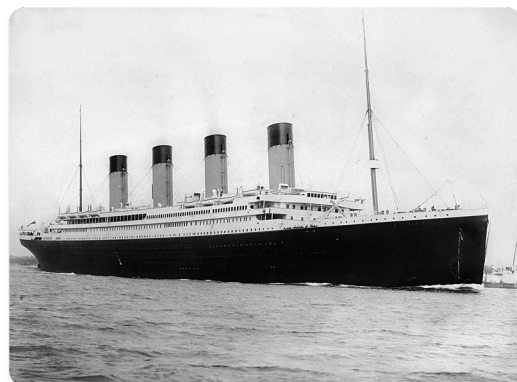


# Samostalni zadatak za vježbu 4

Radit ćemo s podacima o putnicima s Titanika, učitanim iz Excel datoteke. Cilj je analizirati podatke, identificirati i obraditi nedostajuće vrijednosti te izvesti osnovne statističke analize.



- 1. Učitajte podatke iz Excel datoteke "zadatak1.xls"**
  - Učitajte radni list **"Titanic"** u `data.frame` naziva `titanic`
  - Ako je potrebno, instalirajte i aktivirajte odgovarajući paket za rad s Excel datotekama
- 2. Koliko ima opservacija i varijabli?**
- 3. Koje je vrste varijabla `Embarked` s obzirom na mjernu skalu?**
- 4. Izračunajte osnovne statističke mjere za varijablu `Age`**
  - Aritmetička sredina (`mean()`)
  - Medijan (`median()`)
  - Standardna devijacija (`sd()`)
- 5. Ima li nedostajućih vrijednosti?**
  - `is.na()`
- 6. Ako ih ima, u kojim varijablama se nalaze?**
- 7. Izračunajte postotak nedostajućih vrijednosti za svaku varijablu.**
  - `sapply()` ili `mean()` ili `colSums()`
- 8. Kreirajte novi `data.frame` naziva `t`**
  - `t` treba sadržavati sve varijable iz `titanic`, osim one koja ima više od **50% nedostajućih vrijednosti**
  - ručno pogledamo koja varijabla ima više od 50% iz prethodnih izračuna
- 9. Dodajte novu varijablu `Imputirane` u `t`**
  - Varijabla `Imputirane` treba sadržavati podatke iz varijable koja ima više od **10% nedostajućih vrijednosti**
  - ručno pogledamo koja varijabla ima više od 00% iz prethodnih izračuna
- 10. Prikažite prvih 20 vrijednosti varijable `Imputirane`.**
- 11. Zamijenite nedostajuće vrijednosti u `Imputirane` medijanom.**
- 12. Ponovno prikažite prvih 20 vrijednosti varijable `Imputirane`.**

13. Ako u `Imputirane` više nema nedostajućih vrijednosti, uklonite iz `t` varijablu čije ste podatke kopirali u `Imputirane`.
14. Ima li još nedostajućih vrijednosti u `t`?
15. Koliko `t` sada ima opservacija?
16. Uklonite sve opservacije koje još uvijek imaju nedostajuće vrijednosti.
- `complete.cases()` ili `na.omit()`
17. Koliko `t` sada ima opservacija nakon uklanjanja nepotpunih redaka?
18. Uklonite prvi stupac iz `t`.
19. Napravite tablicu frekvencija za varijablu `Embarked`
- Prikazuje broj putnika prema luci ukrcaja
  - `table()`
20. Koliko je bilo različitih luka ukrcaja i koliko se putnika ukrcalo u svakoj?
21. Napravite tablicu kontingencije za varijable `Embarked` i `Survived`
- Spremite rezultat u `tk`
  - `table()`
22. Izračunajte postotke i zaokružite na dvije decimale
- Spremite u `tkp`
  - `proportions()` ili `xtabs()`
23. Dodajte sumarne stupce u `tkp`.
- `addmargins()`
24. Koliki je ukupni udio preživjelih putnika (`Survived: 0 = nije preživio, 1 = preživio`)?
25. Koja luka ima najviše, a koja najmanje preživjelih putnika?
- Izračunat i zaključit iz prethodnih rezultata