

ALI ZIA KHAN

PIAIC AI BATCH 07

CNC022080

NUMPY TASK 2 SOLVED

CrunchieMunchies

You work in marketing for a food company **myCorps**, which is developing a new kind of tasty, wholesome cereal called **CrunchieMunchies**.

You want to demonstrate to consumers how healthy your cereal is in comparison to other leading brands, so you've dug up nutritional data on several different competitors.

Your task is to use *NumPy statistical calculations* to analyze this data and prove that your **CrunchieMunchies** is the healthiest choice for consumers.

Task STEPS

1.First, import numpy.

```
In [1]: import numpy as np
```

2.Look over the **cereal.csv** file. This file contains the reported calorie amounts for different cereal brands. Load the data from the file and save it as **calorie_stats**.

```
In [2]: import csv
with open("cereal.csv", 'r') as f:
    data = list(csv.reader(f, delimiter=","))
    data = np.ravel(np.array(data, dtype=np.int))

print(data)

[ 70 120  70  50 110 110 110 130  90  90 120 110 120 110 110 110 100 110
 110 110 100 110 100 100 110 110 100 120 120 110 100 110 100 110 120 120
 110 110 110 140 110 100 110 100 150 150 160 100 120 140  90 130 120 100
  50  50 100 100 120 100  90 110 110  80  90  90 110 110  90 110 140 100
 110 110 100 100 110]
```

3.There are 60 calories per serving of CrunchieMunchies. How much **higher** is the **average calorie count** of your competition?

Save the answer to the variable **average_calories** and print the variable to the terminal to see the answer.

```
In [3]: average_calories=np.average(data)-60
print(average_calories)

46.883116883116884
```

4.Does the **average calorie count** adequately reflect the distribution of the dataset? Let's sort the data and see.

Sort the data and save the result to the variable **calorie_stats_sorted**. Print the sorted data to the terminal.

```
In [4]: calorie_stats_sorted=np.sort(data)
print(calorie_stats_sorted)

[ 50  50  50  70  70  80  90  90  90  90  90  90  90  90 100 100 100 100 100
 100 100 100 100 100 100 100 100 100 100 110 110 110 110 110 110 110
 110 110 110 110 110 110 110 110 110 110 110 110 110 110 110 110 110
 110 110 110 110 110 120 120 120 120 120 120 120 120 120 120 130 130 140
 140 140 150 150 160]
```

5.Do you see what I'm seeing? Looks like **the majority of the cereals are higher than the mean**. Let's see if the **median** is a better representative of the dataset.

Calculate the median of the dataset and save your answer to **median_calories**. Print the median so you can see how it compares to the mean.

```
In [5]: median_calories=np.median(data)
print(median_calories)
print(median_calories-average_calories)

110.0
63.116883116883116
```

6.While the median demonstrates that **"at least half of our values are over 100 calories"**, it would be more impressive to show that a significant portion of the competition has a higher calorie count than CrunchieMunchies.

Calculate different percentiles and print them to the terminal until you find the lowest percentile that is greater than 60 calories. Save this value to the variable **nth_percentile**.

```
In [6]: nth_percentile=np.percentile(data,4)
print(nth_percentile)

70.0
```

```
In [ ]:
```

7.While the percentile shows us that**"the majority of the competition has a much higher calorie count"**, it's an awkward concept to use in marketing materials.

Instead, let's calculate the percentage of cereals that **"have more than 60 calories per serving"**. Save your answer to the variable **more_calories** and print it to the terminal

```
In [7]: more_calories=(74/77)*100
print(more_calories)

96.1038961038961
```

8.Wow! That's a really high percentage. That's going to be very useful when we promote CrunchieMunchies. But one question is, **how much variation exists in the dataset?** Can we make the generalization that most cereals have around 100 calories or is the spread even greater?

Calculate the amount of variation by finding the **standard deviation** Save your answer to **calorie_std** and print to the terminal. How can we incorporate this value into our analysis?

```
In [8]: calorie_std=np.std(data)
print(calorie_std)

19.35718533390827
```

9.Write a short paragraph that sums up your findings and how you think this data could be used to myCorp's advantage when marketing CrunchieMunchies.

```
In [9]: print("Summing up our findings we can say that we find out average value , but we thought that since we have to show that in market we should do so with some
Summing up our findings we can say that we find out average value , but we thought that since we have to show that in market we should do so with some more re
flecting figures thenwe took out median and consequentlt percentile, but still it could be difficult for common people to let understand sosimply we calculate
d percentage as everyone generally have knowhow about that and the such high percentage can easily help us to showcase the point we want to make and get out
most from market by competeting with others using stastical data.
```

```
In [ ]:
```