

Progetto Analisi Dati – Titanic

L'RMS Titanic era una nave passeggeri britannica che affondò nell'Oceano Atlantico settentrionale nelle prime ore del mattino del 15 aprile 1912, dopo essere entrato in collisione con un iceberg durante il suo viaggio inaugurale da Southampton a New York City. C'erano circa 2.224 passeggeri e membri dell'equipaggio a bordo della nave e più di 1.500 morirono.

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

Train.csv

Descrizione:

- Survival - Survival (0 = No; 1 = Yes). Not included in test.csv file.
 - 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)
-
- 1- Calcolare statistiche descrittive e relativi grafici per tutte le variabili: per esempio percentuali di uomini e donne, percentuali di sopravvissuti per classe usando grafici a torta, istogrammi, bar chart, box-plot, etc.
 - 2- Calcolare correlazione tra tutte le variabili, usare scatter plot per coppie di variabili con interpolazione lineare, usare pairplot ed heatmaps. Gestire outliers e valori nulli se presenti.
 - 3- Scrivere una relazione di 2500 parole con i risultati dell'analisi