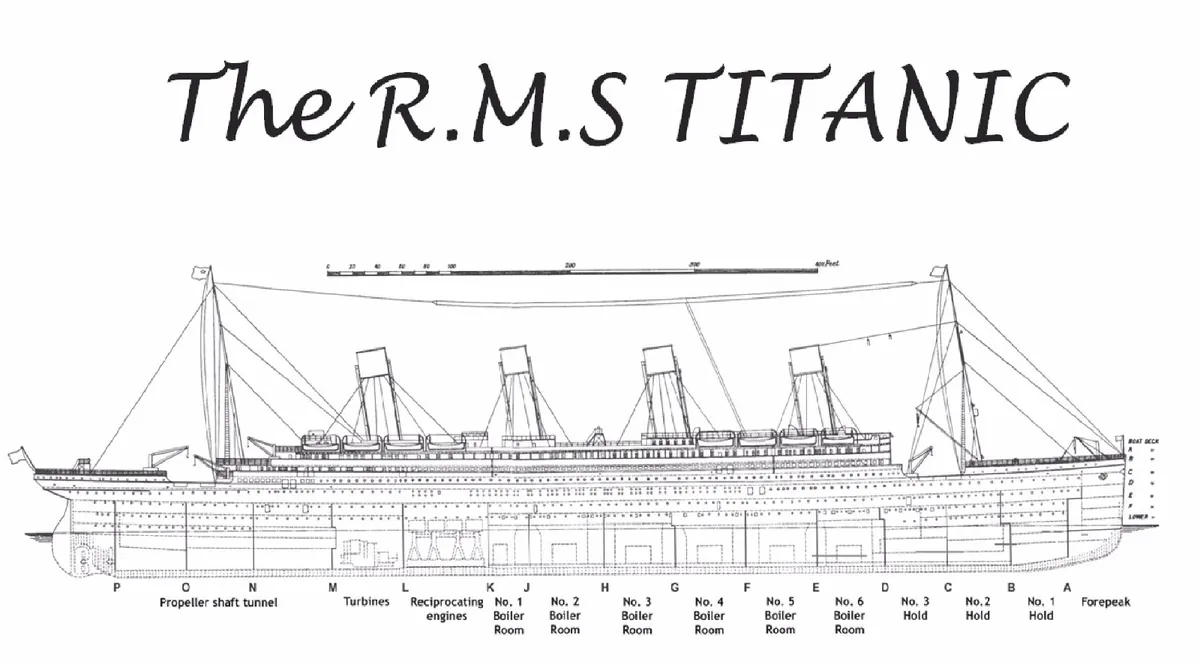
What questions does this data analysis attempt to answer?

* Correlations?
  + Missing a lot of values in the ‘age’ column and the ‘cabin’ column
  + Filling in the missing values in the ‘age’ column
    - A graph of age and age

      Description automatically generated
    - Both male and female age distributions seem to be the same
    - Correlation coefficients:
    - Age and Fare: 0.179
    - Age and Parch: -0.151
    - Age and SibSp: -0.244
    - Age and Pclass: -0.41
    - Age and Pclass seem to have the strongest correlation
* Median age for each Pclass:
  + 1: 39.0
  + 2: 29.0
  + 3: 24.0
* Fill in the age column based on the passenger’s class
* Fare and Pclass: -0.5586
* Get the median and fill in the missing data

Beyond surface-level visualization and analysis:

Filling in the holes in the ‘cabin’ column seems imperative



A few things about the ‘cabin’ column:

* We do NOT need the specific cabin numbers 🡪 just the letter of each cabin is enough

A graph of a distribution of deck

Description automatically generated 🡪 created a new column called ‘Deck’

Deck

NaN 1014

C 94

B 65

D 46

E 41

A 22

F 21

G 5

T 1

Deck – (median Fare of each deck)

Deck

A 35.0771

B 82.2667

C 86.2896

D 52.5542

E 53.1000

F 13.0000

G 16.7000

T 35.5000

Deck – Pclass count

Deck Pclass

A 1 22

B 1 65

C 1 94

D 1 40

2 6

E 1 34

2 4

3 3

F 2 13

3 8

G 3 5

T 1 1