Tariq Baghrous
Tommaso Farneti

Business Intelligence Homework

Data Science - Year 2023/24

Preprocessing

In preparation for analysis, the date format in the database was proactively adjusted to align with the supported format for SQLite functions.

Relational model

1 - Added a new column of the 'datetime' type

```
ALTER TABLE appelli ADD date_appello DATETIME;
```

2 - Updated the new column with the correct dates

```
UPDATE appelli
   SET date_appello = CASE WHEN instr(appelli.dtappello, '/') = 0
   THEN NULL ELSE strftime('%Y-%m-%d', printf('%04d-%02d-%02d', substr(appelli.dtappello, -4),
   substr('0' || substr(appelli.dtappello, instr(appelli.dtappello, '/') + 1, 2), -2),
   substr('0' || substr(appelli.dtappello, 1, instr(appelli.dtappello, '/') - 1), -2) ) ) END;
```

bos_denormalizzato

1 - Changed the date format of the 'dtappello' column

```
UPDATE bos_denormalizzato SET DtAppello = '20' ||
    substr(DtAppello, 7, 2) || '-' ||
    substr(DtAppello, 4, 2) || '-' ||
    substr(DtAppello, 1, 2);
```

Query 1

Distribution of the number of students enrolled in various exam sessions, broken down by year and degree program.

Relational model

```
SELECT strftime('%Y', date_appello) as anno,
       cds,
       count(studente) AS [numero studenti iscritti]
  FROM iscrizioni,
       appelli,
       cds
 WHERE iscrizioni.appcod = appelli.appcod AND
       appelli.cdscod = cds.cdscod
 GROUP BY cds,
          strftime('%Y', date_appello)
 ORDER BY strftime('%Y', date_appello);
```

bos_denormalizzato

```
SELECT strftime('%Y', DtAppello) AS anno,
       CdS,
       count(Studente) as 'numero studenti iscritti'
  FROM bos_denormalizzato
  GROUP by CdS,
           anno
  ORDER by anno;
```

Query 1

Distribution of the number of students enrolled in various exam sessions, broken down by year and degree program.

Output

	anno	cds	numero studenti iscritti
58	2017	[E1601N] SCIENZE DELL'ORGANIZZAZIONE	3111
59	2017	[E1801M] MARKETING, COMUNICAZIONE AZIENDALE E MERCATI GLOBALI	6614
60	2017	[E1802M] ECONOMIA E AMMINISTRAZIONE DELLE IMPRESE	5525
61	2017	[E1803M] ECONOMIA DELLE BANCHE, DELLE ASSICURAZIONI E DEGLI INTERMEDIARI ...	5398
62	2017	[E1901R] SCIENZE DELL'EDUCAZIONE	11729
63	2017	[E2001R] COMUNICAZIONE INTERCULTURALE	3781
64	2017	[E2003P] COMUNICAZIONE E PSICOLOGIA	3
65	2017	[E2004P] SCIENZE PSICOSOCIALI DELLA COMUNICAZIONE	2523

•••

110	2017	[F8801N] SOCIOLOGIA	211
111	2017	[F9201P] TEORIA E TECNOLOGIA DELLA COMUNICAZIONE	779
112	2017	[G8501R] SCIENZE DELLA FORMAZIONE PRIMARIA	13415
113	2017	[H4101D] MEDICINA E CHIRURGIA	7073
114	2017	[H4601D] ODONTOIATRIA E PROTESI DENTARIA	949

- The total number of courses retrieved is 123
- The enrolment timeframe detected is 2016 - 2017, due to lack of data for 2015 and 2018
- Course of studies with the most number of students enrolled in various exam sessions are “Scienze della Formazione Primaria” followed by “Scienze dell’Educazione”

Query 2

Identification of the Top-10 most difficult exams, divided by field of study (the exam with the highest overall pass rate, considering all sessions of the Academic Year).

Relational model

```
SELECT cds, ad, Promossi, Iscritti, rapporto_promossi
FROM (
        SELECT cds.cds, ad.ad,
       sum(iscrizioni.Superamento) AS Promossi,
       count(iscrizioni.Iscrizione) AS Iscritti,
       round(CAST (sum(iscrizioni.Superamento) AS REAL) / count(iscrizioni.Iscrizione), 3) AS rapporto_promossi,
        row_number() OVER(PARTITION BY cds.cds ORDER BY CAST (sum(iscrizioni.Superamento) AS REAL) / count(iscrizioni.Iscrizione)) AS rank
  FROM ad JOIN appelli ON ad.adcod = appelli.adcod
       JOIN cds ON appelli.cdscod = cds.cdscod
       JOIN iscrizioni ON appelli.appcod = iscrizioni.appcod
 WHERE (strftime('%Y', date_appello) = '2016' AND
        strftime('%m', date_appello) >= '10' OR
        strftime('%Y', date_appello) = '2017' AND
        strftime('%m', date_appello) <= '10')
GROUP BY Cds, ad
HAVING count(ad.ad) >= 10
) ranked
WHERE rank <= 10
ORDER BY Cds, rapporto_promossi;
```

- We selected the academic year 2016/17 (from October 2016 to October 2017).
- We considered only those course of studies with at least 10 educational activities.
- rapporto_promossi is determined by the ratio of the number of students who passed the exam to the number of students who registered for the exam.

bos_denormalizzato

```
SELECT CdS, AD, Promossi, Iscritti, rapporto_promossi
FROM (
        SELECT CdS, AD,
       SUM(Superamento) AS Promossi,
       COUNT(Iscrizione) AS Iscritti,
       ROUND(CAST (SUM(Superamento) AS REAL) / COUNT(Iscrizione), 3) AS rapporto_promossi,
        ROW_NUMBER() OVER(PARTITION BY CdS ORDER BY CAST (SUM(Superamento) AS REAL) / COUNT(Iscrizione)) AS rank
  FROM bos_denormalizzato
 WHERE (strftime('%Y', DtAppello) = '2016' AND
        strftime('%m', DtAppello) >= '10' OR
        strftime('%Y', DtAppello) = '2017' AND
        strftime('%m', DtAppello) <= '10')
GROUP BY CdS, AD
HAVING COUNT(AD) >= 10
) ranked
WHERE rank <= 10
ORDER BY CdS, rapporto_promossi;
```

- CAST ensures that the division occurs between floating-point numbers.

Query 2

Identification of the Top-10 most difficult exams, divided by field of study (the exam with the highest overall pass rate, considering all sessions of the Academic Year).

Output

	CdS	AD	Promossi	Iscritti	rapporto_promossi
1	[581] GIURISPRUDENZA	[A5810031] INFORMATICA GIURIDICA	30	71	0.423
2	[581] GIURISPRUDENZA	[A5810057] DIRITTO INDUSTRIALE	11	22	0.5
3	[581] GIURISPRUDENZA	[A5810070] MEDICINA LEGALE	32	63	0.508
4	[581] GIURISPRUDENZA	[A5810017] DIRITTO PROCESSUALE...	178	343	0.519
5	[581] GIURISPRUDENZA	[A5810137] TUTELA ...	13	25	0.52
6	[581] GIURISPRUDENZA	[A5810018] PROCEDURA PENALE	119	226	0.527
7	[581] GIURISPRUDENZA	[A5810130] DIRITTO DELL'UNIONE ...	125	236	0.53
8	[581] GIURISPRUDENZA	[A5810135] STORIA DEL DIRITTO ...	32	60	0.533
9	[581] GIURISPRUDENZA	[A5810011] GIUSTIZIA ...	97	181	0.536
10	[581] GIURISPRUDENZA	[A5810019] ISTITUZIONI DI DIRITT...	258	466	0.554
11	[E0201Q] BIOTECNOLOGIE	[E0201Q001] MATEMATICA	141	268	0.526
12	[E0201Q] BIOTECNOLOGIE	[E0201Q059] BIOCHIMICA PER LE ...	73	136	0.537
13	[E0201Q] BIOTECNOLOGIE	[E0201Q050] METODOLOGIE ...	52	86	0.605
14	[E0201Q] BIOTECNOLOGIE	[E0201Q075] BIOTECNOLOGIE ...	26	42	0.619
15	[E0201Q] BIOTECNOLOGIE	[E0201Q055] COMPOSTI ORGANICI ...	24	38	0.632
16	[E0201Q] BIOTECNOLOGIE	[E0201Q053] ORGANI E FUNZIONI	83	128	0.648
17	[E0201Q] BIOTECNOLOGIE	[E0201Q065] BIOCHIMICA ...	84	126	0.667
18	[E0201Q] BIOTECNOLOGIE	[E0201Q073] GENETICA MOLECOLA...	25	37	0.676

- The period examinated is from October 2016 to October 2017.
- By isolating courses that feature a minimum of 10 didactic activities, we could reasonably compare diverse academic programs
- Considering the overall pass rate for the teaching activities of “Giurisprudenza” seems that the most difficult exam is “Informatica Giuridica”, followed by “Diritto Industriale” and “Medicina Legale”.

Query 3

Identification of degree programs with a high commitment rate, meaning different exam sessions of the same degree program that took place on the same day

Relational model

```
WITH AppelliDiversiStessaData AS (
  SELECT cdscod, COUNT(*) AS GiorniConPiuAppelli
  FROM (
    SELECT cdscod, date_appello
    FROM (
    SELECT DISTINCT date_appello, adcod, cdscod
    FROM appelli
)
    GROUP BY cdscod, date_appello
    HAVING COUNT(DISTINCT adcod) > 1
  )
    GROUP BY cdscod
)
, TotaleAppelli AS (
  SELECT cdscod, COUNT(DISTINCT date_appello) AS TotGiorni
  FROM (
    SELECT DISTINCT date_appello, adcod, cdscod
    FROM appelli
)
  GROUP BY cdscod
)
SELECT cds.cds, adsd.GiorniConPiuAppelli, t.TotGiorni, round((1.0 * adsd.GiorniConPiuAppelli) / t.TotGiorni, 3) AS TassoCommitment
FROM AppelliDiversiStessaData adsd, TotaleAppelli t, cds
WHERE adsd.cdscod = t.cdscod AND adsd.cdscod = cds.cdscod
ORDER by TassoCommitment DESC;
```

bos_denormalizzato

```
WITH AppelliDiversiStessaData AS (
  SELECT CdSCod, COUNT(*) AS GiorniConPiuAppelli
  FROM (
    SELECT CdsCod, DtAppello
    FROM (
    SELECT DISTINCT DtAppello, AdCod, CdsCod
    FROM bos_denormalizzato
)
    GROUP BY CdsCod, DtAppello
    HAVING COUNT(DISTINCT AdCod) > 1
  )
    GROUP BY CdsCod
)
, TotaleAppelli AS (
  SELECT CdsCod, COUNT(DISTINCT DtAppello) AS TotGiorni
  FROM (
    SELECT DISTINCT DtAppello, AdCod, CdsCod
    FROM bos_denormalizzato
)
  GROUP BY CdsCod
)
SELECT CdS, adsd.GiorniConPiuAppelli, t.TotGiorni, round((1.0 * adsd.GiorniConPiuAppelli) / t.TotGiorni, 3) AS TassoCommitment
FROM AppelliDiversiStessaData adsd, TotaleAppelli t, bos_denormalizzato
WHERE adsd.CdsCod = t.CdsCod AND adsd.CdsCod = bos_denormalizzato.CdsCod
GROUP by CdS
ORDER by TassoCommitment DESC;
```

- "AppelliDiversiStessaData" counts the number of days on which multiple distinct exams ('adcod') were held for the same degree program ('cdscod').
- "TotaleAppelli" calculates the total number of distinct days on which exams were conducted for each degree program.
- The final query merges this data to calculate the 'commitment rate'.

Query 3

Identification of degree programs with a high commitment rate, meaning different exam sessions of the same degree program that took place on the same day

Output

	CdS	GiorniConPiuAppelli	TotGiorni	TassoCommitment
1	[581] GIURISPRUDENZA	87	106	0.821
2	[E1901R] SCIENZE DELL'EDUCAZIONE	70	93	0.753
3	[E1803M] ECONOMIA DELLE BANCH...	62	84	0.738
4	[E1802M] ECONOMIA E ...	55	78	0.705
5	[E2001R] COMUNICAZIONE ...	64	91	0.703
6	[E3301M] ECONOMIA E COMMERCIO	56	80	0.7
7	[E1801M] MARKETING, ...	64	95	0.674
8	[F8204B] SCIENZE STATISTICHE ED ...	40	60	0.667
9	[E4101B] SCIENZE STATISTICHE ED...	50	75	0.667
10	[E1401A] SCIENZE DEI SERVIZI ...	54	83	0.651
11	[G8501R] SCIENZE DELLA ...	54	85	0.635
12	[F8501R] SCIENZE PEDAGOGICHE	43	70	0.614
13	[H4101D] MEDICINA E CHIRURGIA	69	117	0.59
14	[E1601N] SCIENZE ...	36	62	0.581
15	[E4102B] STATISTICA E GESTIONE ...	34	61	0.557
16	[F7601M] ECONOMIA DEL TURISMO	30	54	0.556
17	[F8203B] BIOSTATISTICA	26	49	0.531
18	[E2401P] SCIENZE E TECNICHE ...	70	134	0.522

- Considering all the course of studies, the ones which present highest commitment rates are “Giurisprudenza” with a value of 0.821 and “Scienze dell’Educazione” with 0.753;
- On the other hand, the courses with the lowest commitment rates are “Psicologia clinica e neuropsicologia nel ciclo di vita” with 0.167 followed by “Scienze infermieristriche e ostetriche” with 0.182

Query 4

Identification of the Top-3 exams with the highest and lowest average grades, respectively, calculated for each individual course of study

Relational model

```
----- Top 3 media voti maggiore
WITH Classificati AS (
    SELECT appelli.cdscod, appelli.adcod, AVG(iscrizioni.Voto) AS MediaVoti, ROW_NUMBER() OVER (PARTITION BY appelli.cdscod ORDER BY AVG(iscrizioni.Voto) DESC
        ) as Rango
    FROM iscrizioni
    JOIN appelli ON iscrizioni.appcod = appelli.appcod
    WHERE iscrizioni.Superamento = 1 AND iscrizioni.Voto is not null
    GROUP BY appelli.cdscod, appelli.adcod
), CdsAdCount AS (
    SELECT cdscod, COUNT(DISTINCT adcod) AS AdCount
    FROM appelli
    GROUP BY cdscod
    HAVING COUNT(DISTINCT adcod) >= 6
)
SELECT c.cdscod, c.adcod, c.MediaVoti
FROM Classificati c
JOIN CdsAdCount cac ON c.cdscod = cac.cdscod
WHERE c.Rango <= 3
ORDER BY c.cdscod, c.Rango;
```

```
----- Top 3 media voti minore
WITH Classificati AS (
    SELECT appelli.cdscod, appelli.adcod, AVG(iscrizioni.Voto) AS MediaVoti, ROW_NUMBER() OVER (PARTITION BY appelli.cdscod ORDER BY AVG(iscrizioni.Voto) ASC
        ) as Rango
    FROM iscrizioni JOIN appelli ON iscrizioni.appcod = appelli.appcod
    WHERE iscrizioni.Superamento = 1
    AND iscrizioni.Voto is not null
    GROUP BY appelli.cdscod, appelli.adcod
), CdsAdCount AS (
    SELECT cdscod, COUNT(DISTINCT adcod) AS AdCount
    FROM appelli
    GROUP BY cdscod
    HAVING COUNT(DISTINCT adcod) >= 6
)
SELECT c.cdscod, c.adcod, c.MediaVoti
FROM Classificati c JOIN CdsAdCount cac ON c.cdscod = cac.cdscod
WHERE c.Rango <= 3
ORDER BY c.cdscod, c.Rango;
```

bos_denormalizzato

```
-- Top 3 con media voti maggiore
WITH Classificati AS (
    SELECT CdSCod, AdCod, AVG(Voto) AS MediaVoti, ROW_NUMBER() OVER (PARTITION BY CdSCod ORDER BY AVG(Voto) DESC
        ) AS Rango
    FROM bos_denormalizzato
    WHERE Superamento = 1 AND Voto IS NOT NULL
    GROUP BY CdSCod, AdCod
), CdsAdCount AS (
    SELECT CdSCod, COUNT(DISTINCT AdCod) AS AdCount
    FROM bos_denormalizzato
    GROUP BY CdSCod
    HAVING COUNT(DISTINCT AdCod) >= 6
)
SELECT c.CdSCod, c.AdCod, c.MediaVoti
FROM Classificati c
JOIN CdsAdCount cac ON c.CdSCod = cac.CdSCod
WHERE c.Rango <= 3
ORDER BY c.CdSCod, c.Rango;
```

```
-- Top 3 con media voti minore
WITH Classificati AS (
    SELECT CdSCod, AdCod, AVG(Voto) AS MediaVoti, ROW_NUMBER() OVER (PARTITION BY CdSCod ORDER BY AVG(Voto) ASC
        ) AS Rango
    FROM bos_denormalizzato
    WHERE Superamento = 1 AND Voto IS NOT NULL
    GROUP BY CdSCod, AdCod
), CdsAdCount AS (
    SELECT CdSCod, COUNT(DISTINCT AdCod) AS AdCount
    FROM bos_denormalizzato
    GROUP BY CdSCod
    HAVING COUNT(DISTINCT AdCod) >= 6
)
SELECT c.CdSCod, c.AdCod, c.MediaVoti
FROM Classificati c
JOIN CdsAdCount cac ON c.CdSCod = cac.CdSCod
WHERE c.Rango <= 3
ORDER BY c.CdSCod, c.Rango;
```

Query 4

Identification of the Top-3 exams with the highest and lowest average grades, respectively, calculated for each individual course of study

Output

	cds	ad	MediaVoti
1	[581] GIURISPRUDENZA	[A5810234] LA CORTE DI GIUSTIZIA: TECNICHE E STRUMENTI	30.0
2	[581] GIURISPRUDENZA	[A5810172] INTERNATIONAL CRIMINAL JUSTICE	30.0
3	[581] GIURISPRUDENZA	[A5810171] LAW OF BRICS COUNTRIES AND EU	30.0
4	[E0201Q] BIOTECNOLOGIE	[E0201Q074] BIOFISICA	30.0
5	[E0201Q] BIOTECNOLOGIE	[E0201Q058] CHIMICA FISICA	29.0
6	[E0201Q] BIOTECNOLOGIE	[E0201Q063] BIOCHIMICA CELLULARE	28.364
7	[E1301Q] SCIENZE BIOLOGICHE	[E3201Q084] GESTIONE DI BASE DATI	30.0
8	[E1301Q] SCIENZE BIOLOGICHE	[E0201Q063] BIOCHIMICA CELLULARE	30.0
9	[E1301Q] SCIENZE BIOLOGICHE	[E0201Q046] INFORMATICA	30.0

	cds	ad	MediaVoti
1	[581] GIURISPRUDENZA	[A5810081] ECONOMIA AZIENDALE	21.0
2	[581] GIURISPRUDENZA	[A5810001] DIRITTO DEI CONSUMATORI	23.667
3	[581] GIURISPRUDENZA	[A5810130] DIRITTO DELL'UNIONE EUROPEA	25.104
4	[E0201Q] BIOTECNOLOGIE	[E0201Q050] METODOLOGIE BIOCHIMICHE E TECNOLOGIE ...	22.635
5	[E0201Q] BIOTECNOLOGIE	[E1301Q059] ECOLOGIA APPLICATA	23.0
6	[E0201Q] BIOTECNOLOGIE	[E0201Q059] BIOCHIMICA PER LE BIOTECNOLOGIE	23.411
7	[E1301Q] SCIENZE BIOLOGICHE	[E1301Q071] MATEMATICA E STATISTICA	23.926
8	[E1301Q] SCIENZE BIOLOGICHE	[E0201Q067] PATOLOGIA GENERALE	24.273
9	[E1301Q] SCIENZE BIOLOGICHE	[E0201Q065] BIOCHIMICA SISTEMATICA UMANA	24.333

- We first calculated the average grade for each teaching activity within each degree program. The average is computed only for passed exams.
- Then we counted the number of distinct teaching activities for each degree program and includes only those with at least 6 different activities.
- The results are ordered by degree program and rank.
- For example the three exams of Scienze Biologiche with the highest grades on average are “Gestione di Base Dati”, “Biochimica Cellulare” and “Informatica”.

Query 5

Calculate the distribution of '*fast&furious*' students by field of study, defined as students with the highest ratio of 'average grade achieved in passed exams' to 'period of academic activity'

Relational model

```
WITH Media as (
    SELECT cdscod, studente, avg(voto) as MediaEsami
    FROM iscrizioni, appelli
    WHERE iscrizioni.appcod = appelli.appcod AND Superamento == '1' AND voto is NOT NULL
    group by studente
    HAVING count(studente) > 3
    )
, Periodo as (
    SELECT cdscod, studente, (JULIANDAY(MAX(date_appello)) - JULIANDAY(MIN(date_appello))) AS PeriodoAttività
    FROM iscrizioni, appelli
    WHERE iscrizioni.appcod = appelli.appcod AND (Superamento == '1' or Insufficienza == '1')
    GROUP BY studente
)
SELECT cds.cds, m.studente, MediaEsami, PeriodoAttività, round((1.0 * m.MediaEsami) / p.PeriodoAttività, 3) AS TassoFastAndFurious
FROM Media m, Periodo p, cds
WHERE m.studente = p.studente AND cds.cdscod = p.cdscod AND TassoFastAndFurious NOT NULL
ORDER by cds.cds ASC, TassoFastAndFurious DESC;
```

bos_denormalizzato

```
WITH Media as (
    SELECT CdSCod, Studente, avg(Voto) as MediaEsami
    FROM bos_denormalizzato
    WHERE Superamento == '1' AND Voto is NOT NULL
    group by Studente
    HAVING count(Studente) > 3
    )
, Periodo as (
    SELECT CdSCod, Studente, (JULIANDAY(MAX(DtAppello)) - JULIANDAY(MIN(DtAppello))) AS PeriodoAttività
    FROM bos_denormalizzato
    WHERE Superamento == '1' or Insufficienza == '1'
    GROUP BY Studente
)
SELECT CdS, m.Studente, MediaEsami, PeriodoAttività, (1.0 * m.MediaEsami) / p.PeriodoAttività AS TassoFastAndFurious
FROM Media m, Periodo p, bos_denormalizzato
WHERE m.Studente = p.Studente AND bos_denormalizzato.CdSCod = p.CdSCod AND TassoFastAndFurious NOT NULL
ORDER by CdS ASC, TassoFastAndFurious DESC;
```

- "Media" calculates the average exam grades of each student who has completed at least 4 exams, divided by course of study.
- "Periodo" calculates the period of activity for each student., considering passed and failed exams
- The final query merges this data to identify students with the highest Fast&Furious rate.

Query 5

Calculate the distribution of '*fast&furious*' students by field of study, defined as students with the highest ratio of 'average grade achieved in passed exams' to 'period of academic activity'

Output

	cds	studente	MediaEsami	PeriodoAttività	TassoFastAndFurious
1	[581] GIURISPRUDENZA	CE8EF622396E7BE8CF2A124F38D13...	30.0	52.0	0.577
2	[581] GIURISPRUDENZA	6E532A675C8EDF4F697013F1E79FF...	25.0	44.0	0.568
3	[581] GIURISPRUDENZA	7DCBCCCBFF28C468569DDFEFFC8A...	24.25	47.0	0.516
4	[581] GIURISPRUDENZA	83B0085457A81B5AA49037DCFA7B3...	26.0	51.0	0.51
5	[581] GIURISPRUDENZA	A3281DF2B1E241F9228669D62F02C...	27.75	65.0	0.427
6	[581] GIURISPRUDENZA	643317FD9540A468CFD978BF09A9B...	27.2	66.0	0.412
7	[581] GIURISPRUDENZA	C8F891CEB25055123A46DD432F50C...	29.0	75.0	0.387
8	[581] GIURISPRUDENZA	6DED9BE5F18220B9FF5E2016BD338...	29.25	77.0	0.38
9	[581] GIURISPRUDENZA	E4628C4A835E26C3BCE35A686B4B7...	29.25	78.0	0.375
10	[581] GIURISPRUDENZA	C3B04BD6EBA25F5AFB8C5540FE70A...	23.75	66.0	0.36
11	[581] GIURISPRUDENZA	863191A71288E11F18037076D654E0...	27.0	77.0	0.351
12	[581] GIURISPRUDENZA	7512838E4A5044E03DF0153941DB8...	20.5	62.0	0.331
13	[581] GIURISPRUDENZA	C471A3EBB39C8D0A47F805AE217A...	27.25	83.0	0.328
14	[581] GIURISPRUDENZA	5B97FDBC6C89B7D1D304188BBD9D...	28.75	100.0	0.288
15	[581] GIURISPRUDENZA	1141E9E64A01B080D5C811ABD2658...	28.0	98.0	0.286
16	[581] GIURISPRUDENZA	4C310EBDE21AEDF6EAAEB036781D...	27.5	98.0	0.281
17	[581] GIURISPRUDENZA	41A58F787C07200A789CDCBCFBE7C...	20.75	75.0	0.277
18	[581] GIURISPRUDENZA	F4E95005F5F2E1C005EAA7B013EDE...	24.75	93.0	0.266

- Filtering out students with less than 4 passed exams contributed to a more meaningful evaluation of the *Periodo di Attività*
- The best students according to this ratio were from “International economics”, “Economia delle banche, delle assicurazioni e degli intermedi finanziari” and “Marketing, comunicazione aziendale e mercati globali” with values respectively of 7.0, 5.875 and 3.531

Query 6

Identification of the Top-3 '*trial&error*' exams, meaning exams that require the highest number of attempts before passing (average number of attempts (failures) by each student for each session of the course)

Relational model

```sql
WITH superamenti AS (
    SELECT cds.cds, studente, ad.ad
    FROM iscrizioni
    JOIN appelli ON iscrizioni.appcod = appelli.appcod
    JOIN cds ON appelli.cdscod = cds.cdscod
    JOIN ad ON appelli.adcod = ad.adcod
    WHERE Superamento = '1' AND Voto IS NOT NULL
)
, tentativi AS (
    SELECT cds.cds, iscrizioni.studente, ad.ad, COUNT(CASE WHEN Insufficienza = '1' THEN 1 END) AS bocciature
    FROM iscrizioni
    JOIN appelli ON iscrizioni.appcod = appelli.appcod
    JOIN cds ON appelli.cdscod = cds.cdscod
    JOIN ad ON appelli.adcod = ad.adcod
    WHERE Insufficienza = '1' AND
        studente IN (SELECT studente FROM superamenti)
    GROUP BY cds.cds, ad.ad, studente
)
SELECT cds, ad, TrialAndError
FROM (
    SELECT cds, ad, ROUND(AVG(bocciature), 3) AS TrialAndError, ROW_NUMBER() OVER (PARTITION BY cds ORDER BY AVG(bocciature) DESC) AS rank
    FROM tentativi
    GROUP BY cds, ad
) ranked
WHERE rank <= 3
ORDER BY cds, TrialAndError DESC;
```

bos_denormalizzato

```sql
WITH superamenti AS (
    SELECT CdS, Studente, AD
    FROM bos_denormalizzato
    WHERE Superamento = '1' AND Voto IS NOT NULL
)
, tentativi AS (
    SELECT CdS, Studente, AD, COUNT(CASE WHEN Insufficienza = '1' THEN 1 END) AS bocciature
    FROM bos_denormalizzato
    WHERE Insufficienza = '1' AND
        Studente IN (SELECT Studente FROM superamenti)
    GROUP BY CdS, AD, Studente
)
SELECT CdS, AD, TrialAndError
FROM (
    SELECT CdS, AD, ROUND(AVG(bocciature), 3) AS TrialAndError, ROW_NUMBER() OVER (PARTITION BY CdS ORDER BY AVG(bocciature) DESC) AS rank
    FROM tentativi
    GROUP BY CdS, AD
) ranked
WHERE rank <= 3
ORDER BY CdS, TrialAndError DESC;
```

- "*Superamenti*" identifies students who have passed an exam for a specific educational activity in a course of study.
- "*Tentativi*" counts the number of failures in each exam across various educational activities, only for the students in “Superamenti”.
- The final query combines this data to identify exams with the highest Trial&Error rate.

Query 6

Identification of the Top-3 '*trial&error*' exams, meaning exams that require the highest number of attempts before passing (average number of attempts (failures) by each student for each session of the course)

Output

	cds	ad	TrialAndError
1	[581] GIURISPRUDENZA	[A5810003] SCIENZA DELLE FINANZE	1.313
2	[581] GIURISPRUDENZA	[A5810070] MEDICINA LEGALE	1.25
3	[581] GIURISPRUDENZA	[A5810228] STORIA DEL DIRITTO ...	1.212
4	[E0201Q] BIOTECNOLOGIE	[E0201Q073] GENETICA MOLECOLA...	1.5
5	[E0201Q] BIOTECNOLOGIE	[E0201Q001] MATEMATICA	1.434
6	[E0201Q] BIOTECNOLOGIE	[E0201Q055] COMPOSTI ORGANICI ...	1.333
7	[E1301Q] SCIENZE BIOLOGICHE	[E2401P019] PSICOBIOLOGIA DEI ...	2.0
8	[E1301Q] SCIENZE BIOLOGICHE	[E1301Q009] CHIMICA GENERALE	1.629
9	[E1301Q] SCIENZE BIOLOGICHE	[E0201Q065] BIOCHIMICA ...	1.529
10	[E1401A] SCIENZE DEI SERVIZI ...	[E1401A080] DIRITTO DELL'UNIONE...	1.667
11	[E1401A] SCIENZE DEI SERVIZI ...	[E1401A014] STATISTICA ...	1.429
12	[E1401A] SCIENZE DEI SERVIZI ...	[E1401A076] ELEMENTI DI STORIA ...	1.333
13	[E1501N] SCIENZE DEL TURISMO E ...	[E1501N005] SOCIOLOGIA GENERALE	1.444
14	[E1501N] SCIENZE DEL TURISMO E ...	[E1501N071] SOCIOLOGIA DEL ...	1.351
15	[E1501N] SCIENZE DEL TURISMO E ...	[E1501N007] LINGUA E TRADUZION...	1.207
16	[E1601N] SCIENZE ...	[E1801M050] LINGUA - FRANCESE	3.0
17	[E1601N] SCIENZE ...	[E1601N006] INGLESE	1.525
18	[E1601N] SCIENZE ...	[E1601N043] RETI DI IMPRESE E ...	1.5

- The attemps considered only refers to “Insufficienze”
- Across all the courses, the highest “trial&error” rate was found for “Lingua inglese” exam for “Tecniche di laboratorio biomedico” course with an average attempt of 3.2 before passing

Query 7

Academic performance: geographical variation and gender disparities across study levels

- Identification of the average grades based on the students' geographical origin

Relational model

```
SELECT s.cittnaz, COUNT(DISTINCT s.studente) AS NumeroStudenti, round(AVG(i.Voto), 3) AS MediaVoti
FROM studenti s JOIN iscrizioni i ON s.studente = i.studente
JOIN appelli a ON i.appcod = a.appcod
WHERE i.Voto IS NOT NULL
GROUP BY s.cittnaz;
```

bos_denormalizzato

```
SELECT cittnaz, count(distinct(studente)) as NumeroStudenti, round(avg(Voto), 3) as MediaVoti
FROM bos_denormalizzato
WHERE Voto is not null
GROUP BY cittnaz;
```

- Comparison of average grades by gender across different levels of study programs, including 'Bachelor's', 'Master's', and 'Single-cycle Master's' degrees

Relational model

```
SELECT strftime('%m/%Y', a.date_appello) AS data, s.genere, c.tipocorso, ROUND(AVG(i.Voto), 3) AS media_voti
FROM iscrizioni i JOIN appelli a ON i.appcod = a.appcod
JOIN cds c ON a.cdscod = c.cdscod
JOIN studenti s ON i.studente = s.studente
WHERE i.Voto IS NOT NULL
AND i.Superamento = 1
GROUP BY strftime('%m/%Y', a.date_appello), s.genere, c.tipocorso
ORDER BY strftime('%Y', a.date_appello), c.tipocorso;
```

bos_denormalizzato

```
SELECT strftime('%m/%Y', DtAppello) AS data, StuGen AS genere, TipoCorso, ROUND(AVG(Voto), 2) AS media_voti
FROM bos_denormalizzato
WHERE Voto IS NOT NULL
AND Superamento = 1
GROUP BY strftime('%m/%Y', DtAppello), StuGen, TipoCorso
ORDER BY  strftime('%Y', DtAppello), TipoCorso;
```

Query 7

Academic performance: geographical variation and gender disparities across study levels

Output

- Identification of the average grades based on the students' geographical origin

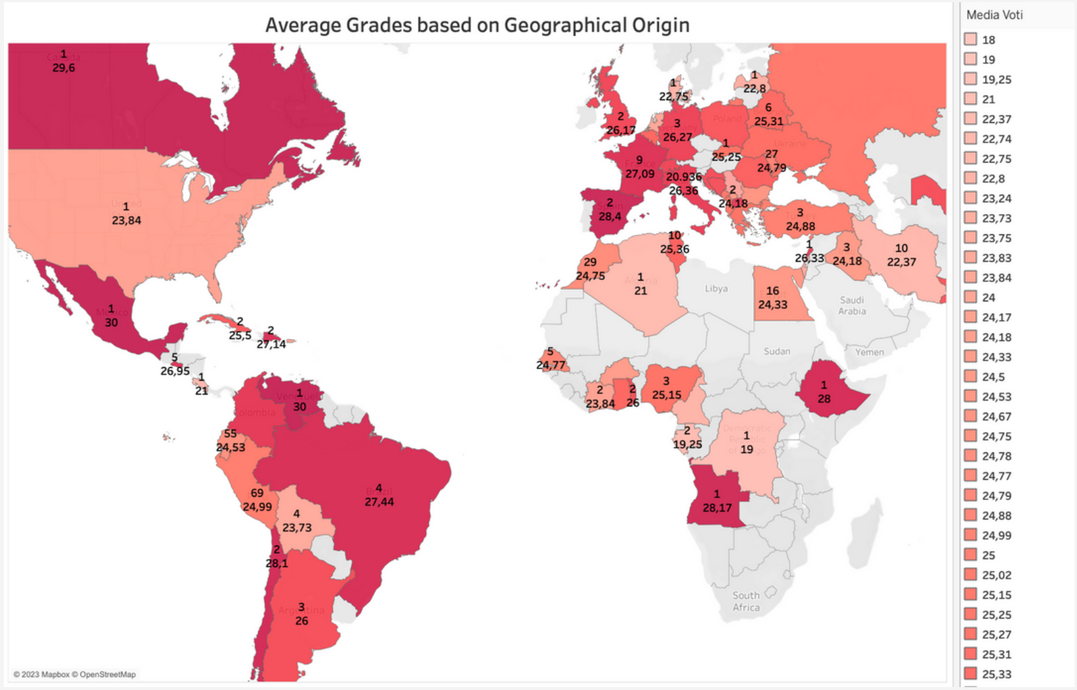
	cittnaz	NumeroStudenti	MediaVoti
1	ALBANIA	155	25.025
2	ALGERIA	1	21.0
3	ANGOLA	1	28.167
4	ARGENTINA	3	26.0
5	BANGLADESH	3	26.143
6	BELGIO	3	25.333
7	BIELORUSSIA	6	25.308
8	BOLIVIA	4	23.733
9	BOSNIA-ERZEGOVINA	3	25.875
10	BRASILE	4	27.438
11	BULGARIA	14	24.5
12	BURKINA FASO (ALTO VOLTA)	1	24.167
13	CAMERUN	9	22.737
14	CANADA	1	29.6
15	CILE	2	28.1
16	CINA	114	23.831
17	COLOMBIA	2	26.714
18	CONGO, REP.DEM. (ZAIRE)	1	19.0

- Comparison of average grades by gender across different levels of study programs, including 'Bachelor's', 'Master's', and 'Single-cycle Master's' degrees

	data	genere	TipoCorso	media_voti
1	09/2015	F	[L2] Corso di Laurea - L2	24.0
2	04/2016	M	[L2] Corso di Laurea - L2	25.0
3	05/2016	M	[L2] Corso di Laurea - L2	28.0
4	07/2016	F	[L2] Corso di Laurea - L2	24.0
5	09/2016	F	[L2] Corso di Laurea - L2	27.4
6	09/2016	M	[L2] Corso di Laurea - L2	27.5
7	11/2016	F	[L2] Corso di Laurea - L2	25.56
8	11/2016	M	[L2] Corso di Laurea - L2	25.07
9	12/2016	F	[L2] Corso di Laurea - L2	26.12
10	12/2016	M	[L2] Corso di Laurea - L2	25.43
11	07/2016	F	[LM5] Laurea Magistrale Ciclo Unico ...	26.0
12	09/2016	F	[LM5] Laurea Magistrale Ciclo Unico ...	29.16
13	09/2016	M	[LM5] Laurea Magistrale Ciclo Unico ...	29.25
14	11/2016	F	[LM5] Laurea Magistrale Ciclo Unico ...	28.12
15	11/2016	M	[LM5] Laurea Magistrale Ciclo Unico ...	27.98
16	12/2016	F	[LM5] Laurea Magistrale Ciclo Unico ...	28.69
17	12/2016	M	[LM5] Laurea Magistrale Ciclo Unico ...	28.49
18	12/2016	F	[LM6] Laurea Magistrale Ciclo Unico ...	28.6

Query 7

- Identification of the average grades based on the students' geographical origin



- Comparison of average grades by gender across different levels of study programs, including 'Bachelor's', 'Master's', and 'Single-cycle Master's' degrees

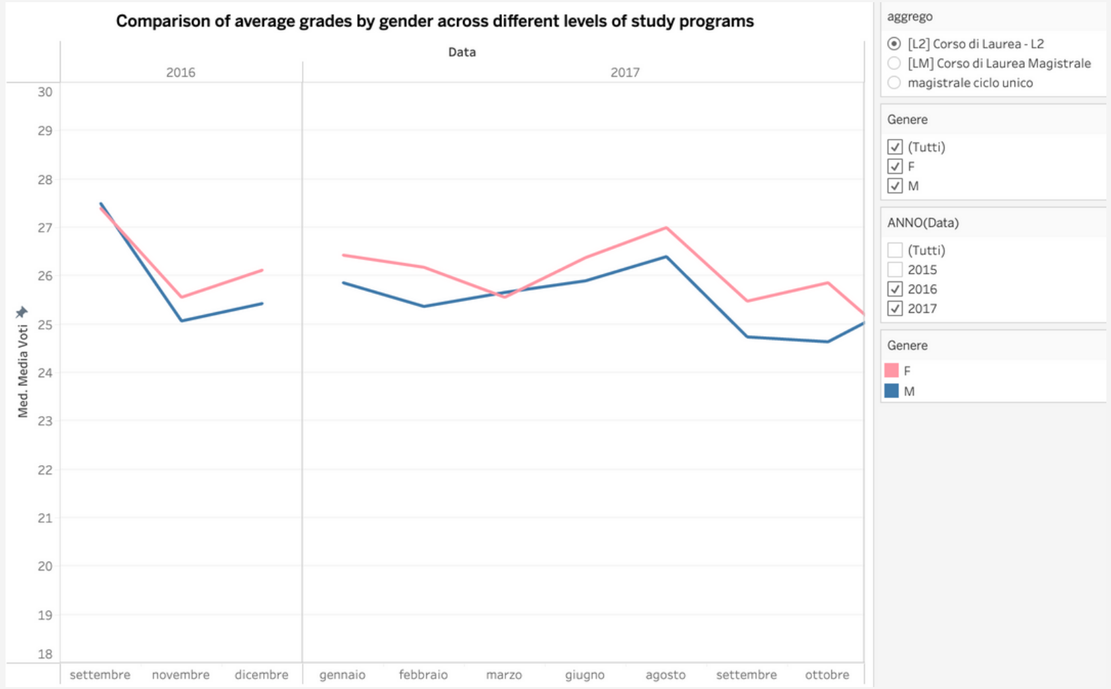


Tableau: Query 7 Dashboard

Conclusion

- Both queries (*bos_denormalizzato* and relational model) yield identical results, emphasizing the consistency of the database.
- The *bos_denormalizzato* queries, though simpler to write, exhibits a relatively longer execution time.
- The database contains an anomaly where sometimes a passed exam is associated with a *null* value under the “voto” attribute.