Task One

Welcome, I'm hoping all is well with you.

Congratulations on reaching the advanced step of AI. Let's quickly revise **"Exploratory Data Analysis" (EDA)** as it is an essential skill every AI engineer must have.

Task:

- Create a repository in Github named **IEEE CS24 ZSB AI Advanced**, then create a folder named **"EDA Revision"**
- This is where you will upload both .py files of the EDA revision.

The 1st part is a revision on Statistics:

part1.py:
https://drive.google.com/file/d/13Qaha4AN5WXXFjyGcS7ngT9pwVkDANmm/view?usp=sharing

Create a function named `calculate()` in `part1.py` that uses Numpy to output the mean, variance, standard deviation, max, min, and sum of the rows, columns, and elements in a 3 x 3 matrix.

The input of the function should be a list containing 9 digits. The function should convert the list into a 3 x 3 Numpy array, and then return a dictionary containing the mean, variance, standard deviation, max, min, and sum along both axes and for the flattened matrix. **In other words, calculate the mean, variance, standard deviation, max, min, and sum 3 times. One time for each column, 2nd time for each row, and a 3rd time for all elements.**

The returned dictionary should follow this format:

```
{
```

```
  'mean': [axis1, axis2, flattened],

  'variance': [axis1, axis2, flattened],

  'standard deviation': [axis1, axis2, flattened],

  'max': [axis1, axis2, flattened],

  'min': [axis1, axis2, flattened],

  'sum': [axis1, axis2, flattened]

}
```

If a list containing less than 9 elements is passed into the function, it should raise a `ValueError` exception with the message: "List must contain nine numbers." The values in the returned dictionary should be lists and not Numpy arrays.

For example, `calculate([0,1,2,3,4,5,6,7,8])`

should return:

```
{

  'mean': [[3.0, 4.0, 5.0], [1.0, 4.0, 7.0], 4.0],

  'variance': [[6.0, 6.0, 6.0], [0.6666666666666666,
0.6666666666666666, 0.6666666666666666], 6.666666666666667],

  'standard deviation': [[2.449489742783178, 2.449489742783178,
2.449489742783178], [0.816496580927726, 0.816496580927726,
0.816496580927726], 2.581988897471611],

  'max': [[6, 7, 8], [2, 5, 8], 8],

  'min': [[0, 1, 2], [0, 3, 6], 0],

  'sum': [[9, 12, 15], [3, 12, 21], 36]

}
```

The 2nd part is a revision on Data Analysis using pandas:

Dataset:
https://drive.google.com/file/d/1WJjEh8_rgD-UT0icWA6q0UHChbE9yRHY/view?usp=sharing

Python File:
https://drive.google.com/file/d/1Y63gOuZRRrri3uxfo9SFOixlkyXSDevK/view?usp=sharing

- In this challenge you must analyze demographic data using Pandas. You are given a dataset of demographic data that was extracted from the 1994 Census database.

You must use Pandas to answer the following questions:

- How many people of each race are represented in this dataset? This should be a Pandas series with race names as the index labels. (`race` column)
- What is the average age of men?
- What is the percentage of people who have a Bachelor's degree?
- What percentage of people with advanced education (`Bachelors`, `Masters`, or `Doctorate`) make more than 50K?
- What percentage of people without advanced education make more than 50K?
- What is the minimum number of hours a person works per week?
- What percentage of the people who work the minimum number of hours per week have a salary of more than 50K?
- What country has the highest percentage of people that earn >50K and what is that percentage?
- Identify the most popular occupation for those who earn >50K in India.

Use the starter code in the file `part2.py`

Update the code so all variables set to "None" are set to the appropriate calculation or code. Round all decimals to the nearest tenth.

Deadline: 6/2/2024 11:59 PM