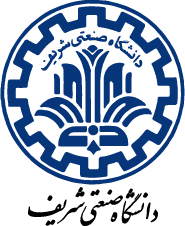
**به نام خدا**



**تمرین شماره­ی3**

**زبان­های شبیه سازی**

**احمد امامی**

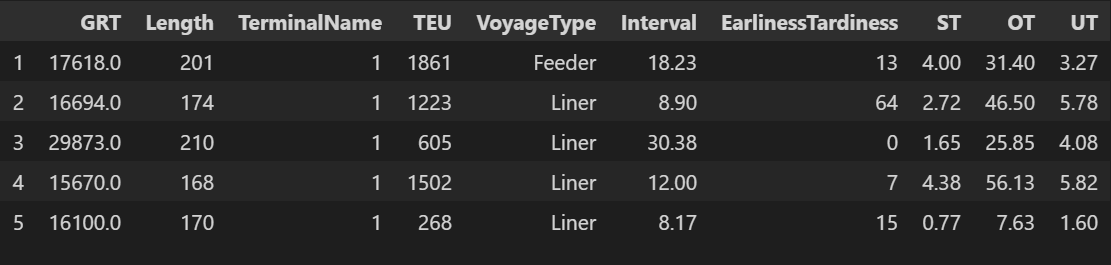
**99207521**

\*\* تمام بررسی­ها و نتایج ارائه شده در این تمرین به کمک پایتون انجام شده است و کد آن به پیوست ضمیمه شده است.

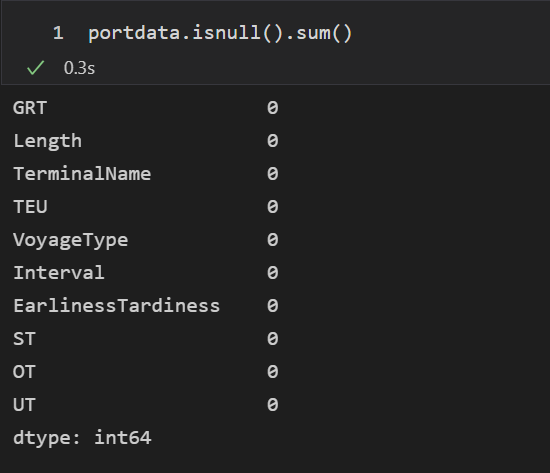
# پیش­پردازش داده­ها Data Cleaning

پیش از پاسخ به پرسش ­های تمرین بهتر است دیتاست مربوطه را بررسی کرده و در صورت وجود داده­ی null و یا داده­های پرت آن­ها را حذف کنیم. زیرا در فرایند فیت کردن توزیع به هرکدام از مشخصه­ها، در صورت وجود داده­های پرت با مشکل مواجه می­شویم.

ابتدا ستون No از دیتاست را حذف می­کنیم زیرا نیازی به آن نخواهیم داشت. 5 سطر ابتدایی دیتاست جدید را مشاهده می­کنید.

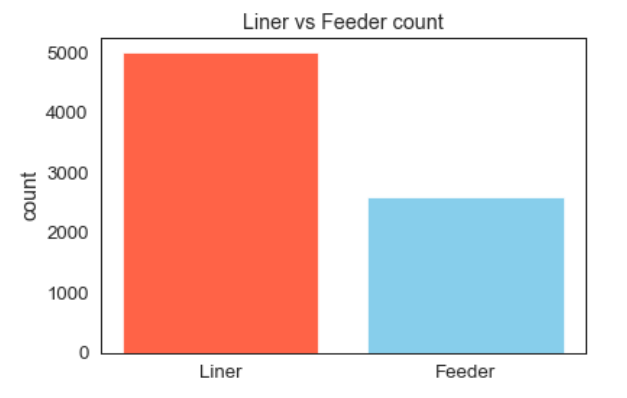


داده­های null را بررسی می­کنیم:



همان­طور که می­بینیم در هیچ یک از ستون­های دیتاست داده­ی null مشاهده نمی­شود.

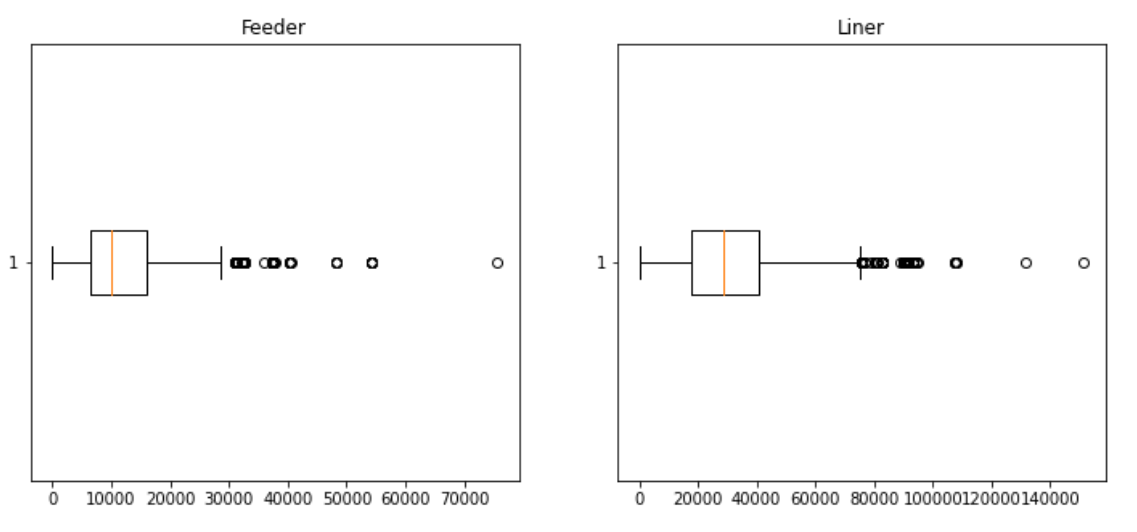
در مرحله­ی بعدی بهتر است که داده­هایمان را به دو دسته تقسیم کنیم. یک قسمت مربوط به کشتی­های لاینر و دیگری مربوط به کشتی­های فیدر. از آنجایی که مشخصات مرتبط با هر کشتی مختص خودش است، بهتر است آن­ها را به صورت جداگانه و در دو دیتاست مختلف مورد بررسی قرار دهیم. همان طور که در تصویر زیر می­بینیم حدودا 5 هزار کشتی لاینر و 2500 عدد نیز فیدر هستند که به صورت مجزا مورد بررسی قرار می­دهیم.



**حذف داده­های پرت**

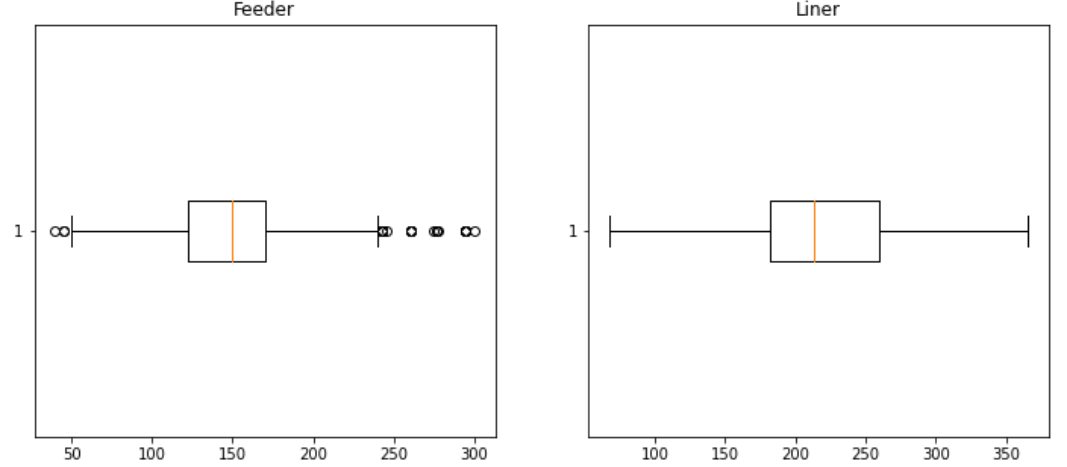
برای حذف داده­های پرت از نمودار جعبه­ای (boxplot) کمک می­گیریم. برای هر مشخصه این نمودار را رسم می­کنیم و داده­های بسیار بزرگ و غیر معمول را حذف می­نماییم.

**GRT**



در ادامه به کمک پایتون oulier های موجود را حذف می­کنیم و بررسی­های موجود را بر روی این مجموعه داده­ی جدید انجام می­دهیم.

**Length**



برای طول کشتی­ها وضعیت بهتری را شاهد هستیم و برای کشتی­های Liner داده­ی پرتی مشاهده نمی­شود. برای کشتی­های Feeder تعدادی مشاهده outlier داریم که تصمیم گرفتیم که آن­ها را حذف بنماییم.

# سوال 1

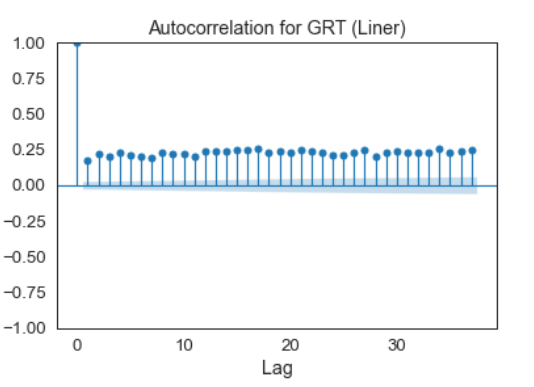
**فرض i.i.d بودن را برای مشخصه­های مختلف در دیتاست بررسی کنید.**

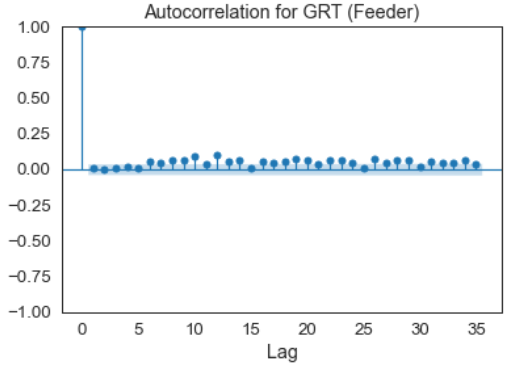
--------------------------------------------------------------------------------------------------------

برای بررسی iid بودن باید چک شود که آیا داده­های مربوطه در هر ستون به یکدیگر مرتبط هستند یا خیر. به بیان دیگر آیا خوهمبستگی autocorrelation میان داده­ها دیده می­شود یا نه. اگر فرض iid بودن برقرار باشد آنگاه داده­های ما مانند یک نویز رفتار می­کنند و بین هیچ لگی همبستگی­ای وجود نخواهد داشت. یکی از بهترین راه­ها برای بررسی این موضوع استفاده از نمودار خودهمبستگی (ACF) می­باشد. در صورت مشاهده­ی یک لگ قابل توجه در این نمودار میتوان نتیجه گرفت که داده­ها دارای خودهمبستگی هستند. نتایج زیر را برای مشخصه­های مختلف دیتاست مربوطه مشاهده می­کنید:

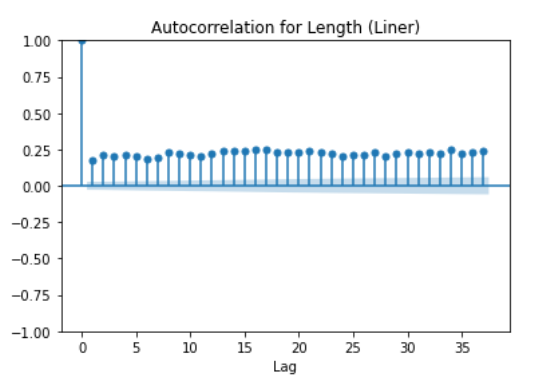
**توناژ ناخالص کشتی GRT**

همان­طور که در تصاویر زیر مشاهده می­شود مقدار همبستگی میان توناژ ناخالص در کشتی­های لاینر و فیدر دارای خودهمبستگی است و در لگ­های متعددی مقادیر قابل توجه مشاهده می­شود. در نتیجه فرض i.i.d بودن در ارتباط با مشخصه­ی توناژ ناخالص کشتی قابل قبول نمی­باشد.

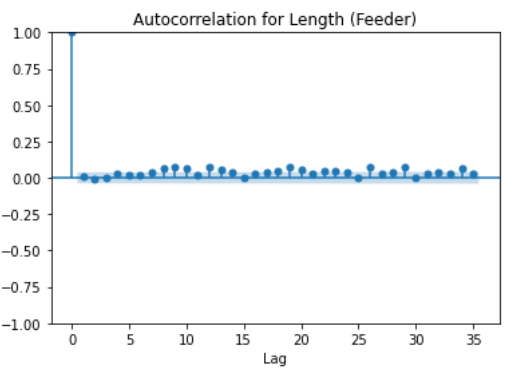




**طول کشتی­ها Length**



برای کشتی­های لاینر واضح است که طول کشتی دارای خودهمبستگی است و فرض i.i.d بودن برقرار نیست.



برای کشتی­های Feeder وضعیت بهتر است ولی کماکان نمیتوان iid بودن را پذیرفت.

