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

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

Step 2. Import the dataset from this address.

Step 3. Assign it to a variable called drinks.

In [2]:
```
dr = pd.read_csv("https://raw.githubusercontent.com/justmarkham/DAT8/master/data/drinks.csv")
dr
```

Out[2]:

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	continent
0	Afghanistan	0	0	0	0.0	AS
1	Albania	89	132	54	4.9	EU
2	Algeria	25	0	14	0.7	AF
3	Andorra	245	138	312	12.4	EU
4	Angola	217	57	45	5.9	AF
...	...	...	...	...	...	...
188	Venezuela	333	100	3	7.7	SA
189	Vietnam	111	2	1	2.0	AS
190	Yemen	6	0	0	0.1	AS
191	Zambia	32	19	4	2.5	AF
192	Zimbabwe	64	18	4	4.7	AF

In [6]:
```
dr.describe(include="all")
```

Out[6]:

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	continent
count	193	193.000000	193.000000	193.000000	193.000000	170
unique	193	NaN	NaN	NaN	NaN	5
top	Afghanistan	NaN	NaN	NaN	NaN	AF
freq	1	NaN	NaN	NaN	NaN	53
mean	NaN	106.160622	80.994819	49.450777	4.717098	NaN
std	NaN	101.143103	88.284312	79.697598	3.773298	NaN
min	NaN	0.000000	0.000000	0.000000	0.000000	NaN
25%	NaN	20.000000	4.000000	1.000000	1.300000	NaN
50%	NaN	76.000000	56.000000	8.000000	4.200000	NaN
75%	NaN	188.000000	128.000000	59.000000	7.200000	NaN
max	NaN	376.000000	438.000000	370.000000	14.400000	NaN

Step 4. Which continent drinks more beer on average?

In [3]:
```
dr.groupby("continent").beer_servings.mean().sort_values(ascending = False)
```

Out[3]:
```
continent
EU    193.777778
SA    175.083333
OC     89.687500
AF     61.471698
AS     37.045455
Name: beer_servings, dtype: float64
```

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

In [5]:
```
dr.groupby("continent").wine_servings.describe(include = "all")
```

Out[5]:

continent	count	mean	std	min	25%	50%	75%	max
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

In [7]:
```
dr.groupby("continent").mean()
```

Out[7]:

continent	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol
AF	61.471698	16.339623	16.264151	3.007547
AS	37.045455	60.840909	9.068182	2.170455
EU	193.777778	132.555556	142.222222	8.617778
OC	89.687500	58.437500	35.625000	3.381250
SA	175.083333	114.750000	62.416667	6.308333

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

In [8]:
```
dr.groupby("continent").median()
```

Out[8]:

continent	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol
AF	32.0	3.0	2.0	2.30
AS	17.5	16.0	1.0	1.20
EU	219.0	122.0	128.0	10.00
OC	52.5	37.0	8.5	1.75
SA	162.5	108.5	12.0	6.85

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

This time output a DataFrame

In [9]:
```
dr.groupby("continent").spirit_servings.agg(['mean', 'min', 'max'])
```

Out[9]:

continent	mean	min	max
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