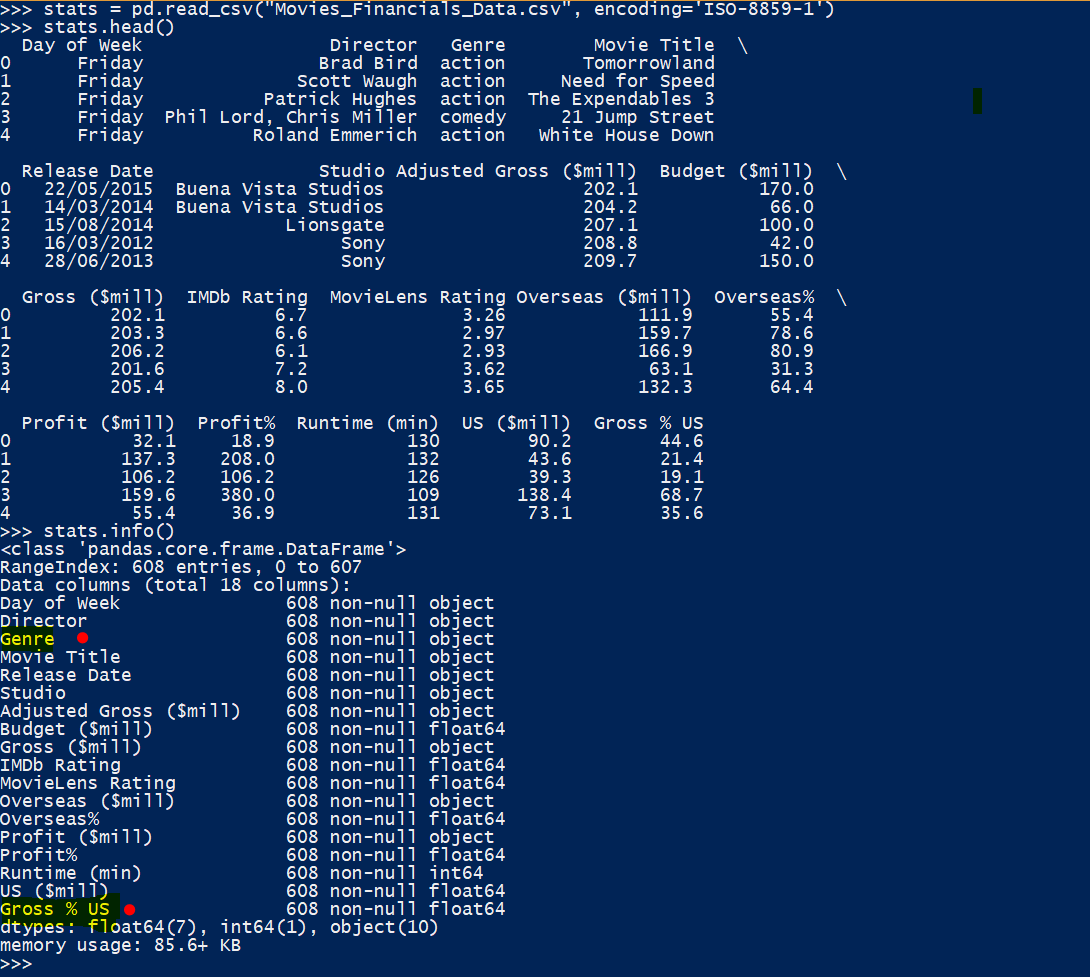
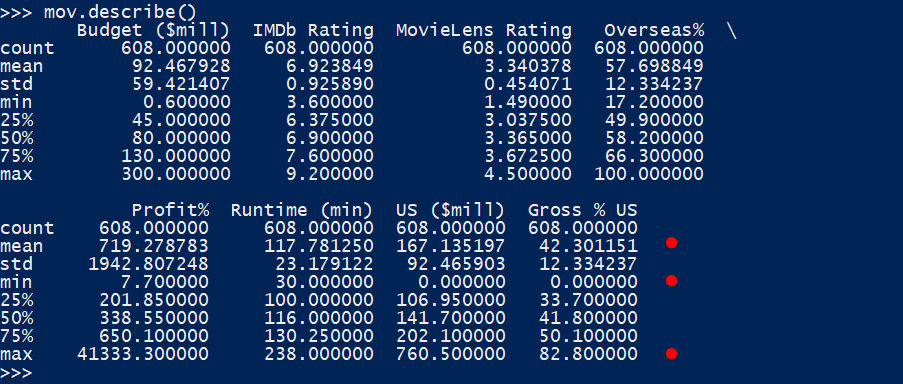
1. Getting a feel of the data-set

Import the ‘Movies\_Financials\_Data.csv’ file in pandas and read overview of the different fields and the types of data. It looks like this:



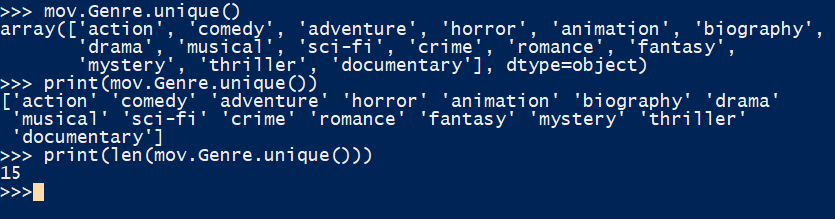
The fields most important to us to solve Challenge-1 have been highlighted. They are ‘Genre’ and ‘Gross % US’.

Now, let’s get a feel of the Gross US% in general for all movies inclusive of all genres:



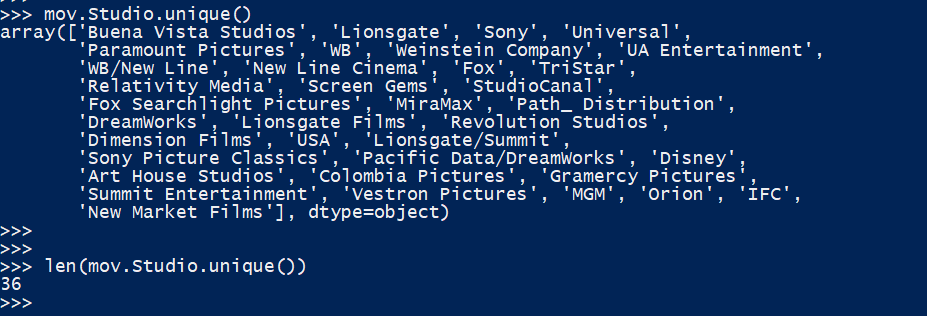
This figure shows us that there are movies that have 0.0% Gross Revenue of the total Gross in US and there is at-least one movie that has got a high of 82.8% on the same. The data is consistent as all the percentages values logically should be between 0% to 100%. Other quartile values can be found in the figure. However, they are not directly useful to us as we want the Gross % US visualized by Genre. So, let’s work towards that.

Let’s check how many genres are available and what are they:



So, 15 genres are available in total and printing the list of unique genres as mentioned above displays all the name.

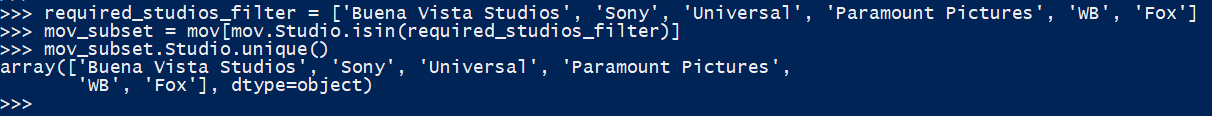
1. Building the visualization step-by-step
   1. Filtering out unnecessary Studios
      1. Check all Studios available:



There are 36 genres available, but our visualization requires only 6 genres. The required genres are:

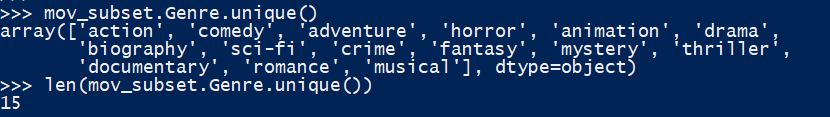


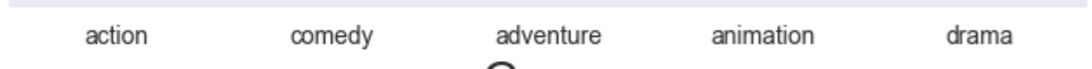
* + 1. Create a subset of the data-set with values relevant to the required Studios



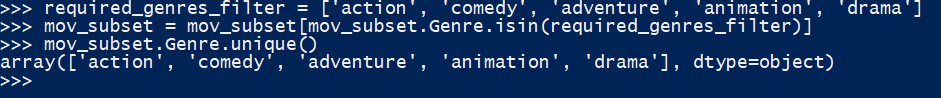
The new subset of the data-sets has records relevant only to the required studios.

* 1. Filtering out unnecessary Genres:
     1. Check all Genres available

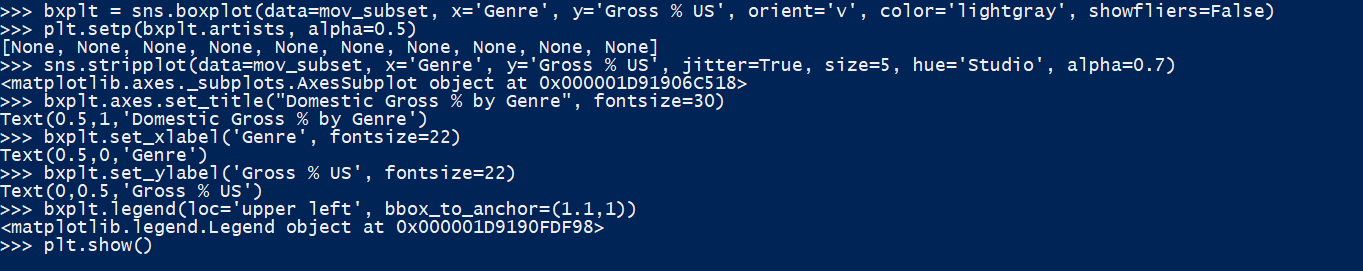
  
There are 15 unique genres as shown above but only 5 are required.



* + 1. Create a further filtered subset of the data-set with required genres only



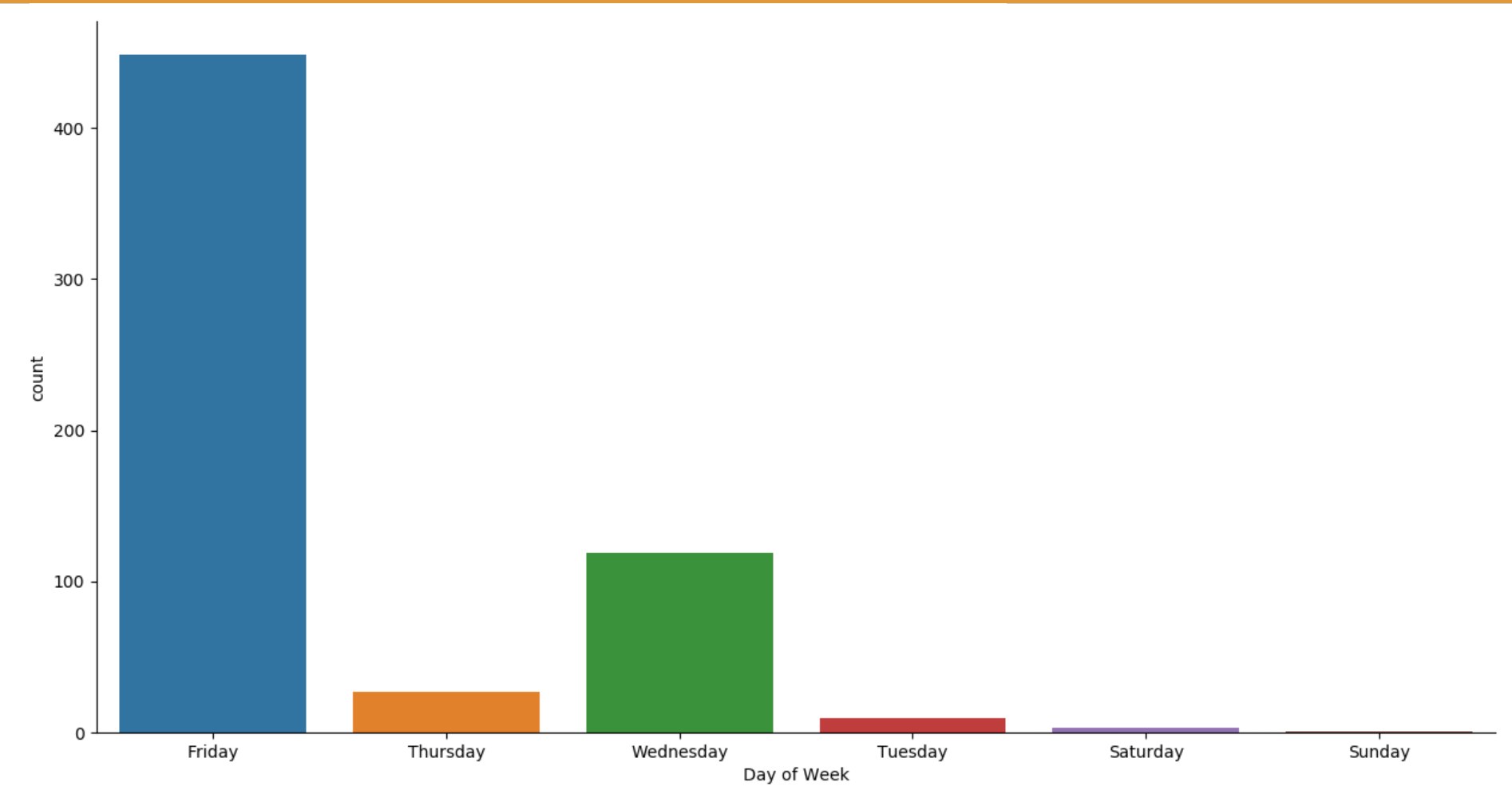
* 1. Building the plots:
     1. Create the Boxplot
     2. Set the transparency
     3. Create the jitter plot
     4. Set the title, font size, align legends



The jitter plot resides over the box plot in this case. It is colored by the Studio(hue=’Studio’). Different labels with font sizes are set using the ‘set\_title’ and ‘set\_xlabel’ and ‘set\_ylabel’. The replacement of legends is an interesting functionality which is implemented using ‘loc’ and ‘bbox\_to\_anchor’ parameters. Here, the upper left corner of the legend box is anchored to the co-ordinates (1.05,1) (where 0,0 is the upper left corner of the plot) of the whole plot object ‘bxplt’.

1. The plot: Please refer to Gross % US by Genre.png in the folder.

Other interesting observations

1. Number of movies launched by days of the week:  
     
   NO movies launched on a Monday and very few on Saturdays and Sundays. On the other end of the spectrum is Friday when the maximum number of movies are launched.