

# PROGETTO SQL

Rappresentazione di genere nella diplomazia

FRANCESCA CARRERA



# Indice

1. Introduzione - pag. 2.
2. Estrazione dei dati - pag. 3-5.
  - a. Colonne della tabella estratta - pag. 5.
3. Trasformazione dei dati - pag. 6-26.
  - a. Fase di Design del Database - pag. 6-8.
  - b. Normalizzazione dei dati - pag. 9-26.
4. Caricamento dei dati - pag. 27.
  - a. Diagramma ER - pag. 28.
5. Analisi dei dati accompagnati dai grafici - pag 29-70.



Obiettivi ONU



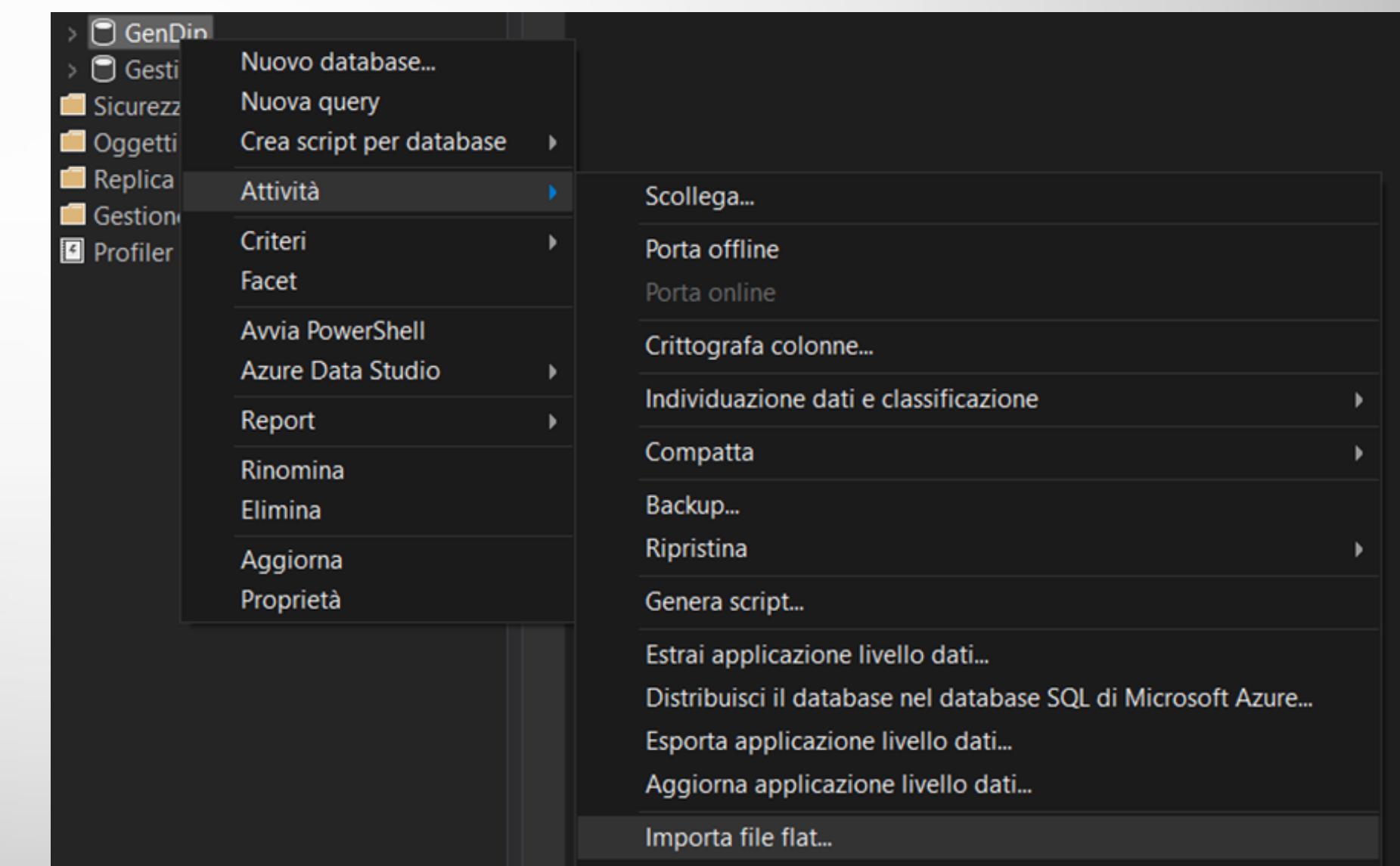
# Introduzione

- Nel set di dati GenDip sono raccolte informazioni riguardo la nomina di uomini, donne e altri diplomatici, da parte degli Stati, per diversi tipi di incarichi di ambasciatori bilaterali.
- In questa presentazione verranno illustrate le fasi del processo di estrazione, trasformazione e caricamento dei dati (ETL), seguite dall'analisi dei dati e dalla rappresentazione visiva delle informazioni attraverso grafici su Excel.
- Ho utilizzato l'implementazione di SQL Server, nella versione Express, all'interno dell'ambiente di sviluppo di SQL Server Management Studio (SSMS). Inoltre, ho impiegato Excel per la rappresentazione del diagramma Entità-Relazione e per la visualizzazione dei grafici.

# Estrazione dei dati

- Dopo aver provveduto alla creazione del database GenDip, sono passata all'importazione del file CSV tramite SSMS.
- Prima di procedere, ho convertito il file di lavoro scaricato dal formato xlsx a CSV.

```
CREATE DATABASE GenDip  
GO  
USE GenDip
```



- Ho definito tutte le colonne come tipo di dato *nvarchar(10)*, tranne per le colonne *cname\_send* e *cname\_receive* che ho lasciato sempre come tipo *nvarchar* ma a 100, ho flaggato l'opzione ‘Consenti valori Null’ e ho nominato la mia tabella di parcheggio *stagingArea*.
- Ho optato per il tipo di dato *nvarchar*, e non per *varchar*, poiché il primo supporta l'archiviazione di caratteri Unicode e il file che sto importando potrebbe contenere caratteri non appartenenti al set di caratteri ASCII standard.  
In questo modo garantisco la corretta rappresentazione di questi eventuali caratteri, semplificando il processo di importazione e riducendo la possibilità di perdita di dati o di errori di conversione.
- Il rovescio della medaglia è che *nvarchar* richiede il doppio dello spazio di archiviazione rispetto a *varchar*.

Nome colonna	Tipo di dati	Chiave primaria	<input checked="" type="checkbox"/> Consenti valori Null
year	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>cname_send</i>	<i>nvarchar(100)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>main_posting</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>title</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>gender</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>cname_receive</i>	<i>nvarchar(100)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>ccode_send</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>ccodealp_send</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>ccodeCOW_send</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>region_send</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>GME_send</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>v2lgfemleg_send</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>FFP_send</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>ccode_receive</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>ccodealp_receive</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>ccodeCOW_receive</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>region_receive</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>GME_receive</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<i>FFP_receive</i>	<i>nvarchar(10)</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

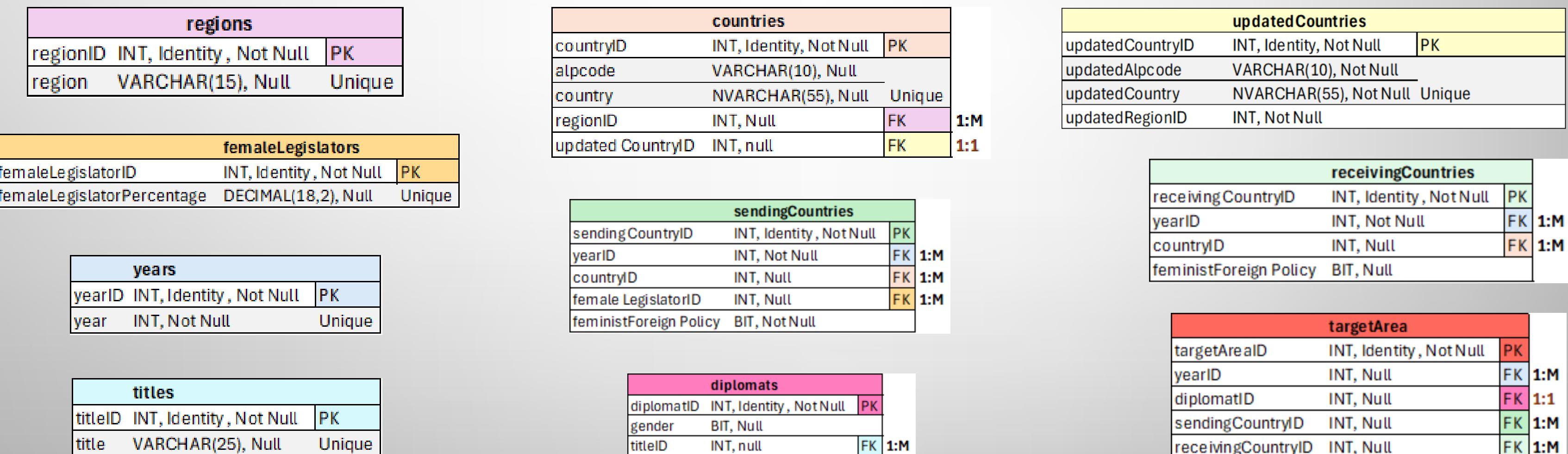
## Colonne della tabella estratta

- **year**  
Anno in cui è avvenuta una determinata missione diplomatica.
- **main\_posting**  
L'incarico principale di un diplomatico per un determinato anno. Ho escluso questa colonna dal mio database di lavoro, a causa della mancanza di chiarezza nelle fonti e nei criteri utilizzati per determinare il dato.
- **title**  
Titolo del diplomatico.
- **gender**  
La classificazione del genere del diplomatico.
- **cname\_send, cname\_receive**  
Paese di invio e di destinazione.
- **v2lgfemleg\_send**  
Si riferisce alla percentuale di legislatori femminili nella camera bassa nel paese di invio.
- **FFP\_send, FFP\_receive**  
Dato che indica se un paese di invio o di destinazione ha dichiarato di aderire ad una politica estera femminista.
- **{ccode\_send, ccodealp\_send, GME\_send, region\_send, ccodeCOW\_send}**  
Identificatori impiegati per fornire informazioni e misure sui paesi di invio in elenco. Di questi terrò in considerazione la regione geografica (*region*) e il codice alpha a tre caratteri (*ccodealp*).
- **{ccode\_receive, ccodealp\_receive, GME\_receive, region\_receive, ccodeCOW\_receive}**  
Questi identificatori corrispondono a quelli utilizzati per i paesi di invio, ma in questa versione sono specificamente associati ai paesi di destinazione.

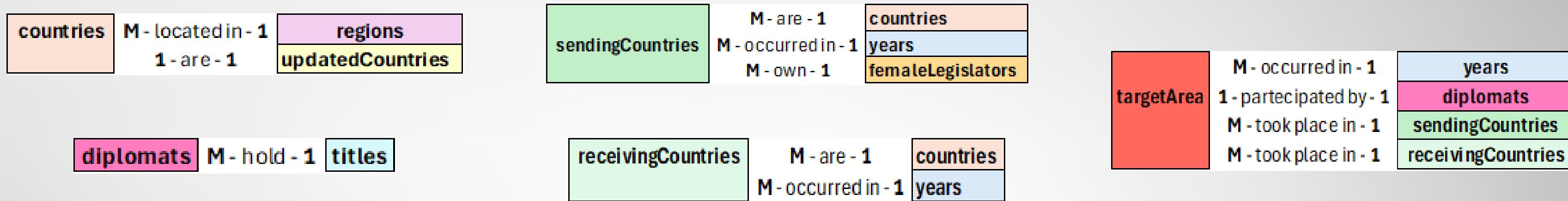
# Trasformazione dei dati

## Fase di design del database

Per iniziare, ho definito la struttura logica del database utilizzando un diagramma ER online. Successivamente, ho trascritto questa struttura su un foglio di Excel per una migliore visualizzazione e organizzazione.



## Tipo di relazione tra entità



- **Caratteristiche delle chiavi esterne.**

Le chiavi esterne (FK) indicate sono principalmente opzionali e possono essere multivaleure.

Tuttavia, ci sono due eccezioni:  
*updatedCountryID* nella tabella *countries* e *diplomatID* nella tabella *targetArea*.

Queste FK, sebbene opzionali, sono univoche poiché sono attributi di entità coinvolte in relazioni uno a uno.

Al contrario, le altre FK sono attributi di entità coinvolte in relazioni uno a molti.

- **Unicità degli attributi.**

Gli attributi *region* di *regions*, *femaleLegislatorPercentage* di *femaleLegislator*, *year* di *years* e *title* di *titles* sono valori univoci, il che significa che ciascun valore in queste colonne è unico all'interno della rispettiva tabella.

D'altra parte, le coppie di attributi *alpcode* e *country* in *countries*, *updatedAlpcode* e *updatedCountry* in *updatedCountries* formano coppie univoche. Questo significa che ogni combinazione di valori in queste colonne è unica all'interno delle rispettive tabelle.

- **Entità supertipo con specializzazione totale sovrapposta.**
  - L'entità *countries* svolge il ruolo di entità supertipo con una specializzazione totale sovrapposta rispetto alle entità figlie *sendingCountries* e *receivingCountries*.
  - Questo significa che l'entità *countries* può essere specializzata sia come paese di invio (tramite l'entità *sendingCountries*) che come paese di destinazione (tramite l'entità *receivingCountries*).
  - Questa specializzazione totale sovrapposta implica che ogni riga nella tabella 'countries' può essere associata sia a una riga nella tabella *sendingCountries* che a una riga nella tabella *receivingCountries*.
  - Gli attributi condivisi tra l'entità supertipo e le entità figlie includono *countryID*, che suggerisce che la specializzazione avviene in base a questo attributo. Inoltre, le entità figlie possono avere attributi aggiuntivi per soddisfare le loro esigenze specifiche, mentre l'entità supertipo può mantenere attributi comuni a tutti i paesi, come *regionID* e *alpcode*.

Creazione tab. *workingArea* con le colonne di mio interesse da tab. *stagingArea*.

```
SELECT IDENTITY(int, 1, 1) AS workingAreaID,  
       year,  
       gender,  
       title,  
       NULL          AS cname_sendID,  
       cname_send,  
       ccodealp_send,  
       region_send,  
       FFP_send,  
       vlgfemleg_send,  
       NULL          AS cname_receiveID,  
       cname_receive,  
       ccodealp_receive,  
       region_receive,  
       FFP_receive  
INTO   dbo.workingArea  
FROM   dbo.stagingArea;
```

- *cname\_sendID* e *cname\_receiveID* fungono da collegamento momentaneo, tra tab. *countries* (che non ho ancora creato) e tab. *workingArea*, per apportare modifiche ai paesi di invio e di destinazione.
- Aggiunta del vincolo alla chiave primaria.

```
ALTER TABLE dbo.workingArea  
ADD CONSTRAINT pk_workingArea PRIMARY KEY (workingAreaID);  
-- PK_<TableName>
```

## Normalizzazione dei dati

Verifica della presenza di caratteri non standard, sistemazione dei NULL e dei tipi di dato delle colonne.

- Ho personalizzato la normalizzazione dei dati a seconda della colonna e del contesto.
- Riporto la normalizzazione della colonna *cname\_send* (che ho applicato allo stesso modo per *cname\_receive*), essendo l'unica colonna in cui ho trovato caratteri non standard che ho eliminato.
  - .
- Faccio una verifica generale dei dati presenti nella colonna.

```
SELECT ccodealp_send  
FROM dbo.workingArea  
GROUP BY ccodealp_send  
ORDER BY ccodealp_send;  
-- OUTPUT '9999' per NULL
```

- Con la funzione PATINDEX() ricerco caratteri non standard, inclusi spazi non standard, nello specifico né numeri, né lettere maiuscole, né lettere minuscole.

```
SELECT ccodealp_send  
FROM dbo.workingArea  
WHERE PATINDEX('%[^0-9a-zA-Z ]%', ccodealp_send) > 0  
GROUP BY ccodealp_send;
```

- Visualizzo e identifico il carattere non standard presente.

```

SELECT RIGHT(ccodealp_send, 1) AS LastChar,
       ASCII(RIGHT(ccodealp_send, 1)) AS LastCharAsciiCode
  FROM dbo.workingarea
 WHERE PATINDEX('%[^0-9a-zA-Z ]%', ccodealp_send) > 0
 GROUP BY RIGHT(ccodealp_send, 1),
          ASCII(RIGHT(ccodealp_send, 1));
-- OUTPUT: LastChar: ' ' - LastCharAsciiCode: 160 (no-break space).

```

- Elimino lo "spazio non separabile".

```

UPDATE dbo.workingArea
SET ccodealp_send = REPLACE(ccodealp_send, CHAR(160), '');

```

- Sostituisco il valore '9999' con NULL.

```

UPDATE dbo.workingArea
SET ccodealp_send = NULL
WHERE ccodealp_send = '9999'

```

- Aggiorno il tipo di dato.

```

ALTER TABLE dbo.workingArea
ALTER COLUMN ccodealp_send VARCHAR(10);

```

- Creazione e popolamento di tab *regions* e *countries*.

```
-- Creazione e popolamento di tab regions.
CREATE TABLE dbo.regions (
    regionID INT IDENTITY(1,1),
    region VARCHAR(15) NULL,
    CONSTRAINT PK_regions PRIMARY KEY (regionID),
        -- PK_TargetTable
    CONSTRAINT UQ_regions_region UNIQUE (region)
        -- UQ_TargetTable_TargetColumn
)
GO
INSERT INTO dbo.regions (region)
VALUES (NULL), ('Africa'), ('Antarctica'),
        ('Asia'), ('Europe'), ('North America'),
        ('Oceania'), ('South America');
```

```
-- Creazione e popolamento di tab countries.
WITH cnameUNION_CTE
    AS (SELECT cname_send,
                ccodealp_send,
                region_send
        FROM dbo.workingArea
        UNION
        SELECT cname_receive,
                ccodealp_receive,
                region_receive
        FROM dbo.workingArea)
SELECT IDENTITY(int, 1, 1) AS countryID,
        NULL AS countryID2,
        cname_send AS country,
        ccodealp_send AS alpcode,
        region_send AS regionID,
        NULL AS updatedCountryID
    INTO dbo.countries
    FROM cnameUNION_CTE
    ORDER BY cname_send;
    SET countryID2 = countryID;

-- Popolamento di countryID2, copia di countryID.
UPDATE dbo.countries
SET countryID2 = countryID;
```

## **Creazione e popolamento di tab *countries***

Impiego la CTE per estrarre valori univoci dei paesi di invio e di destinazione con i loro rispettivi dati, indicanti la regione geografica di appartenenza e il codice alpha a tre caratteri. Utilizzerò poi questa lista per popolare parte della tabella *countries*, a cui ho aggiunto la colonna *countryID*, che fungerà da futura PK, e una sua copia. Ho aggiunto, inoltre, la colonna *updatedCountryID*, il cui ruolo sarà di chiave esterna in riferimento alla PK della tab *updatedCountries*, non ancora creata.

### **Funzione della colonna *countryID2***

Sarà la copia di *countryID*, in versione non *IDENTITY* in modo che sia modificabile. Come *cname\_sendID* e *cname\_receiveID* fungono da collegamento momentaneo di tab *workingArea*, *countryID2* svolge il medesimo ruolo di tab *countries*, per apportare modifiche ai paesi di invio e di destinazione in entrambe le tabelle menzionate.

## Popolamento delle colonne *cname\_sendID* e *cname\_receiveID* di tab *workingArea*.

- Ho creato una *Stored Procedure* per popolare *cname\_sendID* e *cname\_receiveID* con il valore di *countryID*, poichè impiegherò nuovamente l'istruzione.
- Utilizzo l'operatore *INTERSECT* in questo contesto come forma più concisa di verificare la corrispondenza tra le colonne delle due tabelle, senza dover scrivere esplicitamente tutte le condizioni di corrispondenza per ogni colonna, gestendo automaticamente i valori *NULL*.
- Personalizzo poi la procedura per popolare la colonna *cname\_receiveID* allo stesso modo.

```
CREATE PROCEDURE usp_cnameSendID_UPDATE
    -- userStoredProcedure_column_action
AS
BEGIN
    SET NOCOUNT OFF;
    -- implicito ma lo scrivo: desidero vedere il conteggio delle righe.
    UPDATE W
    SET W.cname_sendID = C.countryID
    FROM dbo.workingArea AS W
        LEFT JOIN dbo.countries AS C
            ON EXISTS (SELECT W.cname_send,
                           W.ccodealp_send,
                           W.region_send
                    INTERSECT
                    SELECT C.country,
                           C.alpcode,
                           C.regionid)
END;
GO
EXEC usp_cnameSendID_UPDATE;

-- NB. versione con forma non concisa.
UPDATE W
SET W.cname_sendID = C.countryID
FROM dbo.workingArea AS W
INNER JOIN dbo.countries AS C
    ON ( W.cname_send = C.country
        OR ( W.cname_send IS NULL
              AND C.country IS NULL ) )
    AND ( W.ccodealp_send = C.alpcode
          OR ( W.ccodealp_send IS NULL
                AND C.alpcode IS NULL ) )
    AND ( W.region_send = C.regionID );
```

## Tab countries: verifica dei valori NULL nelle colonne.

```
SELECT countryID2,
       country,
       alpcode,
       regionID
  FROM dbo.countries
 WHERE country IS NULL
   OR alpcode IS NULL
   OR regionID IS NULL;
```

countryID2	country	alpcode	regionID	
1	NULL	NULL	1	✓
15	Azores	NULL	6	✗
118	Korea	NULL	4	✓
121	Kosovo	NULL	5	✓
212	South Ossetia	NULL	4	✗
249	Virgin Islands	NULL	6	✓
252	Yemen	NULL	4	✗

- Azores fa parte del Portogallo,
- Non è meglio specificato se si tratti di Corea del Nord o del Sud,
- Kosovo non ha un *alpcode* riconosciuto in maniera ufficiale,
- South Ossetia fa ufficialmente parte della Georgia,
- Non è meglio specificato se Virgin Islands sia British o Usa,
- Sono presenti diverse versioni del record con *country* Yemen.

Quando necessario ho apportato modifiche ai record con valore NULL nella tab *countries* e *workingArea*.

- Riporto le modifiche del record con *country* Yemen.

```
SELECT countryID2,
       country,
       alpcode,
       regionID
  FROM dbo.countries
 WHERE country LIKE '%Yemen%';
```

- Eliminazione dei record errati da tab *countries*.

```
DELETE FROM dbo.countries
 WHERE countryID2 IN (252, 255);
```

- Tab *workingArea*: controllo nei paesi d'invio.

```
SELECT cname_sendID,
       cname_send,
       ccodealp_send,
       region_send
  FROM dbo.workingArea
 WHERE cname_send LIKE '%Yemen%'
 GROUP BY cname_sendID,
          cname_send,
          ccodealp_send,
          region_send;
```

- Aggiornamento dei record contenenti Yemen nei paesi d'invio.

```
UPDATE dbo.workingArea
 SET cname_sendID = 253,
     ccodealp_send = 'YEM'
 WHERE cname_sendID = 252;
```

```
UPDATE dbo.workingArea
 SET cname_sendID = 254,
     ccodealp_send = 'YEM'
 WHERE cname_sendID = 255;
```

- Aggiornamento non necessario nei paesi di destinazione.

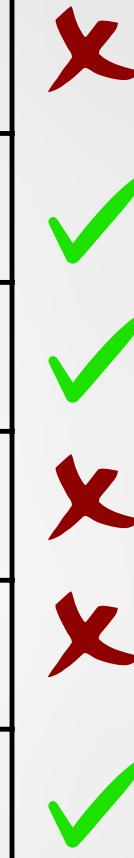
countryID2	country	alpcode	regionID
252	Yemen	NULL	4
253	Yemen	YEM	4
254	Yemen, Arab Republic of	YEM	4
255	Yemen, Arab Republic of	YME	4
256	Yemen, People's Democratic Republic of	YMD	4

## Correzione delle regioni geografiche discordanti nei paesi con valori duplicati..

```
WITH duplicateCountries_CTE
AS (SELECT countryID2,
country,
alpcode,
regionID
FROM dbo.countries
WHERE country IN (SELECT country
FROM dbo.countries
GROUP BY country
HAVING COUNT(*) > 1)),
denseRankRegions_CTE
AS (SELECT countryID2,
country,
alpcode,
regionID,
DENSE_RANK()
OVER (
PARTITION BY country
ORDER BY regionID) AS denseRank_regionID
FROM duplicateCountries_CTE)
SELECT C.countryID2,
C.country,
C.alpcode,
C.regionID,
R.region
FROM dbo.countries AS C
LEFT JOIN denseRankRegions_CTE AS DC
ON C.country = DC.country
LEFT JOIN dbo.regions AS R
ON C.regionID = R.regionID
WHERE DC.denseRank_regionid > 1;
C.alpcode,
C.regionID,
R.region
FROM dbo.countries AS C
LEFT JOIN denseRank_regions AS DC
ON C.country = DC.country
LEFT JOIN dbo.regions AS R
ON C.regionID = R.regionID
WHERE DC.denseRank_regionid > 1;
```

- Nella CTE *duplicateCountries* seleziono i paesi duplicati, utilizzando una sottoquery nella clausola WHERE.
- Nella CTE *denseRank\_regions* utilizzo la funzione finestra *DENSE\_RANK()* per assegnare un rango ai paesi duplicati identificati nella CTE precedente, basato sulla ripetizioni (o non) delle regioni geografiche.
- Nella SELECT finale unisco la tabella *countries* con la CTE *denseRank\_regions*, filtrando dalla stessa i paesi con i valori delle regioni geografiche non ripetute, che risulteranno essere quindi discordanti.

<i>countryID2</i>	<i>country</i>	<i>alpcode</i>	<i>regionID</i>	<i>region</i>
137	Maldives	MDV	2	Africa
138	Maldives	MDV	4	Asia
167	North Macedonia	MKD	5	Europe
168	North Macedonia	MKD	7	Oceania
208	Solomon Islands	SLB	4	Asia
209	Solomon Islands	SLB	7	Oceania



Nella tab *countries* ho cancellato i record errati.

Nella tab *workingArea* ho apportato le modifiche nei paesi d'invio e di destinazione dove necessario.

## Correzione dei codici alpha (a tre caratteri) duplicati.

- Nella CTE *duplicateAlpcode* seleziono i codici presenti più di una volta.
- Nella SELECT unisco la tabella countries con la CTE *duplicateAlpcode*, visualizzando così i dettagli dei paesi con codici alpha duplicati.

- Nella tab *countries* ho cancellato i record errati.
- Nella tab *workingArea* ho apportato le modifiche nei paesi d'invio e di destinazione dove necessario.

<i>countryID2</i>	<i>country</i>	<i>alpcode</i>
14	Azerbaijan	AZE
232	Turkmenistan	AZE
22	Belize	BLZ
100	Honduras	BLZ
42	Central African Empire	CAF
43	Central African Republic	CAF
57	Cyprus	CYP
58	Czechia	CYP

```

WITH duplicateAlpcode_CTE
AS (SELECT alpcode,
           COUNT(alpcode) AS alpcodeCount
      FROM dbo.countries
     GROUP BY alpcode
    HAVING COUNT(alpcode) > 1)
SELECT C.countryID2, C.country, DA.alpcode
  FROM duplicateAlpcode_CTE AS DA
 LEFT JOIN dbo.countries AS C
    ON DA.alpcode = C.alpcode;
  
```

85	Germany	DEU	✓
86	Germany, Federal Republic of	DEU	✓
38	Cambodia	KHM	✓
114	Kampuchea	KHM	✓
119	Korea, Democratic People's Republic of	PRK	✗
166	Niue	PRK	✓
246	Viet Nam	VNM	✓
248	Viet Nam, Republic of	VNM	✓
253	Yemen	YEM	✓
254	Yemen, Arab Republic of	YEM	✓

## Creazione e popolamento di tab *updatedCountries*.

- Ho scaricato da Wikipedia un file con dati aggiornati sui paesi attuali con nome completo, codici alpha a tre caratteri e regione geografica.
- Per snellire il processo ho verificato i dati sul file Excel, apportando le modifiche dove necessario, in modo da rendere i nomi dei paesi grammaticalmente identici a quelli presenti nella tabella *countries*.
- Come fatto in precedenza con la tabella *stagingArea*, ho importato il file CVS intitolando la nuova tabella *updatedCountries* rinominando le colonne tramite finestra grafica.

Nome colonna	Tipo di dati	Chiave primaria	<input checked="" type="checkbox"/> Consenti valori Null
updatedAlpcode	nvarchar(10)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
updatedCountry	nvarchar(100)	<input type="checkbox"/>	<input checked="" type="checkbox"/>
updatedRegionID	nvarchar(10)	<input type="checkbox"/>	<input checked="" type="checkbox"/>

```
ALTER TABLE dbo.updatedCountries  
    ADD updatedCountryID INT IDENTITY(1,1);
```

```
ALTER TABLE dbo.updatedCountries  
    ADD CONSTRAINT PK_updatedCountries PRIMARY  
KEY (updatedCountryID);  
-- PK_TargetTable
```

```
ALTER TABLE dbo.updatedCountries  
    ALTER COLUMN updatedAlpcode VARCHAR(10) NOT NULL;
```

```
ALTER TABLE dbo.updatedCountries  
    ALTER COLUMN updatedCountry NVARCHAR(55) NOT NULL;
```

```
ALTER TABLE dbo.updatedCountries  
    ALTER COLUMN updatedRegionID INT NOT NULL;
```

```
ALTER TABLE dbo.updatedCountries  
    ADD CONSTRAINT UQ_updatedCountries_alpcode_country UNIQUE (updatedAlpcode, updatedCountry);  
-- UQ_TargetTable_TargetColumn1_TargetColumn2
```

- Aggiungo la colonna *updatedCountryID* e la definisco come chiave primaria della tabella.
- Dopo aver verificato che non fossero presenti valori NULL nelle colonne, ho modificato il vincolo rendendolo NOT NULL.
- Ho creato un vincolo di unicità composto dalla combinazione di due colonne.

## Regioni geografiche non combacianti tra tab *countries* e *updatedCountries*

```
WITH unupdatedCountries_CTE
      AS (SELECT country,
                 alpcode,
                 regionID
            FROM dbo.countries
           EXCEPT
        SELECT updatedCountry,
               updatedAlpcode,
               updatedRegionID
          FROM dbo.updatedCountries)
SELECT C.countryID2,
       C.alpcode,
       C.country,
       C.regionID AS      'wrong regionID',
       UC.updatedRegionID 'correct regionID'
  FROM dbo.countries AS C
 INNER JOIN unupdatedCountries_CTE AS CTE
         ON C.country = CTE.country
 LEFT JOIN dbo.updatedCountries AS UC
        ON C.country = UC.updatedCountry
 WHERE UC.updatedRegionID IS NOT NULL
 ORDER BY C.alpcode;
```

- Nella CTE *unupdatedCountries* seleziono, tramite EXCEPT, i paesi presenti nella tab *countries*, ma non in *updatedCountries*.
- Nella SELECT visualizzo i dati ottenuti dalla CTE accompagnandoli, inoltre, alla regione geografica aggiornata, presente nella tab *updatedCountries*.

<i>countryID2</i>	<i>alpcode</i>	<i>country</i>	<i>wrong regionID</i>	<i>correct regionID</i>
80	ATF	French Southern Territories	2	3
56	CUW	Curaçao	8	6
66	EGY	Egypt	4	2
97	GUY	Guyana	2	8
169	MNP	Northern Mariana Islands	4	7
180	PCN	Pitcairn	4	7
228	TON	Tonga	2	7

- 1 NULL
- 2 Africa
- 3 Antarctica
- 4 Asia
- 5 Europe
- 6 North America
- 7 Oceania
- 8 South America

- Correggo i valori errati di *regionID* nella tab *countries* tramite una *Stored Procedure* da riutilizzare quando necessario.

```

CREATE PROCEDURE usp_regionID_UPDATE
    @new_regionID INT,
    @current_countryID INT
AS
BEGIN
    SET NOCOUNT OFF;
    --隐式的，但我还是写出来：我想看行数。
    UPDATE dbo.countries
    SET regionID = @new_regionID
    WHERE countryID2 = @current_countryID;
END;

```

## Paesi non più esistenti.

<i>unupdatedCountry</i>	<i>alpcode</i>	<i>regionID</i>
Central African Empire	CAF	2
Czechoslovakia	CSK	5
German Democratic Republic	DDR	5
Germany, Federal Republic of	DEU	5
Kampuchea	KHM	4
Serbia and Montenegro	SCG	5
USSR	SUN	5
Viet Nam, Democratic Republic of	VDR	4
Viet Nam, Republic of	VNM	4
Yemen, Arab Republic of	YEM	4
Yemen, People's Democratic Republic of	YMD	4
Yugoslavia	YUG	5

```

SELECT country AS 'unupdatedCountry',
       alpcode,
       regionID
  FROM dbo.countries
 WHERE country IS NOT NULL
   AND alpcode IS NOT NULL
EXCEPT
SELECT updatedCountry,
       updatedAlpcode,
       updatedRegionID
  FROM dbo.updatedCountries;

```

- Infine, ho definito i vincoli di chiave primaria, chiave esterna e i vincoli UNIQUE dove necessario in tab *countries* e *workingArea*.

- Tab *years*: creazione e popolamento.

```
SELECT
    IDENTITY(INT, 1,1) AS yearID,
    year
INTO dbo.years
FROM dbo.workingArea
GROUP BY year
ORDER BY year ASC;
```

- Tab *femaleLegislators*: creazione e popolamento.

```
SELECT
    IDENTITY(INT, 1,1) AS femaleLegislatorID,
    v2lgfemleg_send AS femaleLegislatorPercentage
INTO dbo.femaleLegislators
FROM dbo.workingArea
GROUP BY v2lgfemleg_send
ORDER BY v2lgfemleg_send ASC;
```

- Tab *sendingCountries*: creazione e popolamento.

```
SELECT
    IDENTITY(INT, 1,1) AS sendingCountryID,
    Y.yearID,
    W.cname_sendID AS countryID,
    F.femaleLegislatorID,
    W.FFP_send AS feministForeignPolicy
INTO dbo.sendingCountries
FROM dbo.years AS Y
INNER JOIN dbo.workingArea AS W
    ON Y.year = W.year -- NO NULL
LEFT JOIN dbo.femaleLegislators AS F
    ON EXISTS (SELECT W.v2lgfemleg_send
                INTERSECT
                SELECT F.femaleLegislatorPercentage)
GROUP BY Y.yearID,
    W.cname_sendID,
    F.femaleLegislatorID,
    W.FFP_send
ORDER BY W.cname_sendID ASC,
    Y.yearID ASC;
```

- Tab *titles*: creazione e popolamento.

```

CREATE TABLE dbo.titles (
    titleID INT IDENTITY(1,1),
    title NVARCHAR(50) NULL,
    CONSTRAINT PK_titles PRIMARY KEY (titleID),
        -- PK_TargetTable
    CONSTRAINT UQ_titles_title UNIQUE (title)
        -- UQ_TargetTable_TargetColumn
)
GO
INSERT INTO dbo.titles(title)
VALUES (NULL), ('Acting ambassador'), ('Acting chargé d'affaires'),
('Ambassador'), ('Chargé d'affaires'), ('Minister'), ('Other');

```

- Per ciascuna tabella creata, ho provveduto a definire e creare vincoli di chiave primaria e esterna.

- Tab *receivingCountries*: creazione e popolamento.

- Tab *diplomats*: creazione e popolamento.

```

SELECT
    IDENTITY(INT, 1,1) AS diplomatID,
    W.gender,
    T.titleID AS titleID
INTO dbo.diplomats
FROM dbo.workingArea AS W
LEFT JOIN dbo.titles AS T
    ON (W.title = T.titleID
        OR (W.title IS NULL AND T.titleID IS NULL));

```

```

SELECT
    IDENTITY(INT, 1,1) AS receivingCountryID,
    Y.yearID,
    W.cname_receiveID AS countryID,
    W.FFP_receive AS feministForeignPolicy
INTO dbo.receivingCountries
FROM dbo.years AS Y
    INNER JOIN dbo.workingArea AS W
        ON Y.year = W.year
GROUP BY Y.yearID, W.cname_receiveID, W.FFP_receive
ORDER BY W.cname_receiveID ASC,
    Y.yearID ASC;

```

# Caricamento dei dati

- Tab *targetArea*: creazione e popolamento.

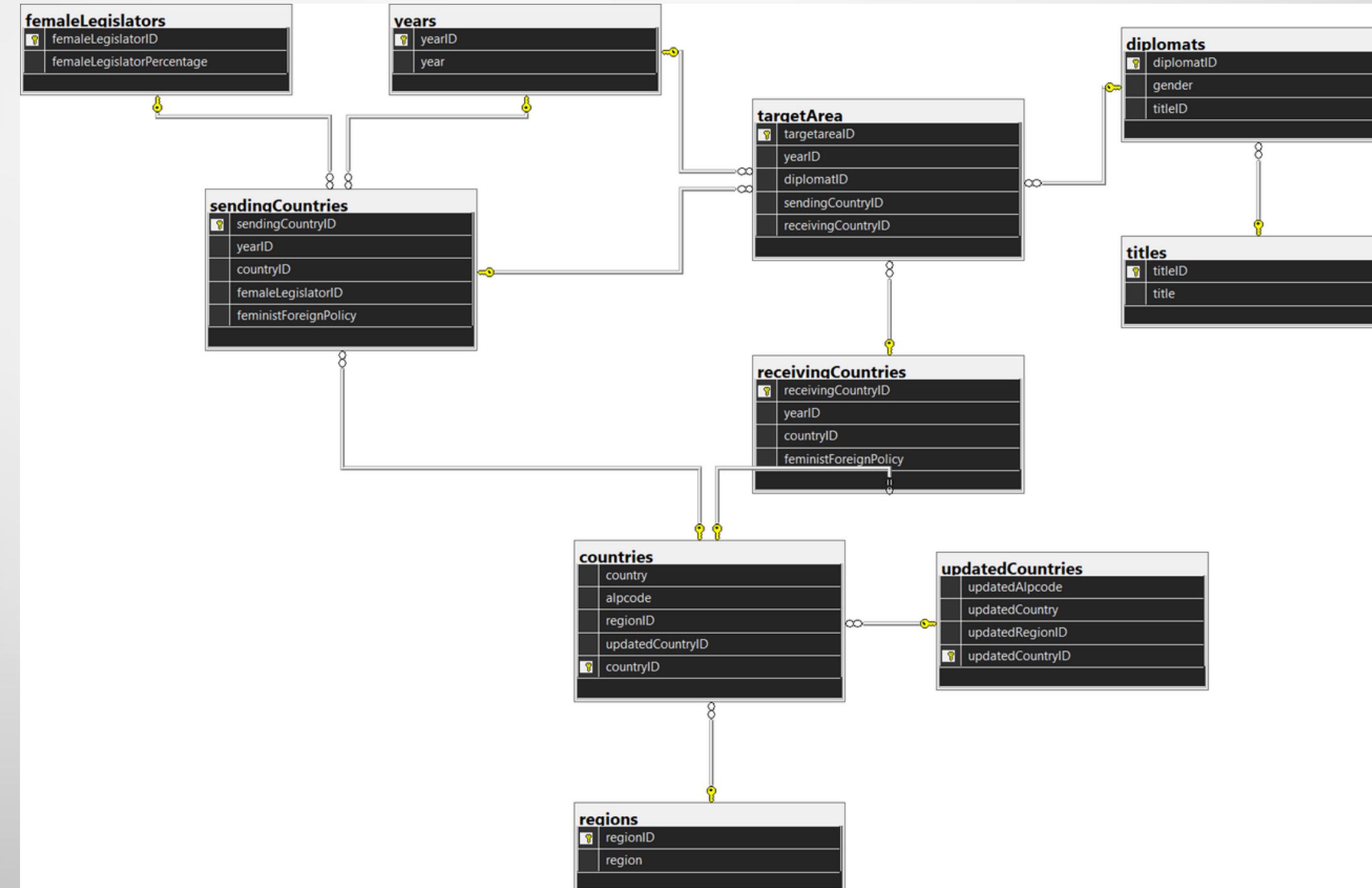
```
SELECT  
    IDENTITY(INT, 1,1) AS targetareaID,  
    Y.yearID,  
    NULL AS diplomatID,  
    S.sendingCountryID,  
    R.receivingCountryID  
INTO dbo.targetArea  
FROM dbo.workingArea AS W  
    LEFT JOIN dbo.years AS Y  
        ON W.year = Y.year  
    LEFT JOIN dbo.femaleLegislators AS F  
        ON W.v2lgfemleg_send = F.femaleLegislatorPercentage  
        OR (W.v2lgfemleg_send IS NULL AND F.femaleLegislatorPercentage IS NULL)  
    LEFT JOIN dbo.sendingCountries AS S  
        ON EXISTS (SELECT Y.yearID,  
                    W.cname_sendID,  
                    F.femaleLegislatorID,  
                    W.FFP_send  
                INTERSECT  
                SELECT S.yearID,  
                    S.countryID,  
                    S.femaleLegislatorID,  
                    S.feministForeignPolicy)  
    LEFT JOIN dbo.receivingCountries AS R  
        ON EXISTS (SELECT Y.yearID,  
                    W.cname_receiveID,  
                    W.FFP_receive  
                INTERSECT  
                SELECT R.yearID,  
                    R.countryID,  
                    R.feministForeignPolicy);
```

- Collego la tabella *diplomats* direttamente a parte, poiché non è collegata alla tabella *targetArea* tramite la tabella *workingArea*.

```
UPDATE T  
SET T.diplomatID = D.diplomatID  
FROM dbo.targetArea AS T  
    INNER JOIN dbo.diplomats AS D  
        ON T.targetareaID = D.diplomatID;
```

- La tabella *targetArea* rappresenta l'area di destinazione o l'obiettivo dei dati trasformati e preparati per l'elaborazione successiva.
- Contiene i dati elaborati e filtrati provenienti dalla tabella *stagingArea*, pronti per essere analizzati e impiegati per la rappresentazione dei grafici.
- Ora i dati sono disponibili nella loro forma finale e il processo *ETL* può considerarsi concluso.

# Diagramma ER



# Analisi dei dati

## **PRIMO FOCUS: genere e titoli dei diplomatici.**

- 1) Conteggio totale e annuale di diplomatici per genere.
  - a) Maggior numero di diplomatici per genere e in quale anno.
  - b) Rapporto di crescita di diplomatici, per genere, del 2021 rispetto al 1968.
  - c) Crescita percentuale maggiore di diplomatici di genere maschile.
- 2) Paesi con più diplomatici di genere femminile per anno e in assoluto.
  - a) Regioni geografiche con più diplomatici di genere femminile in assoluto.
- 3) Conteggio dei titoli per categoria.
  - a) Conteggio di ogni titolo per genere.

## Creazione vista per raggruppare paesi esistenti e datati

- La vista si concentra sui paesi che hanno cambiato nome ma hanno mantenuto la loro identità nazionale fondamentale. Per questo motivo, ho scelto di escludere dalla vista paesi come Yugoslavia, Serbia e Montenegro e Czechoslovakia, poiché hanno subito divisioni che hanno generato nuove entità nazionali.

```
-- vw_unifiedCountries
CREATE VIEW vw_unifiedCountries AS
SELECT countryID, alpcode, regionID,
CASE
    WHEN country = 'Central African Empire' THEN
        'Central African Republic' → DA: id 41 - A: 42
    WHEN country IN ('German Democratic Republic',
                     'Germany, Federal Republic of')
        THEN 'Germany' → DA: id 82 e 84 - A: 83
    WHEN country = 'Kampuchea' THEN
        'Cambodia' → DA: id 111 - A: 37
    WHEN country = 'USSR' THEN
        'Russian Federation' → DA: id 232 - A: 180
    WHEN country IN ('Viet Nam, Democratic Republic of',
                     'Viet Nam, Republic of')
        THEN 'Viet Nam' → DA: id 237 e 238 - A: 236
    WHEN country IN ('Yemen, Arab Republic of',
                     'Yemen, People''s Democratic Republic of')
        THEN 'Yemen' → DA: id 243 e 244 - A: 242
    ELSE country
END AS unifiedCountries
FROM dbo.countries
GROUP BY countryID, alpcode, regionID,
CASE
    WHEN country = 'Central African Empire' THEN
        'Central African Republic' → DA: id 41 - A: 42
    WHEN country IN ('German Democratic Republic',
                     'Germany, Federal Republic of')
        THEN 'Germany' → DA: id 82 e 84 - A: 83
    WHEN country = 'Kampuchea' THEN
        'Cambodia' → DA: id 111 - A: 37
    WHEN country = 'USSR' THEN
        'Russian Federation' → DA: id 232 - A: 180
    WHEN country IN ('Viet Nam, Democratic Republic of',
                     'Viet Nam, Republic of')
        THEN 'Viet Nam' → DA: id 237 e 238 - A: 236
    WHEN country IN ('Yemen, Arab Republic of',
                     'Yemen, People''s Democratic Republic of')
        THEN 'Yemen' → DA: id 243 e 244 - A: 242
    ELSE country
END;
```

- Ho progettato la vista *vw\_unifiedCountries* per raggruppare i paesi esistenti e datati sotto un'unica denominazione, rappresentando il nome attuale e definitivo di ciascun paese. Questo ha reso i dati più chiari e comprensibili, evitando la duplicazione dei paesi con nomi diversi ma con lo stesso significato durante l'analisi dei dati. Inoltre, ciò ha eliminato la necessità di gestire manualmente le varianti dei nomi dei paesi.

<i>countryID</i>	<i>alpcode</i>	<i>regionID</i>	<i>unifiedCountries</i>
37	KHM	4	Cambodia
41	CAF	2	Central African Republic
42	CAF	2	Central African Republic
82	DDR	5	Germany
83	DEU	5	Germany
84	DEU	5	Germany
111	KHM	4	Cambodia

180	RUS	5	Russian Federation
232	SUN	5	Russian Federation
236	VNM	4	Viet Nam
237	VDR	4	Viet Nam
238	VNM	4	Viet Nam
242	YEM	4	Yemen
243	YEM	4	Yemen
244	YMD	4	Yemen

# 1) Conteggio totale e annuale di diplomatici per genere.

```
SELECT COALESCE(CAST(Y.year AS VARCHAR(10)), 'tot') AS year,  
FORMAT(COUNT(CASE  
WHEN D.gender = 1 THEN 1  
END), '#,0') AS femaleDiplomats,  
FORMAT(COUNT(CASE  
WHEN D.gender = 0 THEN 1  
END), '#,0') AS maleDiplomats  
FROM dbo.diplomats AS D  
INNER JOIN dbo.targetArea AS T  
ON D.diplomatID = T.diplomatID  
INNER JOIN dbo.years AS Y  
ON T.yearID = Y.yearID  
GROUP BY Y.year WITH ROLLUP  
-- La clausola WITH ROLLUP consentirà di ottenere  
-- anche il totale generale delle colonne.  
ORDER BY Y.year ASC;
```

year	femaleDiplomats	maleDiplomats
tot	12.321	76.779
1968	29	3.136
1978	81	4.593
1988	162	5.762
1998	413	6.576
2003	674	6.908
2008	964	7.232
2013	1.184	6.885
2014	2.340	12.224
2019	2.992	11.480
2021	3.482	11.983

- *COALESCE(CAST(Y.year AS VARCHAR(10)), 'tot') AS year* restituirà la colonna *year*, che sarà il valore di *Y.year* convertito in stringa se non è nullo, altrimenti restituirà il valore '*tot*'. Con la clausola *WITH ROLLUP*, il valore '*tot*' verrà utilizzato per rappresentare il totale generale.
- Con questo codice creo, inoltre, una vista denominata *VW\_diplomatsByGender* con conteggio annuale di diplomatici per genere (senza formattazione e clausola *WITH ROLLUP*) da riutilizzare.

- 1a) Maggior numero di diplomatici per genere e in quale anno.

```
WITH femaleDiplomats_CTE
AS (SELECT TOP 1 year,
FORMAT(femaleDiplomats, '#,0') AS femDip_MAX
FROM VW_diplomatsByGender
ORDER BY femaleDiplomats DESC)

, maleDiplomats_CTE
AS (SELECT TOP 1 year,
FORMAT(maleDiplomats, '#,0') AS maleDip_MAX
FROM VW_diplomatsByGender
ORDER BY maleDiplomats DESC)

SELECT FD.year, FD.femDip_MAX,
MD.maleDip_MAX, MD.year
FROM femaleDiplomats_CTE AS FD
CROSS JOIN maleDiplomats_CTE AS MD;
```

<i>year</i>	<i>femDip_MAX</i>	<i>maleDip_MAX</i>	<i>year</i>
2021	3.482	12.224	2014

- 1b) Crescita percentuale maggiore di diplomatici di genere maschile.

```
WITH maleDiff_CTE
AS ( SELECT year,
maleDiplomats,
LAG(maleDiplomats)
OVER (
        ORDER BY Year)
        AS male_prevYearCount,
maleDiplomats - LAG(maleDiplomats)
OVER (
        ORDER BY Year) AS male_diff
FROM VW_diplomatsByGender
),
maleDiff_pct_CTE
AS ( SELECT year,
FORMAT(ROUND( (male_diff * 100.0 / NULLIF(male_prevYearCount, 0) )
, 2)
, '0.00') + '%' AS maleDiff_pct
FROM maleDiff_CTE
)
SELECT TOP 1 year, maleDiff_pct
FROM maleDiff_pct_CTE
ORDER BY maleDiff_pct DESC;
```

<i>year</i>	<i>maleDiff_pct</i>
2014	77,55%

## 1b) Crescita percentuale maggiore di diplomatici di genere maschile - DETTAGLIO CODICE

- Output in dettaglio di *maleDiff\_CTE*.

year	maleDiplomats	male_prevYearCount	male_diff
1968	3136	NULL	NULL
1978	4593	3136	1457
1988	5762	4593	1169
1998	6576	5762	814
2003	6908	6576	332
2008	7232	6908	324
2013	6885	7232	-347
2014	12224	6885	5339
2019	11480	12224	-744
2021	11983	11480	503

- Nella CTE *maleDiff\_CTE* viene calcolato il numero di diplomatici di genere maschile per ogni anno e la differenza tra il numero di diplomatici maschili di un determinato anno rispetto a quello precedente.

- Nello specifico:

- con la funzione finestra *LAG* mostro il numero di diplomatici maschili dell'anno precedente per ogni riga.
- con il codice

'maleDiplomats - LAG(maleDiplomats) OVER (ORDER BY Year) AS male\_diff'

calcolo la differenza tra il numero di diplomatici maschili dell'anno corrente e quello dell'anno precedente.

## 1b) Crescita percentuale maggiore di diplomatici di genere maschile - DETTAGLIO CODICE

year	maleDiff_pct
1968	NULL
1978	46,46%
1988	25,45%
1998	14,13%
2003	5,05%
2008	4,69%
2013	-4,80%
2014	77,55%
2019	-6,09%
2021	4,38%

- Output in dettaglio di *maleDiff\_pct\_CTE*.
- Nella CTE *maleDiff\_pct\_CTE* vengono calcolate le percentuali di variazione tra il numero di diplomatici maschili di un anno e quello dell'anno precedente.
- Nello specifico:
  - gestisce i casi in cui il numero di diplomatici maschili dell'anno precedente è zero. Restituisce *NULL* se il valore è zero, altrimenti restituisce il numero di diplomatici maschili dell'anno precedente.  
`'NULLIF(male_prevYearCount, 0)'`
  - calcola la variazione percentuale tra il numero di diplomatici maschili dell'anno corrente e quello dell'anno precedente, *ROUND* 2 per arrotondare il risultato a due cifre decimali.  
`'ROUND( (male_diff * 100.0 / NULLIF(male_prevYearCount, 0) ), 2)'`
  - per formattare il risultato come stringa, aggiungendo il simbolo percentuale alla fine, creando così la rappresentazione della variazione percentuale.  
`'FORMAT( ... , '0.00') + '%' AS maleDiff_pct'`
- Nella *SELECT* finale viene visualizzato il maggior incremento percentuale di diplomatici, maschili con l'anno di riferimento.

## 1c) Rapporto di crescita di diplomatici, per genere, del 2021 rispetto al 1968.

```
WITH femaleDip_CTE AS (
    SELECT
        SUM(CASE
            WHEN year = 2021 THEN femaleDiplomats
            END) AS femaleDiplomats_2021,
        SUM(CASE
            WHEN year = 1968 THEN femaleDiplomats
            END) AS femaleDiplomats_1968
    FROM VW_diplomatsByGender
),
maleDip_CTE AS (
    SELECT
        SUM(CASE
            WHEN year = 2021 THEN maleDiplomats
            END) AS maleDiplomats_2021,
        SUM(CASE
            WHEN year = 1968 THEN maleDiplomats
            END) AS maleDiplomats_1968
    FROM VW_diplomatsByGender
)
SELECT FORMAT(ROUND ( CONVERT(DECIMAL(18,2), femaleDiplomats_2021 ) /
NULLIF( CONVERT(DECIMAL(18,2), femaleDiplomats_1968 ), 0 )
, '0.00', 2) AS femaleDip_growthRatio,
FORMAT(ROUND ( CONVERT(DECIMAL(18,2), maleDiplomats_2021 ) /
NULLIF( CONVERT(DECIMAL(18,2), maleDiplomats_1968 ), 0 )
, '0.00', 2) AS maleDip_growthRatio
FROM femaleDip_CTE
CROSS JOIN maleDip_CTE;
```

femaleDip_growthRatio	maleDip_growthRatio
120,07	3,82

## 1c) Rapporto di crescita di diplomatici, per genere, del 2021 rispetto al 1968 - DETTAGLIO CODICE.

Nella CTE *femaleDip\_CTE* vengono calcolate le somme dei diplomatici di genere femminili per due anni specifici, il 2021 e il 1968.

Nella CTE *maleDip\_CTE* vengono calcolate le somme dei diplomatici di genere maschile per due anni specifici, il 2021 e il 1968.

Nella *SELECT* finale, viene calcolata il tasso di crescita dei diplomatici, di genere maschile e femminile, tra il 1968 e il 2021.

- Nello specifico, nella *SELECT*:  
`'NULLIF(maleDiplomats_1968, 0)'`  
1) Se *maleDiplomats\_1968* è zero, viene restituito *NULL* per evitare l'errore di divisione per zero.

```
'ROUND( ... , maleDiplomats_2021) /  
... ( ... , maleDiplomats_1968), ... ),  
2)'
```

3) Il risultato ottenuto dalla divisione viene arrotondato a due cifre decimali. Tuttavia, questo risultato non mostra solo due decimali, ma aggiunge zeri aggiuntivi alla fine.

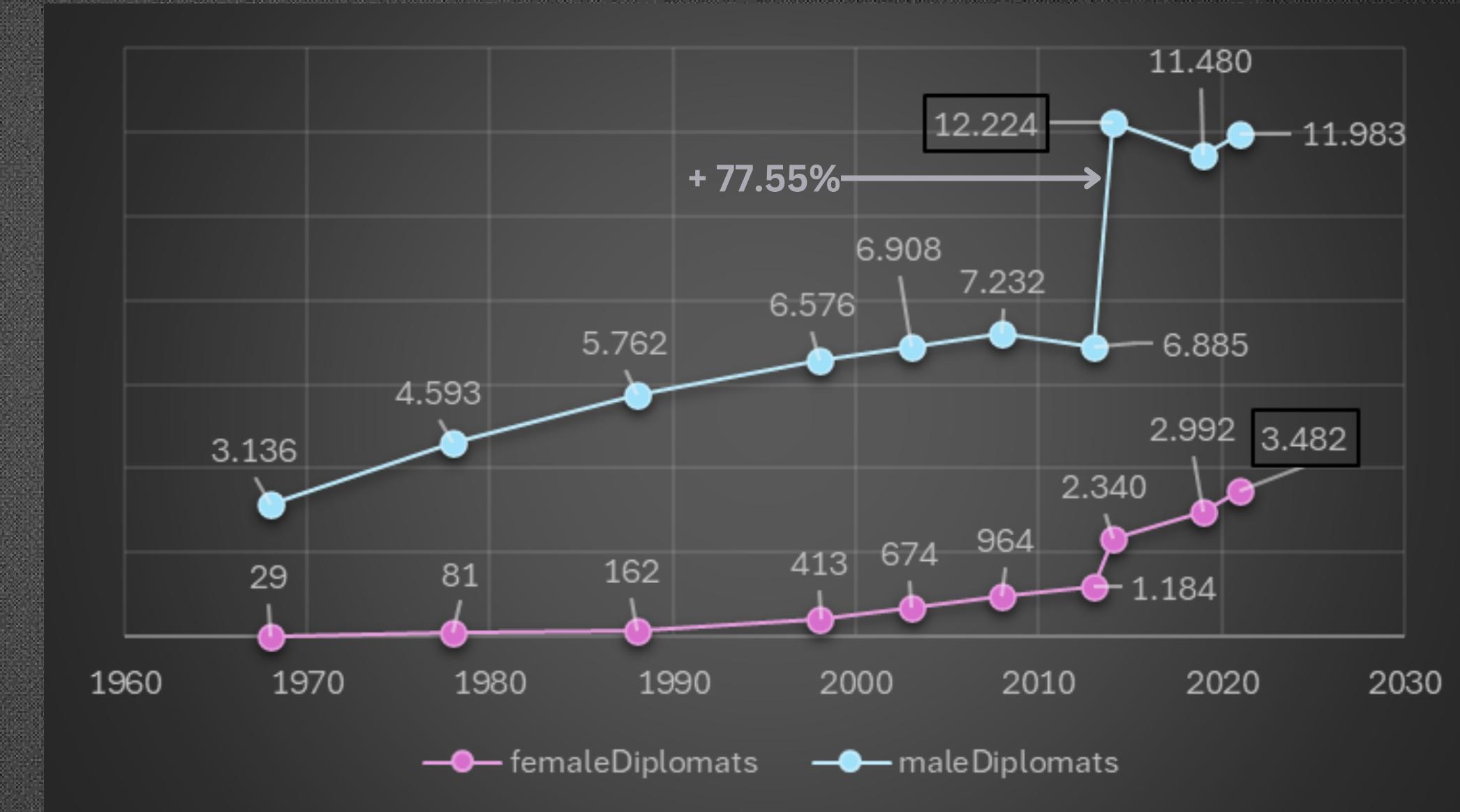
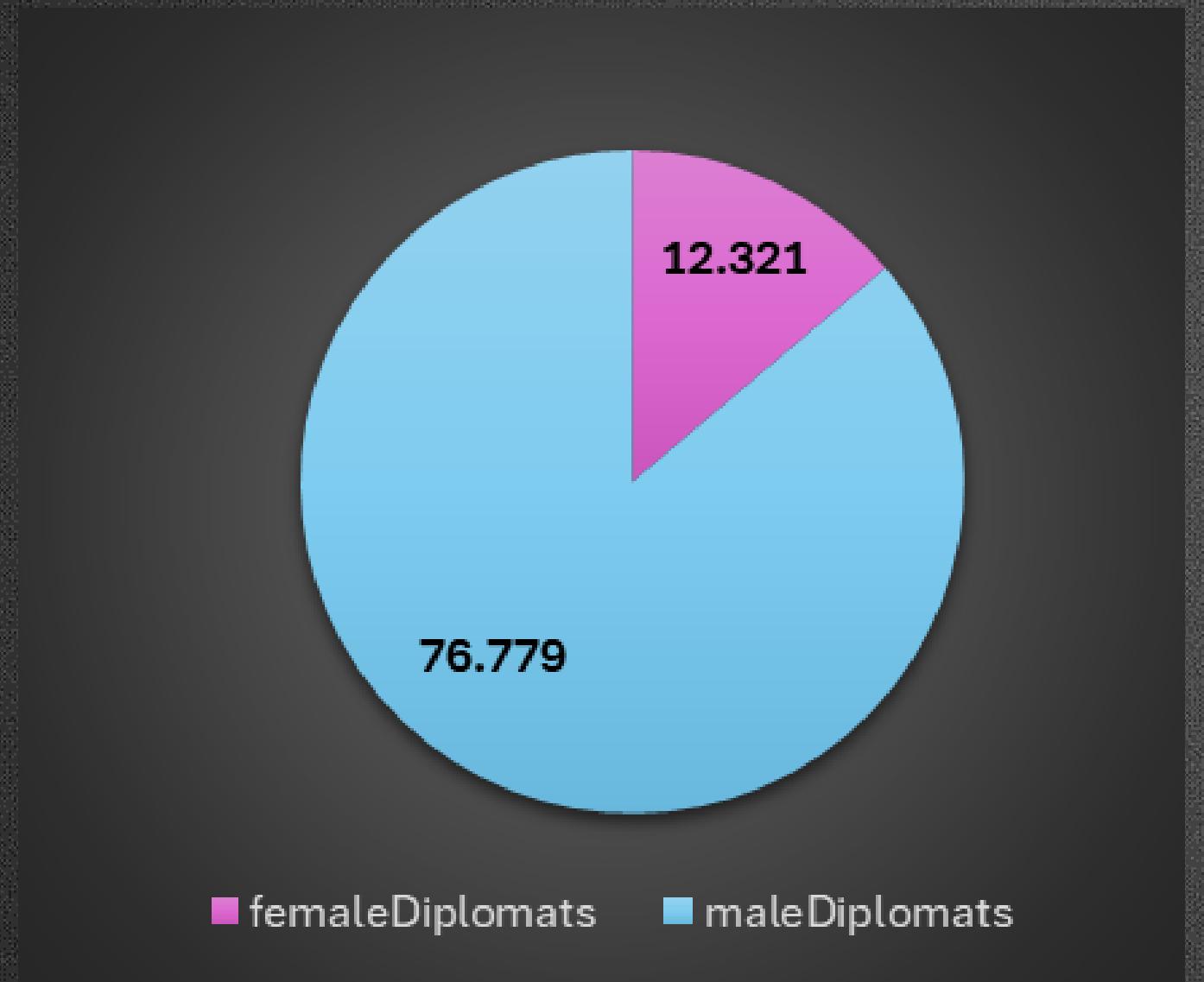
```
'FORMAT(ROUND( ... , maleDiplomats_2021) /  
... ( ... , maleDiplomats_1968), ... ), ... ),  
'0.00')'
```

```
'(CONVERT(DECIMAL(18,2), maleDiplomats_2021) /  
NULLIF(CONVERT(DECIMAL(18,2), maleDiplomats_1968),  
0))'
```

2) *maleDiplomats\_2021* e *maleDiplomats\_1968* vengono convertiti in formato decimale prima di eseguire la divisione.

4) Si formatta il risultato arrotondato come stringa con due cifre decimali.

# 1. Conteggio totale e annuale di diplomatici per genere - GRAFICI.



## Analisi dei dati

- 1) Conteggio totale e annuale di diplomatici per genere.
  - a) Maggior numero di diplomatici per genere e in quale anno.
  - b) Rapporto di crescita di diplomatici, per genere, del 2021 rispetto al 1968.
  - c) Crescita percentuale maggiore di diplomatici di genere maschile.
  
- 1)** Il conteggio totale dei diplomatici di genere femminile è di 12.321, mentre quello maschile è di 76.779.  
  
**c)** Nel 2014 c'è stato un'aumento di 5.339 diplomatici di genere maschile rispetto all'anno precedente, ovvero del 77.55%, numericamente è la crescita annuale più significativa registrata.
  
- a)** Nel 2014 si è registrato il maggior numero di diplomatici di genere maschile, per un totale di 12.224, mentre la maggior presenza di genere femminile si è registrata nel 2021 con 3.482.
  
- b)** Nel 2021 la presenza di diplomatici femminili è aumentata di circa 120 volte rispetto al 1968, ed è aumentata di quasi 4 volte la presenza di diplomatici maschili.

## 2) Paesi con più diplomatici di genere femminile per anno e in assoluto.

### 2 pt1) Paesi con più diplomatici femminili per anno.

```
WITH femDip_count_CTE
AS( SELECT
      VW.unifiedCountries,
      COUNT( CASE
                  WHEN gender = 1 THEN 1
              END )
      DENSE_RANK()
      OVER ( PARTITION BY Y.year
              ORDER BY COUNT( CASE
                  WHEN gender = 1 THEN 1
              END )
              DESC )
      AS ROW_order
  FROM targetArea AS T
  LEFT JOIN dbo.years AS Y
    ON T.yearID = Y.yearID
  LEFT JOIN sendingCountries AS S
    ON T.sendingCountryID = S.sendingCountryID
  LEFT JOIN vw_unifiedCountries AS VW
    ON S.countryID = VW.countryID
  LEFT JOIN dbo.diplomats AS D
    ON T.diplomatID = D.diplomatID
  GROUP BY Y.year,
           VW.unifiedCountries
  HAVING COUNT(CASE
                  WHEN gender = 1 THEN 1
              END)
        > 1)
SELECT year,
       unifiedCountries,
       femDip_count
  FROM femDip_count_CTE
 WHERE ROW_order = 1
 ORDER BY year;
```

- Nella CTE *femDip\_count\_CTE*, si calcola il numero di diplomatici femminili per ogni anno e paese, utilizzando l'alias *femDip\_count* per questa misura.
  - Impiegando la funzione finestra *DENSE\_RANK()*, viene assegnata un'etichetta di ordinamento basata su questo conteggio, classificando le combinazioni anno-paese in base al numero più alto di diplomatici femminili.
- Nella *SELECT* si seleziona la prima combinazione anno-paese in base alla classificazione determinata dalla funzione finestra *DENSE\_RANK()*.
  - Tramite la clausola '*WHERE ROW\_order = 1*' viene assicurato il risultato corrispondente al conteggio più alto di diplomatici femminili.

## 2 pt2) Paesi con più diplomatici femminili in assoluto.

```
SELECT TOP 10 WITH TIES VW.unifiedCountries,
       COUNT(D.gender) AS femDip_count
  FROM targetArea AS T
  LEFT JOIN dbo.years AS Y
    ON T.yearID = Y.yearID
  LEFT JOIN sendingCountries AS S
    ON T.sendingCountryID = S.sendingCountryID
  LEFT JOIN vw_unifiedCountries AS VW
    ON S.countryID = VW.countryID
  LEFT JOIN dbo.diplomats AS D
    ON T.diplomatID = D.diplomatID
 WHERE D.gender = 1
 GROUP BY VW.unifiedCountries
 ORDER BY femDip_count DESC;
      GROUP BY R.region)
SELECT region,
       FORMAT(unformatted_count, '#,0') AS femDip_re
  FROM femDip_regionCount_CTE
 ORDER BY unformatted_count DESC;
```

- Utilizzo la clausola *TOP 10 WITH TIES* in modo che, se dovessero esserci più paesi con lo stesso numero massimo di diplomatici femminili, verranno restituiti tutti questi paesi oltre ai primi 10 (con ordine decrescente).

## 2a) Regioni geografiche con più diplomatici femminili in assoluto.

```
WITH femDip_regionCount_CTE
      AS( SELECT R.region,
                 COUNT(D.gender) AS unformatted_count
            FROM dbo.targetArea AS TA
            LEFT JOIN dbo.sendingCountries AS S
              ON TA.sendingCountryID = S.sendingCountryID
            LEFT JOIN vw_unifiedCountries AS VW
              ON S.countryID = VW.countryID
            LEFT JOIN dbo.regions AS R
              ON VW.regionID = R.regionID
            LEFT JOIN dbo.diplomats AS D
              ON TA.diplomatID = D.diplomatID
            LEFT JOIN dbo.years AS Y
              ON TA.yearID = Y.yearID
            WHERE D.gender = 1
            GROUP BY R.region)
SELECT region,
       FORMAT(unformatted_count, '#,0') AS femDip_regionCount
  FROM femDip_regionCount_CTE
 ORDER BY unformatted_count DESC;
```

2 pt1

<i>year</i>	<i>unifiedCountries</i>	<i>femDip_count</i>
1968	Brazil	3
1978	United States of America	8
1988	Cuba	9
1988	United States of America	9
1998	United States of America	31
2003	United States of America	28
2008	United States of America	46
2013	United States of America	39
2014	Jamaica	73
2019	Canada	90
2021	Madagascar	120

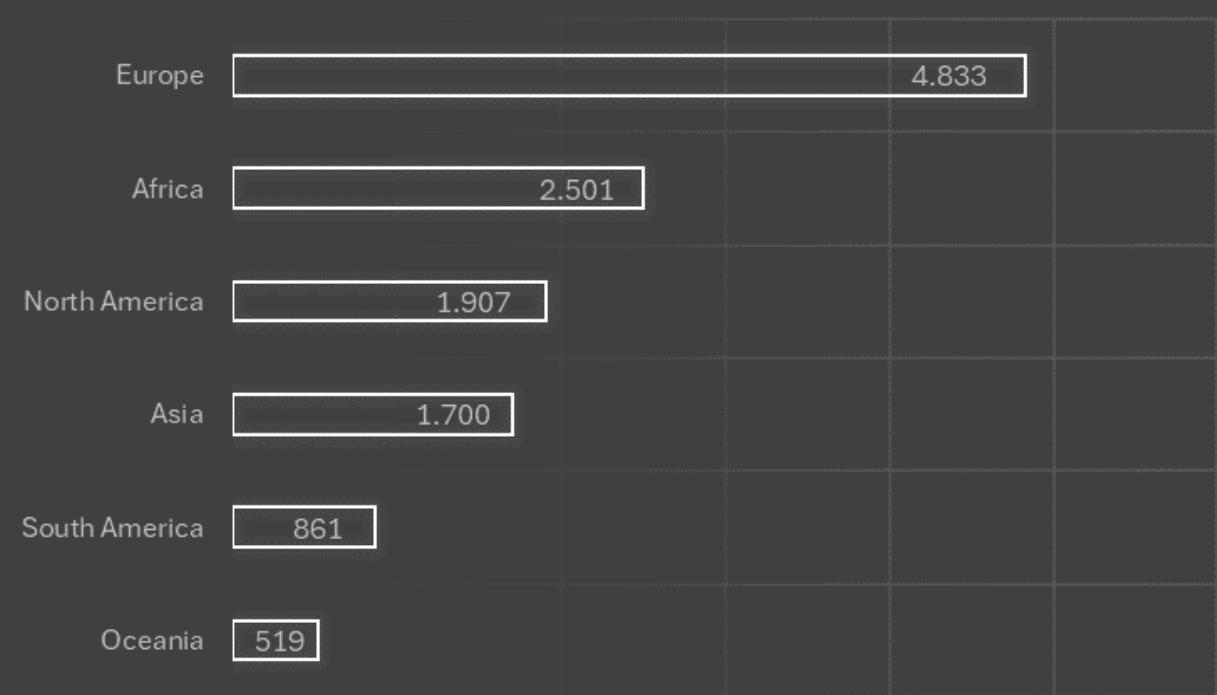
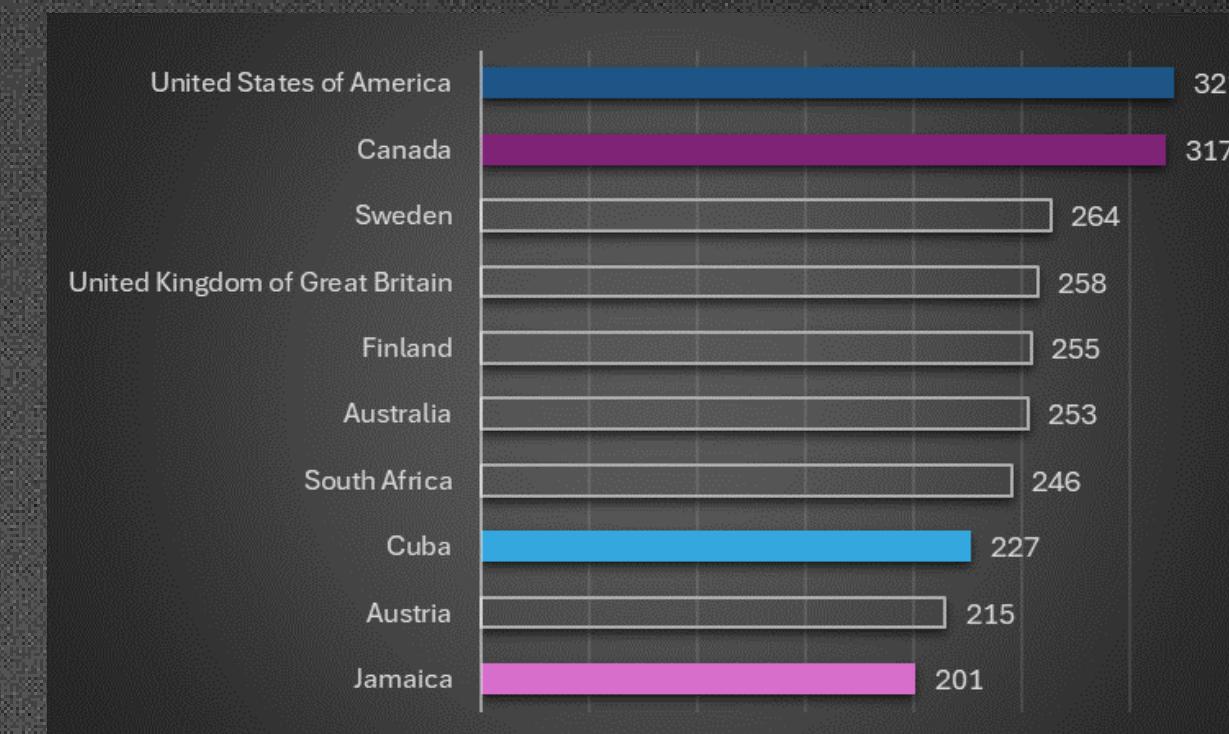
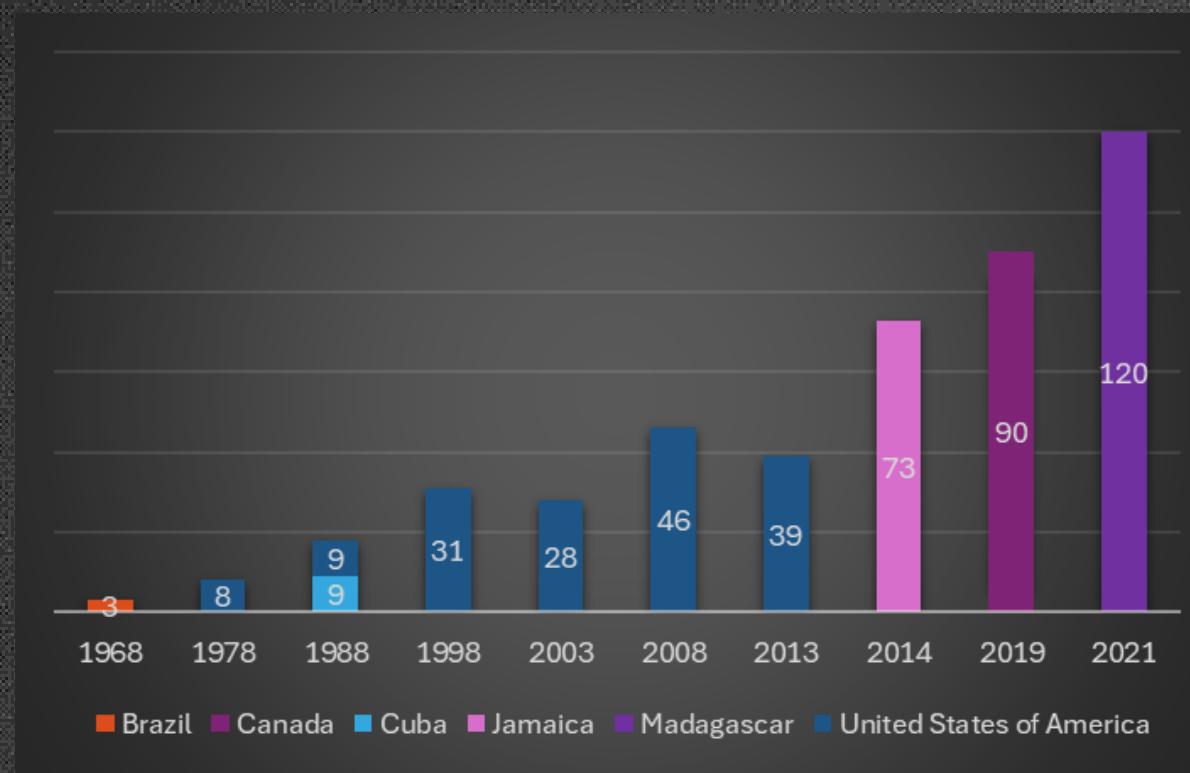
2 pt2

<i>unifiedCountries</i>	<i>femDip_count</i>
United States of America	321
Canada	317
Sweden	264
United Kingdom of Great Britain and Northern Ireland	258
Finland	255
Australia	253
South Africa	246
Cuba	227
Austria	215
Jamaica	201

2a

<i>region</i>	<i>femDip_regionCount</i>
Europe	4.833
Africa	2.501
North America	1.907
Asia	1.700
South America	861
Oceania	519

## 2) Paesi con più diplomatici di genere femminile per anno e in assoluto - GRAFICI.



## Analisi dei dati

- 2) Paesi con più diplomatici di genere femminile per anno e in assoluto.
- a) Regioni geografiche con più diplomatici di genere femminile in assoluto.
- Il Madagascar è il paese con più diplomatici di genere femminile per quanto riguarda il singolo anno, ovvero il 2021 con un totale di 120.
  - Gli U.S.A hanno avuto il maggior numero di diplomatici di genere femminile per 6 anni consecutivi e sono, inoltre, il paese con il maggior numero in assoluto, per un totale di 321.
  - Il 1988 è l'unico anno in cui sono presenti più paesi con il maggior numero di diplomatici di genere femminile, trattasi di Cuba e degli U.S.A. con 9 diplomatici ciascuno.
  - 4 paesi su 10 hanno avuto sia la maggior presenza di diplomatici di genere femminile per almeno un anno, sia il maggior numero in assoluto: si tratta degli U.S.A con 321 diplomatici in totale, Canada con 317, Cuba con 227 e Jamacia con 201.
  - L'Europa è prima per numero di diplomatici di genere femminile con un totale di 4.833, seguita dall'Africa, con un totale di 2.501.

### 3) Conteggio dei titoli per categoria.

```
WITH titles_count_CTE
AS( SELECT
    CASE
        WHEN T.title NOT IN ('Ambassador', 'Chargé d'affaires') THEN 'other titles'
        WHEN T.title IN ('Ambassador', 'Chargé d'affaires') THEN T.title
    END AS title,
    COUNT(TA.diplomatID) AS unformatted_count
FROM dbo.targetArea AS TA
LEFT JOIN diplomats AS D
    ON TA.diplomatID = D.diplomatID
LEFT JOIN dbo.titles AS T
    ON D.titleID = T.titleID
WHERE T.title IS NOT NULL
GROUP BY
    CASE
        WHEN T.title NOT IN ('Ambassador', 'Chargé d'affaires') THEN 'other titles'
        WHEN T.title IN ('Ambassador', 'Chargé d'affaires') THEN T.title
    END)
SELECT title,
    FORMAT(unformatted_count, '#,0') AS titles_count
FROM titles_count_CTE
ORDER BY unformatted_count ASC;
```

- Essendo i titoli di Ambassador e Chargé d'affaires numericamente i più presenti, ho incluso gli altri nell'etichetta 'other titles'.

title	titles_count
other titles	539
Chargé d'affaires	3.750
Ambassador	89.258

### 3a) Conteggio di ogni titolo per genere.

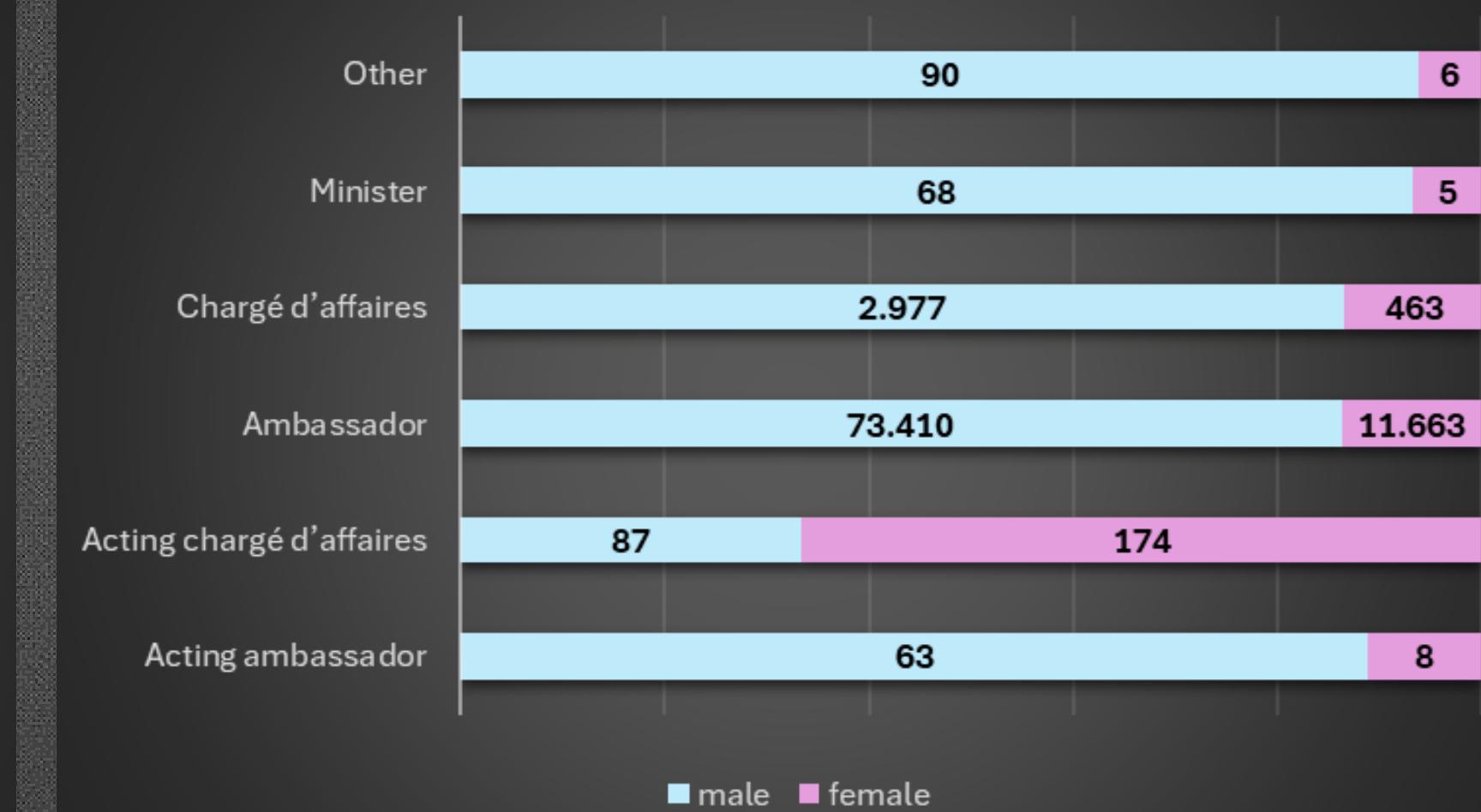
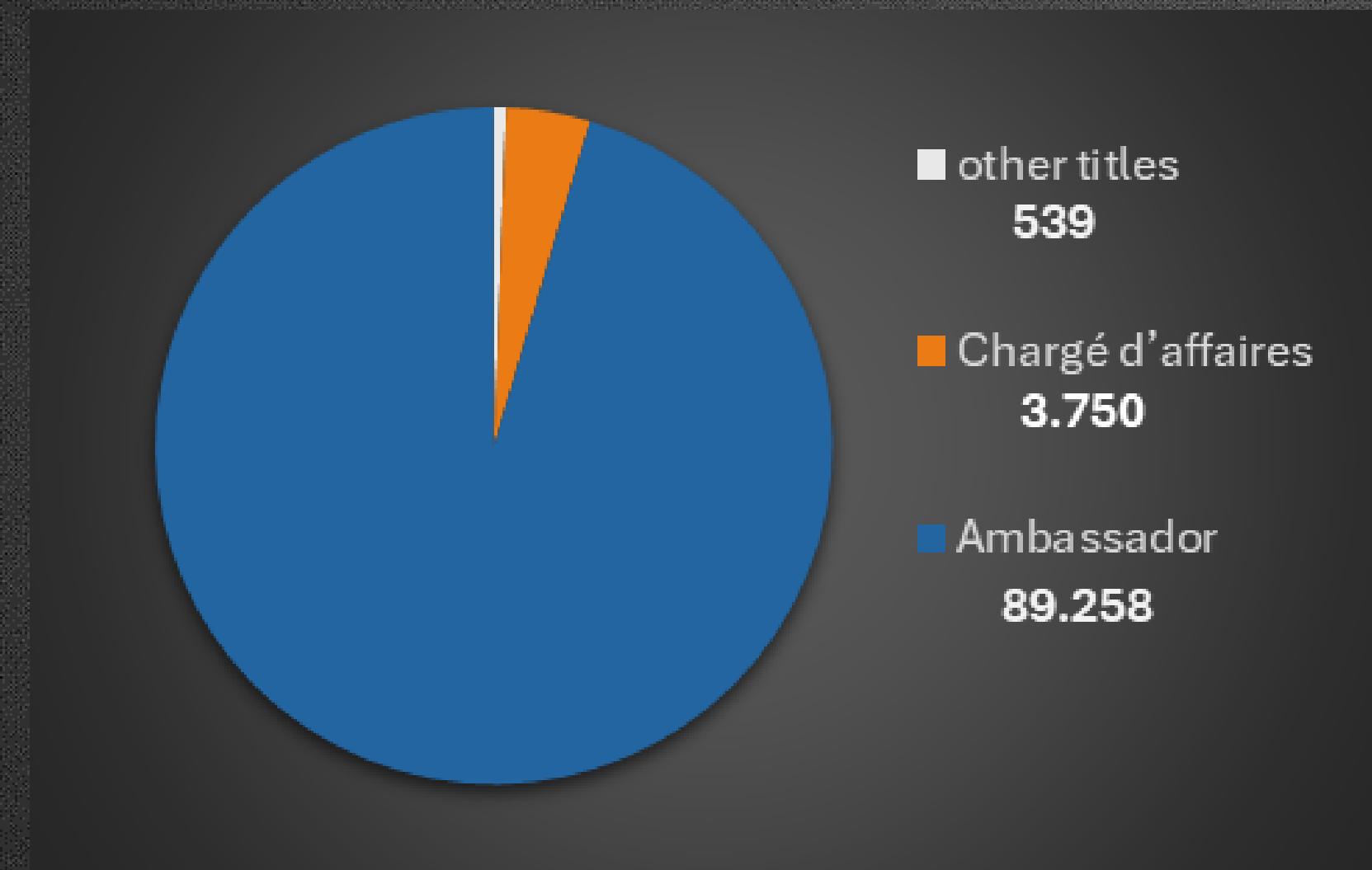
```

SELECT T.title,
CASE
    WHEN D.gender = 0 THEN 'male'
    WHEN D.gender = 1 THEN 'female'
END AS gender,
FORMAT(COUNT(D.gender), '#,0') AS titles_count
FROM dbo.targetArea AS TA
LEFT JOIN diplomats AS D
    ON TA.diplomatID = D.diplomatID
LEFT JOIN dbo.titles AS T
    ON D.titleID = T.titleID
WHERE D.gender IS NOT NULL
    AND T.title IS NOT NULL
GROUP BY T.title,
CASE
    WHEN D.gender = 0 THEN 'male'
    WHEN D.gender = 1 THEN 'female'
END
ORDER BY T.title ASC,
gender ASC;

```

<i>title</i>	<i>gender</i>	<i>titles_count</i>
Acting ambassador	female	8
Acting ambassador	male	63
Acting chargé d'affaires	female	174
Acting chargé d'affaires	male	87
Ambassador	female	11.663
Ambassador	male	73.410
Chargé d'affaires	female	463
Chargé d'affaires	male	2.977
Minister	female	5
Minister	male	68
Other	female	6
Other	male	90

### 3) Conteggio dei titoli per categoria e genere - GRAFICI.



## Analisi dei dati

- 3) Conteggio dei titoli per categoria .
  - a) Conteggio di ogni titolo per genere.

- Il titolo di Ambassador è quello più presente rispetto agli altri, Chargé d'affaires è il secondo più impiegato, ma rimane comunque molto distante a livello quantitativo rispetto al primo. Gli altri titoli sono numericamente poco rilevanti.
  - Il titolo di Ambassador è quello più impiegato dai diplomatici di entrambi i generi, con un totale di 73.410 per i diplomatici di genere maschile e 11.663 per quelli di genere femminile.
  - Acting chargé d'affaires è l'unico titolo maggiormente utilizzato dai diplomatici di genere femminile, con un totale di 174, contro il totale di 87 dei diplomatici di genere maschile.

# Analisi dei dati

## **SECONDO FOCUS: membri del Parlamento di genere femminile e politica estera femminista.**

.

- 4) Media percentuale annuale dei legislatori femminili della Camera Bassa.
  - a) Crescita media annuale dei legislatori femminili della Camera Bassa.
- 5) I paesi con la percentuale annuale maggiore di legislatori femminili.
  - a) TOP 10 paesi con media percentuale più alta in assoluto.
  - b) La media percentuale delle regioni geografiche

## 4) Media percentuale annuale dei legislatori femminili della Camera Bassa.

```
SELECT year,
       FORMAT(ROUND( AVG( COALESCE(FL.femaleLegislatorPercentage,
0)) , '0.00' , 2)
       , '0.00') AS femPct_avg
  FROM
    dbo.targetArea AS TA
  LEFT JOIN dbo.sendingCountries AS SC
    ON TA.sendingCountryID = SC.sendingCountryID
  LEFT JOIN dbo.femaleLegislators AS FL
    ON SC.femaleLegislatorID = FL.femaleLegislatorID
  LEFT JOIN dbo.years AS Y
    ON TA.yearID = Y.yearID
 GROUP BY year
 ORDER BY year ASC;
```

year	femPct_avg
1968	4,96
1978	7,74
1988	10,68
1998	11,88
2003	14,54
2008	17,78
2013	20,49
2014	21,83
2019	25,65
2021	26,21

## 4a) Crescita media annuale dei legislatori femminili della Camera Bassa.

```

WITH femPct_avg_CTE
AS( SELECT year,
          ROUND( AVG( COALESCE(FL.femaleLegislatorPercentage, 0))
                 , 2) AS femPct_avg
    FROM
      dbo.targetArea AS TA
      LEFT JOIN dbo.sendingCountries AS SC
        ON TA.sendingCountryID = SC.sendingCountryID
      LEFT JOIN dbo.femaleLegislators AS FL
        ON SC.femaleLegislatorID = FL.femaleLegislatorID
      LEFT JOIN dbo.years AS Y
        ON TA.yearID = Y.yearID
  GROUP BY year),
avgPct_CTE
AS (SELECT year,
           femPct_avg,
           LAG(femPct_avg)
             OVER (
               ORDER BY year) AS avgPct_prevYear,
           femPct_avg - LAG(femPct_avg)
             OVER (
               ORDER BY year) AS annual_avgPct_growth
    FROM femPct_avg_CTE)
SELECT FORMAT(ROUND( SUM(COALESCE(annual_avgPct_growth, 0)) / (COUNT(year)-1)
                  , 2)
                  , '0.00') AS avgPct_growth
FROM avgPct_CTE;

```

<i>avgPct_growth</i>
2,36

- Output in dettaglio di *avgPct\_CTE*.

<i>year</i>	<i>femPct_avg</i>	<i>avgPct_prevYear</i>	<i>annual_avgPct_growth</i>
1968	4.96	NULL	NULL
1978	7.74	4.96	2.78
1988	10.68	7.74	2.94
1998	11.88	10.68	1.20
2003	14.54	11.88	2.66
2008	17.78	14.54	3.24
2013	20.49	17.78	2.71
2014	21.83	20.49	1.34
2019	25.65	21.83	3.82
2021	26.21	25.65	0.56

#### **4a) Crescita media annuale - DETTAGLIO CODICE.**

- Nella CTE *femPct\_avg\_CTE* visualizzo la media percentuale annuale dei legislatori femminili, appartenenti alla Camera Bassa.
- Nella CTE *avgPct\_CTE* viene espressa la crescita annua media percentuale della presenza di legislatori femminili.
- Nello specifico:
  - con la funzione finestra *LAG* mostro la media percentuale dell'anno precedente per ogni riga.
  - con il codice '*femPct\_avg - LAG(femPct\_avg)* OVER (ORDER BY year) AS *annual\_avgPct\_growth*' calcolo la differenza tra la media percentuale dell'anno corrente e quella dell'anno precedente.
- Nella *SELECT* finale viene calcolata la crescita media annuale delle percentuali di legislatori femminili.
  - Prima si somma le variazioni annuali delle percentuali, poi le si divide per il numero di anni meno uno, ottenendo una media. Infine, si formatta il risultato come percentuale con due cifre decimali.

## 5) I paesi con la percentuale annuale maggiore di legislatori femminili.

```
WITH femPct_CTE
AS( SELECT year,
          VW.unifiedCountries,
          COALESCE(FL.femaleLegislatorPercentage, 0) AS femPct,
          DENSE_RANK()
            OVER (PARTITION BY Y.year
                  ORDER BY COALESCE(FL.femaleLegislatorPercentage, 0)
                  DESC) AS femPct_denseRank
  FROM
    dbo.targetArea AS TA
    LEFT JOIN dbo.sendingCountries AS SC
      ON TA.sendingCountryID = SC.sendingCountryID
    LEFT JOIN vw_unifiedCountries AS VW
      ON SC.countryID = VW.countryID
    LEFT JOIN dbo.femaleLegislators AS FL
      ON SC.femaleLegislatorID = FL.femaleLegislatorID
    LEFT JOIN dbo.years AS Y
      ON TA.yearID = Y.yearID
  GROUP BY
    Y.year,
    VW.unifiedCountries,
    FL.femaleLegislatorPercentage)
SELECT year, unifiedCountries, femPct
FROM femPct_CTE
WHERE femPct_denseRank = 1;
```

year	unifiedCountries	femPct
1968	Germany	30.60
1978	Germany	33.60
1988	Sweden	38.40
1998	Sweden	42.70
2003	Rwanda	48.75
2008	Rwanda	56.25
2013	Rwanda	63.75
2014	Rwanda	63.75
2019	Rwanda	61.25
2021	Rwanda	61.25

## 5) I paesi con la percentuale annuale maggiore di legislatori femminili - DETTAGLIO CODICE:

- Parte dell'output di *femPct\_CTE*.

<i>year</i>	<i>unifiedCountries</i>	<i>femPct</i>	<i>femPct_denseRank</i>
1968	Germany	30.60	1
1968	Russian Federation	28.90	2
1968	Mongolia	21.60	3
1968	Guinea	21.33	4
1968	Czechoslovakia	20.00	5
1968	Hungary	19.80	6
1968	China	17.80	7
1968	Bulgaria	17.10	8
1968	Finland	16.50	9
1968	Albania	16.30	10

- In questo blocco di codice viene calcolata la percentuale di legislatori femminili per ogni anno e paese.
- Nello specifico:
  - Nella CTE *femPct\_CTE* viene attribuito un numero di rango basato sulle percentuali per ogni anno e ogni paese, tramite la funzione finestra *DENSE RANK*.
- Nella *SELECT* finale vengono selezionate solo le righe con il rango uguale a 1, ovvero il paese con le percentuali annuali più alte.

... output continua...

## 5a) TOP 10 paesi con media percentuale più alta in assoluto.

```
SELECT
    TOP 10 VW.unifiedCountries,
    CAST(ROUND( AVG( COALESCE(FL.femaleLegislatorPercentage, 0))
        , 2) AS DECIMAL(18,2)) AS femPct
FROM
    dbo.targetArea AS TA
    LEFT JOIN dbo.sendingCountries AS SC
        ON TA.sendingCountryID = SC.sendingCountryID
    LEFT JOIN vw_unifiedCountries AS VW
        ON SC.countryID = VW.countryID
    LEFT JOIN dbo.femaleLegislators AS FL
        ON SC.femaleLegislatorID = FL.femaleLegislatorID
    LEFT JOIN dbo.years AS Y
        ON TA.yearID = Y.yearID
GROUP BY VW.unifiedCountries
ORDER BY femPct DESC;
```

```
CAST(ROUND( AVG( COALESCE(FL.femaleLegislatorPercentage, 0))
        , 2) AS DECIMAL(18,2)) AS femPct
```

converte la media arrotondata di *femaleLegislatorPercentage* in un numero decimale con due cifre decimali.

<i>unifiedCountries</i>	<i>femPct</i>
Rwanda	50.78
Cuba	43.11
Sweden	41.30
Finland	39.20
Iceland	39.13
Namibia	38.20
Norway	37.35
Mozambique	36.99
South Africa	36.86
Timor-Leste	36.62

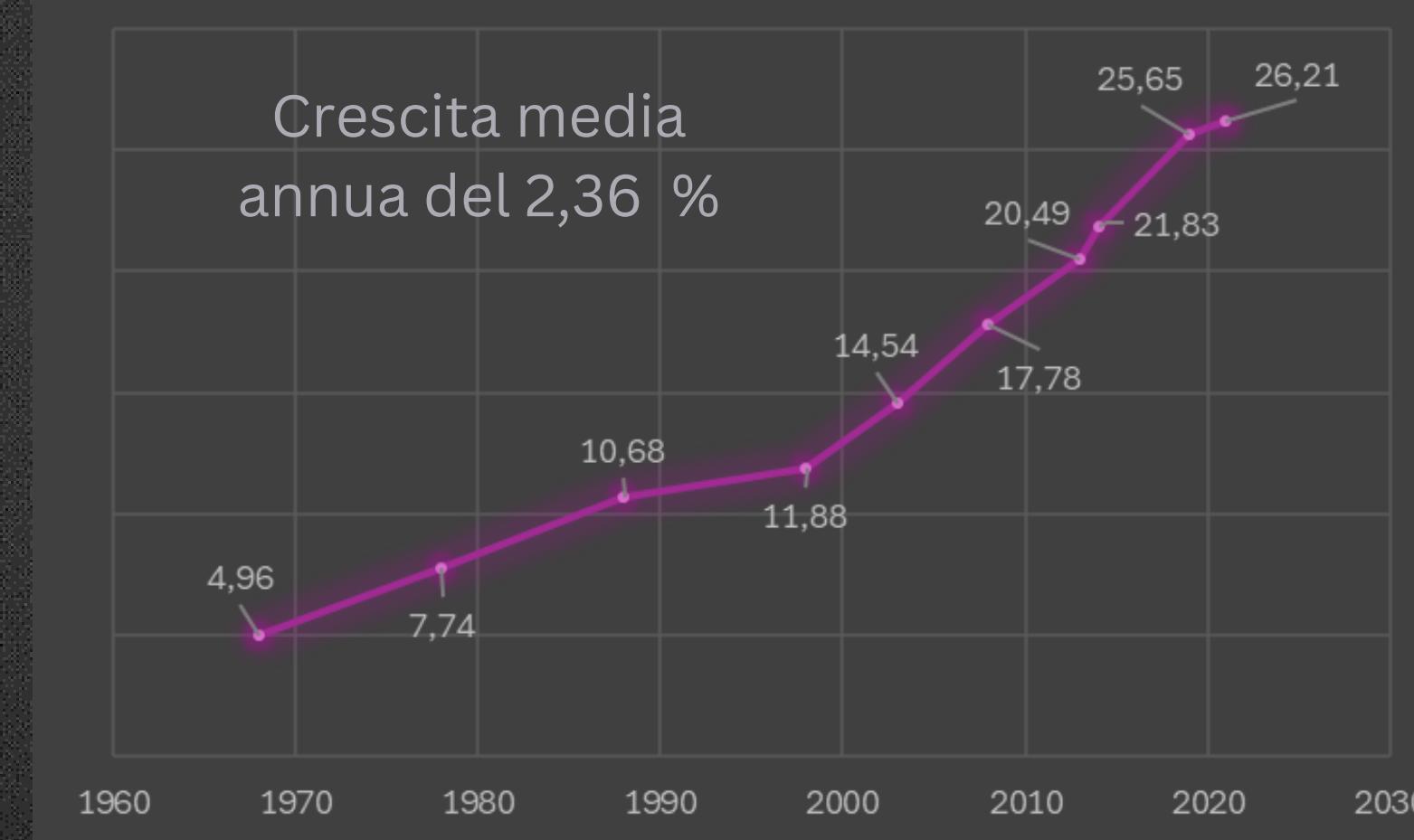
## 5b) La media percentuale delle regioni geografiche.

```
SELECT
    R.region,
    CAST(ROUND( AVG( COALESCE(FL.femaleLegislatorPercentage, 0))
        , 2) AS DECIMAL(18,2)) AS femPct
FROM
    dbo.targetArea AS TA
    LEFT JOIN dbo.sendingCountries AS SC
        ON TA.sendingCountryID = SC.sendingCountryID
    LEFT JOIN vw_unifiedCountries AS VW
        ON SC.countryID = VW.countryID
    LEFT JOIN dbo.regions AS R
        ON VW.regionID = R.regionID
    LEFT JOIN dbo.femaleLegislators AS FL
        ON SC.femaleLegislatorID = FL.femaleLegislatorID
    LEFT JOIN dbo.years AS Y
        ON TA.yearID = Y.yearID
GROUP BY R.region
ORDER BY femPct ASC;
```

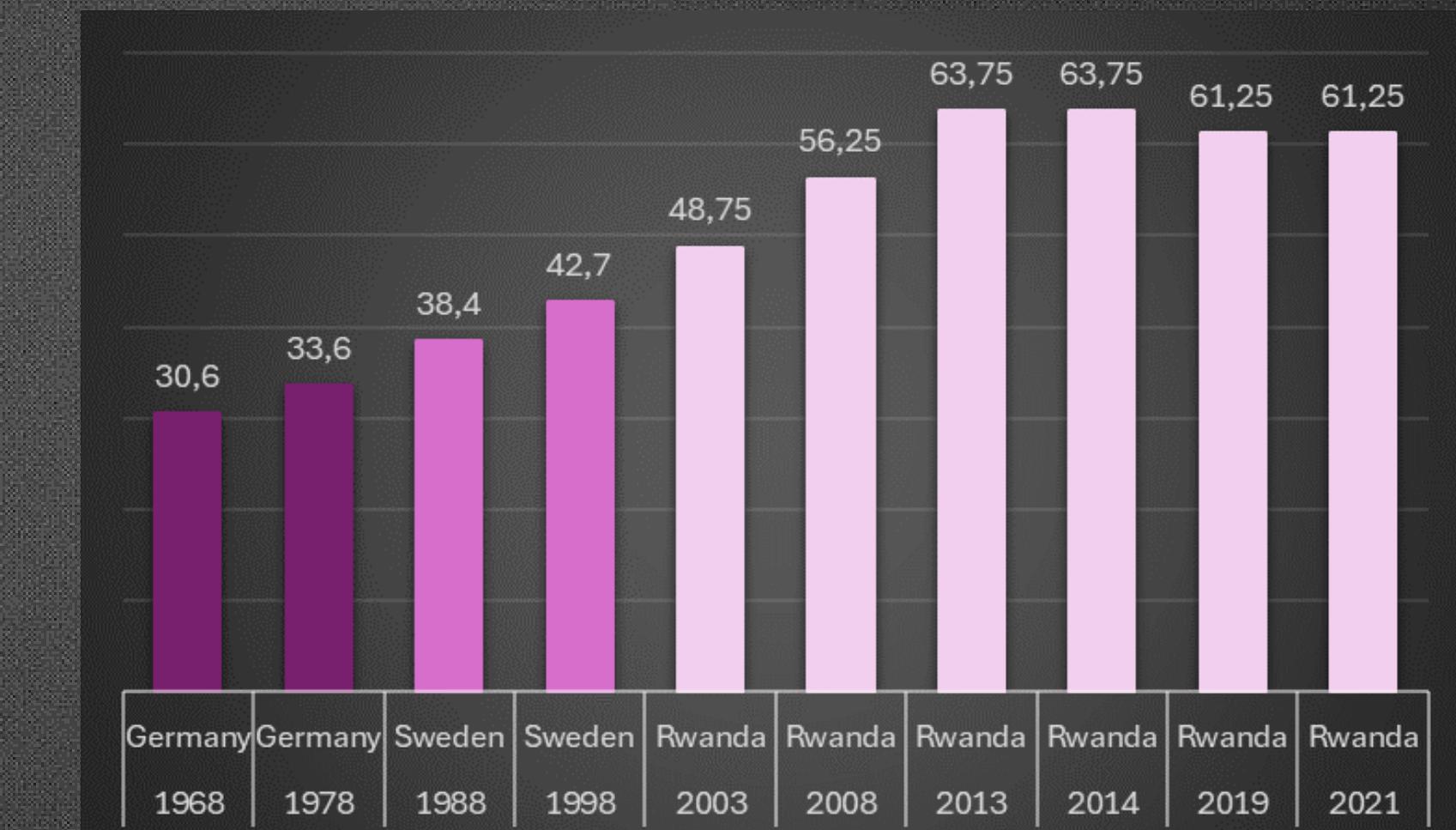
region	femPct
Asia	13.21
Africa	16.73
South America	17.70
Oceania	21.52
North America	22.19
Europe	23.79

## SECONDO FOCUS pt1: membri del Parlamento di genere femminile - GRAFICI.

Media percentuale annuale

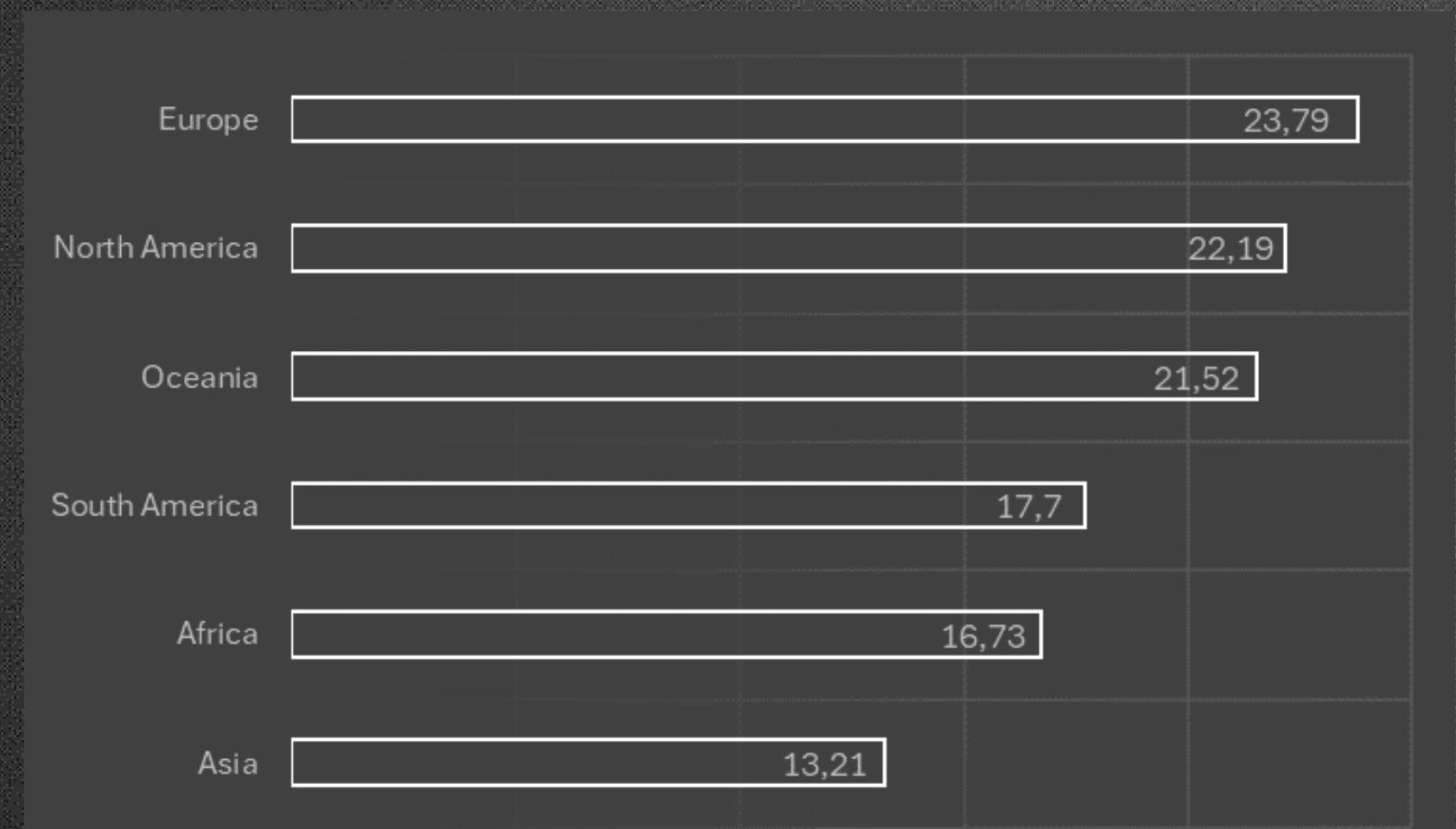


Paesi con presenza media % annuale più alta

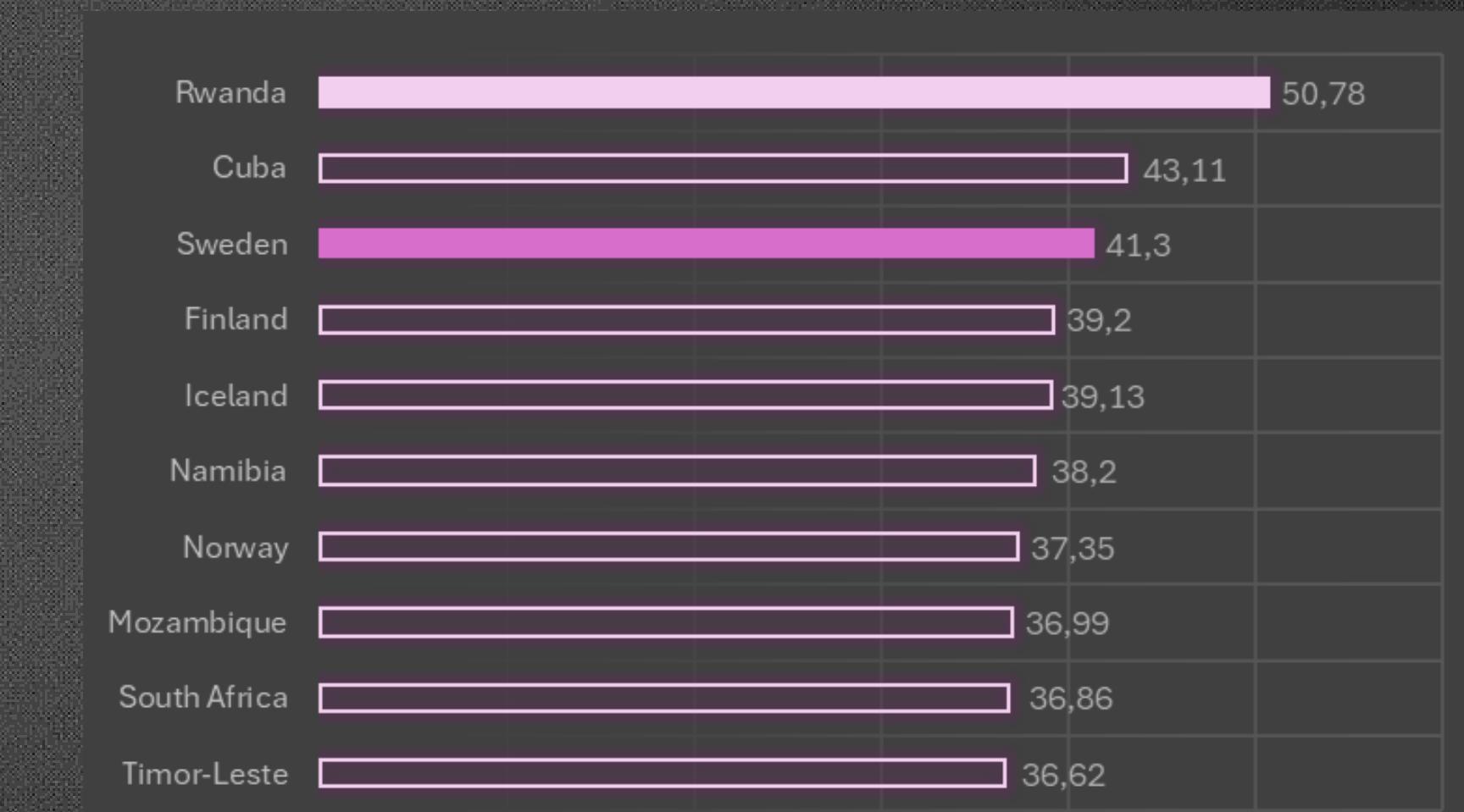


## SECONDO FOCUS pt1: membri del Parlamento di genere femminile - GRAFICI.

Media percentuale delle regioni geografiche



Top 10 paesi con media % maggiore in assoluto



## Analisi dei dati

- 4) Media percentuale annuale dei legislatori femminili della Camera Bassa.
    - a) Crescita media annuale dei legislatori femminili della Camera Bassa.
  - 5) Paesi con la percentuale annuale maggiore di legislatori femminili.
    - a) TOP 10 paesi con media percentuale più alta in assoluto.
    - b) La media percentuale delle regioni geografiche
- 
- Nel 1968 la media percentuale dei membri di genere femminile, appartenenti alla Camera Bassa, era del 4,96%, arrivando al 26,21% nel 2021, con una crescita media annua del 2,36%.
  - Nel 1968 la Germania è stata il primo paese per presenza di membri appartenenti al genere femminile, con una percentuale del 30,6, mantenendo il podio successivamente nel 1978, con una percentuale del 33,6.
  - La Svezia è salita sul podio subito dopo la Germania, per presenza annuale maggiore, con una percentuale di 38,4 nel 1988 e di 42,7 nel 1998, essa risulta, inoltre, essere uno dei dieci paesi con la media percentuale maggiore in assoluto, classificandosi al terzo posto con il valore di 41,3%.

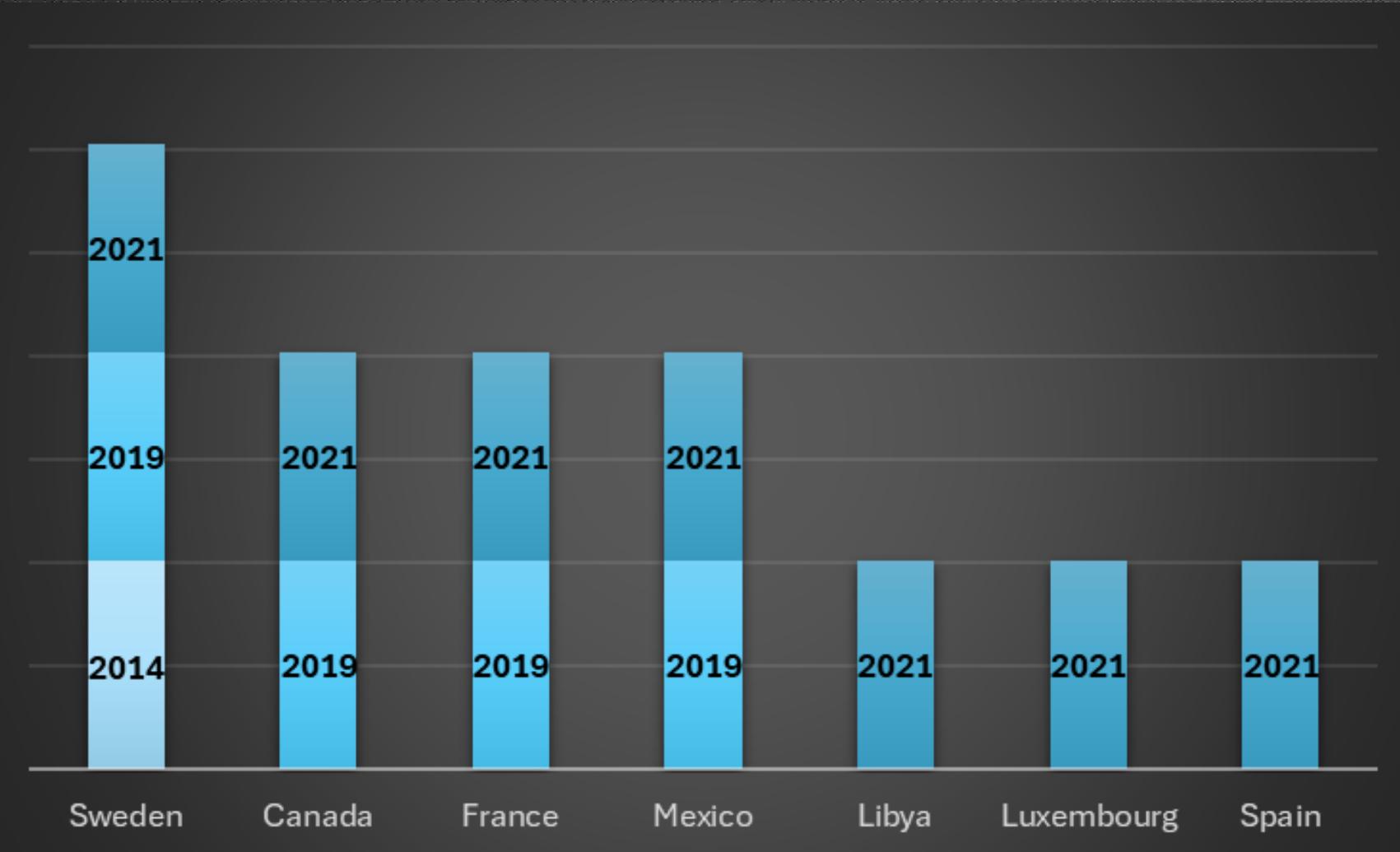
- La Rwanda è il paese con la maggior percentuale di presenza annuale a partire dal 2003, dove è rimasta fissa fino al 2021.
  - Nel 2008 più della metà dei membri della Camera Bassa apparteneva al genere femminile, con una presenza totale di 56,25%.
  - La Rwanda è il primo paese per presenza di membri del Parlamento appartenenti al genere femminile, con una media percentuale del 50,78%.
- Per quanto riguarda le regioni geografiche, al primo posto si trova l'Europa con una media percentuale del 23,79%, seguita dal Nord America, con un valore di 22,19%.

## 6) Paesi d'invio aderenti ad una politica estera femminista.

```
SELECT Y.year,
       unifiedCountries
  FROM dbo.targetArea AS TA
  LEFT JOIN dbo.sendingCountries AS SC
    ON TA.sendingCountryID = SC.sendingCountryID
  LEFT JOIN vw_unifiedCountries AS VW
    ON SC.countryID = VW.countryID
  LEFT JOIN dbo.years AS Y
    ON TA.yearID = Y.yearID
 WHERE SC.feministForeignPolicy = 1
 GROUP BY Y.year,
          VW.unifiedCountries,
          SC.feministForeignPolicy;
 ORDER BY femPct ASC;
```

year	unifiedCountries
2014	Sweden
2019	Canada
2019	France
2019	Mexico
2019	Sweden
2021	Canada
2021	France
2021	Libya
2021	Luxembourg
2021	Mexico
2021	Spain
2021	Sweden

## SECONDO FOCUS pt2: politica estera femminista - GRAFICI.



- L'Europa è prima sia per numero di diplomatici di genere femminile, sia per media percentuale dei membri di genere femminile, appartenenti alla Camera Bassa.

### Analisi dei dati

- Paesi d'invio aderenti ad una politica estera femminista.
  - La Svezia è stato il primo paese ad aderire ad una politica estera femminista e nel 2014 è stato l'unico.
  - Nel 2019 si sono aggiunti il Canada, la Francia e il Messico, con l'aggiunta della Libia, del Lussemburgo e della Spagna nel 2021.
- PARALLELISMI:
  - Svezia, Finlandia, Sud Africa e Cuba risultano essere tra i primi dieci paesi sia per il numero di diplomatici di genere femminile, sia per la media percentuale dei membri di genere femminile, appartenenti alla Camera Bassa.

# Analisi dei dati

## TERZO FOCUS: Italia - analisi su paese specifico.

.

### 7) Conteggio totale e annuale di diplomatici per genere.

- a) Crescita percentuale maggiore di diplomatici di genere maschile.

### 8) Media percentuale annuale della presenza di diplomatici di genere femminile.

- a) Crescita media % annuale dei membri di genere femminile, appartenenti alla Camera Bassa.

### 9) Come si classifica l'Italia per quanto riguarda la media percentuale dei membri di genere femminile, appartenenti alla Camera Bassa.



## 7) Conteggio totale e annuale di diplomatici per genere.

```

SELECT COALESCE( CAST(Y.year AS VARCHAR(10)),
  'tot') AS year,
FORMAT( COUNT( CASE WHEN D.gender = 1 THEN 1
END)
, '#,0') AS femaleDiplomats,
FORMAT( COUNT( CASE WHEN D.gender = 0 THEN 1
END)
, '#,0') AS maleDiplomats
FROM dbo.diplomats AS D
INNER JOIN dbo.targetArea AS T
  ON D.diplomatID = T.diplomatID
INNER JOIN dbo.sendingCountries AS SC
  ON T.sendingCountryID = SC.sendingCountryID
LEFT JOIN vw_unifiedCountries AS VW
  ON SC.countryID = VW.countryID
INNER JOIN dbo.years AS Y
  ON T.yearID = Y.yearID
WHERE VW.alpcode = 'ITA'
GROUP BY Y.year
  WITH ROLLUP -- La clausola WITH ROLLUP consentirà di
ORDER BY Y.year ASC; -- ottenere anche il totale generale delle
colonne.
  
```

year	femaleDiplomats	maleDiplomats
tot	74	1.113
1968	0	71
1978	0	91
1988	1	98
1998	2	115
2003	6	112
2008	8	107
2013	9	105
2014	16	172
2019	15	103
2021	17	139



## 7a) Crescita percentuale maggiore di diplomatici di genere maschile.

```
WITH diplomatsCounts_CTE
    AS( SELECT Y.year,
              COUNT(D.gender) AS maleDiplomats
        FROM dbo.targetArea AS TA
        LEFT JOIN dbo.sendingCountries AS SC
            ON TA.sendingCountryID = SC.sendingCountryID
        LEFT JOIN vw_unifiedCountries AS VW
            ON SC.countryID = VW.countryID
        LEFT JOIN dbo.diplomats AS D
            ON TA.diplomatID = D.diplomatID
        LEFT JOIN dbo.years AS Y
            ON TA.yearID = Y.yearID
    WHERE VW.alpcode = 'ITA'
          AND D.gender = 0
    GROUP BY Y.year
),
maleDiplomatsDiff_CTE
    AS( SELECT year,
              maleDiplomats,
              LAG(maleDiplomats)
                  OVER (
                      ORDER BY Year)
                     AS male_prevYearCount,
              maleDiplomats - LAG(maleDiplomats)
                  OVER (
                      ORDER BY Year) AS male_diff
        FROM
            diplomatsCounts_CTE
),
malePct_CTE
    AS( SELECT year,
              maleDiplomats,
              male_prevYearCount,
              male_diff,
              CAST(ROUND( ((CONVERT(DECIMAL(18,2), male_diff) / male_prevYearCount) * 100)
                         , 2)
                   AS DECIMAL(18,2)) AS pct
        FROM
            maleDiplomatsDiff_CTE
    SELECT CONCAT( MAX(pct), '%' ) AS malePct_MAX
    FROM malePct_CTE;
```

The diagram illustrates the flow of data from the CTEs to the final result. It consists of two main vertical columns separated by a thick black vertical line. The left column contains the code for the `diplomatsCounts_CTE` and `maleDiplomatsDiff_CTE`. The right column contains the code for the `malePct_CTE`. A horizontal line with a plus sign at each end connects the bottom of the left column to the top of the right column. At the bottom of the right column, there is a rectangular box containing the output of the query, which is `malePct_MAX` followed by `63.81%`.



## 8) Media percentuale dei membri di genere femminile, appartenenti alla Camera Bassa.

```
SELECT Y.year,
       CAST(ROUND( AVG( COALESCE(FL.femaleLegislatorPercentage, 0) )
                  , 2)
             AS DECIMAL(18,2)) AS femPct_avg
  FROM dbo.targetArea AS TA
    LEFT JOIN dbo.sendingCountries AS SC
      ON TA.sendingCountryID = SC.sendingCountryID
    LEFT JOIN vw_unifiedCountries AS VW
      ON SC.countryID = VW.countryID
    LEFT JOIN dbo.femaleLegislators AS FL
      ON SC.femaleLegislatorID = FL.femaleLegislatorID
    LEFT JOIN dbo.years AS Y
      ON TA.yearID = Y.yearID
 WHERE VW.alpcode = 'ITA'
 GROUP BY Y.year
 ORDER BY Y.year ASC;
```

year	femPct_avg
1968	2.70
1978	7.50
1988	12.90
1998	11.10
2003	9.84
2008	21.27
2013	28.41
2014	30.95
2019	35.71
2021	35.71



## 8a) Crescita media % annuale della presenza di diplomatici di genere femminile.

```
WITH femPctAvg_CTE
    AS( SELECT Y.year,
                CAST(ROUND (AVG(COALESCE(FL.femaleLegislatorPercentage, 0))
                           , 2)
                     AS DECIMAL(18,2)) AS femPct_avg
        FROM dbo.targetArea AS TA
        LEFT JOIN dbo.sendingCountries AS SC
            ON TA.sendingCountryID = SC.sendingCountryID
        LEFT JOIN vw_unifiedCountries AS VW
            ON SC.countryID = VW.countryID
        LEFT JOIN dbo.femaleLegislators AS FL
            ON SC.femaleLegislatorID = FL.femaleLegislatorID
        LEFT JOIN dbo.years AS Y
            ON TA.yearID = Y.yearID
    WHERE VW.alpcode = 'ITA'
    GROUP BY Y.year),
femPctAvg_annual_growth_CTE
    AS(SELECT year,
          femPct_avg,
          LAG(femPct_avg)
          OVER (
              ORDER BY year) AS femPct_avg_prevYear,
          femPct_avg - LAG(femPct_avg)
          OVER (
              ORDER BY year) AS femPctAvg_annual_growth
    FROM femPctAvg_CTE)
SELECT CAST(ROUND((SUM(femPctAvg_annual_growth)/(COUNT(year)-1)),2) AS DECIMAL(18,2)) AS femPct_avg_growth
FROM femPctAvg_annual_growth_CTE;
```

avgPct_growth
3.67



## 9) Come si classifica l'Italia per quanto riguarda la media percentuale dei membri di genere femminile, appartenenti alla Camera Bassa.

```
WITH femPctAvg_rowNumber_CTE
AS( SELECT unifiedCountries,
          CAST(ROUND( AVG(COALESCE(FL.femaleLegislatorPercentage, 0))
                     , 2)
                AS DECIMAL(18,2)) AS femPct_avg,
          ROW_NUMBER()
            OVER (
              ORDER BY AVG( COALESCE(FL.femaleLegislatorPercentage, 0) )
              DESC) AS femPctAvg_rowNumber
  FROM dbo.targetArea AS TA
    LEFT JOIN dbo.sendingCountries AS SC
      ON TA.sendingCountryID = SC.sendingCountryID
    LEFT JOIN vw_unifiedCountries AS VW
      ON SC.countryID = VW.countryID
    LEFT JOIN dbo.femaleLegislators AS FL
      ON SC.femaleLegislatorID = FL.femaleLegislatorID
    LEFT JOIN dbo.years AS Y
      ON TA.yearID = Y.yearID
  GROUP BY unifiedCountries)

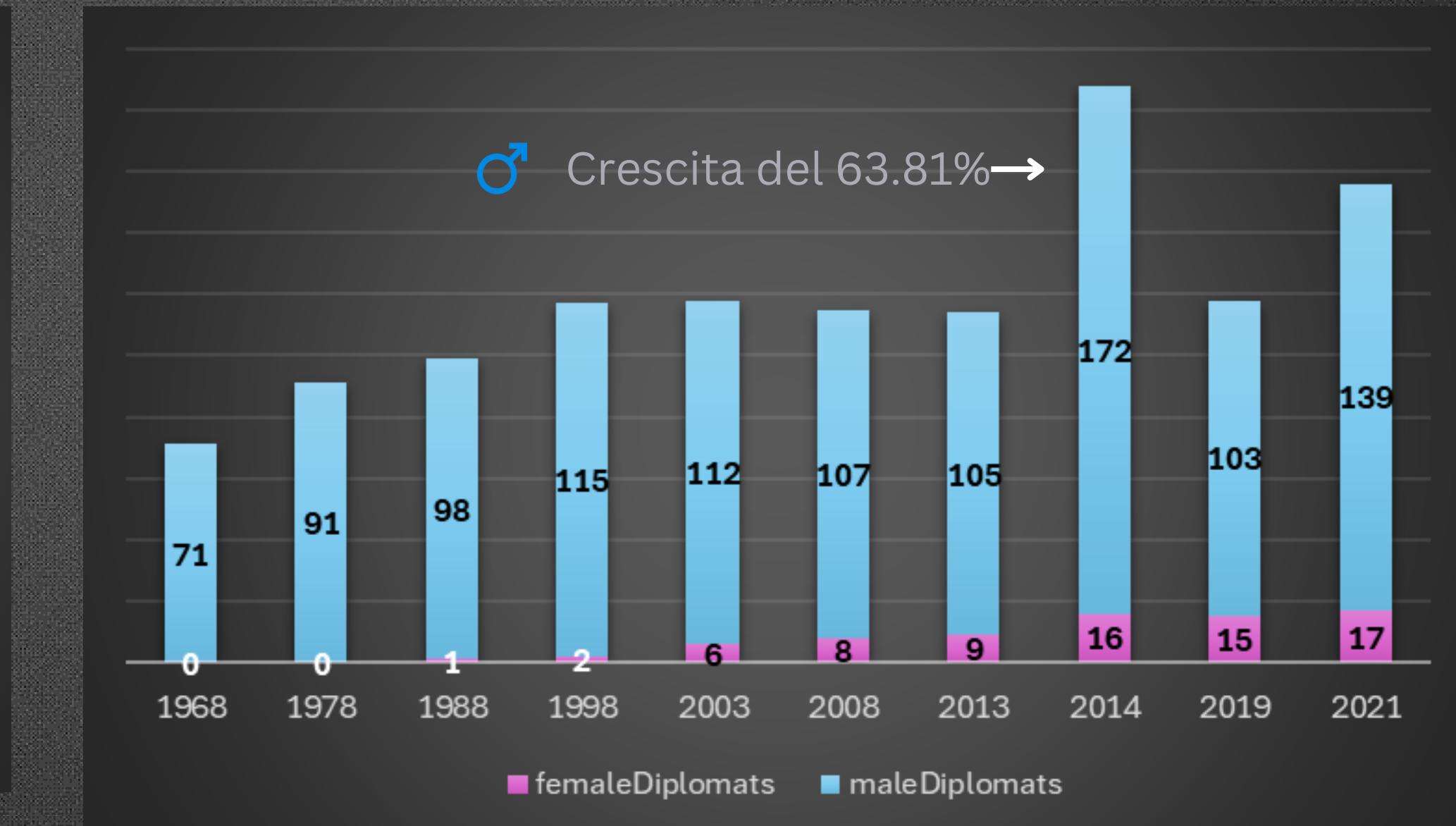
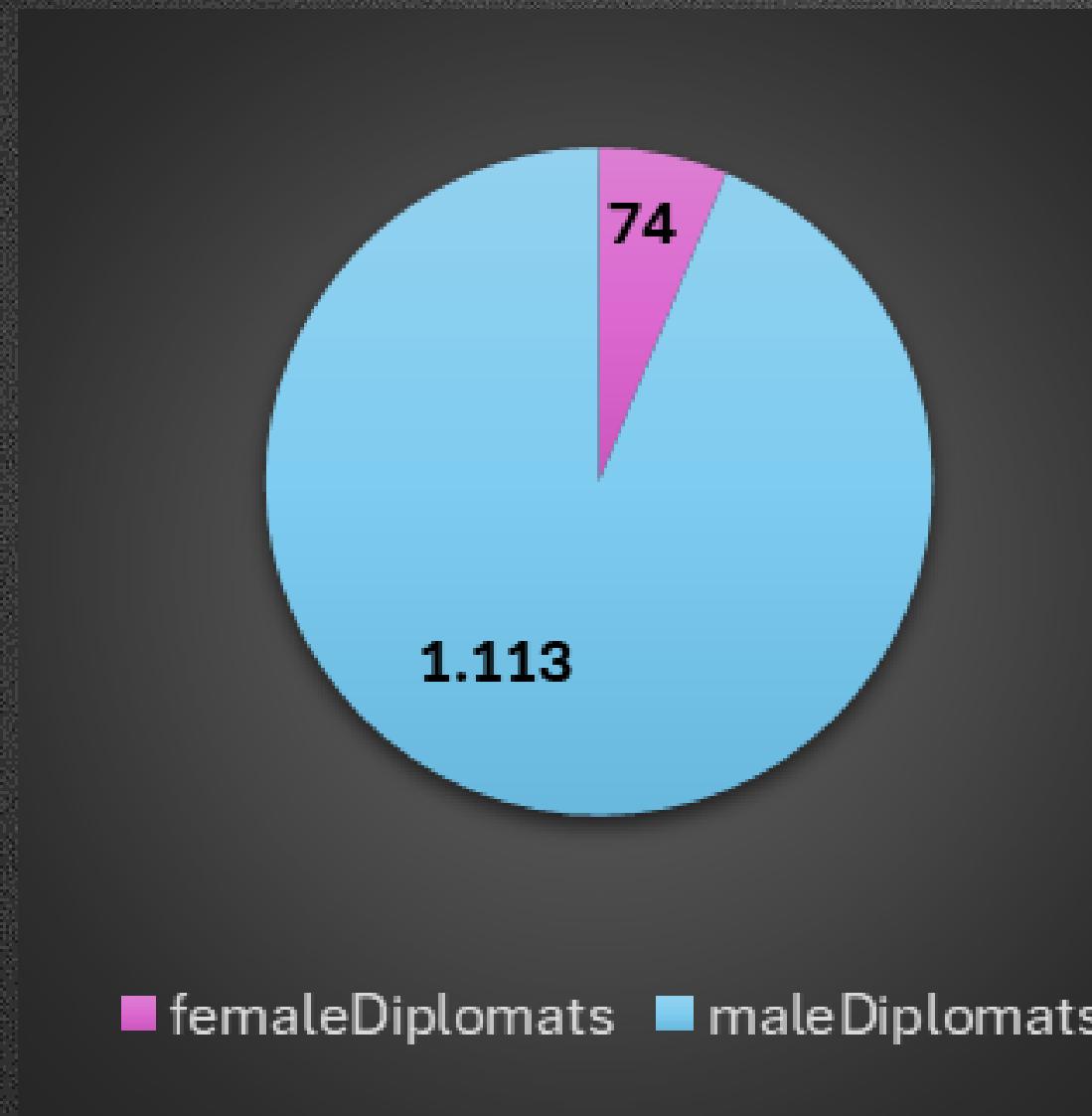
SELECT unifiedCountries, femPct_avg, femPctAvg_rowNumber
FROM femPctAvg_rowNumber_CTE
WHERE unifiedCountries = 'Italy';
```

unifiedCountries	femPct_avg	femPctAvg_rowNumber
Italy	21.60	57



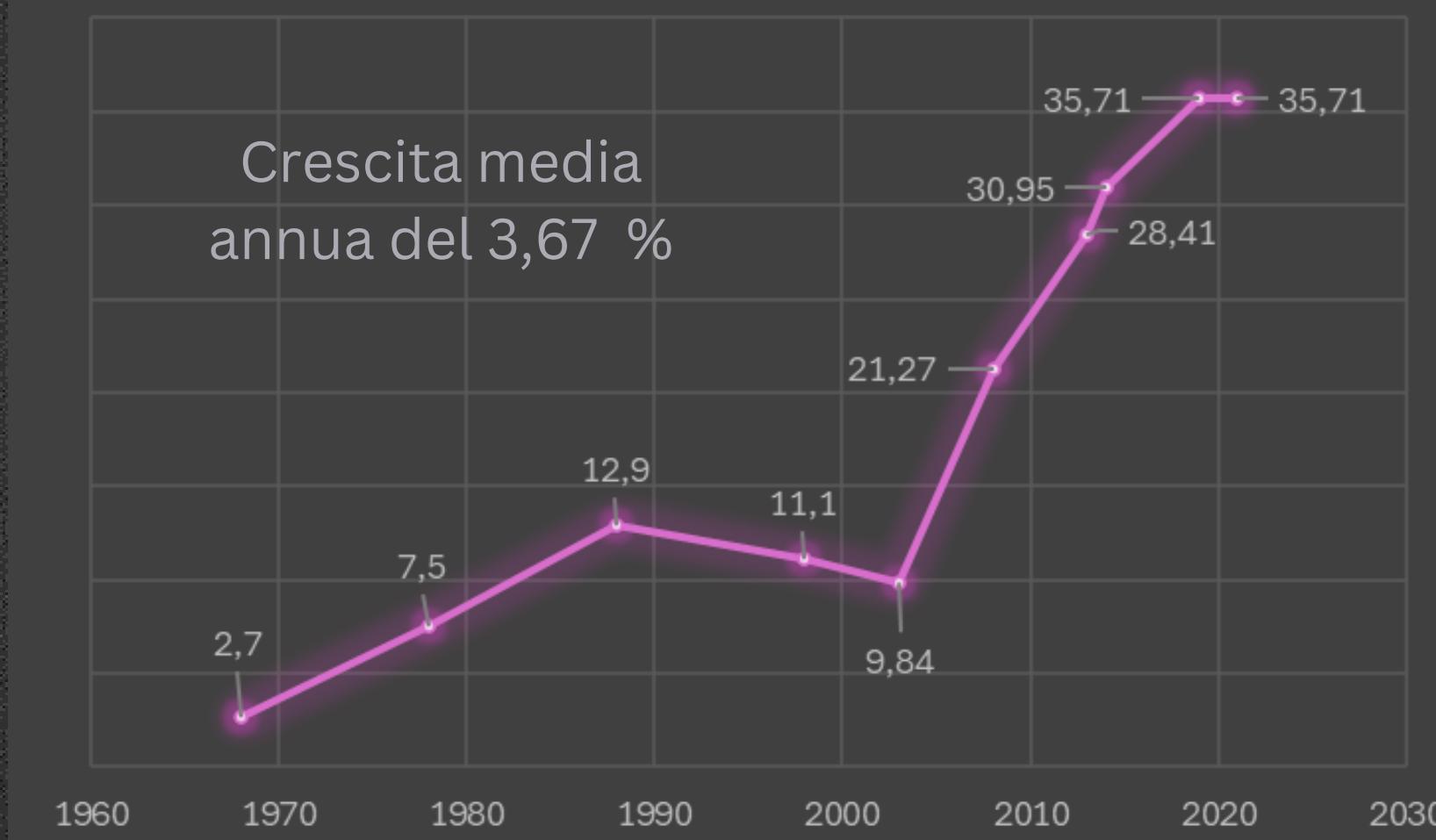
## TERZO FOCUS: Italia - analisi su paese specifico - GRAFICI.

Conteggio totale e annuale di diplomatici per genere.



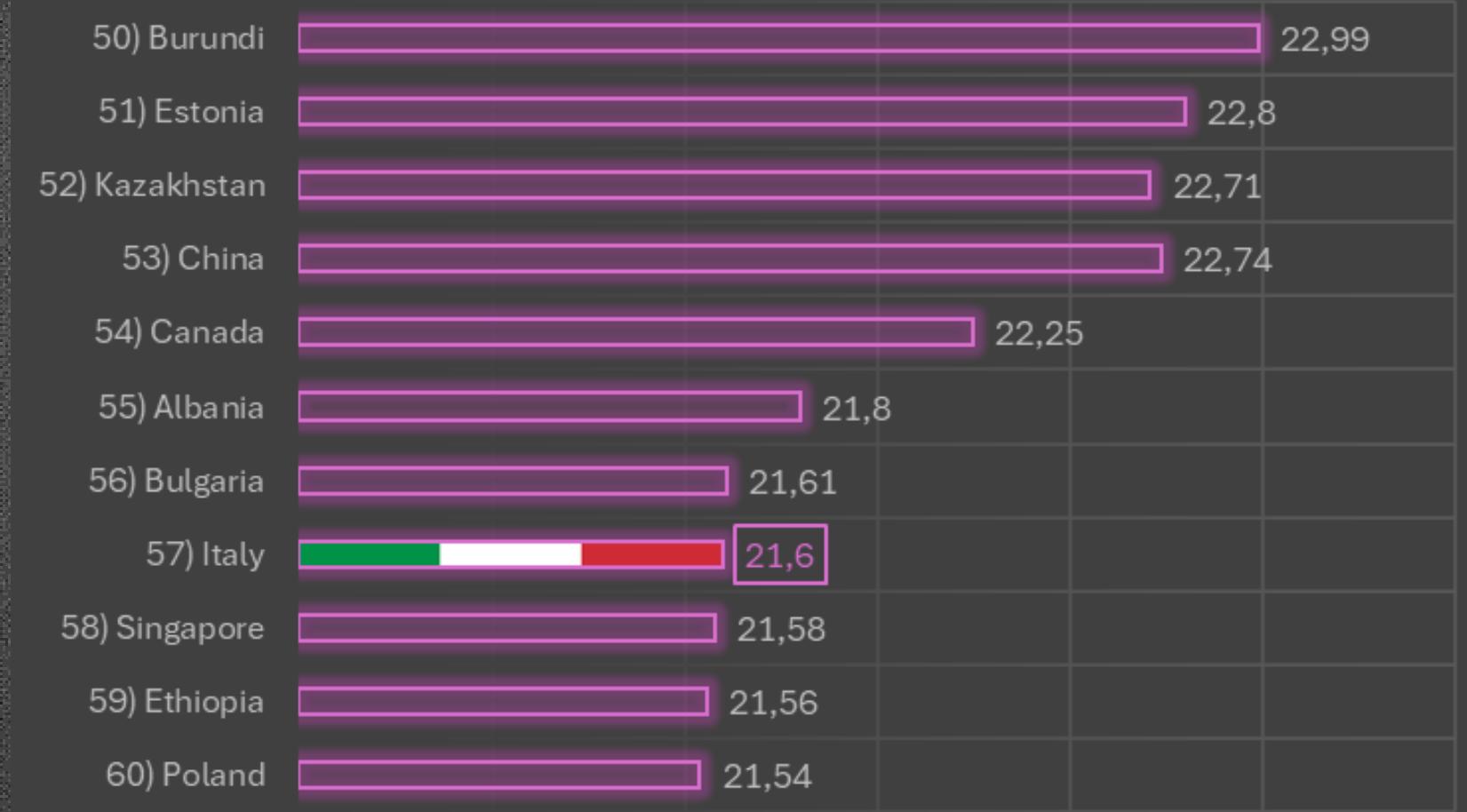
## TERZO FOCUS: Italia - analisi su paese specifico - GRAFICI.

Media percentuale dei legislatori femminili



Media percentuale dei legislatori femminili

Posizione dell'Italia nel mondo



## Analisi dei dati

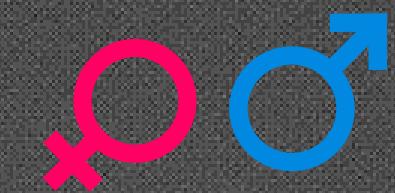
Italia - analisi su paese specifico.

- Nel 1968 la media percentuale dei membri di genere femminile, appartenenti alla Camera Bassa, era del 2,7 %, arrivando al 35,71% nel 2021, con una crescita media annua del 3,67%.
- Per quanto riguarda la media percentuale dei membri di genere femminile, appartenenti alla Camera Bassa, l'Italia raggiunge il 21,60% arrivando al 57esimo posto su 203 paesi.
- Il conteggio totale dei diplomatici di genere femminile è di 74, mentre quello maschile è di 1.113.
- Nel 2021 si è registrato il maggior numero di diplomatici di genere, per un totale di 17, mentre mentre la maggior presenza di genere maschile si è registrata nel 2014 con 172, in quell'anno c'è stato un'aumento di 67 diplomatici di genere maschile rispetto all'anno precedente, ovvero del 63,81%, numericamente è la crescita annuale più significativa registrata.



# PROGETTO SQL

Rappresentazione di genere nella diplomazia



Grazie per l'attenzione

---

FRANCESCA CARRERA