

# Pre Elaborazione e Analisi Esplorativa dei prezzi del Carburante in Italia negli ultimi 20 anni

## 1. FONTI E LICENZE

Scaricamento Dati e Metadati da Fonte Dati:

Ministero dell'ambiente e della sicurezza energetica - Direzione generale infrastrutture e sicurezza.

<https://dgsaie.mise.gov.it/open-data>

Licenza: IODL 2.0 <https://www.dati.gov.it/content/italian-open-data-license-v20>

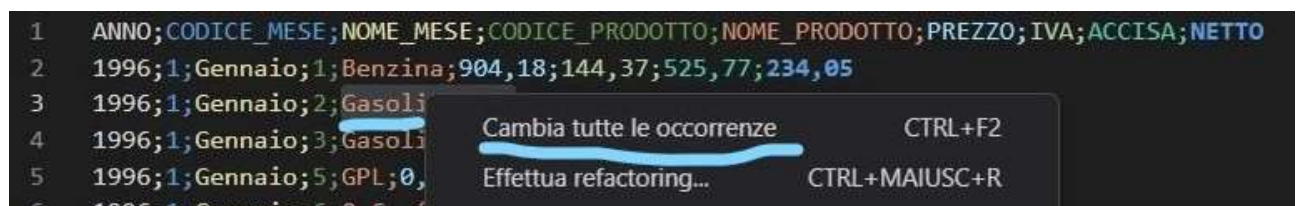
Note (Metadati)

- Il file Excel CSV IT usa come delimitatore il carattere ';' (punto e virgola) e come separatore dei decimali il carattere ',' (virgola).
- I prezzi settimanali sono disponibili dal 2005 ad oggi.
- I prezzi mensili sono disponibili dal 1996 ad oggi.
- I prezzi mensili del GPL sono disponibili da giugno 1997 ad oggi.
- Il prezzo è moltiplicato per un fattore di 1000 per avere una maggiore accuratezza nei decimali.
- Per i carburanti, si tiene esclusivamente conto dei prezzi praticati in modalità self-service.
- Per l'olio combustibile denso BTZ la rilevazione viene effettuata senza IVA.
- Il prezzo del prodotto GPL viene rilevato solo per il settore autotrazione.
- L'accisa del GPL autotrazione viene pubblicata in litri mentre in Gazzetta Ufficiale viene data in kg: il valore di conversione usato è 0.55 (Ministero delle Finanze - 15/01/1999)

## 2. PRIMO SGUARDO

Usando VSCode come TextEditor, che riconosce il formato file, verifico la quantità delle righe e faccio una prima verifica su intestazioni, testi e punteggiature, per modificare eventualmente i dati in tutte le occorrenze e non valore singolo ripetuto volta per volta.

**La punteggiatura e le intestazioni sono idonee per essere caricate in MySQL, mentre per i dati modificherò la nomenclatura "Gasolio auto" in "Diesel".**



```
1 ANNO;CODICE_MESE;NOME_MESE;CODICE_PRODOTTO;NOME_PRODOTTO;PREZZO;IVA;ACCISA;NETTO
2 1996;1;Gennaio;1;Benzina;904,18;144,37;525,77;234,05
3 1996;1;Gennaio;2;Gasolio;904,18;144,37;525,77;234,05
4 1996;1;Gennaio;3;Gasolio;904,18;144,37;525,77;234,05
5 1996;1;Gennaio;5;GPL;0,00;0,00;0,00;0,00
6 1996;1;Gennaio;6;0,00;0,00;0,00;0,00
```

Esclusa l'intestazione risultano 3060 righe.



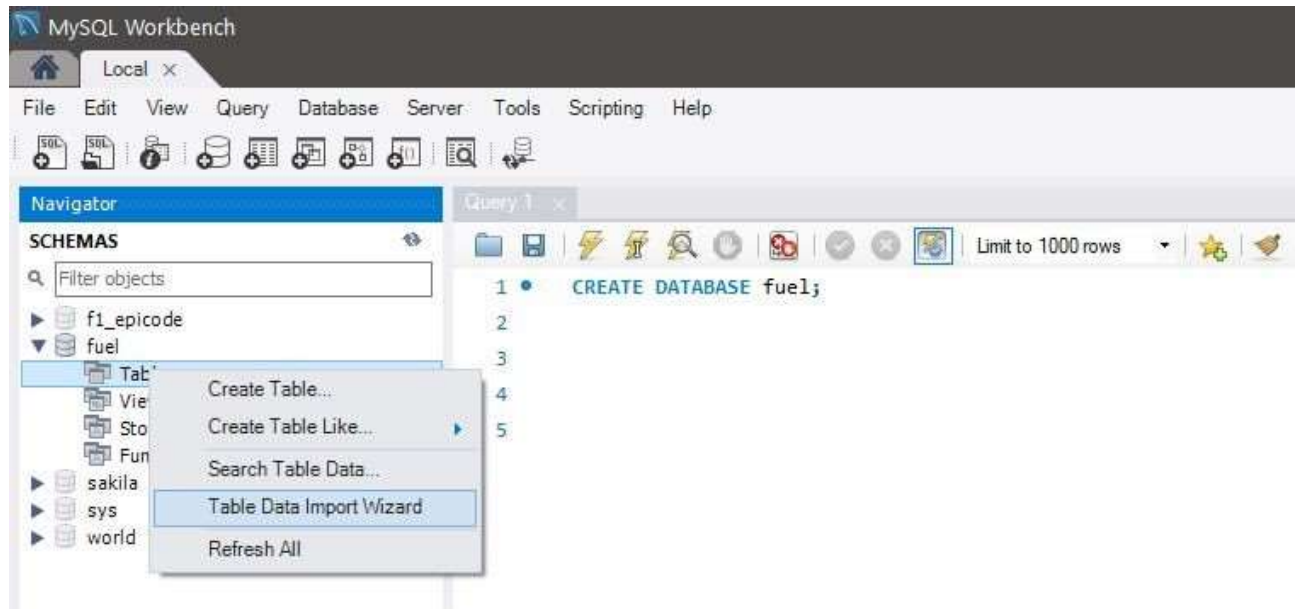
```
CSVLint Query Align Rainbow OFF Ri 3061
```

### 3. CARICAMENTO DEI DATI

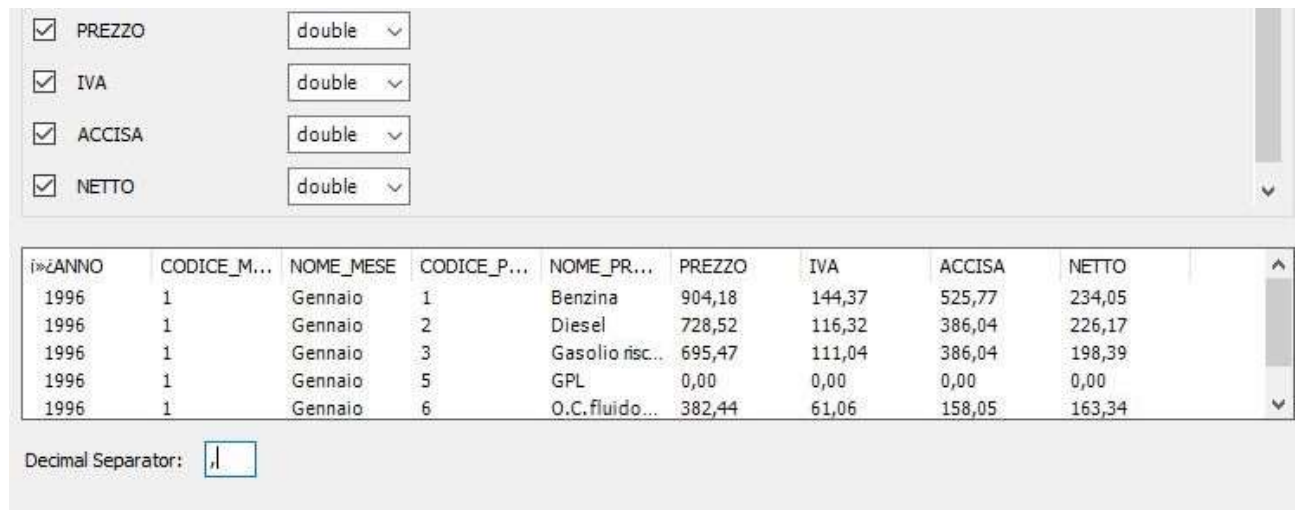
Attraverso MySQL Workbench creo il database e carico la tabella:

Query: "CREATE DATABASE Fuel;"

Caricamento Tabella attraverso la funzione "Table Data Import Wizard"



Verifico e imposto formato dati corretti e la virgola come delimitatore decimale.



In ultimo verifico se i record caricati sono 3060, come da informazione ricevuta tramite il TextEditor.



Perfetto.

#### 4. DATA PREPROCESSING (PRE ELABORAZIONE) & EDA (ANALISI ESPLORATIVA DEI DATI)

Esploro ed analizzo i dati per allineare il dataset all'indagine che intendo sviluppare, cercando di cogliere quante più informazioni interessanti possibili.

(Per motivi grafici presento le query con una formattazione compressa).

Utilizzando il Linguaggio MySQL sviluppo il seguente processo:

Database: Fuel - Table: Carburanti\_it

##### Conteggio righe e colonne

```
8 SELECT
9 (SELECT COUNT(*) FROM carburanti_it)AS nr_righe,
10 (SELECT COUNT(*) FROM information_schema.columns
11 WHERE table_schema = 'fuel' AND table_name = 'carburanti_it') AS nr_colonne;
```

nr_righe	nr_colonne
3060	9

##### Visualizzazione colonne e tipo di dati

```
9 DESCRIBE carburanti_it;
```

Field	Type	Null	Key	Default	Extra
↳ANNO	int	YES	(EMPTY)	(NULL)	(EMPTY)
CODICE_MESE	int	YES	(EMPTY)	(NULL)	(EMPTY)
NOME_MESE	text	YES	(EMPTY)	(NULL)	(EMPTY)
CODICE_PRODOTTO	int	YES	(EMPTY)	(NULL)	(EMPTY)
NOME_PRODOTTO	text	YES	(EMPTY)	(NULL)	(EMPTY)
PREZZO	double	YES	(EMPTY)	(NULL)	(EMPTY)
IVA	double	YES	(EMPTY)	(NULL)	(EMPTY)
ACCISA	double	YES	(EMPTY)	(NULL)	(EMPTY)
NETTO	double	YES	(EMPTY)	(NULL)	(EMPTY)

Notando che vi è un nome di colonna con dei simboli non previsti provvedo a modificarne il nome ed anche il tipo di dato, che invece di “year” risulta “int” benchè sia una data in formato anno(yyyy).

```
12 ALTER TABLE carburanti_it
13 CHANGE ↳ANNO ANNO year;
14 SHOW COLUMNS FROM carburanti_it WHERE field = 'ANNO';
```

Field	Type	Null	Key	Default	Extra
ANNO	year	YES	(EMPTY)	(NULL)	(EMPTY)

## Verifico valori nulli

```
17 SELECT
18   (SELECT COUNT(ANNO) FROM carburanti_it WHERE ANNO IS NULL) AS n_a,
19   (SELECT COUNT(NOME_PRODOTTO) FROM carburanti_it WHERE NOME_PRODOTTO IS NULL) AS n_np,
20   (SELECT COUNT(CODICE_PRODOTTO) FROM carburanti_it WHERE CODICE_PRODOTTO IS NULL) AS n_cp,
21   (SELECT COUNT(CODICE_MESE) FROM carburanti_it WHERE CODICE_MESE IS NULL) AS n_cm,
22   (SELECT COUNT(NOME_MESE) FROM carburanti_it WHERE NOME_MESE IS NULL) AS n_nm,
23   (SELECT COUNT(PREZZO) FROM carburanti_it WHERE PREZZO IS NULL) AS n_p,
24   (SELECT COUNT(IVA) FROM carburanti_it WHERE IVA IS NULL) AS n_i,
25   (SELECT COUNT(ACCISA) FROM carburanti_it WHERE ACCISA IS NULL) AS n_ac,
26   (SELECT COUNT(NETTO) FROM carburanti_it WHERE NETTO IS NULL) AS n_n;
27
```

n_a	n_np	n_cp	n_cm	n_nm	n_p	n_i	n_ac	n_n
0	0	0	0	0	0	0	0	0

## Conteggio valori distinti per colonna.

```
33 SELECT
34   COUNT(DISTINCT ANNO) AS A,
35   COUNT(DISTINCT CODICE_MESE) AS C_M,
36   COUNT(DISTINCT NOME_MESE) AS N_M,
37   COUNT(DISTINCT CODICE_PRODOTTO) AS C_P,
38   COUNT(DISTINCT NOME_PRODOTTO) AS N_P ,
39   COUNT(DISTINCT PREZZO) AS P,
40   COUNT(DISTINCT IVA) AS I,
41   COUNT(DISTINCT ACCISA) AS Ac,
42   COUNT(DISTINCT NETTO) AS N
43 FROM carburanti_it;
44
```

A	C_M	N_M	C_P	N_P	P	I	Ac	N
28	17	17	7	7	2821	2168	201	2803

Noto che

- Più di 12 valori distinti nei campi mese
- 28 valori distinti in quello degli anni
- Stessa quantità di valori distinti nei campi delle colonne “codice\_mese/nome\_mese” e “codice\_prodotto/nome\_prodotto”

Andrò a verificare ed eliminare eventuali valori ridondanti o non appropriati ai miei obiettivi di analisi, considerando che intuisco più valori nelle colonne dei mesi, più anni (basando la mia analisi agli ultimi vent'anni solari), e che probabilmente ci sono delle colonne ridondanti tra prodotti e mesi.

### Verifico collegamento tra Codice\_Mese & Nome\_Mese

```
54 SELECT DISTINCT CODICE_MESE, NOME_MESE
55 FROM carburanti_it;
```

CODICE_MESE	NOME_MESE
11	Novembre
12	Dicembre
13	Anno
14	Primo Trimestre
15	Secondo Trimestre
16	Terzo Trimestre
17	Quarto Trimestre

I mesi testuali della colonna NOME\_MESE sono opportunamente collegati ai dati della colonna CODICE\_MESE da 1 a 12 come da calendario, quindi andrò ad eliminare la colonna NOME\_MESE avendo i codici come riferimenti mensili, ed elimino i valori non utili a fini degli obiettivi, i codici 14-15-16-17.

```
52 ALTER TABLE carburanti_it
53 DROP COLUMN NOME_MESE;
54
55 DELETE FROM carburanti_it
56 WHERE CODICE_MESE BETWEEN 13 AND 17;
57
```

### Verifico collegamento Codice\_Prodotto & Nome\_Prodotto

```
60 SELECT DISTINCT CODICE_PRODOTTO, NOME_PRODOTTO
61 FROM carburanti_it;
```

CODICE_PRODOTTO	NOME_PRODOTTO
1	Benzina
2	Diesel
3	Gasolio riscaldamento
5	GPL
6	O.C. fluido BTZ
7	O.C. ATZ
8	O.C. denso BTZ

Anche in questo caso verifico il collegamento fra i codici e i nomi delle due colonne, essendo entrambi dati univoci propendo per l'eliminazione della colonna dei codici, per una lettura testuale del carburante.

```
63 ALTER TABLE carburanti_it
64 DROP COLUMN CODICE_PRODOTTO;
```



In ultimo vado ad eliminare gli anni che non mi occorrono per avere un dataset composto dai valori degli ultimi vent'anni solari, come da obiettivo preposto.

```
68 DELETE FROM carburanti_it
69 WHERE (ANNO BETWEEN 1996 AND 2002) OR ANNO = 2023;
70
```

A questo punto crea una nuova tabella "fuel\_price\_it", composta da dati filtrati dalla tabella "carburanti\_it" sulla quale sono state eseguite le su scritte query.

Nella nuova tabella riporto:

- I dati dei carburanti che mi occorrono, benzina, gpl, diesel.
- Una colonna con l'incidenza delle tasse sul prezzo finale, "inc\_tax ", formato frazionato.
- Una colonna che somma i valori delle colonne accisa ed iva, "tax".
- Divido i valori economici per 1000, troncandoli a 3 cifre decimali, per convertirli in €, come da prezzi esposti al distributore.

```
72 CREATE TABLE Fuel_Price_It AS(
73     SELECT
74         ANNO AS anno,
75         CODICE_MESE AS mese,
76         NOME_PRODOTTO AS prodotto,
77         TRUNCATE((PREZZO / 1000), 3) AS prezzo,
78         TRUNCATE((IVA / 1000) + (ACCISA / 1000), 3) AS tax,
79         TRUNCATE((NETTO / 1000), 3) AS netto,
80         TRUNCATE((PREZZO - NETTO)/PREZZO, 2) AS inc_tax
81     FROM Carburanti_it
82     WHERE
83         NOME_PRODOTTO = "Benzina"
84         OR NOME_PRODOTTO = "GPL"
85         OR NOME_PRODOTTO = "Diesel");
86
87
```

Ecco la Tabella "fuel\_price\_it", con la quale ritengo finita la fase di Pre Elaborazione.

```
11 SELECT * FROM fuel_price_it LIMIT 9;
12
```

anno	mese	prodotto	prezzo	tax	netto	inc_tax
2003	1	Benzina	1.068	0.719	0.348	0.67
2003	1	Diesel	0.891	0.551	0.339	0.61
2003	1	GPL	0.562	0.25	0.312	0.44
2003	2	Benzina	1.087	0.723	0.364	0.66
2003	2	Diesel	0.906	0.554	0.351	0.61
2003	2	GPL	0.577	0.252	0.324	0.43
2003	3	Benzina	1.095	0.724	0.37	0.66
2003	3	Diesel	0.936	0.559	0.377	0.59
2003	3	GPL	0.591	0.255	0.336	0.43

Dalla Tabella fuel\_price\_it vado ad estrapolare degli indici di confronto tra i 3 carburanti scelti, andando a creare al tabella Eda\_1, formata da un unione di 3 subquery identiche, diverse solo per il riferimento al carburante, della quale riporto in maniera ultra compressa solo la query del carburante “Diesel” per motivi di esposizione grafica.

```
UNION
SELECT * FROM
(SELECT
(SELECT DISTINCT prodotto FROM fuel_price_it WHERE prodotto = "Diesel") AS fuel, -- nome carburante
(SELECT TRUNCATE(AVG(prezzo),3) FROM fuel_price_it
WHERE (anno BETWEEN 2003 AND 2007) AND prodotto = "Diesel") AS 03_07_avg, -- media prezzo 5 anni 03/07
(SELECT TRUNCATE(AVG(prezzo),3) FROM fuel_price_it
WHERE (anno BETWEEN 2008 AND 2012) AND prodotto = "Diesel") AS 08_12_avg, -- media prezzo 5 anni 08/12
(SELECT TRUNCATE(AVG(prezzo),3) FROM fuel_price_it
WHERE (anno BETWEEN 2013 AND 2017) AND prodotto = "Diesel") AS 13_17_avg, -- media prezzo 5 anni 13/17
(SELECT TRUNCATE(AVG(prezzo),3) FROM fuel_price_it
WHERE (anno BETWEEN 2018 AND 2022) AND prodotto = "Diesel") AS 18_22_avg, -- media prezzo 5 anni 18/22
(SELECT TRUNCATE(AVG(prezzo),3) FROM fuel_price_it
WHERE prodotto = "Diesel" AND anno = 2022) AS 22_avg, -- media prezzo 2022, ultimo anno intero
(SELECT TRUNCATE(MIN(prezzo),3) FROM fuel_price_it
WHERE prodotto = "Diesel") AS min_p, -- prezzo minimo
(SELECT DISTINCT anno FROM fuel_price_it
WHERE prodotto = "Diesel" AND prezzo = (SELECT MIN(prezzo)
FROM fuel_price_it WHERE prodotto = 'Diesel')) AS min_p_y, -- anno prezzo minimo
(SELECT TRUNCATE(MAX(prezzo),3) FROM fuel_price_it
WHERE prodotto = "Diesel") AS max_p, -- prezzo massimo
(SELECT DISTINCT anno FROM fuel_price_it
WHERE prodotto = "Diesel" AND prezzo = (SELECT MAX(prezzo)
FROM fuel_price_it WHERE prodotto = 'Diesel')) AS max_p_y, -- anno prezzo massimo
(SELECT TRUNCATE(AVG(tax), 3) FROM fuel_price_it
WHERE prodotto = "Diesel") AS avg_tax, -- media tasse
(SELECT TRUNCATE(AVG(inc_tax),2) FROM fuel_price_it WHERE prodotto = "Diesel") AS avg_it, -- media incidenza tasse
(SELECT TRUNCATE(AVG(inc_tax),2) FROM fuel_price_it WHERE prodotto = "Diesel" AND ANNO = 2022) AS avg_it22, -- media inc tax22
(SELECT MAX(inc_tax) FROM fuel_price_it
WHERE prodotto = "Diesel") AS max_it, -- massimo incidenza tasse
(SELECT MIN(inc_tax) FROM fuel_price_it
WHERE prodotto = "Diesel") AS min_it) AS tb3 -- minimo incidenza tasse
);
```

La tabella Eda\_1 derivante porterà i seguenti dati:

fuel	03_07_avg	08_12_avg	13_17_avg	18_22_avg	22_avg	min_p	min_p_y	max_p	max_p_y	avg_tax	avg_it	avg_it22	max_it	min_it
GPL	0.584	0.696	0.676	0.685	0.813	0.515	2003	0.888	2012	0.256	0.38	0.31	0.47	0.3
Benzina	1.196	1.462	1.593	1.607	1.81	1.032	2003	2.033	2022	0.899	0.61	0.47	0.71	0.41
Diesel	1.049	1.357	1.467	1.516	1.814	0.845	2003	1.972	2022	0.753	0.55	0.41	0.69	0.36

- fuel = carburante
- 03\_07\_avg = media prezzo 5 anni 2003/07
- 08\_12\_avg = media prezzo 5 anni 2008/12
- 13\_17\_avg = media prezzo 5 anni 2013/17
- 18\_22\_avg = media prezzo 5 anni 2018/22
- 22\_avg= media prezzo 2022 (ultimo anno)
- min\_p = prezzo minimo
- min\_p\_y = anno del prezzo minimo
- max\_p = prezzo massimo
- max\_p\_y = anno del prezzo massimo
- avg\_tax = media tasse
- avg\_it = media incidenza tasse
- avg\_it22 = media incidenza tasse 2022 (ultimo anno)
- max\_it = massimo incidenza tasse
- min\_it = minimo incidenza tasse

Infine con i dati dei prezzi della Tabella fuel\_price\_it e i dati sull'auto più venduta degli ultimi 20 anni, Panda Fiat ( fonte [www.truenumbers.it](http://www.truenumbers.it) e [ww.unrae.it](http://ww.unrae.it)), e i consumi previsti dalle 3 motorizzazioni, 1.3Mjt Diesel 4,3L/100km, 1.2 Bifuel GPL 7,2L/100Km, 1.2 Benzina 5L/100Km (fonte [www.ultimatespecs.com](http://www.ultimatespecs.com) e [it.wikipedia.org/wiki/Fiat\\_Panda](http://it.wikipedia.org/wiki/Fiat_Panda)) **ho creato una piccola tabella che mostra le variazioni di spesa per per un viaggio tra Roma e Milano, 655km, (Aereoporti Fiumicino e Malpensa) e per un viaggio di 100km.**

Come nel caso precedente per motivi di esposizione grafica riporto in maniera compressa una delle 3 subquery facente parti dell'unione che formerà la tabella eda\_car.

```
SELECT * FROM
(SELECT
  (SELECT DISTINCT prodotto FROM fuel_price_it WHERE prodotto = "Diesel") AS fuel,
  (SELECT
    TRUNCATE(AVG(prezzo),3) * (SELECT TRUNCATE((655/(100/4.3)),3))
    FROM fuel_price_it WHERE prodotto = "Diesel" AND anno = 2003) AS rom_mil_car_03,
  (SELECT
    TRUNCATE(AVG(prezzo),3) * (SELECT TRUNCATE((655/(100/4.3)),3))
    FROM fuel_price_it WHERE prodotto = "Diesel" AND anno = 2022) AS rom_mil_car_22,
  (SELECT
    TRUNCATE(AVG(prezzo),3) * 4.3
    FROM fuel_price_it WHERE prodotto = "Diesel" AND anno = 2003) AS 100km_car_03,
  (SELECT
    TRUNCATE(AVG(prezzo),3) * 4.3
    FROM fuel_price_it WHERE prodotto = "Diesel" AND anno = 2022) AS 100km_car_22) AS tb4
UNION
```

## Tabella eda\_car

Riporta l'ammontare ( in euro) di spesa carburante per le tratte scelte, nel 2003 e nel 2022, per osservare la variazione avvenuta in 20 anni, per meglio dire l'aumento di spesa riscontrato.

fuel	longtext	rom_mil_car_03	double	rom_mil_car_22	double	100km_car_03	double	100km_car_22	double
Benzina		34.61		59.27		5.28		9.05	
Diesel		24.67		51.09		3.76		7.8	
GPL		17.65		26.62		3.88		5.85	



## 5. Data Visualization Eda

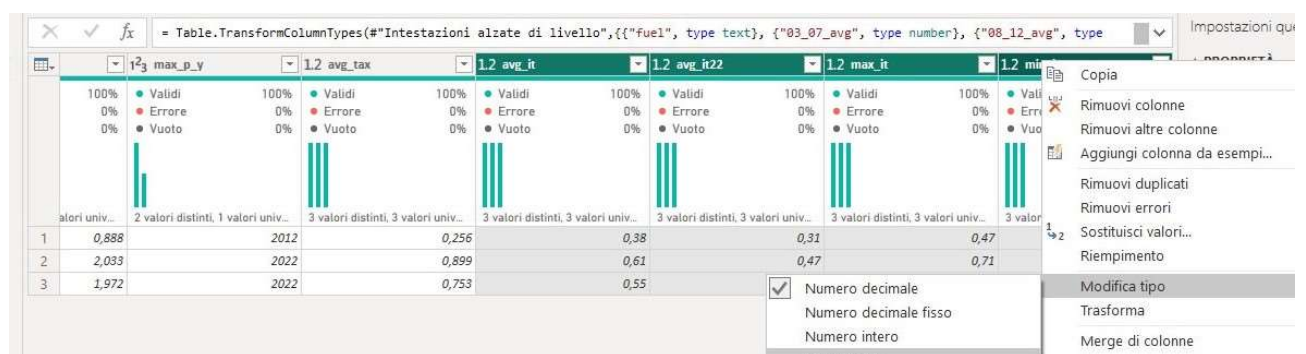
Prima di procedere a creare delle visualizzazioni grafiche dei dati, vado ad impostare un formato idoneo per i campi temporali, mesi e anno, unendoli con una formula di Excel, poiché in MySQL i campi erano separati.

H26          =CONCAT(A26;"-";SE(LUNGHEZZA(B26)=1;CONCAT(0;B26);B26))

	A	B	C	D	E	F	G	H	I
26	2003	9	Benzina	1,071	0,72	0,351	0,67	2003-09	
27	2003	9	Diesel	0,863	0,547	0,316	0,63	2003-09	
28	2003	9	GPL	0,518	0,243	0,275	0,46	2003-09	
29	2003	10	Benzina	1,043	0,715	0,328	0,68	2003-10	
30	2003	10	Diesel	0,86	0,546	0,314	0,63	2003-10	

Carico le tabelle su PowerBi trasformando in Power Query le eventuali occorrenze.

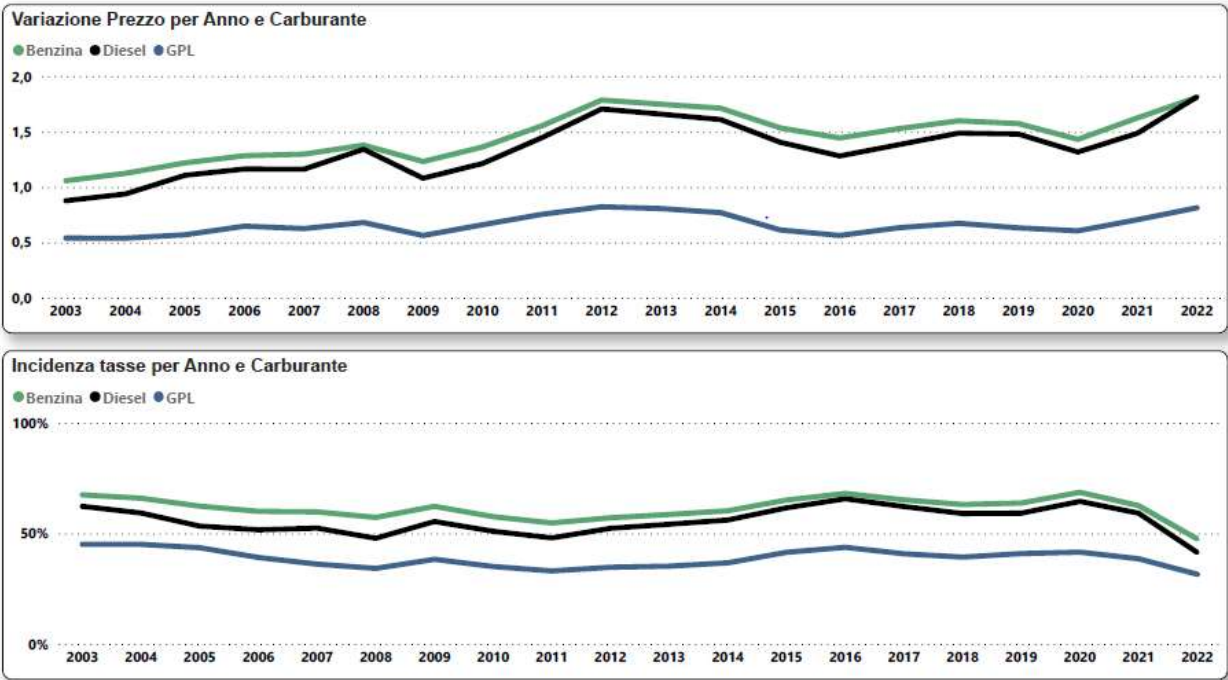
In questo caso cambio il formato in percentuale.



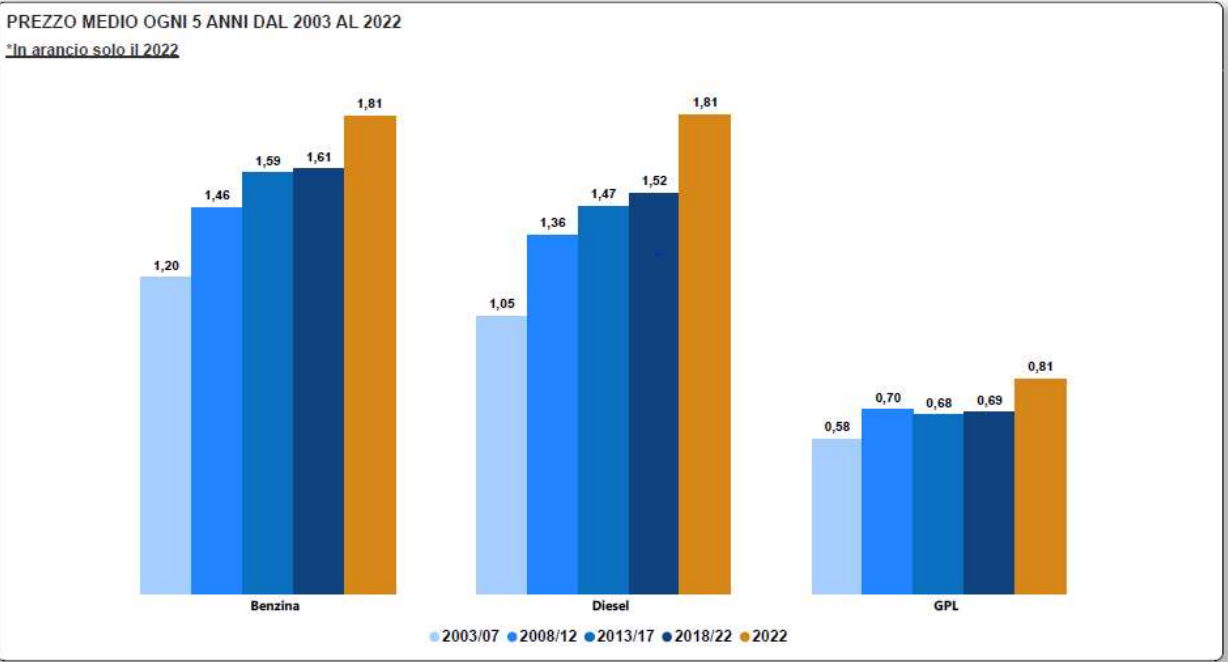
A questo punto utilizzando PowerBi ho creato 3 pagine di report, con dei grafici che in maniera semplice mostrano delle evidenze risultanti dall'analisi, di cui riporto il link della pubblicazione web dell'elaborazione in powerbi:

<https://app.powerbi.com/view?r=eyJrIjojZDdmMTdiN2YtNzBiOS00MTIiLWZhODYtZGRIMjI1M2I4ZjlxliwiZDCl6ImRlZjg5NTIiLTA2OTQtNDQ3OS1hODNiLWZhMWEzYTfkZGQ0ZCIsImMiOiJ9>

Obiettivo: Mostrare graficamente la variazione del prezzo e dell’incidenza delle tasse.



Obiettivo: Mostrare graficamente la variazione media del prezzo per ogni 5 anni e dell’ultimo anno.



Obiettivo: Mostrare Garficamente la variazione di spesa carburante per percorrere due tipologie di percorsi, 100Km generici autostradali e 655Km riguardanti il tratto da aeroporto ad aeroporto tra Roma e Milano, con l’auto più venduta in Italia negli ultimi 20 anni, nelle 3 motorizzazioni scelte, GPL, Diesel e Benzina.

