**Introduction**

**Purpose:** To performa data analysis on a sample Titanic dataset.

This dataset contains demographics and passenger information from 891 of the 2224 passengers and crew on board the Titanic. You can view a description of this dataset on the Kaggle website, where the data was obtained (<https://www.kaggle.com/c/titanic/data>).

**Questions**

What factors made people more likely to survive?

1. Were social-economic standing a factor in survival rate?
2. Did age, regardless of sex, determine your chances of survival?
3. Did women and children have preference to lifeboats (survival)?
4. Did women with children have a better survival rate vs women without children (adults 18+)?

**Purpose:**

**Analyse the data on his eye from the Titanic data set.**

**Required:**

One of the reasons for the loss of life for Titanic passengers was that there were not enough lifeboats for crew and passengers, despite the luck and participation in surviving the sinking, and they were the most surviving children, women and the upper class**.**

**The question to answer is:**

**What factors helped survivors survive?**

### Data Description

(from <https://www.kaggle.com/c/titanic>)

* **survival:** Survival (0 = No; 1 = Yes)
* **pclass:** Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)
* **name:** Name
* **sex:** Sex
* **age:** Age
* **sibsp:** Number of Siblings/Spouses Aboard
* **parch:** Number of Parents/Children Aboard
* **ticket:** Ticket Number
* **fare:** Passenger Fare
* **cabin:** Cabin
* **embarked:** Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)

**Special Notes:**

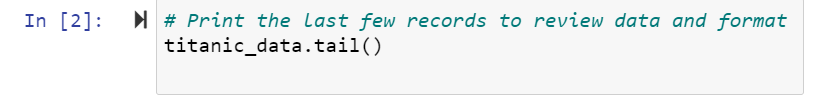
* **Sibling:** Brother, Sister, Stepbrother, or Stepsister of Passenger Aboard Titanic
* **Spouse:** Husband or Wife of Passenger Aboard Titanic (Mistresses and Fiances Ignored)
* **Parent:** Mother or Father of Passenger Aboard Titanic
* **Child:** Son, Daughter, Stepson, or Stepdaughter of Passenger Aboard Titanic

**Here we will provide the initial data, after contacting the following libraries and reading the data file in excel format:**

**Text

Description automatically generated**

**Here we will present some last data for the Titanic Group:**

****

**Data Cleanup**

From the data description and questions to answer, I've determined that some dataset columns will not play a part in my analysis and these columns can therefore be removed. This will decluster the dataset and also help with processing performance of the dataset.

* PassengerId
* Name
* Ticket
* Cabin
* Fare
* Embarked

I'll take a 3 step approach to data clean up:

1. Identify and remove any duplicate entries
2. Remove unnecessary columns
3. Fix missing and data format issues

**Text

Description automatically generated**

**Output:**

**Table

Description automatically generated with medium confidence**

**Step 2 - Remove unnecessary columns**

Columns (PassengerId, Name, Ticket, Cabin, Fare, Embarked) removed



**Output:**

**Table

Description automatically generated**

#### **Step 3 - Fix any missing or data format issues:**

Graphical user interface, text, application

Description automatically generated

Output:

A picture containing table

Description automatically generated

Text

Description automatically generated

Output:

Table

Description automatically generated

A picture containing text

Description automatically generated

Output:

Text

Description automatically generated with medium confidence

Data Exploration and Visualization:

Graphical user interface, text, application

Description automatically generated

Output:

Table

Description automatically generated

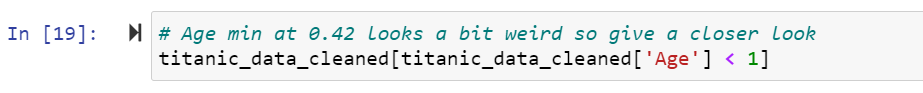
Text

Description automatically generated

Output:

Table

Description automatically generated



Output:

Table

Description automatically generated

Graphical user interface, text

Description automatically generated

Questions :

Q1-Were social-economic standing a factor in survival rate?

Graphical user interface, text

Description automatically generated with medium confidence

Text

Description automatically generated

A picture containing chart

Description automatically generated

Output:

Table

Description automatically generated with medium confidence

Error:in sns object:

Chart

Description automatically generated with medium confidence

Q2-Did age, regardless of sex and class, determine your chances of survival?

Graphical user interface, text, application

Description automatically generated

Output:

Graphical user interface, text, application, email

Description automatically generated

Q3- Did women and children have preference to lifeboats and therefore survival (assuming there was no shortage of lifeboats)?

Graphical user interface, text, application

Description automatically generated

Output:

Graphical user interface, text, application, email

Description automatically generated

Q4-Did women with children have a better survival rate vs women without children (adults 18+)?

* "Women with children" is referring to parents only

Table

Description automatically generated

## Q5- how did children with nannies fare in comparison to children with parents. Did the nanny "abandon" children to save his/her own life?

* Need to review list for children with no parents. These will be children with nannies as stated in the data description
* Compare "normal" survival rate of children with parents against children with nannies

Text

Description automatically generated

Output:

Table

Description automatically generated with medium confidence

Text

Description automatically generated

Output:

Graphical user interface, text, application

Description automatically generated

## And Finally ,Coclusion:

The results of the analysis, although tentative, would appear to indicate that class and sex, namely, being a female with upper social-economic standing (first class), would give one the best chance of survival when the tragedy occurred on the Titanic. Age did not seem to be a major factor. While being a man in third class, gave one the lowest chance of survival. Women and children, across all classes, tend to have a higher survival rate than men in genernal but by no means did being a child or woman guarentee survival. Although, overall, children accompanied by parents (or nannies) had the best survival rate at over 50%.

**Issues:**

* A portion of men and women did not have Age data and were removed from calculations which could have skewed some numbers
* The category of 'children' was assumed to be anyone under the age of 18, using today's North American standard for adulthood which was certainly not the case in the 190

**References:**

* <https://www.kaggle.com/c/titanic/data>