

**ANALISI DEI DATI: COSA SERVE ALLE  
IMPRESE, COSA SERVE ALLA PUBBLICA  
AMMINISTRAZIONE**

**Machine Learning, Intelligenza Artificiale,  
Mining, Statistica, Modelli di  
simulazione: cosa c'è di vecchio nel nuovo,  
cosa c'è di nuovo nel vecchio**

**ARCHIVIAZIONE, DISSEMINAZIONE E  
RIUSO DEI DATI:  
A CHE PUNTO SIAMO?**

*Uno sguardo all'indietro per andare avanti*

## ARCHIVIAZIONE, DISSEMINAZIONE E RIUSO DEI DATI: A CHE PUNTO SIAMO?

*Alla fine degli anni 60 e nella prima metà dei 70 nell'ambito delle scienze sociali si iniziò anche in Italia a produrre dati con il ricorso a ricerche empiriche (survey in primo luogo, ma anche analisi secondaria di dati archiviati per ragioni amministrative). Sull'onda di questo nuovo interesse nel 1975 fu organizzato su iniziativa della Facoltà di Scienze Politiche il Seminario Analisi Quantitativa per la Ricerca Sociale. Ci si proponeva di diffondere anche nelle scienze sociali l'uso degli allora nuovi strumenti per l'elaborazione elettronica dei dati, oltre che le metodologie di analisi statistica. Contestualmente nasceva l'esigenza della loro conservazione e riutilizzo. Quanta strada abbiamo fatto da allora? Quanta dobbiamo ancora farne? Nell'era dei Big Data e del Machine Learning c'è ancora posto per il Data Curation?*

# ANALISI DEI DATI



## ADABI - MASTER UNIVERSITARIO IN ANALISI DATI PER LA BUSINESS INTELLIGENCE E DATA SCIENCE

### Seminario

#### ANALISI DEI DATI: COSA SERVE ALLE IMPRESE, COSA SERVE ALLA PUBBLICA AMMINISTRAZIONE

**Machine Learning, Intelligenza Artificiale, Mining, Statistica, Modelli di  
simulazione: cosa c'è di vecchio nel nuovo, cosa c'è di nuovo nel vecchio**

*Storicamente nell'analisi dei dati si sono contrapposte due impostazioni: quella deduttiva, in cui le idee delimitano e guidano l'elaborazione dei dati (top-down); quella induttiva, in cui le idee nascono e sono guidate dai dati (bottom-up o data driven). Si pone ancora oggi la questione? Se sì, in quali termini? Quali sono le esigenze delle imprese?*

*I professori Rosa Meo del dipartimento di Informatica e Roberto Leombruni del dipartimento di Economia e Statistica "Cognetti de Martiis" interverranno sul tema e ne discuteranno con noi. Il tema sarà presentato da Flavio Bonifacio, project manager del Master che coordinerà gli interventi che seguiranno.*

# ANALISI DEI DATI

9.30 Saluti del Direttore del Master ADABI – **Prof. Mariateresa Giraudò**

## **Introduzione**

9.35 **Flavio Bonifacio**, CEO Metis Ricerche srl, project manager del Master ADABI, *La diffusione della Business Intelligence tra le imprese*

## **Relazioni**

10.00 **Rosa Meo**, Dipartimento di Informatica, Università di Torino, *Il Machine Learning: motore di innovazione guidato dai dati*

10.40 **Roberto Leombruni**, Dipartimento di Economia e Statistica "Cognetti de Martiis", Università di Torino, *La Statistica, il ritorno: dalle schede perforate ai big data*

11.20 **Interventi del pubblico**



Per informazioni

Segreteria Metis Ricerche, tel. 011.538028; e-mail: [segreteria@metis-ricerche.it](mailto:segreteria@metis-ricerche.it)

# ANALISI DEI DATI

**Il pubblico**

**Il Master Adabi**

**Il contesto di riferimento del Master**

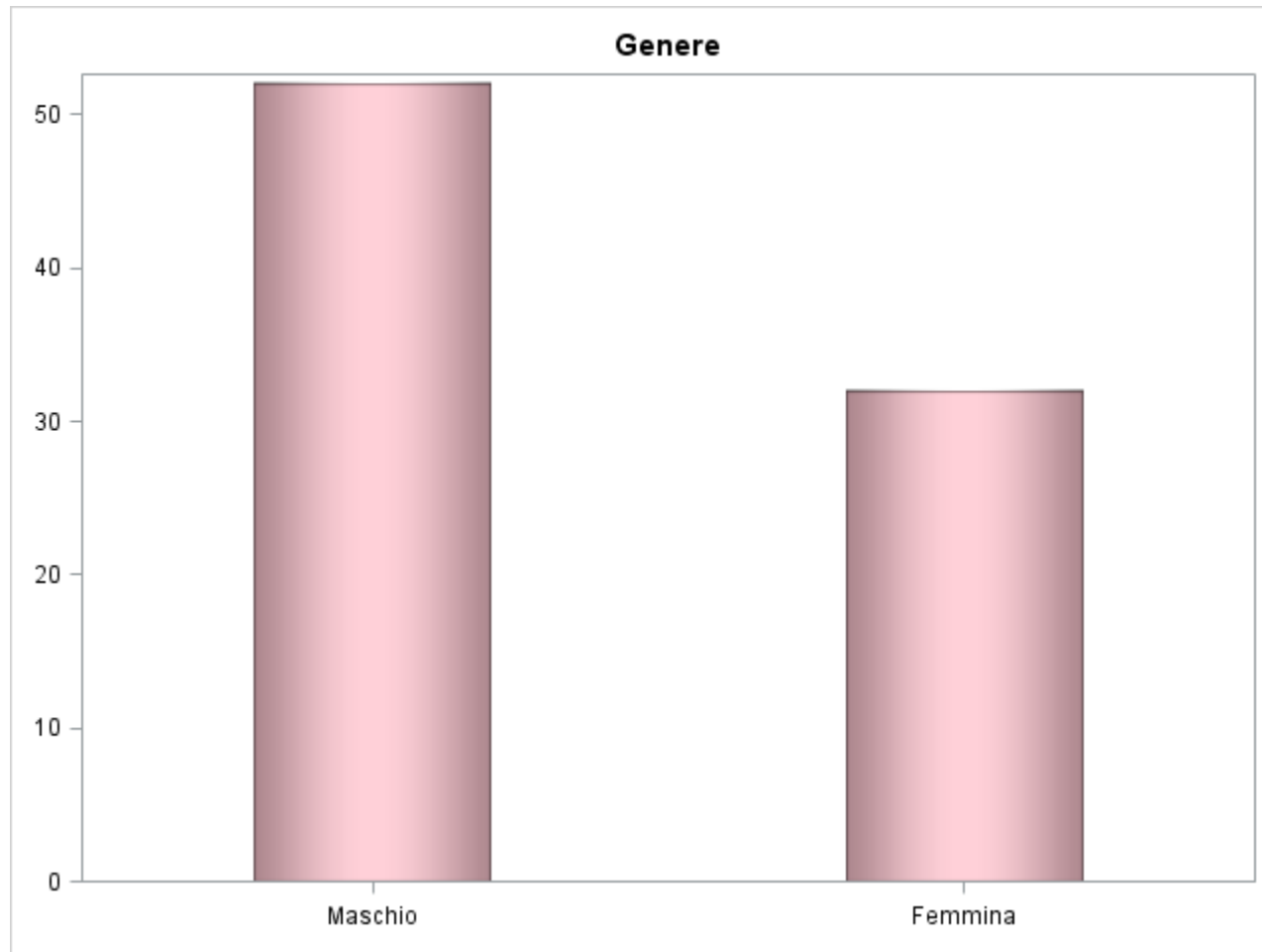
**Le idee e le cose**

**Quale idea abbiamo degli strumenti della  
Business Intelligence**

# ANALISI DEI DATI

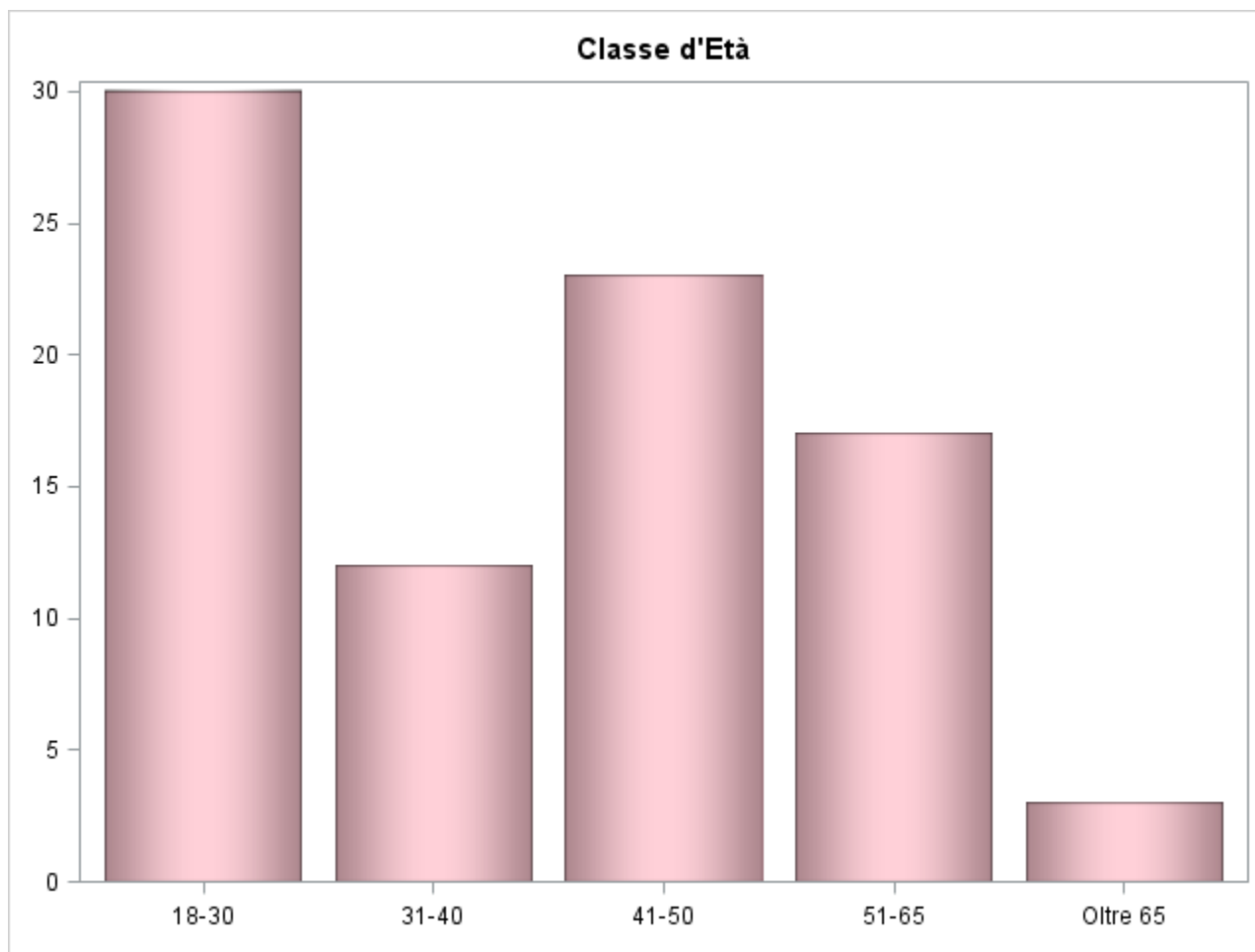
**Il pubblico (campione di 84 persone)**

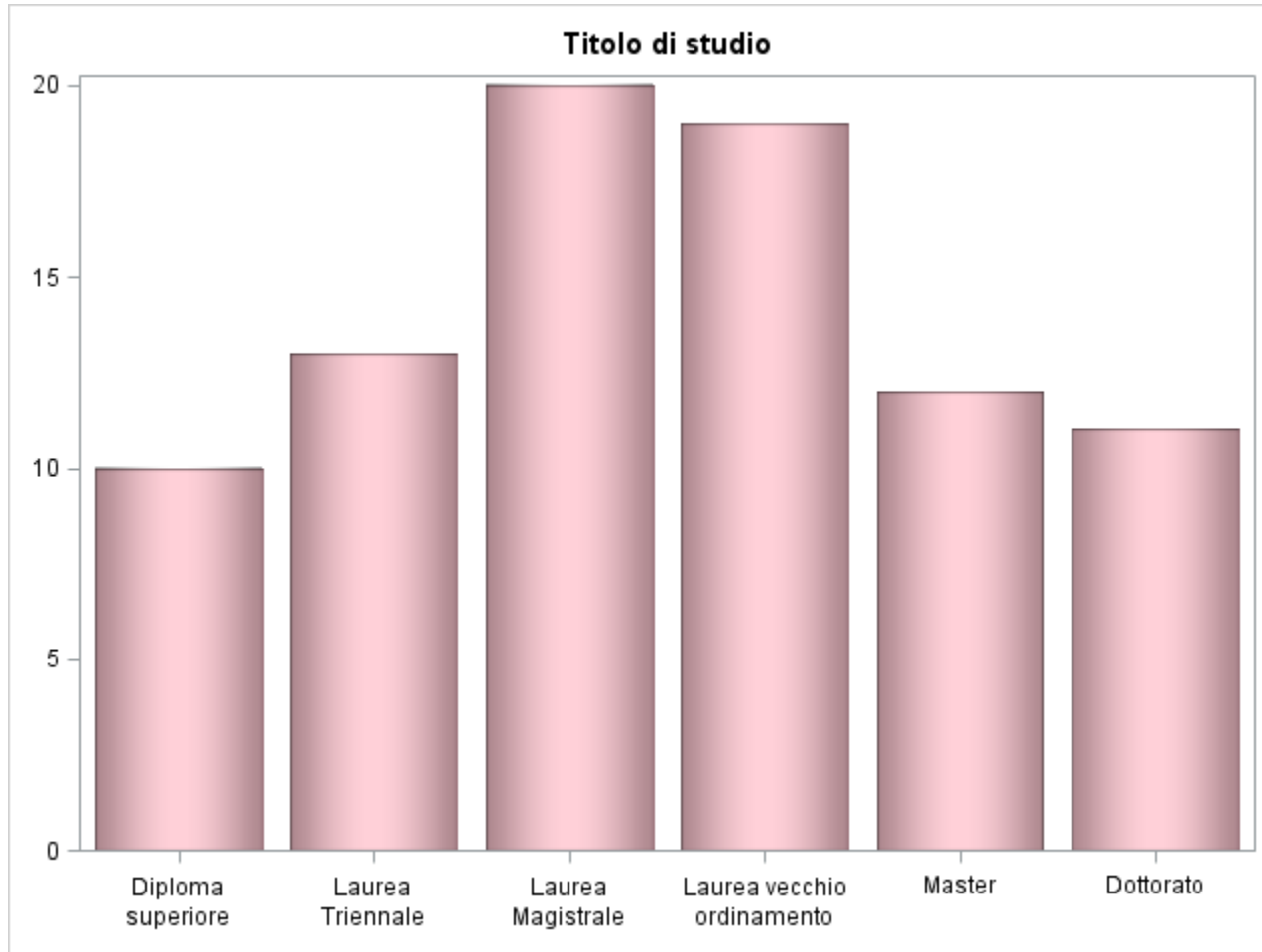
# ANALISI DATI

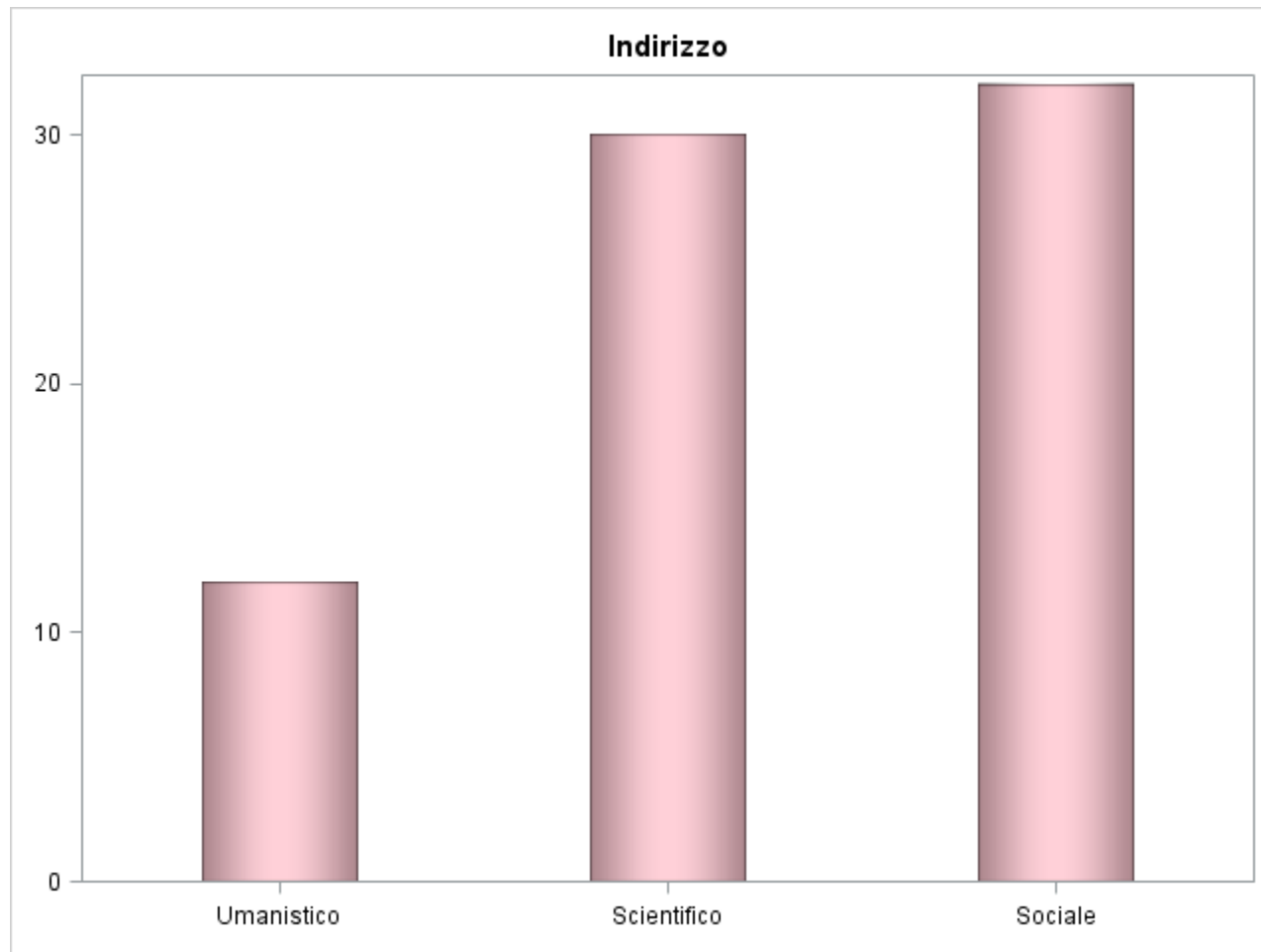


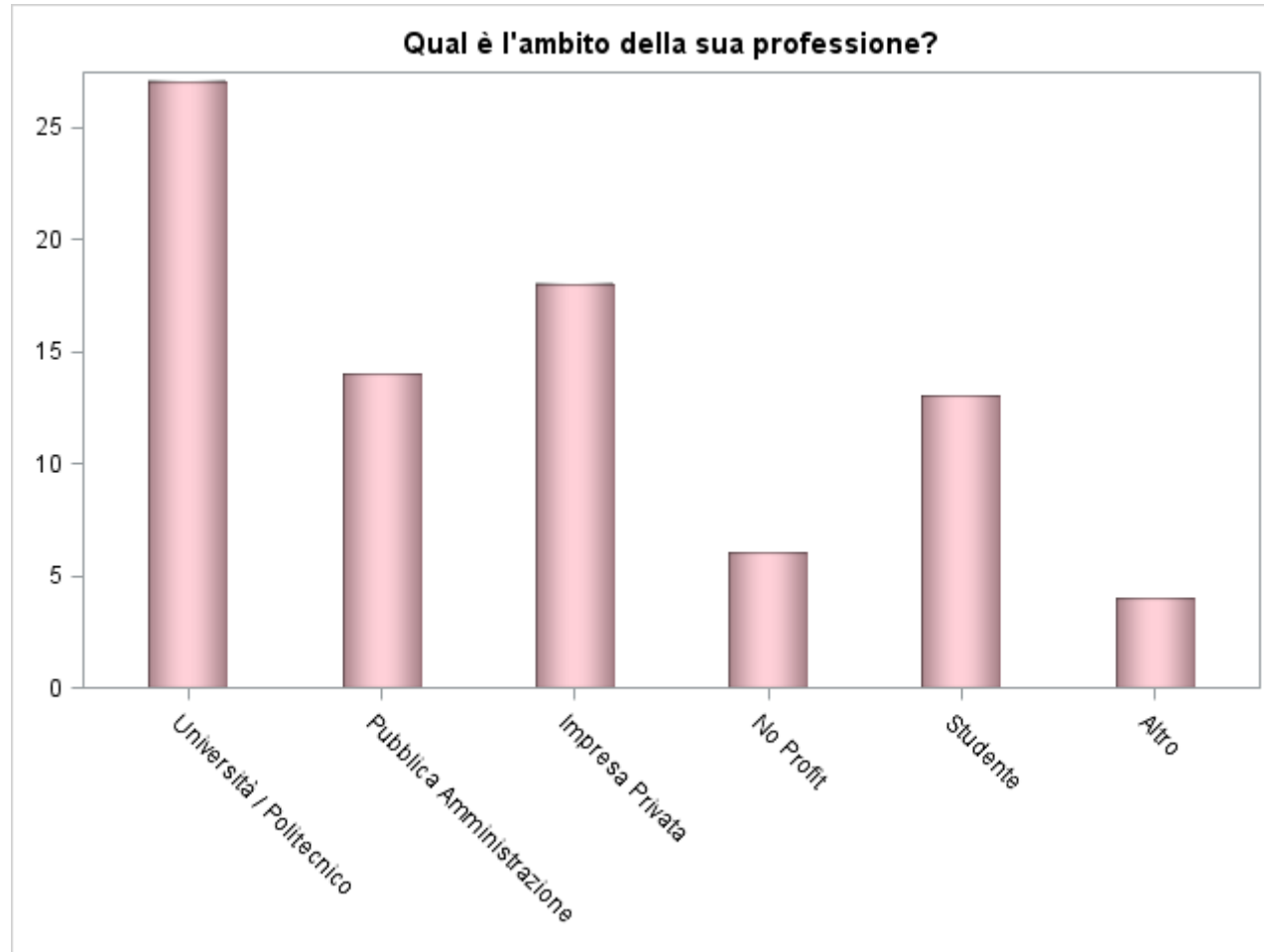


# ANALISI DATI

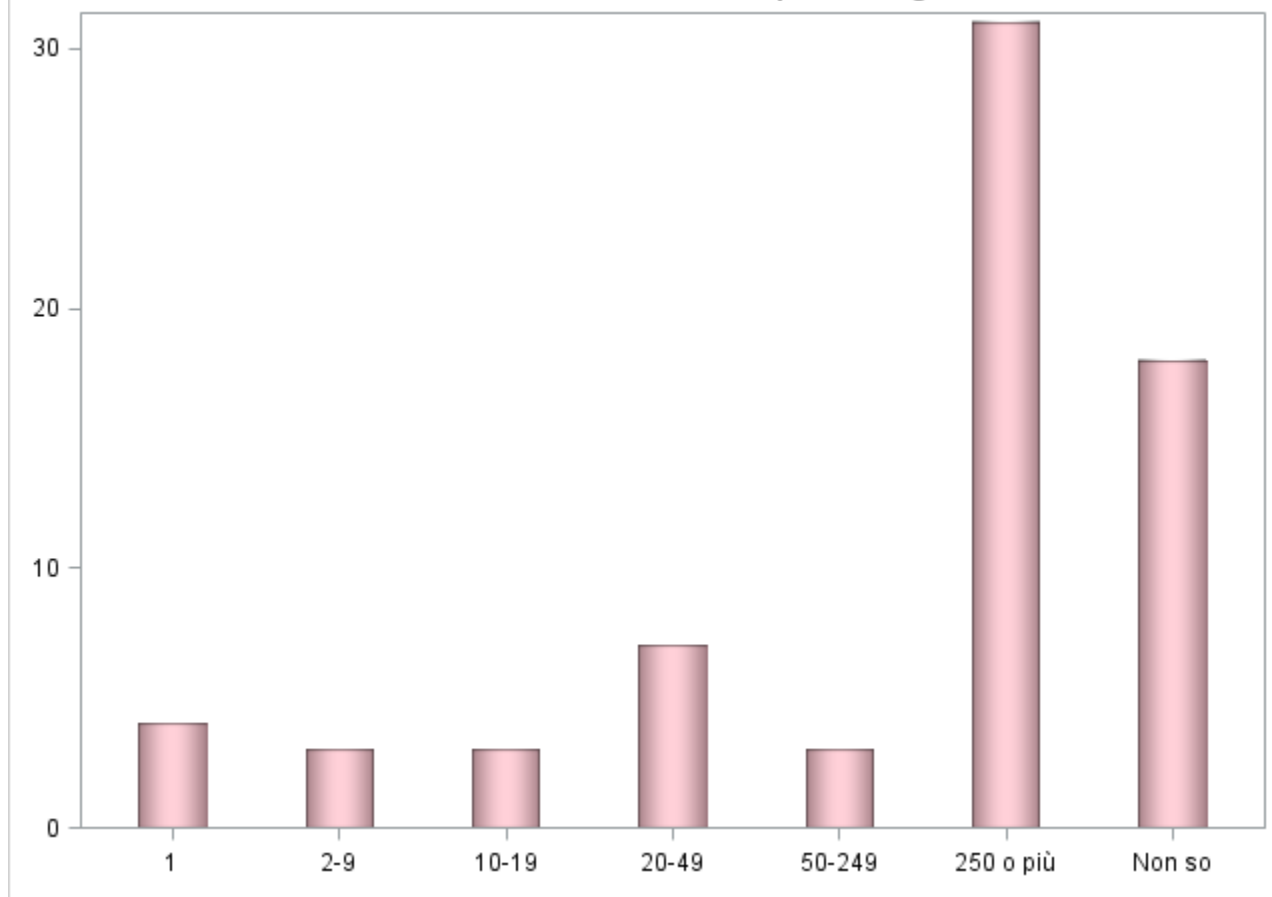




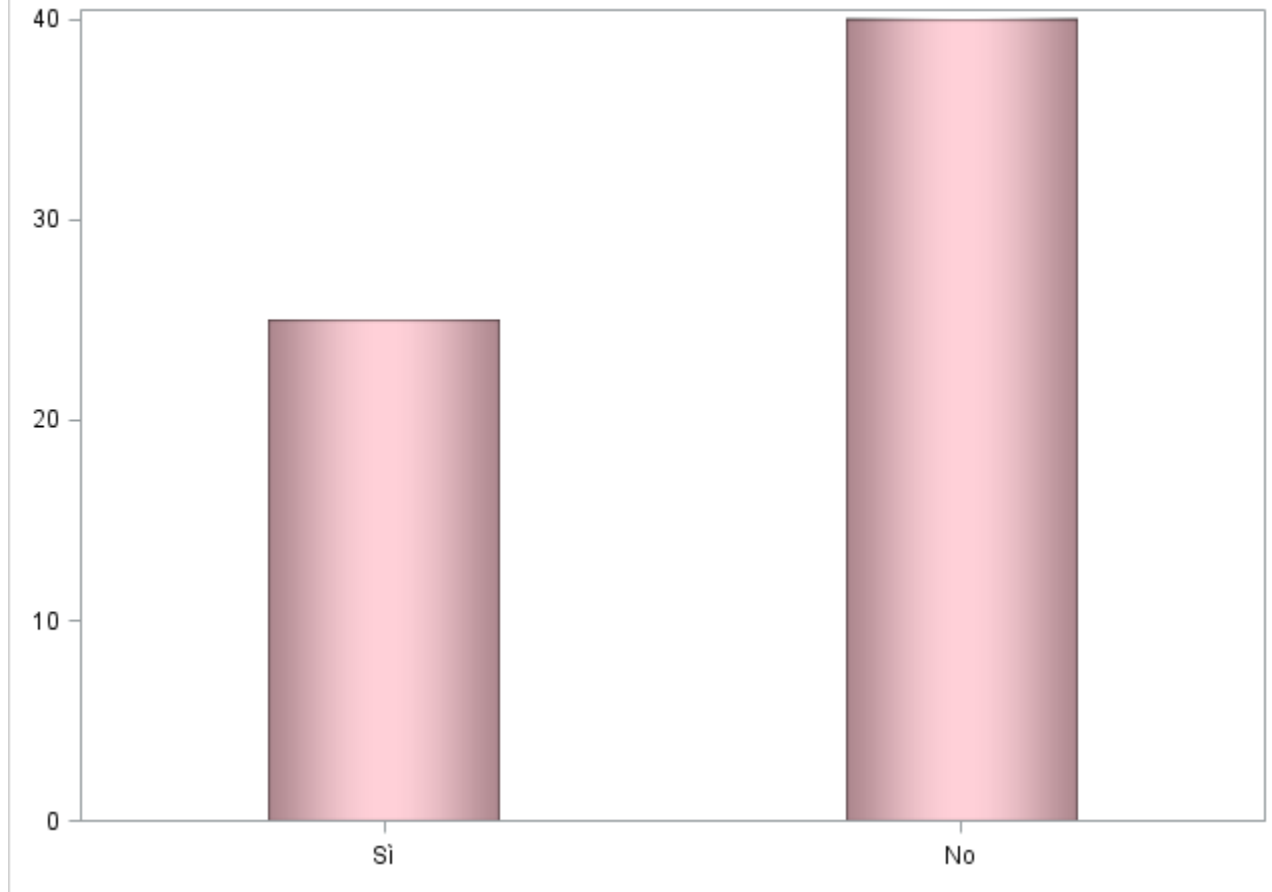




Qual è il numero di addetti della sua impresa / organizzazione?



### La sua impresa / organizzazione è specializzata nel settore ICT?



# ANALISI DEI DATI

*Dal 2012...*

**Master Adabi**

# ANALISI DEI DATI



[HOME](#) [IL MASTER](#) [PROGRAMMA](#) [ISCRIZIONI](#) [ADABI EXPERIENCE](#) [CONTATTACI](#)



UNIVERSITÀ  
DEGLI STUDI  
DI TORINO

IL MASTER È ATTIVATO DAL DIPARTIMENTO DI CULTURE POLITICA E SOCIETÀ - DIPARTIMENTO DI INFORMATICA - DIPARTIMENTO DI MATEMATICA "GIUSEPPE PEANO"  
DIPARTIMENTO DI ECONOMIA E STATISTICA "COGNETTI DE MARTIIS" DELL'UNIVERSITÀ DEGLI STUDI DI TORINO

PROGETTATO CON:



GESTITO DA:



UNIVERSITÀ DEGLI STUDI DI TORINO  
MASTER DI I LIVELLO IN ANALISI DATI PER LA  
BUSINESS INTELLIGENCE E DATA SCIENCE

NEWS

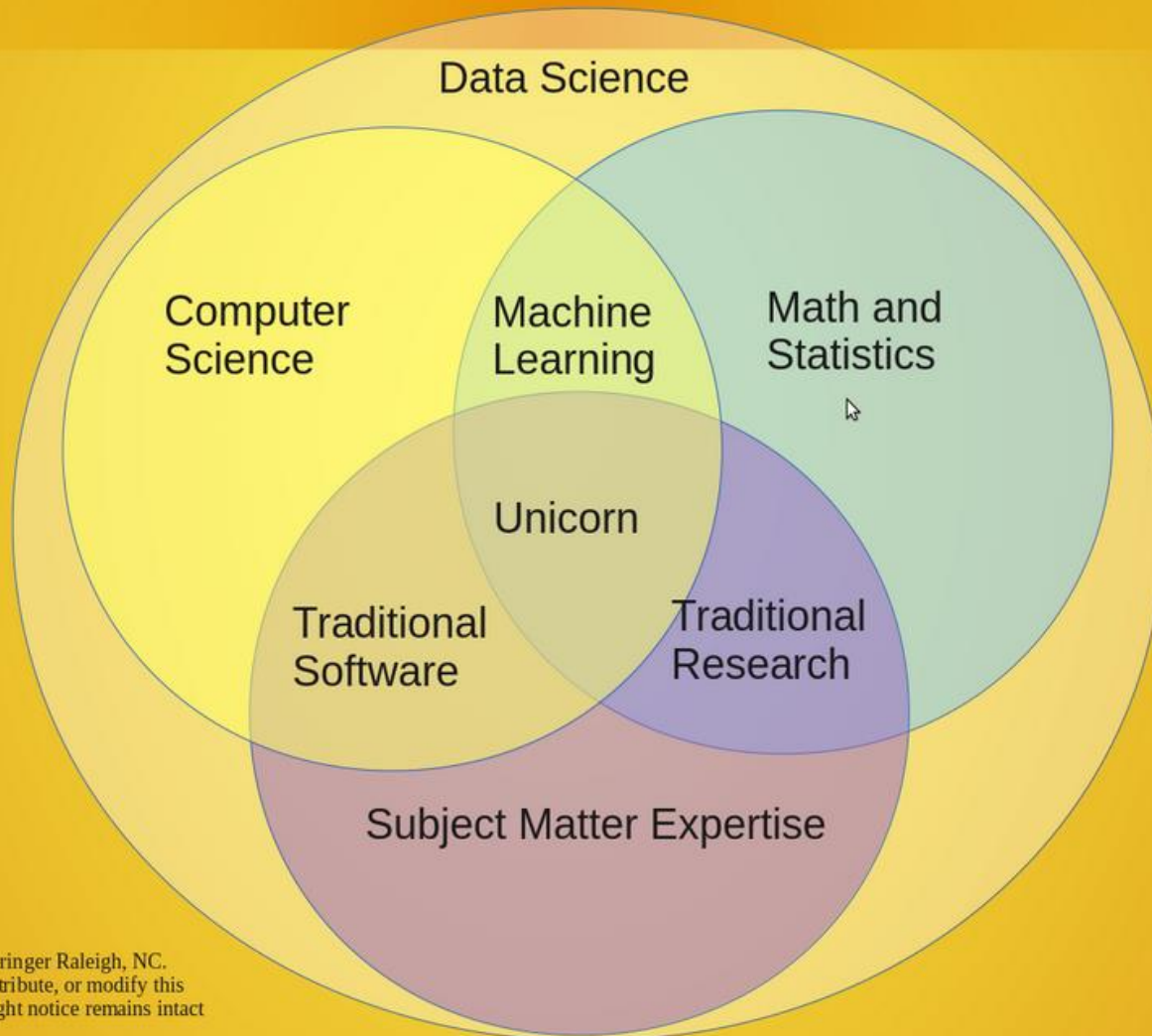
VI EDIZIONE A.A. 2017/2018

ADABI è un Master di I livello dell'Università degli Studi di Torino. Forma professionisti specializzati nell'utilizzo delle tecniche di **Data Science** per tutte le aziende ed enti che posseggono, per la natura della loro Mission, grandi masse di dati. Il Master in Analisi Dati per la Business Intelligence e Data Science ha **carattere pragmatico**. Durante il Master si studieranno metodologie di **Data Quality** e **Data Management**, di **Analisi Statistica dei Dati**, di **Modellazione (Analytics)**, **Segmentazione** e **Scoring**, da un punto di vista operativo, con l'utilizzo di software statistico proprietario (**SAS®**) e open source (Knime, R, Python). Altre importanti keywords che caratterizzano questo Master sono **Data Mining**, **Machine Learning** e **algoritmi di classificazione** (Cluster Analysis, Decision Tree) e di sistemi dell'informazione statistica (**Expert Analysis**).

<http://www.masteradabi.it/>



# ANALISI DEI DATI



Copyright © 2014 by Steven Geringer Raleigh, NC.  
Permission is granted to use, distribute, or modify this  
image, provided that this copyright notice remains intact

# ANALISI DEI DATI

**Interdisciplinarietà**

**Teoria Empirica (si impara facendo)**

**Intergenerazionalità**

**Due anime: pubblica  
(università, csi) e privata  
(imprese)**

# ANALISI DEI DATI

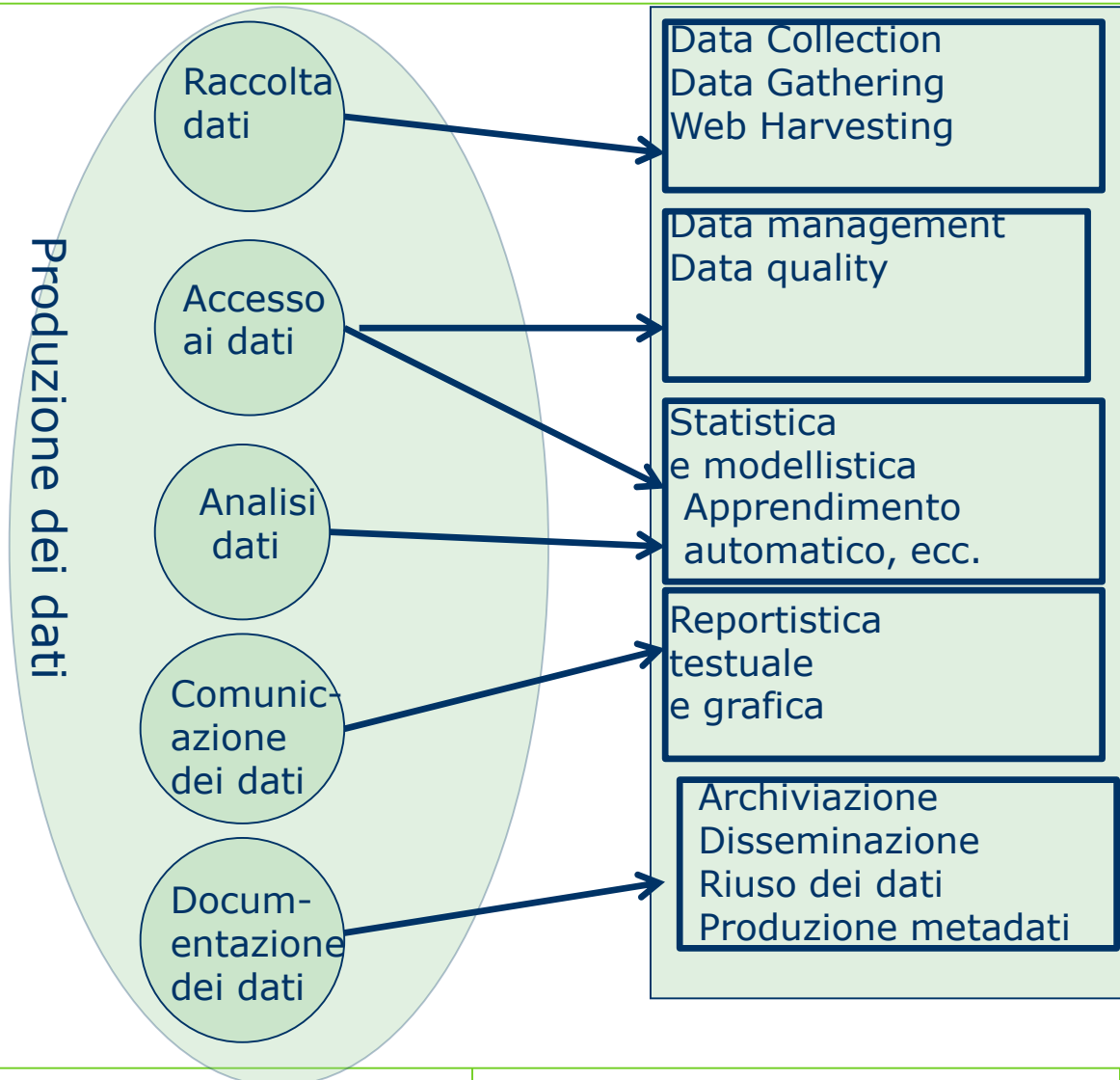
**Data Scientists, the “unicorn” of data: what are they, what do they do and how will they change the world?**

## ANALISI DEI DATI

**In 1962, John Tukey wrote about a shift in the world of statistics, saying, "... as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt...I have come to feel that my central interest is in data analysis..." Tukey is referring to the merging of statistics and computers, at a time when statistical results were presented in hours, rather than the days or weeks it would take if done by hand.**

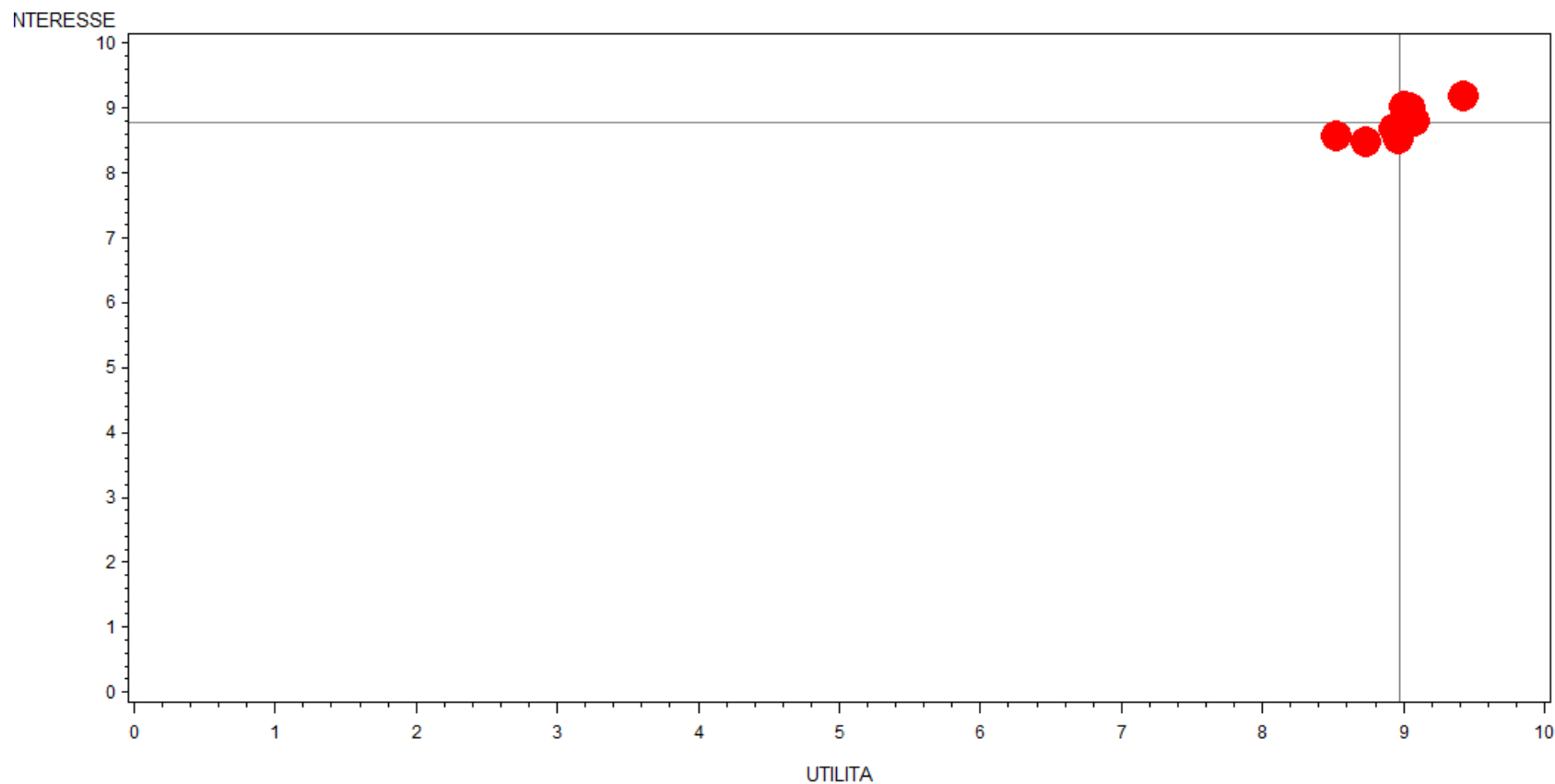
**"The best thing about being a statistician," Mr. Tukey once told a colleague, "is that you get to play in everyone's backyard."**

# ANALISI DEI DATI



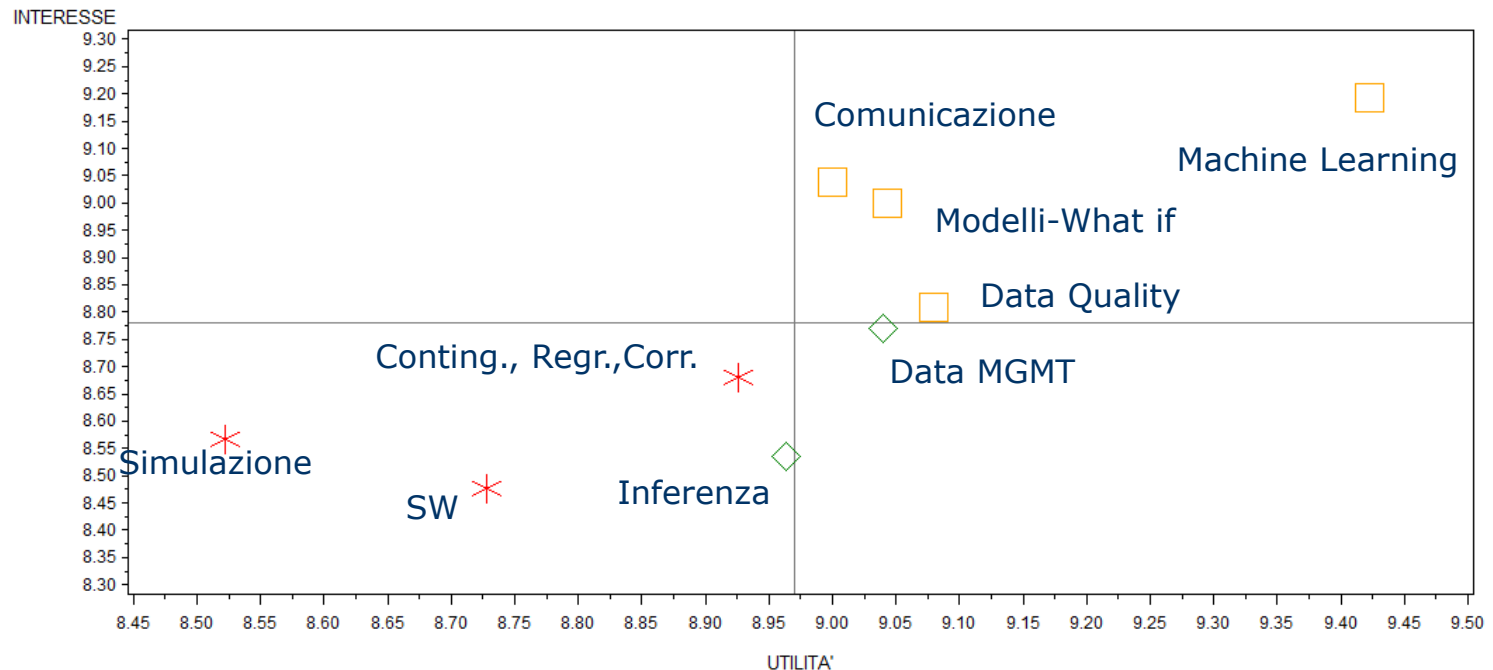
# ANALISI DEI DATI

## Valutazione dei moduli del Master



# ANALISI DEI DATI

## Valutazione dei moduli del Master - zoom



Università/Politecnico (9.86,9.86), n=7

Imprese Private (8.93,9.33), n=17

PA=(8.50,8.50), n=3

# ANALISI DEI DATI

## Il contesto di riferimento del Master



# ANALISI DEI DATI

## Il contesto di riferimento del Master

Per quanto riguarda i titoli di studio il Censis nota ancora una volta che rispetto all'Europa i laureati italiani sono il 13% in meno anche se dal 2014 al 2017 passano dal 23,9 al 26,9% (dal 37,9 al 39,9 in Europa). Gli studenti iscritti all'Università nel 2016-2017 erano 1.654.680, immatricolati 274.339, laureati e diplomati 311.799. Gli iscritti ai Master di primo livello erano (2015-2016) 23.396, ai Master di secondo livello 15.343

*Ci sono delle ragioni però, non rimediabili con semplici, piccoli, ma progressivi e continui spostamenti d'asticella*

Al di là degli squilibri strutturali e territoriali di un'offerta troppo concentrata, nota ancora il Censis, sui percorsi accademici per poter interessare una platea più ampia di studenti, è soprattutto la perdita di valore del titolo di studio superiore che fa male al paese: un titolo che non rende abbastanza dal punto di vista economico, non crea mobilità sociale, non trova sbocchi lavorativi adeguati e sta progressivamente perdendo il suo peso sociale

# ANALISI DEI DATI

Si può leggere la risposta a questo stato di cose con la sempre maggiore disponibilità dei giovani Italiani a muoversi in Europa. L'interscambio con altre università europee è giudicata molto importante da circa il 70% degli Italiani (53% gli europei), come si apprende da dati recenti di Eurbarometro

# ANALISI DEI DATI

## L'economia digitale

La punta della freccia dell'economia digitale in Italia è oggi costituita da una sorta di élite del lavoro che conta circa 750.000 professionisti e specialisti (3% dell'occupazione totale), in crescita del 4,9% tra il 2015 e il 2016 (sul totale dell'occupazione + 1,3%) e del 12,2% rispetto a 6 anni prima (+0,7 l'occupazione totale).

Nello stesso tempo si stima che la domanda inevasa di professioni digitali superi le 60.000 unità.

Le imprese del settore ICT che organizzano corsi di formazione sono il 49,8 % (13% se si considerano tutte le imprese)

Le *vacancies* di professionalità ICT possono solo in parte essere coperte dal sistema di formazione terziaria, se il totale dei laureati in discipline scientifiche con contenuti Digital non raggiunge attualmente le 40.000 unità

Tutto ciò in un paese che nel 2017 contava 25.930.000 unità di forze di lavoro, 23.023.000 occupati e 2.907.000 in cerca di prima occupazione e dove tra il 2007 e il 2017 la componente degli occupati con età compresa tra i 25 e i 34 anni si è ridotta del 27,3%, pari a più di un milione e mezzo di giovani

## ANALISI DEI DATI

### La crescita dell'economia digitale

- 113.000 imprese digitali attive al primo semestre del 2018 cresciute del 18,4% dal 2012
- Raddoppiate le imprese attive nel settore del commercio al dettaglio via internet nell'arco di 6 anni
- Fatturato del mercato digitale in aumento del 2,3%
- Acquisti online aumentati del 16,9% (+1,5% i consumi delle famiglie)

## **ANALISI DEI DATI**

### **CITTADINI, IMPRESE E ICT, Istat, Dicembre, 2017**

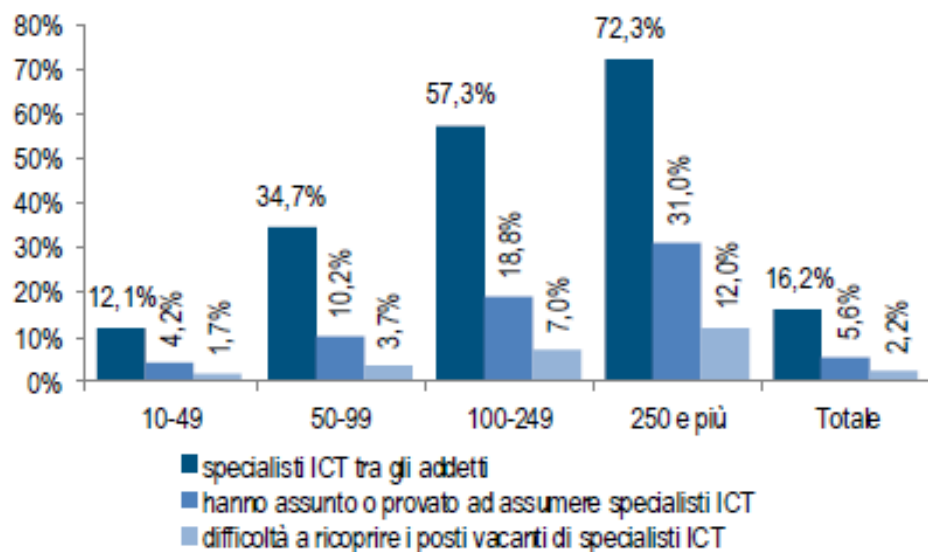
**Il quadro informativo integrato sull'utilizzo delle tecnologie ICT da parte di cittadini e imprese mostra un impiego sempre più diffuso ed evoluto di queste tecnologie nelle attività economiche e nella vita quotidiana. Tuttavia, rimane un gap rilevante dell'Italia rispetto agli altri paesi Ue**

# ANALISI DEI DATI

## Specialisti ICT nelle imprese

FIGURA 4. IMPRESE CHE IMPIEGANO SPECIALISTI ICT O CHE, NELL'ANNO, HANNO ASSUNTO O PROVATO AD ASSUMERE PERSONALE CON COMPETENZE ICT PER CLASSE DI ADDETTI. Anno 2017, valori percentuali sul totale delle imprese con almeno 10 addetti

FIGURA 4. IMPRESE CHE IMPIEGANO SPECIALISTI ICT O CHE, PERSONALE CON COMPETENZE ICT PER CLASSE DI ADDETTI. Ann



# ANALISI DEI DATI

12 CARATTERISTICHE INCLUSE NELL'INDICATORE DI DIGITALIZZAZIONE (a)	CLASSE DI ADDETTI				Totale
	10-49	50-99	100-249	250 e più	
1. la percentuale di addetti che utilizzano computer connessi è superiore al 50%	32,8	35,1	39,5	41,3	33,3
2. presenza di specialisti ICT tra gli addetti dell'impresa	12,1	34,7	57,3	72,3	16,2
3. velocità di download della connessione almeno pari a 30 Mbit/s	20,1	29,8	40,9	58,3	22,1
4. la percentuale di addetti che utilizzano device mobili connessi è superiore al 20%	29,8	28,3	32,8	33,4	29,9
5. utilizzo di sito web dell'impresa	70,4	82,9	86,4	89,1	72,1
6. il sito web dell'impresa offre almeno uno tra i servizi di accesso a cataloghi; tracciabilità ordinazioni; personalizzazione prodotti; personalizzazione contenuti	35,2	46,7	47,4	51,2	36,6
7. utilizzo di social media	42,8	48,6	55,1	66,4	44,0
8. utilizzo di software ERP	32,5	61,6	70,7	79,5	36,5
9. utilizzo di software CRM	29,1	41,6	44,8	53,0	30,9
10. gestione informatica della catena distributiva SCM	10,1	15,4	22,9	34,9	11,3
11. valore delle vendite online almeno pari all'1% dei ricavi totali	7,4	11,1	16,1	25,1	8,2
12. la percentuale delle vendite via web è maggiore dell'1% dei ricavi e quelle del tipo B2C sul totale delle vendite via web è superiore al 10%	4,7	4,8	4,8	6,7	4,8

(a) Indicatori 2. 8. 9. 10 modificati rispetto al 2015 e al 2016

## ANALISI DEI DATI

■ È ampio il divario tra grandi e piccole imprese nel livello di digitalizzazione misurato dall'indicatore composito *Digital intensity indicator*: che, per livelli almeno alti di digitalizzazione, è pari all'11% per le piccole imprese e raggiunge il 47% tra le grandi.



## ANALISI DEI DATI

### Le famiglie

**Il 65,3% delle persone di 6 anni e più si è connesso alla Rete negli ultimi 12 mesi (63,2% nel 2016), mentre circa il 47,6% accede tutti i giorni. Nell'uso di Internet l'età è ancora il principale fattore discriminante: sono i giovani a utilizzarlo di più (oltre il 92% dei 15-24enni) ma la crescita è significativa anche tra i 55-59enni (da 62,7% a 68,2%).**

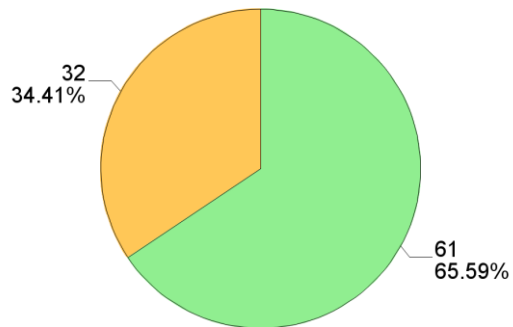
# ANALISI DEI DATI

PROSPETTO 1. PERSONE DI 6 ANNI E PIÙ CHE HANNO USATO INTERNET NEGLI ULTIMI 12 MESI PER SESSO, CLASSE DI ETÀ, TERRITORIO E CONDIZIONE OCCUPAZIONALE (a). Anni 2011-2017, valori per 100 persone di 6 anni e più con le stesse caratteristiche.

SESSO, CLASSI DI ETÀ, RIPARTIZIONI GEOGRAFICHE E CONDIZIONE OCCUPAZIONALE	2011	2012	2013	2014	2015	2016	2017
Maschi	56,7	58,3	60,3	62,4	65,0	67,6	69,5
Femmine	46,7	47,1	49,8	52,8	55,8	59,0	61,3
<b>Totale</b>	<b>51,5</b>	<b>52,5</b>	<b>54,9</b>	<b>57,5</b>	<b>60,2</b>	<b>63,2</b>	<b>65,3</b>
6-10	38,3	40,8	45,1	44,6	43,8	48,2	53,7
11-14	78,1	76,5	80,8	80,9	80,4	82,9	86,1
15-17	89,1	88,5	89,7	91,2	92,0	91,6	93,9
18-19	88,8	88,8	90,0	93,9	92,0	93,0	92,7
20-24	85,8	86,0	85,7	89,3	90,7	91,3	92,4
25-34	77,5	79,2	80,3	83,9	85,1	87,1	89,8
35-44	69,7	69,1	73,5	76,1	80,1	84,3	85,7
45-54	56,2	58,7	61,7	66,1	70,0	75,1	77,6
55-59	42,2	45,3	48,5	52,0	60,4	62,7	68,2
60-64	28,6	31,0	36,0	41,1	45,9	52,2	56,0
65-74	13,8	16,4	19,0	21,2	25,6	28,8	30,8
75 e più	2,7	3,3	3,5	4,4	6,7	7,7	8,8
<b>Totale</b>	<b>51,5</b>	<b>52,5</b>	<b>54,9</b>	<b>57,5</b>	<b>60,2</b>	<b>63,2</b>	<b>65,3</b>
Nord-ovest	56,4	57,3	58,3	61,5	64,6	67,6	69,1
Nord-est	55,9	57,7	60,1	61,5	65,2	66,9	68,0
Centro	54,5	55,1	57,8	60,4	61,6	66,4	67,8
Sud	43,6	43,3	46,6	49,2	53,1	55,9	59,1
Isole	43,9	47,3	49,8	53,0	53,8	55,7	59,6
<b>Italia</b>	<b>51,5</b>	<b>52,5</b>	<b>54,9</b>	<b>57,5</b>	<b>60,2</b>	<b>63,2</b>	<b>65,3</b>

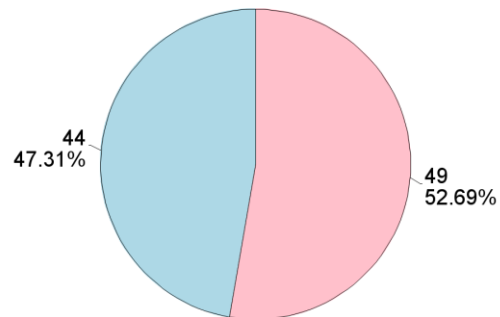
# ANALISI DEI DATI

ISCRITTI



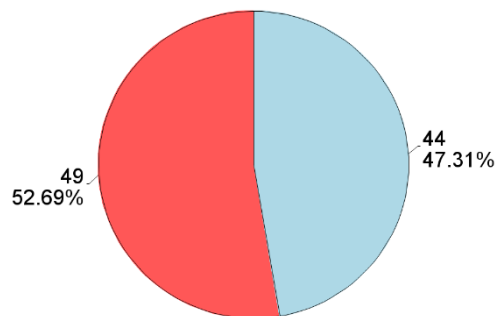
■ Più di 33 anni 
 ■ Meno di 33 anni

ISCRITTI



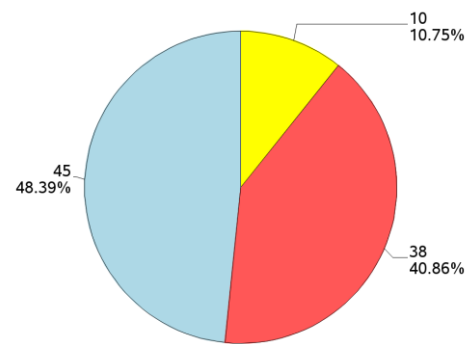
■ Uomo 
 ■ Donna

ISCRITTI



■ Non Occupato 
 ■ Occupato

ISCRITTI



■ SCIENTIFICA 
 ■ SOCIALE 
 ■ UMANISTICA

# ANALISI DEI DATI

## Le idee e le cose

# ANALISI DEI DATI

SAS Culture

Data don't exist apart from the analytics that makes sense of it



## ANALISI DEI DATI

***If we are blinded by darkness, we are also blinded by light. When too much light falls on everything, a special terror results.***  
**Annie Dillard**

# ANALISI DEI DATI

Blinded by the Light <https://www.youtube.com/watch?v=7Iaca30QbOo>

Bruce Springsteen

...

E lei fu accecata dalla luce  
lacerata, persa come un  
peccato, un'altra fuggitiva  
nella notte  
accecata dalla luce  
caduta così in basso da non  
riuscire più a trovarsi, ma  
riuscirà a risolvere tutto

...

...

And she was blinded by the  
light  
Cut loose like a deuce  
another runner in the night  
Blinded by the light  
She got down but she never  
got tight, but she'll make it  
alright

...



#AnalyticsX

sas

**INTERDISCIPLINARY  
INTERDEPENDENCE  
INTERSPACES**

E

X

ANALYTICS EXPERIENCE



cogito  
ergo  
sum



ANALYTICS EXPERIENCE

# ANALISI DEI DATI



## ANALISI DEI DATI

Il bando dato all' *immaginazione* non trovava compenso in nessuna delle superiori certezze conquistate dagli antichi metodi di investigazione. L' errore...era analogo a quello dell' idiota che s' immagina di vedere un oggetto tanto meglio quanto piu' se lo avvicina agli occhi. **Quando si (gli antichi, NdA) regolavano secondo Hog, i loro "fatti" erano sempre ed in ogni caso dei fatti, cosa di ben poca conseguenza anche a presumere che erano fatti e dovevano essere fatti perché apparivano tali. Quando si regolavano secondo l' altro, l' Ariete, il loro ragionamento non era più dritto di quello che può essere dritto un corno dell' animale in questione, giacché mai hanno enunciato un assioma che fosse veramente un assioma...**In ogni caso non credete...che tutti quei dogmatici si sarebbero trovati in imbarazzo se avessero dovuto stabilire a quale dei loro due metodi si doveva la più importante, l' eccelsa tra le loro verità, quella, voglio dire, della legge della gravitazione? Newton la doveva a Kepler. E Kepler ammetteva di aver *indovinato* le tre leggi capitali che condussero il più grande dei matematici inglitch alla scoperta del suo principio, su cui ogni principio fisico si fonda, l' unico che permette di penetrare nel reame della metafisica. Kepler indovinò, cioè a dire *immaginò..*"[Edgar Allan POE, Eureka, 1848]."

# ANALISI DEI DATI

***Le piattaforme tecnologiche hanno imparato a loro spese che per gestire casi complessi non basta mescolare dati e tecnologie, ma serve fiducia nelle capacità dell'Uomo***

Censis, 52° Rapporto sulla situazione sociale del paese, Dicembre 2018

# ANALISI DEI DATI

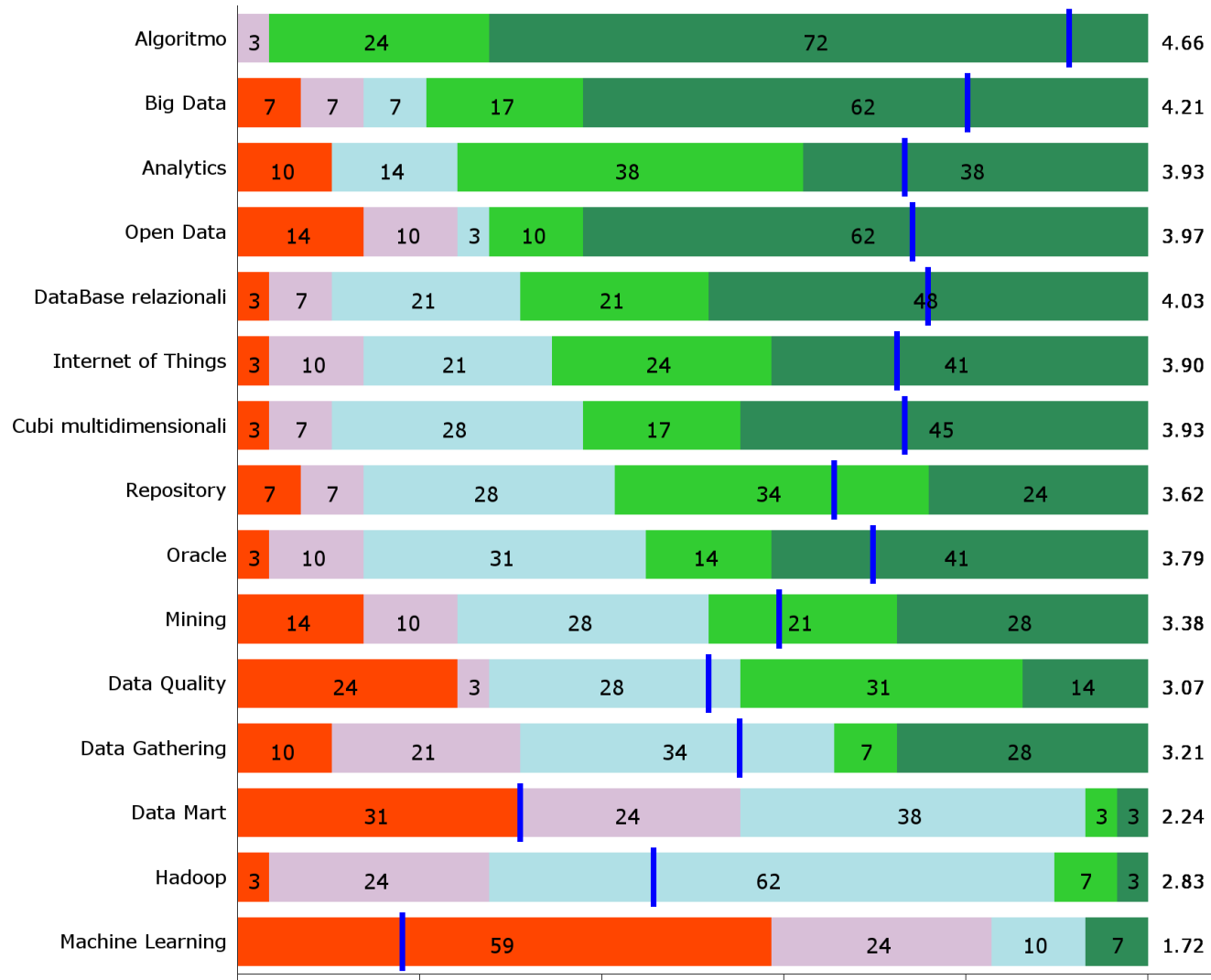
**Quale idea abbiamo degli strumenti della  
Business Intelligence**

# ANALISI DEI DATI

Di seguito sono riportati alcuni termini di uso comune nella Scienza dei dati. Per ognuno vengono riportati due possibili descrizioni. Dovreste indicare con valori da 1 a 5 quale descrizione si avvicina di più al significato del termine. Valori 1 e 2 indicano che propendete per il termine di sinistra, 4 e 5 per il termine di destra. Il valore 3, intermedio, implica che ritenete equivalenti i termini e che preferite non pronunciarvi. Indicate per favore un solo valore.

Analytics	Strumenti per la diagnostica di incongruenze nei dati	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Strumenti di analisi statistica
-----------	---	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	---------------------------------

# ANALISI DEI DATI



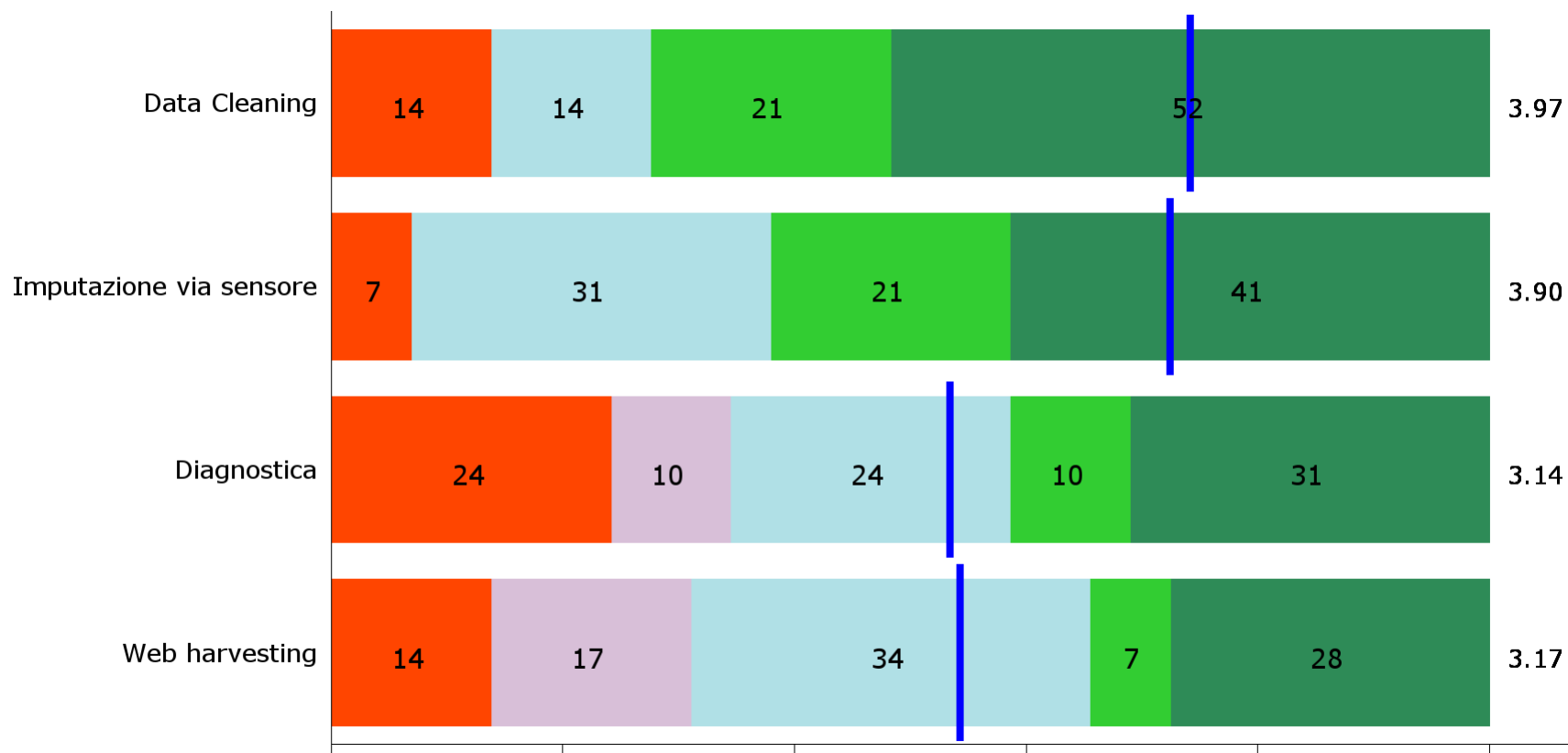
# ANALISI DEI DATI

D16 Di seguito sono riportati alcuni termini di uso comune in ambito Data Gathering/Data Quality. Per ognuno vengono riportati due possibili descrizioni. Dovreste indicare con valori da 1 a 5 quale descrizione si avvicina di più al significato del termine. Valori 1 e 2 indicano che propendete per il termine di sinistra, 4 e 5 per il termine di destra. Il valore 3, intermedio, implica che ritenete equivalenti i termini e che preferite non pronunciarvi. Indicate per favore un solo valore.

Data Cleaning	Procedure per analizzare i dati	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Procedure per correggere i dati
---------------	---------------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	---------------------------------



# ANALISI DEI DATI

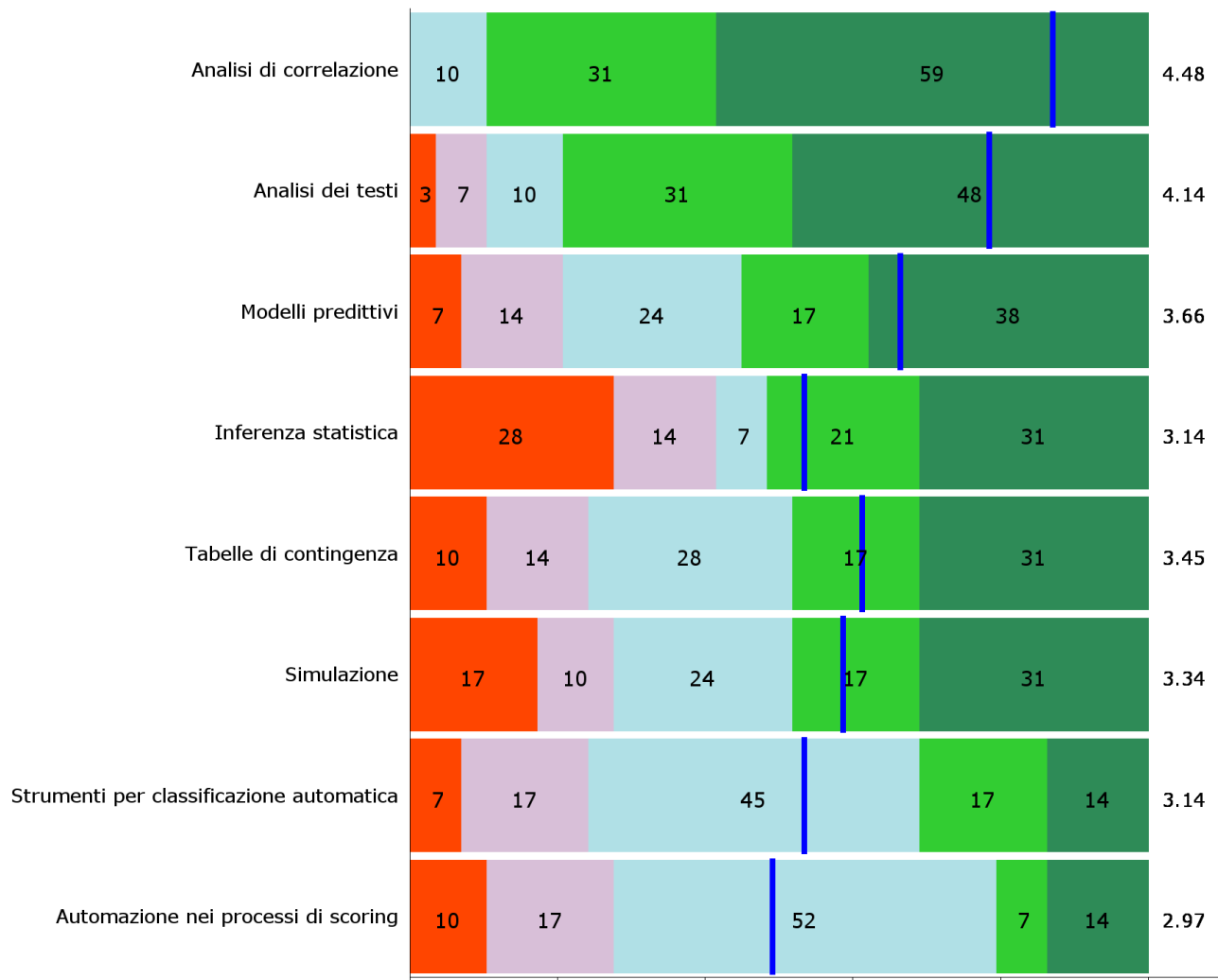


# ANALISI DEI DATI

Di seguito sono riportati alcuni termini di uso comune in ambito Analytics. Per ognuno vengono riportati due possibili descrizioni. Dovreste indicare con valori da 1 a 5 quale descrizione si avvicina di più al significato del termine. Valori 1 e 2 indicano che propendete per il termine di sinistra, 4 e 5 per il termine di destra. Il valore 3, intermedio, implica che ritenete equivalenti i termini e che preferite non pronunciarsi. Indicate per favore un solo valore.

Inferenza statistica	Deduzione di regole per l'analisi dei dati	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Regole per l'uso corretto di dati campionari
----------------------	--	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--

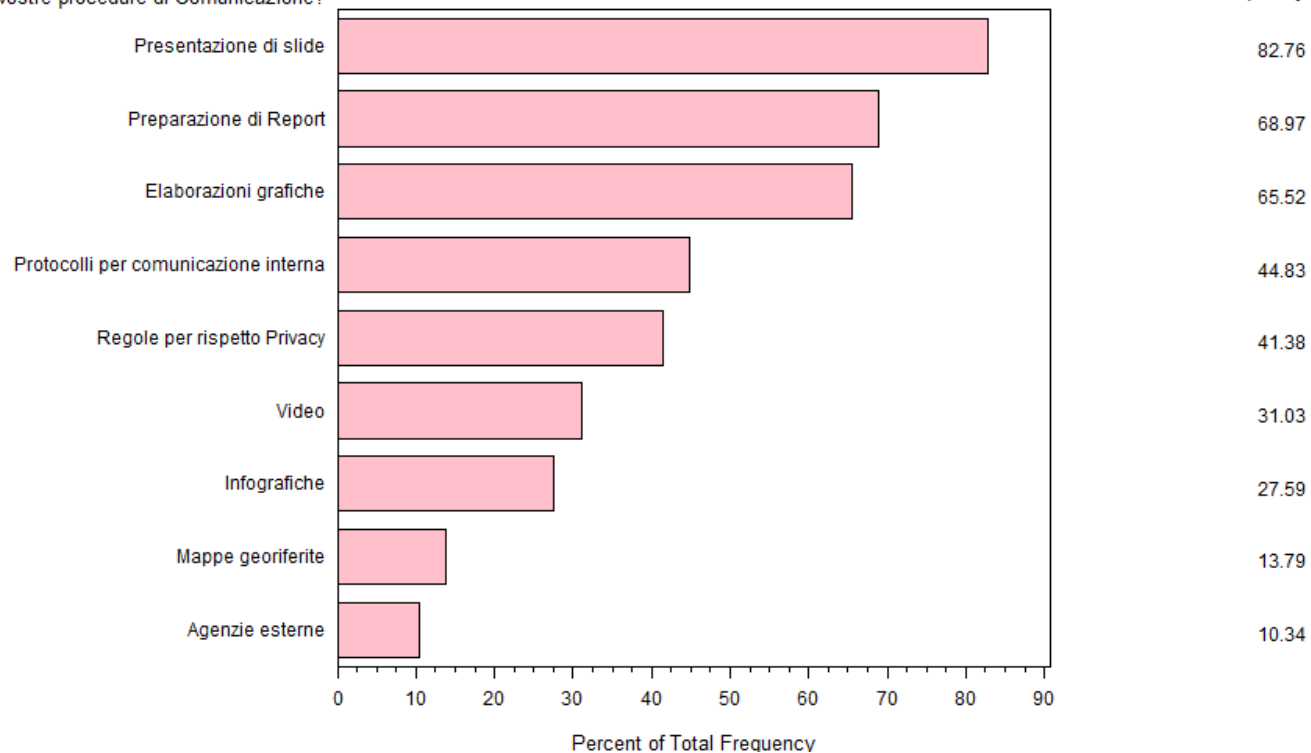
# ANALISI DEI DATI



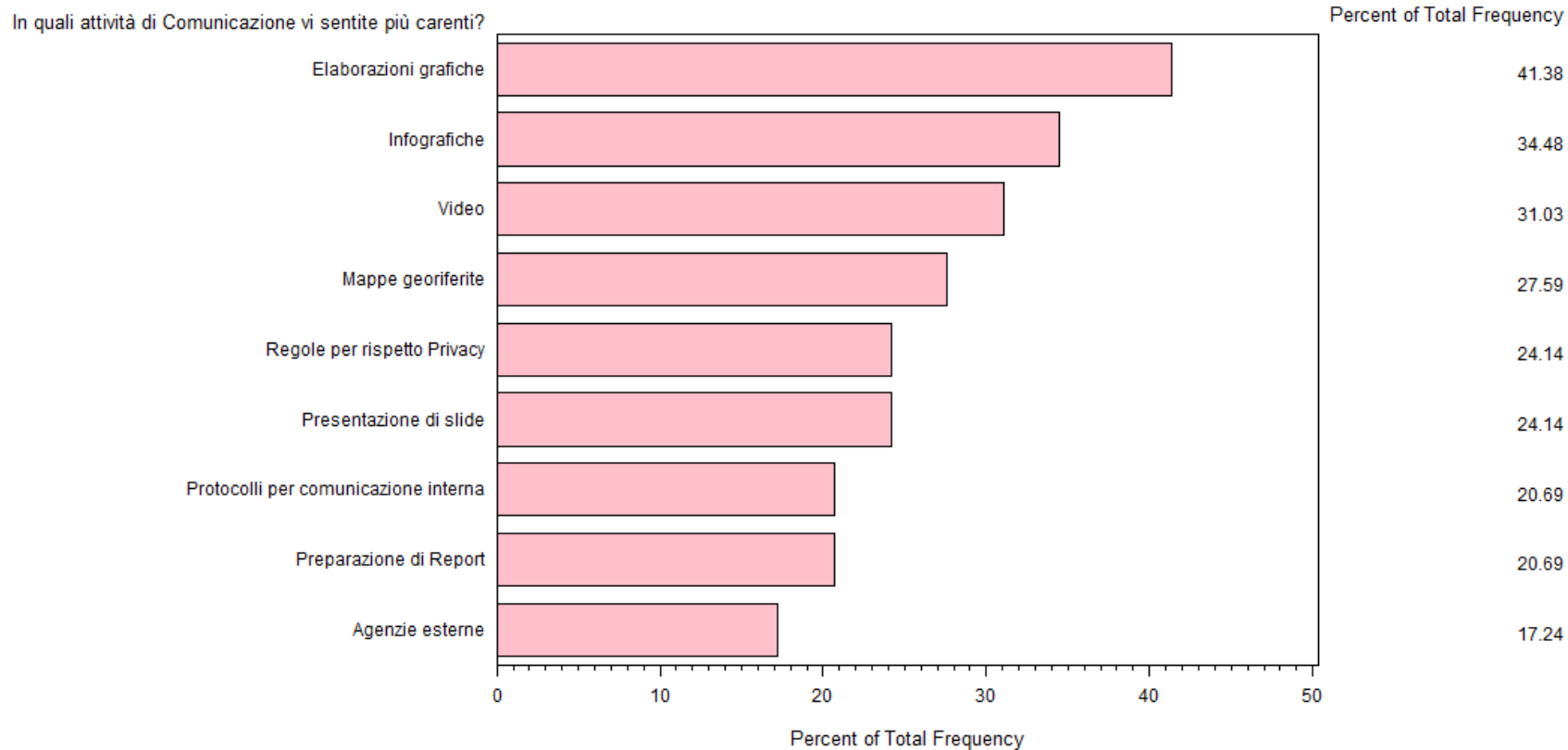
# ANALISI DEI DATI

Quali attività usate svolgere per le vostre procedure di Comunicazione?

Percent of Total Frequency

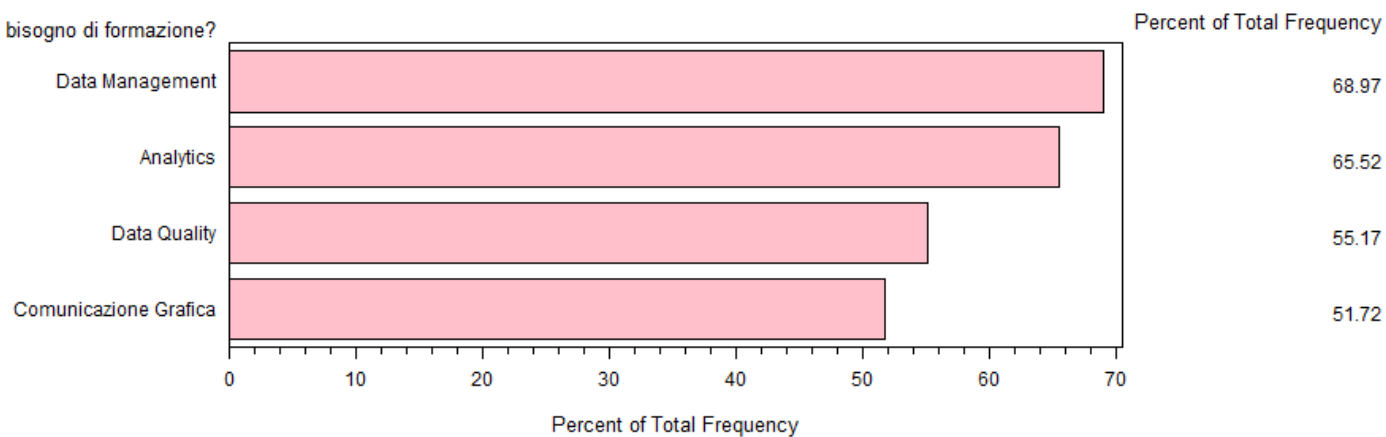


# ANALISI DEI DATI

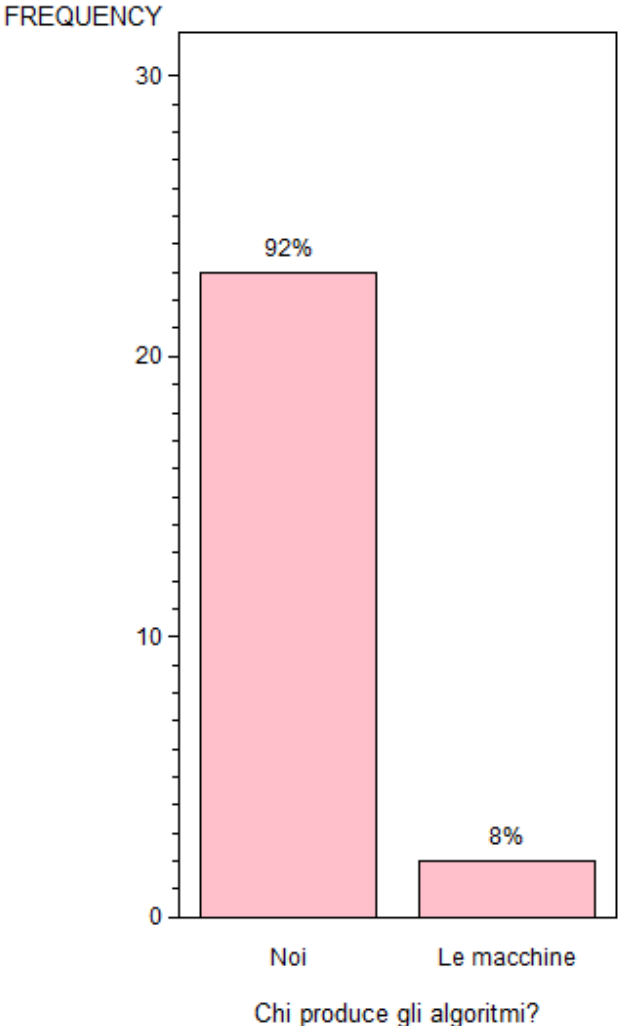


# ANALISI DEI DATI

In quale area ritenete di aver bisogno di formazione?



# ANALISI DEI DATI



# ANALISI DEI DATI

Algoritmo

L'operatività  
della  
macchina  
che funziona  
con  
algoritmi



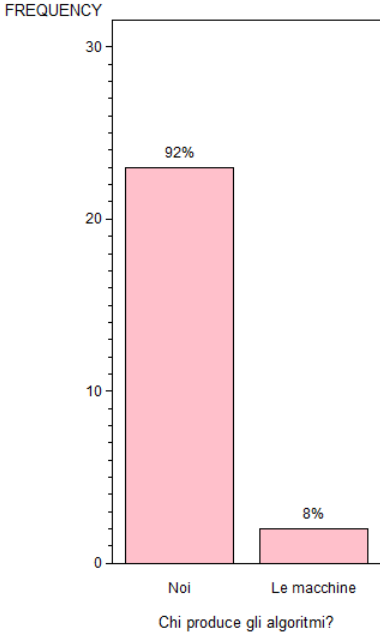
Un'espressio  
ne algebrica  
e/o logica  
resa  
operativa in  
un linguaggio  
di  
programmazi  
one

L'operatività

Chi produce gli algoritmi



# ANALISI DEI DATI



# ANALISI DEI DATI

Machine Learning

Le macchine imparano da sè



Noi impariamo attraverso le macchine

# ANALISI DEI DATI



## ANALISI DEI DATI

Se, come dicono i partecipanti all'indagine:

- L'algoritmo in buona sostanza è una sequenza di istruzioni che vengono passate alla macchina attraverso un programma
- Siamo noi a costruire gli algoritmi

Allora perché gli stessi partecipanti dicono che:

- Le macchine imparano da sé

?