

# Descrizione del dataset


Il dataset, opportunamente anonimizzato, rappresenta lo **storico dei volumi di vendita di due prodotti acquistati da un mercato consumer**.

I dati sono contenuti in un file csv, di 250561 righe e 6 colonne.

In esso sono riportate le serie storiche delle vendite di **due prodotti**:

- su un periodo temporale di quasi **2 anni e mezzo** (dal 01-01-2014 al 19-05-2016) con **granularità temporale giornaliera**
- localizzate geograficamente seconda una gerarchia di 3 dimensioni: **Zona -> Area -> Sottoarea**

Di seguito è riportato un estratto del dataset:



Zona	Area	Sottoarea	Categoria_prodotto	Data	Vendite
Zona_1	Area_1	Sottoarea_1	Prodotto_1	01/01/2014	0
Zona_1	Area_1	Sottoarea_1	Prodotto_1	02/01/2014	5
...	...	...	...	...	...
Zona_27	Area_74	Sottoarea_145	Prodotto_2	19/05/2016	3

## TEMA DI RICERCA:

1. I dati sono utilizzati per fare una previsione GIORNALIERA dei prossimi 10 giorni di vendite, per ogni SOTTOAREA e per ogni CATEGORIA DI PRODOTTO.

## TEMI ALTERNATIVI DI RICERCA

2. I dati sono utilizzati per fare una previsione COMPLESSIVA dei prossimi 10 giorni di vendite, per ogni SOTTOAREA e per ogni CATEGORIA DI PRODOTTO.
3. I dati sono utilizzati per fare una previsione GIORNALIERA dei prossimi 10 giorni di vendite, per ogni AREA e per ogni CATEGORIA DI PRODOTTO.

Si riportano inoltre le seguenti **note**, a supporto dell'utilizzo del dataset nella competizione:

- Il dataset è stato parzialmente pulito (per esempio si noterà che manca la Sottoarea\_51)
- Per alcune aree geografiche, i dati relativi al 2014 presentano una dinamica differente (per esempio nei volumi di vendite dei giorni festivi) rispetto ai dati più recenti. Da prendere in considerazione se, in alcuni casi, sia meglio eliminare tale periodo dall'analisi
- Il dataset permette di creare KPI utili quali:
  - KPI sulla data (esempio giorno della settimana, mese, ..)
  - KPI sulle stagionalità

- KPI sulle festività
- Il dataset non è stato diviso negli insiemi di Train e Test
- Non è necessario disegnare un solo modello per la previsione (per esempio si possono fare tanti modelli quante sono le combinazioni Sottoaree – Categoria\_prodotto, oppure si possono considerare modelli per ogni giorno della settimana, ...)