In this project, we are given the below task:

Pymaceuticals Inc., a burgeoning pharmaceutical company based out of San Diego, specializes in anti-cancer pharmaceuticals. In its most recent efforts, it began screening for potential treatments for squamous cell carcinoma (SCC), a commonly occurring form of skin cancer.

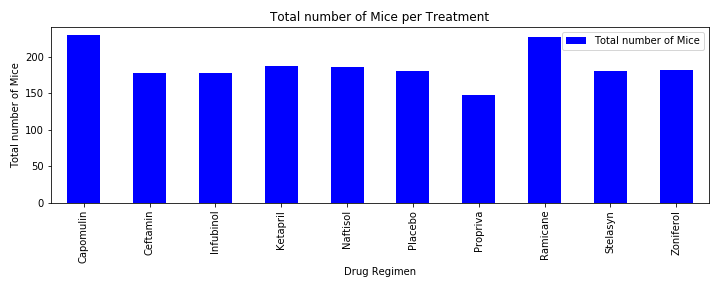
As a senior data analyst at the company, we've been given access to the complete data from their most recent animal study. In this study, 249 mice identified with SCC tumor growth were treated through a variety of drug regimens. Over the course of 45 days, tumor development was observed and measured. The purpose of this study was to compare the performance of Pymaceuticals' drug of interest, Capomulin, versus the other treatment regimens. We have been tasked by the executive team to generate all of the tables and figures needed for the technical report of the study. The executive team also has asked for a top-level summary of the study results.

**Instructions**

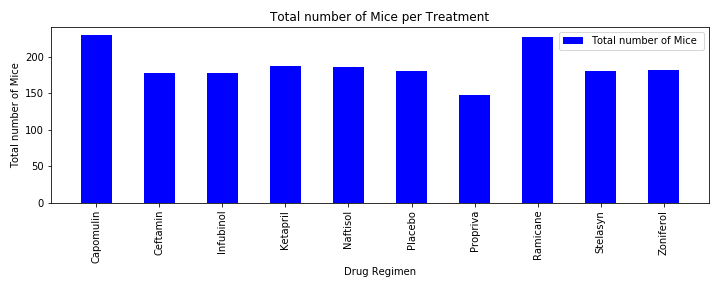
Our tasks are to do the following:

* Before beginning the analysis, check the data for any mouse ID with duplicate time points and remove any data associated with that mouse ID.
* Use the cleaned data for the remaining steps.
* Generate a summary statistics table consisting of the mean, median, variance, standard deviation, and SEM of the tumor volume for each drug regimen.
* Generate a bar plot using both Pandas's DataFrame.plot() and Matplotlib's pyplot that shows the number of total mice for each treatment regimen throughout the course of the study.

Bar plot using Pandas:

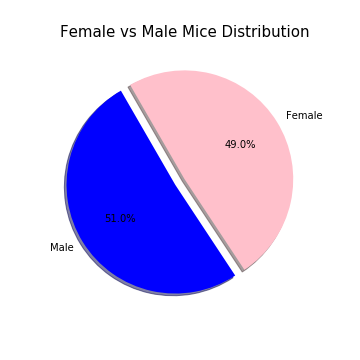
[](https://github.com/UoT-Bootcamp/Python-Matplotlib-Challenge/blob/master/Images/bar_chart_pandas.png)

Bar plot using Matplotlib:

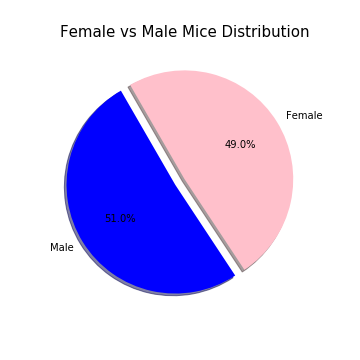
[](https://github.com/UoT-Bootcamp/Python-Matplotlib-Challenge/blob/master/Images/bar_chart_pyplot.png)

* Generate a pie plot using both Pandas's DataFrame.plot() and Matplotlib's pyplot that shows the distribution of female or male mice in the study.

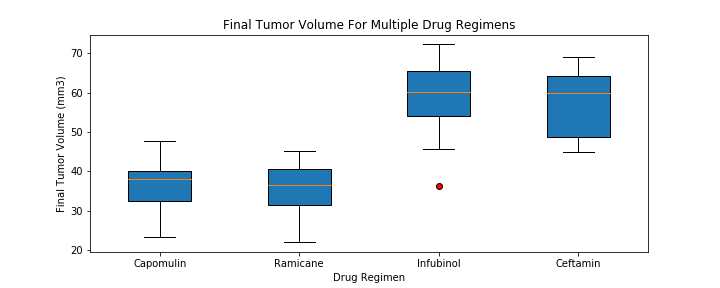
Pie plot using Pandas:

[](https://github.com/UoT-Bootcamp/Python-Matplotlib-Challenge/blob/master/Images/pie_chart_pandas.png)

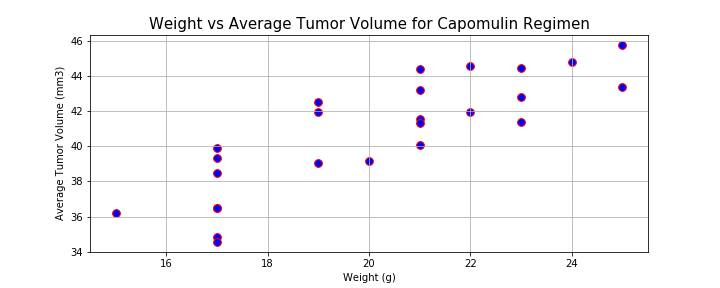
Pie plot using Matplotlib:

[](https://github.com/UoT-Bootcamp/Python-Matplotlib-Challenge/blob/master/Images/pie_chart_pyplot.png)

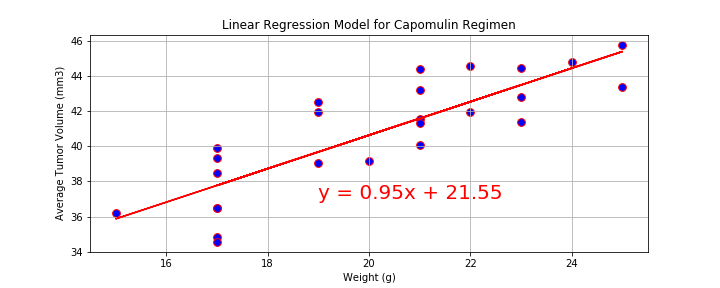
* Calculate the final tumor volume of each mouse across four of the most promising treatment regimens: Capomulin, Ramicane, Infubinol, and Ceftamin. Calculate the quartiles and IQR and quantitatively determine if there are any potential outliers across all four treatment regimens.
* Using Matplotlib, generate a box and whisker plot of the final tumor volume for all four treatment regimens and highlight any potential outliers in the plot by changing their color and style.

[](https://github.com/UoT-Bootcamp/Python-Matplotlib-Challenge/blob/master/Images/box_plot.png)

* Generate a scatter plot of mouse weight versus average tumor volume for the Capomulin treatment regimen.

[](https://github.com/UoT-Bootcamp/Python-Matplotlib-Challenge/blob/master/Images/scatter_plot.png)

* Calculate the correlation coefficient and linear regression model between mouse weight and average tumor volume for the Capomulin treatment. Plot the linear regression model on top of the previous scatter plot.

[](https://github.com/UoT-Bootcamp/Python-Matplotlib-Challenge/blob/master/Images/scatter_plot_linear.png)

Look across all previously generated figures and tables and write at least three observations or inferences that can be made from the data. Include these observations at the top of notebook.

Here are some final considerations:

* We must use proper labeling of your plots, to include properties such as: plot titles, axis labels, legend labels, x-axis and y-axis limits, etc.