# Hands-on Activity 10.1 Data Analysis using Python

## CPE311 - Computational Thinking with Python

Name: Dela Cruz, Gabrielle

Section: CPE22S3

Performed on: 07/09/2024

Submitted on: 07/xx/2024

Submitted to: Engr. Roman M. Richard

Link to Colab: https://colab.research.google.com/drive/1_w47gncSFB8cxw1yx1pSlR1-p_s9bBBD?usp=sharing

### Intended Learning Outcome

1. Perform descriptive and correlation analysis to to analyze the dataset.
2. Interpret the results of descriptive and correlation analysis

### Resources

- Personal Computer
- Jupyter Notebook
- Internet Connection

### Instruction

1. Gather a dataset regarding your identified problem for the ASEAN Data Science Explorer. Make sure that the dataset includes multiple variables.
2. Load the dataset into pandas dataframe.
3. Prepare the data by applying appropriate data preprocessing techniques.
4. Analyze the data using descriptive analysis.
5. Perform correlation analysis.
6. Interpret the results based on the descriptive and correlation analysis.
7. Submit the PDF file.

#### Dataset gathered from UNICEF Dataset:

Diet (6-23 months): https://data.unicef.org/resources/dataset/diets/

Diet (6-23 months): https://data.unicef.org/resources/dataset/diets/

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

diet = pd.read_csv('/content/UNICEF_Expanded_Global_Databases_Diets_6_23months_2023.csv')
diet
```

| | Regions | Abbreviations | Dates | Egg and/or flesh food consumption | Minimum dietary diversity | Minimum meal frequency | Minimum acceptable diet |
|---|---|---|---|---|---|---|---|
| **0** | East Asia and the Pacific | EAP | 2016-2022 | 74 | 55 | 72 | 41 |
| **1** | Eastern Europe and Central Asia | EECA | 2016-2022 | - | - | - | - |
| **2** | Eastern and Southern Africa | ESA | 2016-2022 | 35 | 20 | 50 | 14 |
| **3** | Latin America & the Caribbean | LAC | 2016-2022 | 74 | 62 | - | - |
| **4** | Middle East and North Africa | MENA | 2016-2022 | 60 | 46 | 60 | 28 |
| **5** | North America | NaN | 2016-2022 | - | - | - | - |
| **6** | South Asia | SA | 2016-2022 | 29 | 24 | 42 | 13 |
| **7** | West and Central Africa | WCA | 2016-2022 | 46 | 25 | 37 | 11 |
| **8** | Western Europe | WE | 2016-2022 | - | - | - | - |
| **9** | Australia and New Zealand | NaN | 2016-2022 | - | - | - | - |
| **10** | Central Asia and Southern | NaN | 2016-2022 | 30 | 27 | 43 | 14 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Southern Asia | | | | | | |
| **11** | Eastern Asia and South-eastern Asia | NaN | 2016-2022 | 74 | 56 | 72 | 41 |
| **12** | Latin America and the Caribbean | NaN | 2016-2022 | 74 | 62 | - | - |
| **13** | Northern America and Europe | NaN | 2016-2022 | - | - | - | - |
| **14** | Oceania excluding Australia and New Zealand | NaN | 2016-2022 | 44 | 34 | 46 | 19 |
| **15** | Sub-Saharan Africa | NaN | 2016-2022 | 41 | 23 | 43 | 12 |
| **16** | Western Asia and Northern Africa | NaN | 2016-2022 | - | 39 | - | - |

Next steps: Generate code with `diet` View recommended plots

```
diet.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40 entries, 0 to 39
Data columns (total 7 columns):
 #   Column                              Non-Null Count  Dtype 
---  ------                              --------------  ----- 
 0   Regions                             40 non-null     object
 1   Abbreviations                       8 non-null      object
 2   Dates                               40 non-null     object
 3   Egg and/or flesh food consumption   40 non-null     object
 4   Minimum dietary diversity           40 non-null     object
 5   Minimum meal frequency              40 non-null     object
 6   Minimum acceptable diet             40 non-null     object
dtypes: object(7)
memory usage: 2.3+ KB
```

```
# Sorting the 20 highest "Egg and/or Flesh Food Consumption" values by head() function.
diet.sort_values(by='Egg and/or flesh food consumption', ascending=False).head(20)
```

```
# Sorting the lower 20s of the same column by using the nsmallest() function.
diet['Egg and/or flesh food consumption'] = pd.to_numeric(diet['Egg and/or flesh food con
diet.nsmallest(20, 'Egg and/or flesh food consumption')
```

```
# Getting the higher 20s of the Minimum Dietary Diversity.
diet['Minimum dietary diversity'] = pd.to_numeric(diet['Minimum dietary diversity'], erro
diet.nlargest(20, 'Minimum dietary diversity')
```

```
# Getting the lower 20s of the Minimum Dietary Diversity.
diet.nsmallest(20, 'Minimum dietary diversity')
```

```
# Getting the higher 20s of the Mminimum meal frequency.
diet['Minimum meal frequency'] = pd.to_numeric(diet['Minimum meal frequency'], errors='co
diet.nlargest(20, 'Minimum meal frequency')
```

```
# Getting the lower 20s of the Minimum meal frequency.
diet.nsmallest(20, 'Minimum meal frequency')
```

```
# Getting the higher 20s of the Minimum acceptable diet.
diet['Minimum acceptable diet'] = pd.to_numeric(diet['Minimum acceptable diet'], errors='
diet.nlargest(20, 'Minimum acceptable diet')
```

```
diet.nlargest(20, 'Minimum acceptable diet')
```

```
# Getting the lower 20s of the Minimum acceptable diet.
diet['Minimum acceptable diet'] = pd.to_numeric(diet['Minimum acceptable diet'], errors='
diet.nsmallest(20, 'Minimum acceptable diet')
```

```
# Measures of Central Tendency regarding the Egg and/or flesh food consumption of the die
print('Mean of the Egg and/or flesh food consumption:',diet['Egg and/or flesh food consum
print('Median of the Egg and/or flesh food consumption:',diet['Egg and/or flesh food cons
print('Mode of the Egg and/or flesh food consumption:',diet['Egg and/or flesh food consum
```

```
Mean of the Egg and/or flesh food consumption: 54.285714285714285
Median of the Egg and/or flesh food consumption: 57.5
Mode of the Egg and/or flesh food consumption: 0    74.0
Name: Egg and/or flesh food consumption, dtype: float64
```

```
# Measures of Central Tendency regarding the Minimum dietary diversity of the diet datase
print('Mean of the Minimum dietary diversity:',diet['Minimum dietary diversity'].mean())
print('Median of the Minimum dietary diversity:',diet['Minimum dietary diversity'].median
print('Mode of the Minimum dietary diversity:',diet['Minimum dietary diversity'].mode())
```

```
Mean of the Minimum dietary diversity: 38.43333333333333
Median of the Minimum dietary diversity: 36.5
Mode of the Minimum dietary diversity: 0    62.0
Name: Minimum dietary diversity, dtype: float64
```

```
# Measures of Central Tendency regarding the Minimum meal frequency of the diet dataset.
print('Mean of the Minimum meal frequency:',diet['Minimum meal frequency'].mean())
print('Median of the Minimum meal frequency:',diet['Minimum meal frequency'].median())
print('Mode of the Minimum meal frequency:',diet['Minimum meal frequency'].mode())
```

```
Mean of the Minimum meal frequency: 52.541666666666664
Median of the Minimum meal frequency: 48.5
Mode of the Minimum meal frequency: 0    72.0
Name: Minimum meal frequency, dtype: float64
```

```
# Measures of Central Tendency regarding the Minimum acceptable diet of the diet dataset.
print('Mean of the Minimum acceptable diet:',diet['Minimum acceptable diet'].mean())
print('Median of the Minimum acceptable diet:',diet['Minimum acceptable diet'].median())
print('Mode of the Minimum acceptable diet:',diet['Minimum acceptable diet'].mode())
```

```
Mean of the Minimum acceptable diet: 21.26086956521739
Median of the Minimum acceptable diet: 16.0
Mode of the Minimum acceptable diet: 0    14.0
Name: Minimum acceptable diet, dtype: float64
```

```
# Measures of Dispersion regarding the Egg and/or flesh food consumption.
eggConsumption = diet['Egg and/or flesh food consumption']
range = eggConsumption.max() - eggConsumption.min()
print('The range of the Egg and/or flesh food consumption: ', range)

standardDeviation = eggConsumption.std()
print('The standard deviation of the Egg and/or flesh food consumption: ', standardDeviat

variance = eggConsumption.var()
print('The variance of the Egg and/or flesh food consumption: ', variance)

interquartileRange = eggConsumption.quantile(0.75) - eggConsumption.quantile(0.25)
print('The interquartile range of the Egg and/or flesh food consumption: ', interquartile
```

```
The range of the Egg and/or flesh food consumption:  47.0
The standard deviation of the Egg and/or flesh food consumption:  15.923095601828608
The variance of the Egg and/or flesh food consumption:  253.54497354497357
The interquartile range of the Egg and/or flesh food consumption:  28.75
```

```
# Measures of Dispersion regarding the Minimum dietary diversity.
minimumDietaryDiversity = diet['Minimum dietary diversity']
range = minimumDietaryDiversity.max() - minimumDietaryDiversity.min()
print('The range of the Minimum dietary diversity: ', range)

standardDeviation = minimumDietaryDiversity.std()
print('The standard deviation of the Minimum dietary diversity: ', standardDeviation)

variance = minimumDietaryDiversity.var()
print('The variance of the Minimum dietary diversity: ', variance)
```

```
interquartileRange = minimumDietaryDiversity.quantile(0.75) - minimumDietaryDiversity.qua
print('The interquartile range of the Minimum dietary diversity: ', interquartileRange)
```

```
The range of the Minimum dietary diversity:  51.0
The standard deviation of the Minimum dietary diversity:  15.31662007672426
The variance of the Minimum dietary diversity:  234.59885057471266
The interquartile range of the Minimum dietary diversity:  25.5
```

```
# Measures of Dispersion regarding the Minimum meal frequency.
minimumMealFrequency = diet['Minimum meal frequency']
range = minimumMealFrequency.max() - minimumMealFrequency.min()
print('The range of the Minimum meal frequency: ', range)

standardDeviation = minimumMealFrequency.std()
print('The standard deviation of the Minimum meal frequency: ', standardDeviation)

variance = minimumMealFrequency.var()
print('The variance of the Minimum meal frequency: ', variance)

interquartileRange = minimumMealFrequency.quantile(0.75) - minimumMealFrequency.quantile(
print('The interquartile range of the Minimum meal frequency: ', interquartileRange)
```

```
The range of the Minimum meal frequency:  36.0
The standard deviation of the Minimum meal frequency:  12.161876166285282
The variance of the Minimum meal frequency:  147.91123188405797
The interquartile range of the Minimum meal frequency:  18.5
```

```
# Measures of Dispersion regarding the Minimum acceptable diet.
minimumAcceptableDiet = diet['Minimum acceptable diet']
range = minimumAcceptableDiet.max() - minimumAcceptableDiet.min()
print('The range of the Minimum acceptable diet: ', range)

standardDeviation = minimumAcceptableDiet.std()
print('The standard deviation of the Minimum acceptable diet: ', standardDeviation)

variance = minimumAcceptableDiet.var()
print('The variance of the Minimum acceptable diet: ', variance)

interquartileRange = minimumAcceptableDiet.quantile(0.75) - minimumAcceptableDiet.quantil
print('The interquartile range of the Minimum acceptable diet: ', interquartileRange)
```

```
The range of the Minimum acceptable diet:  33.0
The standard deviation of the Minimum acceptable diet:  11.221802607199107
The variance of the Minimum acceptable diet:  125.92885375494069
The interquartile range of the Minimum acceptable diet:  14.0
```

## Results and Findings of the Data

After treating the dataset with descriptive statistics for determining the positions of data, it can be inferred there is high variability across the regions analyzed, with the Minimum Dietary

be inferred there is high variability across the regions analyzed, with the Minimum Dietary Diversity notably low in areas (especially in ESA, compared to LAC). Additionally, ESA is seen to have the lowest value for all indicators present in the dataset, denoting significant nutrition challenges for this region alone. In comparison, regions like EAP and LAC have relatively higher values for Egg and/or Flesh Food Consumption and Minimum Dietary Diversity, signifying better nutrition statuses for these regions.