# Pymaceuticals Inc. Drug Analysis and Conclusion

**Analysis**

Pymaceuticals Inc., is a new pharmaceutical company that specializes in anti-cancer medications. Recently, it began screening for potential treatments for squamous cell carcinoma (SCC), a commonly occurring form of skin cancer.

* This analysis is conducted on the complete test data from their most recent animal study.
  + In this study, 249 mice who were identified with SCC tumors received treatment with a range of drug regimens.
  + Over the course of 45 days, tumor development was observed and measured.
  + The purpose of this study was to compare the performance of Pymaceuticals’ drug of interest, Capomulin, against the other treatment regimens.
* A combined study results dataframe is created as a first step. This dataframe has results from both the test data csv files ("Mouse\_metadata.csv" and "Study\_results.csv")
* We need to clean the data by identifying the duplicate MouseID and making sure our data is uniquely identified by Mouse ID and Timepoint. Finally, we create a clean DataFrame by dropping the duplicate mouse by its ID.
* The mice count in the clean dataframe is 248 after getting rid of all the duplicates.
* The next step is to generate summary statistics table of mean, median, variance, standard deviation, and SEM of the tumor volume for each regimen.
* We then generate bar plots showing the total number of rows (Mouse ID/Timepoints) for each drug regimen using Pandas and do the same exercise using pyplot as well.
* We also generate a pie plot showing the distribution of female versus male mice using Pandas and pyplots.
* Quartiles, Outliers and Boxplots are calculated that helps us generate a box plot that shows the distrubution of the tumor volume for each treatment group.
* Generate a line plot of tumor volume vs. time point for a single mouse treated with Capomulin.
* Generate a scatter plot of mouse weight vs. the average observed tumor volume for the entire Capomulin regimen.
* Calculate the correlation coefficient and a linear regression model for mouse weight and average observed tumor volume for the entire Capomulin regimen.
* Final analysis, deductions and conclusions based on all the above graphs and correlation coefficient is documented at the end of this notebook after all the calculations are completed.

# Observations and Conclusions

# From the bar graphs we can conclude that over 200 mice were given Capomulin and Ramicane drugs this count is high than any other type of drugs under the study scope.

* Pie charts shows us that the gender distribution of mice as test subjects was a bit more for the male mice at 51%.
* We found one outlier in the analysis for Infubinol drug.
* Capomulin and Ramicane showed better results in reducing the tumor size than the other 2 drugs Infubinol and Ceftamin.
* The line plot of tumor volume vs. timepoint for a single mouse treated with Capomulin shows that the tumor volume decreased with a regular dose of Capomulin drug.
* The scatter plot of mouse weight vs. the average observed tumor volume for the entire Capomulin regimen indicates that as the weight of the mouse increased the average tumor volume increased too.
* The correlation coefficient and a linear regression model calculation for mouse weight and average observed tumor volume for the entire Capomulin regimen shows that there is a strong positive correlation between mouse weight and average tumor volume, it is 0.84. As the mouse weight increases, the average tumor volume increases too.