

دیتاست energy.csv مجموع مصرف برق در تعدادی از ایالت های آمریکا در هر ساعت بین سالهای ۲۰۰۴ تا ۲۰۱۸ را نشان می دهد.

۱- به کمک pandas.to_datetime ، نوع ستون Datetime را به تاریخ تغییر دهید.

۲- به کمک توابع pandas از ستون Datetime سال، ماه و ساعت هر داده را استخراج کنید و در ستونی مجزا ذخیره کنید.

۳- نمودار boxplot یکی از نمودارهای پرکاربرد برای نمایش بازه تغییرات یک متغیر می باشد. به وسیله این تابع که در کتابخانه seaborn وجود دارد، مصرف برق برحسب سال را نشان دهید. (برای این کار باید محور افقی را سال، و محور عمودی را مصرف برق قرار دهید.)

۴- با توجه به شکل نمودار، به نظر شما واریانس مصرف برق در روزهای سال ۲۰۰۴ بیشتر است یا سال ۲۰۰۵؟ درستی ادعا خود را نشان دهید.

۵- نمودار boxplot ساعت-انرژی را رسم کنید. و آن را تحلیل کنید

۶- نمودار boxplot ماه-انرژی را رسم کنید و آن را تحلیل کنید.

۷- ضریب همبستگی مصرف برق و ساعت را در بین ساعت ها چهار صبح تا یک بعد از ظهر را نشان دهید. آیا با نمودار بالا مطابقت دارد؟

۸- ضریب همبستگی مصرف برق و ماه را در بین ماه های دوم تا چهارم و همچنین بین ماه های دهم تا دوازدهم را هر کدام جداگانه به دست آورید. اعداد به دست آمده را با توجه به فصول سال در تقویم میلادی تعبیر کنید.

حال میخواهیم ببینیم که آیا می توان از ضریب همبستگی رابطه علیتی نتیجه گرفت یا خیر.

۹- برای این کار محتوای فایل TV LE Physician را بخوانید. سپس ضریب همبستگی بین متغیرهای عددی را بررسی کنید. به نظر شما آیا رابطه علیتی بین این متغیر ها وجود دارد؟ اگر نه به نظر شما چه چیزی باعث این ضریب همبستگی شده؟