# Introduction

The purpose of this project is to perform data analysis on either the titanic or baseball datasets, it is to see whether we are capable of analysing the data and being able to form questions just by looking at the data. It is to whether we capable of answering the questions using the data provided. The dataset has information on the passengers of the titanic.

I used the python notebook also numpy, pandas, matplotlib.pyplot and seaborn python modules in order to analysis the data

# Questions

1. How many data records where available about the passengers?
2. How many female survivors?
3. How many male survivors?
4. How many female deaths?
5. How many male deaths?
6. How many children survived?
7. How many children did not survive?
8. How many records are missing the age data?

# Data Exploration and visualization

To come up with the questions I looked through the titanic dataset and came up with questions that I was curious to know and looked into the relationship between certain variables. I also used visualization by using pie charts, bar charts and tables to show the relationship between the variables.

I uploaded the titanic data and after doing that I looked for duplicates and checked whether the records where all correctly uploaded. I checked the total amounts of males and females, also created a histogram which depicts the ages. Also checked the mean, standard deviation, maximum and medium basically exploring some descriptive statistics. Also explored the data by calculating the number of missing values and cabin had 687 which was the highest and age was the second highest with 177 missing values.

There was also a look in the number of survivors and deaths between the males and females which also looked into the classes and gave the totals according to the passenger classes. There are pie charts which show the survivors and deaths between the males and females and there is a bar graph which shows the children which are survivors and non- survivors. There is also a look into how many children there are in the dataset plus how many males and females have missing age values.

# Conclusion

From the results age was not a big deal but the passenger class is important as it affects certain things from happening and could have with survival. There were more deaths in the third class then the others and the males had a large number of deaths then females.