**Background info**

The study data 249 mice identified with SCC tumor growth and treated through a variety drug regimens and tumor development was observed and measurement for 45 days. The purpose of the study is Capomulin, versus the other treatment regimens

**Observation 1:**

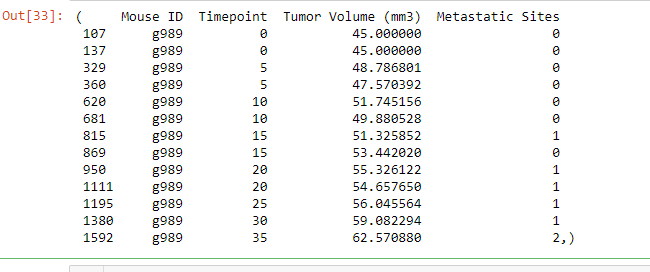
Data Quality issue: Duplicates, redundancies and inconsistencies in the study data.

Two datasets were given first tasks in any analysis in join the analysis join the meta data with the study data and then data clean up. In this case cleanup was done by removing the duplicates.

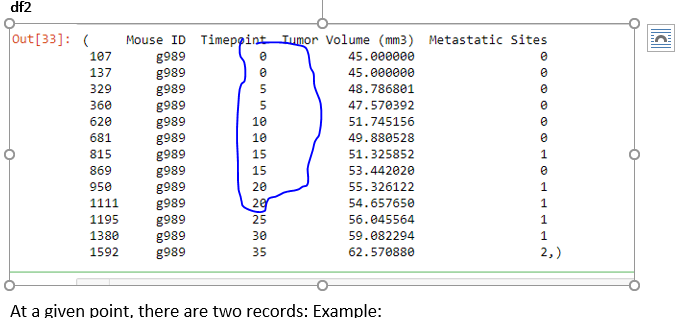
# Example of duplicates:

# df2 = study\_results.loc[study\_results["Mouse ID"] == 'g989'],

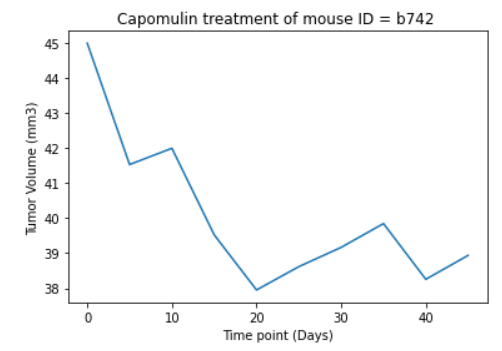
# df2



At a given point, there are two records: Example:

: 

**Observation# 2**: Based on the Generate a line plot of tumor volume vs. time point for a mouse treated with Capomulin on a Mouse ID = b742

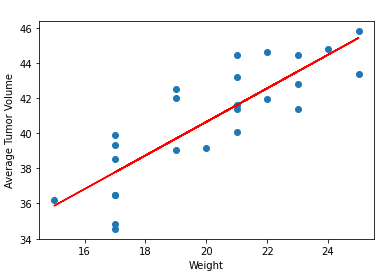


At time point 0 (Day = 0) the tumor volume = 45

At time point 20 (Day =2 0) the tumor volume = 38. This means that treatment is working but then at day 35, tumor volume went up to a volume of 40 mm3 but on day 40 volume is 38.5.

Conclusion: Overall treatment (Capomulin) works on a mouse (b742) and Tumor volume went down.

**Observation #3**: Based on the correlation coefficient and linear regression model for mouse weight and average tumor volume for the Capomulin regimen



The r-value is: 0.7088568047708717

Conclusion: Size of the tumor increases as weight increases