

به نام خداوند بخشنده و مهربان

موضوع : گزارش بازی فوتبال

نام اساتید : جناب آقای دکتر فردپیشه / جناب آقای دکتر فراهانی

شکیلا جابری

۹۹۴۲۲۰۴۶

گزارش کار تمرین ۱ قسمت دوم

در این تمرین **data set** مربوط به مسابقات فوتبال بین المللی به ما داده شده است که در مقابل مسابقات جام جهانی ، جام های قاره ای ، تورنومنت ها ، بازی های دوستانه و ... انجام شده است . به همین دلیل اطلاعاتی مثل نام تیم ها ، محل انجام مسابقات ، میزبان و مهمان ، زمان انجام بازی ها ، تعداد گل های هر بازی ، نتیجه بازی ها و ... قرار داده شده است و تعدادی سؤال مطرح است که با استفاده از برنامه نویسی پایتون باید به آن ها پاسخ دهیم و با استفاده از استنتاج های آماری.

✓ در سؤال اول از ما خواسته شده که بهترین تیم در همه زمان را مشخص کنیم ؟

✓ در سؤال دوم خواسته شده است که تیمی را مشخص کنیم که در دوره های

مختلف بر مسابقات تسلط داشته است ؟

در تمرین قبل با استفاده از **Jupiter notebook** که ها زده شده است

که در این تمرین هم به همین شکل است.

+ در فط اول کتابخانه های **numpy - pandas - matplotlib** -

seaborn ، **data time** ، **import** کردیم که همانند تمرین قبل

میباشد (بجز **DATA TIME**)

+ در فط دوم **Data set** را میخوانیم با استفاده از **file CSV** که توسط

استاد ارسال شد و با استفاده از این **data set** جدول داده ها را خواهیم

داشت و با استفاده از دستور **head(10)** میتوانیم 10 سطر اول مجموعه

داده را بازگردانیم و بخش کوچکی از داده ها را بررسی کنیم .

+ در فط سوم **data set** را بر حسب (سال) و (ماه) مورد بررسی قرار میدهم که اطلاعات در جدول ارسال شده قابل مشاهده میباشد پس برای بررسی بهتر، سال و ماه را از تاریخ جدا میکنیم.

+ در فط چهارم با استفاده از **[total_score]** تعداد گل های زده شده در هر مسابقه را بررسی میکنیم. (در جدول شماره 4 قابل مشاهده است)

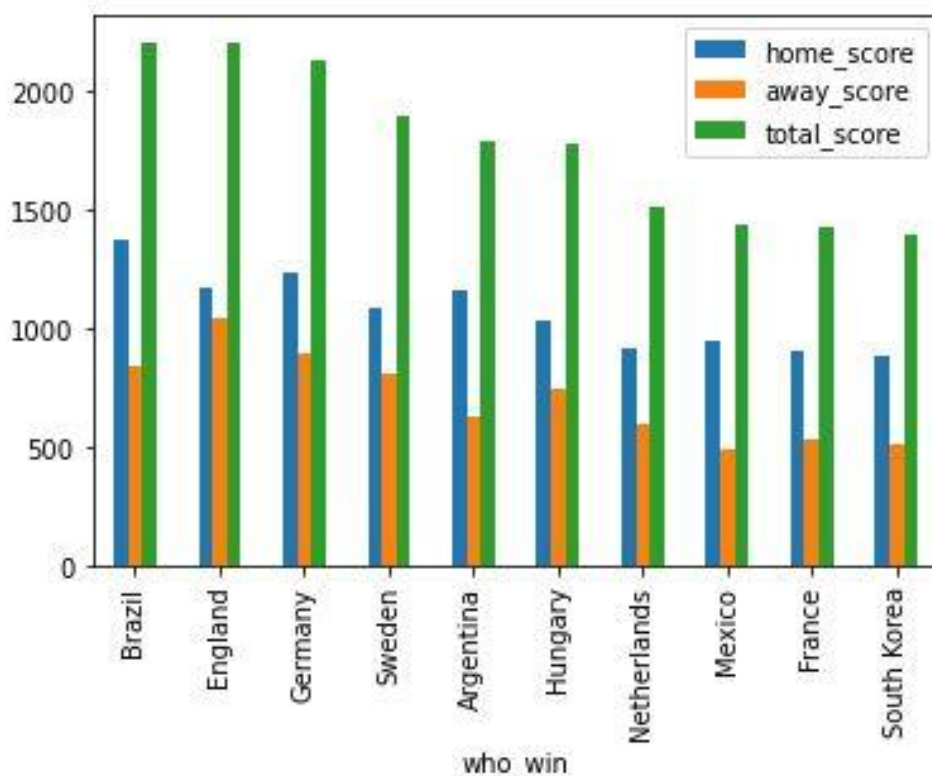
+ در فط پنجم با استفاده از کد

```
Con=[(dataset(home_score))]=data set[away_score]
```

میتوانیم تیم های برنده را مشخص کنیم که در ستون آخر قابل مشاهده است مثلاً در ستون **Who_win** مشخص شده است که در بازی انگلستان و اسکاتلند، تیم انگلستان برنده بازی بوده است.

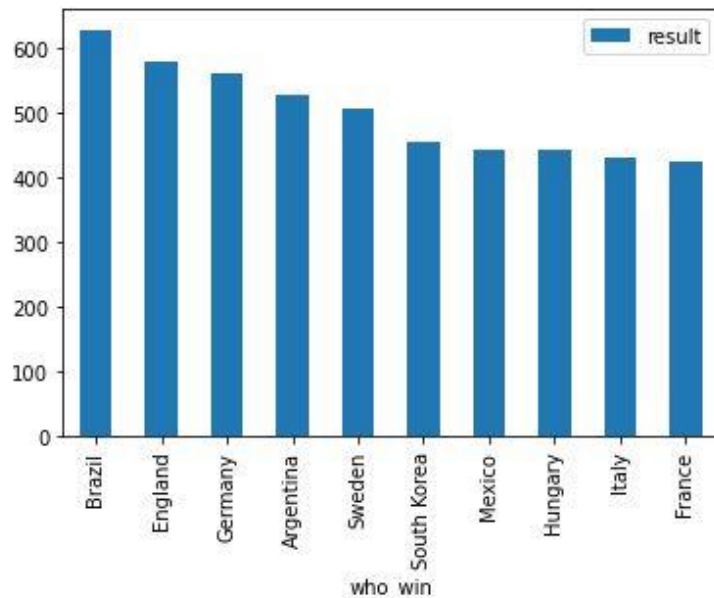
+ در فط ششم سعی کردیم که بهترین تیم در طی مسابقات را مشخص کنیم که با استفاده از دستور **groupby** ابتدا سعی کردیم برنده ها را گروه بندی کنیم و طبق جدول بدست آمده نمودار را رسم میکنیم.

+ که در فط هفتم با استفاده از دستور **plot bar** که برای مصورسازی **data** است و در تمرین قبل این دستور توضیح داده شده است ، سعی کرده ایم تا بهترین تیم و بیشترین گل زده شده را مشخص کنیم



+ در فط هشتم از دستور **count** استفاده کرده ایم تا بتوانیم تیم هایی که بیشترین برد ها را داشته اند را مشخص کنیم که مجدداً از دستور **groupby** استفاده کرده ایم که این دستور در گروه بندی داده ها بسیار کاربردی است ، که طبق جدول بدست آمده قابل مشاهده است که برزیل با 629 برد بیشترین برد را داشته و زامبیا با 331 برد کمترین برد در طول تاریخ را داشته است

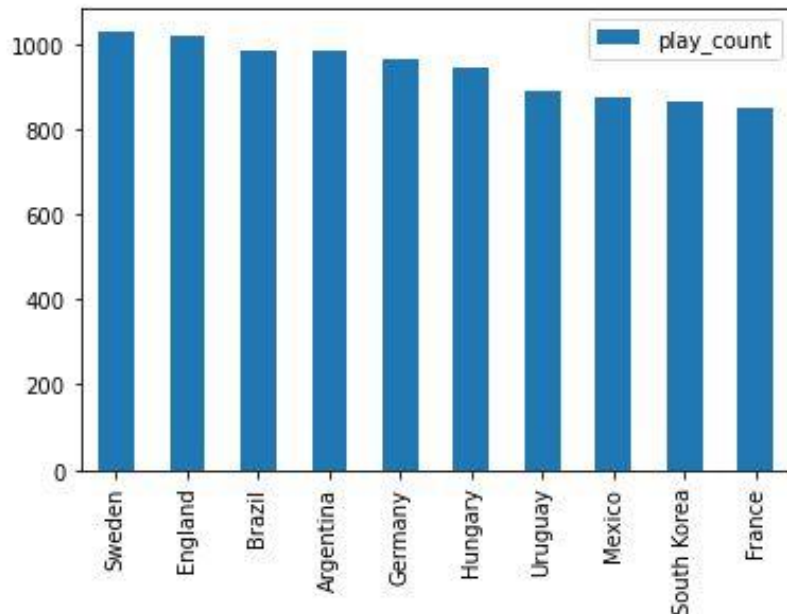
+ در فط نهم مجدداً با استفاده از دستور **plot bar** نمودار را برای فطم بهتر رسم کرده ایم



+ در فط دهم تیم هایی که در یک زمان بیشترین گل را زده اند و بیشترین برد را داشته اند را بررسی کرده ایم و با استفاده از دستور **merge** که از جدولی که بدست آمد میتوان استخراج کرد که باز هم اول برزیل و بعد انگلستان ، هم بیشترین گل را زده اند و هم بیشترین برد را داشته اند

+ در فط یازدهم تیم هایی که بیشترین تعداد مسابقه را داشته اند پیدا میکنیم که باز هم با استفاده از دستور **groupby** مسابقات را گروه بندی کرده ایم و مجدداً از دستور **count** استفاده کردیم (در تمرین قبل آموختیم که وقتی **count** میگیریم یعنی متوجه شویم که در این تیم ها (چه تعدادی) مسابقه برگزار شده است)

در فط دوازدهم مجدداً با استفاده از دستور **plot bar** نمودار مربوط به آن را رسم کرده ایم. قبلاً گفتیم هر کدام از نمودار ها فعالیت مربوط به خود را دارند و با دستور **plot** میتوانیم مصور سازی انجام دهیم که میتواند **bar chart** و باشد (در اینجا از **bar** استفاده کرده ایم)



در فط سیزدهم تیم هایی که بیشترین مسابقه و بیشترین برد را داشته اند را بررسی میکنیم در این قسمت از کد **merge** استفاده میکنیم. لازم به توضیح است که دستکاری یا پیش پردازش داده ها (**data wrangling**) فرآیند پردازش داده ها برای آماده سازی آنها جهت استفاده در قدم بعدی است. برای عمل پیش پردازش میتوان به **merging** ، **grouping** ، **countention** اشاره کرد و این دستکاری در داده ها به این دلیل است که داده ها را به فرض تبدیل کند که برای انجام تحلیل ها و کار با الگوریتم ها مناسب تر هستند

در فط چهاردهم از دستور **sort_value** استفاده کرده ایم که برای مرتب سازی داده ها میباشد ، و هم از **Ascending** که مساوی **false** شده است برای دسترسی به **data frame** با اندیس بولین باید **data frame** ها را با اندیس بولین ایجاد کنیم یعنی **data frame** هایی با اندیس هایی که مقادیر آن **True** و یا **False** باشد.

برای مرتب سازی صعودی یا نزولی بودن داده های ستون دیتا فریم از تابع **sort_values** استفاده میکنیم . قاعده کلی به شکل زیر است :

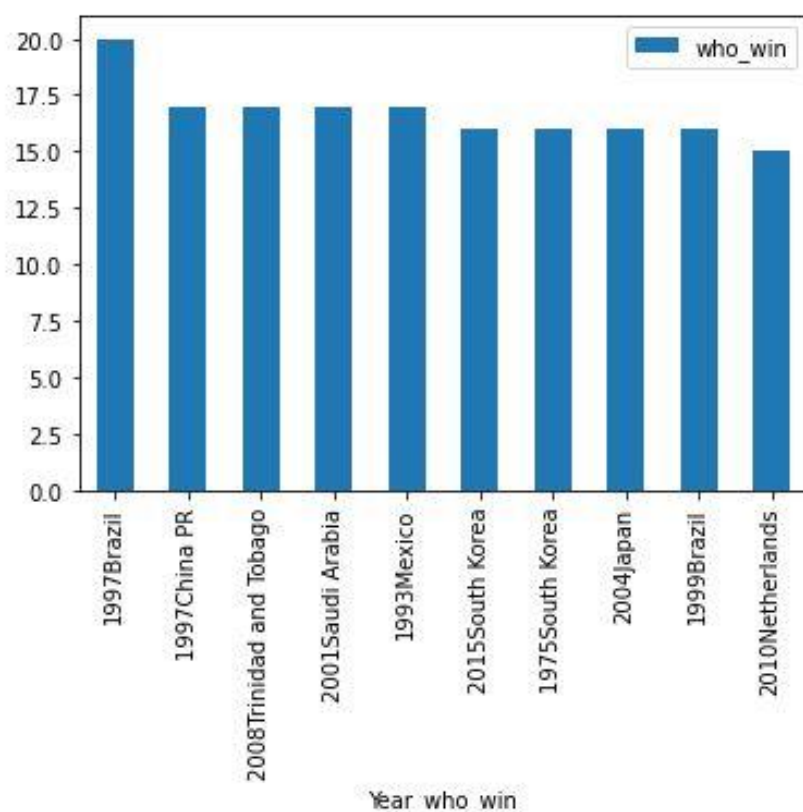
```
Data frame.sort_values(by , ascending=True  
,inplace =false axis=0 kind=quick sort  
,na_position=last)
```

تمامی پارامتر ها بجز **by** دارای مقادیر پیش فرض میباشدند. **by** نام یک و یا مجموعه ای از ستون ها برای مرتب سازی داده ها است.

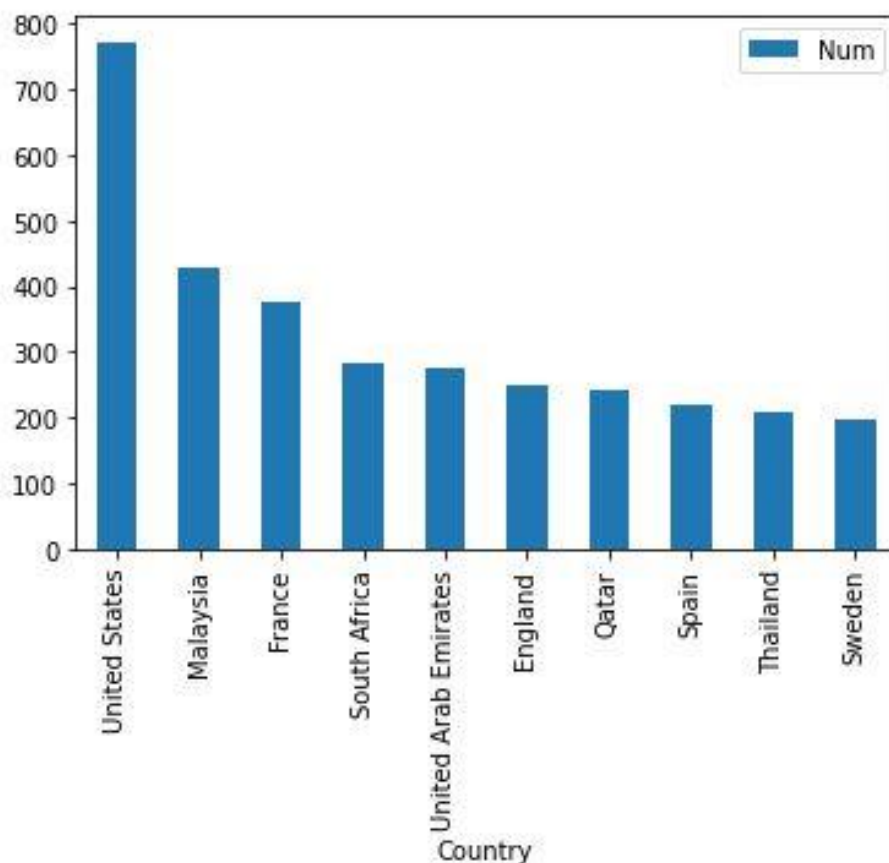
Ascending یک مقدار بولین است که اگر **true** باشد صعودی ؛ و اگر **false** باشد نزولی مرتب میکند.

در فط پانزدهم تیم هایی که بیش از ۸۰۰ مسابقه انجام داده اند و بیشترین برد در هر مسابقه را داشته اند پیدا میکنیم.

در فط نوزدهم سعی کردیم تیم هایی که بیشترین برد را در یکسال داشته اند بررسی کنیم و نمودار آن را در فط ۲۰ رسم کرده ایم



در فط بیست و یکم سعی کرده ایم که بیشترین میزبانی را برای مسابقات سایر کشور ها داشته اند ، پیدا کنیم که با توجه به جدول مشاهده میکنیم آمریکا بیشترین میزبانی را با توجه به عدد **779** داشته است و باز هم با استفاده از **plot bar** نمودار رسم کرده ایم .



✚ خط بیست و سوم توانستیم تیم هایی که بیشترین حضور را در بازی های دوستانه داشته اند پیدا کنیم که با توجه به جدول ، آلمان در مقام اول و مکزیک در مقام آخر قرار دارند.

