🟠

II

Question 1 - Pandas & JSON

You are working on an API in python that will be used by a web app and will also be used by other people in your organization. One of the endpoint accepts GET requests and should return a list of `Teachers` and each teacher should contain a list of `Students`……..see the desired outcome below under **“The desired outcome”**

The data

Initially, you have 2 dataframes in python, (you have no control of where this data comes from), a teachers dataframe (`df_teacher`) and a students dataframe (`df_student`)

✓ df_teacher ···

	name	married	school
0	Pep Guardiola	True	Manchester High School
1	Jurgen Klopp	True	Liverpool High School
2	Mikel Arteta	False	Arsenal High
3	Zinadine Zidane	True	NaN

✓ df_student ···

	teacher	name	age	height
0	Mikel Arteta	Bukayo Saka	21	2.1m
1	Mikel Arteta	Gabriel Martinelli	21	2.1m
2	Pep Guardiola	Jack Grealish	27	2.1m
3	Jurgen Klopp	Roberto Firmino	31	2.1m
4	Jurgen Klopp	Andrew Robertson	28	2.1m
5	Jurgen Klopp	Darwin Nunez	23	2.1m
6	Pep Guardiola	Ederson Moraes	29	2.1m
7	Pep Guardiola	Manuel Akanji	27	2.1m
8	Mikel Arteta	Thomas Partey	29	2.1m

Create these dataframes in python as follows:

```
import pandas as pd #dependency
import numpy as np #dependency

df_teacher = pd.DataFrame({
    "name": ["Pep Guardiola", "Jurgen Klopp", "Mikel Arteta", "Zinadine Zidane"],
    "married": [True, True, False, True],
```

```
    "school": ["Manchester High School", "Liverpool High School", "Arsenal High", np.n
an]
})


df_student = pd.DataFrame({
    "teacher": ["Mikel Arteta", "Mikel Arteta", "Pep Guardiola", "Jurgen Klopp", "Jurg
en Klopp", "Jurgen Klopp", "Pep Guardiola","Pep Guardiola","Mikel Arteta"],
    "name": ["Bukayo Saka", "Gabriel Martinelli", "Jack Grealish", "Roberto Firmino",
 "Andrew Robertson", "Darwin Nunez", "Ederson Moraes", "Manuel Akanji", "Thomas Parte
y"],
    "age": [21, 21, 27, 31, 28, 23, 29, 27, 29],
    "height": ['2.1m','2.1m', '2.1m', '2.1m', '2.1m', '2.1m', '2.1m', '2.1m', '2.1m']

})
```

Do not alter the original dataframe for the final solution. feel free to create a copy and then you alter that one for any debugging you need

You will obviously need pandas and numpy. You can use any other BUILT IN python libraries to solve the solution…*Try* not to use any other 3rd party libraries besides pandas and python.

PART A

Write python code to produce the following output using the above dataframes.

The desired outcome

```
[
    {
        "teacher": "Jurgen Klopp",
        "school": "Liverpool High School",
        "married": true,
        "Students": [
            {
                "student": "Roberto Firmino",
                "age": 31.0,
                "height": "2.1m"
            },
            {
                "student": "Andrew Robertson",
                "age": 28.0,
                "height": "2.1m"
            },
            {
                "student": "Darwin Nunez",
                "age": 23.0,
                "height": "2.1m"
            }
        ]
    },
    {
        "teacher": "Mikel Arteta",
        "school": "Arsenal High",
        "married": false,
        "Students": [
            {
                "student": "Bukayo Saka",
                "age": 21.0,
                "height": "2.1m"
            },
            {
                "student": "Gabriel Martinelli",
                "age": 21.0,
                "height": "2.1m"
            },
            {
                "student": "Thomas Partey",
                "age": 29.0,
                "height": "2.1m"
            }
        ]
    },
    {
        "teacher": "Pep Guardiola",
        "school": "Manchester High School",
        "married": true,
        "Students": [
            {
                "student": "Jack Grealish",
                "age": 27.0,
                "height": "2.1m"
            },
            {
                "student": "Ederson Moraes",
                "age": 29.0,
                "height": "2.1m"
            },
            {
                "student": "Manuel Akanji",
                "age": 27.0,
                "height": "2.1m"
            }
        ]
    }
]
```

PART B

You are told that the `Student` dataframe, df_student, can and will change in the near future, i.e more columns can be ADDED to the df_student, but **none** can be DELETED i.e `teacher` `name`, `age` and `height` will ALWAYS be there

Again, remember, you have no control of the original source of the data.

Any new columns added in df_student, should appear in the final results. for example, if the column `weight` is added as follows:

```
df_student = pd.DataFrame({
            "teacher": ["Mikel Arteta", "Mikel Arteta", "Pep Guardiola", "Jurgen Klop
p", "Jurgen Klopp", "Jurgen Klopp", "Pep Guardiola","Pep Guardiola","Mikel Arteta"],
            "name": ["Bukayo Saka", "Gabriel Martinelli", "Jack Grealish", "Roberto Fi
rmino", "Andrew Robertson", "Darwin Nunez", "Ederson Moraes", "Manuel Akanji", "Thomas
Partey"],
            "age": [21, 21, 27, 31, 28, 23, 29, 27, 29],
            "height": ['2.1m','2.1m', '2.1m', '2.1m', '2.1m', '2.1m', '2.1m', '2.1m',
 '2.1m'],
            "weight": ['80kg','70kg','690kg','73kg','60kg','70kg','80kg','88kg','74k
g',]

        })
```

The desired outcome should now be:

```
[
    {
        "teacher": "Jurgen Klopp",
        "school": "Liverpool High School",
        "married": true,
        "Students": [
            {
                "student": "Roberto Firmino",
                "age": 31.0,
                "height": "2.1m",
                "weight": "73kg"
            },
            {
                "student": "Andrew Robertson",
                "age": 28.0,
                "height": "2.1m",
                "weight": "60kg"
            },
            {
                "student": "Darwin Nunez",
                "age": 23.0,
                "height": "2.1m",
                "weight": "70kg"
            }
        ]
    },
    {
        "teacher": "Mikel Arteta",
        "school": "Arsenal High",
        "married": false,
        "Students": [
            {
                "student": "Bukayo Saka",
                "age": 21.0,
                "height": "2.1m",
                "weight": "80kg"
            },
            {
                "student": "Gabriel Martinelli",
                "age": 21.0,
                "height": "2.1m",
                "weight": "70kg"
            },
            {
                "student": "Thomas Partey",
                "age": 29.0,
                "height": "2.1m",
                "weight": "74kg"
            }
        ]
    },
    {
        "teacher": "Pep Guardiola",
        "school": "Manchester High School",
        "married": true,
        "Students": [
            {
                "student": "Jack Grealish",
                "age": 27.0,
                "height": "2.1m",
                "weight": "690kg"
            },
```

As you can see, weight is included in the outcome.

The catch is you do not know what column(s) is(are) going to be added, the columns could be called `speed`, `weight` or `salary` etc . Multiple columns can be added at once too and they should all be visible the final outcome for each `Student`. This is how this part will be assessed.

Update the code in PART A to cater for these requirements.

NB: if your code in part A already catered for this, well done. No need to do anything else.