



Shahid Beheshti
University

دانشگاه شهید بهشتی
دانشکده علوم ریاضی
گروه علوم کامپیوتر

گزارش تمرین های سری اول

تمرین دوم

درس داده کاوی

اساتید محترم:

جناب آقای دکتر هادی فراهانی و جناب آقای دکتر سعید رضا خردپیشه

آموزشیار محترم: جناب آقای علی شریفی

جواد تدین ۹۹۴۲۲۰۴۱

بهار ۱۴۰۰

در این تمرین مجموعه داده های مربوط به مسابقات فوتبال بین المللی داده شده است در مقابل مسابقات جام جهانی "جام های قاره ای" تورنمنت ها و مسابقات دوستانه به ما داده شده است.

اطلاعات مثل نام تیم ها "محل انجام مسابقه" "مهمان و میزبان" "زمان بازی ها" "تعداد گل ها" نتیجه بازی و... ذخیره شده اند

```
jupyter Football-Tadayon-99422041 Last Checkpoint: a minute ago (autosaved)
File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3
In [4]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import datetime

In [6]: data_set = pd.read_csv('C:\\Users\\Naim\\Pictures\\results.csv')
data_set.head(20)

Out[6]:
```

	date	home_team	away_team	home_score	away_score	tournament	city	country	neutral
0	1872-11-30	Scotland	England	0	0	Friendly	Glasgow	Scotland	False
1	1873-03-08	England	Scotland	4	2	Friendly	London	England	False
2	1874-03-07	Scotland	England	2	1	Friendly	Glasgow	Scotland	False
3	1875-03-06	England	Scotland	2	2	Friendly	London	England	False
4	1876-03-04	Scotland	England	3	0	Friendly	Glasgow	Scotland	False
5	1876-03-25	Scotland	Wales	4	0	Friendly	Glasgow	Scotland	False
6	1877-03-03	England	Scotland	1	3	Friendly	London	England	False
7	1877-03-05	Wales	Scotland	0	2	Friendly	Wrexham	Wales	False
8	1878-03-02	Scotland	England	7	2	Friendly	Glasgow	Scotland	False
9	1878-03-23	Scotland	Wales	9	0	Friendly	Glasgow	Scotland	False
10	1879-01-18	England	Wales	2	1	Friendly	London	England	False

حال به بررسی سوالات خواسته شده می پردازیم

در سوال اول باید بهترین تیم را تعیین کنیم

در سوال دوم باید تعیین کنیم کدام تیم در دوره های مختلف بر مسابقات تسلط داشته و....

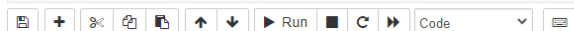
دوباره مثل تمرین قبل data set را با استفاده از یک فایل csv می خوانیم.

با استفاده کردن از دستور . head میتوان به ۱۰ سطر اول دیتا دسترسی پیدا کرد و دید کلی نسبت به داده ها پیدا کرد .

که در جدول موجود در فایل ارسال شده قابل مشاهده است.

برای این که بتوانیم تحلیل بهتری داشته باشیم نیاز به این داریم که داده هارا دسته بندی کنیم

پس به سال و ماه تقسیم می کنیم و جدول بعدی را بدست خواهیم آورد

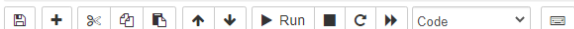


```
In [7]: data_set["Year"] = pd.to_datetime(data_set['date']).dt.year
data_set["Month"] = pd.to_datetime(data_set['date']).dt.month
data_set = data_set.drop(columns=['date'], axis=1)
data_set.head(20)
```

```
Out[7]:
```

	home_team	away_team	home_score	away_score	tournament	city	country	neutral	Year	Month
0	Scotland	England	0	0	Friendly	Glasgow	Scotland	False	1872	11
1	England	Scotland	4	2	Friendly	London	England	False	1873	3
2	Scotland	England	2	1	Friendly	Glasgow	Scotland	False	1874	3
3	England	Scotland	2	2	Friendly	London	England	False	1875	3
4	Scotland	England	3	0	Friendly	Glasgow	Scotland	False	1876	3
5	Scotland	Wales	4	0	Friendly	Glasgow	Scotland	False	1876	3
6	England	Scotland	1	3	Friendly	London	England	False	1877	3
7	Wales	Scotland	0	2	Friendly	Wrexham	Wales	False	1877	3
8	Scotland	England	7	2	Friendly	Glasgow	Scotland	False	1878	3
9	Scotland	Wales	9	0	Friendly	Glasgow	Scotland	False	1878	3
10	England	Wales	2	1	Friendly	London	England	False	1879	1
11	England	Scotland	5	4	Friendly	London	England	False	1879	4
12	Wales	Scotland	0	3	Friendly	Wrexham	Wales	False	1879	4
13	Scotland	England	5	4	Friendly	Glasgow	Scotland	False	1880	3
14	Wales	England	2	3	Friendly	Wrexham	Wales	False	1880	3

در مرحله بعدی گل هایی که در متن مسابقات زده شده را بدست می آوریم و جدول سوم را خواهیم داشت.



```
In [8]: data_set['total_score'] = data_set['home_score'] + data_set['away_score']
data_set.head(20)
```

```
Out[8]:
```

	home_team	away_team	home_score	away_score	tournament	city	country	neutral	Year	Month	total_score
0	Scotland	England	0	0	Friendly	Glasgow	Scotland	False	1872	11	0
1	England	Scotland	4	2	Friendly	London	England	False	1873	3	6
2	Scotland	England	2	1	Friendly	Glasgow	Scotland	False	1874	3	3
3	England	Scotland	2	2	Friendly	London	England	False	1875	3	4
4	Scotland	England	3	0	Friendly	Glasgow	Scotland	False	1876	3	3
5	Scotland	Wales	4	0	Friendly	Glasgow	Scotland	False	1876	3	4
6	England	Scotland	1	3	Friendly	London	England	False	1877	3	4
7	Wales	Scotland	0	2	Friendly	Wrexham	Wales	False	1877	3	2
8	Scotland	England	7	2	Friendly	Glasgow	Scotland	False	1878	3	9
9	Scotland	Wales	9	0	Friendly	Glasgow	Scotland	False	1878	3	9
10	England	Wales	2	1	Friendly	London	England	False	1879	1	3
11	England	Scotland	5	4	Friendly	London	England	False	1879	4	9
12	Wales	Scotland	0	3	Friendly	Wrexham	Wales	False	1879	4	3
13	Scotland	England	5	4	Friendly	Glasgow	Scotland	False	1880	3	9
14	Wales	England	2	3	Friendly	Wrexham	Wales	False	1880	3	5
15	Scotland	Wales	5	1	Friendly	Glasgow	Scotland	False	1880	3	6

در مرحله بعد یعنی مرحله چهارم برنده هر مسابقه را تعیین می کنیم

Jupyter Football-Tadayon-99422041 Last Checkpoint: 39 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

19 Northern Ireland England 0 13 Friendly Belfast Republic of Ireland False 1882 2 13 2 England

```
In [10]: b_team=data_set.groupby('who_win').sum()
b_team=b_team[['home_score','away_score','total_score']].sort_values(by=['total_score'],ascending=False)
b_team=b_team.drop(['no_win'])
b_team.head(20)
```

Out[10]:

who_win	home_score	away_score	total_score
Brazil	1366	836	2202
England	1163	1036	2199
Germany	1233	886	2119
Sweden	1084	806	1890
Argentina	1158	622	1780
Hungary	1030	739	1769
Netherlands	914	598	1512
Mexico	945	483	1428
France	902	524	1426
South Korea	882	512	1394
Spain	850	521	1371
Italy	847	474	1321

Jupyter Football-Tadayon-99422041 Last Checkpoint: 37 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

19 Northern Ireland England 0 13 Friendly Belfast Republic of Ireland False 1882 2 13

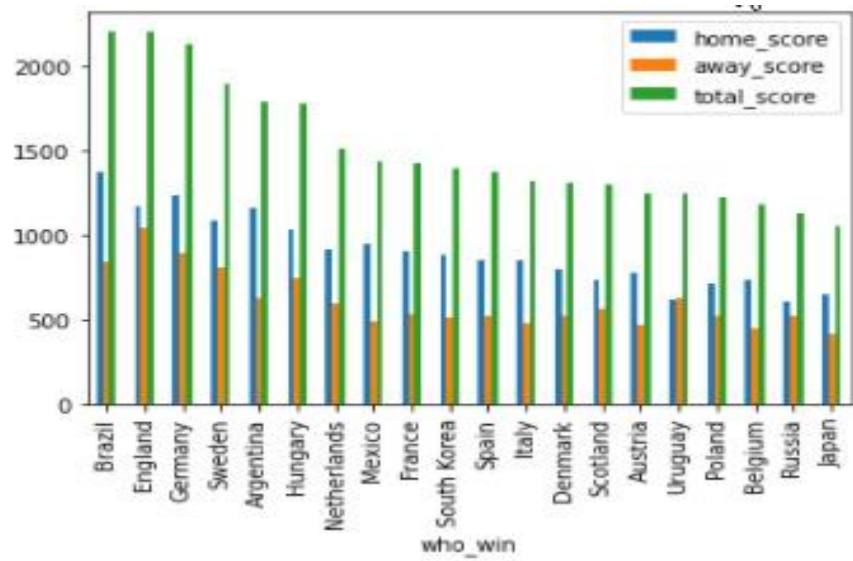
```
In [9]: con=((data_set['home_score']==data_set['away_score']), (data_set['home_score']>data_set['away_score']), (data_set['home_score']<data_set['away_score']))
value=[0,1,2]
value2=['no_win',data_set['home_team'],data_set['away_team']]
data_set['result']=np.select(con,value)
data_set['who_win']=np.select(con,value2)
data_set.head(20)
```

Out[9]:

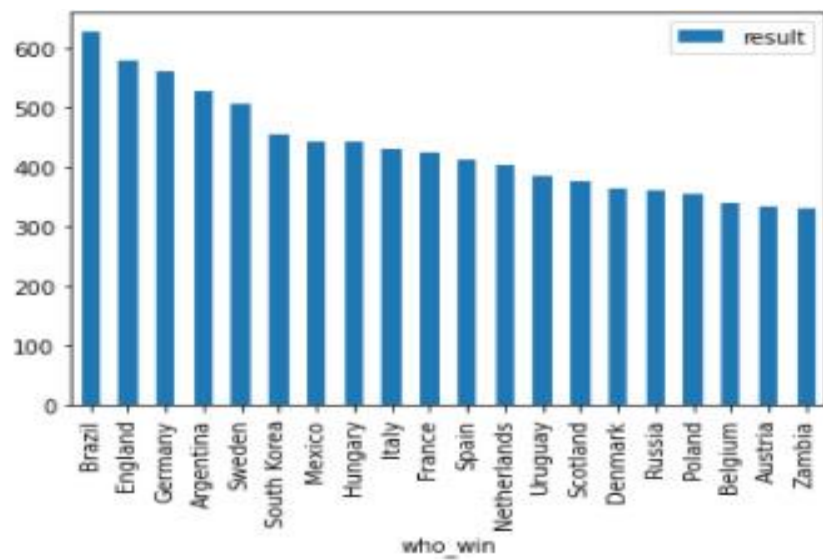
	home_team	away_team	home_score	away_score	tournament	city	country	neutral	Year	Month	total_score	result	who_win
0	Scotland	England	0	0	Friendly	Glasgow	Scotland	False	1872	11	0	0	no_win
1	England	Scotland	4	2	Friendly	London	England	False	1873	3	6	1	England
2	Scotland	England	2	1	Friendly	Glasgow	Scotland	False	1874	3	3	1	Scotland
3	England	Scotland	2	2	Friendly	London	England	False	1875	3	4	0	no_win
4	Scotland	England	3	0	Friendly	Glasgow	Scotland	False	1876	3	3	1	Scotland
5	Scotland	Wales	4	0	Friendly	Glasgow	Scotland	False	1876	3	4	1	Scotland
6	England	Scotland	1	3	Friendly	London	England	False	1877	3	4	2	Scotland
7	Wales	Scotland	0	2	Friendly	Wrexham	Wales	False	1877	3	2	2	Scotland
8	Scotland	England	7	2	Friendly	Glasgow	Scotland	False	1878	3	9	1	Scotland
9	Scotland	Wales	9	0	Friendly	Glasgow	Scotland	False	1878	3	9	1	Scotland

حال نمودار هارا برای نمایش بهتر داده ها رسم می کنیم.

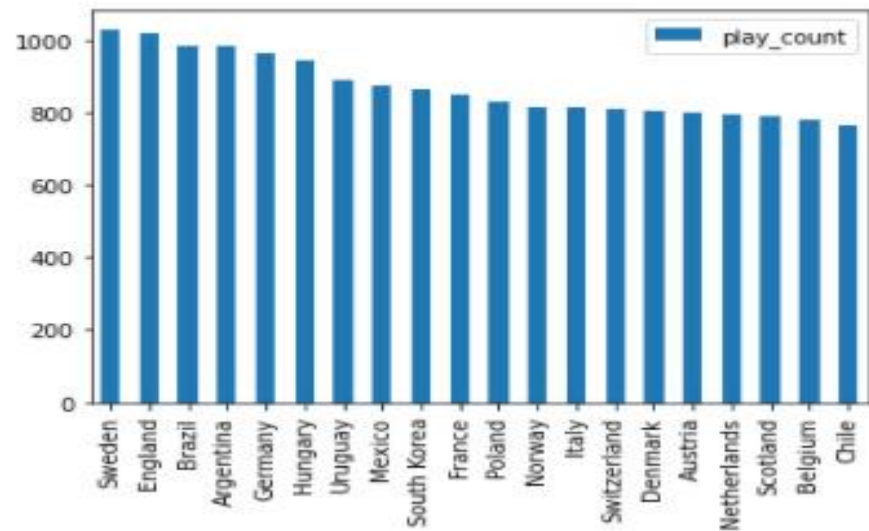
بهترین تیم ها را با بیشترین گل مشخص می کنیم



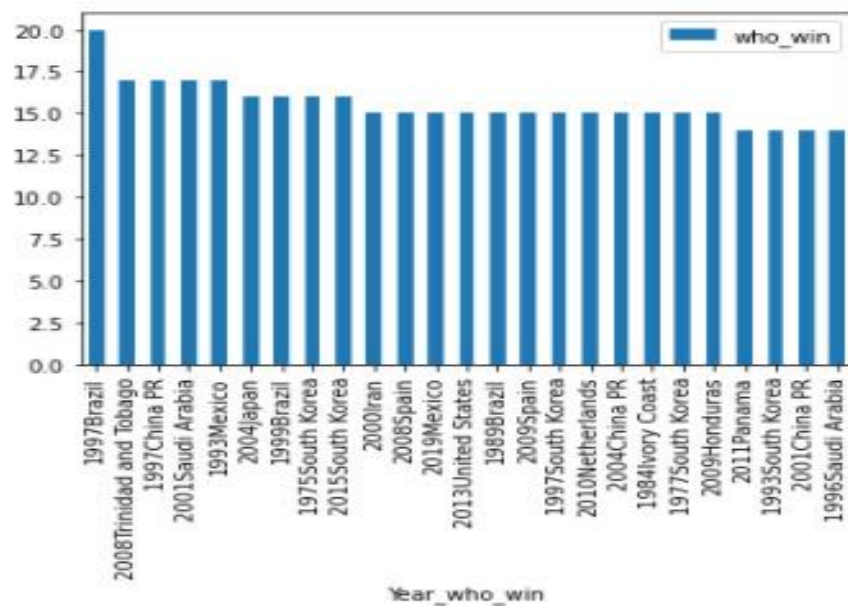
تیم هایی که بیشترین برد های تاریخ را داشته اند پیدا می کنیم.



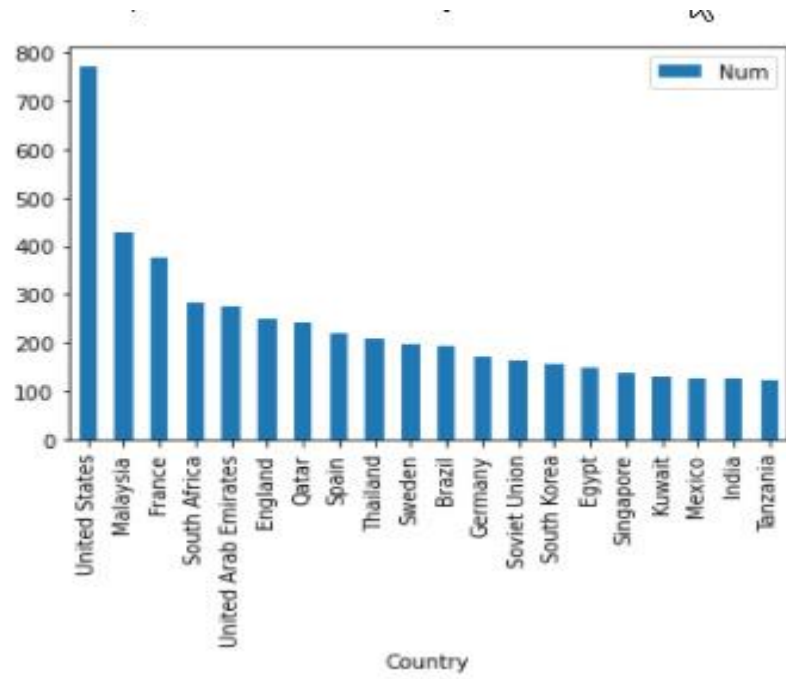
تیم هایی که بیشترین تعداد مسابقه را در طول تاریخ داشته اند مشخص می کنیم



تیم هایی که بیشترین برد را در یک سال داشته اند را نمایش می دهیم



کشورهایی که بیشترین میزبانی را داشته اند مشخص می کنیم



و در آخر هم تیم هایی که بیشترین حضور را داشته اند.

