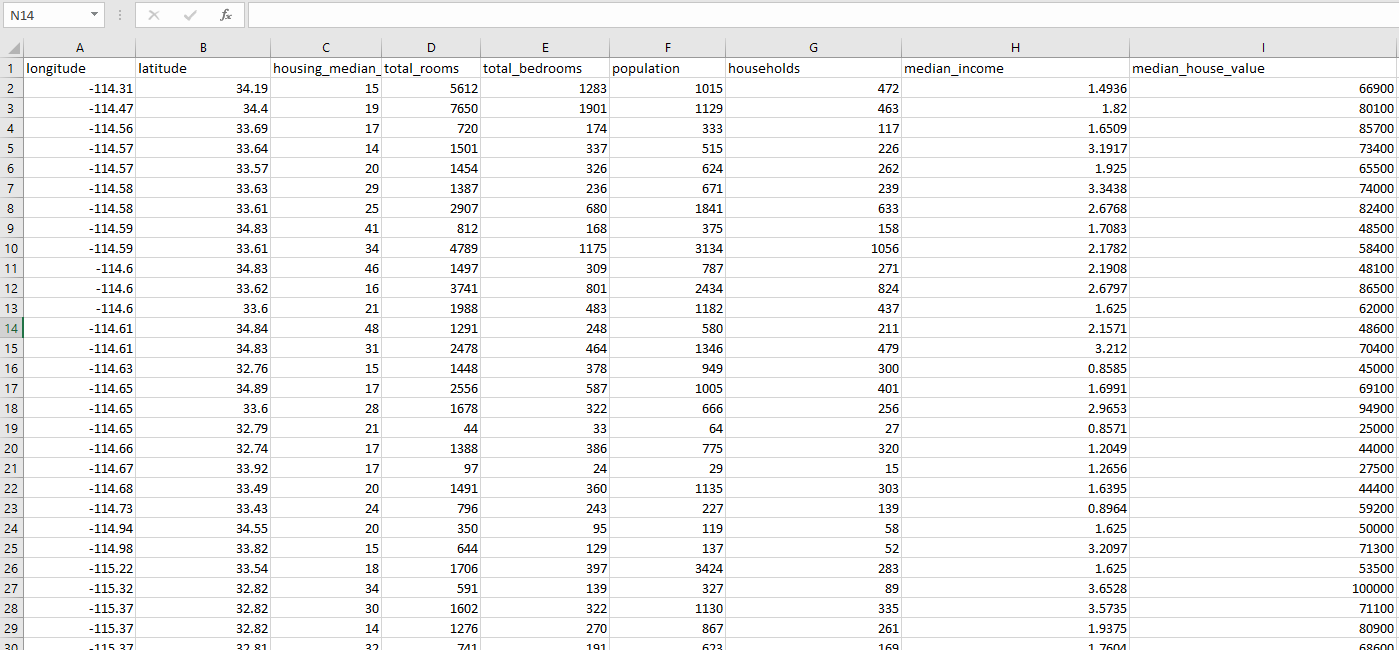
جواب تکلیف اول)

تحقیقات بنده در این فایل دیتا مه به فرم .csv هست به نحو زیر هست.



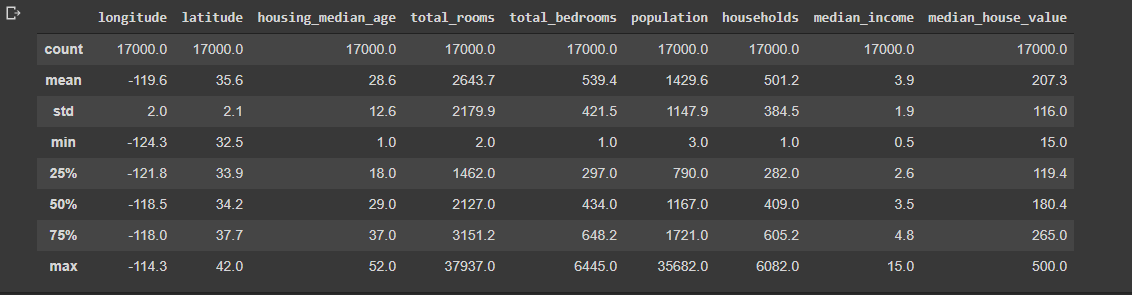
در این فایل دیتا دارای 9 ، title هست که شامل

* Longtitude وlatitude که اطلاعات حاوی عرض جغرافیایی رو در اختیار دارد. یا همان مختصات جغرافیایی رو در اختیار دارد.
* و شامل housing\_median که حاوی اطلاعات متوسط منازلی که در آن منطقه جغرافیایی هستند
* : total\_rooms که حاوی اطلاعت تعداد اتاق ها هست.
* :Total\_bedrooms که حاوی اطلاعات شامل تعداد اتاق خواب ها هست.
* :Population شامل اطلاعات جمعیت در آن منطقه
* Households :شامل اطلاعت تعداد خانواده هایی که در آن منطقه زندگی میکنند.
* Median\_income :شامل اطلاعات این که متوسط میزان حقوق و درآمد هست.
* :Median\_house\_value شامل اطلاعات میزان متوسط ارزش هر خانه هست.

که شامل 17000 دیتا در سطرها و 9 ستون هست.

جواب تکلیف دوم ) Quantile

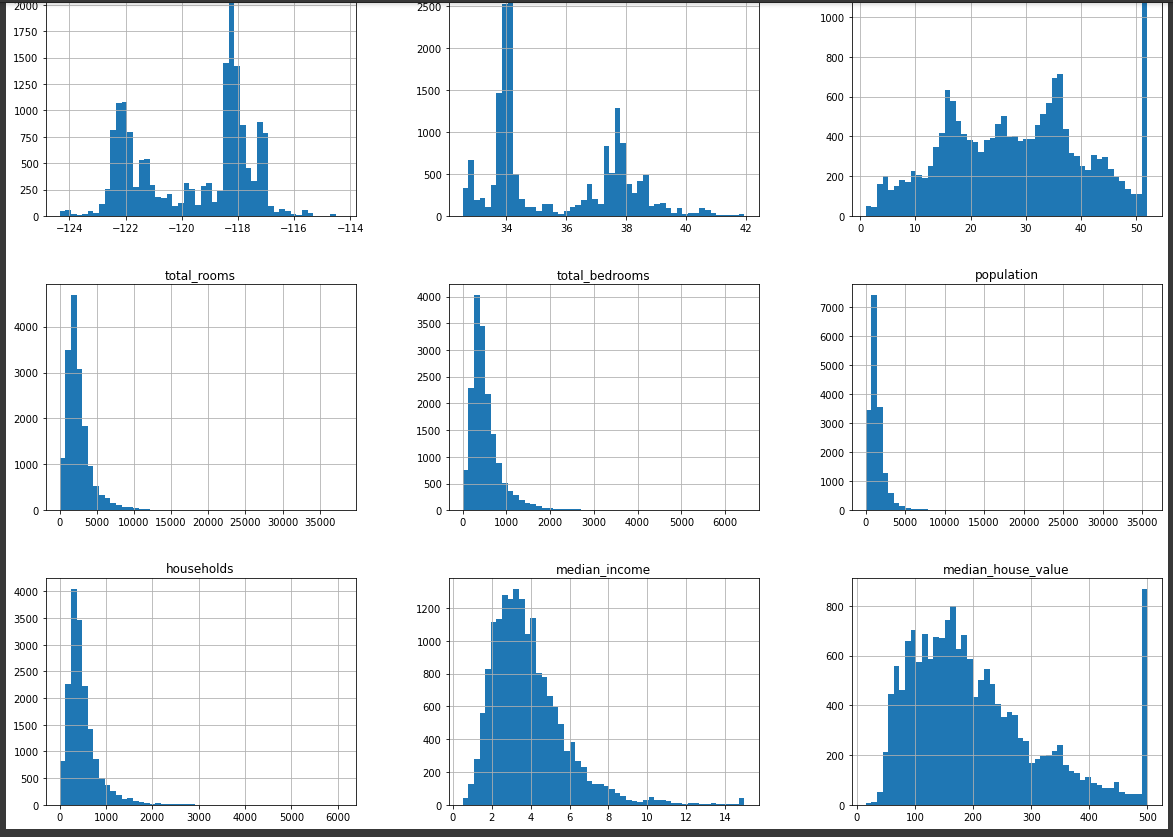
ها همان چارک های ما هستن در علوم آمار   
و مقدار max در اینجا 37937.0 هست که در قسمت total\_rooms که مطابقت دارن چون با یک سرچ ساده در فایل اکسل میتونیم به این نتیجه برسیم و مقدار Max با چارک یا همان Quantile ها نیز مطابقت دارن



جواب تکلیف سوم)

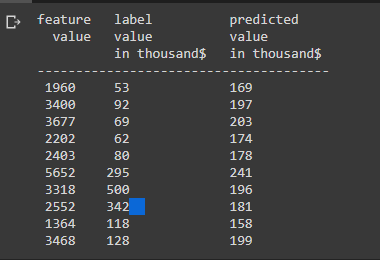
در این قسمت با ران گرفتن قطعه کد به هیستوگرام median\_income میرسییم که قبلا بررسی کرده بودیم که به همون متوسط درآمد میشه تعبیرش کرد.

و واحدش باید به دلار باشه. چون فایل دیتامون داره خانه ها و خانوار های کالیفورنیا (California) رو مورد بررسی قرار میده پس باید و قانونا باید میزان متوسط درآمد ماهینه باشه ولی scale هیستوگرام ما باید 1000 باشه.



جواب تکلیف چهارم)

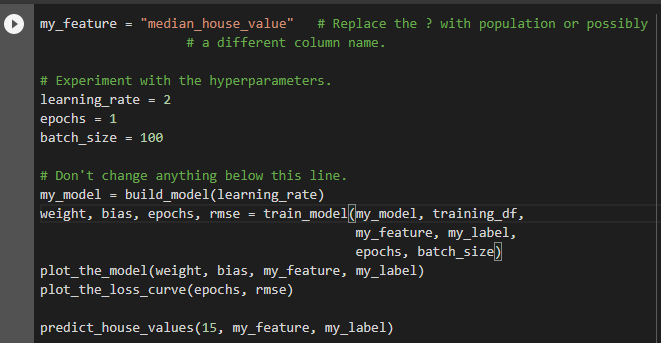
خیر مقدار واقعی و مقدار حدس زده شده ی مدل ما با واقعیت خیلی تفاوت فاحشی دارد و تقریبا در یک حدس خیلی نزدیک بوده ولی بقیه تفاوت خیلی زیاده داشته در واقع accuracy و دقت مدل ما پایین هست.



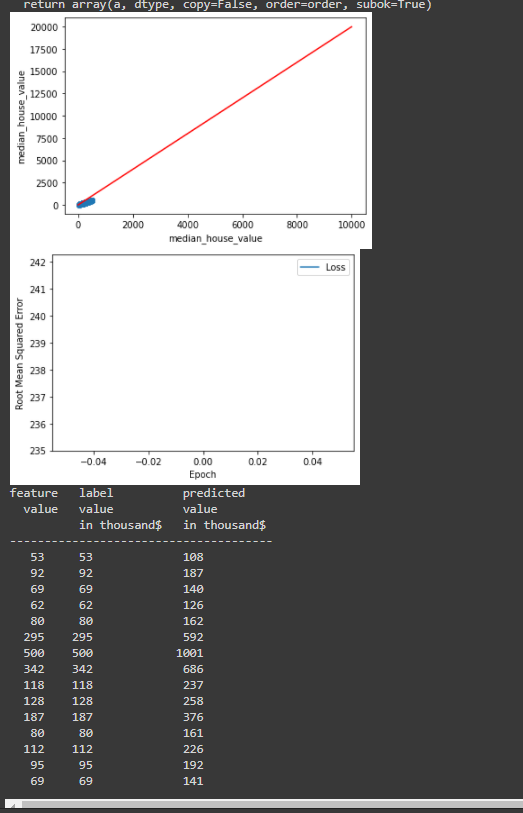
جواب تکلیف پنجم)

مقدار my\_feature را با median\_house\_value تغییر دادیم و دوباره train دادیم مدل را.

کد پس از تغییر:



خروجی کد به طور زیر شد :

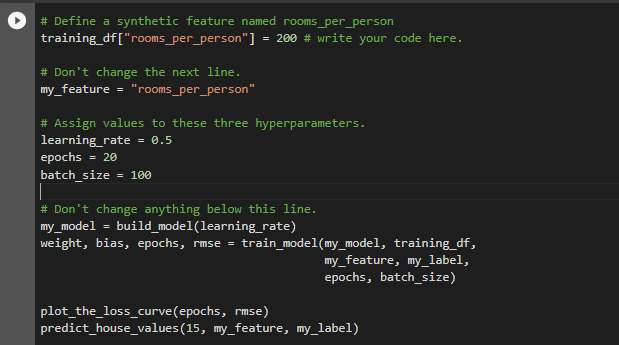


که باز مدل مورد مطالعه ما به خوبی تعلیم ندید و مقدار پیش بینی اونطور که انتظار میرفت موفق نبوده و با قیمت پیش بینی شده فاصله گرفت و نشون میده مقدار دقت یا accuracy پایین هست.

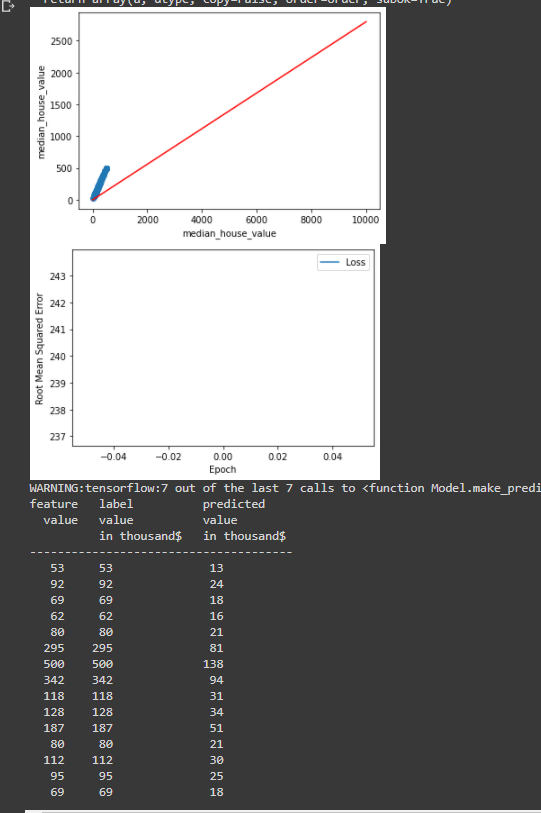
جواب تکلیف ششم)

با تغییر دادن مقدار epochs و learning\_rate و batch\_size به مقادیر زیر در تصویر و تغییر دادن دیتا فریم

Accuracy مدل ما نسبت به خروجی های اخیر بسیا بسیار بالا رفته و تفاوت آنچنان فاحشی نداشته نسبت به خروجی های دیگه که ضمیمه شد و نتیجه میگیریم accuracy بالاتر رفته هست.

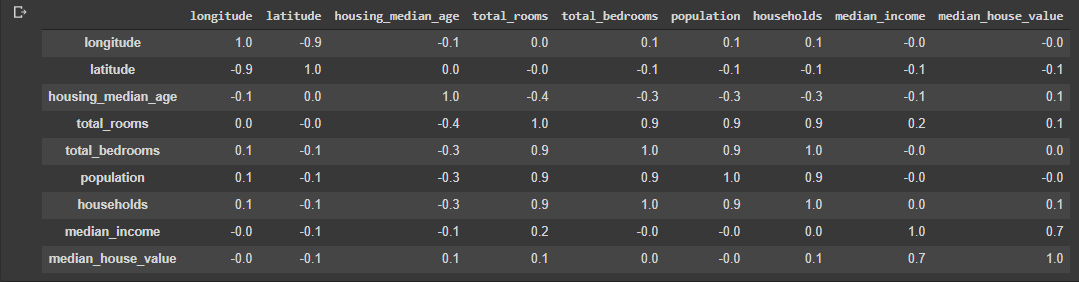


خروجی:



جواب تکلیف هفتم )

ماتریس همبستگی به سادگی جدولی است که همبستگی را نمایش می دهد. این معیار بهترین استفاده در متغیرهایی است که رابطه خطی بین یکدیگر را نشان می دهد. تناسب داده ها را می توان بصورت بصری در یک پراکندگی نشان داد. ... یک ماتریس همبستگی از سطرها و ستونهایی تشکیل شده است که متغیرها را نشان می دهد.



جواب تکلیف هشتم )

با توجه به داکیومنت ها و با توضیحات داخل نوت بوک بنده به این نتیجه رسیدم که بهترین یا شاید بهتر باشه اینطور بیان کنم مناسب ترین feature یا ویژگی از نظر من برای تعلیم و train میتونه feature ، median\_house\_value و median\_income باشن . چون تعلیم و برررسی همین دو متغیر برای رسیدن به این نتیجه که حالا در این سناریو و case برای خرید خونه در منطقه ای مثل کالیفرنیا کفایت میکنه و عوامل دیگه ای مثل population و households زیاد اطلاعات جالبی برای ما نیست و اصلا به درد نمیخورن برای تعلیم و تخمین همینطور lantitude که یک سری info و اطلاعاتی هستن که باید تمیز بشن از دیتاست درواقع و همونطور هم که در طی این پروژه و نوت بوک سعی کردیم به این نتیجه برسیم و با آزمون و خطا به 2 ویژگی ذکر شده پرو بال بیشتری داده میشه چون طبق تعلیمات و ران هایی که خودم گرفتم و مقدار accuracy بسیار بالاتر بودن و به نظر اطلاعات حیاتی تری هستن نسبت به بقیه سطر ها