Milestone report, practicing data story skills:

**Introduction**

The sinking of the RMS Titanic is one of the most infamous shipwrecks in history.  On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships.

One of the reasons that the shipwreck led to such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others, such as women, children, and the upper-class.

:

**The problem I am trying to solve:**

*In this challenge, I am going to complete an analysis of what sorts of people were likely to survive. In particular, applying the tools of machine learning to predict which passengers survived the tragedy.*

**The “client” and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?**

*This project analysis is for Slide Rule’s, “all up” learning assessment.*

**The Datasets I am going to use for this analysis:**

*I will use the Titanic train and test datasets.*

**In brief, outline your approach to solving this problem (knowing that this might change later). What are your deliverables?**

*I will upload my code, documentation and final presentation to github*

Phase one – conducting analysis on the “Train” dataset, in order to answer the following questions:

* + Load the datasets
  + Install the packages
  + What important fields and information does the data set have?
    - Attach sample
  + What are its limitations i.e. what are some questions that you cannot answer with this data set?
    - Give example
  + What kind of cleaning and wrangling is needed?
  + Any preliminary exploration you’ve performed and your initial findings.
  + Determine what kind of prediction model to use.
  + Prototype a prediction model.

Phase two – Apply the model against “Test” dataset

* Create a baseline, with the aim of beating the baseline thru using the model I create.
* Review the accuracy of the model against Train.
* Create visualizations of the outputs

Phase three –Summary and Conclusions.

* Outline of the findings e.g. “single men with quarters on the upper deck rooms survived in more numbers than other men on with quarters on lower decks.”
* Visualizations of the results (charts, etc.)
* Create a presentation (slide deck) summarizing my technical approach, findings and any recommendations.