

سوال ۱

بخش a:

برای پیاده سازی روش Gradient descent به تعداد NUMBER_OF_ITERATION بار گرادیان محاسبه شده و پس از اینکه نرمال شده از β کم شده. فرمول‌های استفاده شده:

$$X\beta = \hat{y}$$

$$(X\beta - y)^T (X\beta - y) = err$$

$$\nabla f(\beta) = X^T X\beta - X^T y$$

$$\beta_{next} = \beta - \alpha \frac{\nabla f(\beta)}{|\nabla f(\beta)|}$$

مقدار α برابر ۱ و مقدار NUMBER_OF_ITERATIONS برابر با ۱۰۰۰۰ در نظر گرفته شده است.

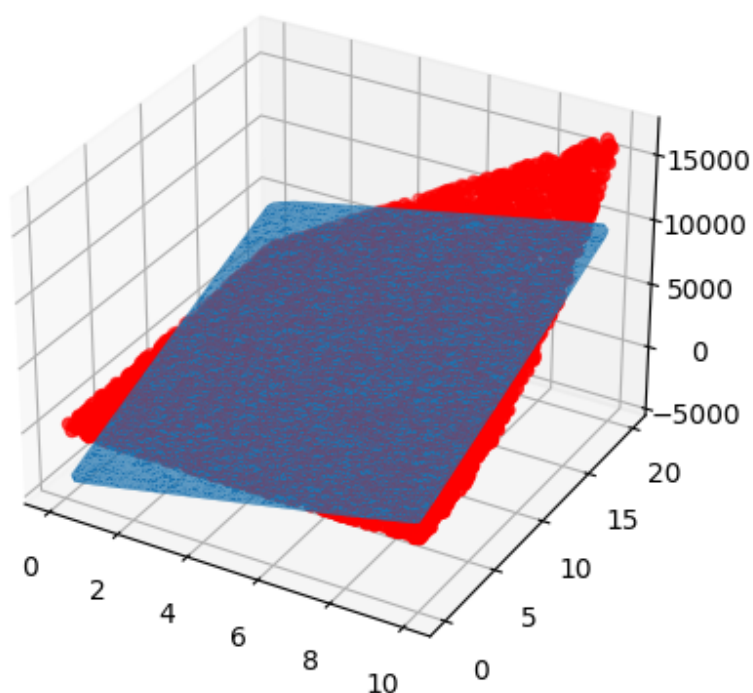
بخش b:

روش Stochastic gradient descent مانند بخش قبل پیاده سازی شده. تفاوت این بخش با بخش قبلی در این بود که در هر پیمایش به جای کم کردن گرادیان از بتا، یکی از داده‌ها به صورت اتفاقی انتخاب شده (فرض کنید داده‌ی i ام) و فرمول زیر اعمال می‌شود:

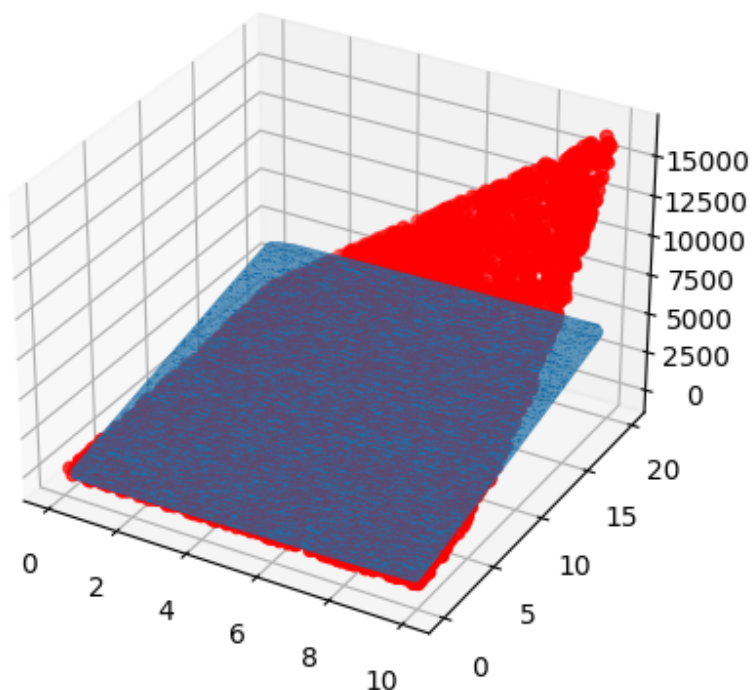
$$\beta_{next,j} = \beta_j - \alpha \frac{\delta err_i}{\delta \beta_j}$$

بخش C:

نتیجه‌ی روش Gradient descent بر روی داده‌های آموزشی: (رنگ قرمز y تست‌ها و رنگ آبی صفحه به دست آمده است)



نتیجه‌ی روش Stochastic gradient descent بر روی داده‌های آموزشی:

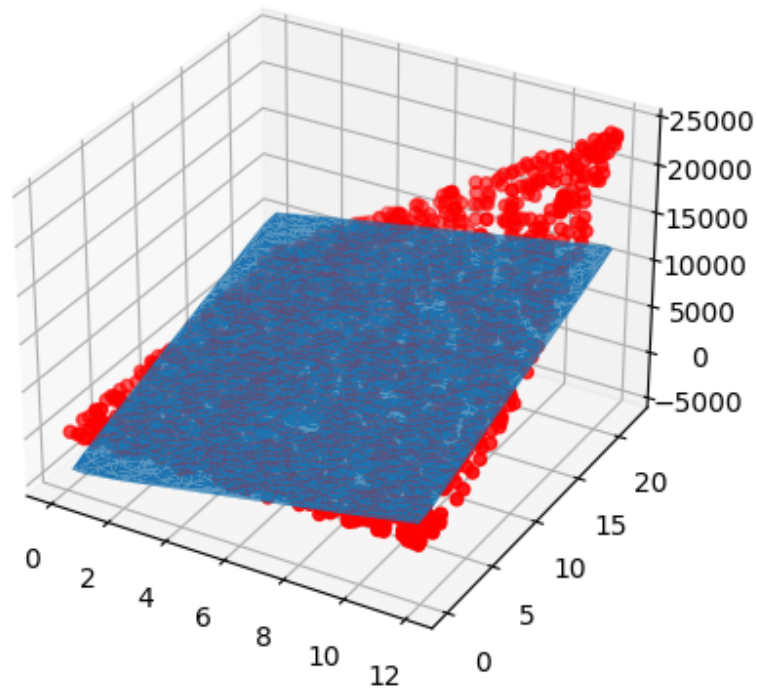


روش a مینیمم محلی پیدا کرده و روی آن ثابت مونده ولی روش b مینیمم محلی‌های زیادی پیدا کرده اما به جستجوی مینیمم کلی خارج شده از آن.

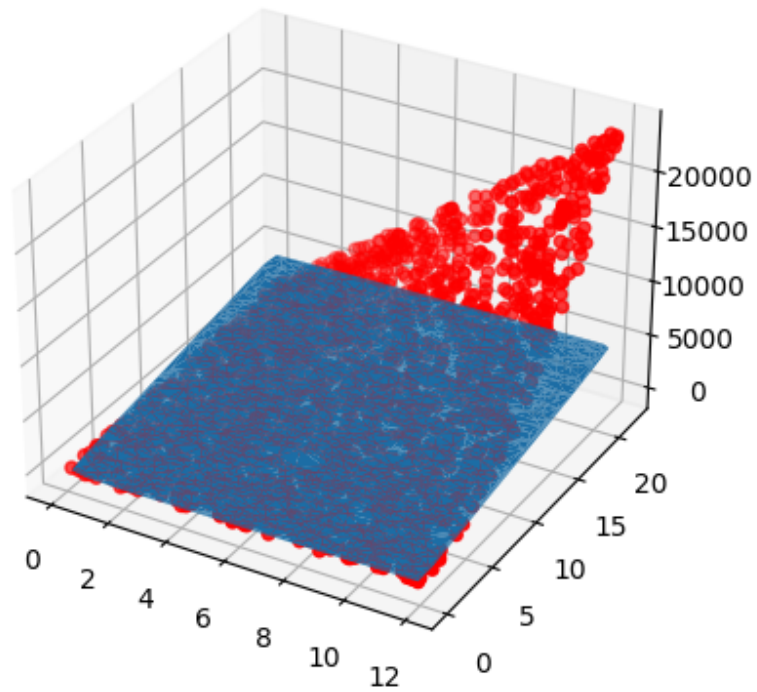
در نهایت اما روش a مینیمم محلی‌ای که پیدا کرده بهتر از حالت نهایی b است (برای داده‌های تمرینی).

بخش d:

نتیجه‌ی روش Gradient descent بر روی داده‌های تست: (رنگ قرمز y تست‌ها و رنگ آبی صفحه به دست آمده است)

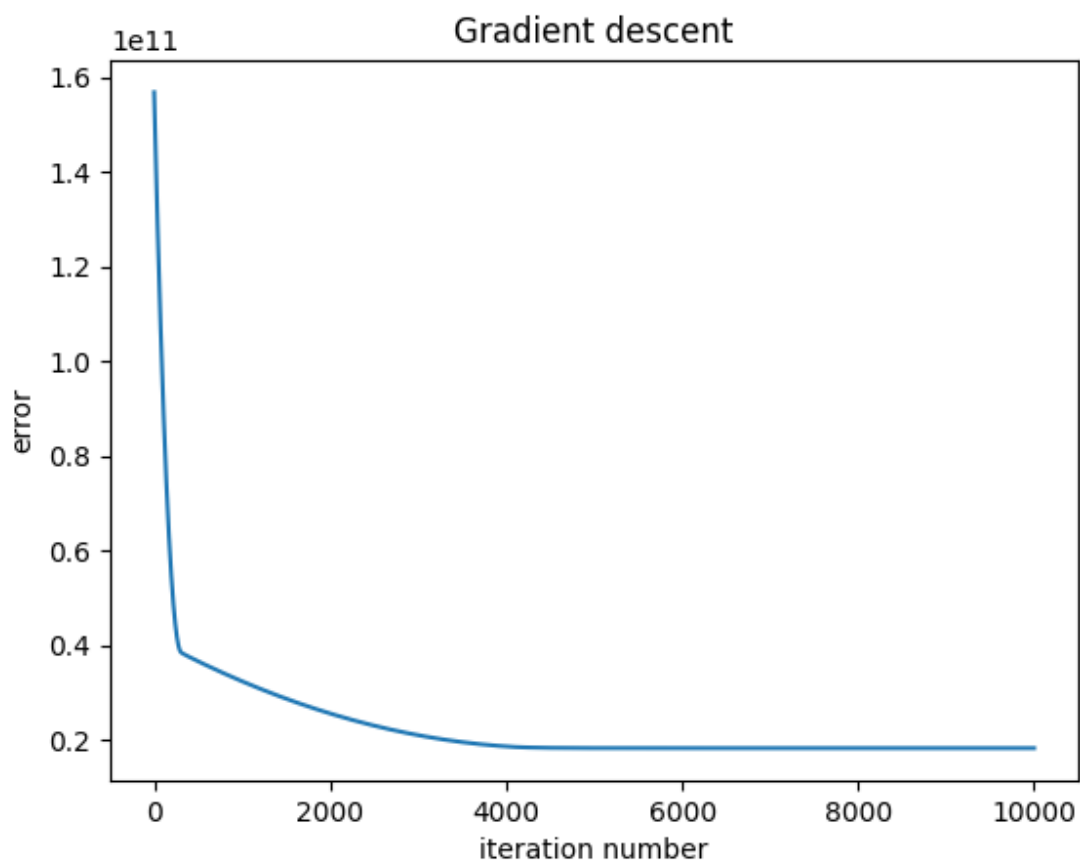


نتیجه‌ی روش Stochastic gradient descent بر روی داده‌های تست:

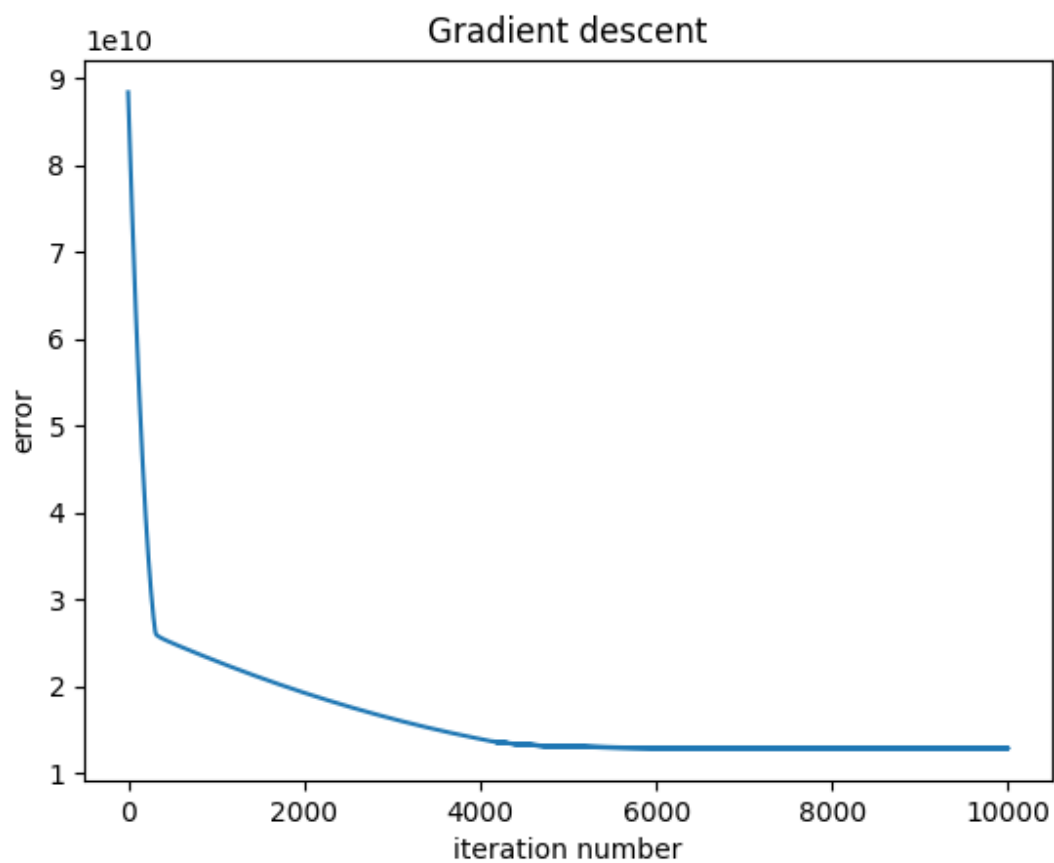


بخش e:

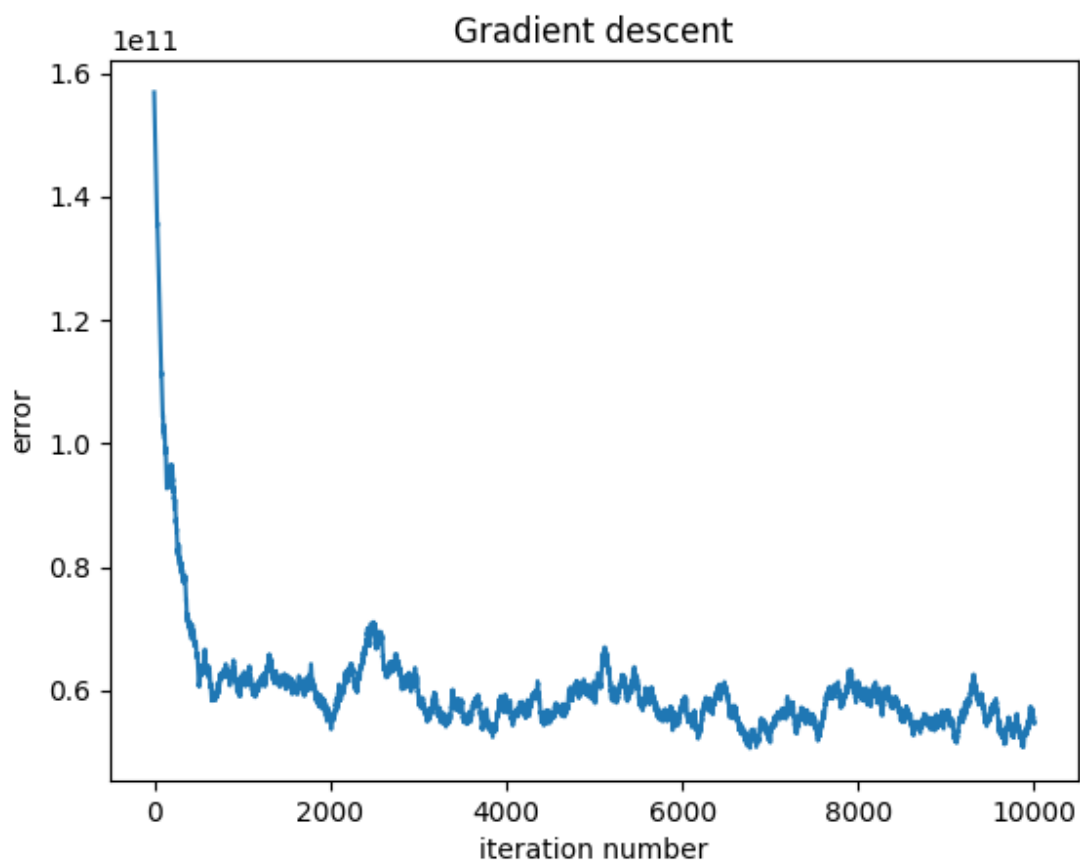
نمودار تابع خطای روش Gradient descent برای داده‌های آموزشی:



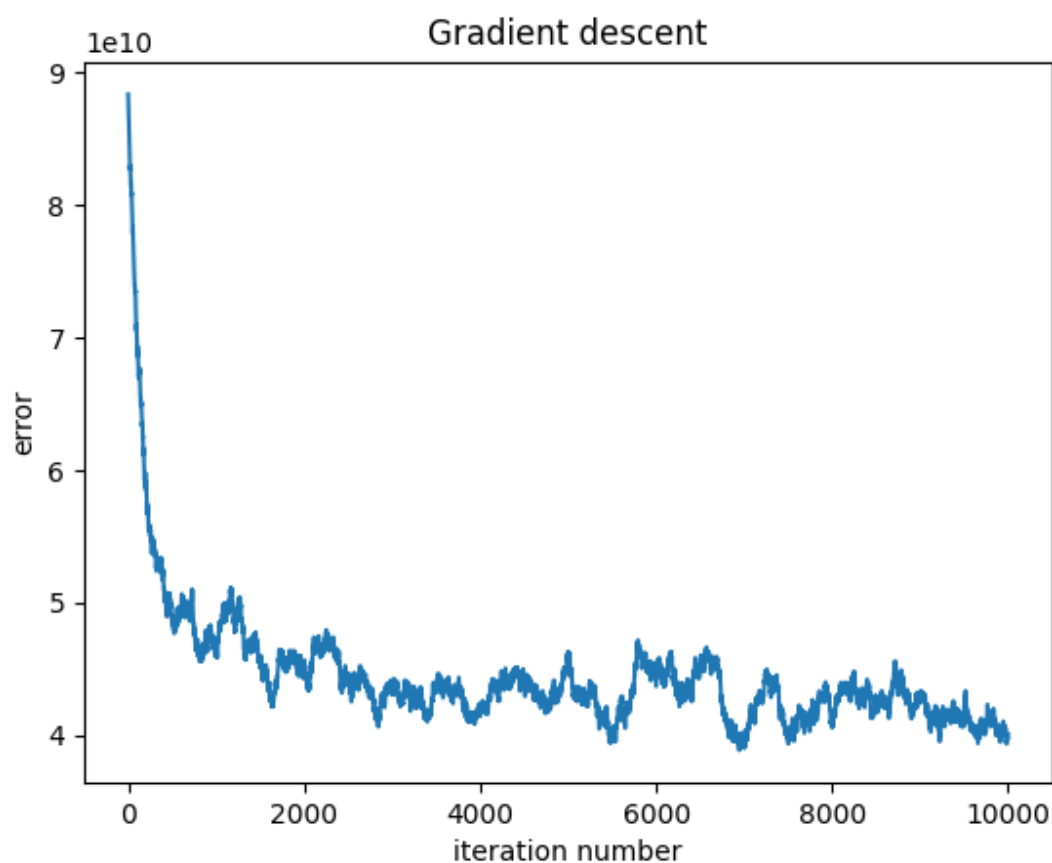
نمودار تابع خطای روش Gradient descent برای داده‌های تست:



نمودار تابع خطای روش Stochastic gradient descent برای داده‌های آموزشی:



نمودار تابع خطای روش Stochastic gradient descent برای داده‌های تست:



همانطور که می‌بینید روش b درون مینیمم‌های محلی گیر نکرده و به دنبال مینیمم تابع می‌گردد.

سوال ۲

در تمام سوالات کوئری‌های خواسته شده بر روی داده‌ها اعمال شده و خروجی نمایش داده شده.
خروجی‌ها در این گزارش به صورت خلاصه بیان شده و کامل آن‌ها توسط کد خروجی داده می‌شود.
برای مدیریت اجرای هر بخش در فایل config.py می‌توانید مشخص کنید که کدام بخش‌ها اجرا شوند.

بخش a:

داده‌های ابتدایی:

```
{'ID': '158023', 'Name': 'L. Messi', 'FullName': 'Lionel Messi', ...  
{'ID': '20801', 'Name': 'Cristiano Ronaldo', 'FullName': 'C. Ronald...
```

داده‌های انتهایی:

```
{'ID': '253763', 'Name': 'R. Dinanga', 'FullName': 'Ricardo Dinang...  
{'ID': '241493', 'Name': 'S. Cartwright', 'FullName': 'Samuel Cartw...
```

بخش b:

تعدادی از داده‌های گم شده:

```
...  
Player 229516's NationalPosition is missing!  
Player 229516's NationalNumber is missing!  
Player 251019's NationalPosition is missing!  
...
```

بخش c:

ماکزیمم، مینیمم و میانگین وزن بازیکنان:

```
Min weight: 50  
Max weight: 110  
Average weight: 75.05241850683491
```

بخش d:

کشورهای دارای بیشترین و بازیکنان:

Country England has maximum number of players, which is 1706

Country Malaysia has minimum number of players, which is 1

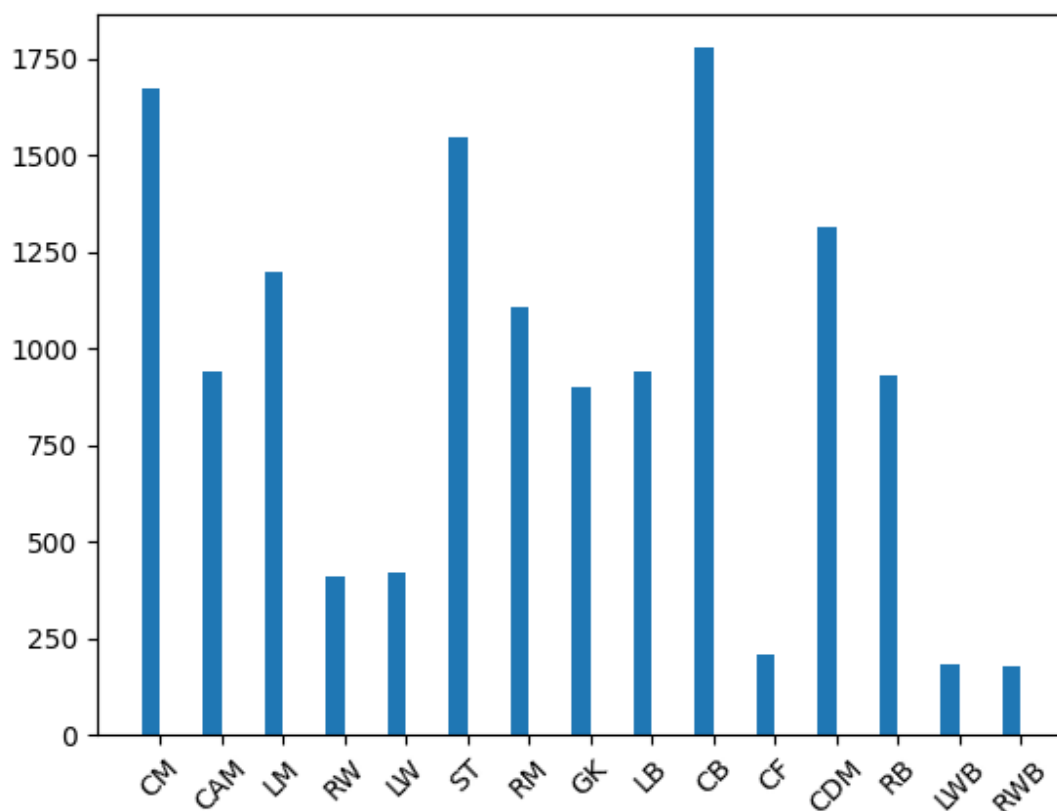
بخش e:

بازیکنانی که رو به رشد نیستند:

Queried players count is 8261 and they are ['169195', '169416', '215333', ...

بخش f:

نمودار بازیکنان سوال قبل بر اساس موقعیت بازی:



بخش g:

باشگاه با بیشترین بازیکنان آینده‌دار

Chelsea has maximum number of promising players which is 14.

بخش h:

تعداد بازیکنانی که ۲۰۲۱ قراردادشان تمام می‌شود و در تیم ملی نیستند برابر با 6727 است.

بخش i:

ارزش بازیکنان زیر ۲۴ سال چلسی:

Player's values: ['121000000', '34500000', '41500000', '43000000',
'29500000', '29000000', '15500000', '10000000', '4800000', '4400000',
'2300000', '500000']

بخش j:

موقعیت، درآمد و باشگاه فعلی آقای هازارد.

E. Hazard's club is Real Madrid, his wage is 350000 and his position is LW.