داده کاوی

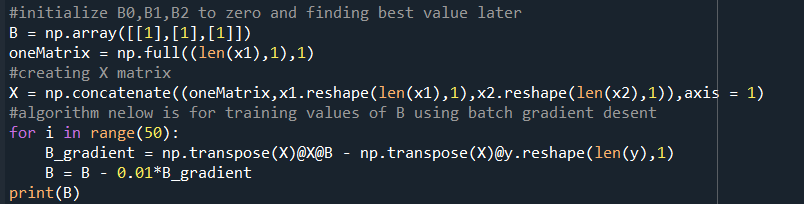
تمرین سری اول

قسمت عملی

سوال اول)

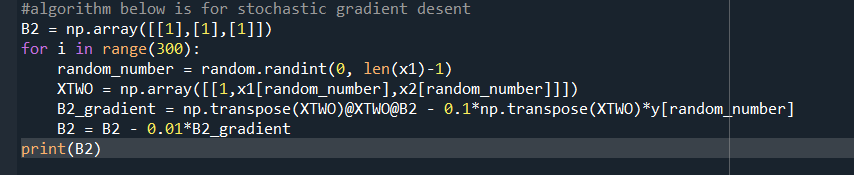
A

در کد زیر ما ماتریس B که شامل b0,b1,b2 هست را مقدار دهی اولیه میکنیم و در ادامه ماتریس X را میسازیم و حال در 50 گام مقدار B را آپدیت میکنیم به این صورت که در هر تکرار حلقه مقدار گرادیان را محاسبه میکنیم و مقدار B با توجه به گرادیان آپدیت میکنیم.

****

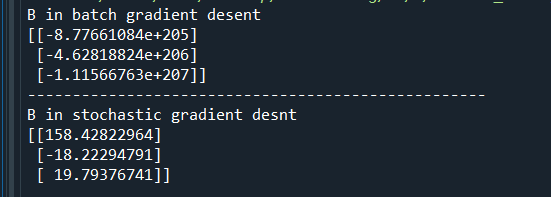
B

در کد زیر همانند کد بالا عمل میکنیم با این تفاوت که ماتریس X ترکیب تمامی داده ها نیست و فقط یک داده به صورت رندوم انتخاب میشود تا نتیجه را در هر دست آپدیت کنیم.



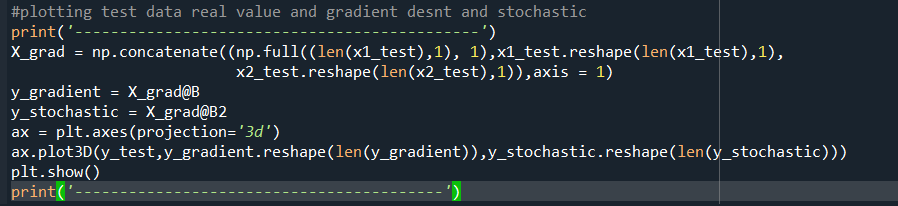
C

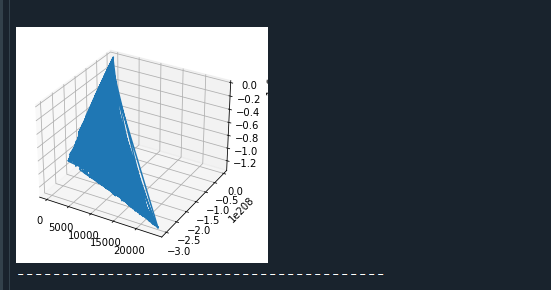
با توجه به خروجی زیر میتوان متوجه شد که روش گرادیان تصادفی نتیجه قابل قبول تری را میتواند به ما بدهد.



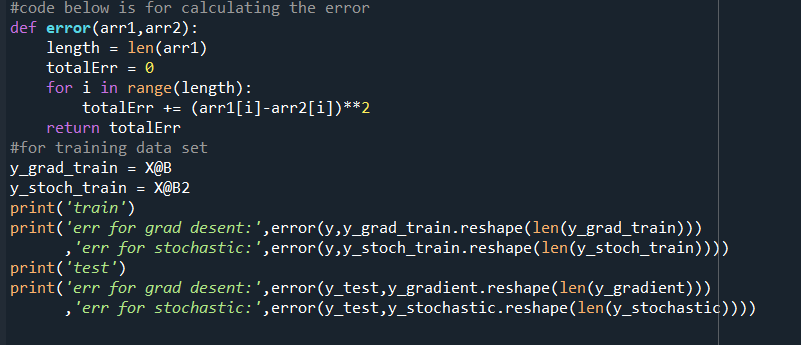
D

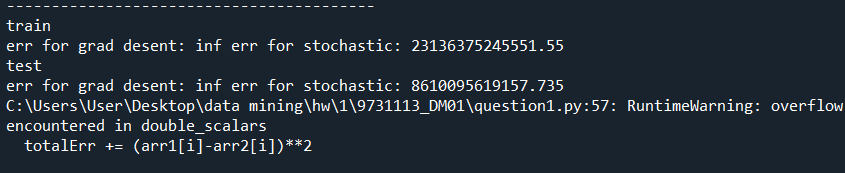
با به دست اوردن مقادیر y تست با توجه به نوع الگوریتم میتوانیم نمودار سه بعدی را رسم نماییم.





E





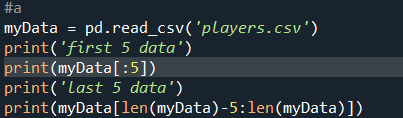
مقدار خطا برای روش گرادیان نزولی بسیار زیاد است و سرریز میکند اما برای روش گرادیان تصادفی نمایش داده شده است.

سوال دوم)

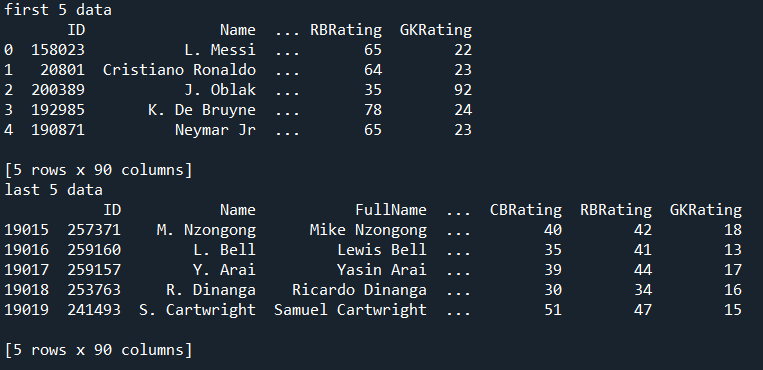
A

کد

خواندن داده ها و نشان دادن 5 داده اول و 5 داده آخر



خروجی



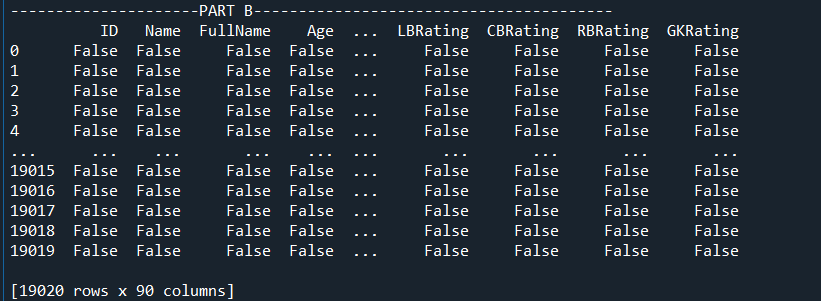
B

کد

با استفاده از isnull و با دیدن خروجی آن هایی که True میباشند به آن معناست که missing value ما هستند.



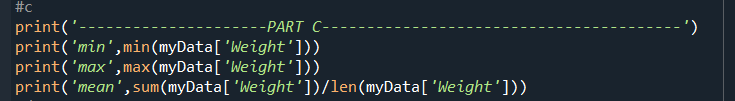
خروجی



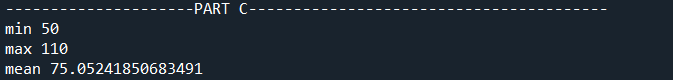
C

کد

محاسبه خروجی با استفاده از توابع min,max,sum



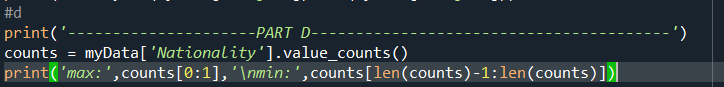
خروجی



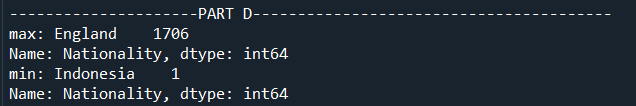
D

کد

تابع value\_counts تعداد تکرار هر value را در لیست پانداسی میدهد که از بزرگ به کوچک مرتب شده هستند.



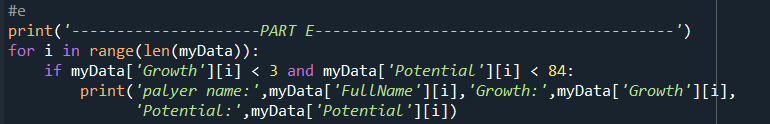
خروجی



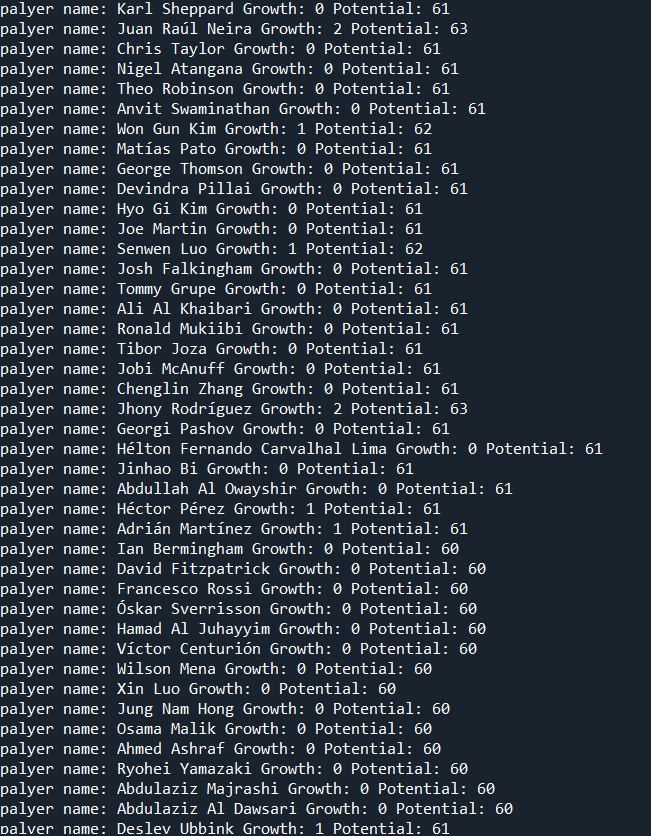
E

کد

یکی یکی داده هارا مورد پردازش قرار میدهیم و هر داده ای که شرط مورد نظر مارا داشت به عنوان خروجی چاپ میشود.



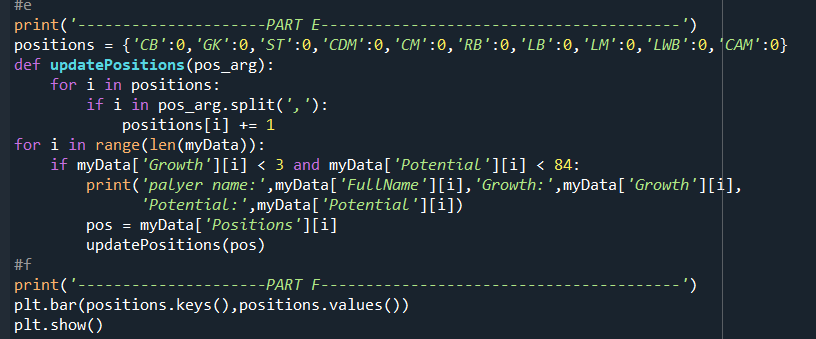
بخشی از خروجی



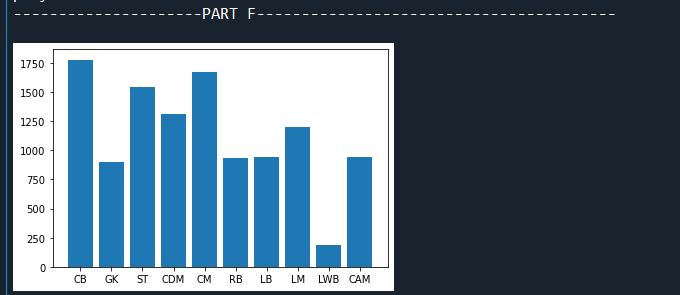
F

کد

از آنجایی که این سوال ادامه سوال قبلی است با استفاده از کد قسمت قبل و اضافه کردن یک دیکشنری که پست های متفاوت در آن نمایش داده شده است تعداد تکرار هر پست را محاسبه میکنیم و آن را روی نمودار میله ای نمایش میدهیم.



خروجی

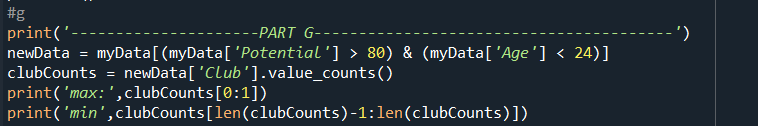


G

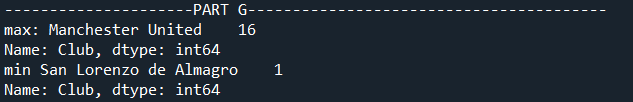
کد

داده جدیدی از فیلتر کردن داده های اولیه به دست می اوریم و با استفاده از تابع value\_counts نام هر باشگاه و تعداد بازیکنان آینده دار آن را محاسبه میکنیم.

توجه شود که برای فیلتر بازیکنان آینده دار آن دسته از بازیکنانی که سن کمتر از 24 و پتانسیل بالای 80 را دارا بودند در نظر گرفته شده است.



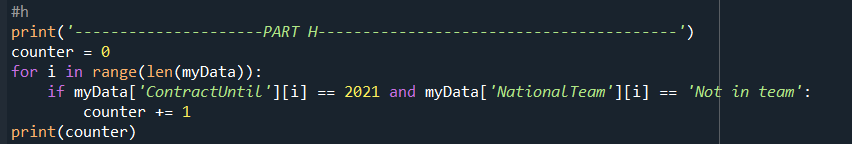
خروجی



H

کد

محاسبه کردن تعداد بازیکنان مورد نظر با قرار دادن دو محدودیت در شرط



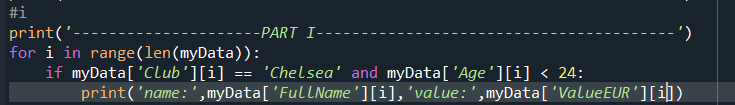
خروجی



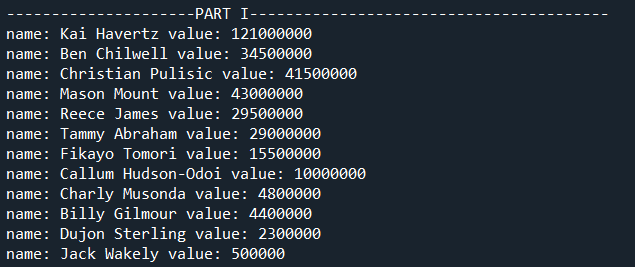
I

کد

پردازش همه داده ها در حلقه و چک کردن شرط مورد نظر



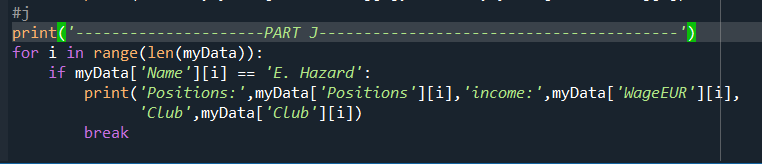
خروجی



J

کد

نشان دادن موارد خواسته شده در صورت تطابق اسم با اسم مورد نظر



خروجی

