

گزارش تمرین شماره 1

واحد درسی داده کاوی

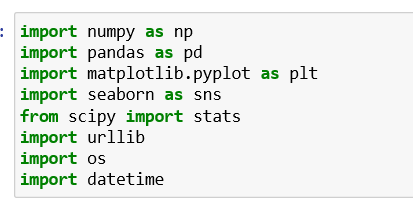
اساتید محترم :

جناب آقای دکتر فراهانی

جناب آقای دکتر خردپیشه

هدیه آشوری 99422022

16/01/1400

* بررسی داده ها

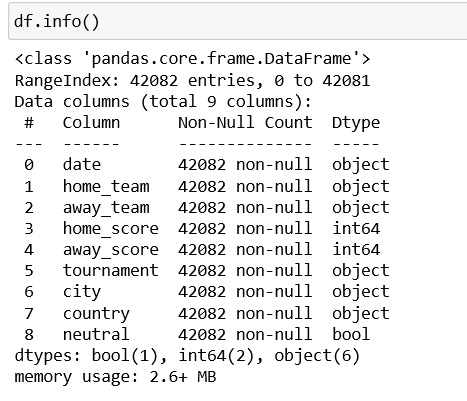
در ابتدا پکیج های مورد نیاز را Import

می کنیم :

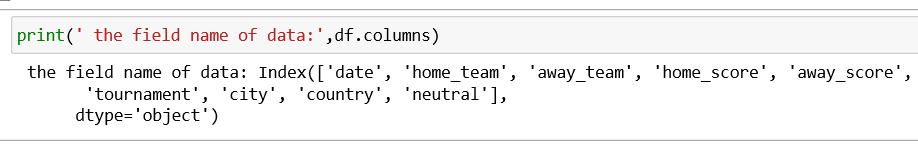
سپس data مورد نظر را با دستور ذیل فراخوانی می کنیم

و

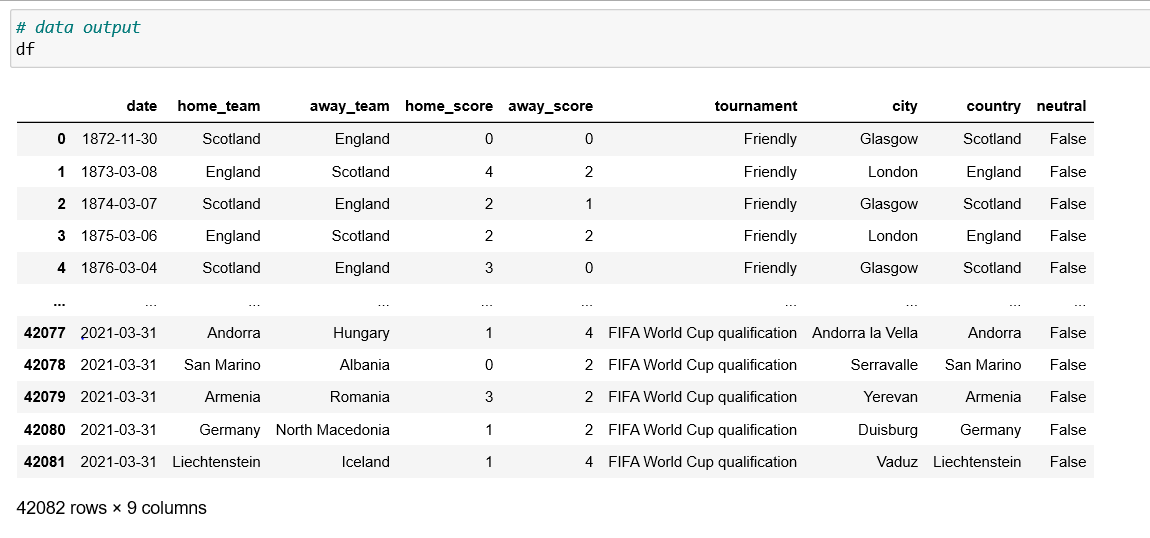
با دستور df.shape ابعاد DataFrame را به دست می آوریم ( 18 ستون و 48895 ردیف )

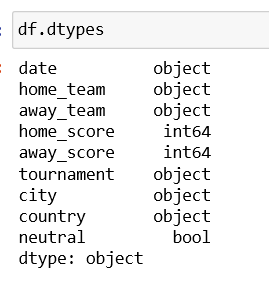
با دستور df.info() خلاصه مختصری از داده ها را به شرح ذیل دریافت می کنیم .

با دستور print(' the field name of data:',df.columns) نام ستون ها را دریافت کردم

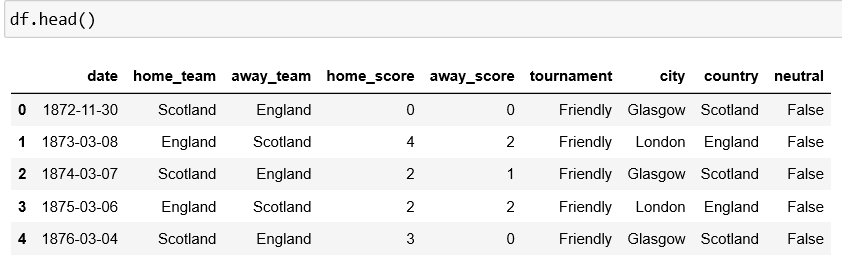


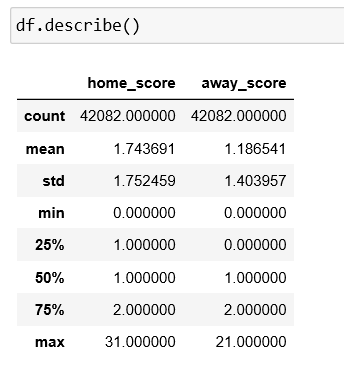
با دستور df جدولی از داده دریافت کردم



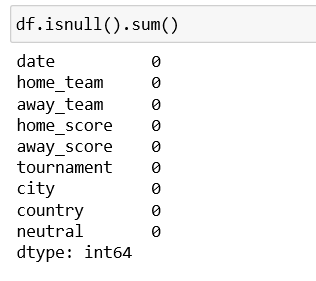
با دستور df.dtypes ، نوع دیتا ها را بررسی کردم

با دستور df.head() ، پنج خط اول داده ها را دریافت کردم

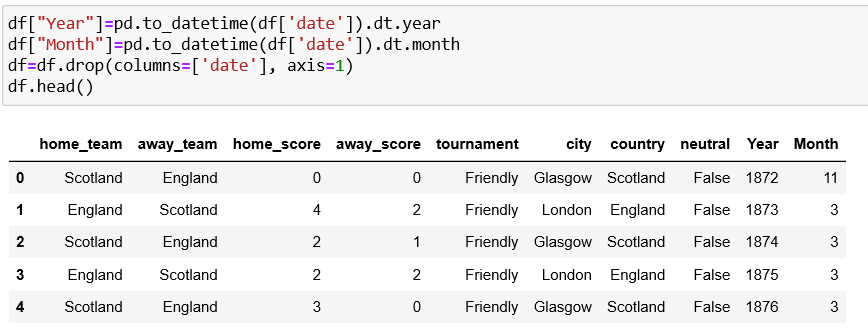




با دستور df.describe () خلاصه ای از اطلاعات عددی data را به دست آوردم

با دستور df.isnull().sum() داده های null را بررسی کردم

برای محاسبه ماه و سال بازی هر تیم از دستور ذیل استفاده کردم ودو ستون به ستون های جدول اضافه کردم



برای محاسبه نمره کلی هر تیم، از رتبه کسب شده در خانه و همچنین رتبه کسب شده در زمین حریف استفاده کردم و ستون Total score به جدول اضافه گردید

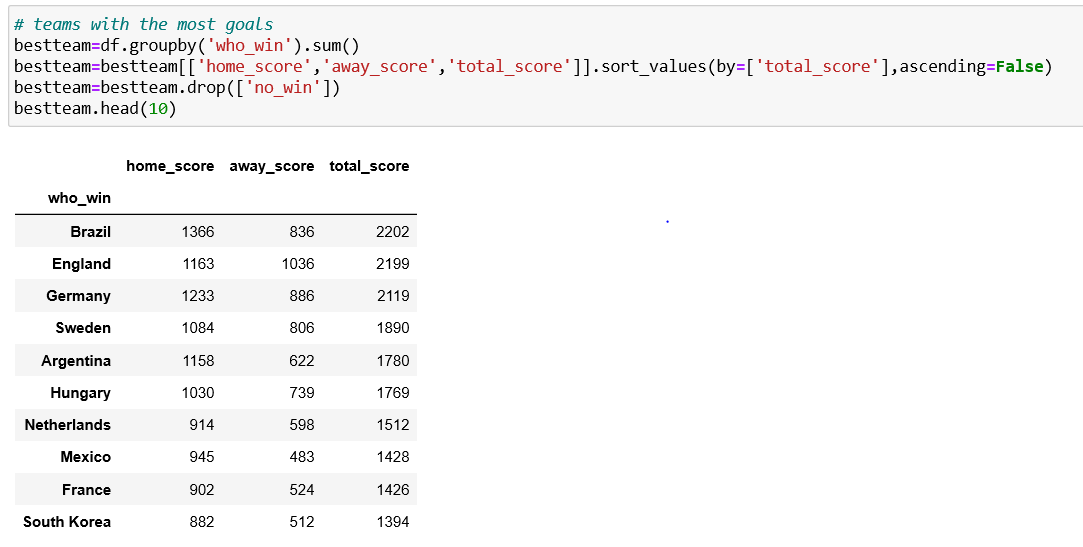


برای محاسبه نتیجه هربازی و اینکه چه تیمی در هربازی برنده شده است از دستور ذیل استفاده کردم و دو ستون دیگر تحت عنوان result و who-win به جدول داده ها اضافه شد

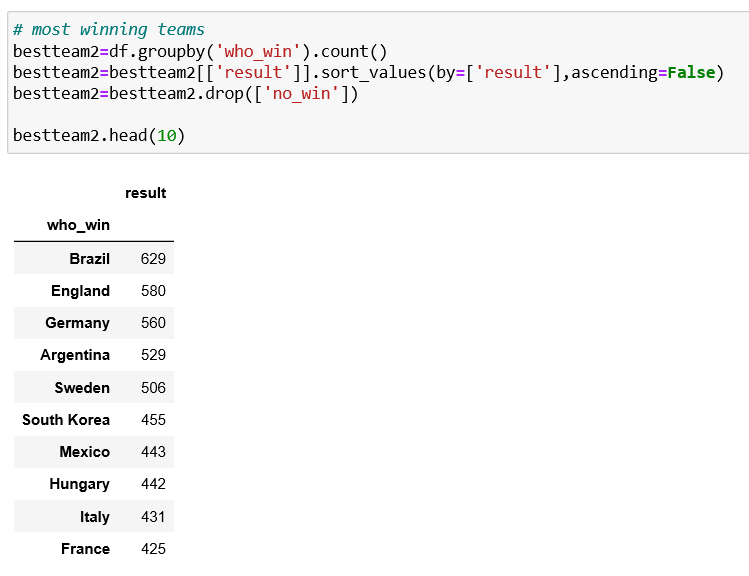


**Who is the best team of all time?**

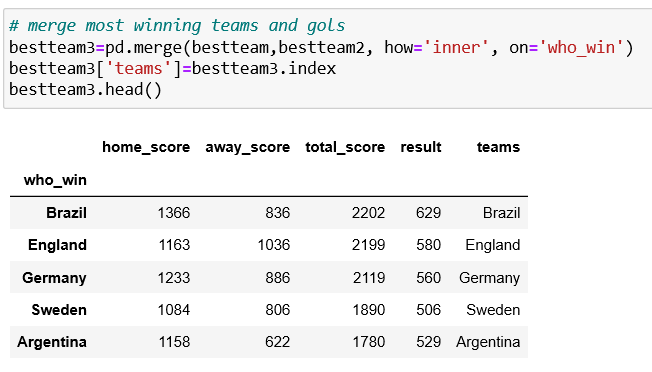
1. آیا تیمی که بیشترین گل را می زند می تواند بهترین باشد !



1. آیا تیمی که بیشترین برد را داشته می تواند بهترین باشد ؟



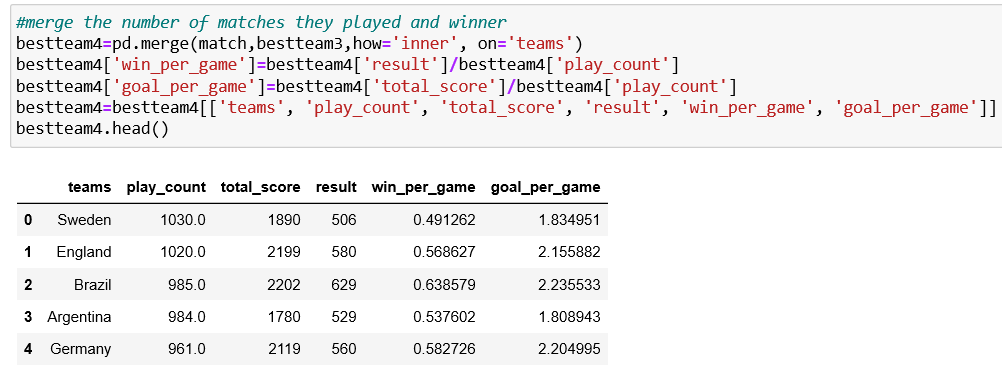
برای کنار هم قرار دادن نتیجه دو سوال بالا و تحلیل بهتر از این دستور استفاده کردم



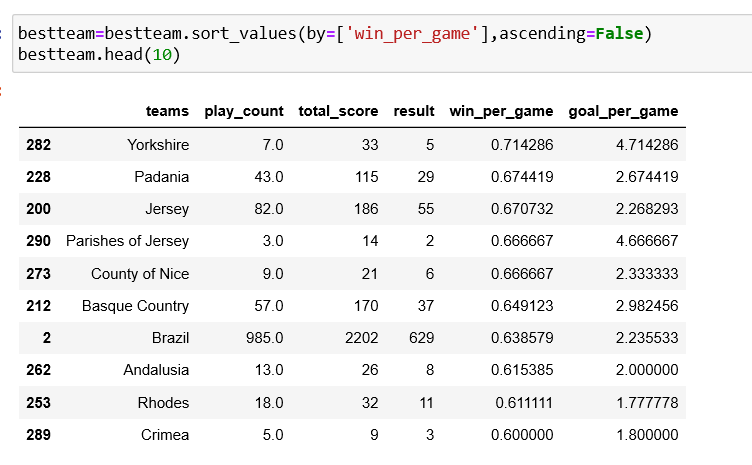
برای بدست آوردن میزان بازی هر تیم از کد دستوری ذیل استفاده شد



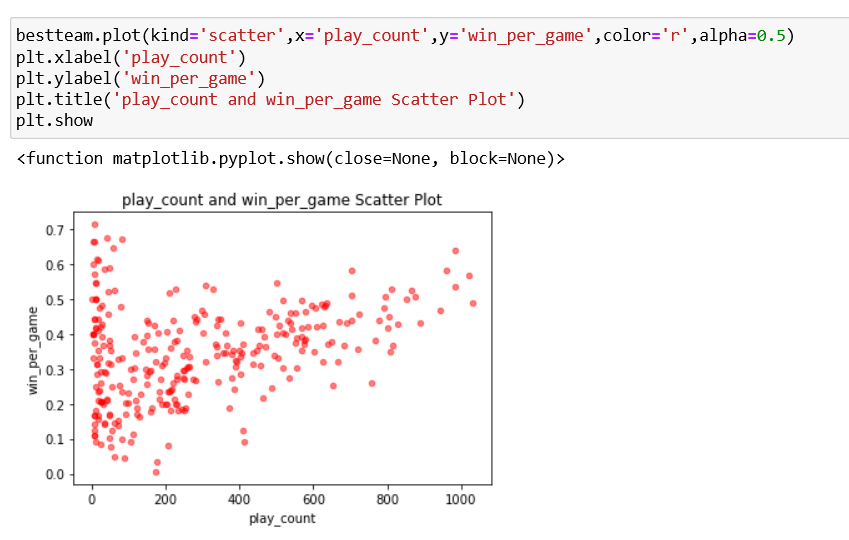
مجددا ستون play\_count را در جدول کنار داده بیشترین گل، بیشترین برد و .... با کد دستوری زیر گذاشته شد



برای بدست آوردن بهترین تیم از نظر بردن در بازی داده ها را sort می کنیم .



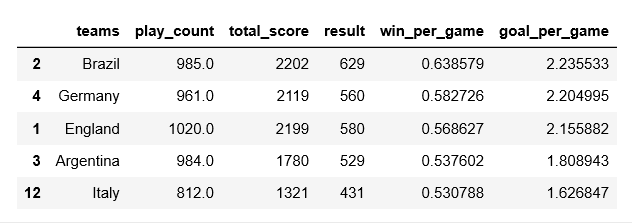
نمودار جدول فوق را با Scatter Plot رسم می کنیم.

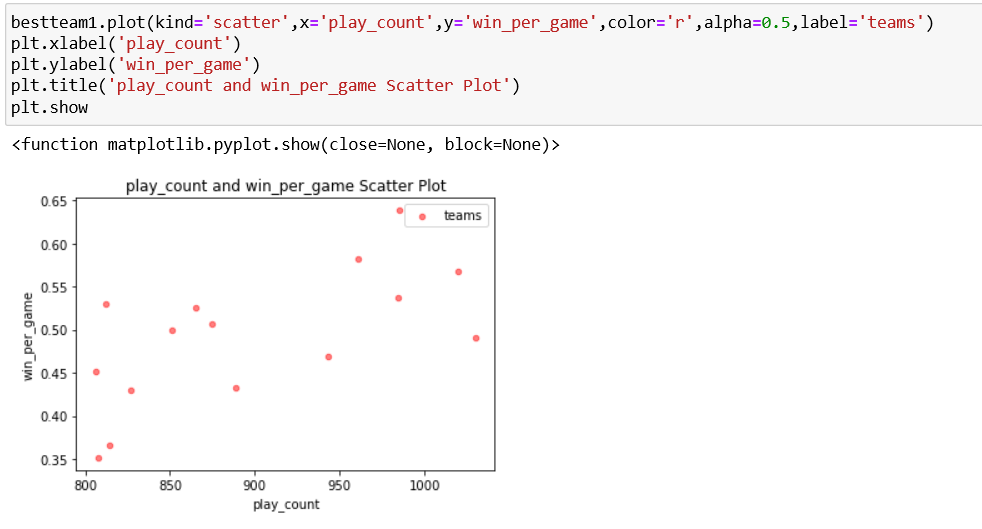


برای تشخیص بهترین تیم، میزان برد هر تیم و میزان بازی هر تیم را با هم لحاظ کردم با کد دستوری ذیل



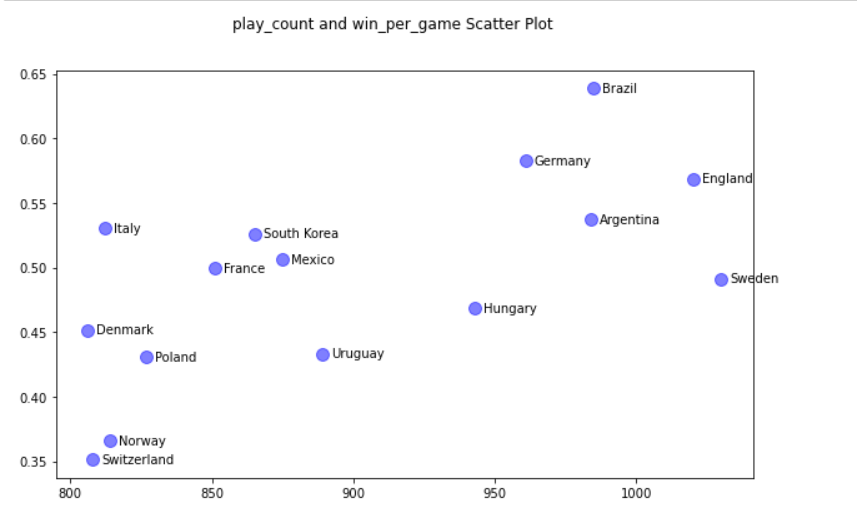
و جدول داده ها نیز به صورت ذیل است





برای نمایش بهتر نموداری با میزان برد هر تیم، نام هرتیم و میزان بازی آن با توجه به جدول رسم کردم



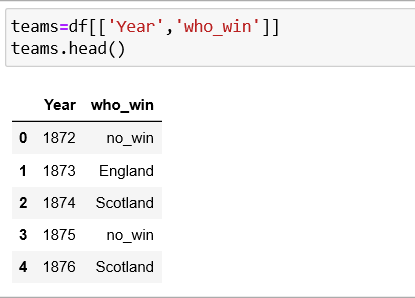


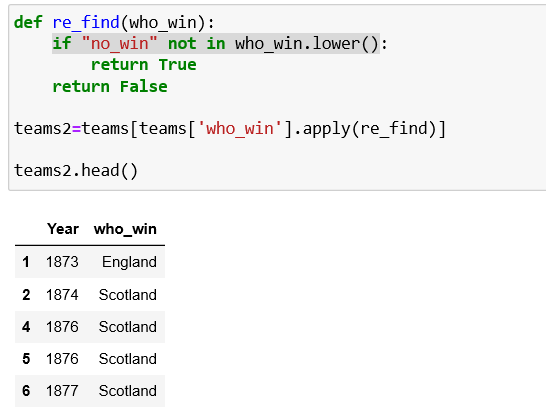
**Which teams dominated different eras of football**

مجددا با دستور df.head() پنج سطر جدول را مشاهده می کنیم



فراخوانی ستون های تیم برنده و سال



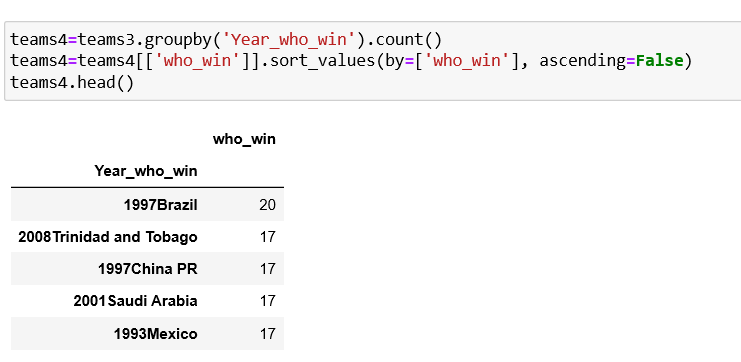


در این قسمت تابعی نوشتم برای اینکه no-win را در سطح پایینی جدول دیده شود

در این قسمت برای دریافت نتیجه سال برد هرتیم را کنار اسم آن قید کردم و ستون جدیدی ایجاد شد



و در نهایت sort داده



**What trends have there been in international football throughout the ages - home**

**advantage, total goals scored, distribution of teams' strength etc**

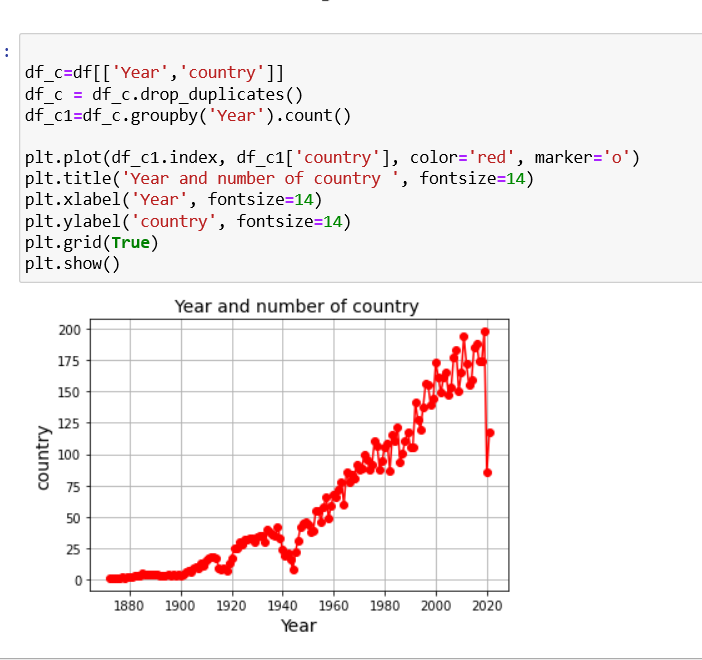


تعریف تابعی برای محاسبه برد در خانه و برد در خانه حریف



نتیجه نهایی با نمودار Pie

**how has the number of countries changed**

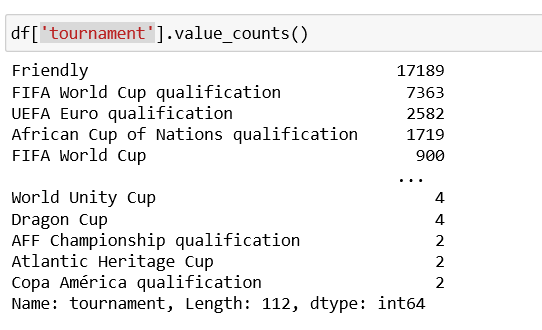


**Which countries host the most matches where they themselves are not participating in**

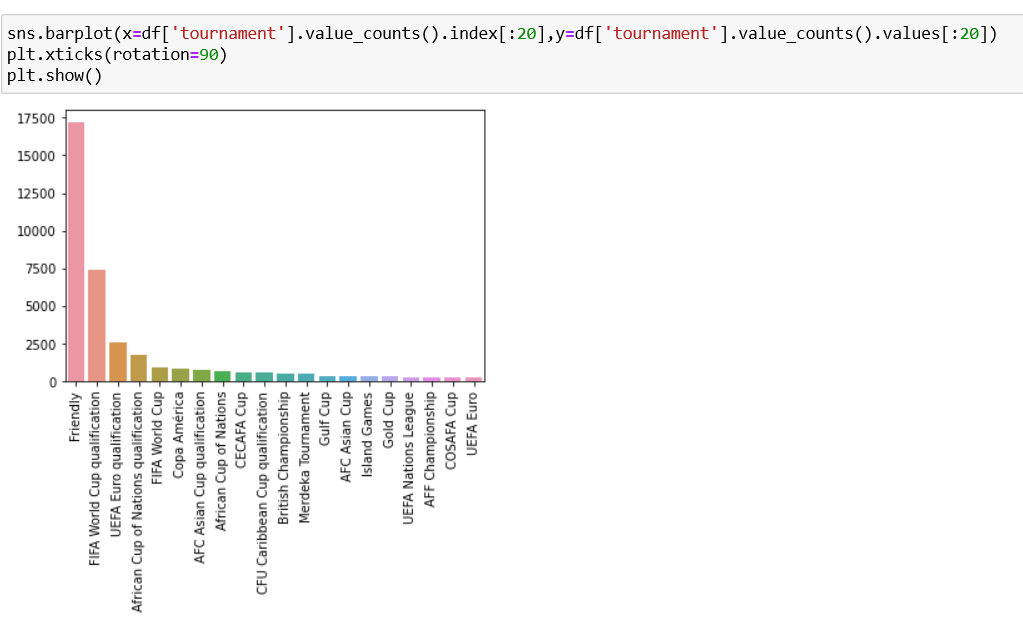


How much, if at all, does hosting a major tournament help a country's chances in the

Tournament

شمارش تعداد هر مسابقه

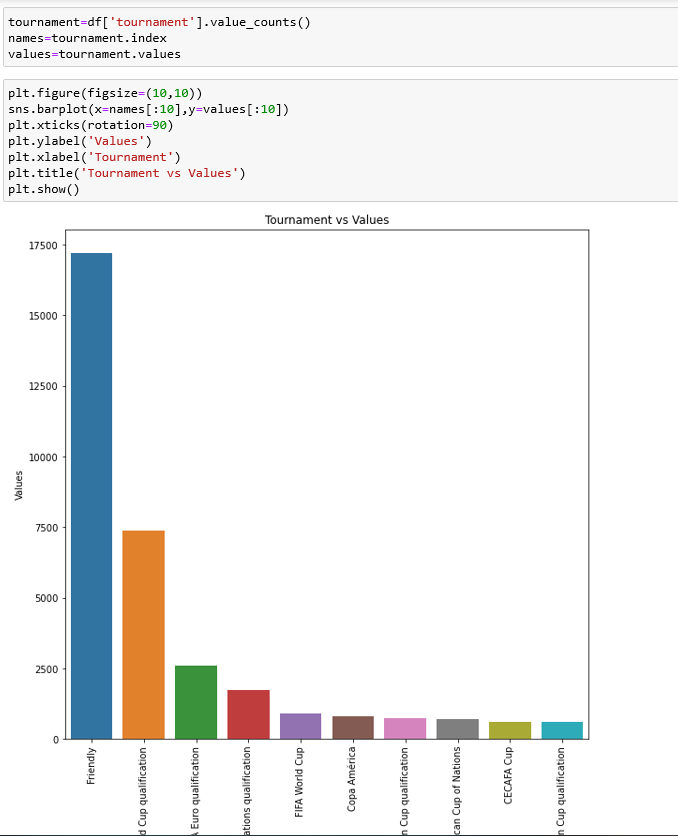
نمودار تعداد مسابقات

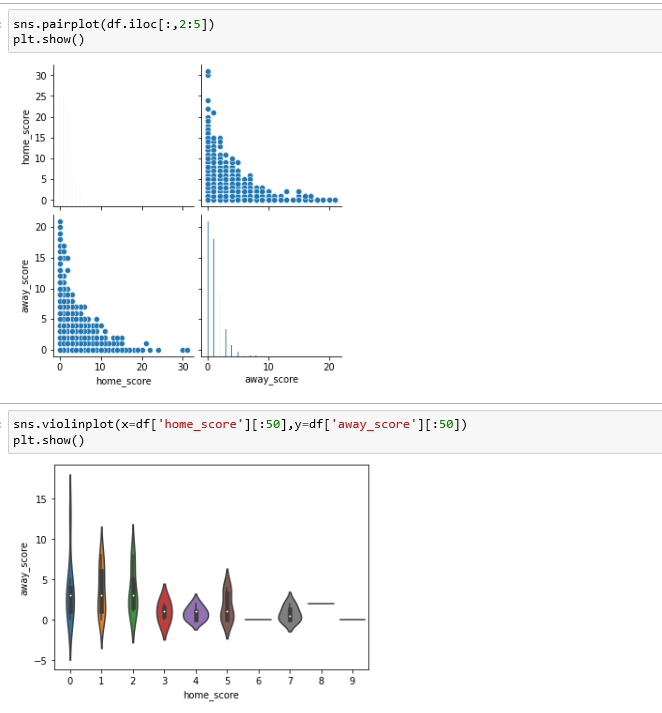


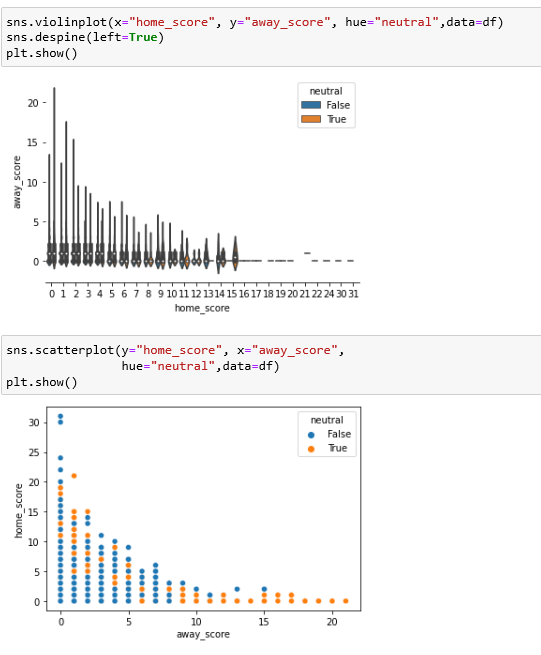
دریافت اطلاعات مسابقات



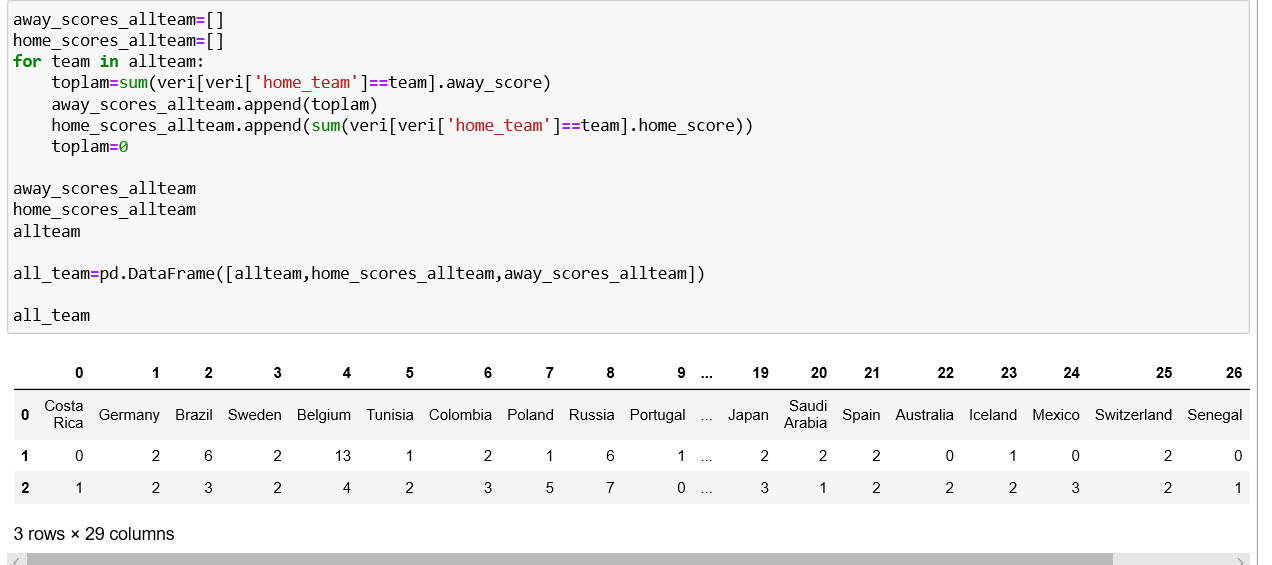
محاسبه طول داده مسابقات

نمودار مسابقات

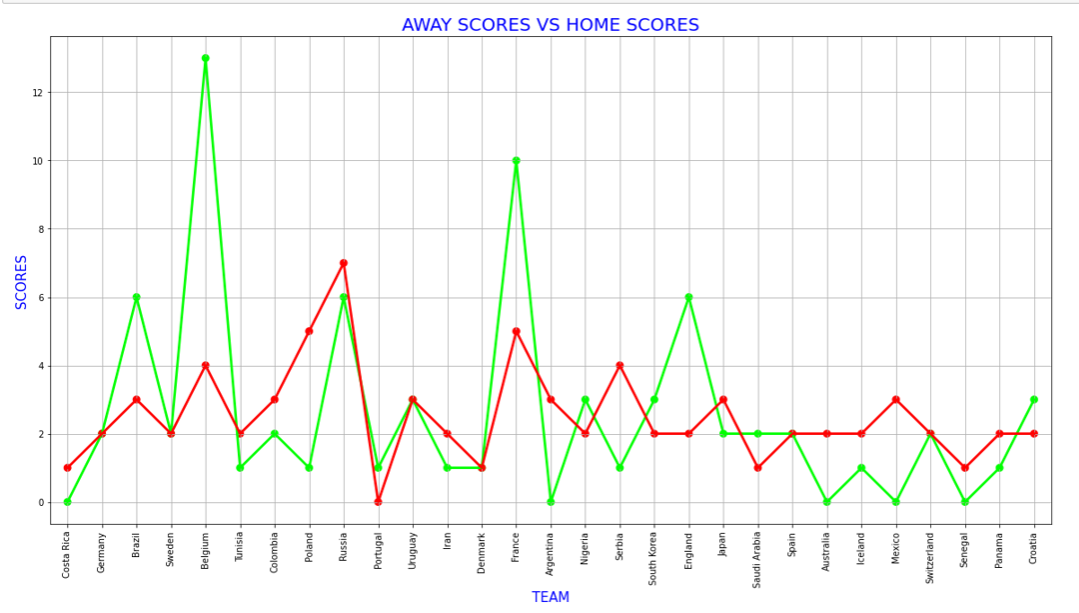




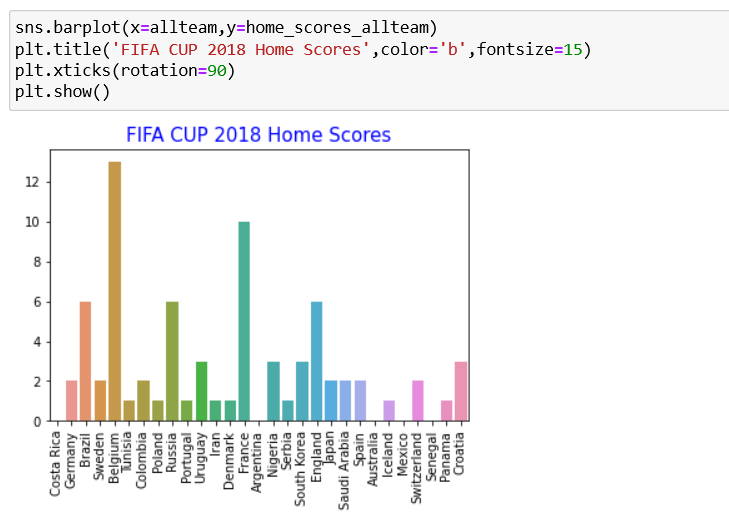
برای به دست آوردن همه مسابقات تیم ها در خانه و همه مسابقات تیم ها در خانه حریف

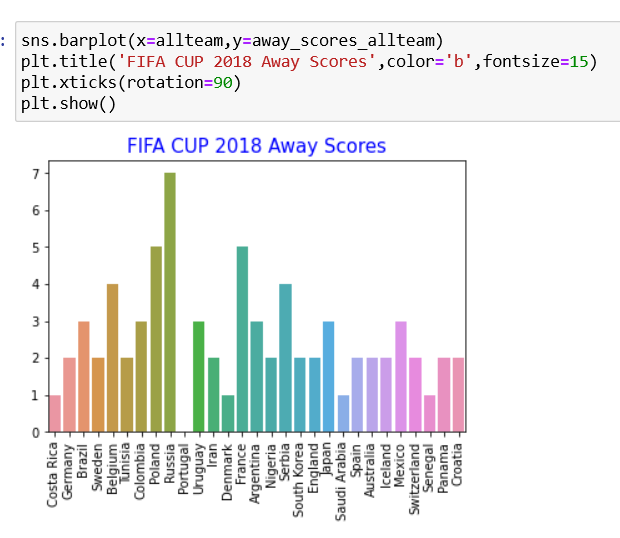


نمودار pointplot از بازی تیم ها در خانه و خانه حریف



نمودار plotbar از homescore





نمودار plotbar از awayscore