estimate_100g | collagen-meat-protein-ratio_100g | cocoa_100g | chlorophyl_100g | carbon-footprint_100g | nutrition-score-fr_100g | nutrition-score-uk_100g |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 3087 | http://world-en.openfoodfacts.org/product/0000... | openfoodfacts-contributors | 1474103866 | 2016-09-17T09:17:46Z | 1474103893 | 2016-09-17T09:18:13Z | Farine de blé noir | NaN | 1kg | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 1 | 4530 | http://world-en.openfoodfacts.org/product/0000... | usda-ndb-import | 1489069957 | 2017-03-09T14:32:37Z | 1489069957 | 2017-03-09T14:32:37Z | Banana Chips Sweetened (Whole) | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | 14.0 | 14.0 |
| 2 | 4559 | http://world-en.openfoodfacts.org/product/0000... | usda-ndb-import | 1489069957 | 2017-03-09T14:32:37Z | 1489069957 | 2017-03-09T14:32:37Z | Peanuts | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | 0.0 | 0.0 |
| 3 | 16087 | http://world-en.openfoodfacts.org/product/0000... | usda-ndb-import | 1489055731 | 2017-03-09T10:35:31Z | 1489055731 | 2017-03-09T10:35:31Z | Organic Salted Nut Mix | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | 12.0 | 12.0 |
| 4 | 16094 | http://world-en.openfoodfacts.org/product/0000... | usda-ndb-import | 1489055653 | 2017-03-09T10:34:13Z | 1489055653 | 2017-03-09T10:34:13Z | Organic Polenta | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 356022 | 99567450 | http://world-en.openfoodfacts.org/product/9956... | usda-ndb-import | 1489059076 | 2017-03-09T11:31:16Z | 1491244499 | 2017-04-03T18:34:59Z | Mint Melange Tea A Blend Of Peppermint, Lemon ... | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | 0.0 | 0.0 |
| 356023 | 9970229501521 | http://world-en.openfoodfacts.org/product/9970... | tomato | 1422099377 | 2015-01-24T11:36:17Z | 1491244499 | 2017-04-03T18:34:59Z | Leba pickle flavor potato chips | NaN | 50 g | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 356024 | 9977471758307 | http://world-en.openfoodfacts.org/product/9977... | openfoodfacts-contributors | 1497018549 | 2017-06-09T14:29:09Z | 1500730305 | 2017-07-22T13:31:45Z | Biscottes bio | NaN | 300g | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 356025 | 9980282863788 | http://world-en.openfoodfacts.org/product/9980... | openfoodfacts-contributors | 1492340089 | 2017-04-16T10:54:49Z | 1492340089 | 2017-04-16T10:54:49Z | Tomates aux Vermicelles | NaN | 67g | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 356026 | 999990026839 | http://world-en.openfoodfacts.org/product/9999... | usda-ndb-import | 1489072709 | 2017-03-09T15:18:29Z | 1491244499 | 2017-04-03T18:34:59Z | Sugar Free Drink Mix, Peach Tea | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

356027 rows × 163 columns

```
In [6]:  df.describe(include="all")
```

Out[6]:

| | code | url | creator | created_t | created_datetime | last_modified_t | last_modified_datetime | product_name | generic_name | quantity | ... | fruits-vegetables-nuts_100g | fruits-vegetables-nuts-estimate_100g | collagen-meat-protein-ratio_100g | cocoa_100g | chlorophyl_100g | carbon-footprint_100g | nutrition-score-fr_100g | nutrition-score-uk_100g |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 3.560010e+05 | 356001 | 356024 | 3.560240e+05 | 356017 | 3.560270e+05 | 356027 | 338515 | 57714 | 119285 | ... | 3228.000000 | 404.000000 | 182.000000 | 1383.000000 | 0.0 | 278.000000 | 254856.000000 | 254856.000000 |
| unique | 3.558390e+05 | 356001 | 3890 | 2.248120e+05 | 224752 | 2.169360e+05 | 216836 | 249245 | 42451 | 15563 | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| top | 7.065080e+10 | http://world-en.openfoodfacts.org/product/0000... | usda-ndb-import | 1.489077e+09 | 2017-03-09T10:37:09Z | 1.439142e+09 | 2015-08-09T17:35:42Z | Ice Cream | Pâtes alimentaires au blé dur de qualité supér... | 500 g | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| freq | 3.000000e+00 | 1 | 169868 | 2.000000e+01 | 20 | 3.000000e+01 | 30 | 411 | 201 | 5285 | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| mean | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | 33.392680 | 60.360124 | 15.362637 | 52.102675 | NaN | 335.790664 | 9.166137 | 8.98065 |
| std | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | 32.906834 | 29.262350 | 3.692658 | 19.028361 | NaN | 423.244871 | 8.999870 | 9.15175 |
| min | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | 0.000000 | 0.000000 | 8.000000 | 6.000000 | NaN | 0.000000 | -15.000000 | -15.000000 |
| 25% | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | 0.000000 | 45.000000 | 12.000000 | 33.000000 | NaN | 82.650000 | 1.000000 | 1.00000 |
| 50% | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | 25.000000 | 58.000000 | 15.000000 | 52.000000 | NaN | 190.950000 | 10.000000 | 9.00000 |
| 75% | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | 55.000000 | 93.000000 | 15.000000 | 70.000000 | NaN | 378.700000 | 16.000000 | 16.00000 |
| max | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | 100.000000 | 100.000000 | 25.000000 | 100.000000 | NaN | 2842.000000 | 40.000000 | 40.00000 |

11 rows × 163 columns

Step 4. See the first 5 entries

```
In [3]:  df.head(5)
```

Out[3]:

| | code | url | creator | created_t | created_datetime | last_modified_t | last_modified_datetime | product_name | generic_name | quantity | ... | fruits-vegetables-nuts_100g | fruits-vegetables-nuts-estimate_100g | collagen-meat-protein-ratio_100g | cocoa_100g | chlorophyl_100g | carbon-footprint_100g | nutrition-score-fr_100g | nutrition-score-uk_100g | glycemic-index_100g |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 3087 | http://world-en.openfoodfacts.org/product/0000... | openfoodfacts-contributors | 1474103866 | 2016-09-17T09:17:46Z | 1474103893 | 2016-09-17T09:18:13Z | Farine de blé noir | NaN | 1kg | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 1 | 4530 | http://world-en.openfoodfacts.org/product/0000... | usda-ndb-import | 1489069957 | 2017-03-09T14:32:37Z | 1489069957 | 2017-03-09T14:32:37Z | Banana Chips Sweetened (Whole) | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | 14.0 | 14.0 | NaN |
| 2 | 4559 | http://world-en.openfoodfacts.org/product/0000... | usda-ndb-import | 1489069957 | 2017-03-09T14:32:37Z | 1489069957 | 2017-03-09T14:32:37Z | Peanuts | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | 0.0 | 0.0 | NaN |
| 3 | 16087 | http://world-en.openfoodfacts.org/product/0000... | usda-ndb-import | 1489055731 | 2017-03-09T10:35:31Z | 1489055731 | 2017-03-09T10:35:31Z | Organic Salted Nut Mix | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | 12.0 | 12.0 | NaN |
| 4 | 16094 | http://world-en.openfoodfacts.org/product/0000... | usda-ndb-import | 1489055653 | 2017-03-09T10:34:13Z | 1489055653 | 2017-03-09T10:34:13Z | Organic Polenta | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

5 rows × 163 columns

Step 5. What is the number of observations in the dataset?

```
In [6]:  df.shape[0]
```

Out[6]:  356027

Step 6. What is the number of columns in the dataset?

```
In [8]:  df.shape[1]
```

Out[8]:  163

Step 7. Print the name of all the columns.

```
In [13]:  df.columns
```

```
Out[13]:  Index(['code', 'url', 'creator', 'created_t', 'created_datetime',
         'last_modified_t', 'last_modified_datetime', 'product_name',
         'generic_name', 'quantity',
         ...
         'fruits-vegetables-nuts_100g', 'fruits-vegetables-nuts-estimate_100g',
         'collagen-meat-protein-ratio_100g', 'cocoa_100g', 'chlorophyl_100g',
         'carbon-footprint_100g', 'nutrition-score-fr_100g',
         'nutrition-score-uk_100g', 'glycemic-index_100g',
         'water-hardness_100g'],
        dtype='object', length=163)
```

Step 8. What is the name of 105th column?

```
In [17]:  df.columns[105]
```

Out[17]:  '-fructose_100g'

Step 9. What is the type of the observations of the 105th column?

```
In [20]:  df.dtypes[105]
```

Out[20]:  dtype('float64')

Step 10. How is the dataset indexed?

```
In [21]:  df.index
```

Out[21]:  RangeIndex(start=0, stop=356027, step=1)

Step 11. What is the product name of the 19th observation?

```
In [23]:  df["product_name"][105]
```

Out[23]:  'Split Pea Soup Mix'

**The END**