




# To host or not to host, that is the question



ANDREA BALDININI  
JACOPO BONANNO  
SABRINA INFANTE



# Overbooking

Che cos'è

Possibilità data alle strutture ricettive di accettare un numero superiore di prenotazioni rispetto alla loro vera capacità.

I proprietari di tali strutture ricorrono a questa pratica per prepararsi ad eventuali cancellazioni

---

# Road map

## Analisi esplorativa

- Esplorazione dataset
- Studio delle distribuzioni
- Studio delle correlazioni
- Principal Component Analysis
- Analisi Cluster

## Modelli di classificazione

- Algoritmo
- LDA
- Regressione Logistica
- KNN
- SVM
- Random Forest

## Conclusioni

- Modello migliore
- Variabili decisive

# Analisi Esplorativa

---

# Il dataset

## Le variabili

- Number of adults
- Number of children
- Number of weekend nights
- Number of week nights
- Lead time
- Arrival month
- Number of previous cancellations
- Number of previous booking not cancelled
- Average price per room
- Number of special requests
- **Booking status** → Variabile oggetto di studio

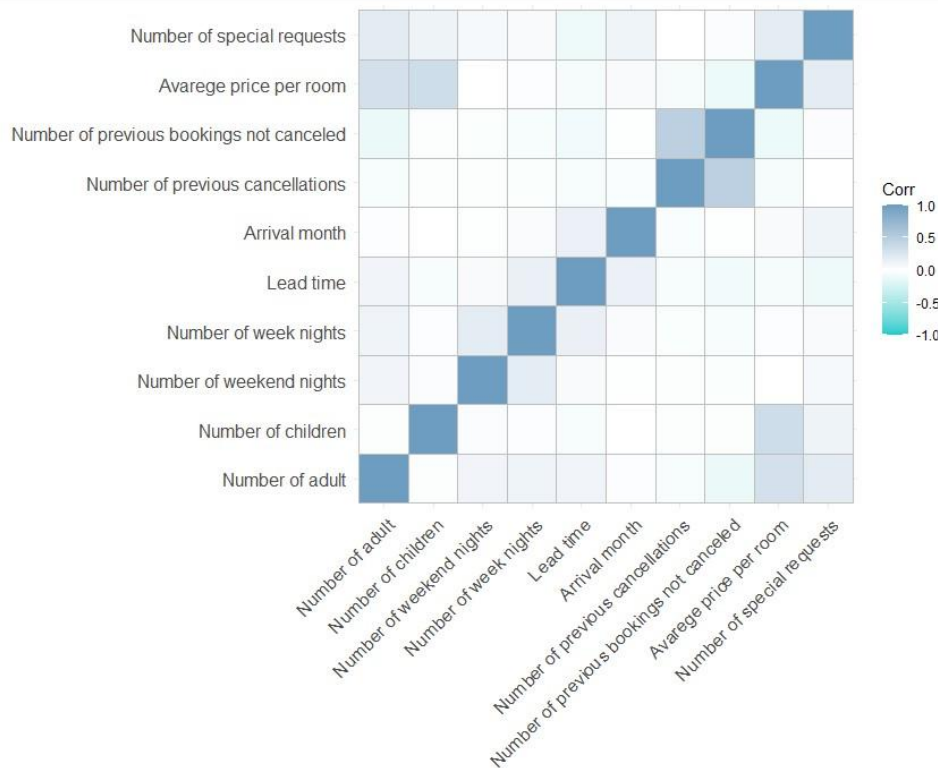
## Caratteristiche dataset

- Numero di osservazioni: 36.275
- Numero di variabili: 11
- Numero di valori NA: 0

# PairPlot



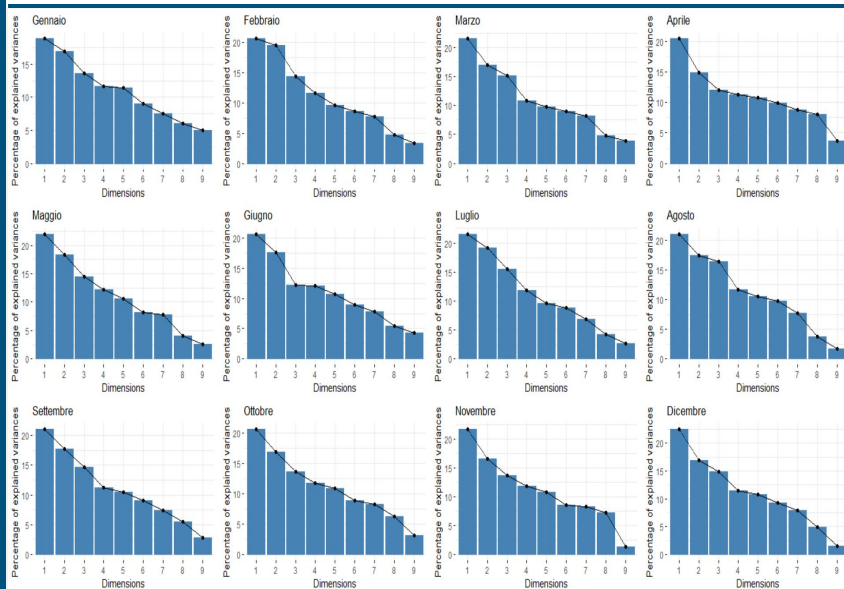
# Correlazioni



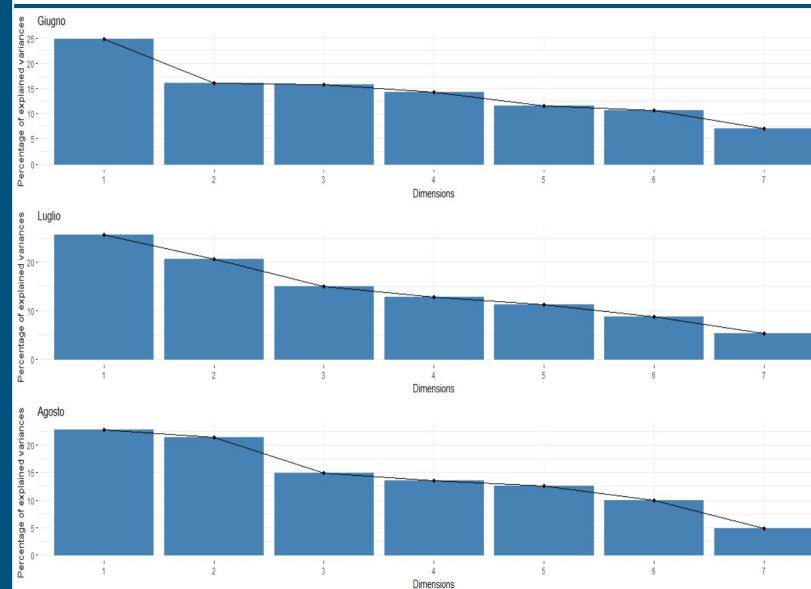
- Moderata e positiva correlazione tra number of previous bookings not canceled e number of previous cancellations (0.468)
- Inferiore correlazione tra il number of children ed average price per room (0.338)
- Diffusa incorrelazione o leggermente negativa tra le altre variabili
- Quelle correlate si può supporre dovranno essere rimosse dall'analisi, ma questa decisione deriva da una PCA

# Pca - Varianza spiegata

## Var. spiegata per mese modello generale



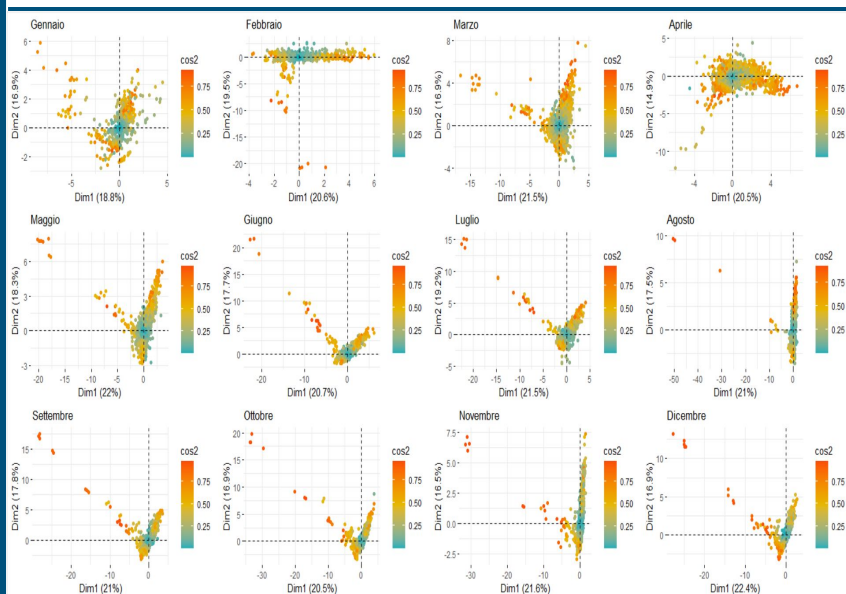
## Var. spiegata per mesi target e modello ridotto





# Pca - Scatter Plot

## Plot per mese modello generale

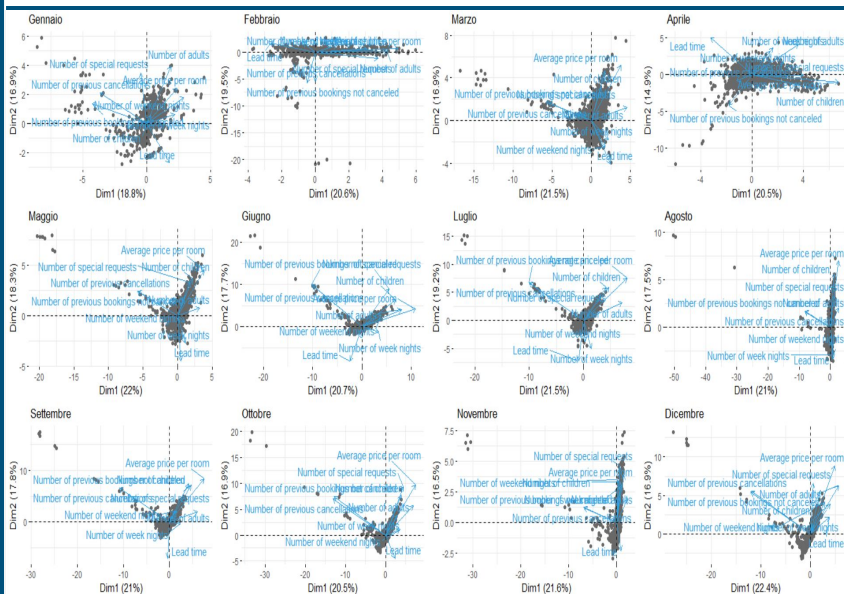


## Plot per mesi target e modello ridotto

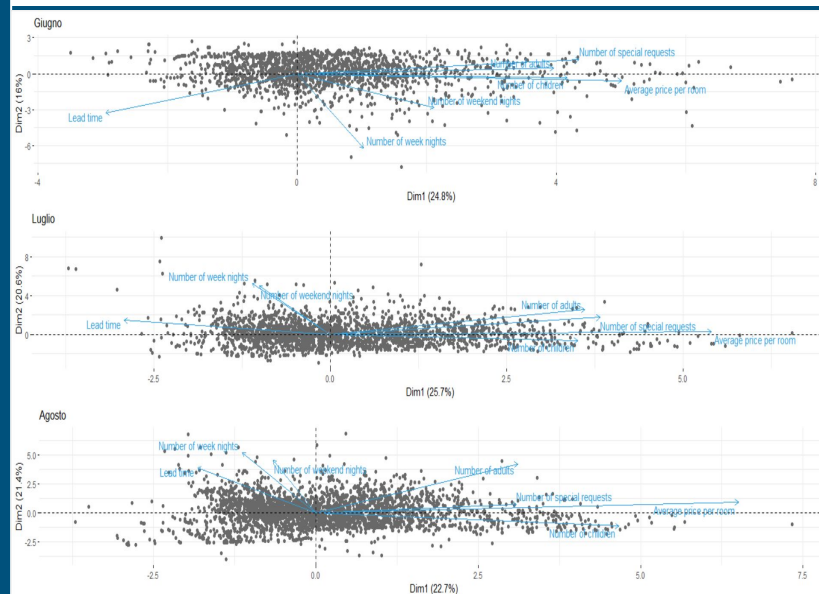


# Pca - Biplot

## Biplot per mese modello generale

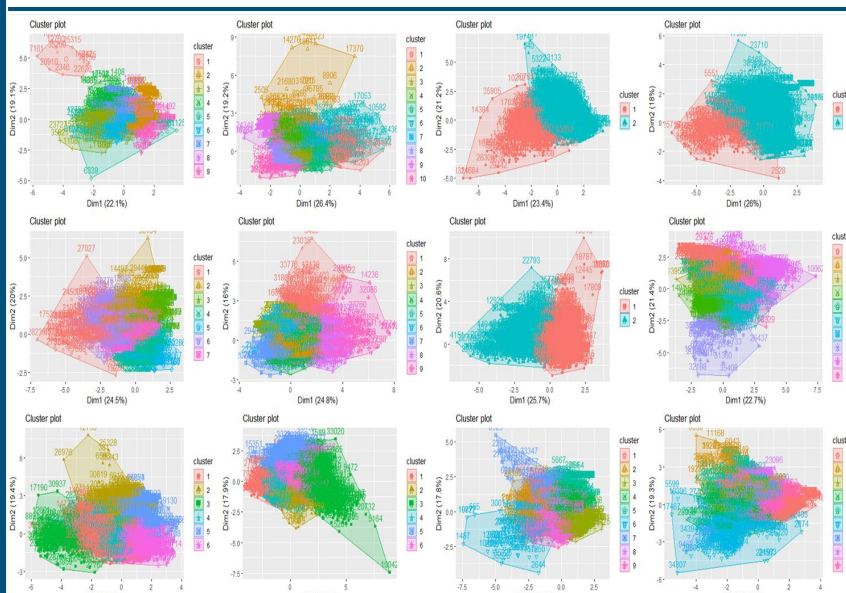


## Biplot per mesi target e modello ridotto

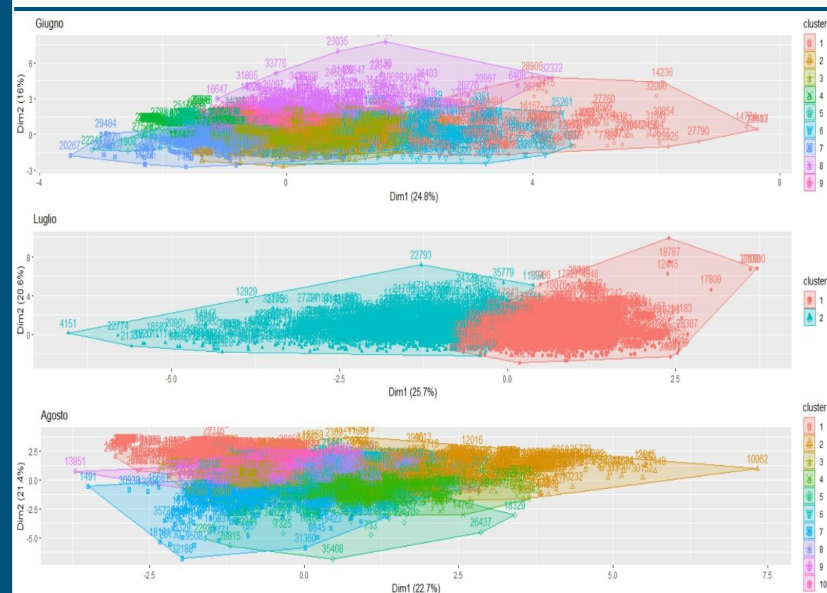


# K-means

## K-means per mese



## K-means per mesi target

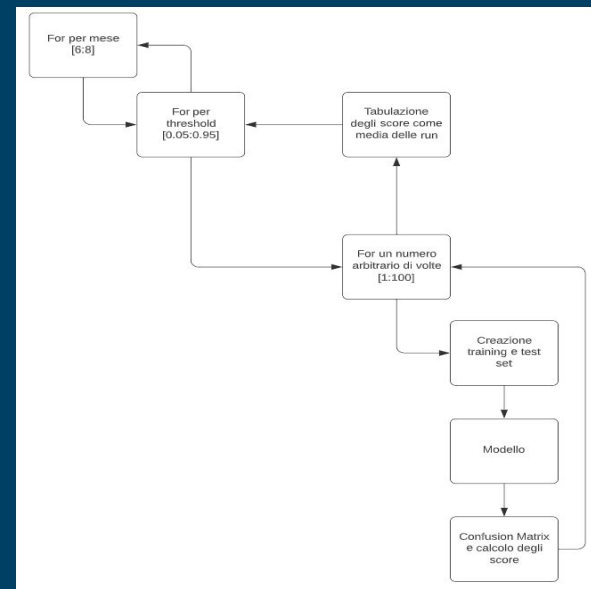


# Modelli di classificazione

---

# Algoritmo validativo

Metodologia con la quale  
abbiamo implementato ciascun  
modello

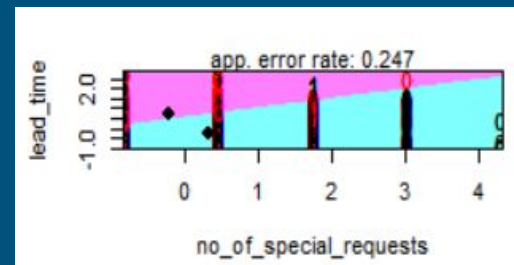
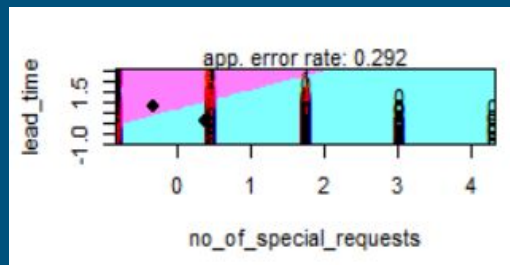
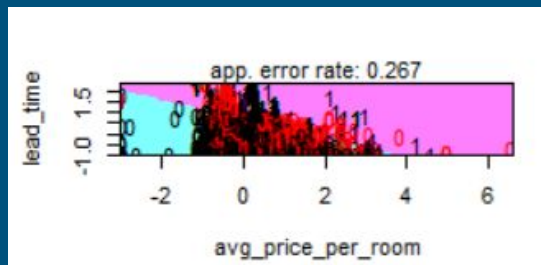


Threshold	Giugno				Luglio				Agosto			
	Precision	Accuracy	Sensitivity	Specificity	Precision	Accuracy	Sensitivity	Specificity	Precision	Accuracy	Sensitivity	Specificity
0.05	0.82	0.86	0.82	0.88	0.77	0.78	0.75	0.81	0.70	0.77	0.72	0.80
0.10	0.82	0.85	0.81	0.88	0.76	0.77	0.73	0.81	0.72	0.79	0.73	0.82
0.15	0.84	0.87	0.84	0.89	0.75	0.77	0.75	0.79	0.71	0.77	0.71	0.81
0.20	0.83	0.87	0.85	0.88	0.79	0.81	0.78	0.83	0.71	0.77	0.71	0.81
0.25	0.85	0.88	0.86	0.90	0.74	0.77	0.75	0.78	0.73	0.79	0.75	0.82
0.30	0.83	0.88	0.86	0.88	0.74	0.77	0.74	0.79	0.71	0.77	0.70	0.81
0.35	0.83	0.87	0.85	0.88	0.78	0.80	0.77	0.83	0.75	0.80	0.74	0.84
0.40	0.85	0.88	0.87	0.90	0.75	0.77	0.75	0.80	0.70	0.77	0.71	0.80
0.45	0.82	0.86	0.85	0.87	0.74	0.76	0.72	0.79	0.76	0.80	0.73	0.85
0.50	0.87	0.88	0.84	0.91	0.74	0.76	0.74	0.78	0.76	0.82	0.79	0.84
0.55	0.86	0.88	0.85	0.91	0.74	0.79	0.81	0.77	0.74	0.81	0.78	0.83
0.60	0.84	0.87	0.85	0.89	0.80	0.79	0.72	0.85	0.74	0.81	0.79	0.82
0.65	0.87	0.89	0.85	0.91	0.79	0.80	0.75	0.84	0.75	0.81	0.76	0.84
0.70	0.86	0.88	0.82	0.91	0.74	0.77	0.74	0.79	0.72	0.79	0.74	0.82
0.75	0.83	0.86	0.82	0.89	0.76	0.78	0.75	0.80	0.75	0.79	0.68	0.86
0.80	0.85	0.87	0.83	0.90	0.74	0.74	0.67	0.80	0.74	0.79	0.72	0.84
0.85	0.86	0.88	0.85	0.91	0.73	0.75	0.71	0.78	0.73	0.79	0.73	0.82
0.90	0.85	0.87	0.83	0.90	0.77	0.76	0.68	0.83	0.74	0.79	0.72	0.84
0.95	0.84	0.86	0.80	0.90	0.77	0.78	0.73	0.82	0.72	0.77	0.70	0.82

# LDA

	Giugno				Luglio				Agosto			
	Pre	Acc	Sen	Spe	Pre	Acc	Sen	Spe	Pre	Acc	Sen	Spe
Value	0.97	0.79	0.99	1.00	1.00	0.73	1.00	1.00	0.90	0.77	1.00	1.00
Threshold	0.90	0.65	0.05	0.95	0.90	0.45	0.05	0.95	0.90	0.90	0.05	0.95

## Partition Plot



# Regressione Logistica

	Giugno				Luglio				Agosto			
	Pre	Acc	Sen	Spe	Pre	Acc	Sen	Spe	Pre	Acc	Sen	Spe
Value	0.94	0.81	0.99	1.00	0.98	0.74	1.00	1.00	0.91	0.77	0.99	1.00
Threshold	0.90	0.55	0.05	0.95	0.90	0.45	0.05	0.95	0.90	0.50	0.05	0.95

## Regressori e coefficienti

	No of adults	No of weekend nights	No of week nights	Lead time	Avg price per room	No of special requests
Giugno	0.57	0.41	0.18	1.43	0.91	-1.52
Luglio	0.10	0.08	0.10	0.97	0.48	-1.00
Agosto	0.04	0.07	0.16	1.30	0.66	-0.83

# K-Nearest Neighbor & Support Vector Machines

## K-NN

	Giugno				Luglio				Agosto			
	Pre	Acc	Sen	Spe	Pre	Acc	Sen	Spe	Pre	Acc	Sen	Spe
Value	0.87	0.89	0.87	0.91	0.80	0.81	0.81	0.85	0.76	0.82	0.79	0.86
Threshold	0.65	0.65	0.40	0.65	0.60	0.20	0.55	0.60	0.45	0.50	0.50	0.75

## SVM

	Giugno				Luglio				Agosto			
	Pre	Acc	Sen	Spe	Pre	Acc	Sen	Spe	Pre	Acc	Sen	Spe
Value	0.96	0.86	1.00	0.99	0.98	0.81	1.00	1.00	0.97	0.81	0.99	1.00
Threshold	0.95	0.35	0.05	0.95	0.85	0.40	0.05	0.95	0.85	0.35	0.05	0.95



# Random Forest

	Giugno				Luglio				Agosto			
	Pre	Acc	Sen	Spe	Pre	Acc	Sen	Spe	Pre	Acc	Sen	Spe
Value	1.00	0.91	0.99	1.00	0.98	0.86	0.98	0.99	0.98	0.88	0.97	0.99
Threshold	0.90	0.60	0.05	0.90	0.80	0.45	0.05	0.95	0.90	0.50	0.05	0.90

## Variable Importance

	Giugno		Luglio		Agosto	
	MDGini	MDAccuracy	MDGini	MDAccuracy	MDGini	MDAccuracy
Lead time	535.23	190.97	433.39	235.86	688.32	291.00
No of special requests	172.75	157.75	217.62	222.99	379.44	191.07
Average price per room	312.47	128.39	338.26	158.58	178.98	178.98

Pre	0.95
Acc	0.76
Sen	0.99
Spe	1.00

Pre	0.97
Acc	0.82
Sen	1.00
Spe	1.00

Pre	0.99
Acc	0.88
Sen	0.98
Spe	1.00



Pre	0.94
Acc	0.77
Sen	0.99
Spe	1.00

Pre	0.81
Acc	0.84
Sen	0.82
Spe	0.87

# Conclusioni

- Modello migliore: Random forest
  - Accuracy mesi target: 91%, 86%, 88%
  - Variabili chiave per classificazione:
    - Lead time
    - Number of special requests
    - Average price per room
-

# Grazie per l'attenzione

Andrea Baldinini:

[S1110402@studenti.univpm.it](mailto:S1110402@studenti.univpm.it)

Jacopo Bonanno:

[S1112151@studenti.univpm.it](mailto:S1112151@studenti.univpm.it)

Sabrina Infante:

[S1113999@studenti.univpm.it](mailto:S1113999@studenti.univpm.it)

---