

Progetto Statistica Numerica

Dataset: Brain Weight in Humans (Anubhab Swain) Hotel Bookings Analysis (The Devastator)

Federico Malservigi

0001115961

2024-2025

I Dataset selezionati

Hotel Bookings Analysis (The Devastator)

Descrizione:

Il dataset contiene informazioni dettagliate sulle prenotazioni alberghiere, inclusi dati sui clienti, durata del soggiorno, richieste speciali, tariffe, cancellazioni, e altro.

Per l'analisi sono state selezionate solo alcune colonne, ritenute rilevanti:

Colonne utilizzate:

- lead_time, stays_in_weekend_nights, stays_in_week_nights, adults, children babies, adr, total_of_special_requests (numeriche)
- is_canceled (categorica)

Dimensione finale del dataset:

118.896 righe × 9 colonne

Dataset per la Regressione

Brain Weight in Humans (Anubhab Swain)

Descrizione:

Dataset derivato da uno studio medico contenente dati su dimensioni della testa e peso del cervello, oltre a genere e fascia d'età.

Variabili utilizzate per la regressione:

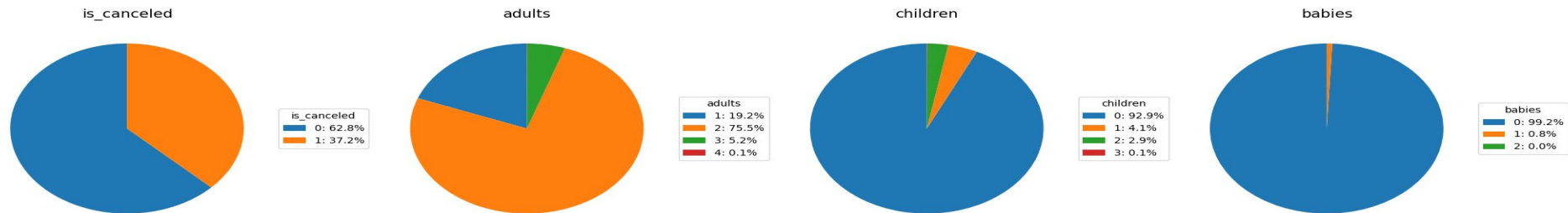
- Head Size (cm³) (numerica continua)
- Brain Weight (grams) (numerica continua)

Dimensione finale del dataset:

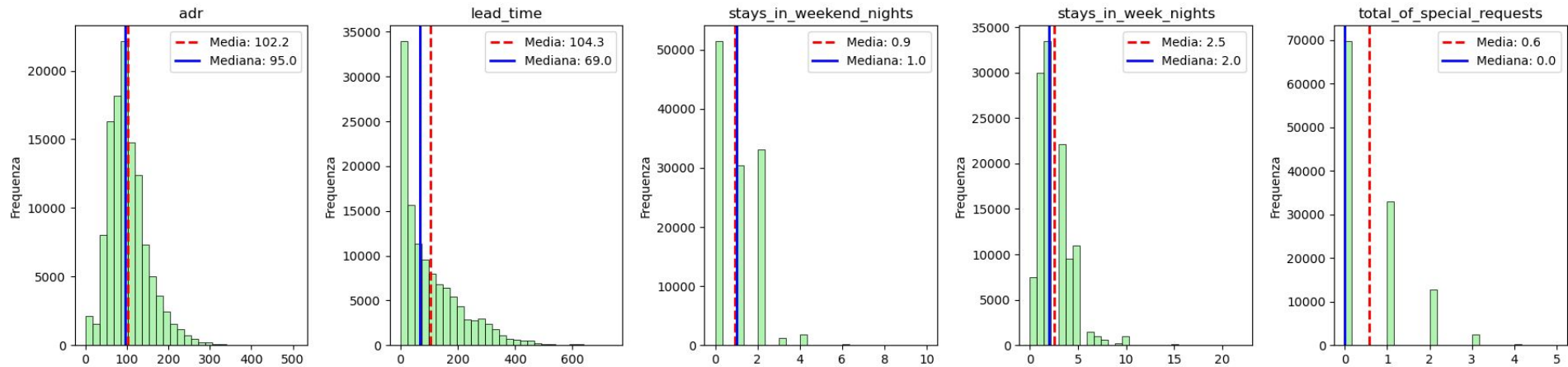
237 righe × 2 colonne

EDA Univariata (Hotel Bookings Analysis)

Diagrammi a torta

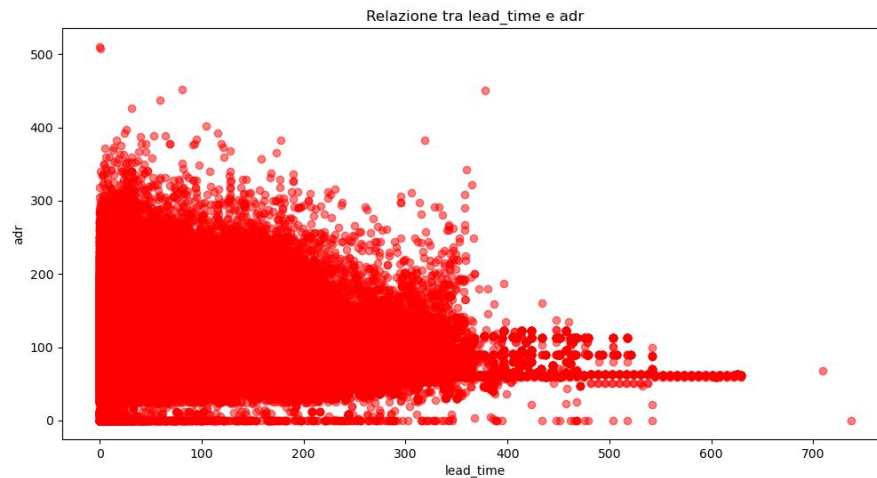


Istogrammi



EDA Bivariata (Hotel Bookings Analysis)

In questo Scatter plot possiamo notare un aumento del Prezzo medio giornaliero della camera quando il numero di giorni tra la data di prenotazione e la data di arrivo è più breve, sottolineando una proporzionalità inversa tra le due variabili

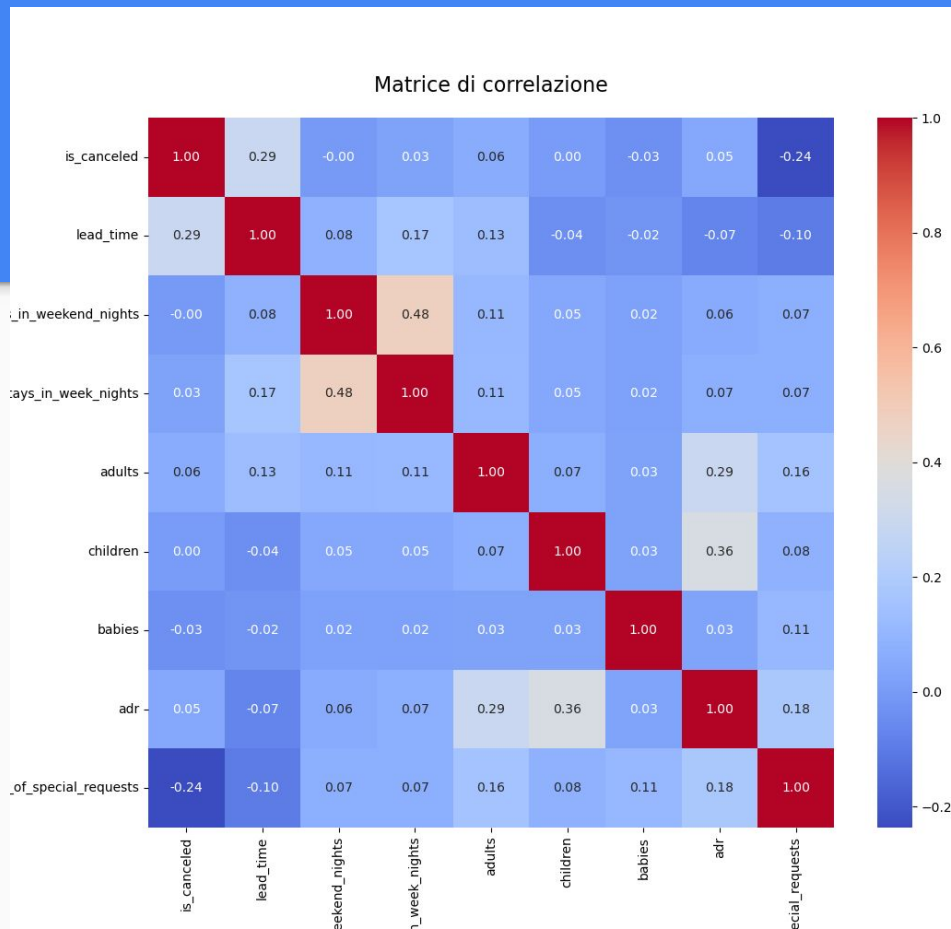


EDA Multivariata

(Hotel Bookings Analysis)

Le correlazioni più forti confermano che:

- **Lead time** influisce positivamente sul rischio di cancellazione.
- **Richieste speciali** influiscono negativamente sul rischio di cancellazione.
- **Prezzo medio giornaliero della camera** è influenzato positivamente soprattutto dal numero di adulti e bambini



Classificazione (Hotel Bookings Analysis)

Per la classificazione ho ridotto il dataset.

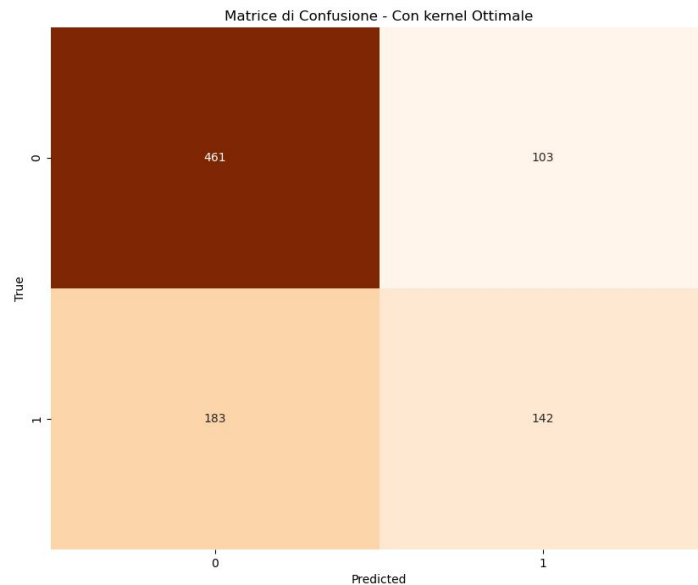
Dimensione 5% dell'originale (5924 righe)

Utilizzo la variabile target `is_cancelled`

Miglior test/validation size: 15% (70% training)

Configurazione ottimale: {'kernel': 'linear', 'C': 10}

Accuratezza sul test set: 0.6783



Tuning degli Iperparametri (Hotel Bookings Analysis)

Kernel	Degree	C	Gamma	Accuratezza Media sulla Validation Set
Linear	N/A	10	N/A	0.6977
Poly	2	10	N/A	0.6696
Poly	3	10	N/A	0.6727
Rbf	N/A	10	'scale'	0.6609

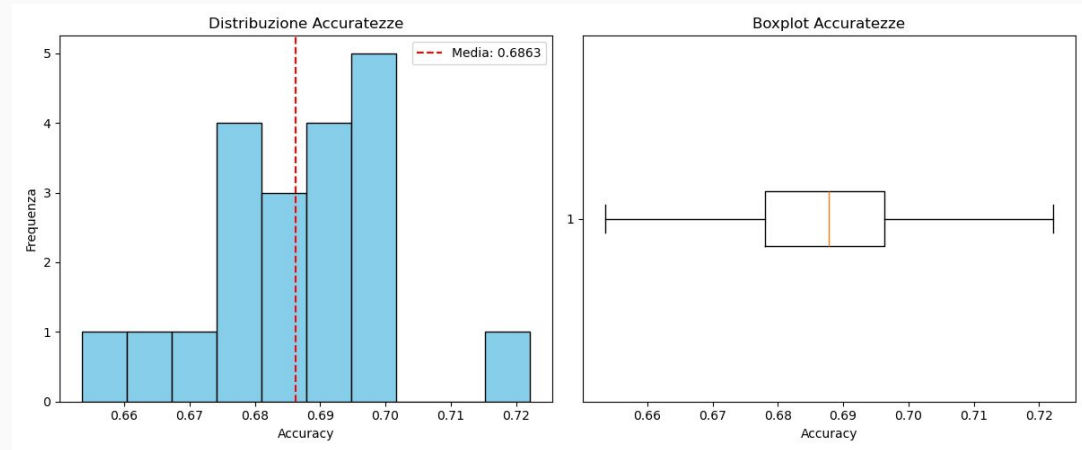
Studio statistico sui risultati della valutazione

(Hotel Bookings Analysis)

L'addestramento del del modello è stato fatto 20 volte. Di ogni metrica di errore abbiamo quanti un SRS(20)

Campioni (k): 20
Media accuratezze: 0.6863
Deviazione standard: 0.0146
Minimo: 0.6535
Massimo: 0.7222
Mediana: 0.6879

=== INFERENZA STATISTICA ===
Intervallo di confidenza al 95.0%:
(0.6793, 0.6933)



Regressione Lineare Semplice

(Brain Weight in Humans)

Variable indipendente -> Head Size(cm^3)

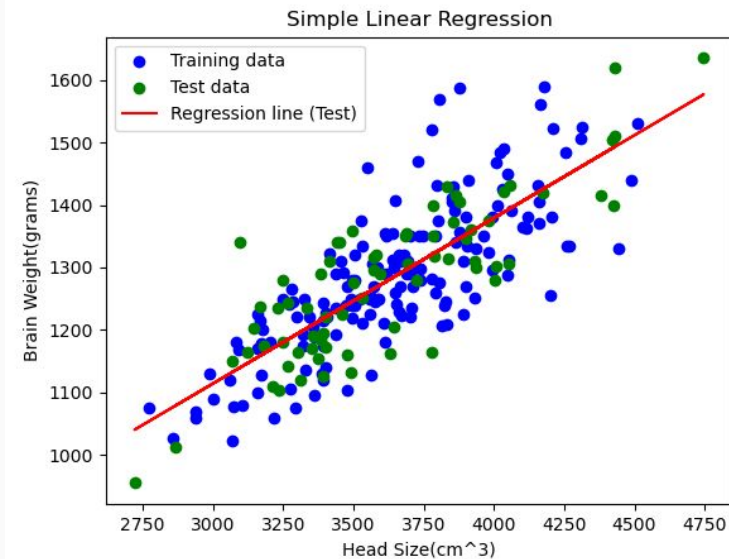
Variable dipendente -> Brain Weight (grams)

Intercetta (B_0): 319.83

Pendenza (B_1): 0.2648

Coefficiente di semplice di determinazione (R^2)(test set): 0.70

MSE: 4617.92



Analisi di Normalità dei residui (Test set)

(Brain Weight in Humans)

Shapiro-Wilk p-value: 0.8087

I residui seguono una distribuzione normale (non rifiutiamo H_0)

