



UNIVERSITÀ DEGLI STUDI DELLA BASILICATA  
SCUOLA DI INGEGNERIA

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA INFORMATICA E DELLE TECNOLOGIE  
DELL'INFORMAZIONE

**ANALISI E CONCLUSIONI  
PROGETTO PER IL CORSO DI BIG DATA**

Gruppo 1:

Antonella Bonelli	68791
Marcello Pizzuti	68844
Michael Pio Stolfi	68787
Simona Calocero	68395

ANNO ACCADEMICO 2022/2023

# Sommario

<b>Introduzione .....</b>	<b>1</b>
<b>Analisi dei notebook.....</b>	<b>2</b>
1.1 Notebook "Data Cleaning" .....	2
1.2 Notebook "Data Analytics" .....	3
<b>Analisi visuali .....</b>	<b>6</b>
2.1 Workbook "PrezziConsumo".....	6
2.2 Workbook "PrezziCarburanti".....	6
<b>Risultati e Conclusioni .....</b>	<b>8</b>
3.1 Query relative alle elezioni presidenziali.....	8
3.2 Query relative alla disoccupazione.....	13
3.2 Query relative al reddito pro capite.....	14
3.3 Query relative alla vaccinazione COVID .....	15
3.4 Query relative all'inflazione .....	16
3.5 Query relative al cambiamento di voti tra presidenziale e legislative.....	17
3.6 Risultati relativi alle correlazioni.....	18
3.7 Risultati relativi alle analisi visuali .....	30
3.8 Conclusioni .....	32

# Introduzione

La seguente relazione descrive il codice e le procedure utilizzate per svolgere il progetto di analisi per il corso di Big Data che si propone individuare possibili motivazioni alla base delle proteste verificatesi in Francia, in seguito all'approvazione della legge che ha aumentato l'età pensionabile da 62 a 64 anni. Sono stati analizzati dataset riguardanti: le elezioni presidenziali e legislative del 2022, l'andamento del consumo e del prezzo del carburante, tasso di vaccinazione covid e tasso di disoccupazione al fine di individuare possibili correlazioni e fornire risultati.

Le tecnologie utilizzate per lo sviluppo delle query richieste. Per l'elaborazione e l'estrazione dei dati dai dataset è stato impiegato Python insieme a librerie specializzate nella manipolazione e analisi di dati come Pandas. Jupyter Notebook è un ambiente interattivo grazie al quale è stato possibile sviluppare e condividere il codice. Lo strumento scelto per l'analisi visuale, Tableau, che consente la creazione di grafici e dashboard per rappresentare in modo chiaro ed intuitivo i dati.

Nel Capitolo 1, si tratteranno nello specifico l'analisi dei dati attraverso notebook dedicati. Il notebook "Data Cleaning", che si concentra sul processo di pulizia e preparazione dei dati per l'analisi successiva, e il notebook "Data Analytics", che approfondisce l'analisi dei dati e l'estrazione di informazioni significative.

Nel Capitolo 2, si esplorerà l'aspetto visuale dell'analisi dei dati attraverso due workbook, sviluppati con Tableau: "PrezziConsumo" e "PrezziCarburanti". A partire dall'analisi preliminare, questi workbook facilitano la comprensione delle tendenze e dei modelli presenti nei dati.

Infine, nel Capitolo 3, verranno discusse le conclusioni derivanti dalle analisi svolte. Saranno esposte le possibili ragioni alla base delle proteste, utilizzando le informazioni e i risultati ottenuti dalle fasi precedenti.

# Capitolo 1

## Analisi dei notebook

Per effettuare l'analisi dei dati si è utilizzato un Jupyter Notebook. In realtà, data la grande quantità di dati, e la assoluta necessità di migliorarne la qualità in vista delle elaborazioni successive, sono stati creati due Jupyter Notebook:

- 1) Il notebook denominato "Data Cleaning" si occupa di caricare in maniera opportuna i dati dalle sorgenti (in questo caso i file csv) in DataFrame Pandas, effettuare una serie di operazioni atte a migliorare la qualità dei DataFrame (tali operazioni non sono standard e cambiano da DataFrame a DataFrame) e infine salvare tali DataFrame in nuovi csv rinominati con il prefisso "work";
- 2) Il notebook denominato "Data Analytics" si occupa di caricare in DataFrame Pandas i dati elaborati dal notebook precedente e scritti nei csv "work", e con questi DataFrame vengono effettuate prima le query e poi le correlazioni descritte nella specifica di progetto.

### 1.1 Notebook "Data Cleaning"

Come descritto sopra, il Notebook "Data Cleaning" si occupa di migliorare la qualità dei dati; per ottenere questo risultato una volta caricati i file csv in memoria e salvati in DataFrame Pandas essi subiscono una serie di elaborazioni che passano anche attraverso delle funzioni ausiliarie.

La prima funzione è la seguente:

```
filter_dataset_by_regions(file_path_dataset, sep=';', dtype={})
```

tale funzione si occupa di caricare i dati dai file csv e filtrarli per le regioni assegnate al presente gruppo.

La seconda funzione è la seguente:

```
float_converter(s)
```

quest'ultima è una semplice funzione pensata per essere utilizzata con il metodo apply() di Pandas, e converte ogni valore della Series su cui è applicata da una stringa contenente un valore decimale, dove il separatore della parte decimale è la virgola, a una stringa contenente un valore decimale, dove il separatore della parte decimale è il punto. Questo perché i dataset di lavoro sono di origine francese e in Europa il separatore decimale è la virgola, mentre per Pandas affinché una stringa possa essere convertita in un float deve avere come separatore il punto.

La terza funzione è la seguente:

```
thousands_converter(s)
```

anche quest'ultima è una semplice funzione pensata per essere utilizzata con il metodo apply() di Pandas, e converte ogni valore della Series su cui è applicata da una stringa contenente un valore avente come separatore delle migliaia la virgola, in una stringa contenente un valore senza alcun separatore. Questo perché nei dataset di lavoro è presente la virgola come separatore delle migliaia, da convenzione francese.

La quarta funzione è la seguente:

```
substring_datetime(s)
```

anche quest'ultima è una semplice funzione pensata per essere utilizzata con il metodo `apply()` di Pandas (o `map()`), e separa in ogni valore di tipo `datetime` della Series su cui è applicata la parte della data dalla parte dell'orario e conserva solo la prima.

Infine, la quinta funzione è la seguente:

```
price_normalizer(n)
```

anche quest'ultima è una semplice funzione pensata per essere utilizzata con il metodo `apply()` di Pandas, il suo scopo è appunto normalizzare i prezzi del dataset dei carburanti. Dopo la fase di Analisi Visuale si è infatti notato che quest'ultimo aveva molti dati il cui valore effettivo era corrotto, ovvero il punto era stato mal posizionato, e questa funzione risolve questo problema permettendo di salvare molti dati che altrimenti sarebbero stati scartati.

Funzioni ausiliarie a parte ogni DataFrame contenente dati provenienti da file csv viene sottoposto a una serie di elaborazioni e migliorie, queste come detto cambiano da DataFrame a DataFrame. Le operazioni più comuni sono: il `rename` delle colonne, necessario per rinominare le colonne e tradurle dal francese all'italiano; il `drop` delle colonne inutili; e il `reindex` delle colonne, utile a riordinarle in modo che si presentino nelle prime posizioni quelle di maggior interesse.

Per quanto concerne il dataset relativo al reddito pro capite esso è stato creato ad-hoc a partire dai dati relativi al reddito per dipartimenti esposti sul quotidiano *journaldunet*. In particolare, il dataset si compone di tre colonne, ossia: regione del dipartimento, nome del dipartimento, reddito pro capite del dipartimento; e durante il caricamento il dataset viene elaborato esattamente allo stesso modo in cui vengono elaborati gli altri dataset.

Per quanto riguarda invece il dataset relativo all'inflazione, è stato anch'esso generato appositamente a partire dalla cartina interattiva messa a disposizione dal sito web *francetvinfo.fr*. Il dataset contiene soltanto le due colonne di interesse, ovvero nome del dipartimento e indice dei prezzi al consumo.

## 1.2 Notebook "Data Analytics"

Come accennato in precedenza il Notebook "Data Analytics" si occupa di rispondere alle richieste della specifica sia in termini di query sia in termini di correlazioni; per ottenere questo risultato una volta caricati i file csv lavorati (ovvero quelli con il prefisso "work", che sono il risultato delle elaborazioni del notebook precedente) in memoria e salvati in DataFrame Pandas essi subiscono una serie di elaborazioni che passano anche attraverso delle funzioni ausiliarie che saranno descritte qui di seguito, per poi passare alla descrizione delle query e, in ultima analisi, a quella delle correlazioni.

La prima funzione è la seguente:

```
percentage_calculator(row)
```

quest'ultima è una semplice funzione pensata per essere utilizzata con il metodo `apply()` di Pandas, ed è specifica per i DataFrame delle elezioni presidenziali e legislative ove vi sono

le colonne "espressi" e "voti". Essa ragiona per righe: se esiste almeno un voto espresso allora ritorna la percentuale dei voti a favore sui voti espressi dividendo appunto il valore della colonna "voti" e della colonna "espressi" per la corrente riga.

Vi sono poi le funzioni utilizzate nel calcolo delle query, la prima funzione è la seguente:

```
get_departments_min_max_votes_perc(df_dataset, surname=None,  
min=False)
```

quest'ultima si occupa di calcolare e stampare i venti dipartimenti con la maggiore o la minore percentuale di voti per un candidato specificato.

La seconda funzione è la seguente:

```
presidential_legislative_difference(candidates, party)
```

quest'ultima si occupa di calcolare la differenza in valore assoluto tra le percentuali dei voti espressi per uno o più candidati alle presidenziali e i voti espressi per il relativo partito alle legislative e infine stampa i dieci dipartimenti con maggiore differenza di voti.

Vi sono poi le funzioni utilizzate nel calcolo delle correlazioni, la prima funzione è la seguente:

```
get_correlation(df, value, choice_y, x_label, y_label,  
is_legislative=False, labels_departments=False)
```

quest'ultima si occupa di effettuare tutta una serie di operazioni preliminari e ripetitive sui DataFrame per poi estrapolare da questi le Series da mandare in input alla funzione successiva che effettua il vero e proprio calcolo della correlazione.

La seconda funzione è la seguente:

```
correlation_calculator(x, y, x_label, y_label,  
labels_departments=False)
```

quest'ultima si occupa di calcolare i coefficienti di correlazione per i vettori di dati inseriti, e disegnare il grafico dei dati e della loro regressione lineare.

È importante notare che questa estrema modularizzazione ha permesso il riuso di molto codice nella esecuzione delle query. Le prime cinque query sono state espletate chiamando opportunamente la funzione `get_departments_min_max_votes_perc`. La query numero quattordici è stata espletata chiamando opportunamente la funzione `presidential_legislative_difference`. Le altre query intermedie effettuano operazioni simili ma su dataset diversi, le operazioni tipiche sono: groupby, filtriaggi per data, ordinamento dei valori di alcune colonne, e slicing dei DataFrame.

Per quanto riguarda le correlazioni invece, a parte la funzione `get_correlation` che mette a fattor comune moltissimo codice e riduce il calcolo delle correlazioni a una chiamata, il cuore del calcolo risiede nella funzione `correlation_calculator`.

Quest'ultima funzione disegna il grafico dei dati e della loro regressione lineare grazie a Matplotlib e Scipy; calcola grazie a Pandas il coefficiente di correlazione di Pearson r, di Spearman p, e di Kendall τ; infine calcola grazie a Scikit-learn il coefficiente di

determinazione  $R^2$ . Per ogni correlazione in base ai dati restituiti dalla funzione precedente è stata poi effettuata una considerazione finale.

Il coefficiente di correlazione  $r$  di Pearson è quello che risponde a regole più restrittive. Si può usare solo se la relazione è lineare e se non ci sono outliers. Quando tutte queste ipotesi sono verificate e le due variabili sono entrambe quantitative, va bene considerare Pearson. Se invece almeno una di queste ipotesi non è verificata allora per ottenere risultati validi dall'analisi della correlazione è necessario utilizzare un test non parametrico. I test non parametrici, infatti, sono meno restrittivi di quelli parametrici, non richiedono che la relazione sia lineare, inoltre sono più robusti rispetto agli outliers. Si è infatti osservata la quantità degli outliers e se la relazione fosse lineare o meno. Per verificare quest'ultima proprietà si è tenuto conto del coefficiente di determinazione stesso, infatti, fornisce informazioni sulla bontà di adattamento del modello (che nel nostro caso era lineare) ai dati. Nel caso di presenza di molti outliers o di un coefficiente di determinazione molto piccolo (vicino a zero) si è tenuto in considerazione il coefficiente di correlazione di Spearman. In ogni caso che si sia preso in considerazione Pearson o Spearman si è valutata la forza della correlazione confrontando il valore assoluto del dato ottenuto con uno degli intervalli di Evans. Così si è potuto valutare sempre in maniera robusta verso e intensità della correlazione tra i dati.

# **Capitolo 2**

## **Analisi visuali**

In questo capitolo analizzeremo le metodologie utilizzate per la generazione dei grafici relativi agli andamenti dei prezzi.

### **2.1 Workbook “PrezziConsumo”**

Il primo workbook prodotto riguarda l’andamento dei prezzi al consumo per panieri di diverso genere nel periodo compreso tra a Aprile 2022 e Febbraio 2023.

Sulle ascisse si ha un asse temporale a cadenza mensile, mentre sulle ordinate si trovano i valori relativi all’indice dei prezzi al consumo a cui è stata applicata la funzione di aggregazione media in modo da ottenere per ogni mese un valore indicativo per tutti i prezzi delle diverse classi di consumo.

Per la rappresentazione si è optato per un grafico a linea, una scelta naturale per la visualizzazione di un andamento temporale, in cui sono stati evidenziati i punti di interpolazione mostrandoli e annotandoli con i relativi valori. Il valore corrispondente al mese di Gennaio 2023, mese di inizio delle proteste, è stato messo in risalto applicando un colore diverso. Per migliorare la percezione dell’andamento complessivo è stata aggiunta una linea in corrispondenza del valore medio.

A seguito della generazione del grafico è stato osservato un andamento discendente anomalo nel mese di Marzo 2023, che si è scoperto essere dovuto al fatto che la quantità di dati presenti nel dataset relativamente a tale mese era notevolmente inferiore rispetto ai dati riguardanti gli altri mesi, per cui si è deciso di ignorare tali dati e non riportarli nel grafico.

### **2.2 Workbook “PrezziCarburanti”**

Il secondo workbook riguarda invece l’andamento dei prezzi dei diversi tipi di carburante, in particolare diesel, SP98, SP95, GPL, E10 e E85, nel periodo che va dal 1° aprile 2022 al 14 Aprile 2023. Il workbook contiene tre worksheet relativi rispettivamente ad andamento giornaliero, andamento a media mobile a 30 giorni e andamento a media mobile a 60 giorni.

In tutti e tre i worksheet è presente un asse delle ascisse comune su cui il tempo è indicato con cadenza giornaliera, mentre sulle ordinate sono raffigurati separatamente i valori relativi al prezzo di ognuno dei carburanti sopra citati, filtrati ad escludere i valori nulli.

Anche in questo caso è stato scelto un grafico a linee, preferendo però mostrare le annotazioni soltanto in corrispondenza di massimi e minimi, per evitare sovrapposizioni e conseguente perdita di leggibilità, data l’interpolazione molto fitta. Coerentemente con quanto fatto nel primo workbook si è scelto di evidenziare i giorni di Gennaio 2023 marcandoli con un colore differente.

Nel primo worksheet ai valori dei prezzi è stata applicata la funzione di aggregazione media per tener conto dei prezzi giornalieri delle varie stazioni di servizio.

Nel secondo e nel terzo worksheet è stata invece applicata una funzione di aggregazione di media mobile sui valori dei prezzi, calcolata nel primo caso basandosi per ogni punto sui 30 valori precedenti, ovvero sui 30 giorni precedenti, e nel secondo caso basandosi sui precedenti 60 valori.

I grafici relativi ai diversi tipi di carburante vengono mostrati separatamente in modo da mettere in risalto un confronto sui loro andamenti complessivi anziché un confronto tra i loro effettivi costi che si sarebbe percepito sovrapponendoli su un unico grafico.

# Capitolo 3

## Risultati e Conclusioni

In questo capitolo verranno riportati i risultati relativi ai calcoli delle query e delle correlazioni, nonché i grafici generati. Infine, saranno esposte le conclusioni.

### 3.1 Query relative alle elezioni presidenziali

I 20 dipartimenti con la maggiore percentuale di voti per Macron sono:

Tabella 1 - Primo Turno

	Dipartimento	% Voti a favore
1	Hauts-de-Seine	37,11
2	Paris	35,34
3	Ille-et-Vilaine	34,50
4	Yvelines	33,41
5	Morbihan	32,61
6	Manche	32,57
7	Finistère	32,21
8	Calvados	31,16
9	Côtes-d'Armor	31,02
10	Indre-et-Loire	30,99
11	Bas-Rhin	30,70
12	Orne	30,48
13	Val-de-Marne	29,10
14	Loiret	28,92
15	Marne	28,63
16	Côte-d'Or	28,55
17	Seine-Maritime	27,95
18	Loir-et-Cher	27,92
19	Haut-Rhin	27,85
20	Somme	27,80

Tabella 2 - Secondo Turno

	Dipartimento	% Voti a favore
1	Paris	85,11
2	Hauts-de-Seine	80,39
3	Val-de-Marne	74,48
4	Seine-Saint-Denis	73,72
5	Yvelines	71,05
6	Ille-et-Vilaine	70,94
7	Finistère	67,50
8	Val-d'Oise	66,15
9	Essonne	65,43
10	Côtes-d'Armor	62,90
11	Morbihan	62,81
12	Indre-et-Loire	62,72
13	Calvados	60,29
14	Manche	59,61
15	Bas-Rhin	59,00
16	Loiret	57,57
17	Côte-d'Or	57,27
18	Doubs	57,16
19	Seine-et-Marne	56,97
20	Seine-Maritime	55,26

I 20 dipartimenti con la minore percentuale di voti per Macron sono:

Tabella 3 - Primo Turno

	Dipartimento	% Voti a favore
27	Nord	26,37
28	Val-d'Oise	26,09
29	Eure	26,05
30	Indre	26,03
31	Moselle	26,01
32	Aube	25,59
33	Nièvre	25,51
34	Vosges	25,25
35	Seine-et-Marne	25,00
36	Meuse	24,91
37	Jura	24,88
38	Pas-de-Calais	24,61
39	Oise	24,30
40	Yonne	24,14
41	Territoire de Belfort	24,07
42	Ardennes	23,64
43	Haute-Marne	23,33
44	Haute-Saône	22,42
45	Aisne	22,09
46	Seine-Saint-Denis	20,27

Tabella 4 - Secondo Turno

	Dipartimento	% Voti a favore
27	Haut-Rhin	52,90
28	Nord	52,83
29	Cher	52,56
30	Marne	52,09
31	Territoire de Belfort	51,44
32	Indre	51,40
33	Moselle	50,46
34	Nièvre	49,89
35	Somme	49,00
36	Eure	48,62
37	Yonne	48,41
38	Aube	48,29
39	Vosges	47,58
40	Oise	47,27
41	Meuse	44,40
42	Ardennes	43,33
43	Haute-Saône	43,12
44	Haute-Marne	43,04
45	Pas-de-Calais	42,50
46	Aisne	40,09

I 20 dipartimenti con la maggiore percentuale di voti per Le Pen sono:

*Tabella 5 - Primo Turno*

	Dipartimento	% Voti a favore
1	Aisne	39,25
2	Pas-de-Calais	38,68
3	Haute-Marne	36,60
4	Ardennes	36,02
5	Meuse	35,11
6	Haute-Saône	34,60
7	Aube	32,97
8	Somme	32,79
9	Eure	32,29
10	Oise	32,27
11	Vosges	32,23
12	Yonne	31,25
13	Marne	30,57
14	Moselle	30,37
15	Nord	29,27
16	Nièvre	29,20
17	Indre	28,53
18	Eure-et-Loir	28,16
19	Cher	27,89
20	Loir-et-Cher	27,77

*Tabella 6 - Secondo Turno*

	Dipartimento	% Voti a favore
1	Aisne	59,91
2	Pas-de-Calais	57,50
3	Haute-Marne	56,96
4	Haute-Saône	56,88
5	Ardennes	56,67
6	Meuse	55,60
7	Oise	52,73
8	Vosges	52,42
9	Aube	51,71
10	Yonne	51,59
11	Eure	51,38
12	Somme	51,00
13	Nièvre	50,11
14	Moselle	49,54
15	Indre	48,60
16	Territoire de Belfort	48,56
17	Marne	47,91
18	Cher	47,44
19	Nord	47,17
20	Haut-Rhin	47,10

I 20 dipartimenti con la minore percentuale di voti per Le Pen sono:

*Tabella 7 - Primo Turno*

	Dipartimento	% Voti a favore
27	Jura	26,29
28	Loiret	25,59
29	Bas-Rhin	25,30
30	Manche	25,53
31	Côte-d'Or	24,44
32	Doubs	24,07
33	Calvados	23,76
34	Seine-et-Marne	23,56
35	Morbihan	22,00
36	Côtes-d'Armor	21,79
37	Indre-et-Loire	21,54
38	Finistère	18,58
39	Essonne	17,79
40	Val-d'Oise	17,20
41	Ille-et-Vilaine	17,05
42	Yvelines	13,77
43	Seine-Saint-Denis	11,88
44	Val-de-Marne	11,82
45	Hauts-de-Seine	08,37
46	Paris	05,54

*Tabella 8 - Secondo Turno*

	Dipartimento	% Voti a favore
27	Seine-Maritime	44,74
28	Seine-et-Marne	43,03
29	Doubs	42,84
30	Côte-d'Or	42,73
31	Loiret	42,43
32	Bas-Rhin	41,00
33	Manche	40,39
34	Calvados	39,71
35	Indre-et-Loire	37,29
36	Morbihan	37,19
37	Côtes-d'Armor	37,10
38	Essonne	34,57
39	Val-d'Oise	33,85
40	Finistère	32,50
41	Ille-et-Vilaine	29,06
42	Yvelines	28,95
43	Seine-Saint-Denis	26,28
44	Val-de-Marne	25,52
45	Hauts-de-Seine	19,61
46	Paris	14,89

I 20 dipartimenti con la maggiore percentuale di voti per Mélenchon sono:

Tabella 9 - Primo Turno

	Dipartimento	% Voti a favore
1	Seine-Saint-Denis	49,09
2	Val-d'Oise	33,17
3	Val-de-Marne	32,67
4	Paris	30,08
5	Essonne	28,12
6	Seine-et-Marne	25,87
7	Haute-de-Seine	25,77
8	Yvelines	22,89
9	Ille-et-Vilaine	22,20
10	Nord	21,95
11	Finistère	21,47
12	Seine-Maritime	21,17
13	Territoire de Belfort	20,96
14	Meurthe-et-Moselle	20,89
15	Indre-et-Loire	20,77
16	Côtes-d'Armor	20,26
17	Doubs	20,06
18	Jura	19,89
19	Oise	19,30
20	Calvados	19,15

### 3.2 Query relative alla disoccupazione

I 20 dipartimenti con la maggiore e la minore disoccupazione sono:

Tabella 10 - Maggiore Disoccupazione

	Dipartimento	Disoccupati
1	Nord	742160
2	Paris	579890
3	Seine-Saint-Denis	510960
4	Pas-de-Calais	366630
5	Haute-de-Seine	345410
6	Val-de-Marne	325230
7	Seine-et-Marne	322530
8	Seine-Maritime	315880
9	Val-d'Oise	310620
10	Yvelines	305330
11	Essonne	275490
12	Bas-Rhin	256170
13	Ille-et-Vilaine	247890
14	Moselle	244840
15	Finistère	210390
16	Oise	197550
17	Haut-Rhin	175520
18	Loiret	170820
19	Morbihan	169330
20	Calvados	160110

Tabella 11 - Minore Disoccupazione

	Dipartimento	Disoccupati
46	Territoire de Belfort	36570
45	Haute-Marne	36650
44	Nièvre	38980
43	Meuse	40430
42	Indre	48670
41	Haute-Saône	50780
40	Jura	52010
39	Orne	57800
38	Cher	70040
37	Loir-et-Cher	72960
36	Ardennes	73880
35	Yonne	74130
34	Aube	86660
33	Vosges	87300
32	Manche	91570
31	Eure-et-Loir	99370
30	Côte-d'Or	115970
29	Doubs	123800
28	Saône-et-Loire	125840
27	Marne	138120

### 3.2 Query relative al reddito pro capite

I 20 dipartimenti con il maggiore e il minore reddito pro capite sono:

Tabella 12 - Maggiore Reddito

	Dipartimento	€ Reddito
1	Paris	47695
2	Hauts-de-Seine	45141
3	Yvelines	40347
4	Essonne	30974
5	Val-de-Marne	30973
6	Haut-Rhin	30716
7	Seine-et-Marne	29278
8	Bas-Rhin	28421
9	Doubs	28238
10	Val-d'Oise	28184
11	Ille-et-Vilaine	28102
12	Côte-d'Or	27964
13	Marne	27283
14	Oise	27056
15	Territoire de Belfort	26971
16	Loiret	26782
17	Eure-et-Loir	26608
18	Indre-et-Loire	26597
19	Morbihan	26395
20	Calvados	26201

Tabella 13 - Minore Reddito

	Dipartimento	€ Reddito
46	Seine-Saint-Denis	21565
45	Nièvre	22091
44	Indre	22186
43	Ardennes	22227
42	Aisne	22513
41	Pas-de-Calais	22572
40	Haute-Marne	22705
39	Orne	22804
38	Vosges	23398
37	Meuse	23490
36	Haute-Saône	23563
35	Cher	23745
34	Somme	23772
33	Yonne	23883
32	Aube	24388
31	Nord	24628
30	Saône-et-Loire	24868
29	Manche	25041
28	Côtes-d'Armor	25144
27	Loir-et-Cher	25275

### 3.3 Query relative alla vaccinazione COVID

I 20 dipartimenti con la maggiore e la minore percentuale di vaccinazioni COVID sono:

Tabella 14 - Maggiore percentuale vaccinazione

	Dipartimento	% Copertura
1	Manche	85,5
2	Côtes-d'Armor	85,0
3	Finistère	84,1
4	Morbihan	83,9
5	Calvados	83,1
6	Loir-et-Cher	82,9
7	Paris	82,5
8	Nièvre	82,3
9	Orne	82,3
10	Indre	82,2
11	Pas-de-Calais	81,5
12	Saône-et-Loire	81,2
13	Indre-et-Loire	81,2
14	Somme	81,1
15	Meuse	81,1
16	Cher	81,0
17	Haute-Marne	80,7
18	Seine-Maritime	80,6
19	Vosges	80,5
20	Eure	80,3

Tabella 15 - Minore percentuale vaccinazione

	Dipartimento	% Copertura
46	Seine-Saint-Denis	65,9
45	Val-d'Oise	71,5
44	Val-de-Marne	72,6
43	Haut-Rhin	73,1
42	Doubs	73,3
41	Seine-et-Marne	74,1
40	Essonne	74,3
39	Territoire de Belfort	75,5
38	Oise	75,7
37	Aube	75,9
36	Nord	77,0
35	Moselle	77,1
34	Meurthe-et-Moselle	77,2
33	Hauts-de-Seine	77,2
32	Jura	77,3
31	Bas-Rhin	77,7
30	Loiret	77,8
29	Yvelines	77,9
28	Côte-d'Or	78,2
27	Marne	78,3

### 3.4 Query relative all'inflazione

I 10 dipartimenti con la maggiore e la minore inflazione sono:

Tabella 16 - Maggiore Inflazione

	Dipartimento	Indice Prezzo
1	Paris	120,0
2	Hauts-de-Seine	113,3
3	Val-de-Marne	110,0
4	Seine-Saint-Denis	108,1
5	Yvelines	105,5
6	Yonne	104,6
7	Val-d'Oise	103,5
8	Essonne	103,4
9	Ardennes	102,0
10	Seine-et-Marne	102,0

Tabella 17 - Minore Inflazione

	Dipartimento	Indice Prezzo
37	Meuse	97,8
38	Indre-et-Loire	97,7
39	Finistère	97,7
40	Eure-et-Loir	97,4
41	Territoire de Belfort	97,4
42	Orne	97,3
43	Côtes-d'Armor	97,2
44	Loir-et-Cher	97,0
45	Haut-Rhin	96,9
46	Ille-et-Vilaine	96,4

### 3.5 Query relative al cambiamento di voti tra presidenziale e legislative

I 10 dipartimenti con il maggior cambiamento di voti tra il primo turno delle presidenziali e il primo turno delle legislative sono:

*Tabella 18 - Macron - ENS*

	Dipartimento	% Differenza Voti
1	Orne	13,6
2	Ardennes	9,2
3	Meuse	9,0
4	Vosges	7,5
5	Haute-Marne	6,5
6	Somme	5,3
7	Aube	5,0
8	Territoire de Belfort	4,6
9	Oise	4,3
10	Marne	4,3

*Tabella 19 - Le Pen - RN*

	Dipartimento	% Differenza Voti
1	Ardennes	12,25
2	Vosges	8,9
3	Côtes-d'Armor	7,8
4	Territoire de Belfort	7,1
5	Manche	6,9
6	Haut-Rhin	6,9
7	Orne	6,9
8	Bas-Rhin	6,7
9	Aisne	6,1
10	Morbihan	6,1

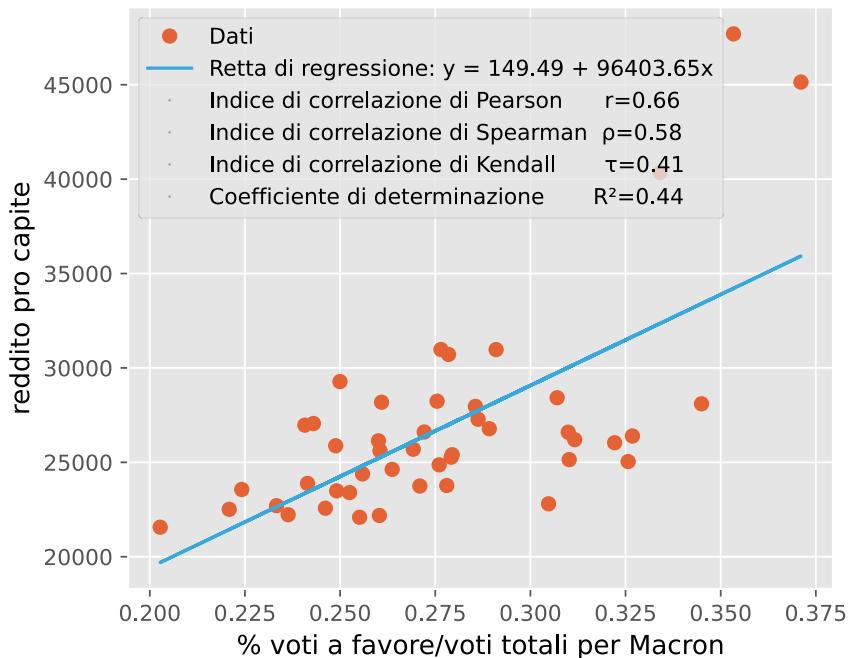
*Tabella 20 - Melenchon – Nupes*

	Dipartimento	% Differenza Voti
1	Seine-Saint-Denis	43,7
2	Val-de-Marne	33,5
3	Paris	32,8
4	Essonne	29,4
5	Val-d'Oise	28,7
6	Ille-et-Vilaine	28,4
7	Seine-Maritime	26,8
8	Finistère	26,5
9	Seine-et-Marne	25,9
10	Meurthe-et-Moselle	25,3

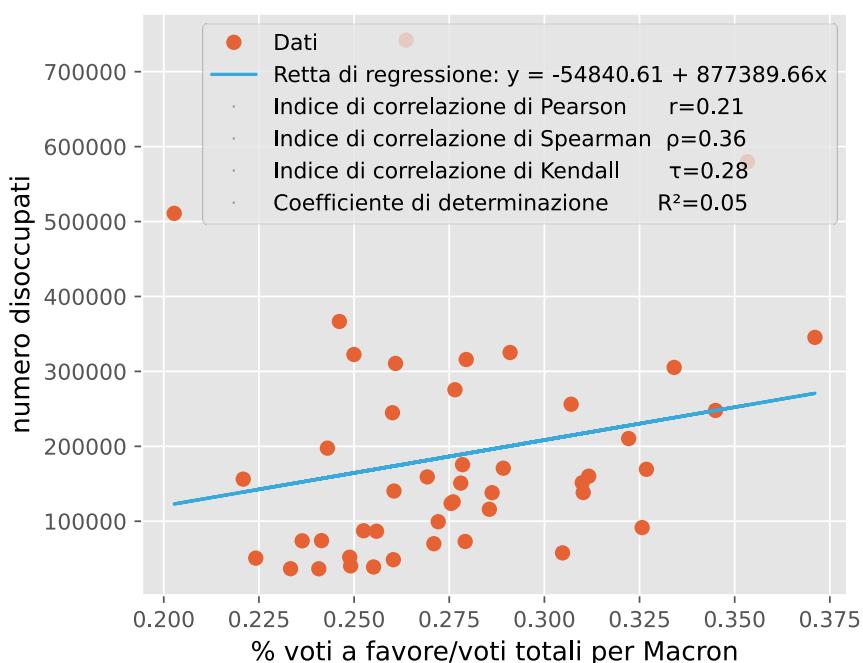
### 3.6 Risultati relativi alle correlazioni

#### Correlazioni relative alle elezioni presidenziali

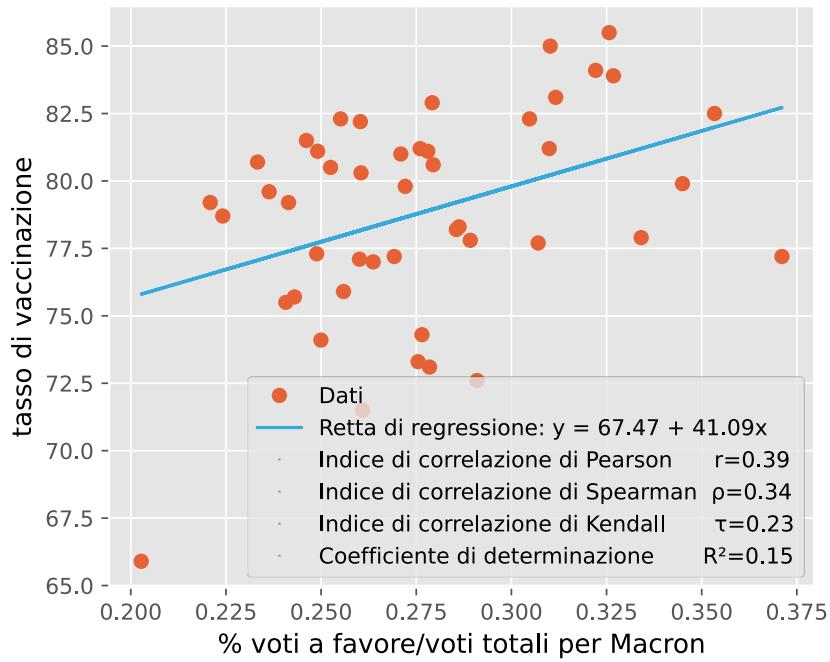
Il coefficiente di correlazione secondo Pearson tra il voto a Macron nel primo turno e il reddito pro capite è il seguente:



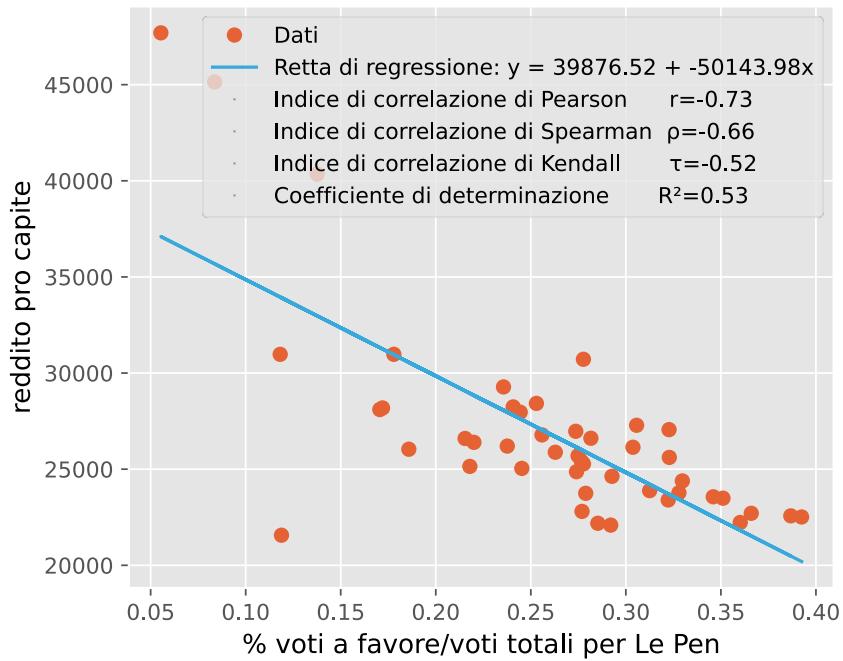
Il coefficiente di correlazione secondo Spearman tra il voto a Macron nel primo turno e disoccupazione è il seguente:



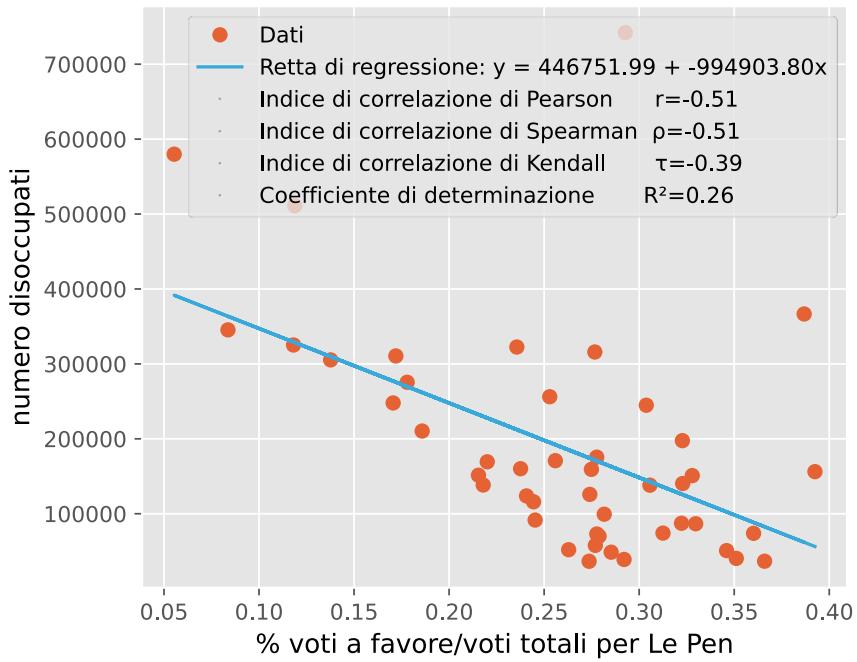
Il coefficiente di correlazione secondo Spearman tra il voto a Macron nel primo turno e il tasso di vaccinazione è il seguente:



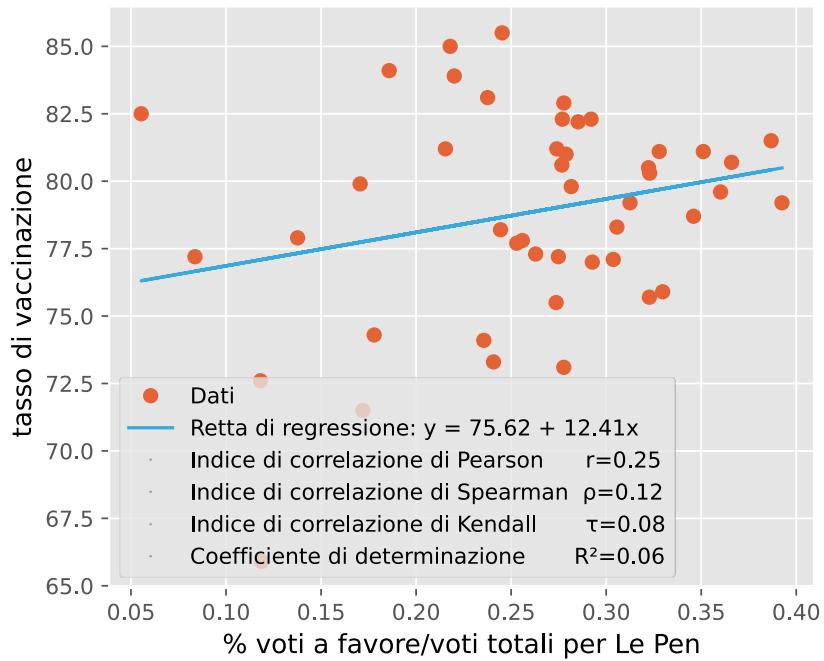
Il coefficiente di correlazione secondo Pearson tra il voto a Le Pen nel primo turno e il reddito pro capite è il seguente:



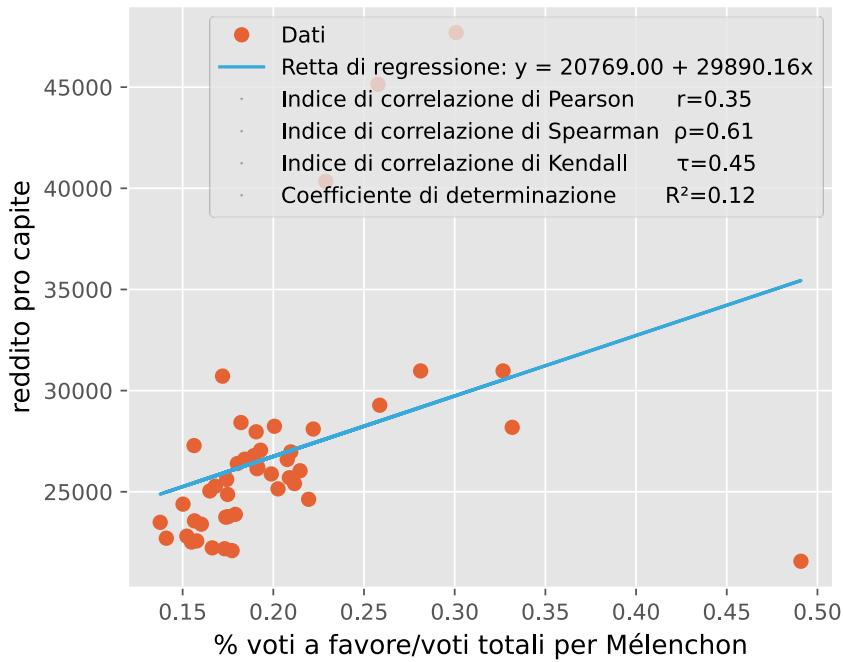
Il coefficiente di correlazione secondo Pearson tra il voto a Le Pen nel primo turno e disoccupazione è il seguente:



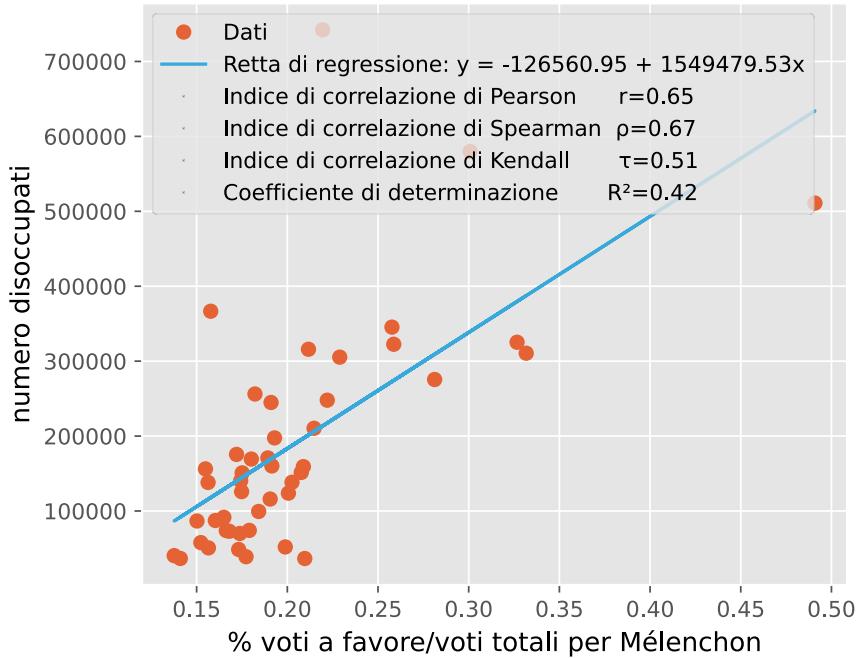
Il coefficiente di correlazione secondo Spearman tra il voto a Le Pen nel primo turno e il tasso di vaccinazione è il seguente:



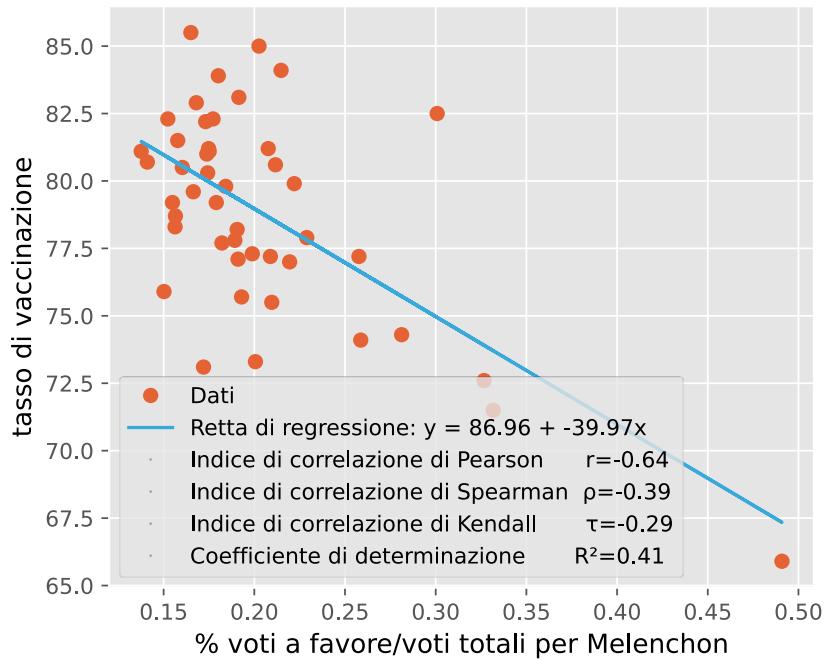
Il coefficiente di correlazione secondo Spearman tra il voto a Mélenchon nel primo turno e il reddito pro capite è il seguente:



Il coefficiente di correlazione secondo Pearson tra il voto a Mélenchon nel primo turno e disoccupazione è il seguente:

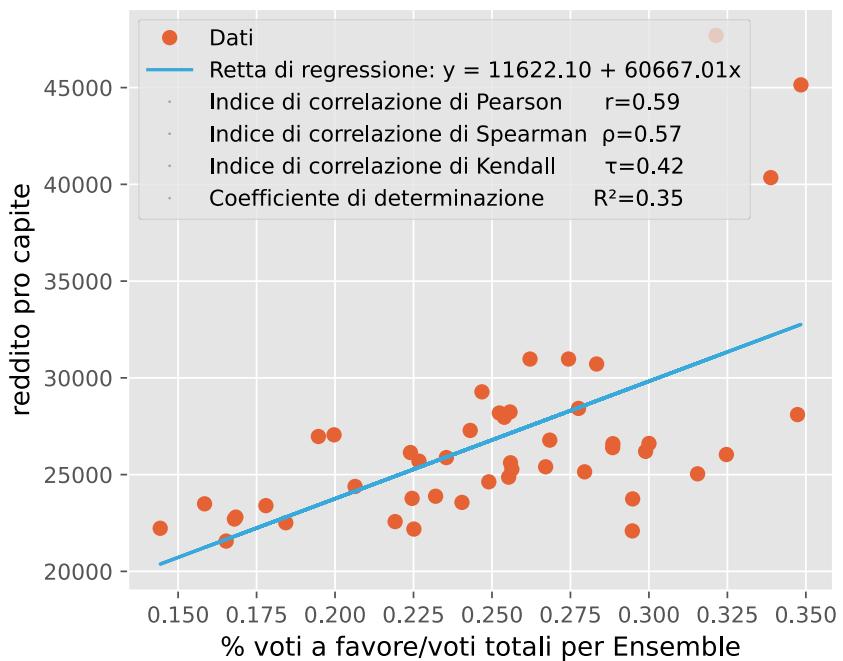


Il coefficiente di correlazione secondo Pearson tra il voto a Melenchon nel primo turno e il tasso di vaccinazione è il seguente:

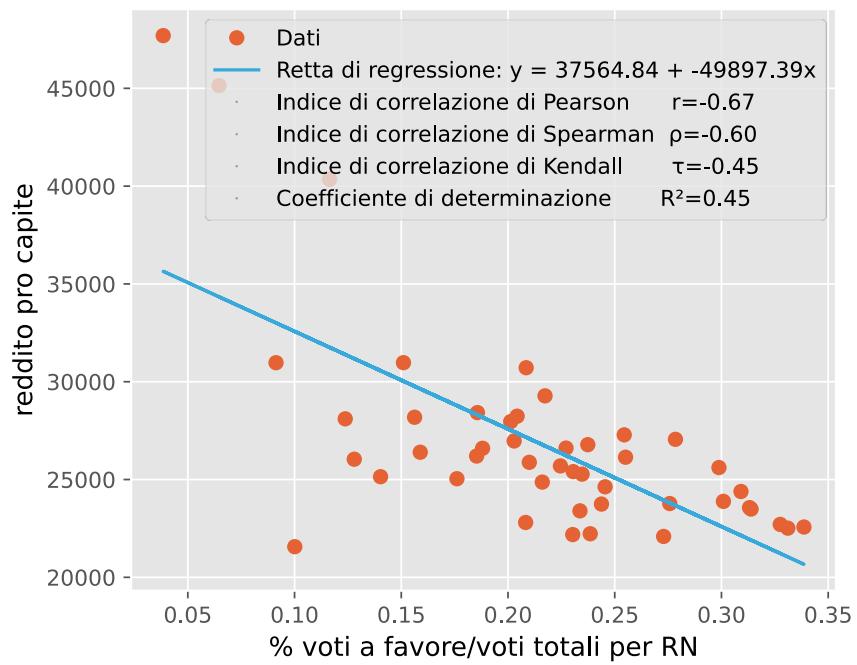


### Correlazioni relative alle elezioni legislative

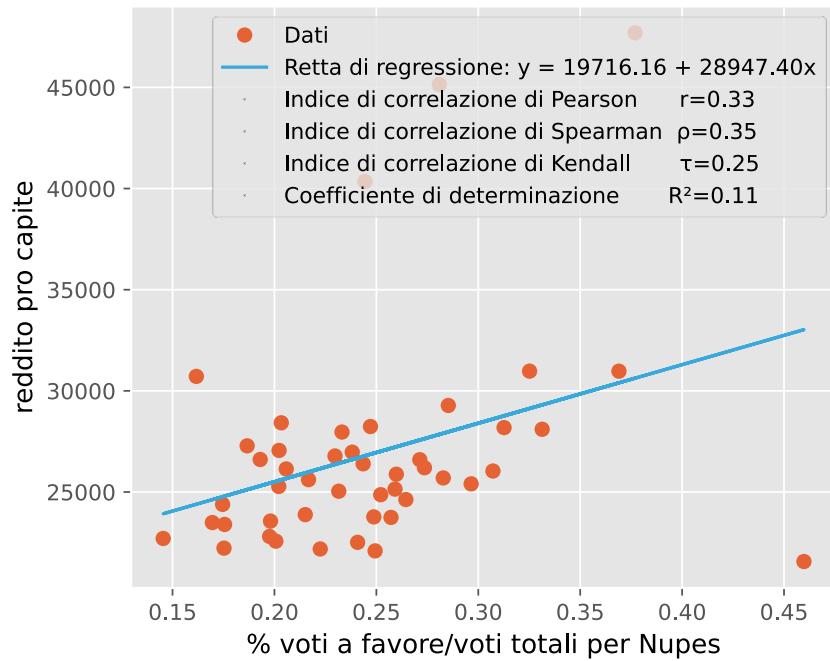
Il coefficiente di correlazione secondo Pearson tra il voto a Ensemble nel primo turno e il reddito pro capite è il seguente:



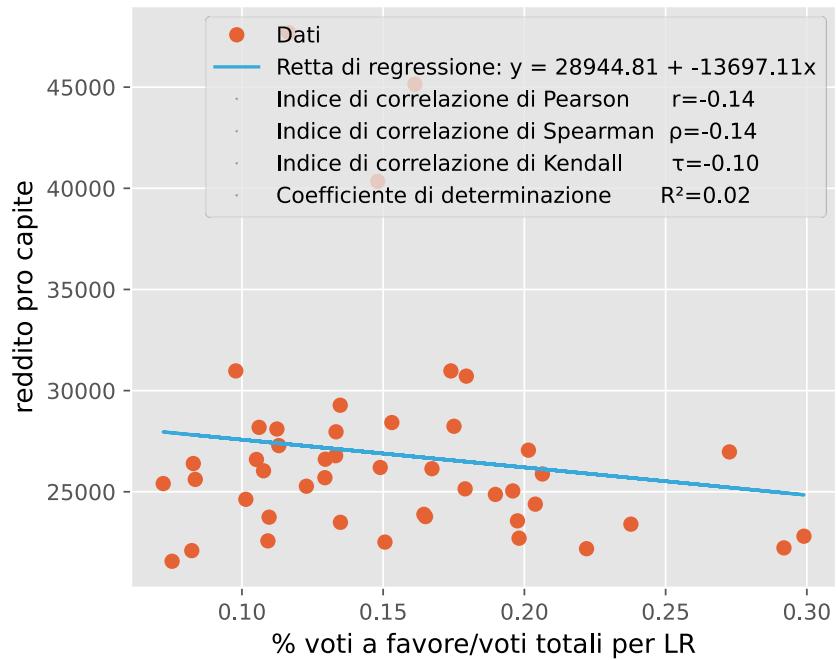
Il coefficiente di correlazione secondo Pearson tra il voto a RN nel primo turno e il reddito pro capite è il seguente:



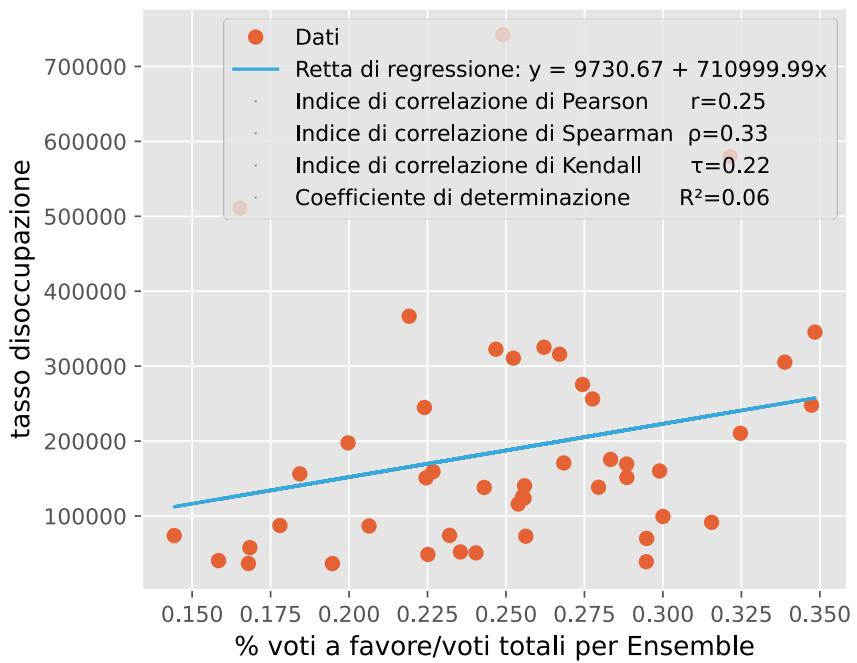
Il coefficiente di correlazione secondo Spearman tra il voto a Nuples nel primo turno e il reddito pro capite è il seguente:



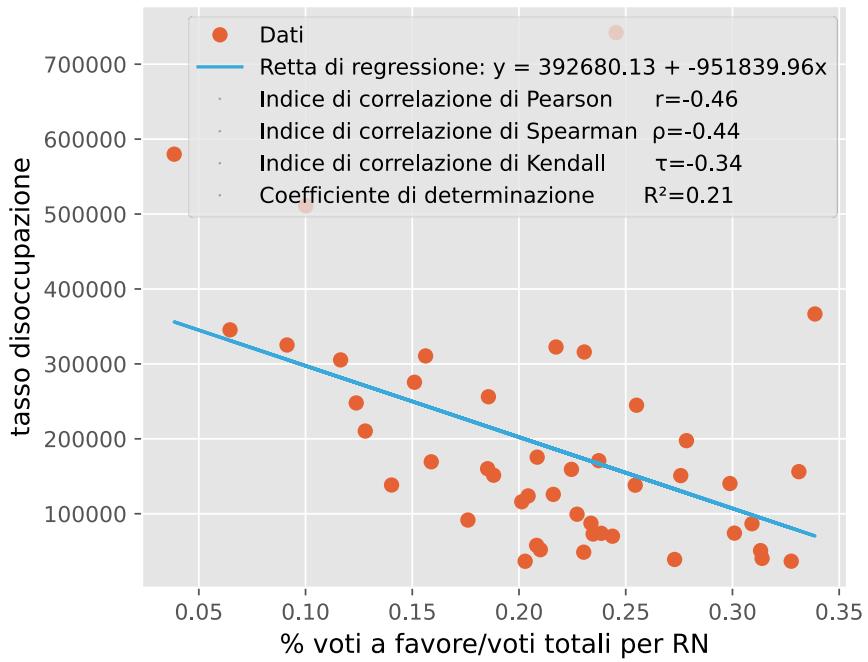
Il coefficiente di correlazione secondo Spearman tra il voto a LR nel primo turno e il reddito pro capite è il seguente:



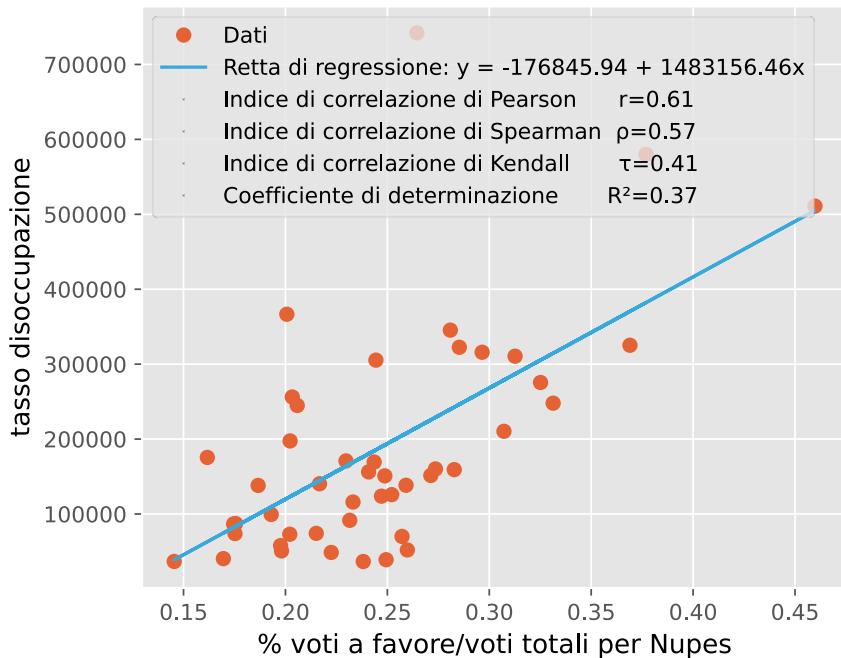
Il coefficiente di correlazione secondo Spearman tra il voto a Ensemble nel primo turno e disoccupazione è il seguente:



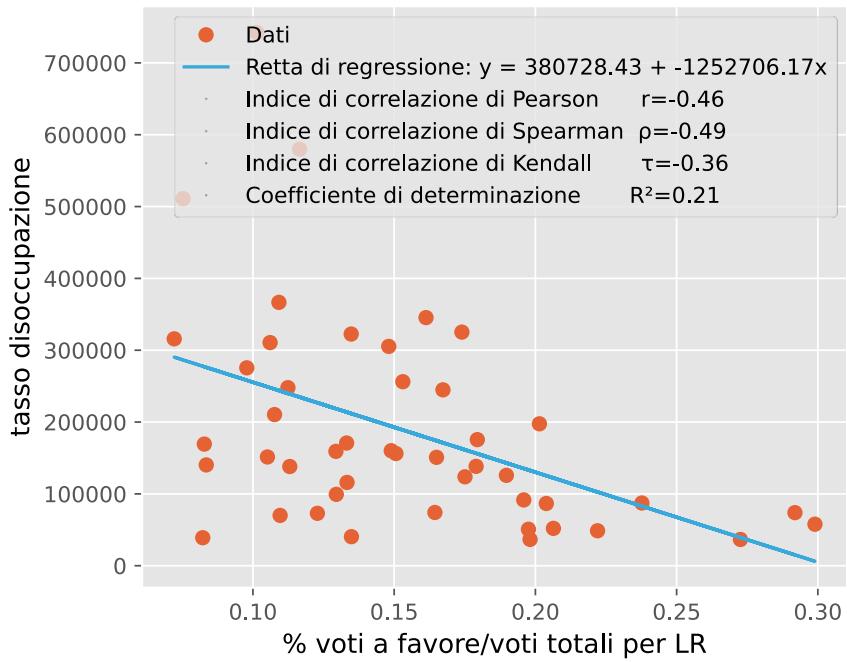
Il coefficiente di correlazione secondo Pearson tra il voto a RN nel primo turno e disoccupazione è il seguente:



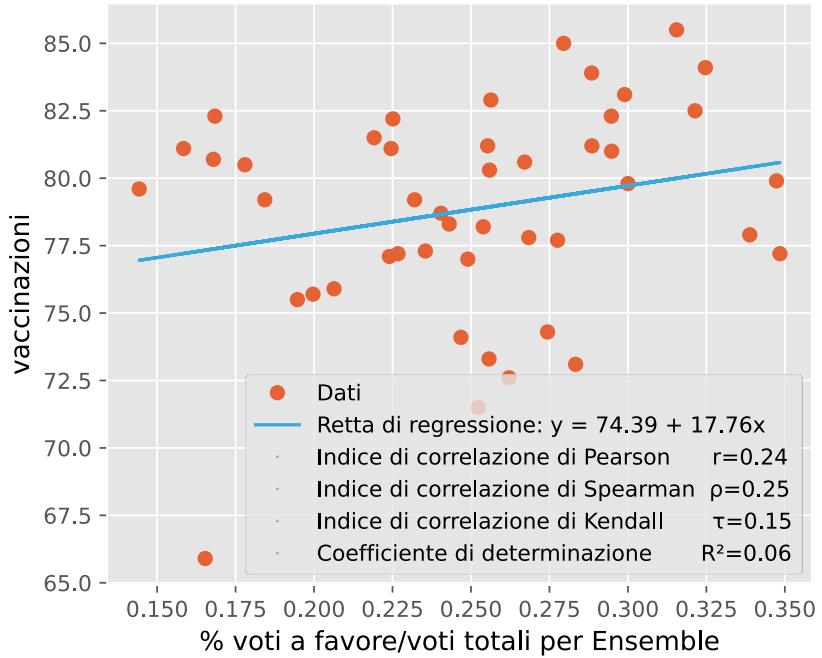
Il coefficiente di correlazione secondo Pearson tra il voto a Nuples nel primo turno e disoccupazione è il seguente:



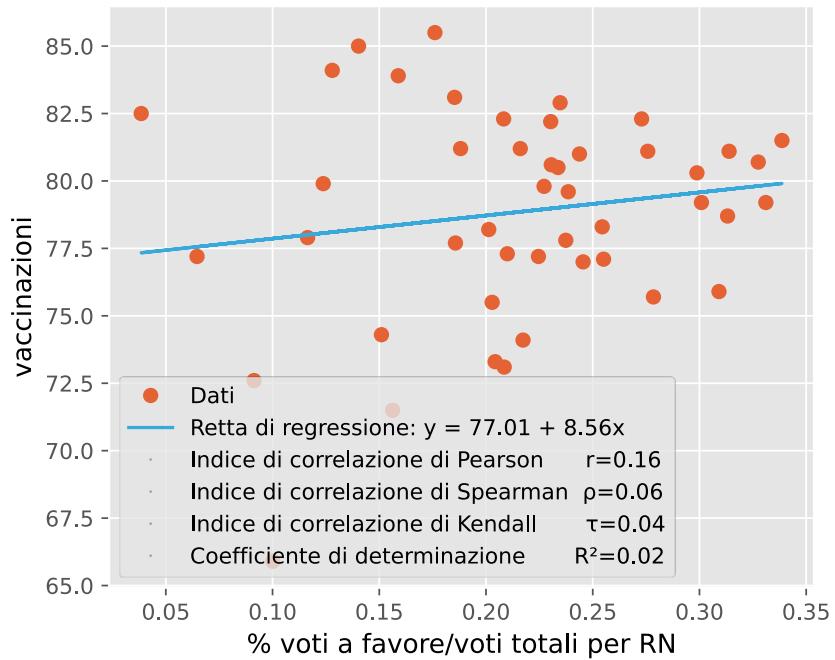
Il coefficiente di correlazione secondo Pearson tra il voto a LR nel primo turno e disoccupazione è il seguente:



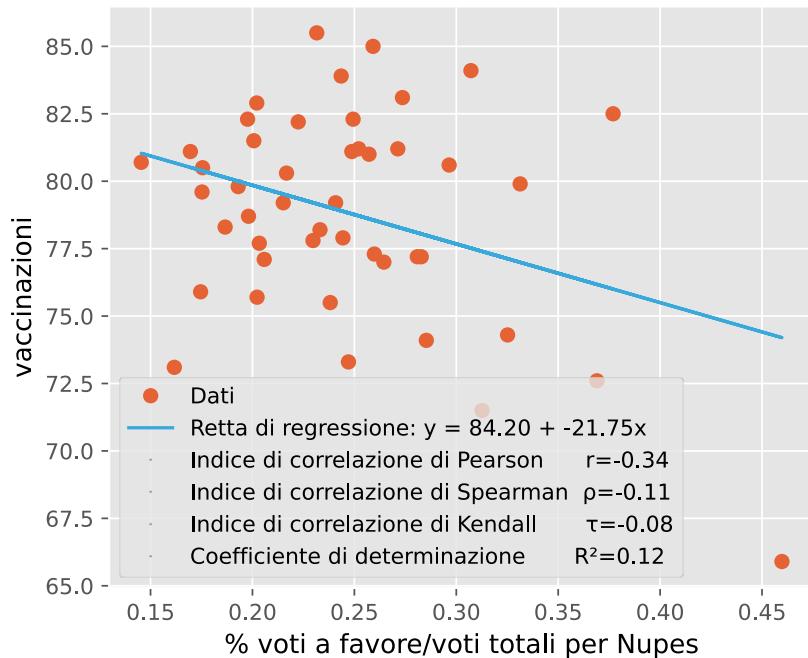
Il coefficiente di correlazione secondo Spearman tra il voto a Ensemble nel primo turno e il tasso di vaccinazione è il seguente:



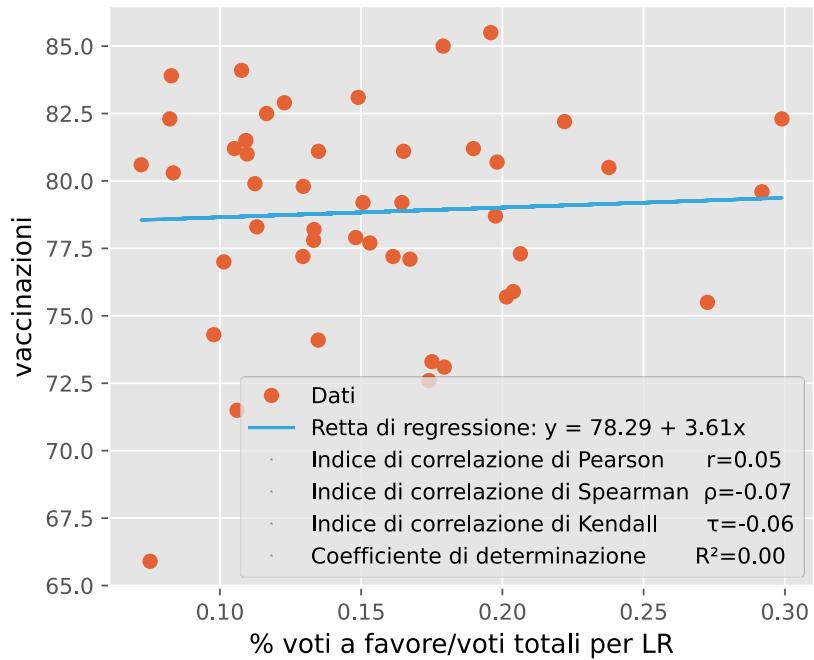
Il coefficiente di correlazione secondo Spearman tra il voto a RN nel primo turno e il tasso di vaccinazione è il seguente:



Il coefficiente di correlazione secondo Spearman tra il voto a Nuples nel primo turno e il tasso di vaccinazione è il seguente:

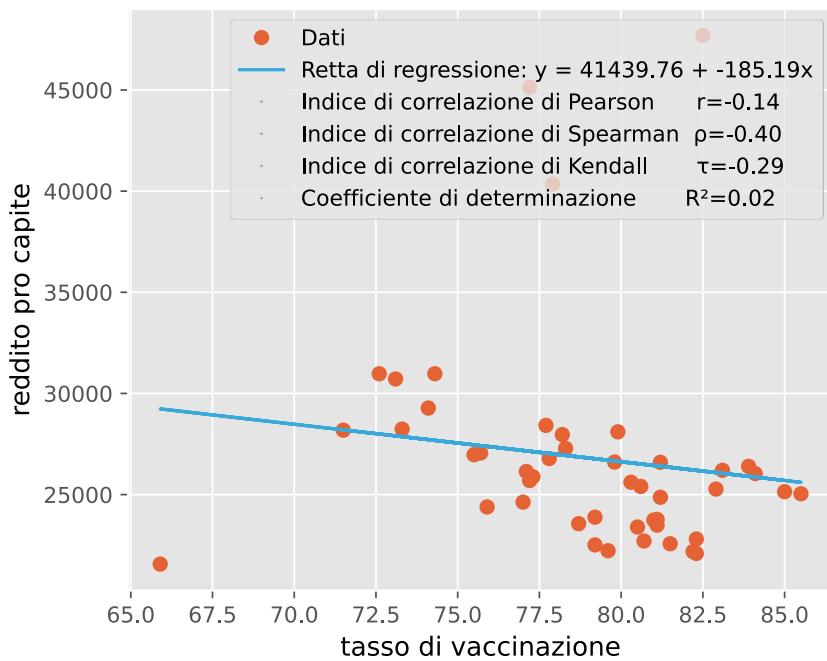


Il coefficiente di correlazione secondo Spearman tra il voto a LR nel primo turno e il tasso di vaccinazione è il seguente:



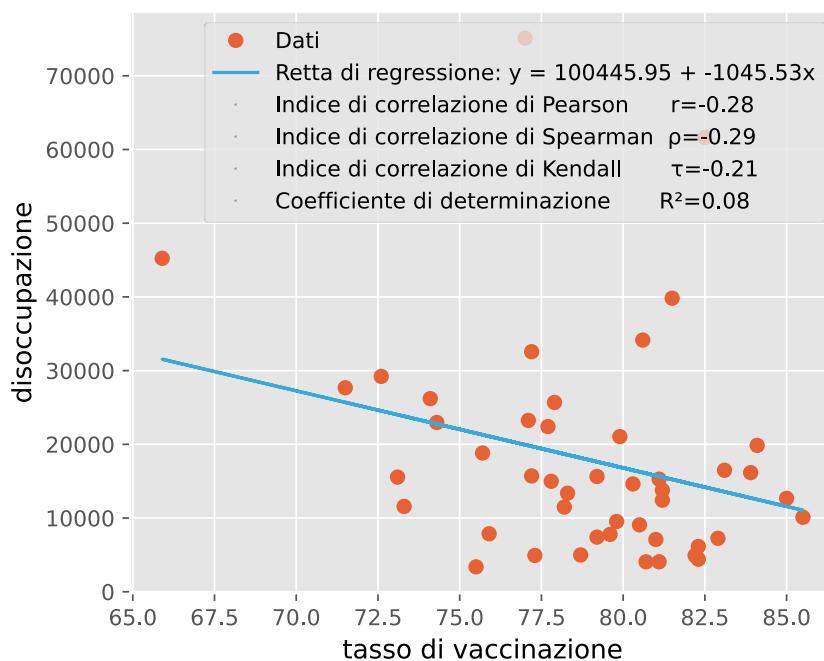
### Correlazioni relative al tasso di vaccinazione e il reddito pro capite

Il coefficiente di correlazione secondo Spearman tra il tasso di vaccinazione e il reddito pro capite è il seguente:



## Correlazioni relative al tasso di vaccinazione e il tasso di disoccupazione

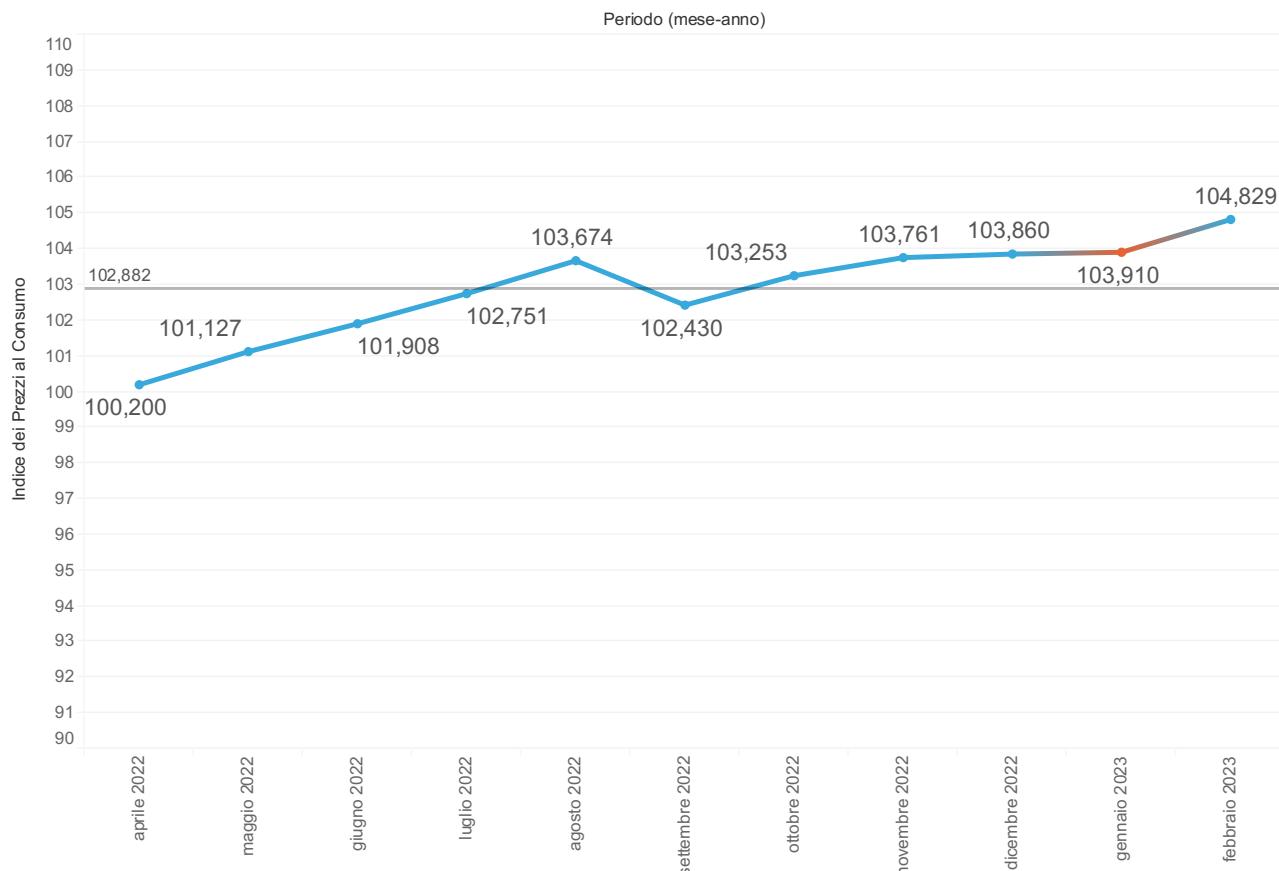
Il coefficiente di correlazione secondo Spearman tra il tasso di vaccinazione e il tasso di disoccupazione è il seguente



### 3.7 Risultati relativi alle analisi visuali

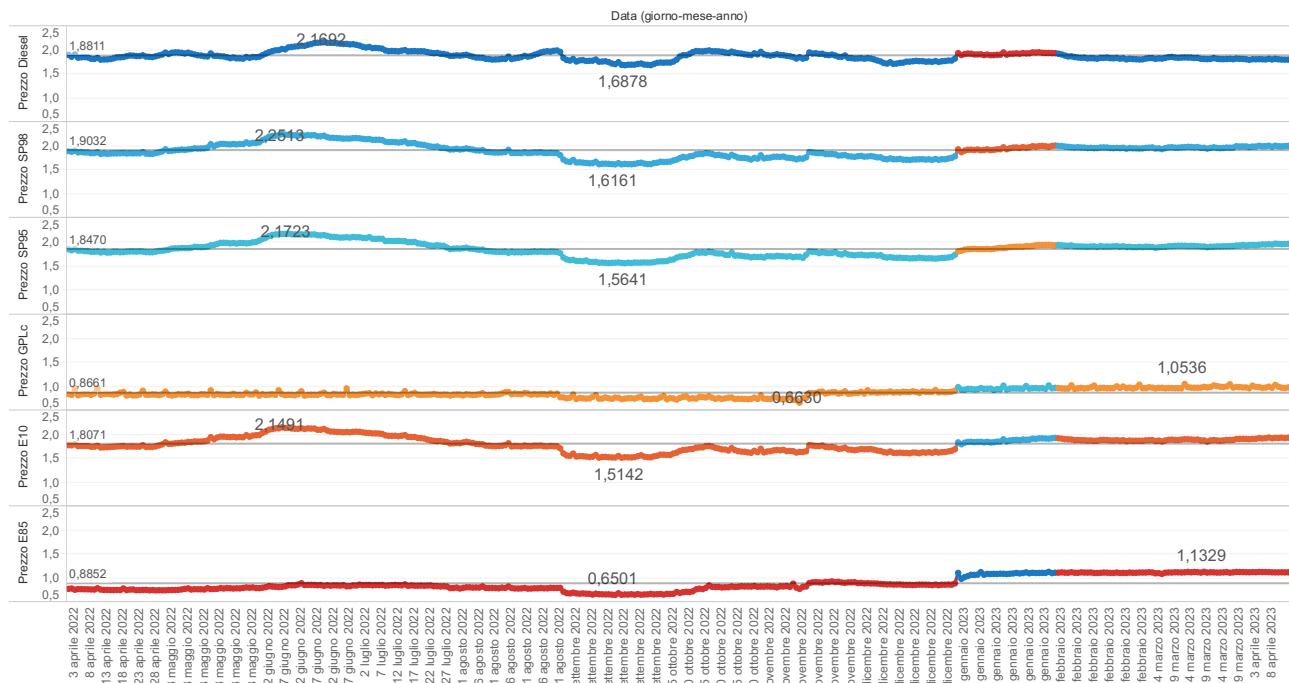
Il grafico generato a partire dai dati relativi all'indice dei prezzi al consumo è il seguente:

Andamento Prezzi Consumo

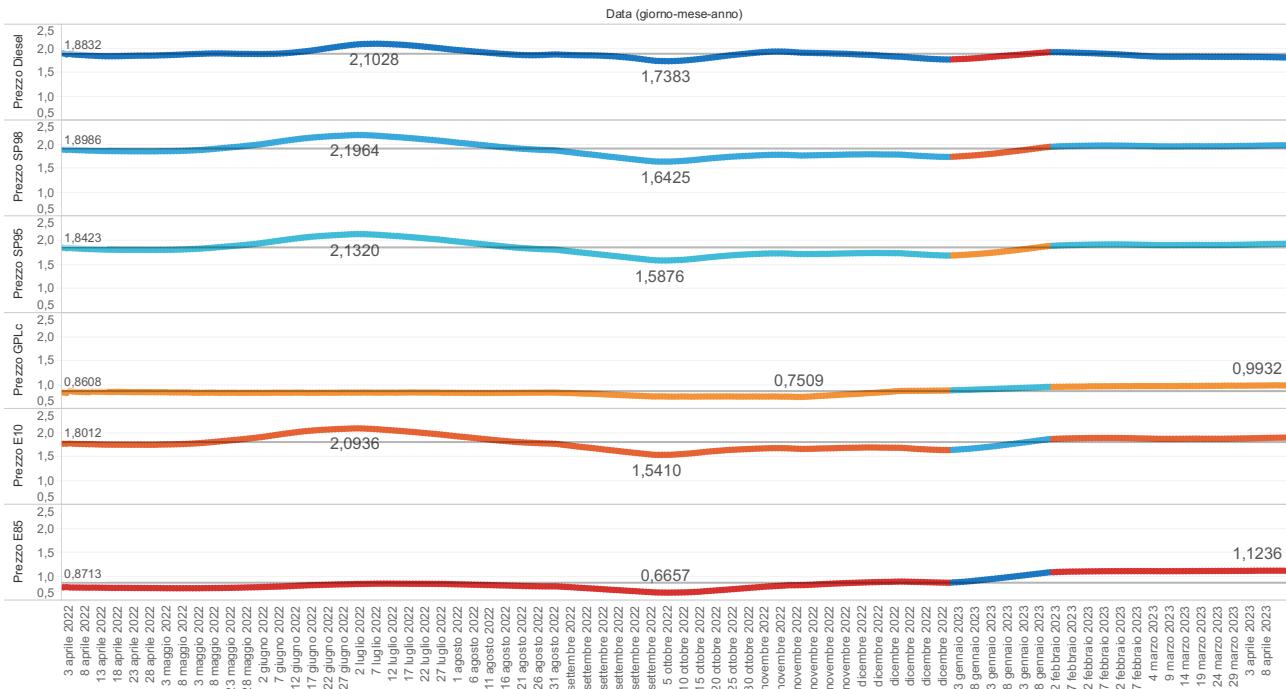


I tre grafici elaborati a partire dai dati relativi ai prezzi dei carburanti sono:

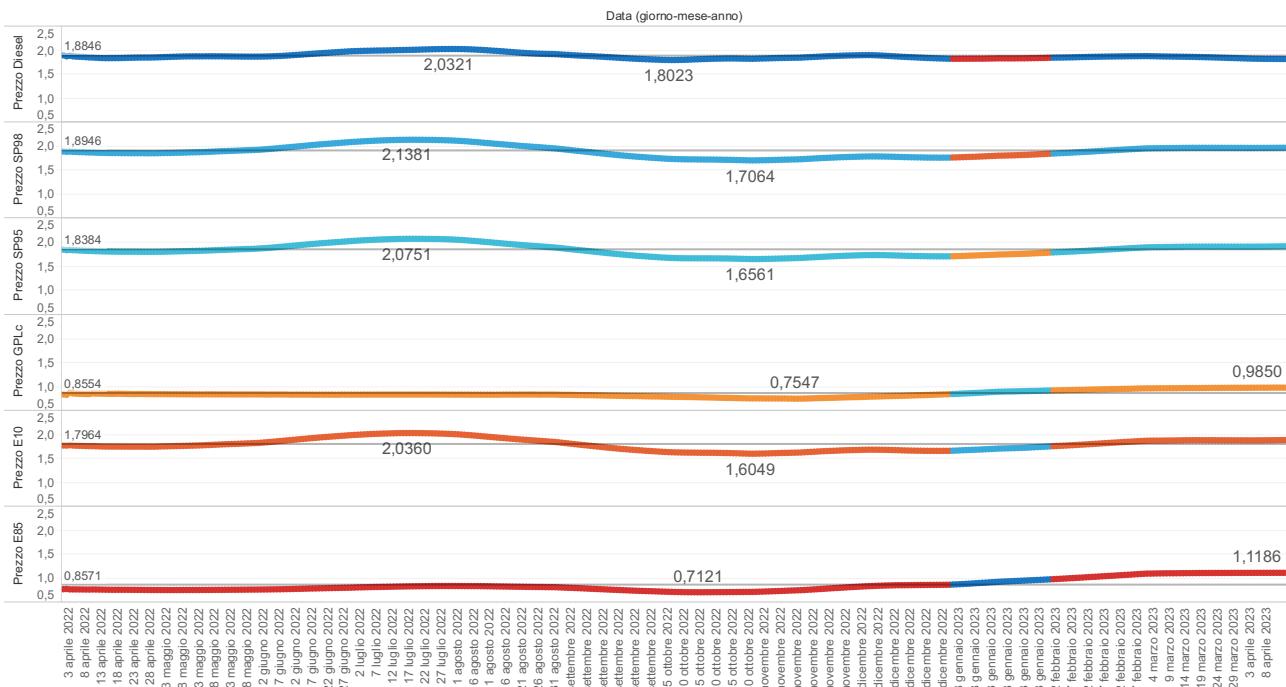
Andamento Prezzi Carburanti



### Andamento Prezzi Carburanti media mobile 30gg



### Andamento Prezzi Carburanti media mobile 60gg



### 3.8 Conclusioni

Le analisi svolte finora consentono di affermare che le elezioni francesi del 2022 riflettono delle tendenze significative. In primo luogo, si osserva che Macron, il vincitore delle presidenziali, ha ottenuto un ampio sostegno nel Nord-Ovest della Francia, in particolare nelle regioni di Normandia, Bretagna e grande parte dell'Ile de France, al contrario le regioni del Nord-Est, come Grand Est e Borgogna sono quelle in cui ha riscosso il minor favore. Analizzando dal punto di vista economico, le regioni sopra elencate, sia attraverso le query sulla disoccupazione per dipartimento che la correlazione tra voto a favore di un candidato e reddito pro capite, è emersa la tendenza degli elettori appartenenti alle regioni con un reddito pro capite più alto a sostenere Macron.

In secondo luogo, la candidata di opposizione Le Pen ha ottenuto la maggior percentuale di voto proprio nelle regioni dove Macron ha ottenuto risultati elettorali minori, allo stesso tempo non guadagnando consensi nelle regioni centrali. Per le stesse analisi svolte per il candidato precedente, per Le Pen si osserva che gli elettori che l'hanno sostenuta sono perlopiù provenienti da regioni con la maggior disoccupazione e minor reddito pro capite.

Il candidato per la lista Nupes, Melechon ha ricevuto la maggior percentuale di voto da regioni centrali. Dalle correlazioni positive forti sia per reddito che per disoccupazione emerge che ha raccolto un insieme eterogeneo di elettori.

Per quanto riguarda le elezioni legislative tenutesi a giugno 2022 i partiti associati ai candidati hanno raggiunto pressoché gli stessi risultati riscontrati per le presidenziali, i dati ottenuti dall'analisi riguardanti il partito LR non hanno avuto particolare peso. Tuttavia, è importante notare che per le regioni: Grand Est, Hauts-de-France e Normandia c'è stato il più rilevante spostamento di opinione tra gli elettori, che di conseguenza hanno votato alle legislative diversamente rispetto alle presidenziali.

Questo cambiamento può essere motivato dai risultati delle query sull'inflazione; infatti, tra i dipartimenti più colpiti si trovano proprio Grand Est e Ile de France.

Inoltre, dai grafici sul consumo e prezzi carburante si evince una generale tendenza dei costi nel crescere. Nel dettaglio, a giugno 2022 si è verificato il massimo aumento dei prezzi per quasi tutti i carburanti; per il consumo, invece, continua la crescita che porterà poi tra luglio e agosto a superare la media dei prezzi.

L'analisi dei dati sulle vaccinazioni ha prodotto dei dati coerenti con le politiche a riguardo di ciascun rappresentante politico, ma non risulta legata a reddito o disoccupazione.

Concludendo, si ritiene che molti fattori abbiano contribuito all'insorgere delle proteste, come il crescente costo della vita che ha portato i sostenitori di Macron a perdere fiducia e che ha rafforzato le motivazioni degli oppositori. Inoltre, dai grafici dell'andamento dei prezzi del carburante si nota un lieve picco a inizio gennaio, che si è subito trasformato in un andamento costante, ma i costi si sono stabilizzati al di sopra della media, questo potrebbe essere uno dei fattori scatenanti delle proteste.