

# HOMework BI 2018/2019

*Stefano Rola, Eugenio Andreose.*

# 1. Who is playing?

```
SELECT substr(DtAppello, length(DtAppello) - 1, 2) AS
```

```
AnnoAccademico
```

```
,CdS AS CorsoLaurea
```

```
,AD AS AttivitàDidattica
```

```
,count(*) AS NumeroStudenti
```

```
FROM bos_denormalizzato
```

```
GROUP BY CdSCod
```

```
,CdS
```

```
,AdCod
```

```
,AD
```

```
,AnnoAccademico
```

```
ORDER BY AnnoAccademico
```

```
,CdS;
```

Anno Accademico	Corso di Laurea	Attività Didattica	Numero Studenti
15	SCIENZE E TECNICHE PSICOLOGICHE	PSICOLOGIA DINAMICA	1
16	GIURISPRUDENZA	DIRITTO AMMINISTRATIVO	3
16	GIURISPRUDENZA	GIUSTIZIA AMMINISTRATIVA	5

# Who is playing?

- Formato data non agevole per le funzioni di SQLite;
- Dati riguardanti gli anni 2015 (un solo record), 2016 e 2017;
- Differenza tra DB denormalizzato e DB normalizzato nella trattazione del formato data.



## 2. I'll never graduate


```
CREATE VIEW top_dieci AS
```

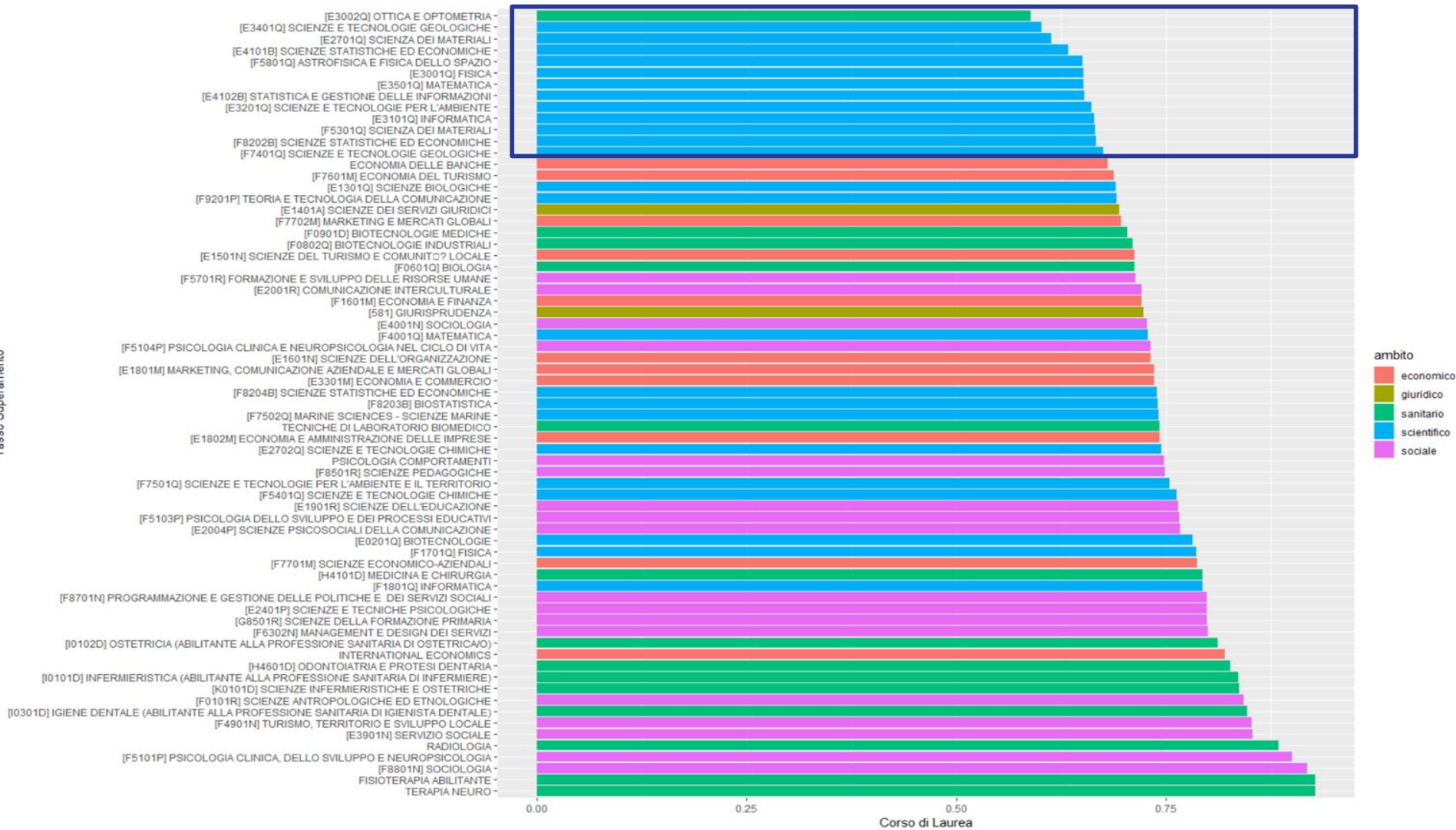
```
SELECT CdS,AD, round(sum(Superamento*1.00)/sum(Iscrizione*1.00),2) AS Tax_sup  
FROM bos_denormalizzato  
GROUP BY CdS,AD  
ORDER BY CdS, Tax_sup;
```

```
SELECT a.CdS, a.AD, a.Tax_sup  
FROM top_dieci AS a  
INNER JOIN top_dieci as b  
ON a.CdS = b.CdS AND (a.Tax_sup >= b.Tax_sup)  
GROUP BY a.CdS, a.AD,a.Tax_sup  
HAVING count(*) <= 10  
ORDER BY a.CdS, a.Tax_sup;
```

Corso di studio	Attività didattica	Tasso superamento
GIURISPRUDENZA	INFORMATICA GIURIDICA	0.42
GIURISPRUDENZA	DIRITTO COSTITUZIONALE EUROPEO ( I DIRITTI FONDAMENTALI) MOD. JEAN MONNET	0.46
GIURISPRUDENZA	[A5810018] PROCEDURA PENALE	0.47

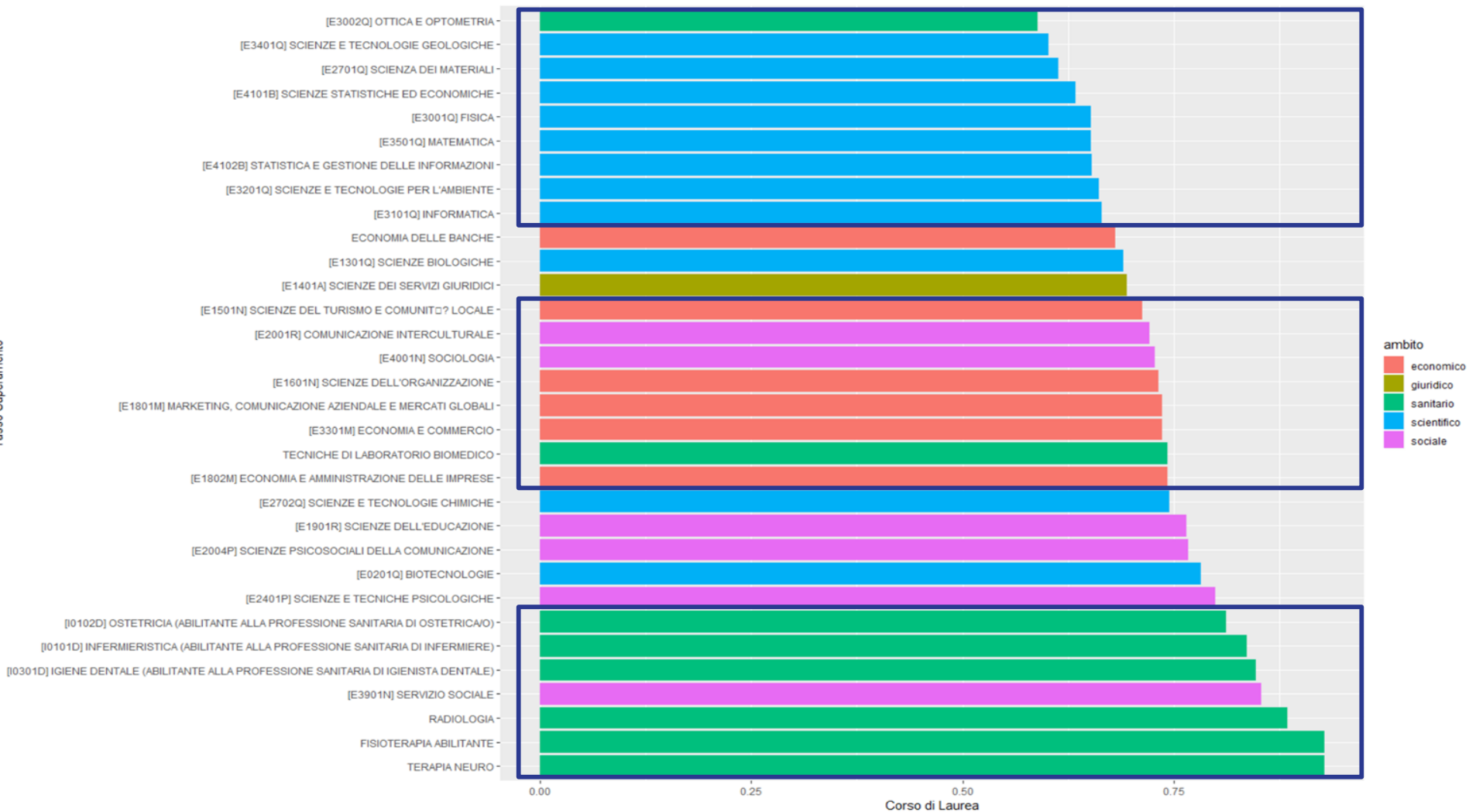
# I'll never graduate

- Per stilare la top-10 entro ogni corso di laurea abbiamo utilizzato una “*Self Join*”;
  - Non grosse differenze tra DB denormalizzato e DB normalizzato;
  - Il corso di studio “Scienze dell'Educazione” è codificato da due diversi identificatori.
- 



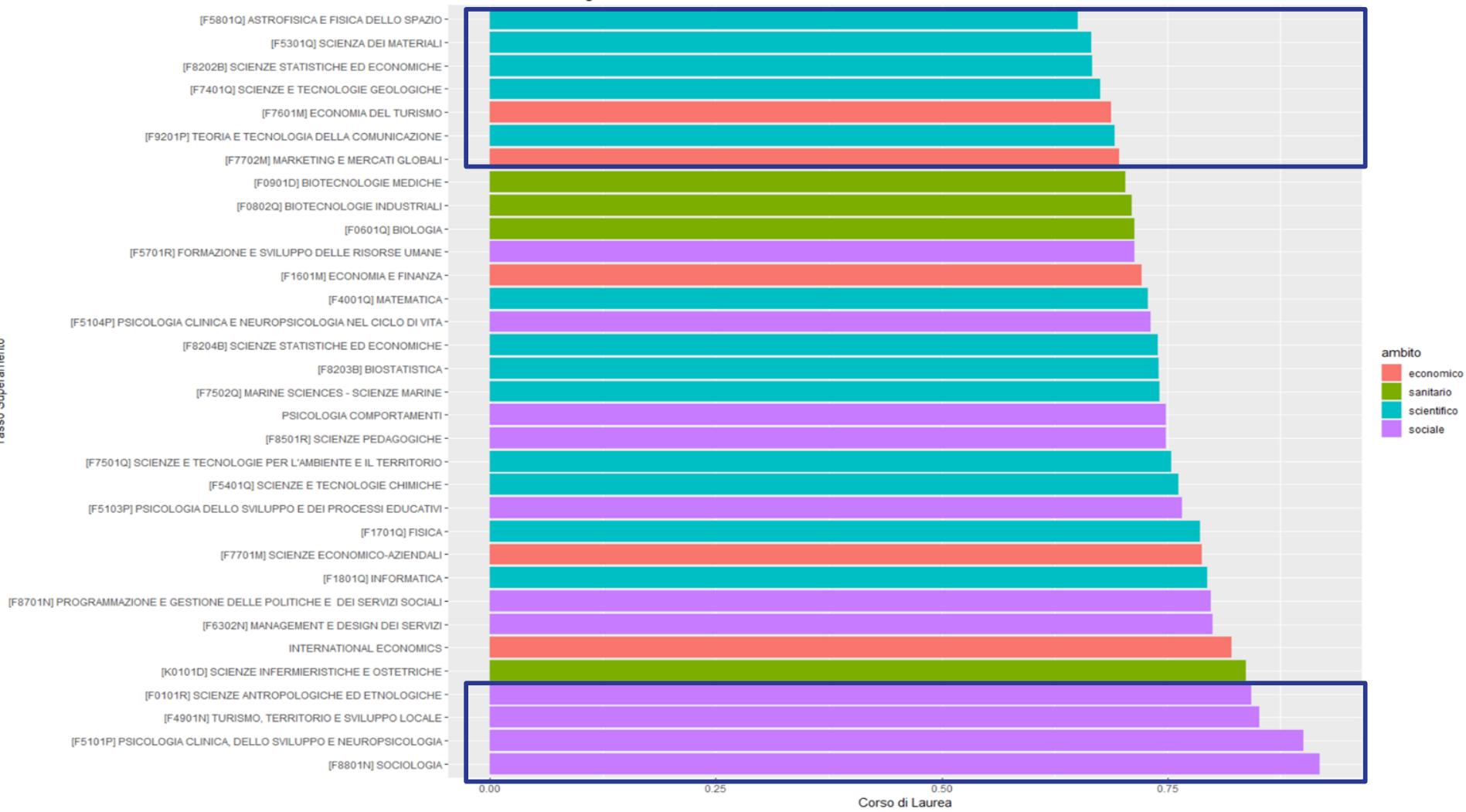
# Corsi di Laurea Triennale

Tasso Superamento



# Corsi di Laurea Magistrale

Tasso Superamento



ambito



Corso di Laurea



### 3. Commitment

```
CREATE VIEW com AS
```

```
SELECT DtAppello, CdS, AD  
FROM bos_denormalizzato  
GROUP BY DtAppello, CdS, AD  
ORDER BY DtAppello;
```

```
CREATE VIEW com1 AS
```

```
SELECT DtAppello, CdS, Count(*) AS num_app_cds_un_giorno  
FROM com  
GROUP BY DtAppello, CdS  
ORDER BY num_app_cds_un_giorno DESC;
```

```
SELECT CdS, round((num_app_cds_un_giorno > 1)*1.00/count(*)*1.00,2) AS Commitment_Tax  
FROM com1  
GROUP BY CdS  
ORDER BY Commitment_Tax DESC;
```

Corso di studi	Commitment Tax
GIURISPRUDENZA	0.82
SCIENZE DELL'EDUCAZIONE	0.75
ECONOMIA DELLE BANCHE, DELLE ASSICURAZIONI E DEGLI INTERMEDIARI FINANZIARI	0.74

# Commitment

- Tasso come numero di eventi sul totale della popolazione;
- Codice con più viste per migliorare l'interpretabilità;
- Sarebbe forse stato più informativo un'ulteriore suddivisione per anno di corso.




## 4. Hardest/Easiest exams

```
CREATE VIEW voto_medio_norm AS
SELECT cds.cds as CdS, ad.ad as Ad, ROUND(AVG(iscrizioni.voto),2) AS media_voto, COUNT(*)
FROM iscrizioni
      JOIN appelli ON appelli.appcod = iscrizioni.appcod
      JOIN ad ON ad.adcod = appelli.adcod
      JOIN cds ON cds.cdscod = appelli.cdscod
GROUP BY appelli.cdscod, appelli.adcod
HAVING media_voto NOT NULL AND count(*) > 5
ORDER BY appelli.cdscod, media_voto DESC;
```

```
CREATE VIEW top_3_norm AS
SELECT a.CdS, substr(a.Ad, -(length(a.Ad) - instr(a.Ad,']'))) as Ad, a.media_voto
FROM voto_medio_norm as a
      INNER JOIN voto_medio_norm as b on a.CdS=b.CdS AND a.media_voto<=b.media_voto
GROUP BY a.CdS, a.Ad, a.media_voto
HAVING count(*) <=3
ORDER BY a.CdS, a.media_voto DESC;
```

```
SELECT CdS, GROUP_CONCAT(Ad || " | " || media_voto) as top_3
FROM top_3_norm
GROUP BY CdS;
```



Corso di Studi	top_3	bottom_3
SCIENZE BIOLOGICHE	LABORATORIO DI CHIMICA   28.36, ANALISI DI FUNZIONI GENICHE   28.22, GENETICA   28.22	BIOCHIMICA SISTEMATICA UMANA   24.33, PATOLOGIA GENERALE   24.27, MATEMATICA E STATISTICA   23.93
SCIENZE STATISTICHE ED ECONOMICHE	DATA MINING   29.12, DATA SCIENCE E MODELLI STATISTICI PER IL TRATTAMENTO DEI DATI NON STRUTTURATI   28.5, ANALISI DEI DATI   28.4	ALGEBRA LINEARE   22.96, ECONOMIA AZIENDALE   22.5, ANALISI MATEMATICA I   22.12
ECONOMIA E COMMERCIO	BILANCIO   28.76, ECONOMIA DELLE AZIENDE PUBBLICHE   28.31, ECONOMIA DEL MERCATO MOBILIARE   26.87	MATEMATICA GENERALE I   23.74, STATISTICA PER L'AZIENDA   22.34, METODI MATEMATICI   21.74

Distorsione dei risultati causata da:

- Esami a scelta erogati da altri corsi di laurea (pochi iscritti, argomenti di studio inusuali)
- 498 esami superati con voto NULL

# 5. Fast & Furious

```
SELECT substr(CdS, -(length(CdS) - instr(CdS, ' '))) as cds, Studente,  
       primo_appello, ultimo_appello,  
       round(avg_voto,2) as avg_voto, diff,  
       num_appelli,  
       avg_voto/diff as ff_index  
FROM(  
  SELECT Studente, CdS, min(date.date) as primo_appello, max(date.date) as  
    ultimo_appello, julianday(max(date.date)) - julianday(min(date.date)) as diff,  
    avg(Voto) as avg_voto, count(*) as num_appelli  
  FROM bos_denormalizzato  
    JOIN date on date.date_denorm = bos_denormalizzato.DtAppello  
  GROUP BY Studente  
)  
WHERE ff_index not NULL AND num_appelli > 5 AND diff > 60  
ORDER BY ff_index DESC, cds;
```



Corso di Studi	Studente	primo_appello	ultimo_appello	avg_voto	diff	num_appelli	ff_index
SCIENZE DEL TURISMO E COMUNITA' LOCALE	945***B64	2017-05-10	2017-07-14	28.67	65	6	0.44
TURISMO, TERRITORIO E SVILUPPO LOCALE	9B2***F83	2017-07-11	2017-09-15	29.2	66	6	0.44
MEDICINA E CHIRURGIA	34C***9BB	2017-01-13	2017-03-23	29	69	6	0.42
SCIENZE DELLA FORMAZIONE PRIMARIA	3BA***E7F	2017-07-13	2017-09-23	30	72	6	0.42

- Problemi nella gestione della data (differenze normalizzato - denormalizzato);
- Chi ha dato meno esami è avvantaggiato: potrebbe essere utile inserire un peso per il numero di appelli superati;
- Analisi su studenti fast&furious potrebbe risultare più efficiente se si considerasse l'intera carriera accademica e non un solo anno.

## 6. Trial & Error

```
SELECT      Cds as Corso_di_Studi,
            Ad as Attività_didattica,
            AVG(te_index) as avg_te_index
FROM
    (SELECT      Cds,
                 Ad,
                 Studente,
                 SUM(Insufficienza+Ritiro) as te_index
    FROM bos_denormalizzato
    GROUP BY Cds, Ad, Studente
    )
GROUP BY Cds, Ad
ORDER BY avg_te_index DESC
```

Corso_di_Studio	Attività_didattica	avg_te_index
TECNICHE DI LABORATORIO BIOMEDICO (ABILITANTE ALLA PROFESSIONE SANITARIA DI TECNICO DI LABORATORIO)	ORGANI E FUNZIONI	2.07
COMUNICAZIONE INTERCULTURALE	STORIA MEDIOEVALE	2
SCIENZE DELL' ORGANIZZAZIONE	BILANCIO E ANALISI ECONOMICO FINANZIARIA	2

# Trial & Error

- Top 3 degli esami che richiedono in assoluto più tentativi prima del superamento;
- Numero di tentativi visto come somma di Insufficienze e Ritiri per ogni studente;
- Il massimo dei tentativi richiesti è di 2.07, alcuni esami ne richiedono 2 ma la maggior parte delle volte ne basta 1.



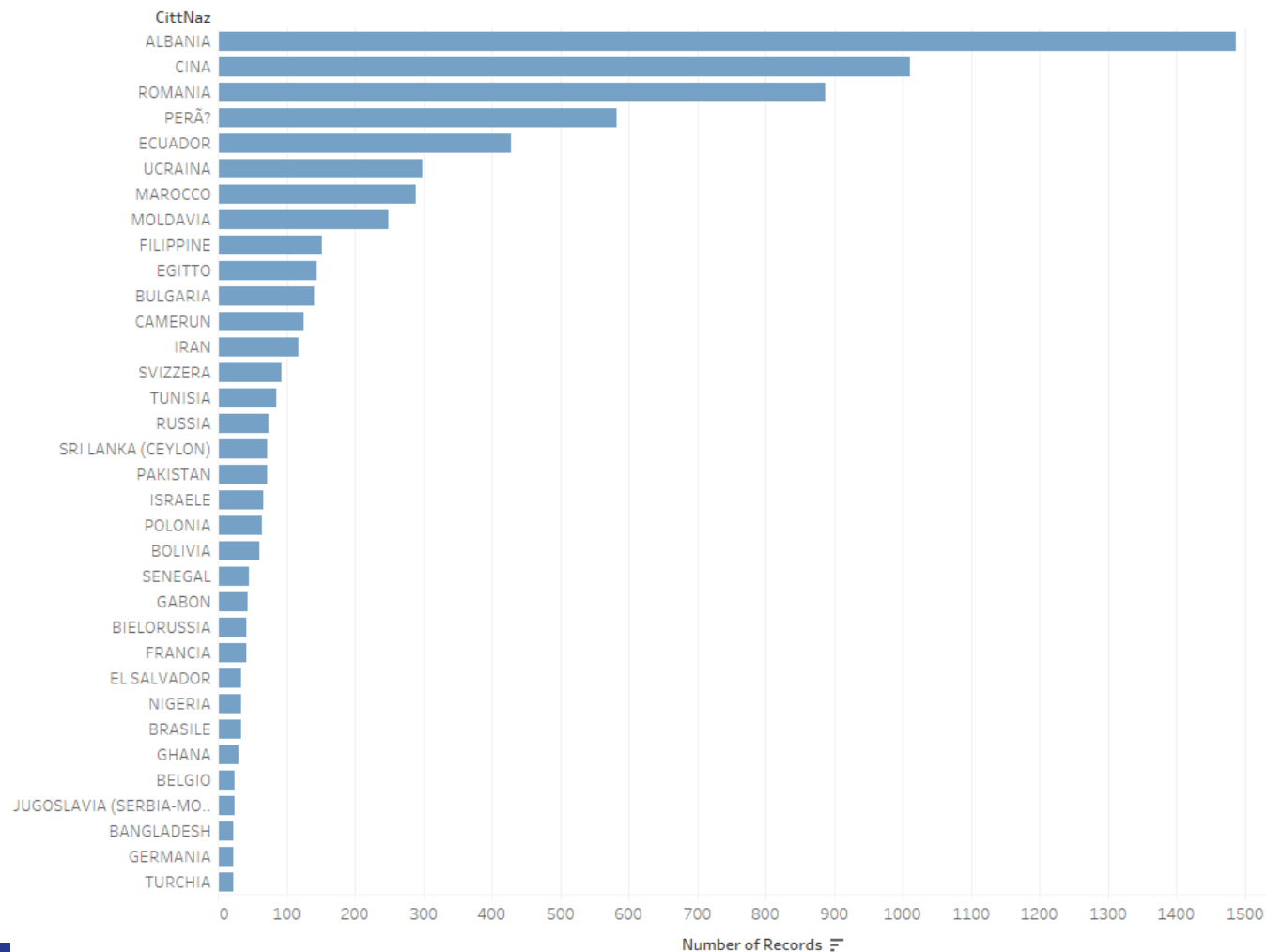


## 7. Foreigners students

```
SELECT bos_denormalizzato.CdS,  
       CittNaz,  
       COUNT(*) as numero_iscritti,  
       round( AVG(Voto) ,2) as media_voto,  
       round( (sum(Superamento)*1.00/sum(Iscrizione)*1.00) ,2) as tasso_superamento  
FROM bos_denormalizzato  
      JOIN date on bos_denormalizzato.DtAppello = date.date_denorm  
GROUP BY bos_denormalizzato.CdS, CittNaz  
HAVING COUNT(*) >= 5 AND tasso_superamento NOT NULL AND CittNaz != 'ITALIA' AND media_voto NOT NULL  
ORDER BY bos_denormalizzato.CdS, numero_iscritti desc
```

Corso di Studi	Cittadinanza	numero_iscritti	media_voto	tasso_superamento
GIURISPRUDENZA	ALBANIA	163	26.08	0.59
MARKETING, COMUNICAZIONE AZIENDALE E MERCATI GLOBALI	CINA	229	23.74	0.4
INFORMATICA	PERU'	54	24.34	0.59

## Distribuzione degli studenti per nazionalità



GRAZIE PER L'ATTENZIONE

