

Coding e Fondamenti di Intelligenza Artificiale

A.A. 2026/27 — Lezione 8: Dati

Vincenzo Nardelli



vincenzo.nardelli@unicatt.it · vincnardelli.com/cia

Dplyr - funzioni base

select()



Subset variabili
(colonne)

filter()



Subset osservazioni
(righe)

mutate()



Crea nuove variabili

summarise()



Riassumi

Dplyr - group by



group_by()
Raggruppa per

+

mutate()
Crea nuove variabili



summarise()
Riassumi

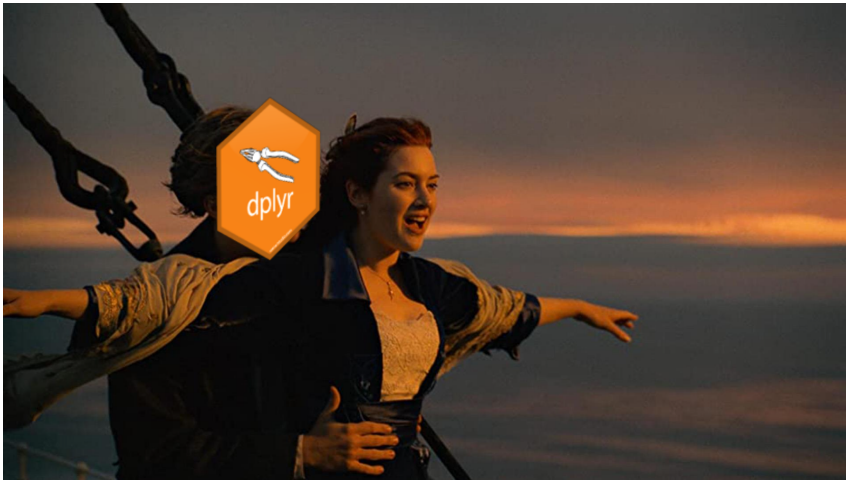




- ▶ Quanti ordini sono stati pagati con PayPal?
- ▶ Quale corriere consegna più in fretta, in giorni medi?
- ▶ I clienti Gold spendono di più per ordine dei clienti Basic?

LAB

Analisi dati Titanic



- ▶ Quante persone oltre i 50 anni erano sul titanic?
- ▶ Quale era la percentuale di uomini e donne tra gli over 50?
- ▶ In valore assoluto ci sono più uomini tra gli over 50 o tra gli under 50? Quante donne sono under 50?
- ▶ In percentuale sono sopravvissuti più passeggeri in prima o terza classe?
- ▶ Quale è stato il costo del biglietto minimo, medio e massimo, per ogni classe e per ogni sesso?
- ▶ I sopravvissuti hanno pagato di più il viaggio?

LAB

LAB - Titanic (continua)

- ▶ Le famiglie più numerose hanno avuto una percentuale più alta di sopravvivenza?
- ▶ C'è differenza nelle percentuali di sopravvivenza rispetto al porto d'imbarco?

Extra!

- ▶ Chi ha pagato i 10 biglietti più costosi?
- ▶ Quali sono i biglietti associati a più passeggeri?