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 [1]:
```
import pandas as pd
import numpy as np
food = pd.read_csv('en.openfoodfacts.org.products.tsv', sep='\t')
```

```
C:\Users\Yasin\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3165: 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.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

Step 4. See the first 5 entries

In [2]:
```
food.head()
```

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
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
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
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
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
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

5 rows × 163 columns

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

In [3]:
```
food.shape
```

Out[3]:
```
(356027, 163)
```

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

In [4]:
```
food.shape
```

Out[4]:
```
(356027, 163)
```

Step 7. Print the name of all the columns.

In [5]:
```
food.columns
```

Out[5]:
```
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 [6]:
```
food.columns[104]
```

Out[6]:
```
'-glucose_100g'
```

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

In [7]:
```
food.dtypes["-glucose_100g"]
```

Out[7]:
```
dtype('float64')
```

Step 10. How is the dataset indexed?

In [8]:
```
food.index
```

Out[8]:
```
RangeIndex(start=0, stop=356027, step=1)
```

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

In [9]:
```
food["product_name"][18]
```

Out[9]:
```
'Lotus Organic Brown Jasmine Rice'
```