



ANALISI DEI DATI SPAZIALI PER LE APPLICAZIONI ECONOMICHE

START IT NOW!

Obiettivi

- Studiare una variabile d'interesse tramite un'analisi *descrittiva* e un'analisi *spaziale*
- Verificare quale dei due *approcci* spieghi meglio la variabile
- Evidenziare eventuali differenze tra i *software* (R e Python) utilizzati per lo studio

Dati e Fonti Statistiche

- Variabile d'interesse: Tasso di occupazione giovanile (15–29 Anni), valutata nelle 107 province italiane
- Anno di Riferimento: 2022
- Fonte dati: "<https://www.istat.it/it/archivio/285730>"

Metodologia

Ai fini dell'analisi è stato utilizzato principalmente il software Python, tramite il launcher Colab di Google.

Come supporto e per la verifica della coerenza dei risultati è stato fondamentale anche l'utilizzo di R.

Questo duplice approccio ci ha permesso di comprendere le principali differenze tra i due applicativi, nonché le potenzialità di Python.

Nella sola fase di elaborazione dei dati è stato utilizzato anche il software KNIME, che tramite un workflow ci ha permesso di selezionare le variabili d'interesse dal dataset di partenza.

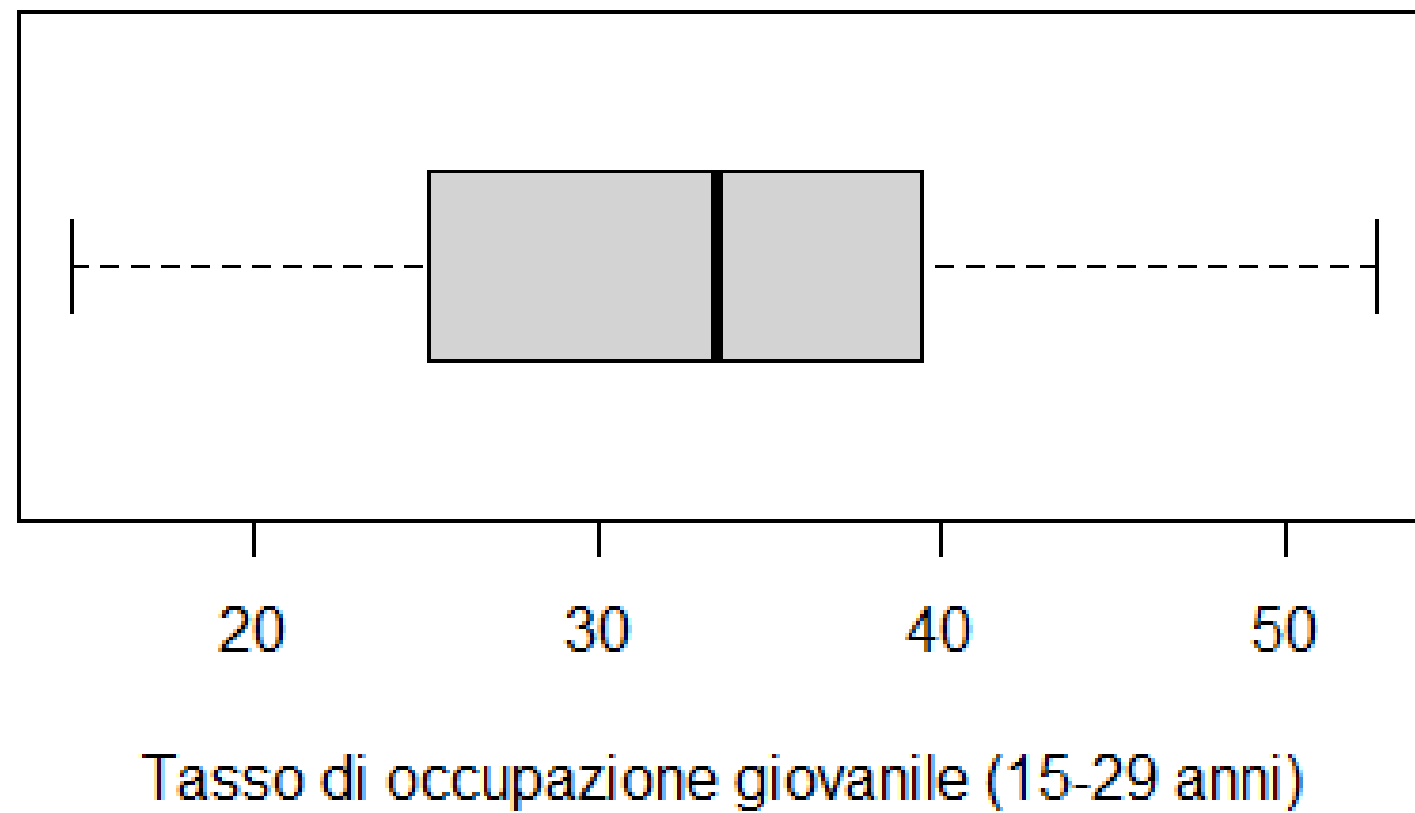
Variabile d'interesse

Tasso di occupazione giovanile (15-29 anni)

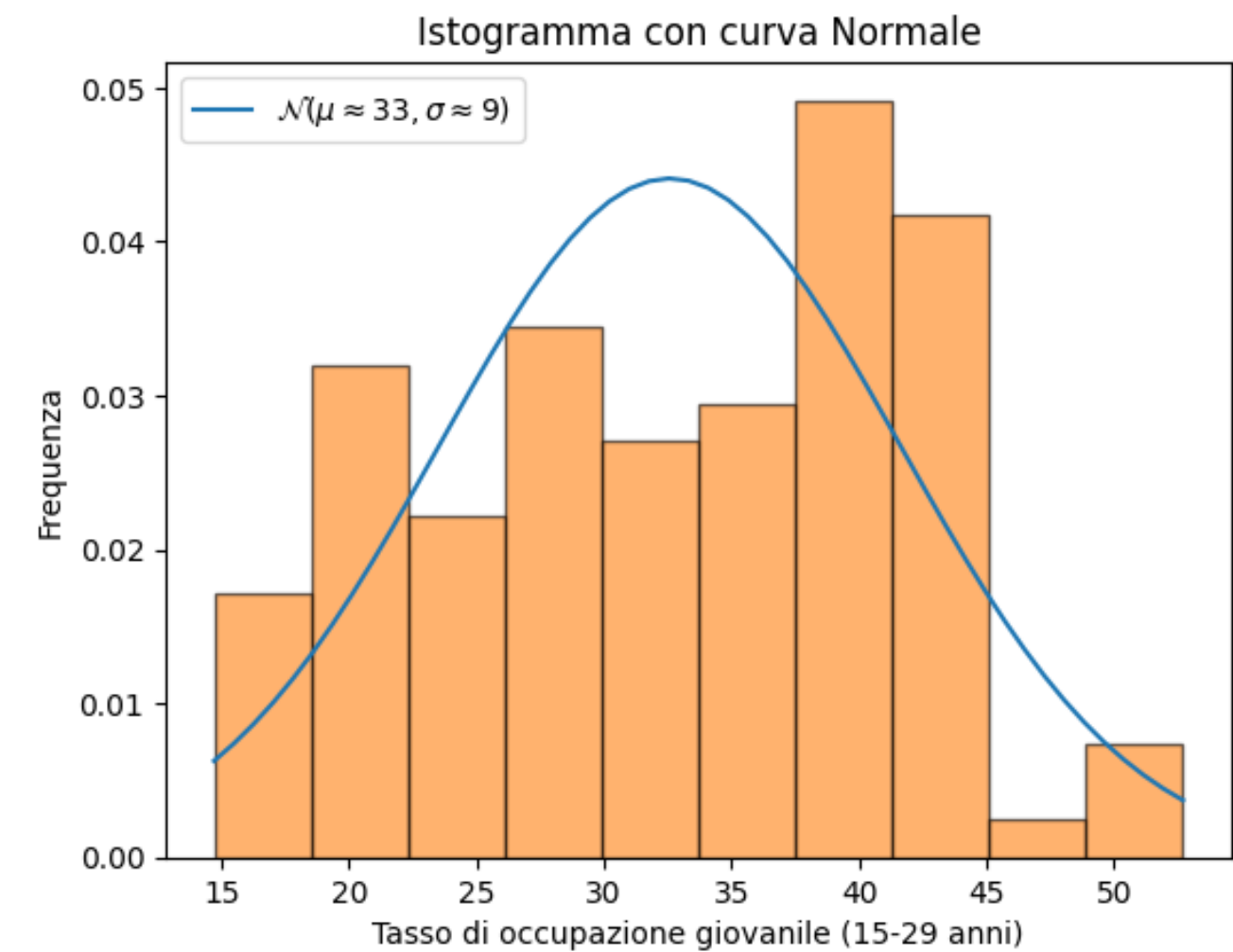
La variabile d'interesse esprime
la percentuale (%) di occupati di 15-29 anni
sulla popolazione totale italiana di 15-29 anni

Analisi Descrittiva

Boxplot



Istogramma



Selezione variabile "competenze"

Risultati modello competenza alfabetica:

OLS Regression Results						
=====						
Dep. Variable:	Q('Tasso di occupazione giovanile (15-29 anni)')	R-squared:	0.468			
Model:	OLS	Adj. R-squared:	0.463			
Method:	Least Squares	F-statistic:	92.38			
Date:	Tue, 12 Dec 2023	Prob (F-statistic):	4.56e-16			
Time:	15:35:28	Log-Likelihood:	-353.18			
No. Observations:	107	AIC:	710.4			
Df Residuals:	105	BIC:	715.7			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
		coef	std err	t	P> t	[0.025 0.975]

Intercept		63.3266	3.263	19.406	0.000	56.856 69.797
Q('Competenza alfabetica non adeguata (studenti classi III scuola secondaria primo grado)')		-0.8778	0.091	-9.612	0.000	-1.059 -0.697

Risultati modello competenza numerica:

OLS Regression Results						
=====						
Dep. Variable:	Q('Tasso di occupazione giovanile (15-29 anni)')	R-squared:	0.669			
Model:	OLS	Adj. R-squared:	0.665			
Method:	Least Squares	F-statistic:	211.8			
Date:	Tue, 12 Dec 2023	Prob (F-statistic):	6.23e-27			
Time:	15:35:28	Log-Likelihood:	-327.86			
No. Observations:	107	AIC:	659.7			
Df Residuals:	105	BIC:	665.1			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
		coef	std err	t	P> t	[0.025 0.975]

Intercept		60.3737	1.976	30.554	0.000	56.456 64.292
Q('Competenza numerica non adeguata (studenti classi III scuola secondaria primo grado)')		-0.7082	0.049	-14.554	0.000	-0.805 -0.612

Stima del modello OLS

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Tasso di occupazione giovanile (15-29 anni)    R-squared:                0.772
Model:                  OLS                                           Adj. R-squared:           0.753
Method:                 Least Squares                                F-statistic:              41.46
Date:                  Tue, 12 Dec 2023                               Prob (F-statistic):       3.65e-28
Time:                  15:35:30                                       Log-Likelihood:           -307.88
No. Observations:      107                                           AIC:                      633.8
Df Residuals:          98                                           BIC:                      657.8
Df Model:              8
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	34.9933	10.879	3.217	0.002	13.405	56.582
Competenza numerica non adeguata (studenti classi III scuola secondaria primo grado)	-0.3011	0.096	-3.120	0.002	-0.493	-0.110
Laureati e altri titoli terziari (25-39 anni)	-0.1126	0.129	-0.871	0.386	-0.369	0.144
Mobilità dei laureati italiani (25-39 anni)	0.1151	0.040	2.862	0.005	0.035	0.195
Organizzazioni non profit	0.0940	0.036	2.641	0.010	0.023	0.165
Partecipazione alla formazione continua	0.0575	0.296	0.194	0.846	-0.529	0.644
Partecipazione elettorale	0.2492	0.071	3.511	0.001	0.108	0.390
Persone con almeno il diploma (25-64 anni)	-0.2029	0.110	-1.843	0.068	-0.421	0.016
Raccolta differenziata dei rifiuti urbani	0.0910	0.042	2.184	0.031	0.008	0.174

```

=====
Omnibus:                1.188    Durbin-Watson:              1.869
Prob(Omnibus):          0.552    Jarque-Bera (JB):          0.762
Skew:                   0.180    Prob(JB):                  0.683
Kurtosis:               3.203    Cond. No.                  3.35e+03
=====

```


Stima del modello Stepwise

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Tasso di occupazione giovanile (15-29 anni)    R-squared:                0.770
Model:                  OLS                                           Adj. R-squared:           0.756
Method:                 Least Squares                                F-statistic:              55.84
Date:                  Tue, 12 Dec 2023                               Prob (F-statistic):       9.42e-30
Time:                  15:35:32                                       Log-Likelihood:           -308.29
No. Observations:      107                                           AIC:                      630.6
Df Residuals:          100                                           BIC:                      649.3
Df Model:              6
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	34.2422	10.775	3.178	0.002	12.865	55.619
Competenza numerica non adeguata (studenti classi III scuola secondaria primo grado)	-0.2898	0.094	-3.070	0.003	-0.477	-0.103
Mobilità dei laureati italiani (25-39 anni)	0.1123	0.035	3.197	0.002	0.043	0.182
Organizzazioni non profit	0.0999	0.034	2.954	0.004	0.033	0.167
Partecipazione elettorale	0.2539	0.070	3.623	0.000	0.115	0.393
Persone con almeno il diploma (25-64 anni)	-0.2497	0.089	-2.799	0.006	-0.427	-0.073
Raccolta differenziata dei rifiuti urbani	0.0910	0.041	2.198	0.030	0.009	0.173

```

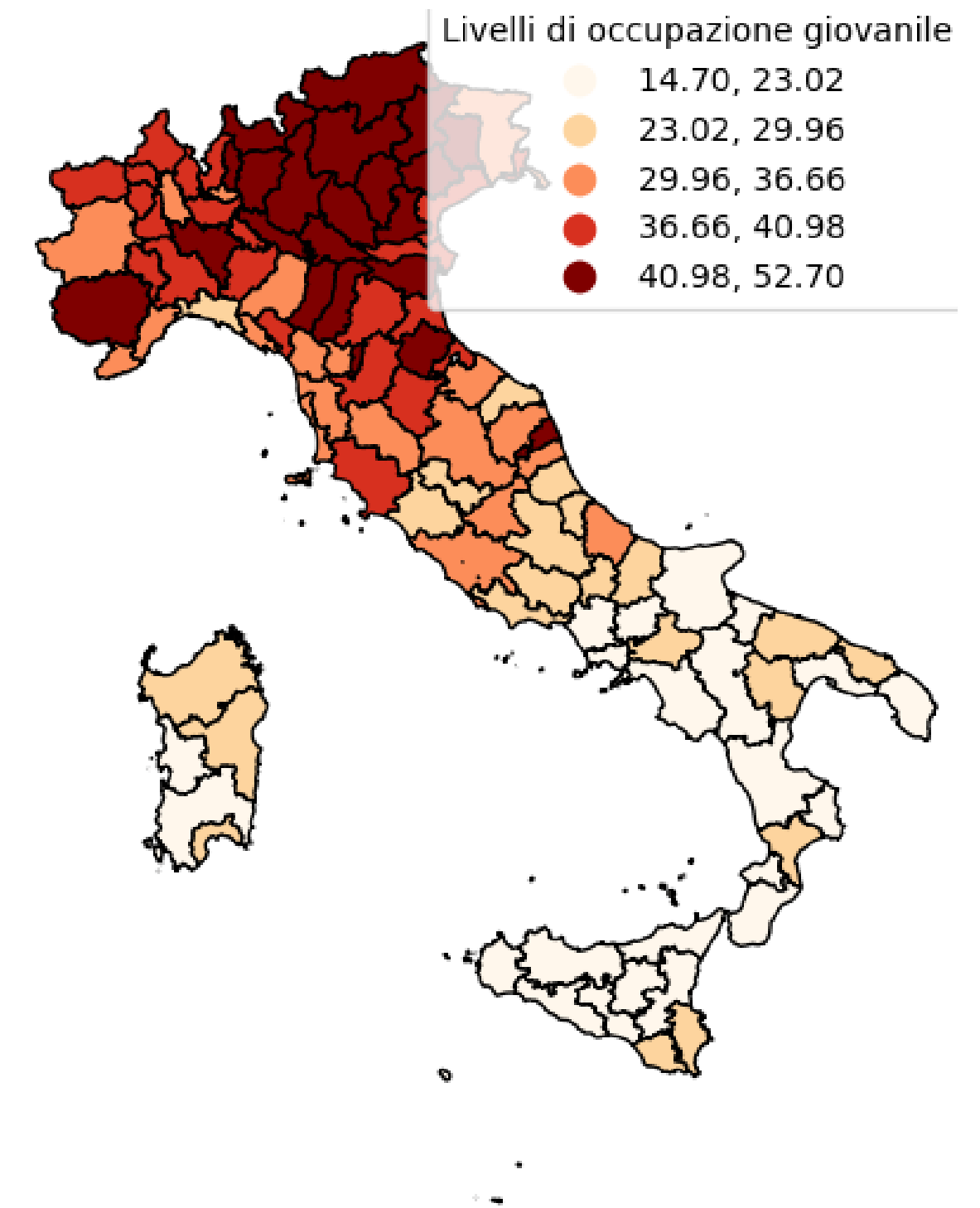
=====
Omnibus:                2.459    Durbin-Watson:              1.888
Prob(Omnibus):          0.292    Jarque-Bera (JB):          1.871
Skew:                   0.279    Prob(JB):                  0.392
Kurtosis:               3.328    Cond. No.                  3.27e+03
=====

```

Analisi Spaziale

Si procede con l'analisi spaziale visualizzando la mappa dell'Italia, suddivisa per province.

In base ai suoi valori, il "Tasso di occupazione giovanile" assume colori diversi per ogni area: notiamo come i colori più intensi si concentrano maggiormente nelle zone del Centro e del Nord-Italia, a sostegno del fatto che in queste zone c'è un Tasso di occupazione giovanile maggiore rispetto al Sud.



Confronto tra criteri

QUEEN



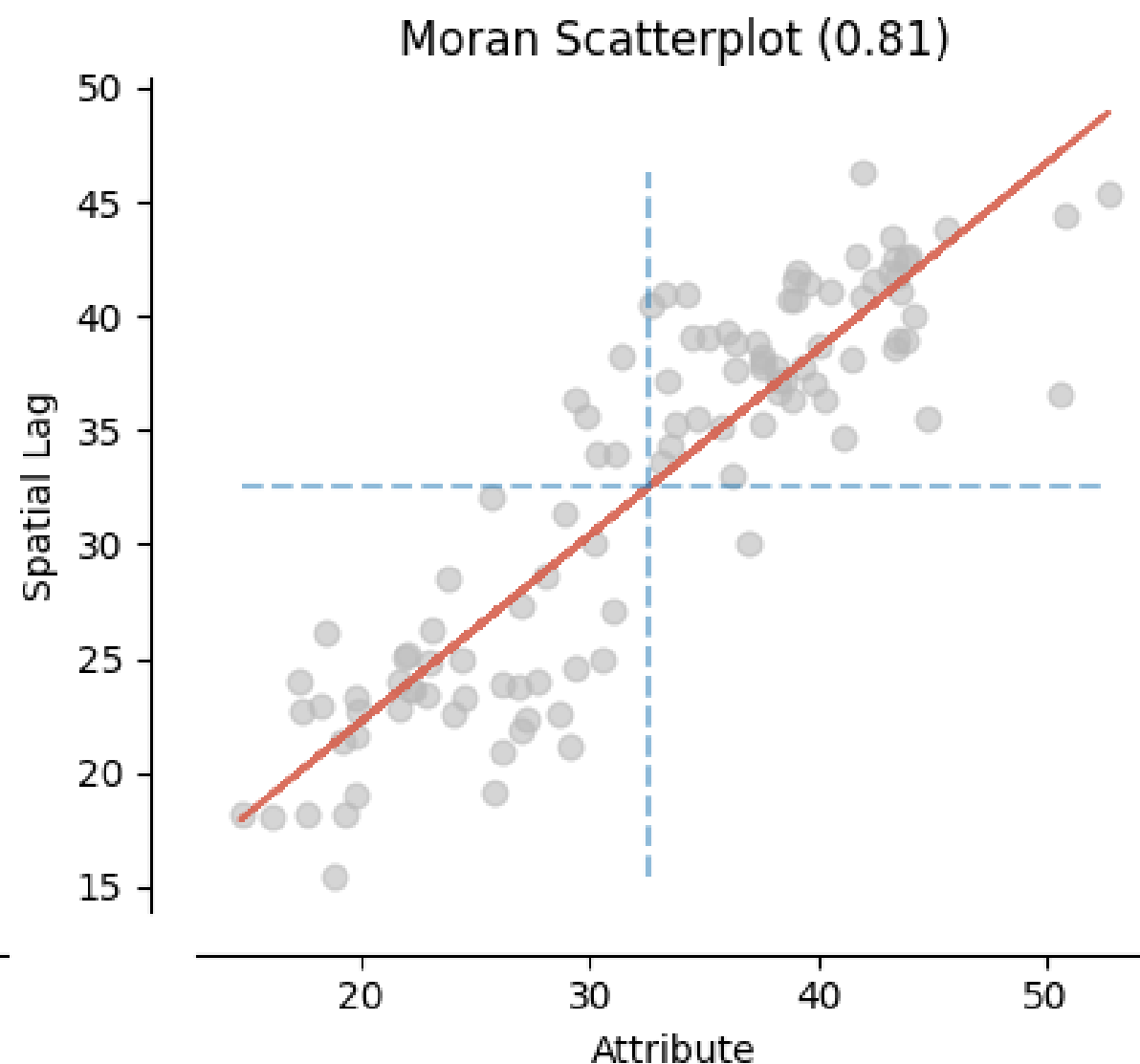
ROOK



DISTANCE



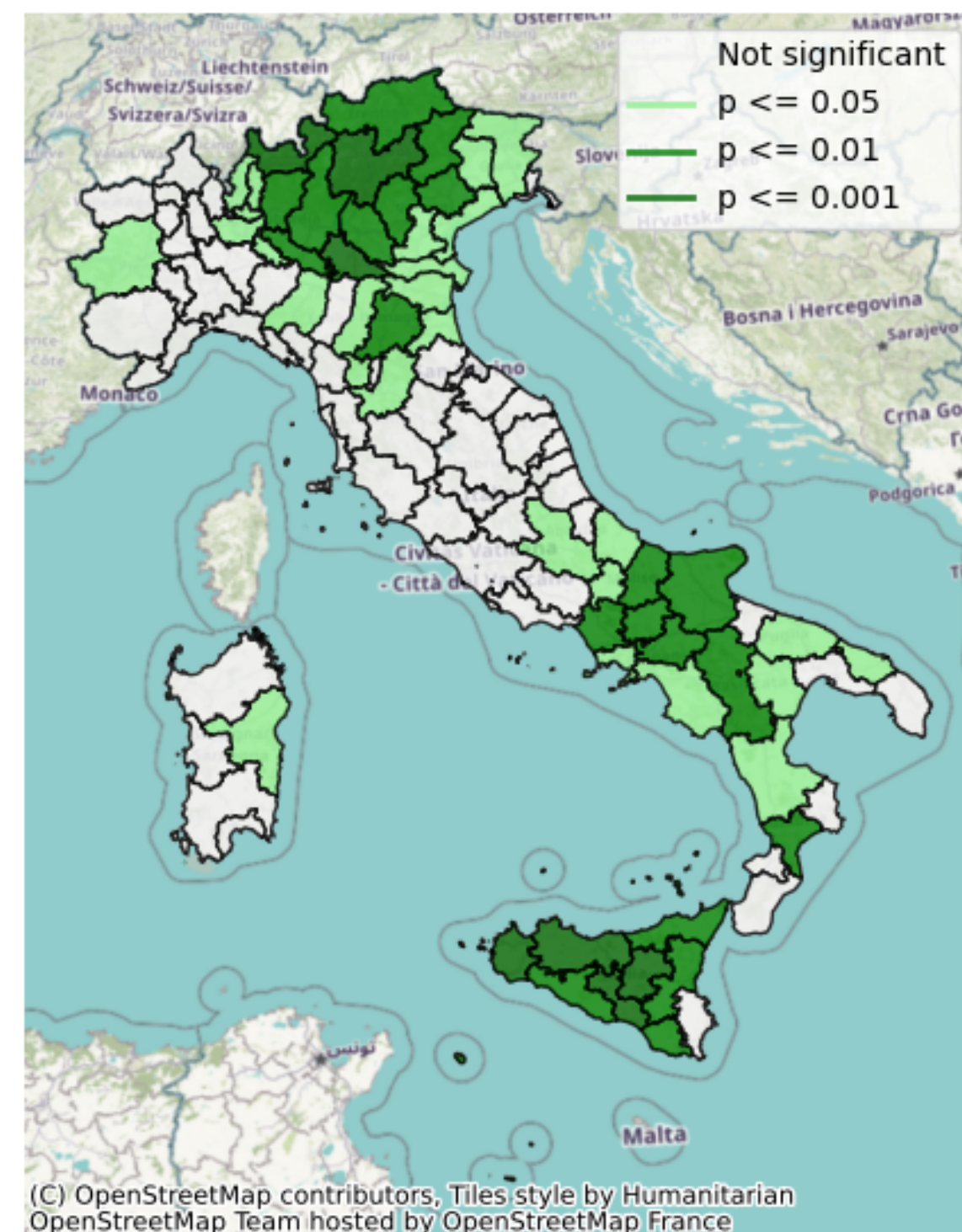
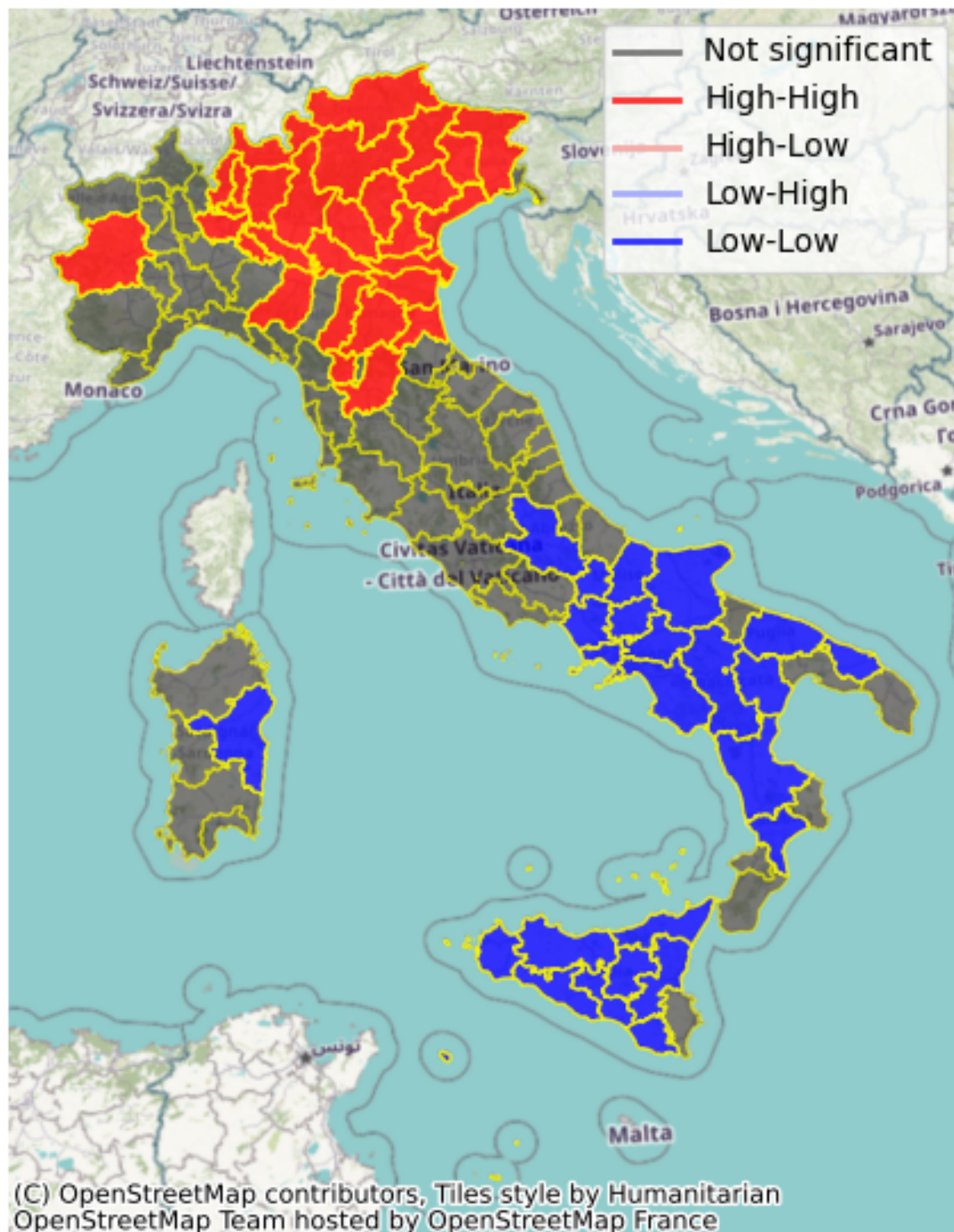
Indici di Autocorrelazione Spaziale



L'Indice I di Moran e l'Indice C di Geary sono utilizzati per valutare l'autocorrelazione spaziale nei dati geografici. L'Indice di Moran misura la presenza di pattern spaziali nei dati e assume valori da -1 a 1, mentre l'Indice di Geary varia tra 0 a 2. Valori rispettivamente diversi da 0 e 1, indicano la presenza di autocorrelazione spaziale. Nel modello considerato, entrambi gli indici rivelano una forte autocorrelazione spaziale.

Moran(I)	p-value(I)	Geary(C)	p-value(C)
0.814599	2.219563e-34	0.169654	0.001

Moran locali



Lagrange Multiplier Tests

LMtest	p-value
LMerr	0.005386
LMlag	1.04e-05
RLMerr	0.2362
RLMlag	0.000296
SARMA	0.02983

Test di ipotesi LMlag

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Test di ipotesi LMerr

$$H_0: \lambda = 0$$

$$H_1: \lambda \neq 0$$

Test di ipotesi LMerr

$$H_0: \rho = 0; \lambda = 0$$

$$H_1: \rho \neq 0; \lambda \neq 0$$

L'Indice di Moran e di Geary colgono la relazione spaziale anche se questa non dovesse essere di tipo lineare.

Gli LMtest, invece, sono utili per iniziare a individuare quali modelli spaziali utilizzare basandosi sul presupposto che la relazione sia di tipo lineare, dato che suddetti modelli derivano dalla regressione lineare multipla, che per definizione coglie una relazione lineare.

Secondo i risultati ottenuti il valore di rho è significativo (anche per il test robusto), mentre lambda lo risulta solo per il test non robusto.

Anche il test SARMA, che valuta congiuntamente rho e lambda, ci restituisce un buon p-value.

Effetti del modello SAR

Dopo aver verificato la correttezza di diversi modelli, il SAR è quello che restituisce i risultati migliori. Di seguito ne vengono mostrati gli effetti.

Variabili	Effetti		
	Diretti	Indiretti	Totali
Competenza numerica non adeguata (studenti classi III scuola secondaria primo grado)	-0.17862599	-0.12808584	-0.3067118
Mobilità dei laureati italiani (25-39 anni)	0.06728913	0.04825045	0.1155396
Organizzazioni non profit	0.06677428	0.04788127	0.1146556
Partecipazione elettorale	0.14593731	0.10464604	0.2505833
Persone con almeno il diploma (25-64 anni)	-0.17799610	-0.12763417	-0.3056303
Raccolta differenziata dei rifiuti urbani	0.07112262	0.05099930	0.1221219

Variabili	p-value		
	Diretti	Indiretti	Totali
Competenza numerica non adeguata (studenti classi III scuola secondaria primo grado)	0.031627	0.058659	0.032479
Mobilità dei laureati italiani (25-39 anni)	0.040421	0.063220	0.037543
Organizzazioni non profit	0.011625	0.033337	0.012290
Partecipazione elettorale	0.038494	0.059678	0.035707
Persone con almeno il diploma (25-64 anni)	0.028608	0.040566	0.024482
Raccolta differenziata dei rifiuti urbani	0.035395	0.088877	0.044006

Conclusioni

- L'approccio spaziale ci ha condotto ad un modello che ha spiegato in maniera significativamente migliore il 'Tasso di occupazione giovanile' per l'anno 2022 all'interno delle province italiane
- Il modello SAR è risultato il *migliore* per studiare l'andamento della variabile d'interesse nell'anno di riferimento
- I risultati forniti dall' utilizzo di Python sono *coerenti* con quelli forniti dall'utilizzo di R



Francesco Palmisano

END SLIDE