19CSE305 ML - Pandas Exercise Questions - 30th Sep 2022

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

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

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

```
import pandas as pd

tsv_file='/content/en.openfoodfacts.org.products.tsv'

food=pd.read_csv(tsv_file,sep='\t')
```

Task 4. See the first 5 entries

```
food.head(5)
```

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

```
print(len(food))
print(food.shape)
```

```
1588
(1588, 163)
```

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

3	16087	[illegible]	[illegible]	[illegible]	1.489056e+09	[illegible]

```
print(len(food.columns))
```

```
163
```

5 rows × 163 columns

Task 7. Print the name of all the columns.

```
print(food.columns)
```

```
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)
```

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

```
name = food.columns[105]
name
```

```
'-fructose_100g'
```

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

```
print(type(name))
```

```
<class 'str'>
```

Task 10. How is the dataset indexed?

```
print(food.index)
```

```
RangeIndex(start=0, stop=1588, step=1)
```

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

```
print(food['product_name'][19])
```

```
Organic Oat Groats
```

Task 12. How many columns in the dataset tend have null values?

```
print(food.isna().all(axis=0).sum())
```

```
64
```

Task 13. Find the number of null values present under each column?

```
r = pd.DataFrame(food.isnull().sum())
r.columns = ['Count']
r
```

	Count
code	0
url	0

Task 14. Create a dictionary with keys to be column names in the dataset and values to number of values under each column. Print the dictionary by iterating through its elements using a for loop.

```
out = food.count().to_dict()
for key,value in out.items():
    print(key,"->",value)
```

```
-fructose_100g -> 0
-lactose_100g -> 0
-maltose_100g -> 0
-maltodextrins_100g -> 0
starch_100g -> 2
polyols_100g -> 0
fiber_100g -> 1070
proteins_100g -> 1398
casein_100g -> 0
serum-proteins_100g -> 0
nucleotides_100g -> 0
salt_100g -> 1313
sodium_100g -> 1313
alcohol_100g -> 3
vitamin-a_100g -> 831
beta-carotene_100g -> 0
vitamin-d_100g -> 25
vitamin-e_100g -> 0
vitamin-k_100g -> 10
vitamin-c_100g -> 832
vitamin-b1_100g -> 28
vitamin-b2_100g -> 28
vitamin-pp_100g -> 37
vitamin-b6_100g -> 24
vitamin-b9_100g -> 7
folates_100g -> 10
vitamin-b12_100g -> 23
biotin_100g -> 0
pantothenic-acid_100g -> 19
silica_100g -> 0
bicarbonate_100g -> 0
potassium_100g -> 85
chloride_100g -> 0
calcium_100g -> 899
phosphorus_100g -> 13
iron_100g -> 944
magnesium_100g -> 10
zinc_100g -> 9
copper_100g -> 2
```

```
manganese_100g -> 9
fluoride_100g -> 0
selenium_100g -> 8
chromium_100g -> 0
molybdenum_100g -> 0
iodine_100g -> 0
caffeine_100g -> 0
taurine_100g -> 0
ph_100g -> 0
fruits-vegetables-nuts_100g -> 3
fruits-vegetables-nuts-estimate_100g -> 1
collagen-meat-protein-ratio_100g -> 0
cocoa_100g -> 2
chlorophyl_100g -> 0
carbon-footprint_100g -> 0
nutrition-score-fr_100g -> 978
nutrition-score-uk_100g -> 978
glycemic-index_100g -> 0
water-hardness_100g -> 0
```

Task 15. Display the values associated with all columns from row 108 to 208?

```
for i in range(108,208):
    print(" ",i," ",food.iloc[i])
```

```
  108   code                                                        36252
url                        http://world-en.openfoodfacts.org/product/0000...
creator                                                           tacinte
created_t                                                  1422221701.0
created_datetime                                       2015-01-25T21:35:01Z
                                                  ...
carbon-footprint_100g                                              NaN
nutrition-score-fr_100g                                            NaN
nutrition-score-uk_100g                                            NaN
glycemic-index_100g                                                NaN
water-hardness_100g                                                NaN
Name: 108, Length: 163, dtype: object
  109   code                                                        36269
url                        http://world-en.openfoodfacts.org/product/0000...
creator                                                    usda-ndb-import
created_t                                                  1489055651.0
created_datetime                                       2017-03-09T10:34:11Z
                                                  ...
carbon-footprint_100g                                              NaN
nutrition-score-fr_100g                                           23.0
nutrition-score-uk_100g                                           23.0
glycemic-index_100g                                                NaN
water-hardness_100g                                                NaN
Name: 109, Length: 163, dtype: object
  110   code                                                        36276
url                        http://world-en.openfoodfacts.org/product/0000...
creator                                                    usda-ndb-import
created_t                                                  1489055678.0
```

```
created_datetime                                    2017-03-09T10:34:38Z
                                                    ...
carbon-footprint_100g                                               NaN
nutrition-score-fr_100g                                            15.0
nutrition-score-uk_100g                                            15.0
glycemic-index_100g                                                 NaN
water-hardness_100g                                                 NaN
Name: 110, Length: 163, dtype: object
  111   code                                                        36986
url                       http://world-en.openfoodfacts.org/product/0000...
creator                                                  usda-ndb-import
created_t                                                   1489055649.0
created_datetime                                    2017-03-09T10:34:09Z
                                                    ...
carbon-footprint_100g                                               NaN
nutrition-score-fr_100g                                            30.0
nutrition-score-uk_100g                                            30.0
glycemic-index_100g                                                 NaN
water-hardness_100g                                                 NaN
Name: 111, Length: 163, dtype: object
  112   code                                                        38058
url                       http://world-en.openfoodfacts.org/product/0000...
creator                                                  usda-ndb-import
created_t                                                   1489055651.0
created_datetime                                    2017-03-09T10:34:11Z
                                                    ...
carbon-footprint_100g                                               NaN
nutrition-score-fr_100g                                             NaN
nutrition-score-uk_100g                                             NaN
glycemic-index_100g                                                 NaN
```

✓ 0s completed at 11:12 AM