

بخش اول قسمت دوم. گزارشی شامل موارد زیر :

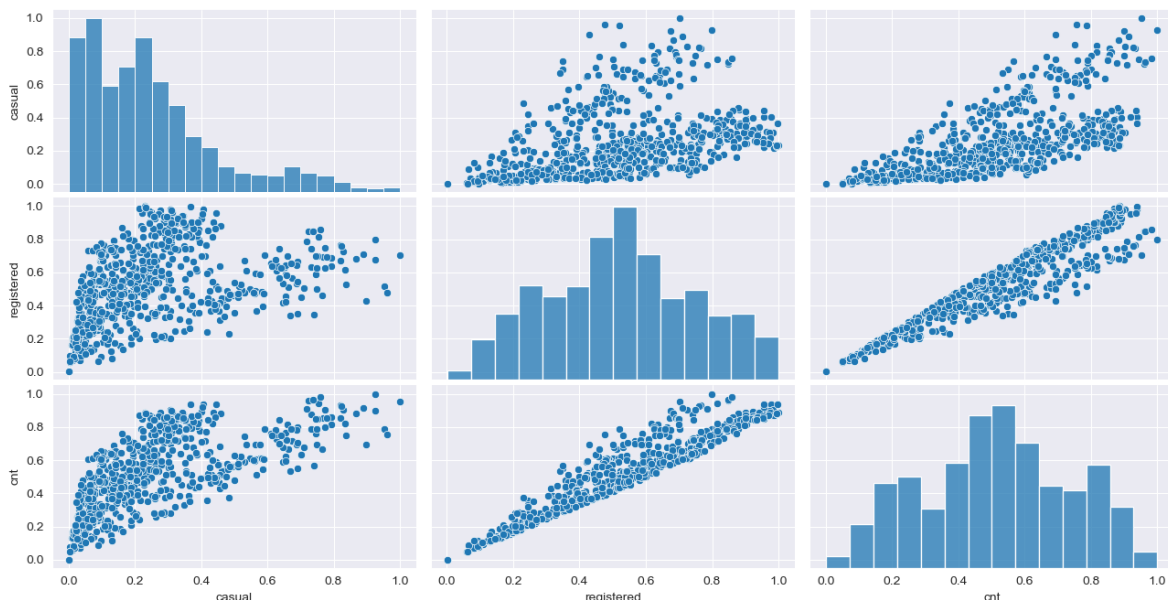
(نمودارها با استفاده از پکیج matplotlib , seaborn رسم شوند)

## a. رسم نمودار PairPlot برای داده‌ها و تحلیل آن

در ابتدا باید این نکته ذکر شود به دلیل اینکه تعداد زیادی ویژگی داریم بنابراین از رسم تمامی نمودارها صرف نظر می شود و تنها چند نمودار آخر بررسی می شود، البته دلیل دیگری برای عدم نمایش بسیاری از نمودارها وجود دارد که با بررسی مقادیر Correlation که در جدول زیر قرار داده شده است (برای دسترسی به تمامی مقادیر می توانید کد مربوط را اجرا کنید) می توان پی برد. در آخر نمودار PairPlot با توجه به دو ویژگی (Registered ، Casual) و یک خروجی (Cnt) ترسیم شده است که ویژگی Registered تعداد افراد عضو این سیستم و ویژگی Casual تعداد افراد معمولی و ثبت نام نشده و خروجی Cnt مجموع این دو را نشان می دهد. این نمودارها همبستگی زیاد Registered و Cnt را نشان می دهند (با افزایش Registered (افراد عضو سیستم (شمار دوچرخه های افراد عضو سیستم) مقادیر Cnt (مجموع شمار دوچرخه ها) افزایش پیدا می کند) و از آن طرف همبستگی نه چندان زیاد Casual و Cnt را داریم. البته این تحلیل از مقادیر Correlation مربوطه در جدول هم دریافت می شود. می توان این طور دریافت کرد که حتی با داشتن دو ویژگی می توانیم کار را تمام کنیم.

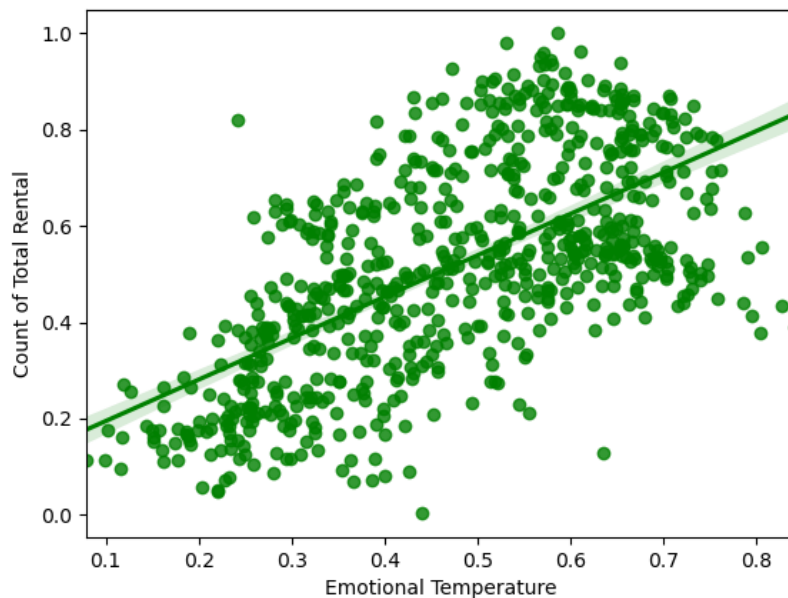
جدول مقادیر Correlation متغیرها

Between weathersit and temp	-0.12060224
Between weathersit and hum	0.5910446
Between weathersit and registered	-0.26038771
⋮	⋮
Between registered and cnt	0.94551692
Between casual and cnt	0.67280443

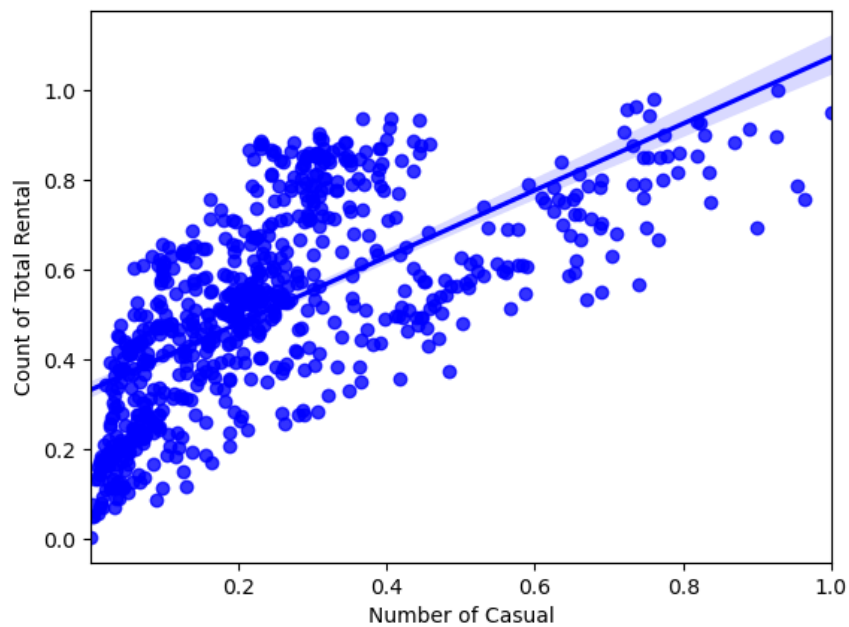


**b. به تعداد ویژگی‌ها نمودار RegPlot تحویل دهید و هر کدام را تحلیل کنید.**

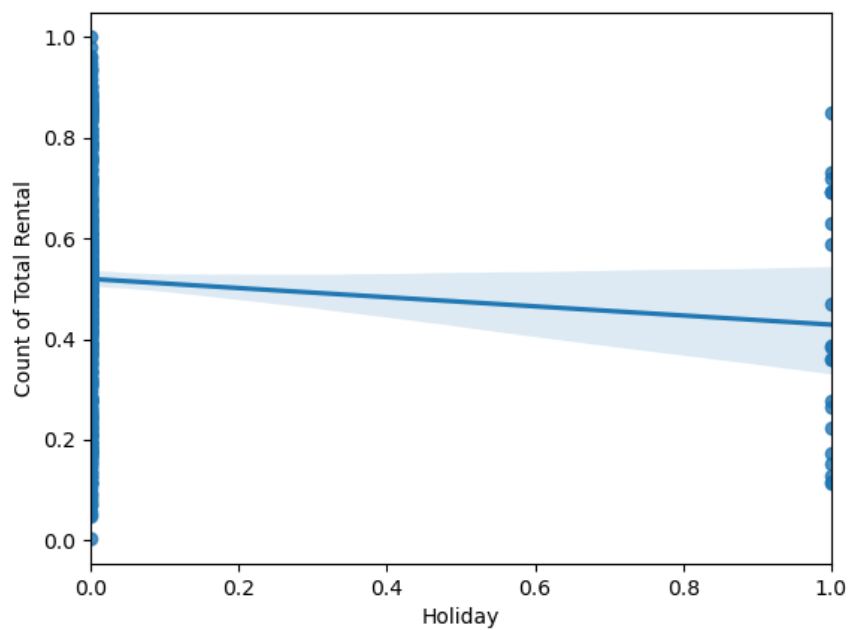
باتوجه به اینکه هدف این بخش شناسایی ویژگی‌هایی است تنها قابلیت ارتباط خطی با هدف (خروجی) را دارند بنابراین نمودار تمامی 13 ویژگی ترسیم شده است.



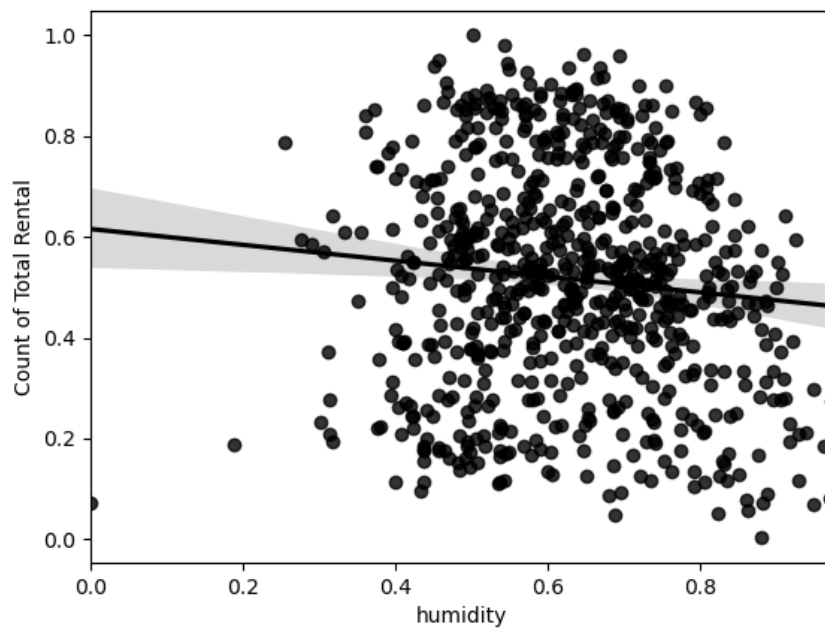
با توجه نمودار بالا و میزان همبستگی 63% به نظر می‌رسد این ویژگی گزینه خوبی برای پیش بینی است ولی ویژگی‌های دیگر را هم بررسی خواهیم کرد.



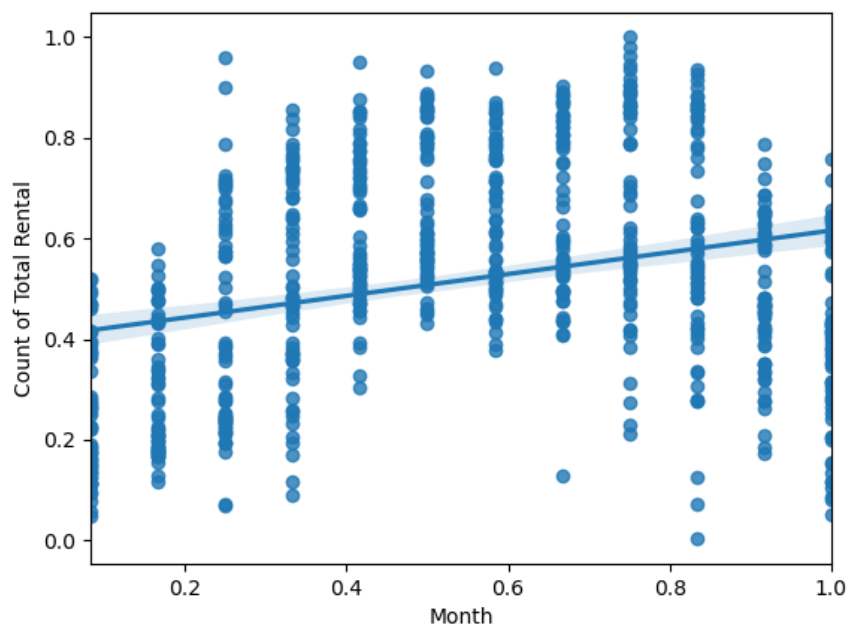
با توجه نمودار بالا و میزان همبستگی 67% به نظر می‌رسد این ویژگی هم گزینه خوبی برای پیش بینی است.



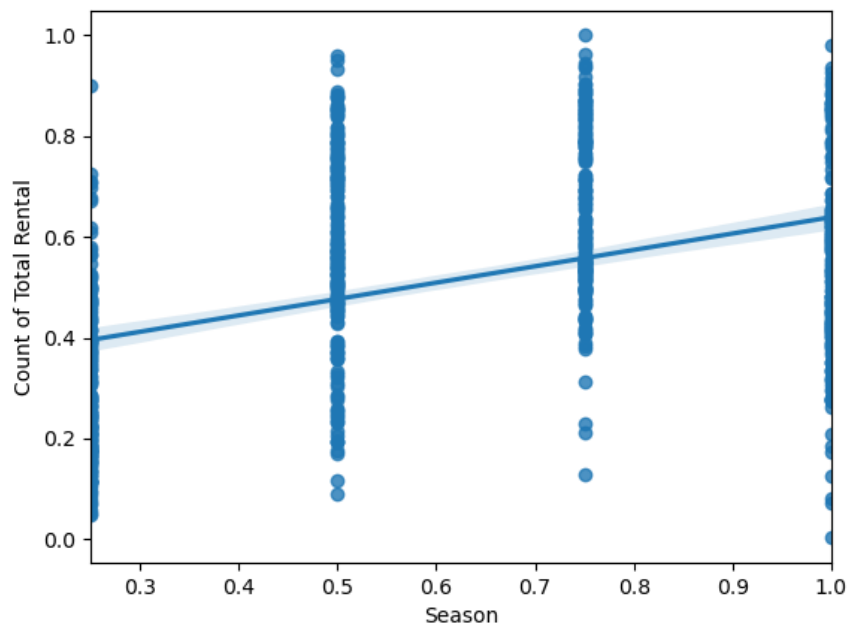
با توجه به نمودار بالا واضح است به دلیل نداشتن ارتباط خطی نمی‌توان خطی را برازش کرد.



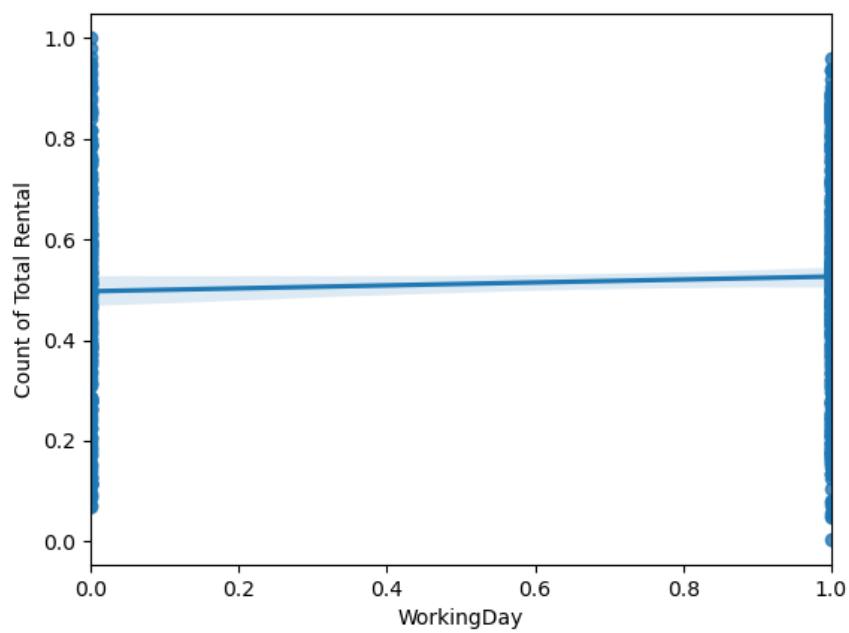
با توجه به نمودار بالا و میزان همبستگی منفی 1% این ویژگی نمی‌تواند گزینه خوبی برای پیش بینی باشد.



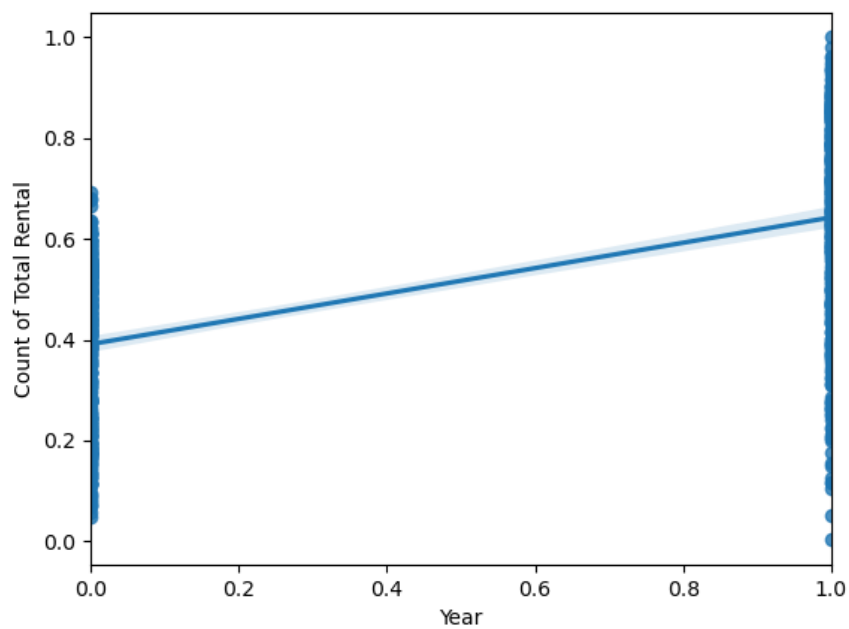
با توجه به نمودار بالا واضح است به دلیل نداشتن ارتباط خطی نمی‌توان خطی را برازش کرد.



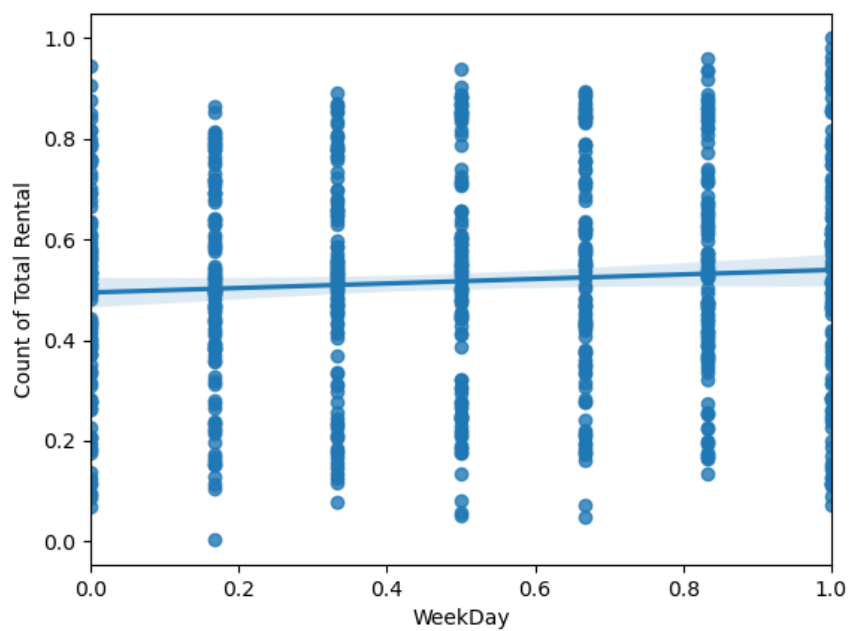
با توجه به نمودار بالا واضح است به دلیل نداشتن ارتباط خطی نمی‌توان خطی را برازش کرد.



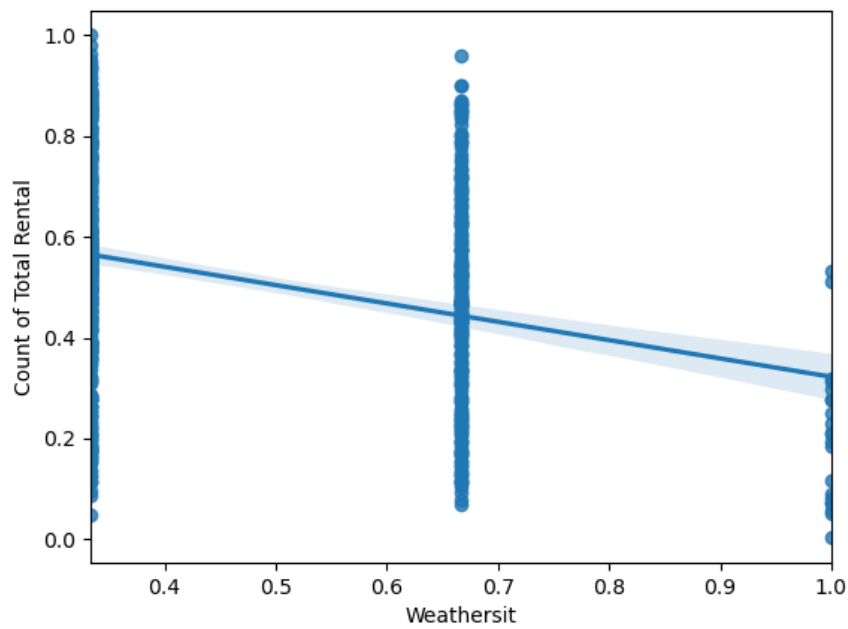
با توجه به نمودار بالا واضح است به دلیل نداشتن ارتباط خطی نمی‌توان خطی را برازش کرد.



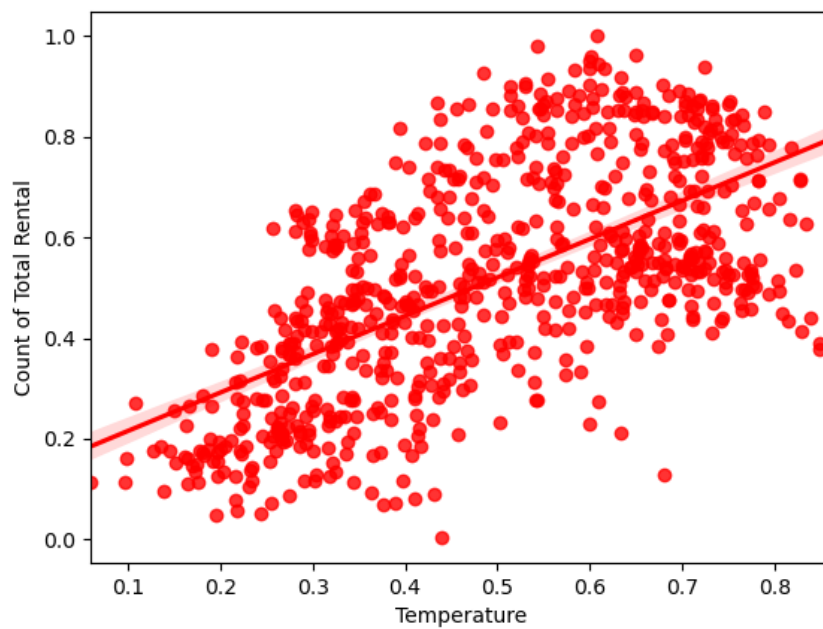
با توجه به نمودار بالا واضح است به دلیل نداشتن ارتباط خطی نمی‌توان خطی را برازش کرد.



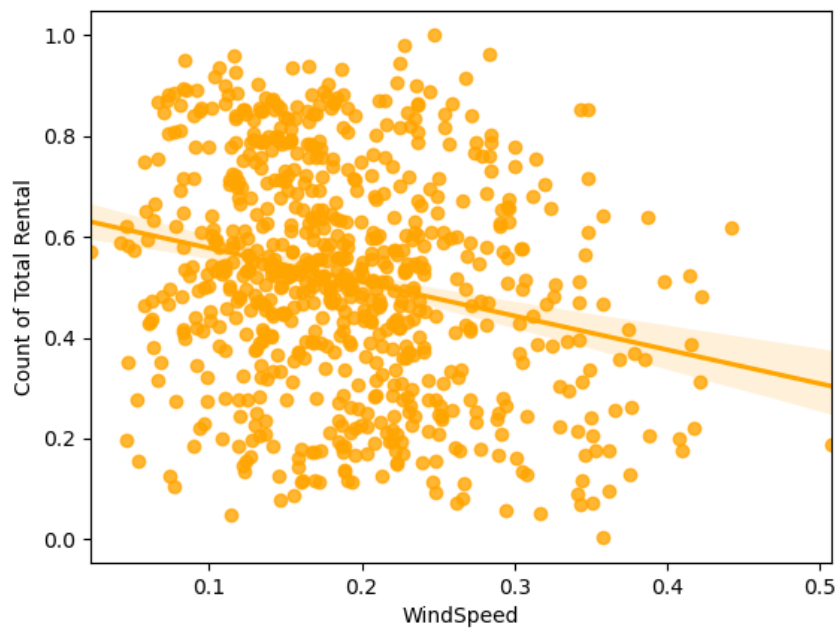
با توجه به نمودار بالا واضح است به دلیل نداشتن ارتباط خطی نمی‌توان خطی را برازش کرد.



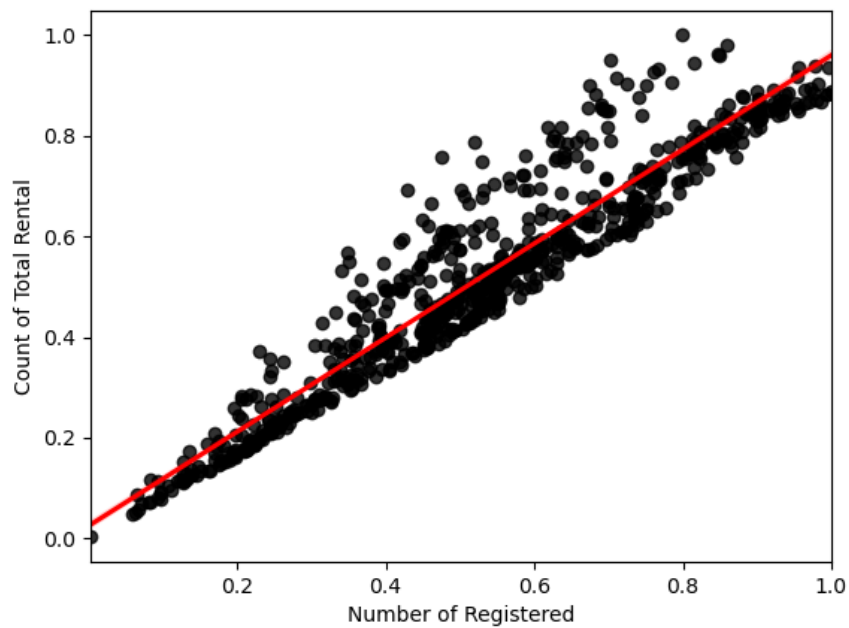
با توجه به نمودار بالا واضح است به دلیل نداشتن ارتباط خطی نمی‌توان خطی را برازش کرد.



با توجه نمودار بالا و میزان همبستگی 62% به نظر می‌رسد این ویژگی گزینه خوبی برای پیش‌بینی است ولی ویژگی‌های دیگر را هم بررسی خواهیم کرد.



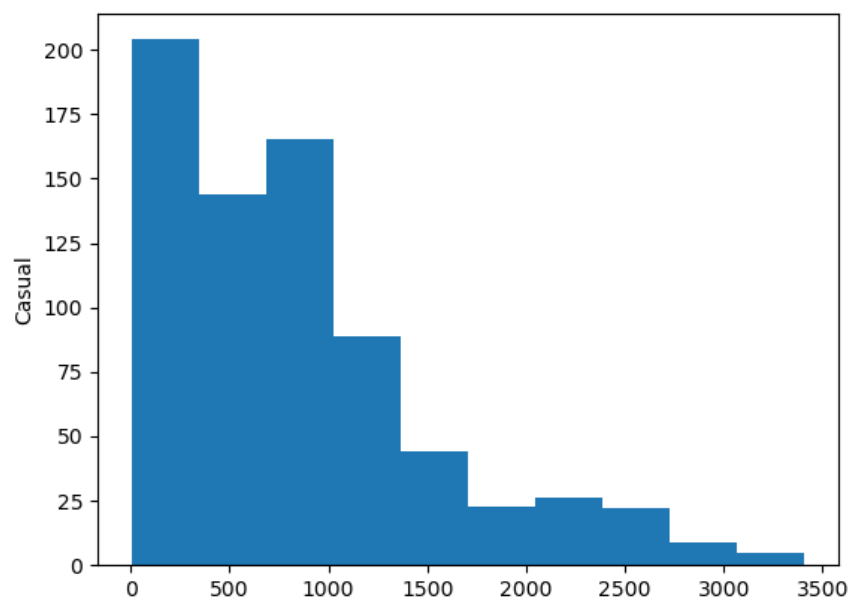
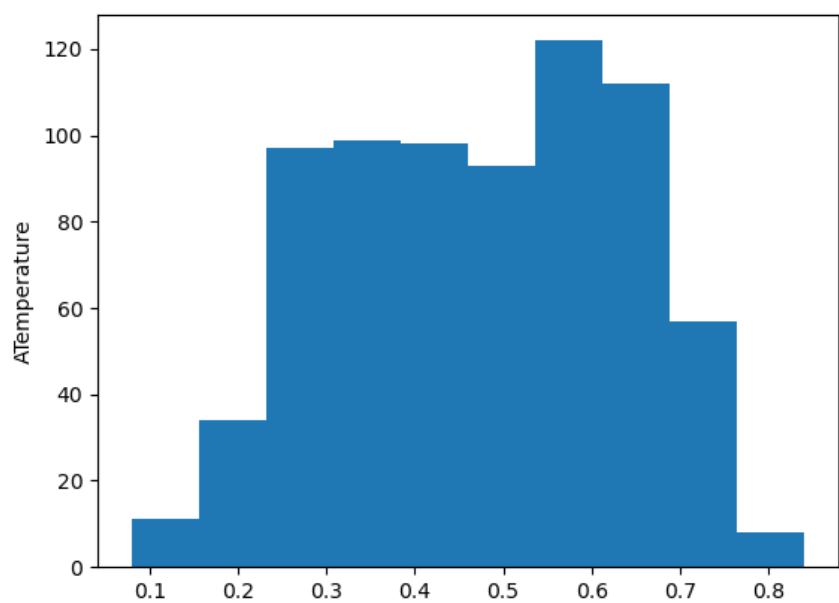
با توجه به نمودار بالا و میزان همبستگی منفی 23% این ویژگی نمی تواند گزینه خوبی برای پیش بینی باشد.

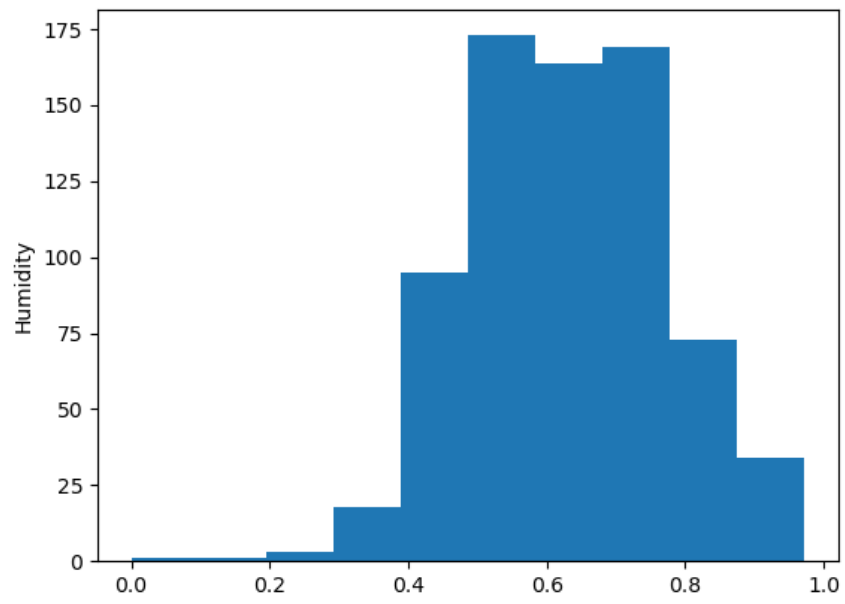
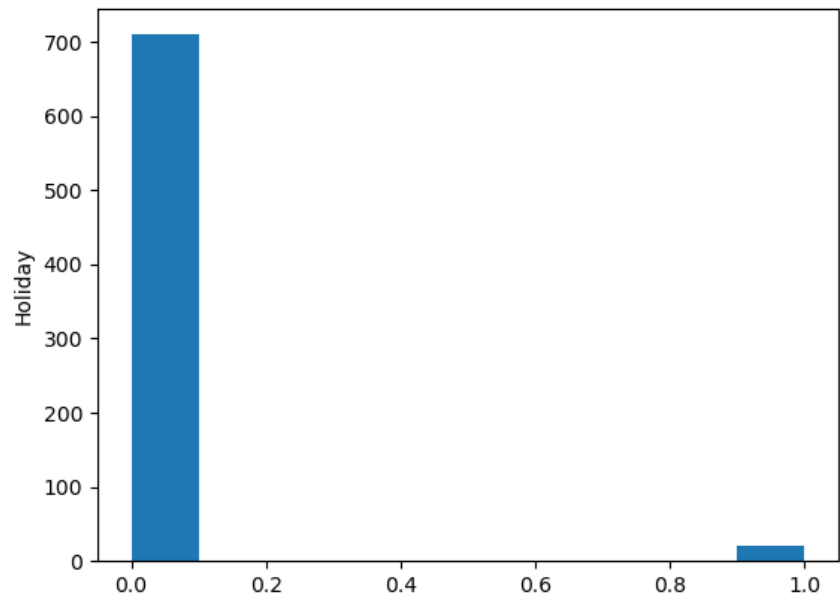


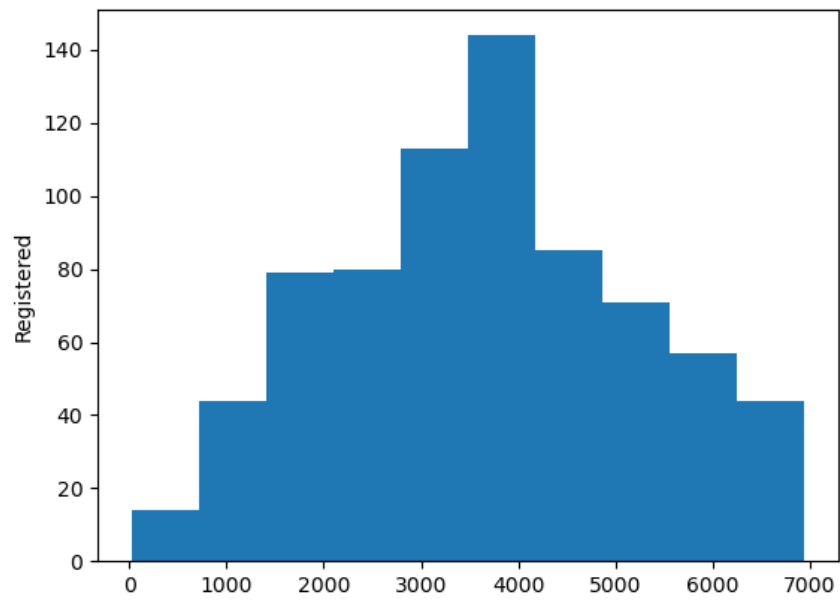
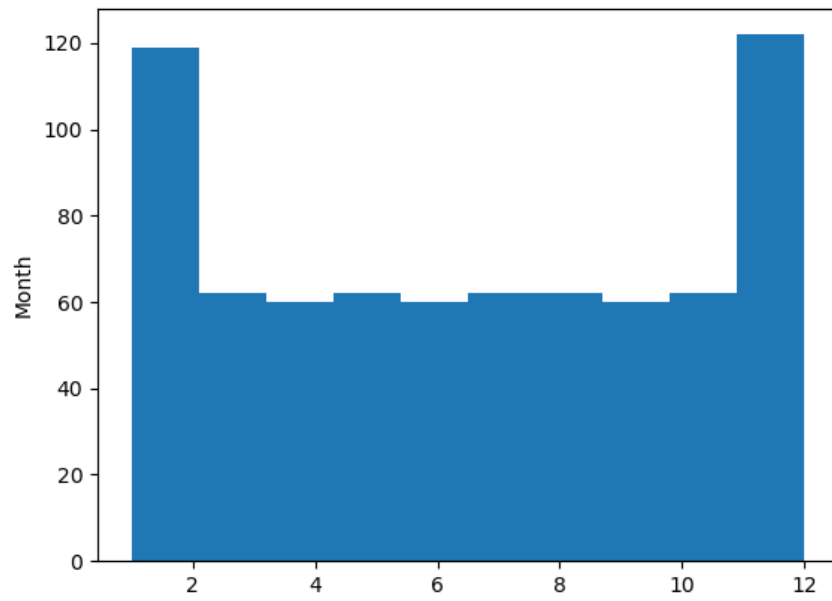
با توجه نمودار بالا و میزان همبستگی 94% به نظر می رسد این ویژگی گزینه خیلی خوبی برای پیش بینی است.

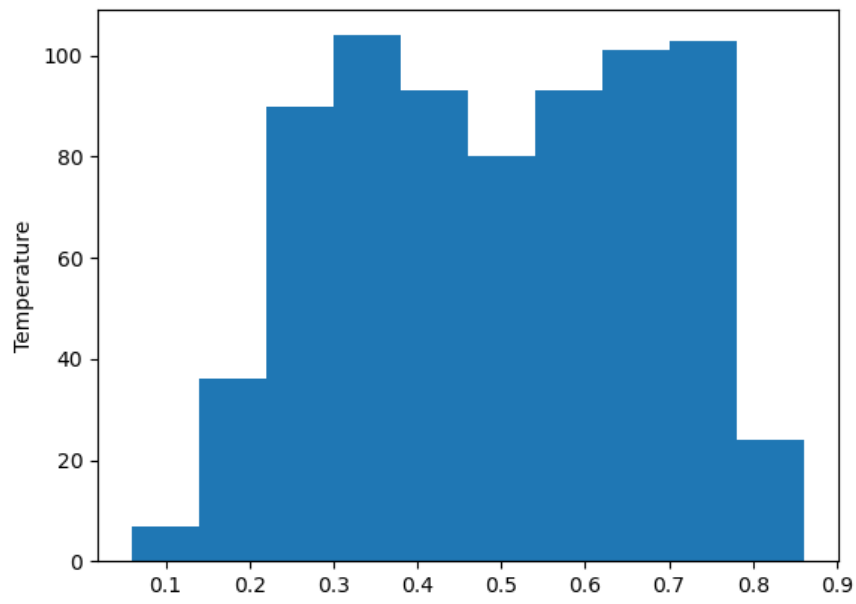
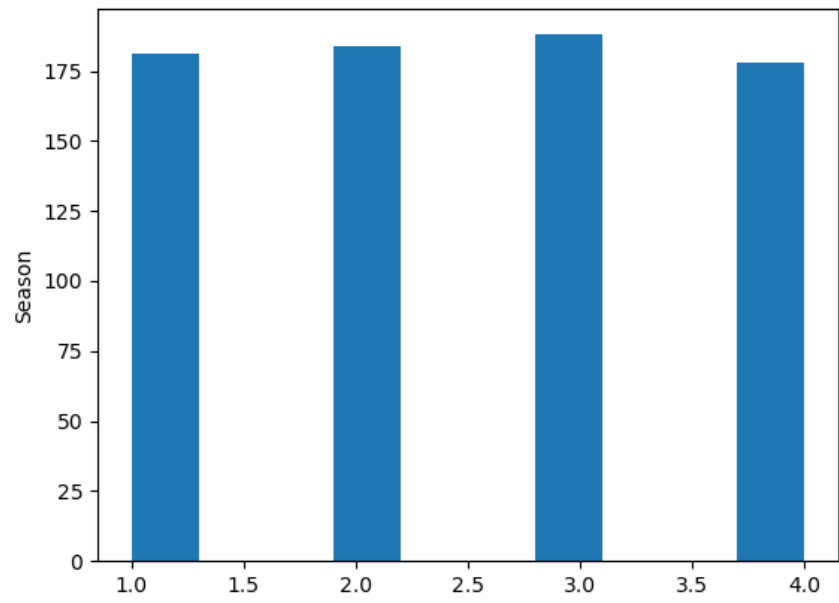


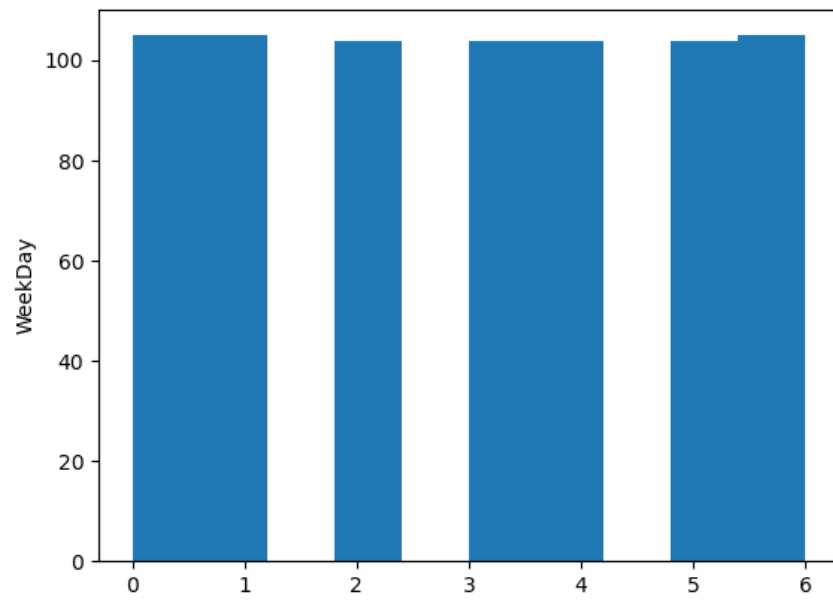
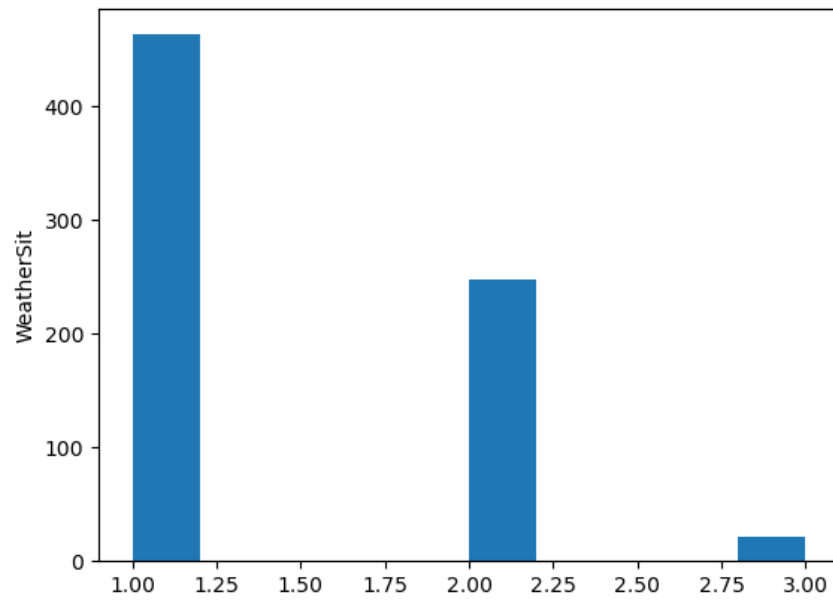
c. برای هر ویژگی نمودار هیستوگرام آن را رسم کنید.

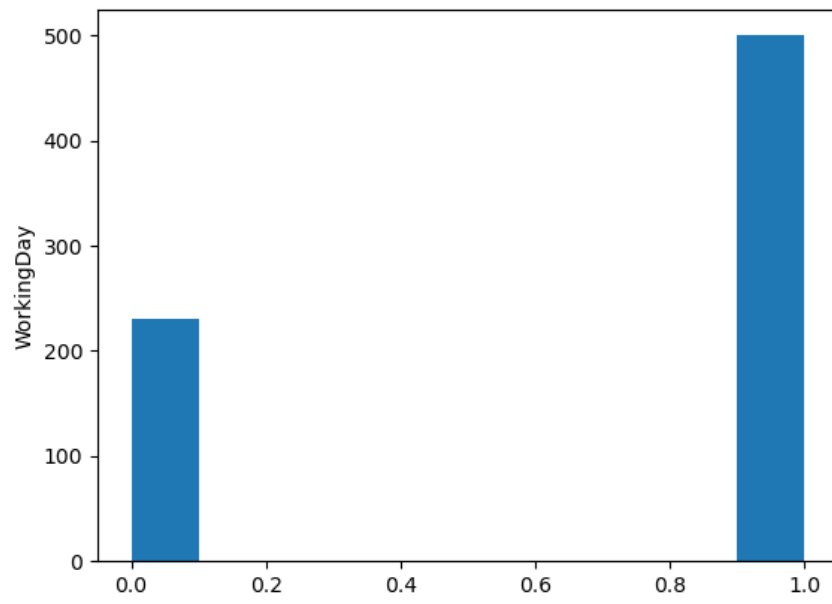
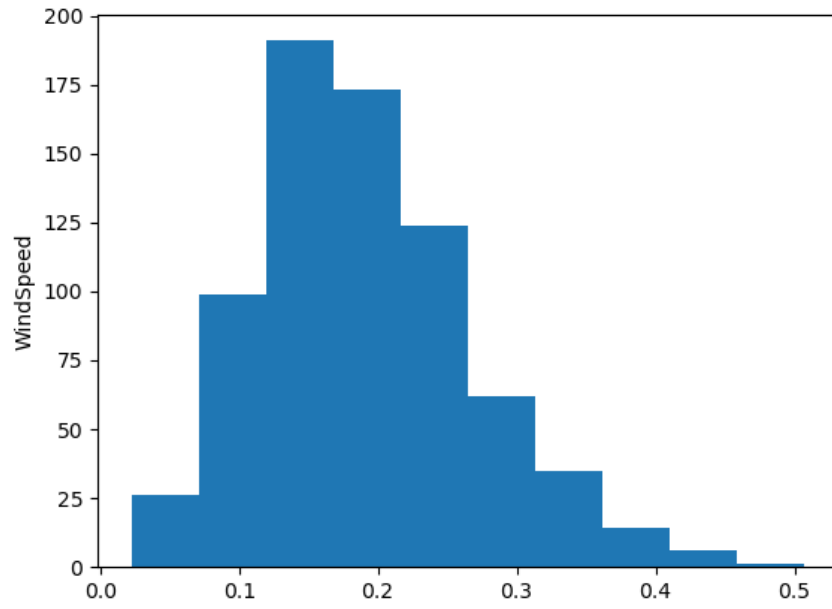


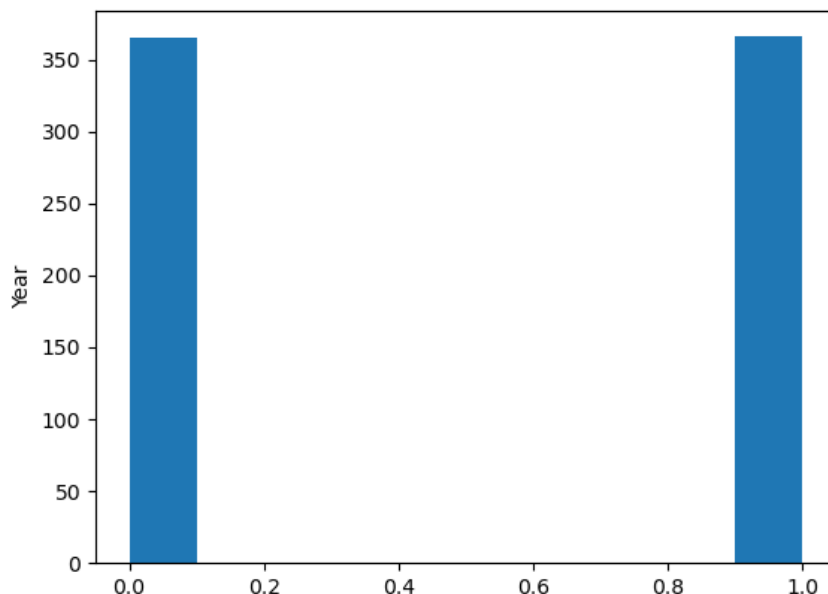












### d. از نمودار StripPlot و یا SwarmPlot برای تحلیل دوبه دوی ویژگی‌ها استفاده کنید.

با توجه به گزارش‌های بالا، به طریقی نمودارها دوبه دو تحلیل شده‌اند و نیازی به رسم این نمودارها نمی‌باشد. (هرچند برای رسم PairPlot هم با توجه به بالا بودن تعداد ویژگی‌ها از رسم صرف نظر شد و به شیوه‌ای دیگر تحلیل انجام شد ولی در نهایت چند نمودار مهم آن ترسیم شد).

### e. بخش نتیجه‌گیری

با توجه به فعالیت‌ها و گزارش‌هایی که قسمت‌ها بالا صورت گرفت به نظر می‌رسد از بین کل ویژگی‌ها فقط 4 ویژگی (Temperature، Emotional Temperature، Casual، Registered) به مقدار هدف مرتبط بوده است و درحقیقت این 4 ویژگی میزان وابستگی بیشتری داشتند و به نظر می‌رسد از بین این 4 ویژگی با داشتن Registered با می‌توان مقدار هدف را برای یک داده جدید به خوبی پیش‌بینی کرد. همچنین با توجه به این نتایج به نظر می‌رسد مدل خطی برای این کار کافی می‌باشد.