Data Analytics Questions

You were given a 'practice_dataset.csv' dataset, that contains data about average salary of some school graduates. Please read in this file here, and explore it.

In [ ]:
```
#please code here
```

Task 1.

In the dataset there is a 'School Type' column that has numircal values: they are IDs for keys that are given in 'school_type.json' file. Please, map over these IDs to replace them with their keys. Here is an expected outcome:

From:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Career 90th Percentile Salary
0	Massachusetts Institute of Technology (MIT)	1	$72,200.00	$126,000.00	$76,800.00	$99,200.00	$168,000.00	$220,000.00
1	California Institute of Technology (CIT)	1	$75,500.00	$123,000.00	NaN	$104,000.00	$161,000.00	NaN
2	Harvey Mudd College	1	$71,800.00	$122,000.00	NaN	$96,000.00	$180,000.00	NaN
3	Polytechnic University of New York, Brooklyn	1	$62,400.00	$114,000.00	$66,800.00	$94,300.00	$143,000.00	$190,000.00
4	Cooper Union	1	$62,200.00	$114,000.00	NaN	$80,200.00	$142,000.00	NaN

To:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Career 90th Percentile Salary
0	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00	$76,800.00	$99,200.00	$168,000.00	$220,000.00
1	California Institute of Technology (CIT)	Engineering	$75,500.00	$123,000.00	NaN	$104,000.00	$161,000.00	NaN
2	Harvey Mudd College	Engineering	$71,800.00	$122,000.00	NaN	$96,000.00	$180,000.00	NaN
3	Polytechnic University of New York, Brooklyn	Engineering	$62,400.00	$114,000.00	$66,800.00	$94,300.00	$143,000.00	$190,000.00
4	Cooper Union	Engineering	$62,200.00	$114,000.00	NaN	$80,200.00	$142,000.00	NaN

In [ ]:
```
#please code here
```

Task 2

We defined a function that takes any 'School Type' value, and estimates rounded average 'Mid-Career Median Salary'for it. However, our funciton is not working. Please find an error and try to fix it.

In [ ]:
```
def function_1(school_type):
    result=round(df[df['School Type']==school_type]['Mid-Career Median Salary'].mean(),
    return result
```

In [ ]:
```
function_1('Engineering')
```

If you fix an error, apply this function to values 'Engineering', 'Party', 'Liberal Arts' and print output of the function, the end result must looks like this:

```
print(function_1('Engineering'))
print(function_1('Party'))
print(function_1('Liberal Arts'))
```

```
103842.11
84685.0
89378.72
```

In [ ]:
```
#please code here
```

Task 3

According to the National Occupational Employment and Wages Estimates, the average salary in the United States is 56,310 USD annaully. Iterate over 'Starting Median Salary' column and assign value 'more than national average' if it is more than 56,310 USD, else 'less than national average'. The result is supposed to be as such:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Career 90th Percentile Salary
0	Massachusetts Institute of Technology (MIT)	Engineering	more than national average	126000	$76,800.00	$99,200.00	$168,000.00	$220,000.00
1	California Institute of Technology (CIT)	Engineering	more than national average	123000	NaN	$104,000.00	$161,000.00	NaN
2	Harvey Mudd College	Engineering	more than national average	122000	NaN	$96,000.00	$180,000.00	NaN
3	Polytechnic University of New York, Brooklyn	Engineering	more than national average	114000	$66,800.00	$94,300.00	$143,000.00	$190,000.00
4	Cooper Union	Engineering	more than national average	114000	NaN	$80,200.00	$142,000.00	NaN

In [ ]:
```
#please code here
```

Now, display all state schools that have less than national average salary.

In [ ]:
```
#please code here
```

Task4

You might have realised that some columns have missing values. Display all rows that has at least one missing value in any column. Then, consider how would you handle these missing values? Please, describe below your thoughts

In [ ]:
```
#please opine here
```

In [ ]:
```
#please code here
```

Task 5

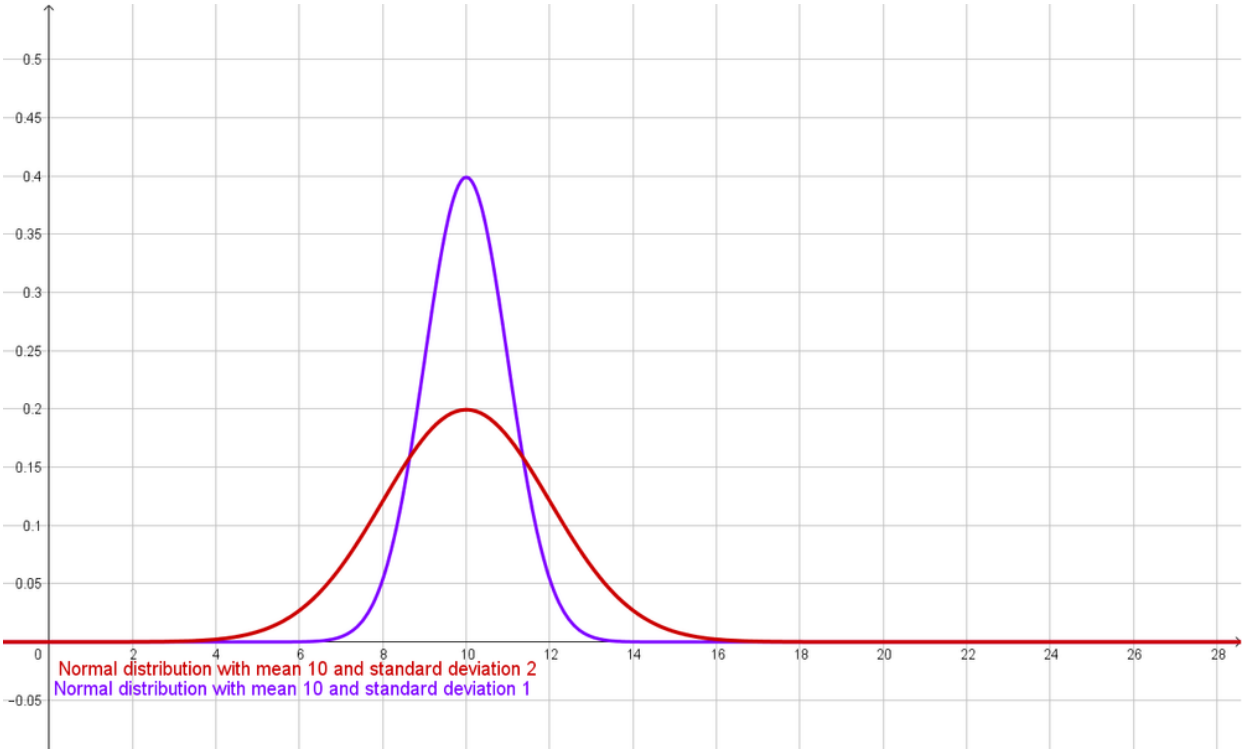
Please visit this web page: https://en.wikipedia.org/wiki/College_and_university_rankings_in_the_United_States It contains several tables, that show some university rankings. Scroll till you reach "Forbes college rankings" field. First, scrape HTML table from this field and save as forbes_ranking. Then, write a function that creates new column "Ranked on Forbes" in practice_dataset and accepts Boolean values (True or False) based on the fact whether this university in forbes_ranking or not. Final result should look as follows:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Career 90th Percentile Salary	Ranked on Forbes
0	Massachusetts Institute of Technology (MIT)	Engineering	more than national average	126000	$76,800.00	$99,200.00	$168,000.00	$220,000.00	True
1	California Institute of Technology (CIT)	Engineering	more than national average	123000	NaN	$104,000.00	$161,000.00	NaN	False
2	Harvey Mudd College	Engineering	more than national average	122000	NaN	$96,000.00	$180,000.00	NaN	False
3	Polytechnic University of New York, Brooklyn	Engineering	more than national average	114000	$66,800.00	$94,300.00	$143,000.00	$190,000.00	False
4	Cooper Union	Engineering	more than national average	114000	NaN	$80,200.00	$142,000.00	NaN	False

In [ ]:
```
#please code here
```

Task 6

If time spent by website visitors on two difernet landing pages could be drawn as below, so that average time spent is the same for both. Based on the graph solely, what do you think what landing page (red or purple) performes better and why?



In [ ]:
```
#please opine here
```

Task 7

if a die is thrown 6 times, what is the probability of 3 of the numbers being even numbers?

In [ ]:
```
#please answer here
```