

Lavorare coi dati

dati.trentino.it



David Leoni

DISI, Università di Trento

david.leoni@unitn.it

<http://it.linkedin.com/in/davidleoni>

slides v1.0



Outline

- Tabella da dati.trentino →



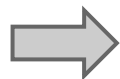
- Integrazione dati
- Strumenti utili

CASO TIPICO

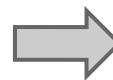
Comma Separated Values

```
assunzioni_prov_trento_nota_trimestrale_2_2013.csv - Blocco note
File Modifica Formato Visualizza ?
Gruppo,Categoria,II trimestr
Sesso,Maschi,16892,"48,49842
Sesso,Femmine,17938,"51,5015
Sesso,Totale,34830,100,32246
Cittadinanza,Italiani,23822,
Cittadinanza,Stranieri,11008
```

TUTTI I PROGRAMMI RIESCONO AD APRIRE I FILE CSV!



Programma



Infografica



1) CSV da dati trentino →



iscrizioni online primo anno 2013-2014



Dataset Organizzazioni Categorie Apps Informazioni FAQ

Dati Aperti del Trentino. Tutti i dati che cercavi del Sistema Trentino.

Home / Organizzazioni / PAT S. Istruzione e form. ... / Iscrizioni al primo anno ... / iscrizioni online primo anno ...

iscrizioni online primo anno 2013-2014

Download

URL: http://www.vivoscuola.it/c/document_library/get_file?uuid=5a82053e-a1a2-4e40-9ac0-84854753e511&groupId=10137

Dal riassunto del dataset

Numero iscrizioni online/cartecee divise per comunità di valle 2013-2014 e per i seguenti tipi di scuole: primarie di primo grado secondarie di primo grado...

Sorgente: Iscrizioni al primo anno degli istituti scolastici per modalità - 2013-2014

Grid Graph Map 16 records « 0 - 100 » Search data ...

COMUNI...	Primaria ...	Primaria ...	Sec I gra...	Sec I gra...	Sec II gr...	Sec II gr...	Form Pr...	Form Pr...
Comunit t...	6	167		218		157	6	55
Comunit ...	77	9	70	11	25	20		32
Comunit ...	51	154		260	33	80	3	51
Comunit ...	49	503	13	587		239	16	173
Comunit ...	11	93		120				

Anteprima come tabella

Significato celle vuote?



<http://dati.trentino.it/dataset/iscrizioni-online-primo-anno-2013-2014/resource/9dde53c0-64f4-405c-b418-6061adae2688>

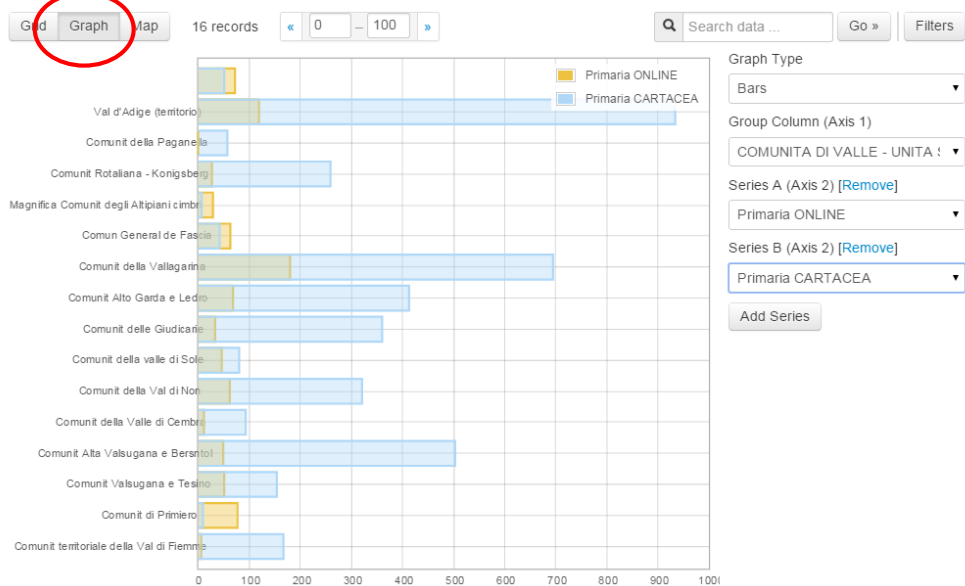
iscrizioni online primo anno 2013-2014

URL: http://www.vivoscuola.it/c/document_library/get_file?uuid=5a82053e-a1a2-4e40-9ac0-84854753e511&groupId=10137

Dal riassunto del dataset

Numero iscrizioni online/cartecee divise per comunità di valle 2013-2014 e per i seguenti tipi di scuole: primarie di primo grado secondarie di primo grado...

Sorgente: Iscrizioni al primo anno degli istituti scolastici per modalità - 2013-2014



2) integrazione dati

Imprese attive

anno;codEnte;valore
2000;74;72
2000;75;167
2000;76;21
2000;78;19
2000;79;190
2000;80;38
2000;81;65
2000;82;48



Elenco dei Comuni della PAT

comu	descriz_1	comv	descriz_2
022001	Ala	10	Vallagarina
022002	Albiano	5	Valle di Cembra
022003	Aldeno	15	Territorio Val d'Adige
022004	Amblar	6	Val di Non
022005	Andalo	14	Paganella
022006	Arco	9	Alto Garda e Ledro
022007	Avio	10	Vallagarina
022009	Baselga di Pin	4	Alta Valsugana e Bersntol
022011	Bedollo	4	Alta Valsugana e Bersntol
022012	Bersone	8	Giudicarie
022013	Besenello	10	Vallagarina



http://dati.trentino.it/dataset/elenco-dei-comuni-della-provincia-autonoma-di-trento/resource/comuni_comunita_codici

<http://dati.trentino.it/dataset/impres-attive/resource/586f3dbb-f1aa-4559-8131-7052cc1be30d>

Imprese attive

<http://dati.trentino.it/dataset/impres-attive/resource/586f3dbb-f1aa-4559-8131-7052cc1be30d>



Dataset Organizzazioni C

Dati Aperti del Trentino. Tutti i

anno;codEnte;valore

2000;74;72

2000;75;167

2000;76;21

2000;78;19

2000;79;190

2000;80;38

2000;81;65

2000;82;48

Cos'è il
CodEnte ??

il separatore
è un punto e
virgola

🏠 / Organizzazioni / PAT S. Statistica / Imprese attive / imprese-attive

impres-attive

URL: <http://www.statweb.provincia.tn.it/indicatoristrutturalisubpro/exp.aspx?idind=87&info=d&fmt=csv>

Imprese attive

Grid Graph Map

1000 records

« 0 - 100 »

🔍 Se

anno;codEnte;val...

2000;1001;1646

2000;1002;828

2000;1003;1484

2000;1004;2997

2000;1005;729

2000;1006;2636

2000;1007;1366

2000;1008;3015

2000;1009;3202

2000;100;14

2000;1010;5228

Elenco dei Comuni della PAT

http://dati.trentino.it/dataset/elenco-dei-comuni-della-provincia-autonoma-di-trento/resource/comuni_comunita_codici



Dataset Organizzazioni Categorie Apps Informazioni FAQ

Dati Aperti del Trentino. Tutti i dati che cercavi del Sistema Trentino.

Organizzazioni / PAT S. Statistica / Elenco dei Comuni della ... / comuni_comunita_codici

comuni_comunita_codici

Download

URL: <http://dati.trentino.it/dataset/701bb7af-7e07-4cc5-8291-f3ea891d963b/resource/fefec39b-fc68-4b02-a5e9-00e7154e0e79/download...>

Dal riassunto del dataset

Elenco dei Comuni della Provincia Autonoma di Trento con relativi codici di appartenenza con relativi codici di...

Sorgente: Elenco dei Comuni della Provincia Autonoma di Trento

Grid Graph Map 217 records 0 - 100 Filters

comu	descriz_1	com	descriz_2
022007	Avio	1	Vallagarina
022009	Baselga di Pin	4	Alta Valsugana e Bersntol
022011	Bedollo	4	Alta Valsugana e Bersntol
022012	Bersone	8	Giudicarie

Iniziano tutti con 022...

Manca la 'è' finale. Ma sarà vero?





Importiamo in OpenOffice CALC

Encoding 'Unicode UTF-8'

Importazione testo - [codicicomunicomunita.csv]

Importa

Tipo di carattere: Unicode (UTF-8)

Lingua: Predefinita - Italiano (Italia)

Dalla riga: 1

Opzioni di sillabazione

Larghezza fissa

Separato

Tabulazione Virgola Altri

Punto e virgola Spazio

Raggruppa i separatori di campo Separ. di testo: "

Altre opzioni

Campo tra virgolette come testo

Individua numeri speciali

Campi

Tipo colonna: Standard

Standard	Standard	Standard	Standard
3	022002 Albiano	5	Valle di
4	022003 Aldeno	15	Territori
5	022004 Amblar	6	Val di Nè
6	022005 Andalo	14	Paganella
7	022006 Arco	9	Alto Gard
8	022007 Avio	10	Vallagar
9	022009 Baselga di Pinè	4	Alta Valè

Sorpresa, la 'è' c'è

7	022006	Arco
8	022007	Avio
9	022009	Baselga di Pinè
10	022011	Bedollo
11	022012	Bersone

Integrazione di tabelle

Imprese
attive

anno;codEnte;valore
2000;74;72
2000;75;167
2000;76;21
2000;78;19
2000;79;190
2000;80;38
2000;81;65
2000;82;48



Codice ente

comu	descriz_1	comv	descriz_2
022001	Ala	10	Vallagarina
022002	Albiano	5	Valle di Cembra
022003	Aldeno	15	Territorio Val d'Adige
022004	Amblar	6	Val di Non
022005	Andalo	14	Paganella
022006	Arco	9	Alto Garda e Ledro
022007	Avio	10	Vallagarina
022009	Baselga di Pin	4	Alta Valsugana e Bersntol
022011	Bedollo	4	Alta Valsugana e Bersntol
022012	Bersone	8	Giudicarie
022013	Besenello	10	Vallagarina

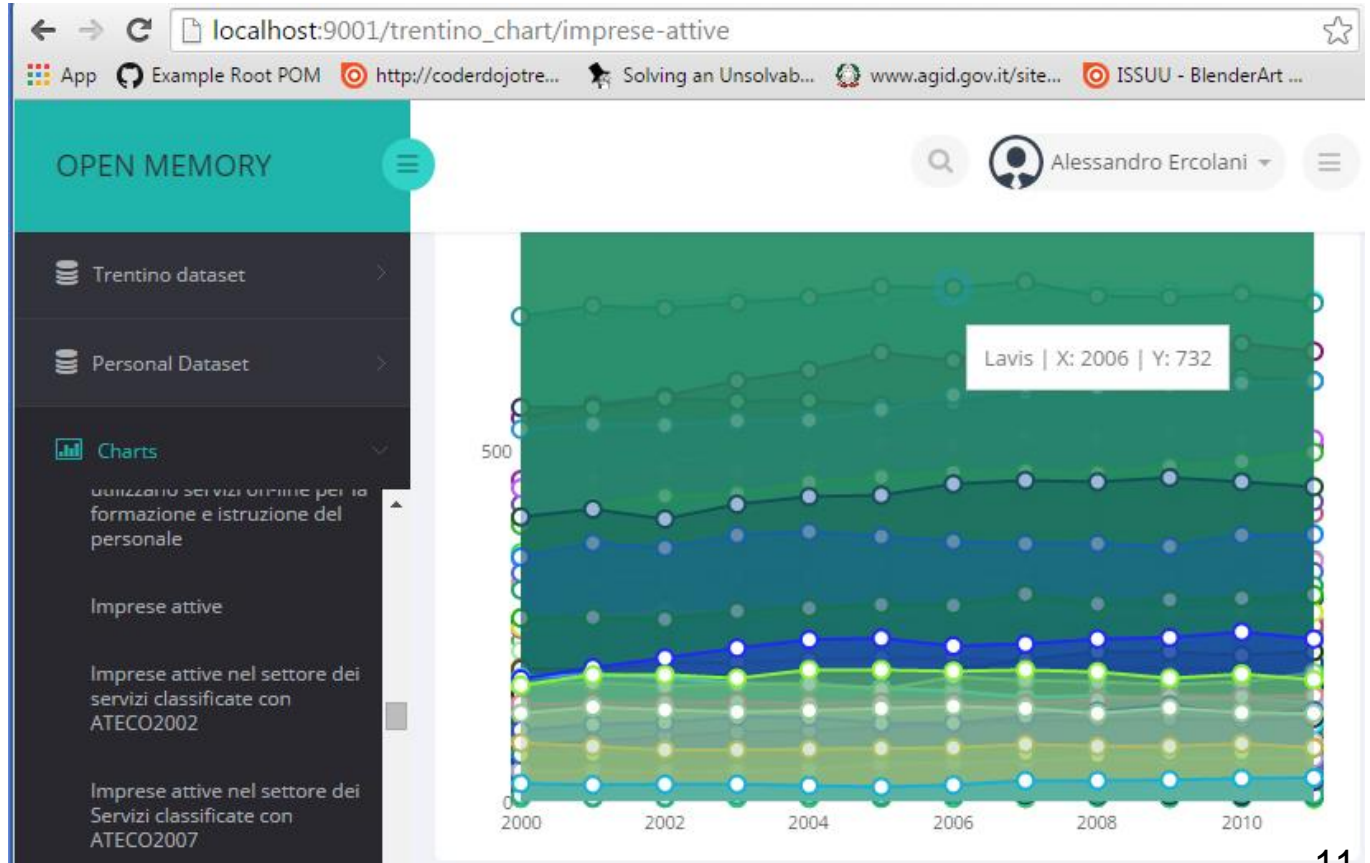


Imprese attive per
comune di
appartenenza

Anno	Comune	Imprese attive
2000	Trento	8906
2000	Rovereto	2773
2000	Levico	446

Open Memory

- Storico dei dati
- Statistiche
- Grafici
- Open source





- Pulizia dati
- Disambiguazione nomi
- Tracciamento provenienza
- Open source

OpenDataRise http://dati.trentino.it/Farmacie nella Provincia di Trento/FARM001 Link Permanente

Disambiguazione del valore degli attributi

Passo 4 di 6

172 righe

Vista a: Mostra: 5 10 25 50 righe

FARMACIA	COMUNE	LATITUDINE_P	LONGITUDINE_P
Nome Stringa []	Comune Località	Latitudine Numero a precisione singola	Longitudine Numero a precisione singola
SILVESTRI LUCIANA	Revo	46.412354	11.052054
RAVAGNANI CHIARA	Ala	45.759007	11.002856
FRANZELLIN VITTORIO E C. SNC	Ala	45.74791	11.041461
BETTINAZZI - R. & T. PHARMA SNC	Arco	45.930424	10.907002
FAMIGLIA COOPERATIVA VALLAGARNA	Avio	45.73216	10.919864
BIALO MICHELE	Borgo Valsugana	46.01531	11.43459
CALLIANO SAS DI VITULLO GHERTA	Calliano	45.934583	11.094259
FRANZELLIN GIANCARLO	Cavalese	46.290806	11.462147
ALL'ANGELO SNC	Denno	46.271946	11.052403
GIRARDI DI DOTT. C. TADDEI & C. SNC	Folgaria	45.909016	11.199658



Esempio dataset farmacie:

Perchè LATITUDINE_P?

Colonna *FARMACIA* significa *Nome*

<p>FARMACIA</p> <p>Nome</p> <p>Stringa []</p>	<p>COMUNE</p> <p>Comune</p> <p>Località</p>	<p>LATITUDINE_P</p> <p>Latitudine</p> <p>Numero a precisione singola</p>
BETTINAZZI - R. & T. PHARMA SNC	Arco	45.930424
FAMIGLIA COOPERATIVA VALLAGARINA	Avio	45.73216

Cos'è ARCO?

Trentino EntityBase

3) Altri tool

csv → grafico interattivo	<u>Data Wrapper</u> , <u>Raw</u>
tabella pdf → tabella	<u>Tabula</u>
sito web → tabella	<u>import.io</u>

Altri link utili:

<https://alessandriaopendata.wordpress.com/strumenti/>

csv → grafico interattivo

DataWrapper

ABZV Datawrapper

+ Nuovo grafico

IT

Accedi / Registrati

1 Carica i dati ✓

2 Verifica e descrivi ✓

3 Visualizza

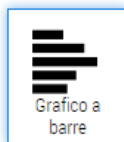
4 Pubblica ed esporta

Chart type

Refine

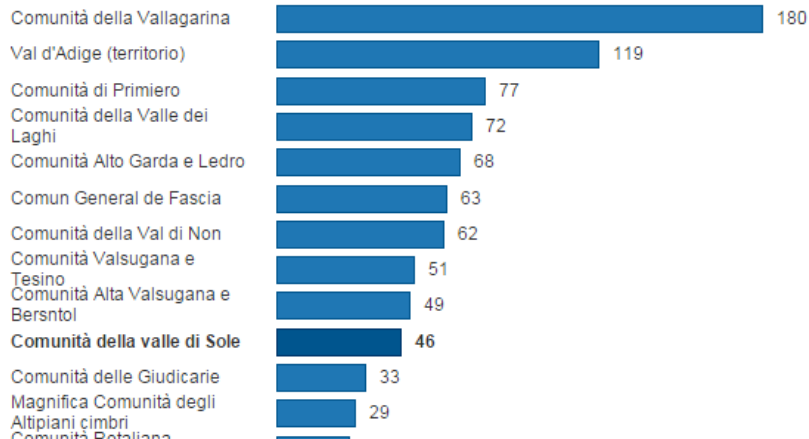
Annotate

Design



Iscrizioni scuola primaria 2013-14

Primaria ONLINE



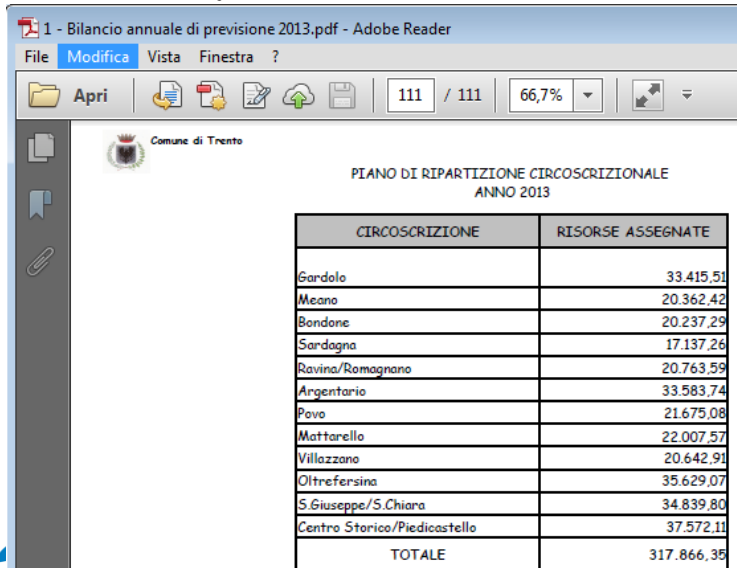
Suggerisci: Nel caso in cui la visualizzazione non fosse come che ti aspettavi, dovresti cercare di [trasporre i dati](#).

PDF → Tabella

Tabula

Difficile fare copia e incolla da PDF...

...ma estrarre con Tabula è facile!!!



1 - Bilancio annuale di previsione 2013.pdf - Adobe Reader

File Modifica Vista Finestra ?

Apri [Icons] 111 / 111 66,7%

Comune di Trento

PIANO DI RIPARTIZIONE CIRCOSCRIZIONALE ANNO 2013

CIRCOSCRIZIONE	RISORSE ASSEGNATE
Gardolo	33.415,51
Meano	20.362,42
Bondone	20.237,29
Sardagna	17.137,26
Ravina/Romagnano	20.763,59
Argentario	33.583,74
Povo	21.675,08
Mattarello	22.007,57
Villazzano	20.642,91
Oltrefersina	35.629,07
S.Giuseppe/S.Chiana	34.839,80
Centro Storico/Piedicastello	37.572,11
TOTALE	317.866,35

Extracted tabular data

CIRCOSCRIZIONE	RISORSE ASSEGNATE
Gardolo	33.415,51
Meano	20.362,42
Bondone	20.237,29
Sardagna	17.137,26
Ravina/Romagnano	20.763,59
Argentario	33.583,74
Povo	21.675,08
Mattarello	22.007,57
Villazzano	20.642,91
Oltrefersina	35.629,07
S.Giuseppe/S.Chiana	34.839,80
Centro Storico/Piedicastello	37.572,11
TOTALE	317.866,35

Copy to Clipboard

Download CSV

Close

Advanced Options:

Extraction Method:

Original

Spreadsheet

Download Data As

sito web → tabella

import.io

The screenshot shows the Wikipedia article page for 'List of best-selling fiction authors'. The URL in the browser is en.wikipedia.org/wiki/List_of_best-selling_fiction_authors. The article title is 'List of best-selling fiction authors'. Below the title, there is a summary: 'From Wikipedia, the free encyclopedia'. The main text states: 'This page provides a list of **best-selling fiction authors** to date and in repeated by reliable sources. "Best-selling" refers to the estimated number of copies of books that have been sold. For a few authors, including Miguel de Cervantes, Alexandre Dumas, Jules Verne and Leon Uris, no exact figure could be found, although the following figures are available:'. Below the text is a table with three columns: 'Author', 'Min. estimated sales', and 'Max. estimated sales'. A green arrow points from this table towards the right side of the image.

Author	Min. estimated sales	Max. estimated sales
William Shakespeare	2 billion ^[2]	4 billion ^[3]
Agatha Christie	2 billion ^[4]	4 billion ^[3]
Barbara Cartland	500 million ^[5]	1 billion ^[6]
Danielle Steel	500 million ^[7]	800 million ^[8]
Harold Robbins	750 million ^[9]	750 million ^[10]
Georges Simenon	500 million ^[11]	700 million ^[12]
Sidney Sheldon	370 million ^[13]	600 million ^[14]

import io

http://en.wikipedia.org/wiki/List_of_best-selling_fiction_authors

#	link_1	link_2	link_3	link_4	link_5
1	William Sh...	[2]	[2]	[3]	Plays
2	Agatha Ch...	[4]	[4]	[3]	Whodunits
3	Barbara C...	[5]	[5]	[6]	Romance
4	Danielle St...	[7]	[7]	[8]	
5	Harold Ro...	[9]	[9]	[10]	Adventure
6	Georges S...	[11]	[11]	[12]	
7	Sidney She...	[13]	[13]	[14]	Suspense
8	Enid Blyton	[15]	[15]	[16]	Children's liter...
9	Dr. Seuss	[17]	[17]	[18]	
10	Gilbert Pat...	[19]	[19]	[20]	

Thank you!

