Tutorial 8 - Concat Dataframes

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

In [2]:
```
india_weather = pd.DataFrame({
    "city": ["mumbai","delhi","banglore"],
    "temperature": [32,45,30],
    "humidity": [80,60,78]
})
india_weather
```

Out[2]:

	city	temperature	humidity
0	mumbai	32	80
1	delhi	45	60
2	banglore	30	78

In [3]:
```
us_weather = pd.DataFrame({
    "city": ["new york","chicago","orlando"],
    "temperature": [21,14,35],
    "humidity": [68,65,75]
})
us_weather
```

Out[3]:

	city	temperature	humidity
0	new york	21	68
1	chicago	14	65
2	orlando	35	75

In [4]:
```
df = pd.concat([india_weather,us_weather])
df
```

Out[4]:

	city	temperature	humidity
0	mumbai	32	80
1	delhi	45	60
2	banglore	30	78
0	new york	21	68
1	chicago	14	65
2	orlando	35	75

In [5]:
```
# Observe que o index fica de cada dados original. Caso se queira utilizar o index
 contínuo:
df = pd.concat([india_weather,us_weather],ignore_index=True)
df
```

Out[5]:

	city	temperature	humidity
0	mumbai	32	80
1	delhi	45	60
2	banglore	30	78
3	new york	21	68
4	chicago	14	65
5	orlando	35	75

In [6]:
```
# Gerando keys para o Dtaframe concatenado (com 'ignore_index=True' não funciona)
df = pd.concat([india_weather,us_weather],keys=['india','us'],ignore_index=False)
df
```

Out[6]:

		city	temperature	humidity
india	**0**	mumbai	32	80
	1	delhi	45	60
	2	banglore	30	78
us	**0**	new york	21	68
	1	chicago	14	65
	2	orlando	35	75

In [7]:
```
# loc = location
df.loc['us']
```

Out[7]:

	city	temperature	humidity
0	new york	21	68
1	chicago	14	65
2	orlando	35	75

In [8]:
```
# Vamos agora analisar 2 dataframe (temperature and windspeed) das mesmas cidades q
ue estão separados e queremos colocar
# na mesma coluna
temp_df = pd.DataFrame({
    "city": ["mumbai","delhi","banglore"],
    "temperature": [32,45,30],
    })
temp_df
```

Out[8]:

	city	temperature
0	mumbai	32
1	delhi	45
2	banglore	30

In [9]:
```
windspd_df = pd.DataFrame({
    "city": ["mumbai","delhi","banglore"],
    "windspeed": [7,12,9],
    })
windspd_df
```

Out[9]:

	city	windspeed
0	mumbai	7
1	delhi	12
2	banglore	9

In [10]:
```
# Veja o que acontece quando utilizamos o comando 'concat'
df = pd.concat([temp_df,windspd_df],sort=False)
df
```

Out[10]:

	city	temperature	windspeed
0	mumbai	32.0	NaN
1	delhi	45.0	NaN
2	banglore	30.0	NaN
0	mumbai	NaN	7.0
1	delhi	NaN	12.0
2	banglore	NaN	9.0

In [11]:
```
# Google "concat" documentation nos pandas para ver mais opções.
# No exemplo anterior se adicionou mais linhas (axis=0) e na verdade queremos inser
ir o dado como colunas (axis=1) adicionais
df = pd.concat([temp_df,windspd_df],axis=1)
df
```

Out[11]:

	city	temperature	city	windspeed
0	mumbai	32	mumbai	7
1	delhi	45	delhi	12
2	banglore	30	banglore	9

In [12]:
```
# É possível juntar "series"
s = pd.Series(["Humid", "Dry", "Rain"], name="event")
s
```

Out[12]:
```
0    Humid
1      Dry
2     Rain
Name: event, dtype: object
```

In [13]:
```
df = pd.concat([temp_df, s], axis=1)
df
```

Out[13]:

	city	temperature	event
0	mumbai	32	Humid
1	delhi	45	Dry
2	banglore	30	Rain