به نام خدا

دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده برق و کامپیوتر

**هوش مصنوعی پاییز 98**

**پروژه صفر**

**پیش بینی قیمت خانه**

**نام و نام خانوادگی**

**علیرضا زارع نژاد اشکذری**

**شماره دانشجویی**

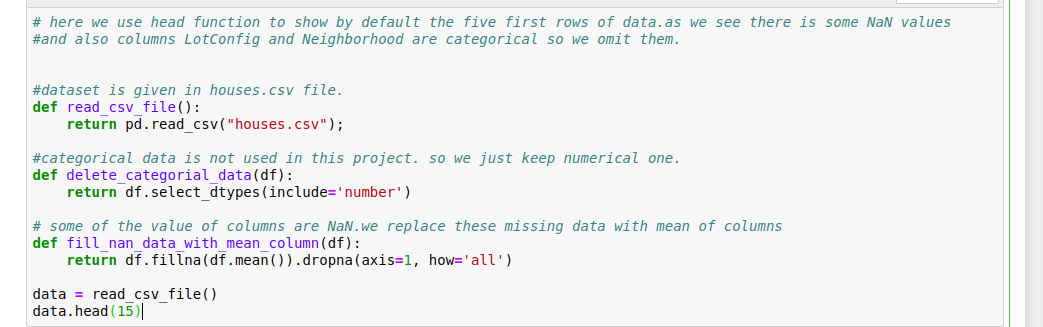
**810196474**

# شرح مختصر پروژه

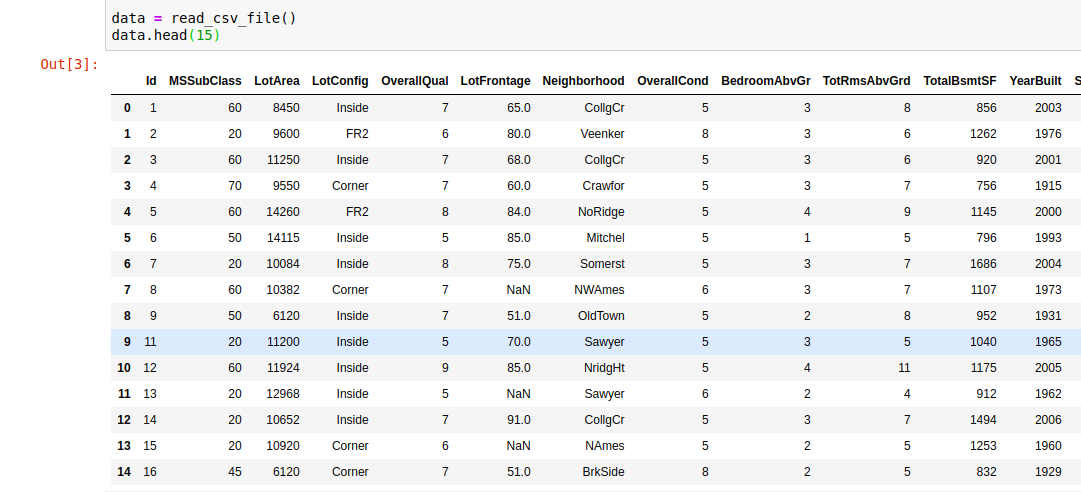
هدف از انجام این پروژه طراحی یک مدل برای پیش بینی قیمت خانه می باشد. این مدل به عنوان ورودی تعدادی از مشخصه های مربوط به یک خانه را دریافت کرده، و به عنوان خروجی قیمت تخمین زده شده برای خانه را برمی گرداند . همچنینن آشنایی با کتابخانه های pandas ، numpy ، matplotlib نیز در پایتون از دیگر اهداف این پروژه می باشد.

# داده های ورودی و بخش اول پروژه

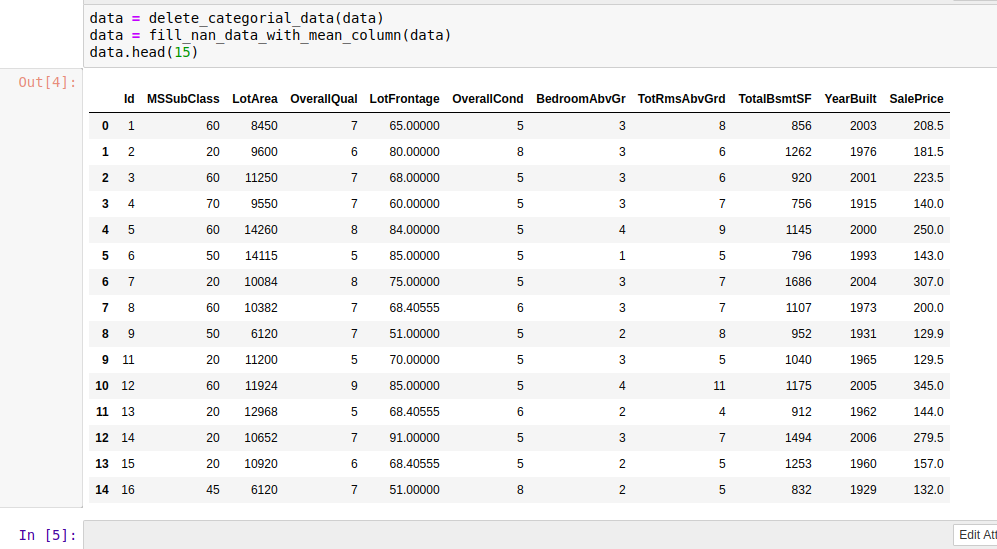
مجموعه داده ها در یک فایل اکسل به نام houses.csv داده شده است.مطابق زیر ابتدا با استفاده از panda فایل ورودی را می‌خوانیم و خروجی آن را به عنوان یک دیتا فریم در data می ریزیم. فایل داده شده شامل ۱۱۳۴ سطر و ۱۱ ستون می‌باشد. هر کدام از ستون‌ها مختص یکی از خصیصه های خانه می‌باشد. با توجه به آنکه داده‌های می‌توانند طبقه ای و یا داده‌ای باشند و در این پروژه نیار به کار با داده‌های طبقه ای نیستیم لذا ستون مرتبط به آن را حذف می‌کنیم . همچنین مقدار بعضی از خانه‌های جدول NaN می باشد. راه برخورد با آن جایگزین کردن میانگین ستون مورد نظر به جای آن هاست. کد مرتبط در زیر آورده شده است.



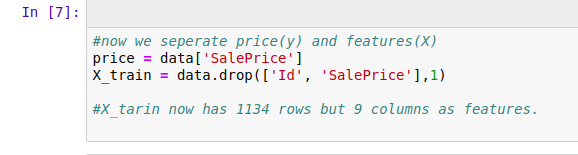
در زیر با استفاده از تابع head چند سطر ابتدایی ورودی را بدون تغییر خواهیم دید.



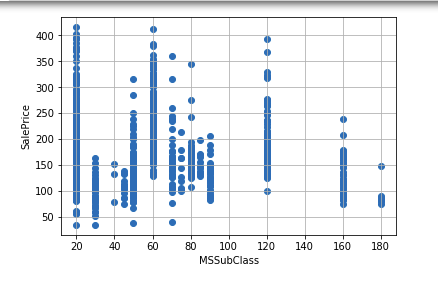
همچنین پس از صدا کردن توابع بالا موارد اصلاح شده در سطر های زیر قابل مشاهده است.

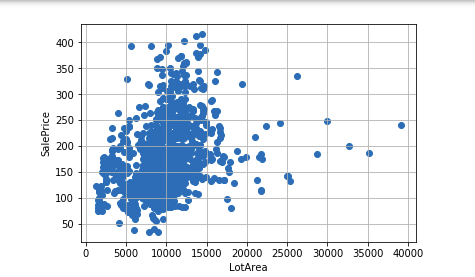


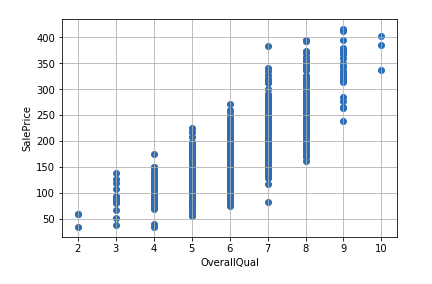
برای راحتی کار داده‌های قیمت خانه و ویژگی‌های مورد نظر را در price و X\_train ذخیره می کنیم.

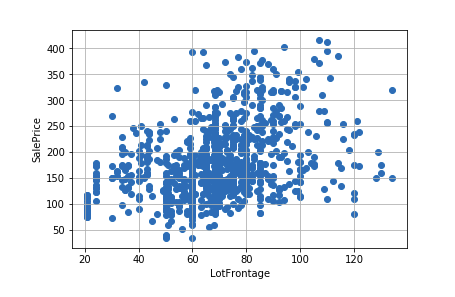


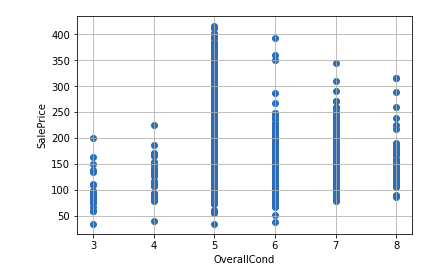
در انتها نمودار قیمت واقعی با توجه به هر کدام از ۹ ویژگی یعنی ستون‌های مرتبط را آورده‌ایم که با استفاده از scatter آورده شده است.

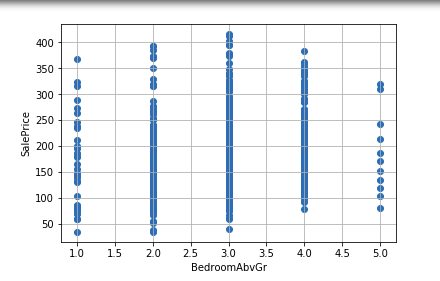


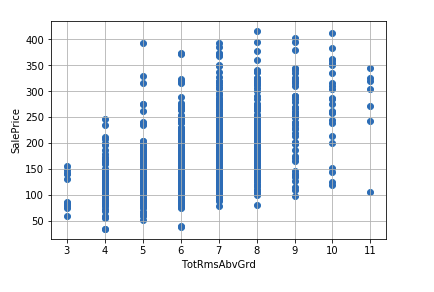


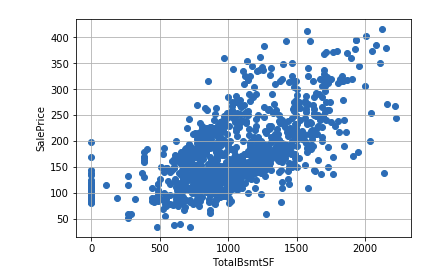


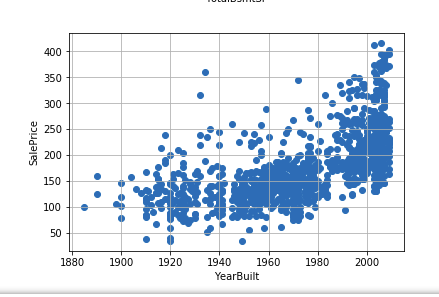




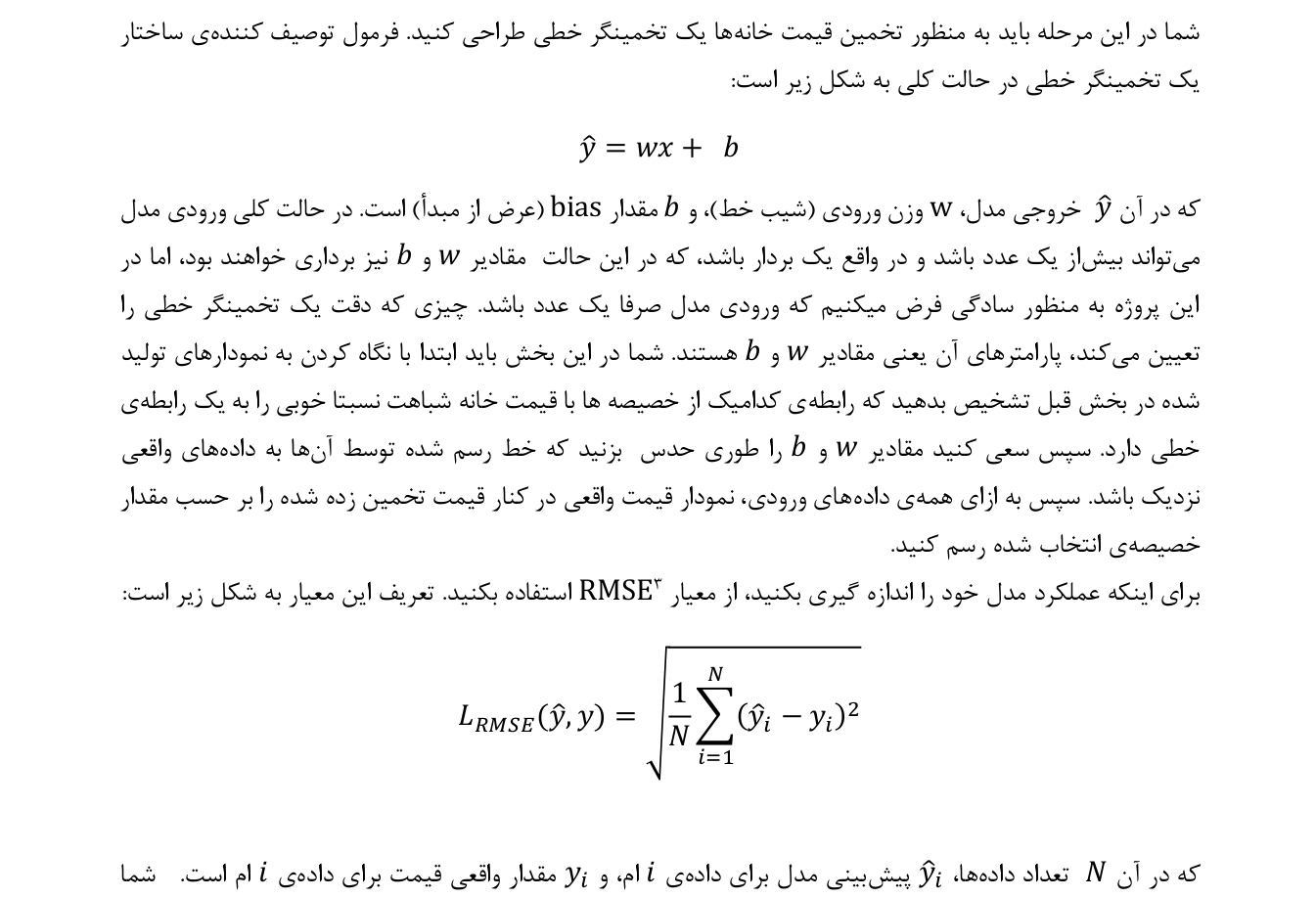




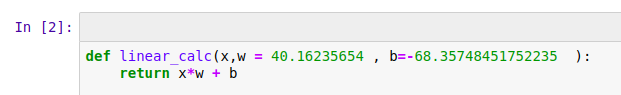


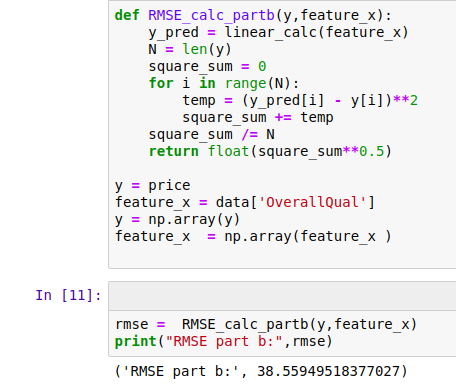


# بخش دوم تقریب گر خطی

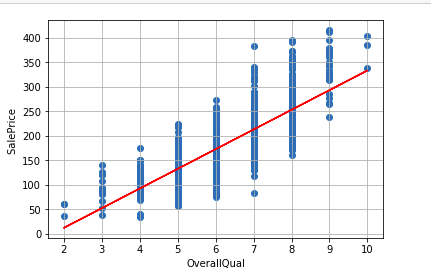


مقادیر b , w به صورت زیر حدس زده شده است و برای تعیین عملکرد مدل rmse با توجه به فرمول بالا محاسبه شد. هم چنین خروجی تقریبا ۳۸ بدست آمد که کمتر از ۹۰ می باشد.



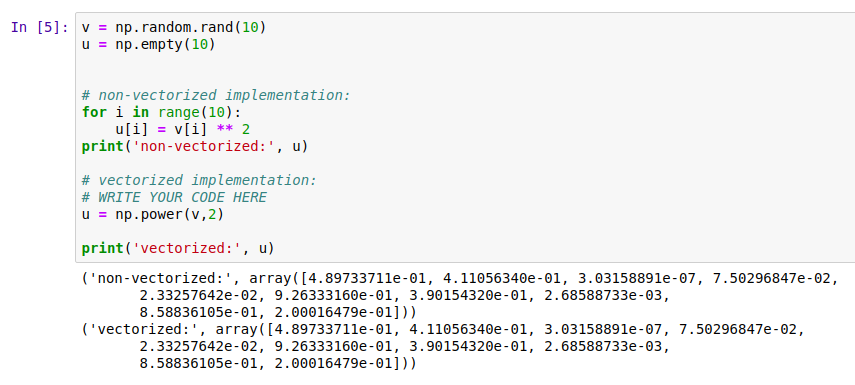


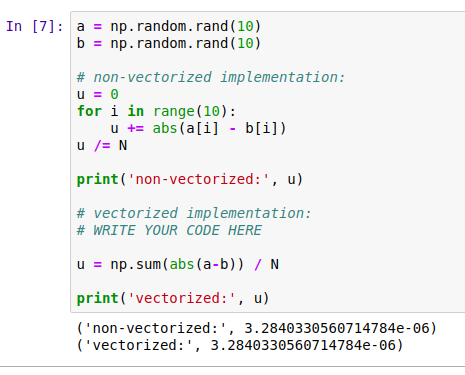
حال نمودار قیمت واقعی را در کنار قیمت تخمین زده شده به صورت زیر خواهیم داشت.

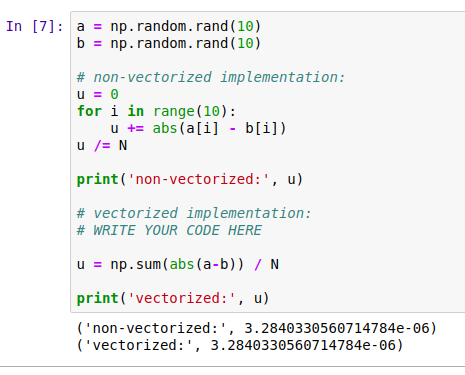


# بخش سوم آشنایی با کتابخانه‌ها

مهم ترین ویژگی این کتابخانه یک کلاس قدرتند نماینگر آرایه های N بعدی و مجموعه بسیار بزرگی از توابع پیچیده از جمله توابع broadcasting است. همچنین در این بخش با محیط jupyter نیزآشنا شدیم و در انتها یک تمرین مطق زیر از Numpy حل گردید.

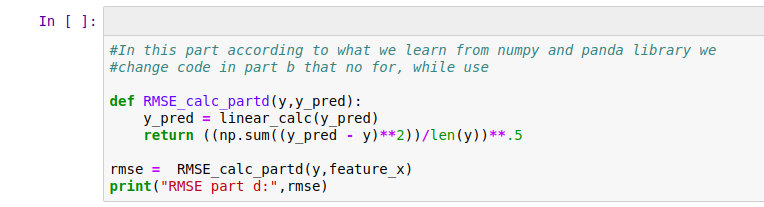






# بخش چهارم اصلاح کد قسمت دوم بدون حلقه

با استفاده از Numpy کد قسمت دوم بدون هیچ گونه for بازنویسی شد.



# بخش پنجم آشنایی با

در این مدل پارامتری وجود ندارد که مقدار بهینه ی آن را یاد گرفته و برای خانه های جدید با استفاده از آن ها تخمین را محاسبه کنیم بلکه تخمین قیمت هر خانه جدید به صورت مستقیم از روی ده هایی که داریم محاسبه می شود.

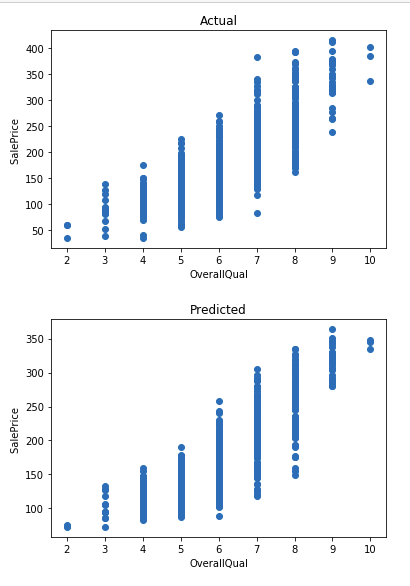
به این صورت که ابتدا نزدیک ترین K نقطه را به خانه مورد نظر می یابیم سپس میانگین قیمت این خانه ها را به عنوان قیمت خانه جدید گزارش می کنیم. ابتدا مقدار همه خصیصه ها را استاندارد سازی می کنیم تا خصیصه هایی که مقادیر بزرگتری دارند اثر سایر خصیصه ها را از بین نبرند.



برای تست تابع knn ابتدا آمدم تمام داده ها را به عنوان ورودی به تابع دادم. خروجی این تابع قیمت پیش بینی شده به ازای هر سطر می باشد.سپس مقدار rmse را به ازای y قیمت واقعی و y\_pred قیمت پیش بینی شده به دست آوردم که این مقدار حدود ۲۵ شد.



همچنین نمودار جهت مقایسه قیمت واقعی و پیش‌بینی شده بر حسب مشخصه overalqual به صورت ز خواهد بود.



**پایان**