

# Лабораторна робота № 0

Виконав студент групи КМ-91мп

Галета М.С.

```
In [1]: 1 import numpy as np
        2 import pandas as pd
        3 import matplotlib.pyplot as plt
        4
        5 %matplotlib inline
```

```
In [2]: 1 df = pd.read_csv("Top_100_Tennis_Players-2007_Men.csv", header=0)
```

Максим Галета

Третя літера імені - К. Її порядковий номер 15

Остання літера прізвища - А. Її порядковий номер 1

```
In [3]: 1 K = 15
        2 N = 1
```

## Виведення перших значень

In [4]: 1 df.head(K+N)

Out[4]:

	Rank	Name	Country	Pts	Singles Record (Career)	Winning Percentage	Career Earnings	Link to Wikipedia
0	1	Roger Federer	Switzerland	7290	594-143	80.60%	\$41721169	<a href="http://en.wikipedia.org/wiki/Roger_Federer">http://en.wikipedia.org/wiki/Roger_Federer</a>
1	2	Rafael Nadal	Spain	4875	309-73	80.89%	\$19470198	<a href="http://en.wikipedia.org/wiki/Rafael_Nadal">http://en.wikipedia.org/wiki/Rafael_Nadal</a>
2	3	Andy Roddick	USA	2980	436-136	76.22%	\$14069938	<a href="http://en.wikipedia.org/wiki/Andy_Roddick">http://en.wikipedia.org/wiki/Andy_Roddick</a>
3	4	Nikolay Davydenko	Russia	2825	303-202	60.00%	\$8637353	<a href="http://en.wikipedia.org/wiki/Nikolay_Davydenko">http://en.wikipedia.org/wiki/Nikolay_Davydenko</a>
4	5	Fernando Gonzalez	Chile	2755	294-160	64.76%	\$6742996	<a href="http://en.wikipedia.org/wiki/Fernando_Gonzalez">http://en.wikipedia.org/wiki/Fernando_Gonzalez</a>
5	6	Tommy Robredo	Spain	2645	320-194	62.26%	\$6538755	<a href="http://en.wikipedia.org/wiki/Tommy_Robredo">http://en.wikipedia.org/wiki/Tommy_Robredo</a>
6	7	Novak Djokovic	Serbia	2540	158-59	72.81%	\$8035331	<a href="http://en.wikipedia.org/wiki/Novak_Djokovic">http://en.wikipedia.org/wiki/Novak_Djokovic</a>
7	8	Ivan Ljubicic	Croatia	2270	340-223	60.39%	\$7090571	<a href="http://en.wikipedia.org/wiki/Ivan_Ljubicic">http://en.wikipedia.org/wiki/Ivan_Ljubicic</a>
8	9	James Blake	USA	2150	275-165	62.50%	\$5848841	<a href="http://en.wikipedia.org/wiki/James_Blake">http://en.wikipedia.org/wiki/James_Blake</a>
9	10	Tommy Haas	Germany	2135	416-224	65.00%	\$8753953	<a href="http://en.wikipedia.org/wiki/Tommy_Haas">http://en.wikipedia.org/wiki/Tommy_Haas</a>
10	11	Andy Murray	UK	2125	124-60	67.39%	\$2592451	<a href="http://en.wikipedia.org/wiki/Andy_Murray_%28tennis%29">http://en.wikipedia.org/wiki/Andy_Murray_%28te...</a>
11	12	David Nalbandian	Argentina	1805	272-134	67.00%	\$8893109	<a href="http://en.wikipedia.org/wiki/David_Nalbandian">http://en.wikipedia.org/wiki/David_Nalbandian</a>
12	13	Mario Ancic	Croatia	1775	189-121	60.97%	\$3694825	<a href="http://en.wikipedia.org/wiki/Mario_Ancic">http://en.wikipedia.org/wiki/Mario_Ancic</a>
13	14	Tomas Berdych	Czech Republic	1715	167-108	60.73%	\$4007728	<a href="http://en.wikipedia.org/wiki/Tomas_Berdych">http://en.wikipedia.org/wiki/Tomas_Berdych</a>
14	15	Richard Gasquet	France	1620	149-98	60.32%	\$3592219	<a href="http://en.wikipedia.org/wiki/Richard_Gasquet">http://en.wikipedia.org/wiki/Richard_Gasquet</a>
15	16	David Ferrer	Spain	1615	242-156	60.80%	\$5733751	<a href="http://en.wikipedia.org/wiki/David_Ferrer">http://en.wikipedia.org/wiki/David_Ferrer</a>

## Препроцесинг

```
In [5]: 1 for i in range(df["Country"].values.shape[0]):  
2         df["Country"].values[i] = df["Country"].values[i].strip()
```

```
In [6]: 1 df["Winning Percentage"] = np.array(df["Winning Percentage"].values, dtype="<U5").astype(float)
```

```
In [7]: 1 values = np.array(df["Career Earnings"].values, dtype=str)  
2 for i in range(values.shape[0]):  
3     if values[i] != "nan":  
4         values[i] = values[i][1:]  
5 df["Career Earnings"] = np.array(values, dtype=float)
```

```
In [8]: 1 values = np.array(df["Singles Record (Career)"].values, dtype=str)  
2 for i in range(values.shape[0]):  
3     if values[i] != "nan":  
4         string = ""  
5         for ch in values[i]:  
6             if ch == "-":  
7                 break  
8             else:  
9                 string += ch  
10        values[i] = string  
11 df.insert(4, "Number of Wins", np.array(values, dtype=float))
```

In [9]: 1 df.head(K+N)

Out[9]:

	Rank	Name	Country	Pts	Number of Wins	Singles Record (Career)	Winning Percentage	Career Earnings	Link to Wikipedia
0	1	Roger Federer	Switzerland	7290	594.0	594-143	80.60	41721169.0	<a href="http://en.wikipedia.org/wiki/Roger_Federer">http://en.wikipedia.org/wiki/Roger_Federer</a>
1	2	Rafael Nadal	Spain	4875	309.0	309-73	80.89	19470198.0	<a href="http://en.wikipedia.org/wiki/Rafael_Nadal">http://en.wikipedia.org/wiki/Rafael_Nadal</a>
2	3	Andy Roddick	USA	2980	436.0	436-136	76.22	14069938.0	<a href="http://en.wikipedia.org/wiki/Andy_Roddick">http://en.wikipedia.org/wiki/Andy_Roddick</a>
3	4	Nikolay Davydenko	Russia	2825	303.0	303-202	60.00	8637353.0	<a href="http://en.wikipedia.org/wiki/Nikolay_Davydenko">http://en.wikipedia.org/wiki/Nikolay_Davydenko</a>
4	5	Fernando Gonzalez	Chile	2755	294.0	294-160	64.76	6742996.0	<a href="http://en.wikipedia.org/wiki/Fernando_Gonzalez">http://en.wikipedia.org/wiki/Fernando_Gonzalez</a>
5	6	Tommy Robredo	Spain	2645	320.0	320-194	62.26	6538755.0	<a href="http://en.wikipedia.org/wiki/Tommy_Robredo">http://en.wikipedia.org/wiki/Tommy_Robredo</a>
6	7	Novak Djokovic	Serbia	2540	158.0	158-59	72.81	8035331.0	<a href="http://en.wikipedia.org/wiki/Novak_Djokovic">http://en.wikipedia.org/wiki/Novak_Djokovic</a>
7	8	Ivan Ljubicic	Croatia	2270	340.0	340-223	60.39	7090571.0	<a href="http://en.wikipedia.org/wiki/Ivan_Ljubicic">http://en.wikipedia.org/wiki/Ivan_Ljubicic</a>
8	9	James Blake	USA	2150	275.0	275-165	62.50	5848841.0	<a href="http://en.wikipedia.org/wiki/James_Blake">http://en.wikipedia.org/wiki/James_Blake</a>
9	10	Tommy Haas	Germany	2135	416.0	416-224	65.00	8753953.0	<a href="http://en.wikipedia.org/wiki/Tommy_Haas">http://en.wikipedia.org/wiki/Tommy_Haas</a>
10	11	Andy Murray	UK	2125	124.0	124-60	67.39	2592451.0	<a href="http://en.wikipedia.org/wiki/Andy_Murray_%28tennis%29">http://en.wikipedia.org/wiki/Andy_Murray_%28te...</a>
11	12	David Nalbandian	Argentina	1805	272.0	272-134	67.00	8893109.0	<a href="http://en.wikipedia.org/wiki/David_Nalbandian">http://en.wikipedia.org/wiki/David_Nalbandian</a>
12	13	Mario Ancic	Croatia	1775	189.0	189-121	60.97	3694825.0	<a href="http://en.wikipedia.org/wiki/Mario_Ancic">http://en.wikipedia.org/wiki/Mario_Ancic</a>
13	14	Tomas Berdych	Czech Republic	1715	167.0	167-108	60.73	4007728.0	<a href="http://en.wikipedia.org/wiki/Tomas_Berdych">http://en.wikipedia.org/wiki/Tomas_Berdych</a>
14	15	Richard Gasquet	France	1620	149.0	149-98	60.32	3592219.0	<a href="http://en.wikipedia.org/wiki/Richard_Gasquet">http://en.wikipedia.org/wiki/Richard_Gasquet</a>
15	16	David Ferrer	Spain	1615	242.0	242-156	60.80	5733751.0	<a href="http://en.wikipedia.org/wiki/David_Ferrer">http://en.wikipedia.org/wiki/David_Ferrer</a>

# Завдання 1

Обчислити та вивести:

- загальну кількість гравців з країни з кодом  $N$ , де  $N$  – це порядковий номер в списку групи. Коди країн: 1 – Іспанія, 2 – США, 3 – Росія, 4 – Чехія, 5 – Німеччина, 6 – Сербія, 7 – Франція, 8 – Хорватія, 9 – Аргентина, 10 – Швеція, 11 – Чилі, 12 – Італія, 13 – Австрія, 14 – Перу, 15 – Данія
- середнє значення призових для гравців, у яких відсоток перемог лежить в діапазоні від  $M$  до  $M+10$ , де  $M$  – це помножений на 10 залишок від ділення порядкового номеру першої літери прізвища на 9

a)

Порядковий номер в списку групи 4. 4 - Чехія

```
In [10]: 1 df.loc[df["Country"] == "Czech Republic"].groupby("Country", as_index=False)["Name"].count().rename(columns={"Name":
```

Out[10]:

	Country	Number of players
0	Czech Republic	4

b)

Максим Галета

Перша літера прізвища - Г. Її порядковий номер 4

**4:9 = 0 (ост. 4) => остача 4**

**M = 4\*10**

In [11]:

```
1 M = 40
```

In [12]:

```
1 avg_earn = np.mean(df.loc[(df["Winning Percentage"] >= M) & (df["Winning Percentage"] <= (M+10))]["Career Earnings"])
2 d = {
3     "Winning Percentage": ["[40.00 - 50.00]"],
4     "Average earnings": ["{:0.2f}".format(avg_earn)]
5 }
6 pd.DataFrame(data=d)
```

Out[12]:

	Winning Percentage	Average earnings
0	[40.00 - 50.00]	1784348.58

## Завдання 2

Побудувати на одному графіку:

- середній рейтинг гравців по країнах
- середню кількість перемог гравців по країнах

In [13]:

```
1 avg_ranks = df.groupby("Country", as_index=False)["Country", "Rank"].mean()
2 avg_wins = df.groupby("Country", as_index=False)["Country", "Number of Wins"].mean()
3
4 ranks = avg_ranks["Rank"].values
5 wins = avg_wins["Number of Wins"].values
6 countries = avg_ranks["Country"].values
```

```
In [14]: 1 plt.figure(figsize=(17, 8))
2         ax = plt.subplot()
3
4         ax.bar(np.arange(countries.shape[0])-.0.15, ranks, width=0.25, label="Average ranks")
5         ax.bar(np.arange(countries.shape[0])+0.15, wins, width=0.25, label="Average wins")
6
7         plt.xticks(range(countries.shape[0]), countries, rotation=90)
8         plt.legend()
9         plt.grid()
10        plt.show()
```



