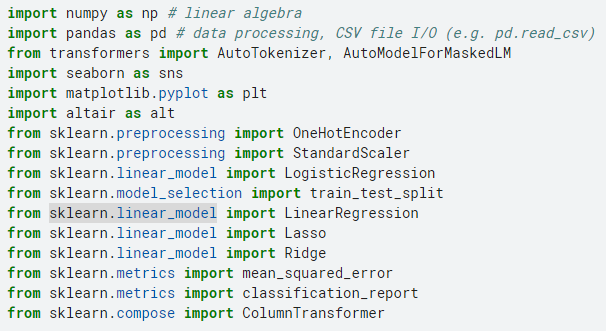
داده کاوی

تمرین شماره دو

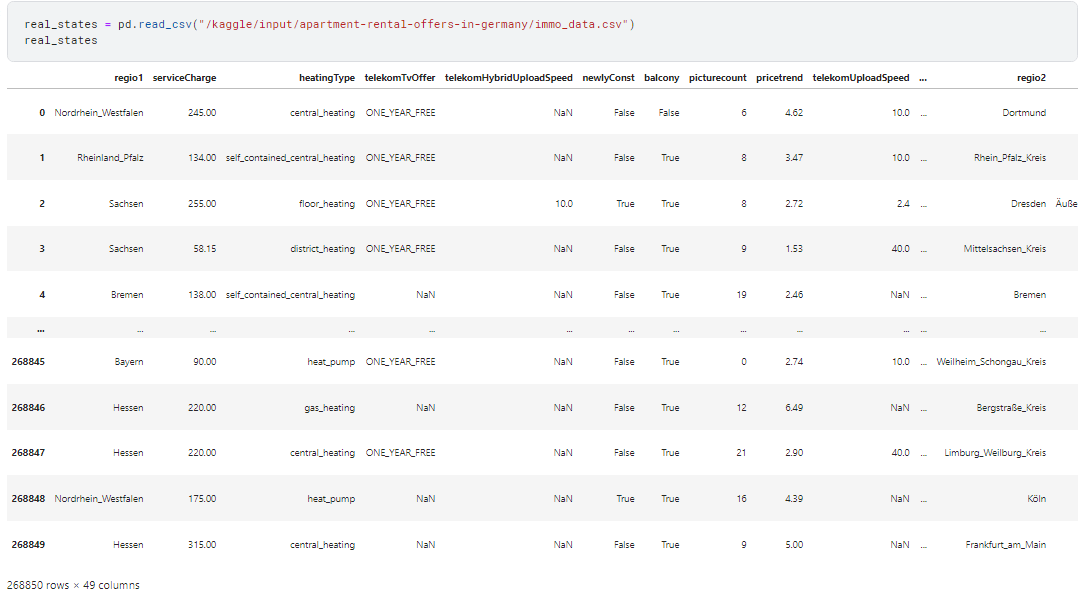
بخش دوم

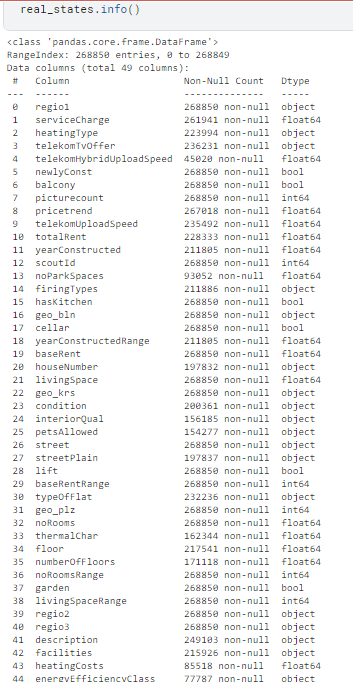
ابراهیم شکیباپور

ایمپورت پکیج های مورد نیاز



خواندن دیتاست



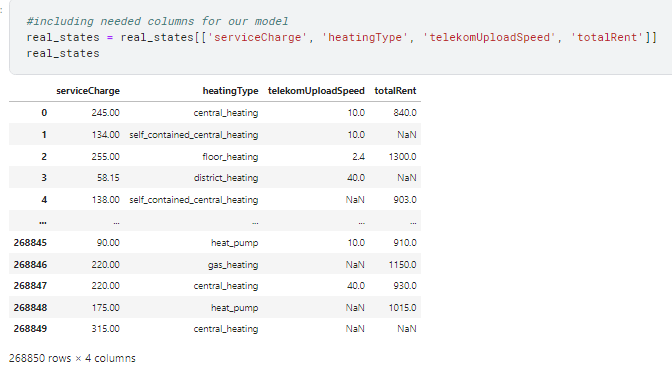


**فقدان داده زیادی در ستون های دیتافریم داریم**

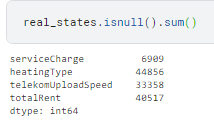
۱ – مدل رگرسیون تنها با سه تا از ویژگی ها:

serviceCharge, heatingType, telekomUploadSpeed

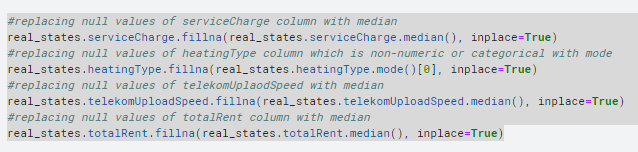
مدل رگرسیون با تابع MSE و بدون استفاده از پکیج:



ستون های که برای پیاده سازی مدل مدنظر داریم.

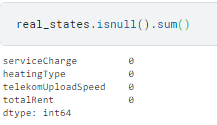


وضعیت این ستون ها از نظر داده های پوچ و فقدان داده

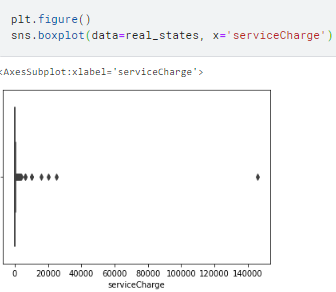


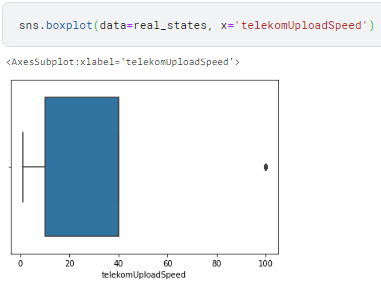
جایگذاری و پرکردن آنها با داده های مناسب

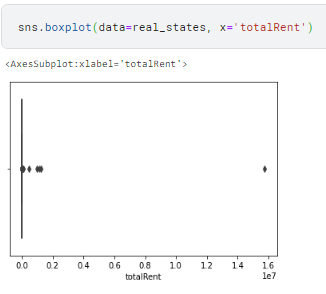
ستون های عددی را با مدیان و ستون غیرعددی و categorical را نیز با مد پر کردیم.



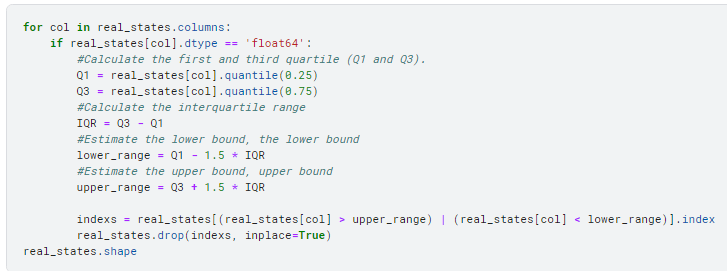
بررسی داده های پرت یا outliers



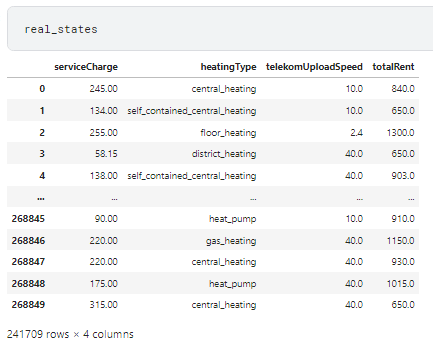


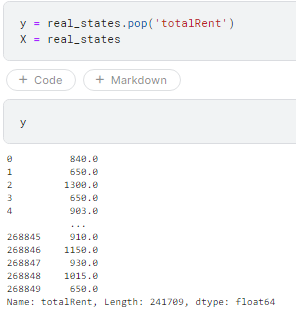


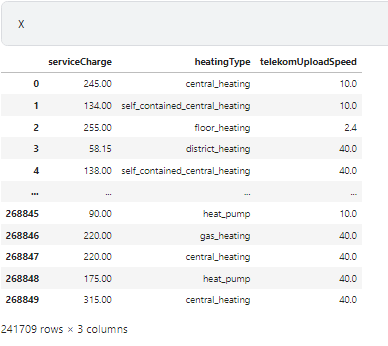
حذف داده های پرت



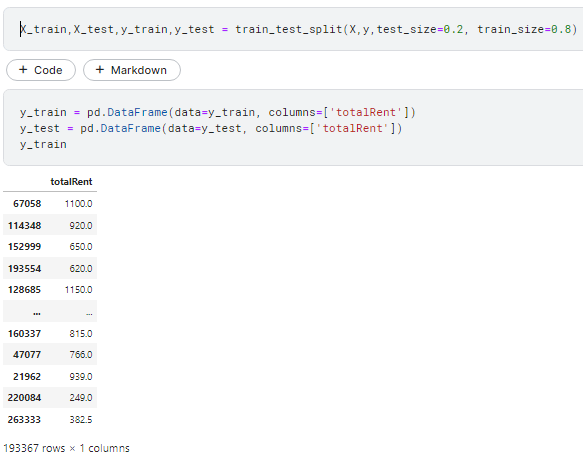
تقسیم دیتافریم به داده های آموزش و تست



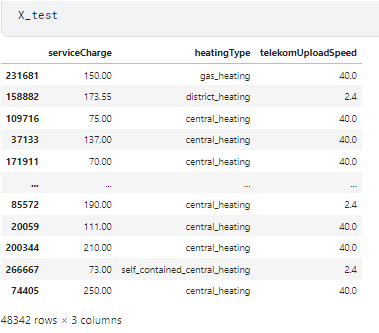


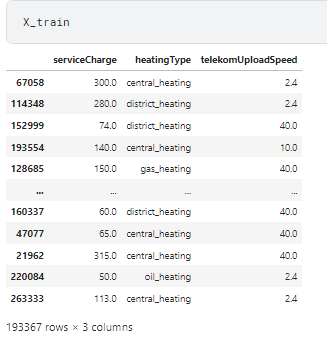


نسبت داده های آموزش به تست هشتاد به بیست



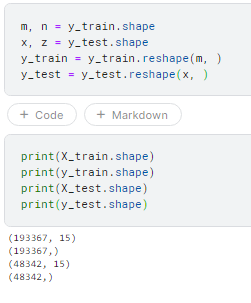






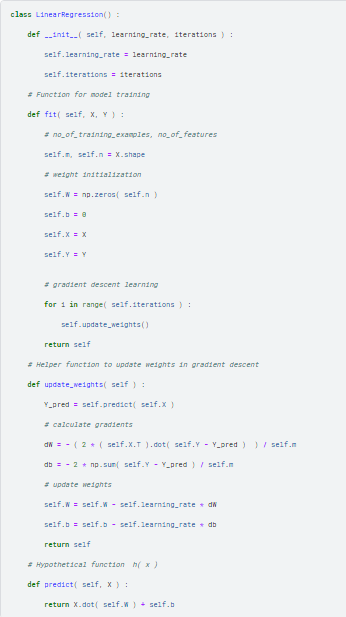
پیش پردازش



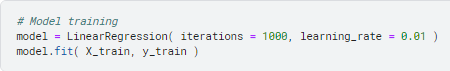


هم اکنون داده ها ی آموزش و تست برای اعمال مدل آماده می باشند.

پیاده سازی مدل رگرسیون با تابع خطای MSE و بدون استفاده از پکیج

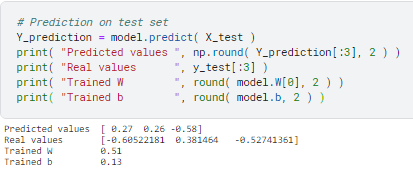


آموزش مدل

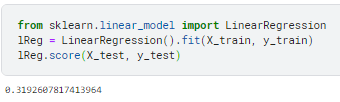


تخمین روی داده تست

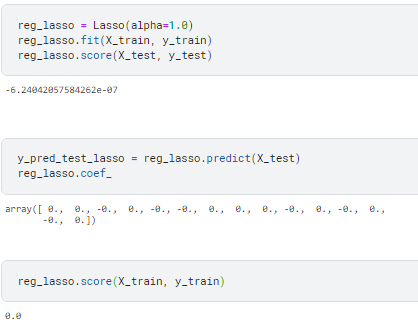




مقایسه مدل پیاده سازی شده با مدل با استفاده از پکیج sklearn



Lasso



Ridge

