فاز2 :

در این قسمت ابتدا دیتای مورد نظر را از دیتاهای پوچ و صفر خالی میکنیم و سپس به بررسی داده های پرت داده می پردازیم و آن هارا دسته بندی میکنیم.

بعد از انجام این عملیات داده ها بر اساس قیمت و ریجن 2 مورد بررسی قرار می دهیم تا یک شمای کلی از فراوانی داده ها بر روی این دو ویژگی مشاهده کنیم.

برای بررسی شیوه های گرمایشی و قیمت هر خانه ای که این ویژگی را دارد ، نمودار های خاص خودش را رسم میکنیم.

در ادامه هم به بررسی فراوانی قیمت پایه هر خانه و مقدار فضایی که دارد بررسی صورت میگیرد.

و شروع به فیلتر کردن کد می کنیم و نمودار را به طور مکرر رسم می کنیم تا بهترین نتیجه را ببینیم و سعی می کنیم نقاط پرت را که برای df معنی ندارد حذف کنیم.

از مجموعه داده، متغیرهایی که به "totalRent" بی ربط هستند عبارتند از 'scellar'، 'floor' و 'garden' بنابراین من آن را حذف می کنم.