Ex - GroupBy

Introduction:

GroupBy can be summarized as Split-Apply-Combine.

Special thanks to: https://github.com/justmarkham for sharing the dataset and materials.

Check out this Diagram

Step 1. Import the necessary libraries

```
import pandas as pd
```

+ Mã + Văn bản

Step 2. Import the dataset from this address.

Step 3. Assign it to a variable called drinks.

```
drinks = pd.read_csv('drinks.csv')
```

Step 4. Which continent drinks more beer on average?

```
step4 = drinks.groupby('continent')['beer_servings'].mean().idxmax()
print(step4)
```

```
EU
```

Step 5. For each continent print the statistics for wine consumption.

```
step5 = drinks.groupby('continent')['wine_servings'].describe()
print(step5)
```

```
           count        mean        std  min   25%    50%     75%    max
continent
AF          53.0   16.264151  38.846419  0.0   1.0    2.0   13.00  233.0
AS          44.0    9.068182  21.667034  0.0   0.0    1.0    8.00  123.0
EU          45.0  142.222222  97.421738  0.0  59.0  128.0  195.00  370.0
OC          16.0   35.625000  64.555790  0.0   1.0    8.5   23.25  212.0
SA          12.0   62.416667  88.620189  1.0   3.0   12.0   98.50  221.0
```

Step 6. Print the mean alcohol consumption per continent for every column

```
step6 = drinks.groupby('continent')[['beer_servings', 'spirit_servings', 'wine_servings', 'total_litres_of_pure_alcohol']].mean()
print(step6)
```

```
           beer_servings  spirit_servings  wine_servings  \
continent
AF             61.471698        16.339623      16.264151
AS             37.045455        60.840909       9.068182
EU            193.777778       132.555556     142.222222
OC             89.687500        58.437500      35.625000
SA            175.083333       114.750000      62.416667

           total_litres_of_pure_alcohol
continent
AF                             3.007547
AS                             2.170455
EU                             8.617778
OC                             3.381250
SA                             6.308333
```

Step 7. Print the median alcohol consumption per continent for every column

```
step7 = drinks.groupby('continent')[['beer_servings', 'spirit_servings', 'wine_servings', 'total_litres_of_pure_alcohol']].median()
print(step7)
```

```
           beer_servings  spirit_servings  wine_servings  \
continent
AF                  32.0              3.0            2.0
AS                  17.5             16.0            1.0
EU                 219.0            122.0          128.0
OC                  52.5             37.0            8.5
SA                 162.5            108.5           12.0

           total_litres_of_pure_alcohol
continent
AF                                 2.30
AS                                 1.20
EU                                10.00
OC                                 1.75
SA                                 6.85
```

Step 8. Print the mean, min and max values for spirit consumption.

This time output a DataFrame

```
step8 = drinks.groupby('continent')['spirit_servings'].agg(['mean', 'min', 'max'])
print(step8)
```

```
                mean  min  max
continent
AF         16.339623    0  152
AS         60.840909    0  326
EU        132.555556    0  373
OC         58.437500    0  254
SA        114.750000   25  302
```