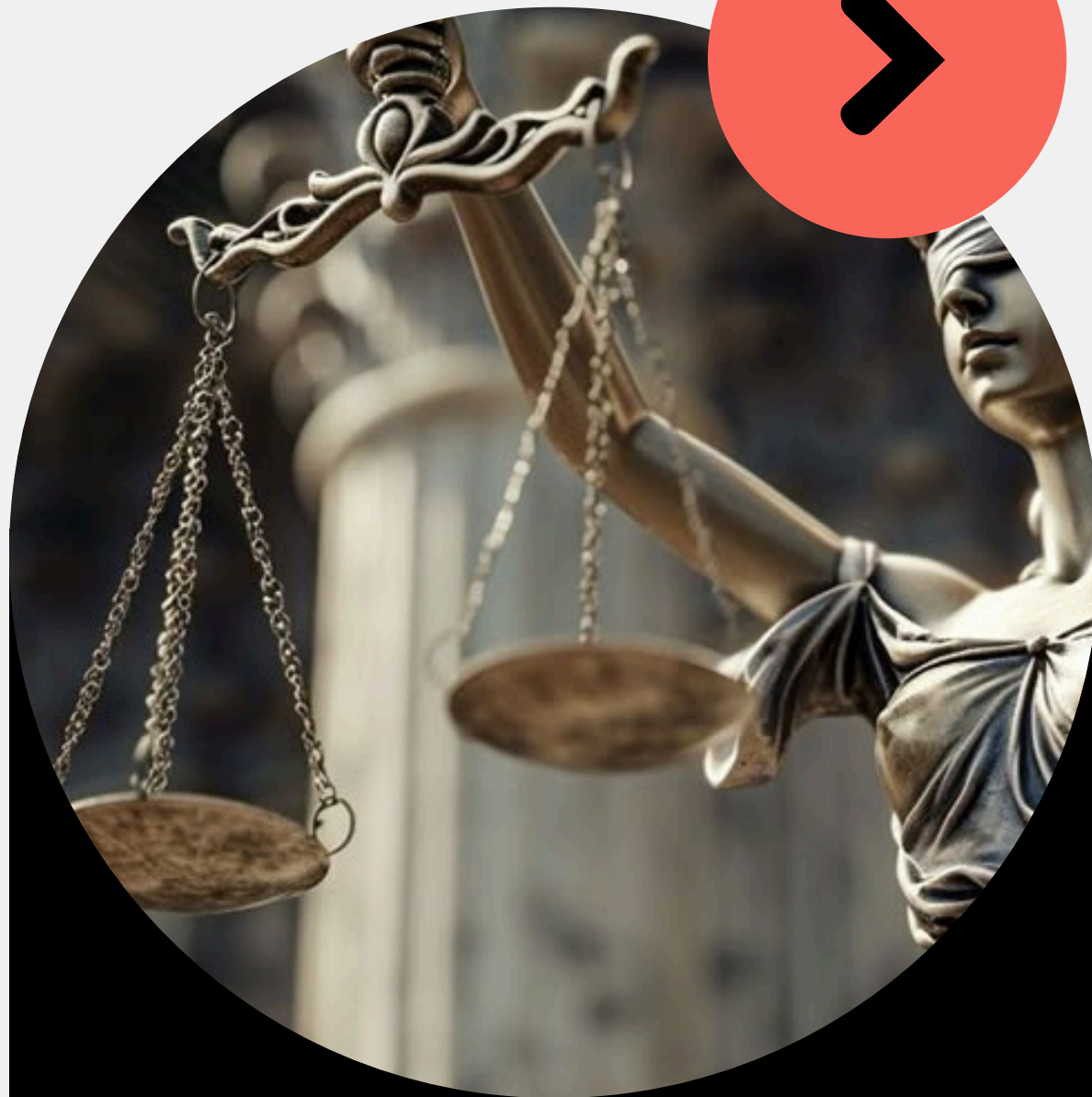


SQL – Gender Diplomacy Information



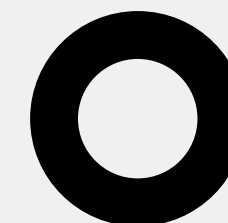


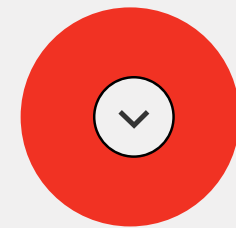
Introduzione

Questo progetto mira ad analizzare la disparità di genere tra i diplomatici dal 1968 al 2021. Per comprendere le varie voci dei dati ho utilizzato il codebook per associarli ai valori testuali.

L'obiettivo è capire dove sono stati fatti progressi e quali regioni necessitano di miglioramenti, esaminando:

1. L'evoluzione del divario di genere nel tempo, confrontando il numero di diplomatici uomini e donne
2. I paesi con una politica estera femminista, analizzando chi invia e riceve più donne nelle missioni diplomatiche
3. La distribuzione degli incarichi principali, main posting e concurrent accreditations
4. Le aree geografiche con maggiore disparità, identificando le regioni con il più alto squilibrio di genere

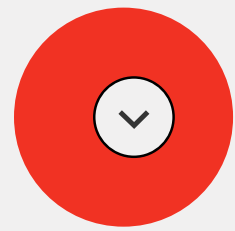




Pulizia dei dati

Per garantire un'analisi accurata, ho effettuato prima un processo di pulizia dati su Excel, nel foglio di lavoro originario. Ho rimosso tutti i duplicati e le celle vuote. Dopodichè, ho formattato le colonne con valori numerici e salvato il foglio di lavoro con i dati puliti in formato CSV.



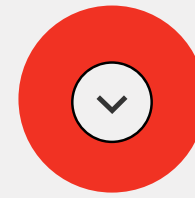


Creazione tabella

Dopo aver creato un nuovo database (geninformation) su PostgreSQL, creo la tabella gendip_dataset con tutti i nominativi delle colonne che la compongono.



```
CREATE TABLE IF NOT EXISTS gendip_dataset
(
  ID SERIAL PRIMARY KEY,
  year integer,
  cname_send character varying(200),
  main_posting integer,
  title integer,
  gender integer,
  cname_receive character varying(200),
  ccode_send integer,
  ccodealp_send character varying(100),
  ccodecow_send integer,
  region_send integer,
  gme_send integer,
  v2lgfemleg_send numeric(10,2),
  ffp_send integer,
  ccode_receive integer,
  ccodealp_receive character varying(200),
  ccodecow_receive integer,
  region_receive integer,
  gme_receive integer,
  ffp_receive integer
);
```



Importazione dei dati

Durante l'importazione dei dati del file CSV in PostgreSQL, ho riscontrato un errore eseguendo COPY, che mi suggeriva di eseguire il comando sul lato client. Per risolvere, ho usato il comando `\copy` da psql, che ha permesso di caricare il file correttamente.



```
geninformation=# \COPY gendip_dataset (year, cname_send, main_posting, title, gender, cname_receive, ccod
e_send, ccodealp_send, ccodecow_send, region_send, gme_send, v2lgfemleg_send,ffp_send, ccode_receive, cco
dealp_receive, ccodecow_receive, region_receive, gme_receive, ffp_receive) FROM 'C:\Temp\GenDip_Dataset_p
ublic_mainpostings_anonymous_1968-2021_2023-05-30.csv' DELIMITER ';' CSV HEADER;
COPY 94509
geninformation=#
```







Query 1

```
--1. Si è evoluto nel tempo il divario di genere?  
-- Nel primo step andremo ad analizzare se anno per anno la maggioranza di genere è uomo o donna  
-- Raggruppiamo gli anni singoli (dal 1968 al 2021)  
-- Per per ogni anno creiamo due colonne per sapere il conteggio del genere uomo=0 e donna=1  
  
WITH Gendercount AS (  
  SELECT  
    year,  
    SUM(CASE WHEN gender = 0 THEN 1 ELSE 0 END) AS male_count,  
    SUM(CASE WHEN gender = 1 THEN 1 ELSE 0 END) AS female_count  
  FROM gendip_dataset  
  WHERE gender IN (0,1) --escludiamo i valori mancanti (99)  
  GROUP BY YEAR  
)  
SELECT  
  year,  
  male_count,  
  female_count,  
  CASE  
    WHEN male_count > female_count THEN 'male'  
    WHEN female_count > male_count THEN 'female'  
    ELSE 'equal'  
  END AS most_frequent_gender  
FROM Gendercount  
ORDER BY year;
```





Analisi query 1

	year integer 	male_count bigint 	female_count bigint 	most_frequent_gender text 
1	1968	3135	29	male
2	1978	4592	81	male
3	1988	5761	161	male
4	1998	6572	413	male
5	2003	6908	674	male
6	2008	7232	964	male
7	2013	6884	1184	male
8	2014	12073	2332	male
9	2019	11479	2990	male
10	2021	11981	3482	male




```
1  --1. Dai risultati mostrati dalla tabella possiamo notare che il divario di genere nei diplomatici
2  -- non si è ridotto nel tempo, anzi sembra essersi amplificato.
3  -- Sebbene il numero di donne sia aumentato gradualmente, la crescita degli uomini è stata
4  -- ancora più marcata, mantenendo un forte squilibrio.
5  -- Questo indica che, nonostante una maggiore presenza femminile, le opportunità per le donne nella
6  -- diplomazia sono rimaste molto più limitate rispetto agli uomini.
```




Query 2


```
-- 2. I paesi che inviano i propri diplomatici all'estero promuovono una politica estera femminista?  
  
-- Step 1. Sapere la quantità totale dei paesi presenti nell'anno 1968 e nell'anno 2021 per fare  
-- un confronto dei dati nei rispettivi anni
```

```
SELECT COUNT(DISTINCT cname_send) AS numero_paesi  
FROM gendip_dataset  
WHERE year = 1968;
```

	numero_paesi 
1	135

1968

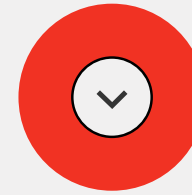
```
SELECT COUNT(DISTINCT cname_send) AS numero_paesi  
FROM gendip_dataset  
WHERE year = 2021;
```

	numero_paesi 
1	197

2021



Query 2



```
-- Step 2. Seleziono solo i paesi che non promuovono una politica estera femminista (no)
```

```
SELECT DISTINCT  
cname_send AS paese,  
CASE WHEN ffp_send = 0 THEN 'No' ELSE 'Si'  
END AS politica_femminista  
FROM gendip_dataset  
WHERE ffp_send = 0  
AND year = 1968;
```

1968

2021

```
SELECT DISTINCT  
cname_send AS paese,  
CASE WHEN ffp_send = 0 THEN 'No' ELSE 'Si'  
END AS politica_femminista  
FROM gendip_dataset  
WHERE ffp_send = 0  
AND year = 2021;
```

cname_send =
paese di invio del
diplomatico

ffp_send = se i
paesi di invio
promuovono una
politica
femminista

Query 2

135 paesi su 135 hanno votato no

paese character varying (200)	politica_femminista text
Kenya	No
DominicanRepublic	No
Somalia	No
NewZealand	No
Mexico	No
Cuba	No
Jamaica	No
Afghanistan	No
Chad	No
Gambia	No
Sweden	No
Yugoslavia	No
CostaRica	No
Uruguay	No
Portugal	No
ElSalvador	No
SaudiArabia	No
Egypt	No
GermanDemocraticRepublic	No
SouthAfrica	No
Madagascar	No
Italy	No

1968

190 paesi su 197 hanno votato no,
gli altri 7 hanno votato a favore

	paese character varying (200)	politica_femminista text
169	Niue	No
170	Jamaica	No
171	Chad	No
172	Zimbabwe	No
173	Gambia	No
174	CostaRica	No
175	Uruguay	No
176	Bahamas	No
177	Czechia	No
178	Slovenia	No
179	Portugal	No
180	ElSalvador	No
181	NorthMacedonia	No
182	SaudiArabia	No
183	VietNam	No
184	Egypt	No
185	Croatia	No
186	SouthAfrica	No
187	Micronesia,FederatedStatesof	No
188	Madagascar	No
189	Italy	No
190	Qatar	No

2021



Analisi query 2

--2. I paesi che inviano i propri diplomatici all'estero promuovono una politica estera femminista?

--2. I risultati mostrano che nel 1968, nessun paese promuoveva una politica estera femminista, mentre nel 2021 solo 7 paesi hanno dichiarato di farlo, evidenziando un progresso molto limitato in questo ambito. Al contrario, i paesi che non promuovono una politica femminista sono aumentati, passando da 135 nel 1968 a 190 nel 2021. La maggior parte dei paesi non ha adottato ancora approcci concreti in ambito diplomatico per promuovere la parità di genere.



Query 3

--3. Quali aree geografiche mostrano maggiori disparità di genere?
--Step 1. Creo un'altra tabella dove converto i valori numerici dei continenti in testo
--Step 2. Dopo aver creato la tabella region_send_dip la unisco alla tab principale gendip_dataset (innerjoin)
--Step 3. Faccio la differenza dei risultati finali e li confronto

✓ **WITH** disparità **AS** (
 SELECT
 r.region_send_text **AS** aree_geografiche, --seleziona i continenti in testo dalla tabella region_send_dip
 COUNT(*) **FILTER** (**WHERE** g.gender = 0) **AS** uomini, --calcola il n. di uomini separatamente
 COUNT(*) **FILTER** (**WHERE** g.gender = 1) **AS** donne --calcola il n. di donne separatamente
 FROM gendip_dataset g
 INNER JOIN region_send_dip r **ON** g.region_send = r.region_send_id
 WHERE g.gender **IN** (0,1) --escludo i valori numerici (99) mancanti dalla colonna gender
 AND g.region_send != 9999 --escludo i valori numerici mancanti dalla colonna continenti
 GROUP BY r.region_send_text --raggruppo i continenti
)
 SELECT
 aree_geografiche,
 uomini,
 donne,
 uomini - donne **AS** differenza
 FROM disparità
 ORDER BY differenza **DESC**;







Query 3

```
CREATE TABLE region_send_dip (  
  region_send_id INTEGER PRIMARY KEY,  
  region_send_text VARCHAR(150)  
);  
  
INSERT INTO region_send_dip (region_send_id, region_send_text)  
VALUES  
  (0, 'Africa'),  
  (1, 'Asia'),  
  (2, 'Central and Noth America'),  
  (3, 'Europe'),  
  (4, 'Middle East'),  
  (5, 'Nordic Countries'),  
  (6, 'Oceania'),  
  (7, 'South America'),  
  (9999, 'missing');
```



Analisi query 3

-- 3. La tabella evidenzia che in Asia, Africa e Medio Oriente presentano le disparità di genere più forti,
-- con un numero significativamente maggiore di uomini rispetto alle donne.
-- Europa, pur mostrando una disparità di 20.459 persone, appare minore rispetto ad altre aree perchè ha
-- un numero più elevato di donne rispetto a regioni come l'Asia (dove le donne sono solo 1268).
-- Le aree con la minore disparità includono i paesi nordici e Oceania, dove la differenza tra uomini e donne
-- è meno marcata.

	aree_geografiche character varying (150) 	uomini bigint 	donne bigint 	differenza bigint 
1	Europe	24279	3820	20459
2	Asia	13941	1268	12673
3	Africa	13560	2449	11111
4	Middle East	8374	511	7863
5	South America	5661	846	4815
6	Central and Noth America	6329	1905	4424
7	Nordic Countries	2854	1008	1846
8	Oceania	1619	503	1116






Query 4

```
--4.Quali incarichi sono assegnati più frequentemente a uomini e donne?

SELECT
CASE
WHEN gender = 0 THEN 'uomo'
WHEN gender = 1 THEN 'donna'
END AS genere,
SUM(CASE WHEN main_posting = 1 THEN 1 ELSE 0 END) AS tot_main_posting,
SUM(CASE WHEN main_posting = 0 THEN 1 ELSE 0 END) AS tot_concurrent_accreditations
FROM gendip_dataset
WHERE gender IN (0,1)
GROUP BY gender;
```

Totale
incarichi dal
1968 al 2021



	genere 	tot_main_posting 	tot_concurrent_accreditations 
1	uomo	61708	14906
2	donna	8263	4047

Query 4

```
1  --Incarico principale nel 1968
2
3  ✓ SELECT
4  CASE
5  WHEN gender = 0 THEN 'Uomo'
6  WHEN gender = 1 THEN 'Donna'
7  END AS genere,
8  SUM (CASE WHEN main_posting = 1 THEN 1 ELSE 0 END) AS tot_main_posting,
9  SUM (CASE WHEN main_posting = 0 THEN 1 ELSE 0 END) AS tot_concurrent_accreditations
10 FROM gendip_dataset
11 WHERE gender IN (0,1)
12 AND year = 1968
13 GROUP BY gender;
```

Data Output Messages Notifications



	genere text	tot_main_posting bigint	tot_concurrent_accreditations bigint
1	Uomo	3135	0
2	Donna	29	0

Risultati
dell'anno
1968



Query 4

```
1  --Incarico principale nel 2021
2
3  ✓ SELECT
4  CASE
5  WHEN gender = 0 THEN 'Uomo'
6  WHEN gender = 1 THEN 'Donna'
7  END AS genere,
8  SUM (CASE WHEN main_posting = 1 THEN 1 ELSE 0 END) AS tot_main_posting,
9  SUM (CASE WHEN main_posting = 0 THEN 1 ELSE 0 END) AS tot_concurrent_accreditations
10 FROM gendip_dataset
11 WHERE gender IN (0,1)
12 AND year = 2021
13 GROUP BY gender;
```

Data Output Messages Notifications

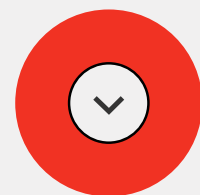


	genere text	tot_main_posting bigint	tot_concurrent_accreditations bigint
1	Uomo	6632	5349
2	Donna	1781	1701

Risultati
dell'anno 2021



Analisi query 4



- 4. I dati mostrano una forte disparità di genere nelle assegnazioni diplomatiche nel periodo 1968-2021.
- Nel totale, gli uomini hanno ricevuto un numero significativamente maggiore di incarichi principali rispetto alle donne.
- Anche per le accreditazioni concorrenti - ovvero quando un diplomatico oltre a lavorare nella sede principale è accreditato per rappresentare il proprio paese anche in altri stati- gli uomini superano le donne, evidenziando una presenza femminile ridotta in questo settore.
- Tuttavia, l'analisi nel tempo mostra un miglioramento nella rappresentanza femminile, sebbene la disparità persista.

Link alla cartella drive



<https://drive.google.com/drive/folders/1hAJd-xHHTinEf5ZuI19jZrczax74D2Hn?usp=sharing>