

# Progetto SQL di Valentina Menon

**Gender and Diplomatic  
Representation**



Il set di dati GenDip di Birgitta Niklasson e Ann Towns mappa la misura in cui gli Stati nominano uomini e donne per diversi tipi di incarichi di ambasciatori bilaterali. I dati coprono oltre 200 Paesi e 10 anni specifici tra il 1968 e il 2021.

Link alla cartella Drive contenente lo script SQL del progetto:  
<https://drive.google.com/drive/folders/1U8pR7rekigSVH2055DMzFwb-4woDyP?usp=sharing>

Nella prima parte del progetto ho analizzato i dati a livello globale, mentre nella seconda parte ho concentrato l'analisi sui dati relativi all'Italia.



OBIETTIVO: analizzare la disparità di genere nel corso del tempo

QUERY: raggruppamento per anno e per genere e ordinare i dati per anno in ordine crescente

```
⊖ SELECT year, gender, COUNT(*)  
FROM sqlgendip  
GROUP BY year, gender  
ORDER BY year ASC
```

	ABC year ▾	ABC gender ▾	123 COUNT(*) ▾
1	1968	0	3.136
2	1968	99	459
3	1968	1	29
4	1978	0	4.593
5	1978	99	997
6	1978	1	81
7	1988	0	5.762
8	1988	99	394
9	1988	1	162
10	1998	0	6.576
11	1998	99	679
12	1998	1	413
13	2003	99	516
14	2003	0	6.908
15	2003	1	674
16	2008	0	7.232
17	2008	1	964
18	2008	99	325
19	2013	0	6.885
20	2013	99	934
21	2013	1	1.184
22	2014	0	12.224
23	2014	1	2.340
24	2014	99	690
25	2019	0	11.480
26	2019	1	2.992
27	2019	99	10
28	2021	0	11.983
29	2021	1	3.482
30	2021	99	405



anno	donne	gen. mancante	uomini	Grand Total	% donne su totale	% gen. mancante su totale	% uomini su totale
1968	29	459	3,136	3,624	0.80%	12.67%	86.53%
1978	81	997	4,593	5,671	1.43%	17.58%	80.99%
1988	162	394	5,762	6,318	2.56%	6.24%	91.20%
1998	413	679	6,576	7,668	5.39%	8.85%	85.76%
2003	674	516	6,908	8,098	8.32%	6.37%	85.31%
2008	964	325	7,232	8,521	11.31%	3.81%	84.87%
2013	1,184	934	6,885	9,003	13.15%	10.37%	76.47%
2014	2,340	690	12,224	15,254	15.34%	4.52%	80.14%
2019	2,992	10	11,480	14,482	20.66%	0.07%	79.27%
2021	3,482	405	11,983	15,870	21.94%	2.55%	75.51%
Grand Total	12,321	5,409	76,779	94,509	13.04%	5.72%	81.24%

Per facilitare la lettura dei dati, creo una tabella pivot su Excel e calcolo le percentuali di donne, uomini e genere mancante rispetto al totale.

Nel corso degli anni aumenta in generale il numero di diplomatici. Aumenta sì la presenza femminile, ma non in modo proporzionale rispetto a quella maschile: gli uomini sono in media oltre l’80% del totale; tale percentuale cala solo leggermente a favore delle donne nel corso degli anni, senza però raggiungere la parità.

Dal 2008 al 2021, ad eccezione del 2013, si nota un aumento della presenza delle donne in contemporanea al diminuire della percentuale di persone per le quali il genere risulta mancante nel dataset. Questo potrebbe far supporre un miglioramento nell’inserimento dei dati.

In alternativa, è possibile lanciare una query per eseguire le stesse operazioni direttamente con SQL: sostituire i valori 0, 1 e 99 con valori testuali e calcolare le percentuali rispetto al totale per ogni valore del campo *gender*.

Con la prima SELECT individuo il conteggio totale e lo rinomino con l'alias *total* in modo da poterlo utilizzare temporaneamente come campo nella seconda parte della query. Con la seconda SELECT calcolo le percentuali rispetto al totale per ogni valore del campo *gender*. Raggruppo per anno e per genere e sostituisco i valori 0, 1 e 99 con valori testuali.

```
WITH total AS
(SELECT sqlgendip.year, COUNT(*) as total
FROM sqlgendip
GROUP BY sqlgendip.year)
SELECT sqlgendip.year,
CASE
WHEN gender = '0' THEN 'uomini'
WHEN gender = '1' THEN 'donne'
WHEN gender = '99' THEN 'mancante'
END AS gender,
COUNT(*)/total as percentuale
FROM sqlgendip
LEFT JOIN total
ON total.year = sqlgendip.year
GROUP BY total, sqlgendip.year, gender
```

Il questo modo l'analisi risulta più efficiente e chiara, in quanto leggibile più facilmente grazie anche all'utilizzo dell'alias: con un'unica query è possibile riunirne diverse.

OBIETTIVO: analizzare la disparità di genere confrontando le aree geografiche

QUERY: raggruppamento per genere e per region\_send e ordinare i dati per region\_send in ordine crescente

```
SELECT region_send, gender, COUNT(*)  
FROM sqlgendip  
GROUP BY gender, region_send  
ORDER BY region_send ASC
```

	ABC region_send	ABC gender	123 COUNT(*)
1	0	0	13.593
2	0	99	1.468
3	0	1	2.451
4	1	0	13.989
5	1	99	1.350
6	1	1	1.268
7	2	0	6.341
8	2	99	258
9	2	1	1.907
10	3	99	1.306
11	3	0	24.305
12	3	1	3.825
13	4	0	8.383
14	4	1	512
15	4	99	444
16	5	0	2.862
17	5	99	154
18	5	1	1.008
19	6	0	1.635
20	6	99	261
21	6	1	504
22	7	0	5.671
23	7	99	168
24	7	1	846

Per una visualizzazione più chiara, ripeto il conteggio confrontando le aree geografiche e isolando i dati di ognuno dei tre valori del campo 'gender' con query distinte.

```
⊖ SELECT region_send, COUNT(*)  
  FROM sqlgendip  
 WHERE gender = '0'  
 GROUP BY region_send  
 ORDER BY region_send ASC
```

```
⊖ SELECT region_send, COUNT(*)  
  FROM sqlgendip  
 WHERE gender = '1'  
 GROUP BY region_send  
 ORDER BY region_send ASC
```

```
⊖ SELECT region_send, COUNT(*)  
  FROM sqlgendip  
 WHERE gender = '99'  
 GROUP BY region_send  
 ORDER BY region_send ASC
```

	ABC region_send ▼	123 COUNT(*) ▼	123 COUNT(*) ▼	123 COUNT(*) ▼
1	0	13.593	2.451	1.468
2	1	13.989	1.268	1.350
3	2	6.341	1.907	258
4	3	24.305	3.825	1.306
5	4	8.383	512	444
6	5	2.862	1.008	154
7	6	1.635	504	261
8	7	5.671	846	168

Unendo gli screenshot, ottengo 3 colonne con conteggi rispettivamente per uomini, donne e genere mancante. Le regioni geografiche n. 2, 5 e 6 (Central and North America, Nordic countries, Oceania) sono quelle in cui, seppur in assenza di parità di genere, le donne sono presenti in misura maggiore, rappresentando circa il 20-25% del totale, e la presenza degli uomini si attesta intorno al 70%. Al contrario, le regioni n. 1 e 4 (Asia e Middle East) sono quelle in cui la presenza delle donne è minima: il genere femminile costituisce solo il 5-7% circa del totale, mentre gli uomini superano l'80% e sfiorano il 90%.



OBIETTIVO: capire se e come è presente disparità di genere tra uomo e donna per ognuno dei titoli

QUERY: raggruppamento per titolo e ordinare i dati per titolo in ordine crescente. Lancio una query per isolare i dati relativi agli uomini e una per i dati relativi alle donne.

```
⊖ SELECT title, COUNT(*)  
  FROM sqlgendip  
 WHERE gender = '0'  
 GROUP BY title  
 ORDER BY title ASC
```

```
⊖ SELECT title, COUNT(*)  
  FROM sqlgendip  
 WHERE gender = '1'  
 GROUP BY title  
 ORDER BY title ASC
```

	ABC title	123 COUNT(*)	123 COUNT(*)
1	1	2.977	463
2	2	68	5
3	3	73.410	11.663
4	96	87	174
5	97	63	8
6	98	90	6
7	99	84	2

Unisco gli screenshot e ottengo 2 colonne con conteggi rispettivamente per uomini e per donne. Su Excel calcolo le percentuali di donne e di uomini rispetto al totale per ognuno dei titoli.


title	uomini	donne	totale	% uomini su totale	% donne su totale
1	2,977	463	3,440	86.54%	13.46%
2	68	5	73	93.15%	6.85%
3	73,410	11,663	85,073	86.29%	13.71%
96	87	174	261	33.33%	66.67%
97	63	8	71	88.73%	11.27%
98	90	6	96	93.75%	6.25%
99	84	2	86	97.67%	2.33%

Dai dati emerge che i titoli 2 (*minister, internuncios*), 98 (*other*; per esempio: *head of mission, head of section, secretary, consul, diplomatic representative, envoy, deputy high commissioner, bureau chief, etc.*) e 99 (*mancante*) sono quelli ricoperti in percentuale minore da donne (tra il 2.33% e il 6.85%). Invece, il titolo 96 (*acting chargé d'affaires*) è ricoperto in maggioranza da donne, quasi al 67% contro il 33.33% di uomini.

OBIETTIVO: capire se e come è presente la disparità di genere in Italia nel corso del tempo

QUERY: raggruppamento per anno e per genere e ordino i dati per genere in ordine crescente

```
⊖ SELECT gender, year, COUNT(*)  
FROM sqlgendip  
WHERE cname_send = 'Italy'  
GROUP BY year, gender  
ORDER BY gender ASC
```

	 gender	year	COUNT(*)
1	0	1968	71
2	0	1978	91
3	0	1988	98
4	0	1998	115
5	0	2003	112
6	0	2008	107
7	0	2013	105
8	0	2014	172
9	0	2019	103
10	0	2021	139
11	1	1988	1
12	1	1998	2
13	1	2003	6
14	1	2008	8
15	1	2013	9
16	1	2014	16
17	1	2019	15
18	1	2021	17
19	99	1968	7
20	99	1978	10
21	99	1988	3
22	99	1998	6
23	99	2008	6
24	99	2013	6
25	99	2014	7

anno	donna	mancante	uomo	Grand Total	% donne su totale	% gen. mancante su totale	% uomini su totale
1968		7	71	78	0.00%	8.97%	91.03%
1978		10	91	101	0.00%	9.90%	90.10%
1988	1	3	98	102	0.98%	2.94%	96.08%
1998	2	6	115	123	1.63%	4.88%	93.50%
2003	6		112	118	5.08%	0.00%	94.92%
2008	8	6	107	121	6.61%	4.96%	88.43%
2013	9	6	105	120	7.50%	5.00%	87.50%
2014	16	7	172	195	8.21%	3.59%	88.21%
2019	15		103	118	12.71%	0.00%	87.29%
2021	17		139	156	10.90%	0.00%	89.10%
<b>Grand Total</b>	74	45	1113	1232	6.01%	3.65%	90.34%

Per facilitare la lettura dei dati, creo una tabella pivot su Excel e calcolo le percentuali di donne, uomini e genere mancante rispetto al totale.

Negli anni 1968, 1978 e 1988 la percentuale di donne in Italia nel dataset analizzato è pari a zero, mentre si riscontra la presenza di diplomatici classificati con genere mancante. Potremmo forse dedurre che in quest'ultimo caso si tratti di donne, in quanto negli anni successivi la percentuale relativa al genere mancante diminuisce e parallelamente quella delle donne aumenta, fino ad arrivare al 2019 e al 2021, anni nei quali non vengono classificati diplomatici con genere mancante e le donne raggiungono oltre il 10% del totale. È altresì vero che questo aspetto potrebbe rispecchiare pure un cambio dei tempi e delle società in direzione di una maggiore presenza di genere femminile, sebbene emerga come processo lento e difficoltoso, data la presenza degli uomini in media sempre sopra l'87% del totale, anche in anni recenti.



OBIETTIVO: capire se e come è presente disparità di genere tra uomo e donna in Italia per ognuno dei titoli

QUERY: raggruppamento per titolo e ordinare i dati per titolo in ordine crescente

```
⊖ SELECT title, COUNT(*)  
  FROM sqlgendip  
 WHERE gender = '0'  
 AND cname_send = 'Italy'  
 GROUP BY title  
 ORDER BY title ASC
```

```
⊖ SELECT title, COUNT(*)  
  FROM sqlgendip  
 WHERE gender = '1'  
 AND cname_send = 'Italy'  
 GROUP BY title  
 ORDER BY title ASC
```

	ABC title ▼	123 COUNT(*) ▼
1	1	10
2	3	1.102
3	98	1

	ABC title ▼	123 COUNT(*) ▼
1	3	74

I diplomatici inviati dall'Italia rientrano in 3 categorie di titoli:

- chargé d'affaires (*title 1*)
- ambassador, high commissioners papal nuncios, designate ambassador, ambassador-at-large, secretary of the people's bureau (*title 3*)
- other (*title 98*)

La maggioranza è costituita da uomini; le donne, che sono 74, ricoprono solo incarichi che fanno parte della categoria del *title 3*.