

**Santiago de Cali, agosto 12 del 2025.**

**Proyecto Bootcamp “TalentoTech”:**

Curso de Análisis de Datos Básico

**Nombre del proyecto:**

Colombianos registrados en el exterior

**Estudiantes:**

Carlos David Ramírez Ospina

Marcos Ronaldo Beltrán Guerra

**Docentes:**

Ing. Jonathan Giraldo

Ing. Dnilson Olaya

## **1. Introducción**

El presente análisis se enfoca en la exploración y comprensión de la base de datos de colombianos registrados en consulados en el exterior, un conjunto de datos crucial para entender los patrones demográficos y de migración de la población colombiana a nivel global.

Este dataset, con un total de 1.747.917 registros (del cual se eliminaron 678.873), proporciona información detallada que incluye variables como el país de residencia, la edad, el nivel académico, el estado civil, el género, y el área de conocimiento de los registrados. El análisis preliminar ha revelado la necesidad de una limpieza y preparación de datos exhaustiva, debido a la presencia de valores atípicos y errores de entrada.

El objetivo de este estudio es transformar los datos brutos en insights significativos que puedan ser de utilidad para la toma de decisiones, especialmente para entidades gubernamentales y/o sociales como el Ministerio de Relaciones Exteriores. A través de este análisis, se buscará identificar las tendencias migratorias, los perfiles sociodemográficos de los colombianos en el exterior y las áreas de conocimiento predominantes, lo que permitirá generar recomendaciones para el desarrollo de políticas públicas y servicios de apoyo a la diáspora colombiana.

## **2. Objetivo del Análisis:**

El objetivo es identificar patrones que se evidencien luego de la limpieza y el análisis de los datos, para considerar la implementación de estrategias o campañas sociales que favorezcan a la población colombiana.

## **3. Descripción del Data set:**

- Descripción General: La base de datos contiene "información de la población colombiana residente y registrada en las diferentes misiones consulares en el exterior".
- Tamaño Inicial: El conjunto de datos tiene un tamaño inicial de 1,747,917 filas y 17 columnas.
- Columnas Principales: Las columnas incluyen datos como País, Ciudad de Residencia, Código ISO país, Oficina de registro, Grupo edad, Edad (años), Área Conocimiento, Sub Área Conocimiento, Nivel Académico, Estado Civil, Género, Etnia de la persona, Estatura (CM), Ciudad de Nacimiento, Localización, Fecha de Registro y Cantidad de personas.
- Tipo de Datos: La mayoría de las columnas son de tipo object (texto), mientras que algunas, como Edad (años), Estatura (CM) y Cantidad de personas son de tipo numérico (int64).

#### **4. Limpieza y Análisis Preliminar de los Datos:**

- **Valores Nulos:** Un análisis inicial muestra la presencia de valores nulos, que deberán ser tratados para asegurar la integridad del estudio. Se puede usar `df.isnull().sum()` para identificar la cantidad de nulos por columna.
- **Datos Atípicos:** Se han identificado datos atípicos en la columna Edad (años), con valores de -1. Estos valores, al ser biológicamente imposibles, deben ser limpiados o imputados de forma inteligente.
- **Datos Duplicados:** Es necesario verificar la existencia de filas duplicadas para evitar sesgos en el análisis, lo cual se puede hacer con `df.duplicated().sum()`.
- **Corrección de Codificación:** Al cargar el archivo CSV, se ha utilizado una codificación utf-8-sig para manejar correctamente los caracteres especiales y acentos en los nombres de las columnas.

## 5. Análisis e Imputación de Datos:

- Métodos de Imputación: Para manejar los datos faltantes en la columna Edad (años), una estrategia propuesta es la imputación por grupos, utilizando la mediana de la edad de perfiles que coincidan en Género, Nivel Académico, Estado Civil y País.
- Visualización: El análisis inicial de los datos se puede complementar con estadísticas descriptivas (`df.describe()`) para entender la distribución de las variables numéricas y con visualizaciones para identificar patrones de manera más efectiva.

*Fuente de datos:* <https://www.datos.gov.co/Estad-sticas-Nacionales/Colombianos-en-el-exterior-de-25-a-40-a-os/3k8j-4d42>

- Tamaño Inicial: 1.747.917 filas y 17 columnas.

## 6. Preparación de los Datos:

- Inicialmente, se importa el dataset y se visualiza su contenido para tener una primera aproximación a la estructura y variables disponibles.
- A continuación, se realiza una búsqueda de datos nulos en las distintas columnas del conjunto de datos.
- Si bien no se identifican valores nulos, se evidencian inconsistencias en los datos relacionados con la edad, ya que se encuentran registros con valores como “-1” y “139”. De manera similar, en la columna de estatura también se detectan valores inválidos como “-1”.
- Posteriormente, se procede a analizar los datos correspondientes a la edad, con el fin de tomar decisiones informadas respecto a las filas que contienen estos valores atípicos.
- Se detectan y corrigen errores tipográficos en la columna socioeconomic\_status, asegurando la coherencia en las categorías.
- Luego, se elimina la columna denominada **estatura (cm)**, debido a que esta variable no se considera relevante para el análisis posterior.
- Continuando con el proceso de limpieza de datos, se identifican fechas registradas con el año 1900. Este valor es considerado un error, ya que representa un dato

atípico dentro del contexto del dataset. Por tanto, se evalúa la cantidad total de registros que contienen fechas con el año igual a 1900, con el objetivo de determinar su impacto en el análisis.

## **7. Análisis Exploratorio**

**Código Python (para Jupyter Notebook o Colab)**

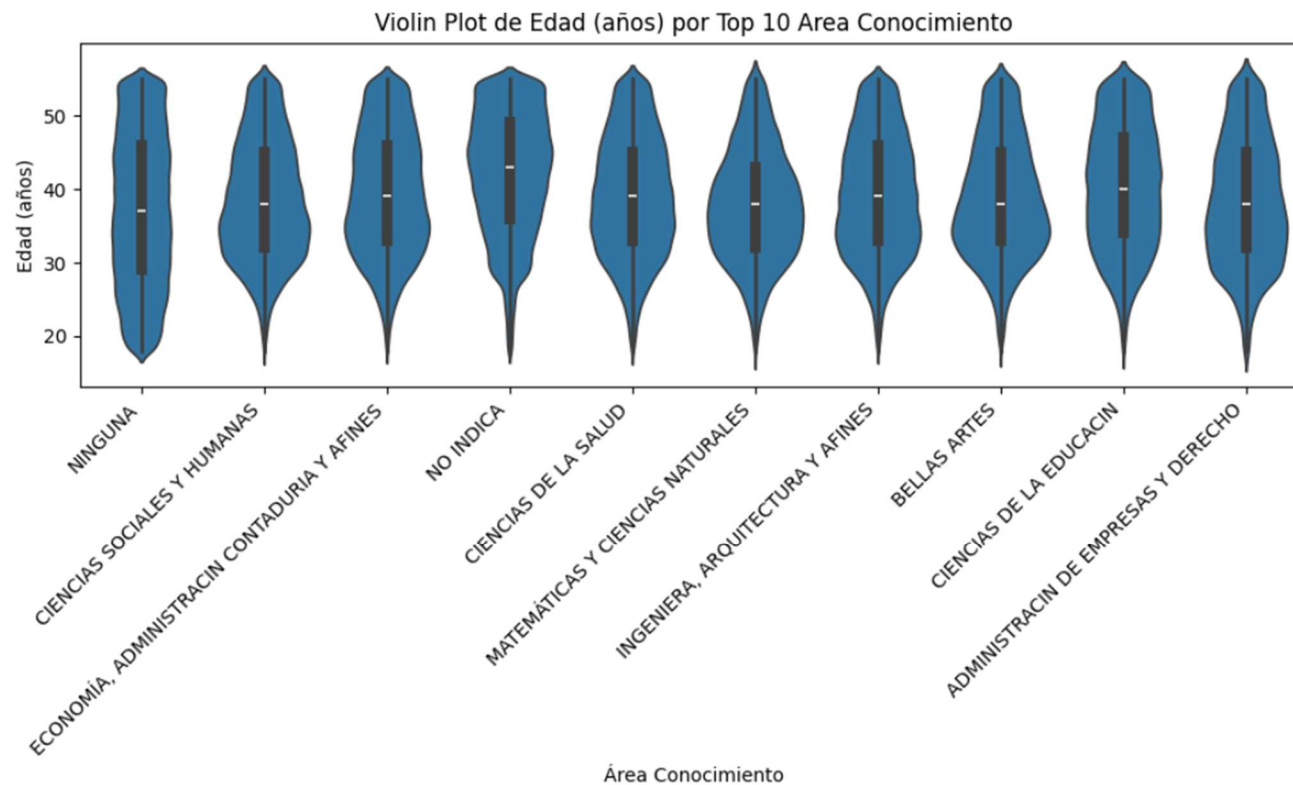
**Link de google colab del analisis de datos.**

**<https://colab.research.google.com/drive/16RsQrd-mUeWDEGwGVeUisEkeIq2xlnqK?usp=sharing#scrollTo=1s-yIN0fziG8>**



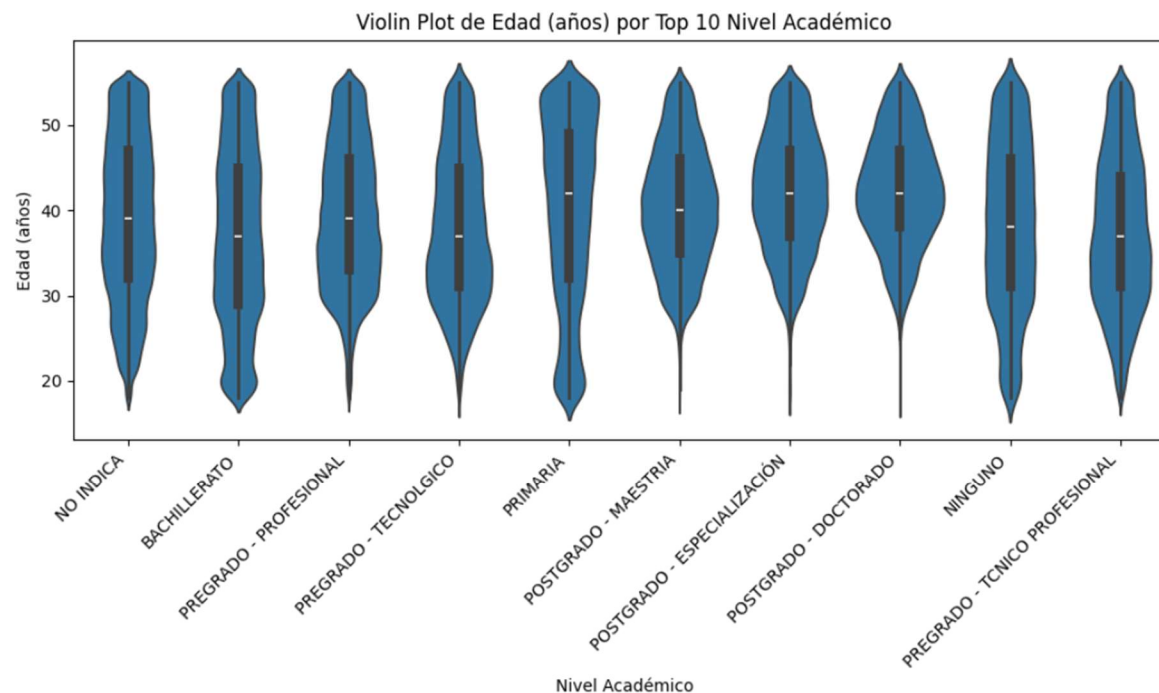
## 7.1. Visualización 1:

- El gráfico muestra la distribución de la edad (en años) en las 10 principales áreas de conocimiento. En general, la mayoría de las personas se concentran en un rango de **30 a 50 años**, con medianas cercanas a los 35-40 años en casi todas las áreas.



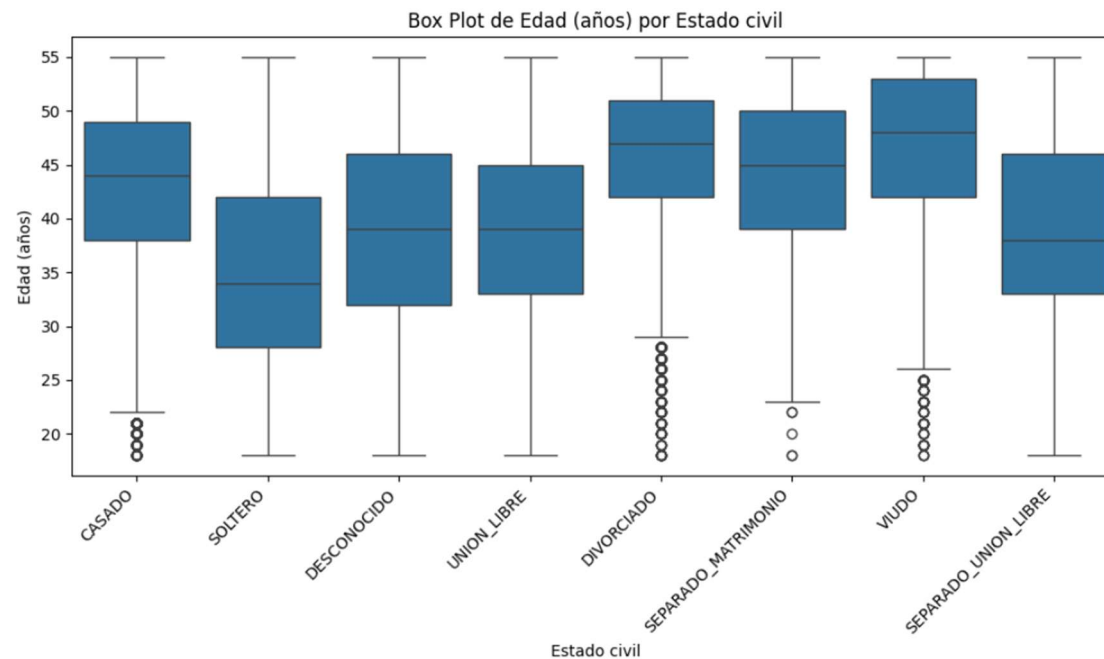
## 7.2. Visualización 2:

- El análisis revela que los niveles académicos más altos (doctorado, maestría y especialización) están asociados a edades promedio más elevadas, mientras que los niveles intermedios (pregrado y tecnológico) concentran personas más jóvenes. Los grupos sin nivel declarado o con primaria presentan gran diversidad de edades. En general, la franja de 30 a 50 años es predominante en todos los niveles académicos.



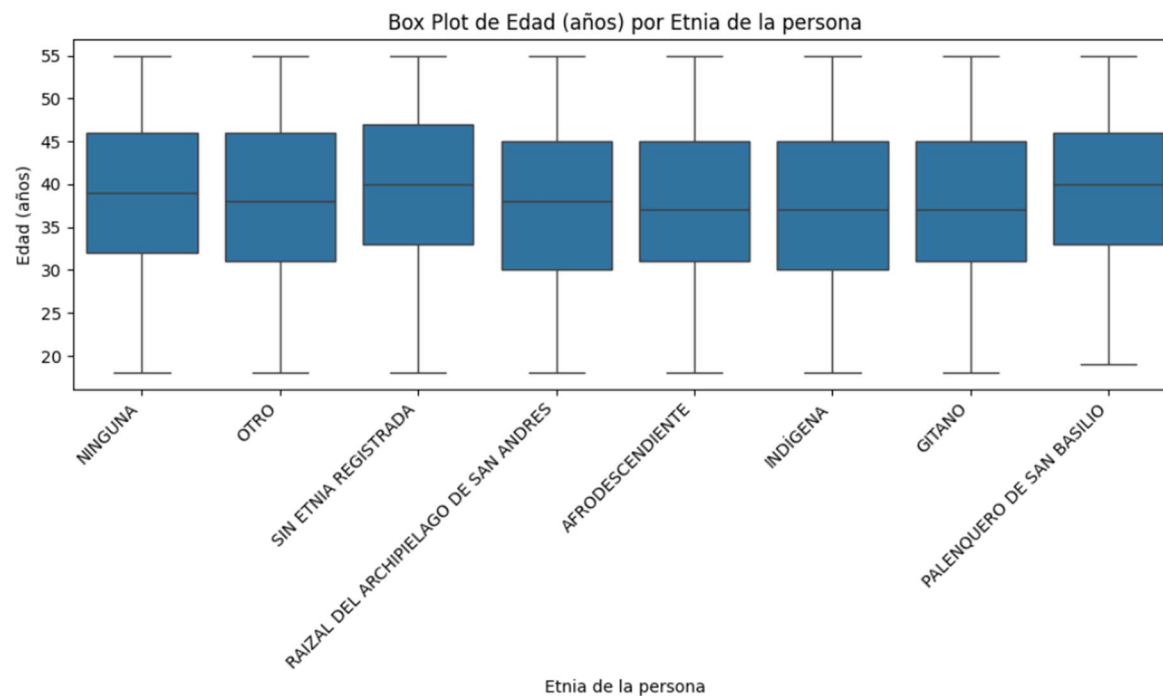
### 7.3. Visualización 3:

- El análisis muestra que los estados civiles asociados a rupturas o pérdidas (viudo, divorciado, separado) presentan edades promedio más altas, cercanas o superiores a los 45 años. Por el contrario, el estado de soltero se asocia a las edades más bajas (alrededor de 34 años), mientras que unión libre y casado se ubican en un rango intermedio. Los grupos muestran cierta dispersión de edades, destacando casos atípicos que no siguen la tendencia general.



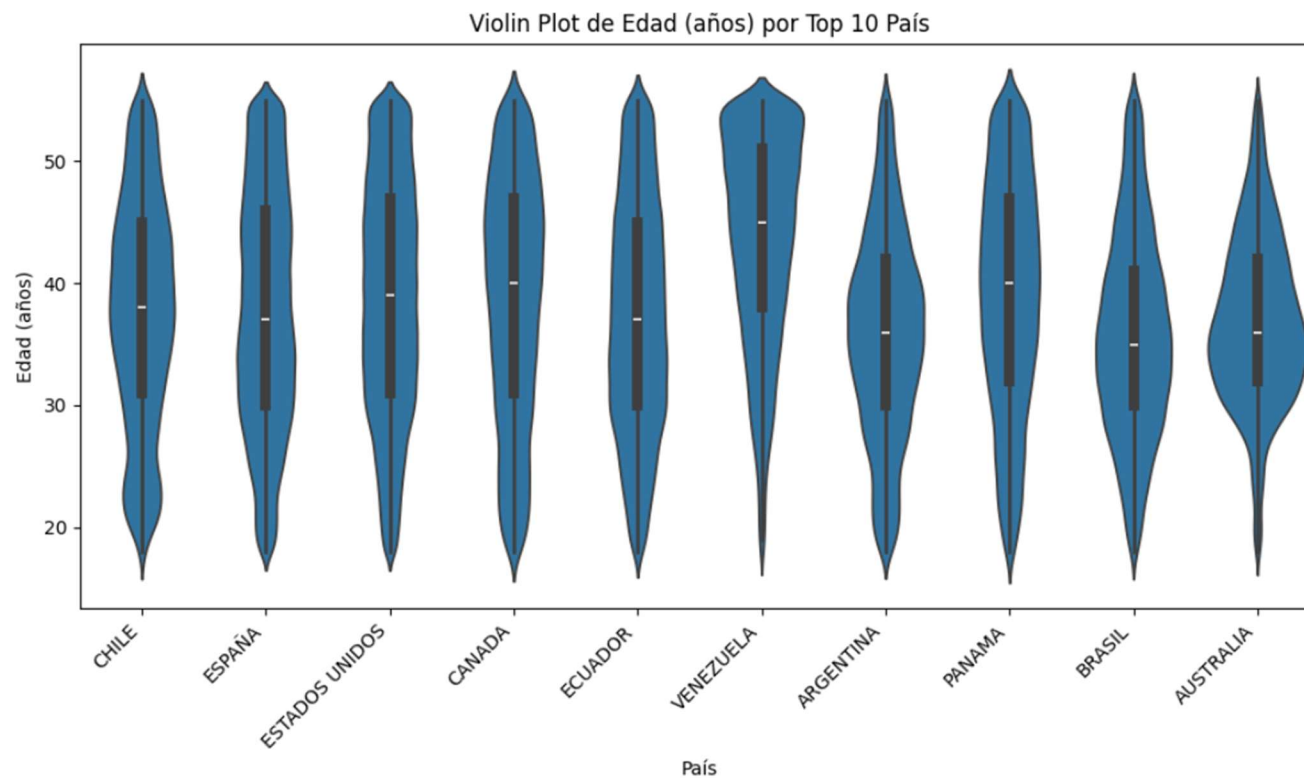
## 7.4. Visualización 4:

- La edad promedio de los registrados varía poco entre etnias, situándose mayoritariamente entre los 35 y 40 años. Las etnias *Raizal* y *Palenquera* muestran medianas ligeramente superiores, mientras que *Indígena* y *Afrodescendiente* presentan edades algo menores. La dispersión es similar, lo que sugiere que la variable etnia no influye significativamente en la edad.



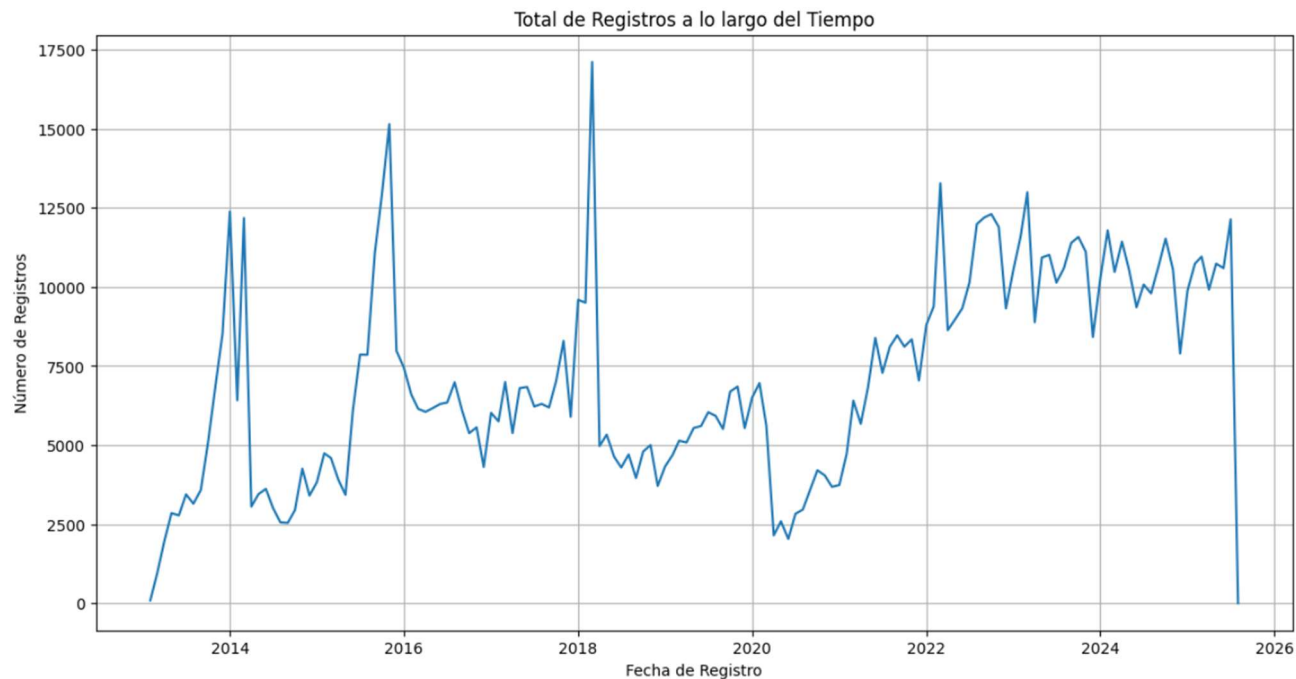
## 7.5. Visualización 5:

- La mayoría de los registros se concentran en edades de 30 a 50 años. Países como Chile, Ecuador y Venezuela presentan edades promedio más altas, mientras que Australia y Brasil concentran personas más jóvenes. Aunque hay variabilidad interna, las diferencias entre países son notorias en la distribución de edades.



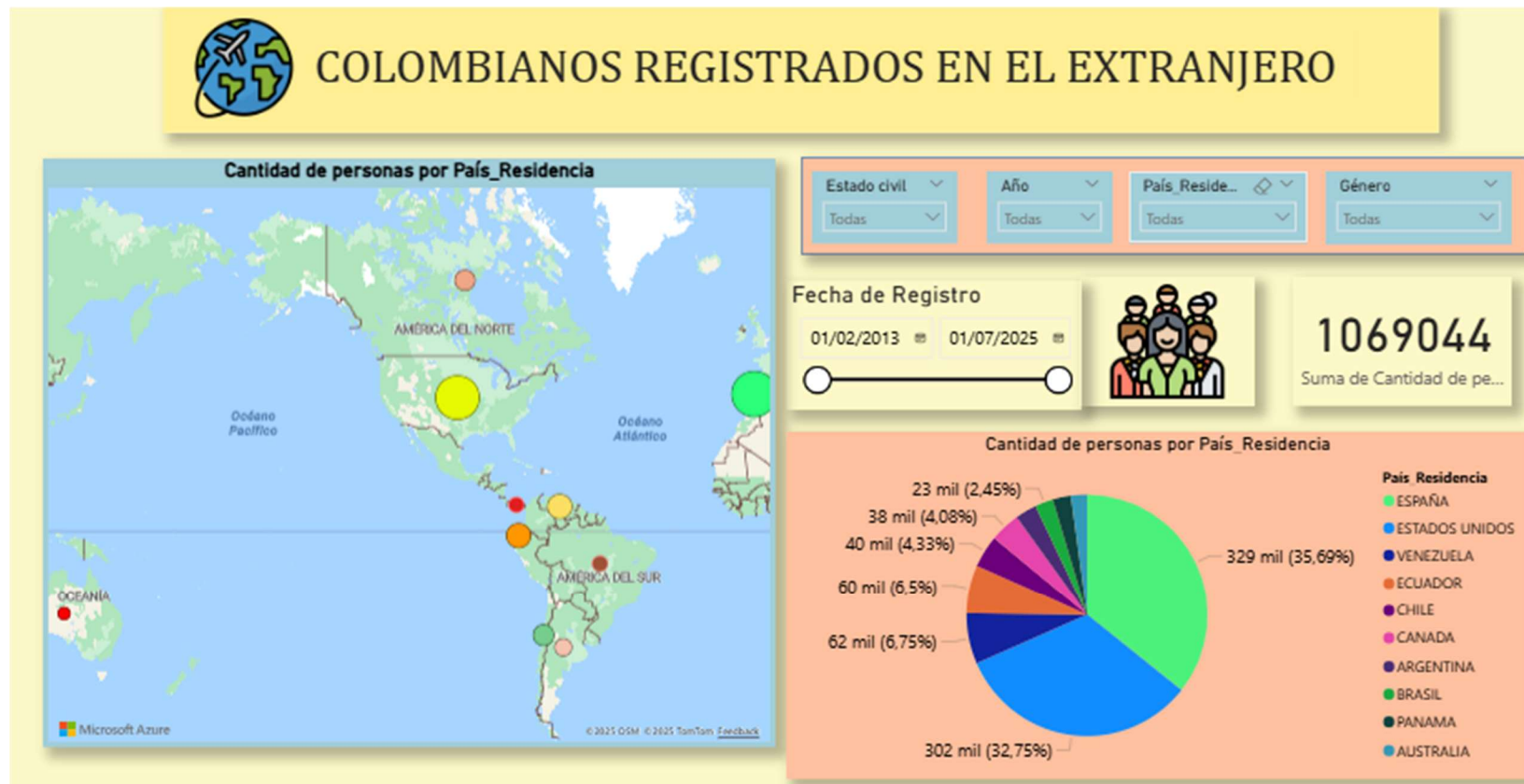
## 7.6. Visualización 6:

- El número de registros ha mostrado picos relevantes en años específicos (2015, 2018 y 2022), probablemente vinculados a campañas o cambios normativos. Tras 2022, el volumen se mantiene alto con una leve caída reciente, posiblemente por cierre de periodo de inscripción.



## 8. Visualizaciones en POWER BI:

8.1. ¿Cuáles son los 10 países que registran mayor cantidad de colombianos residentes?



## 8.2. Cuál es la cantidad de colombianos residentes en el extranjero, en función de su nivel académico?

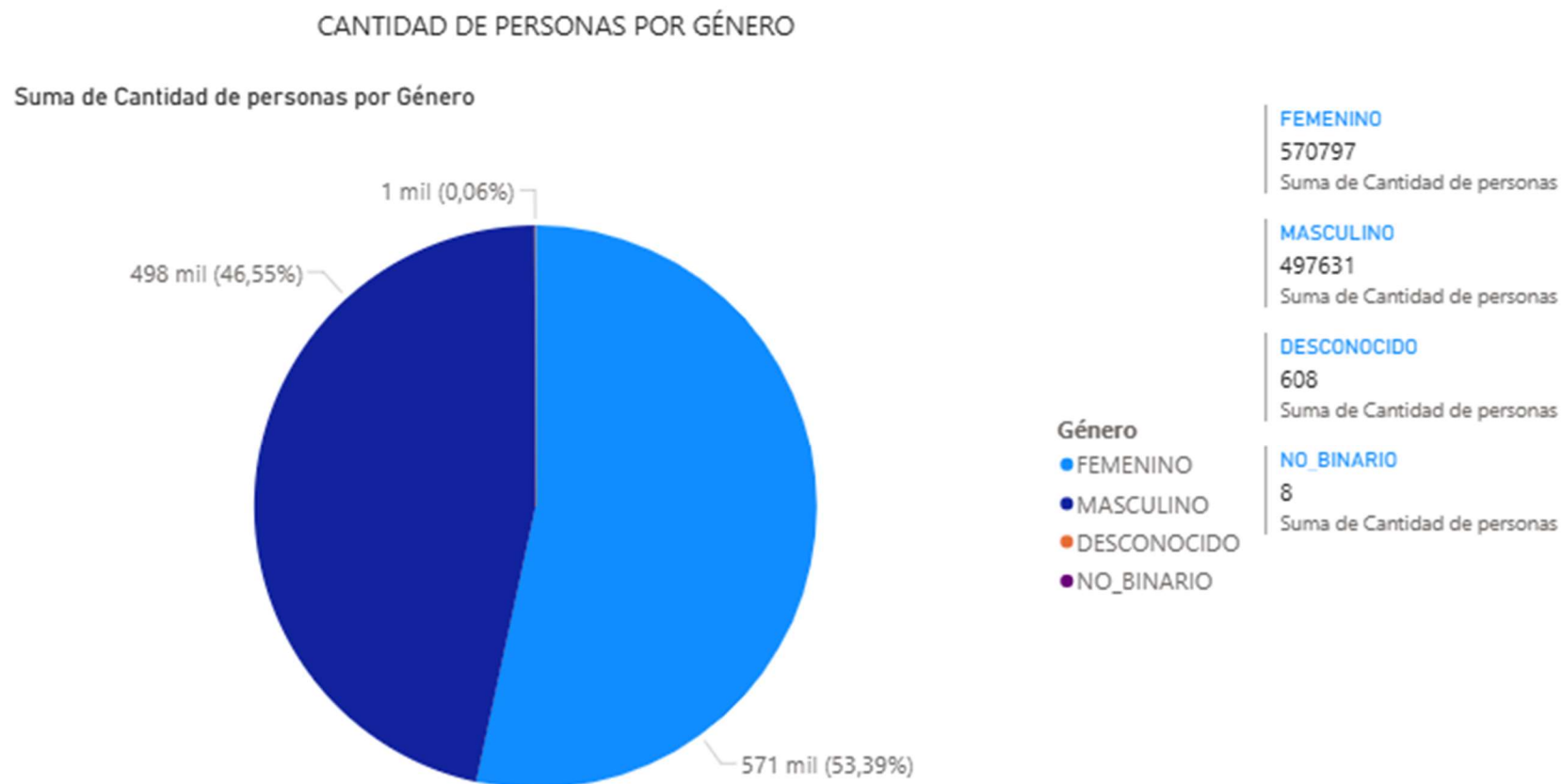




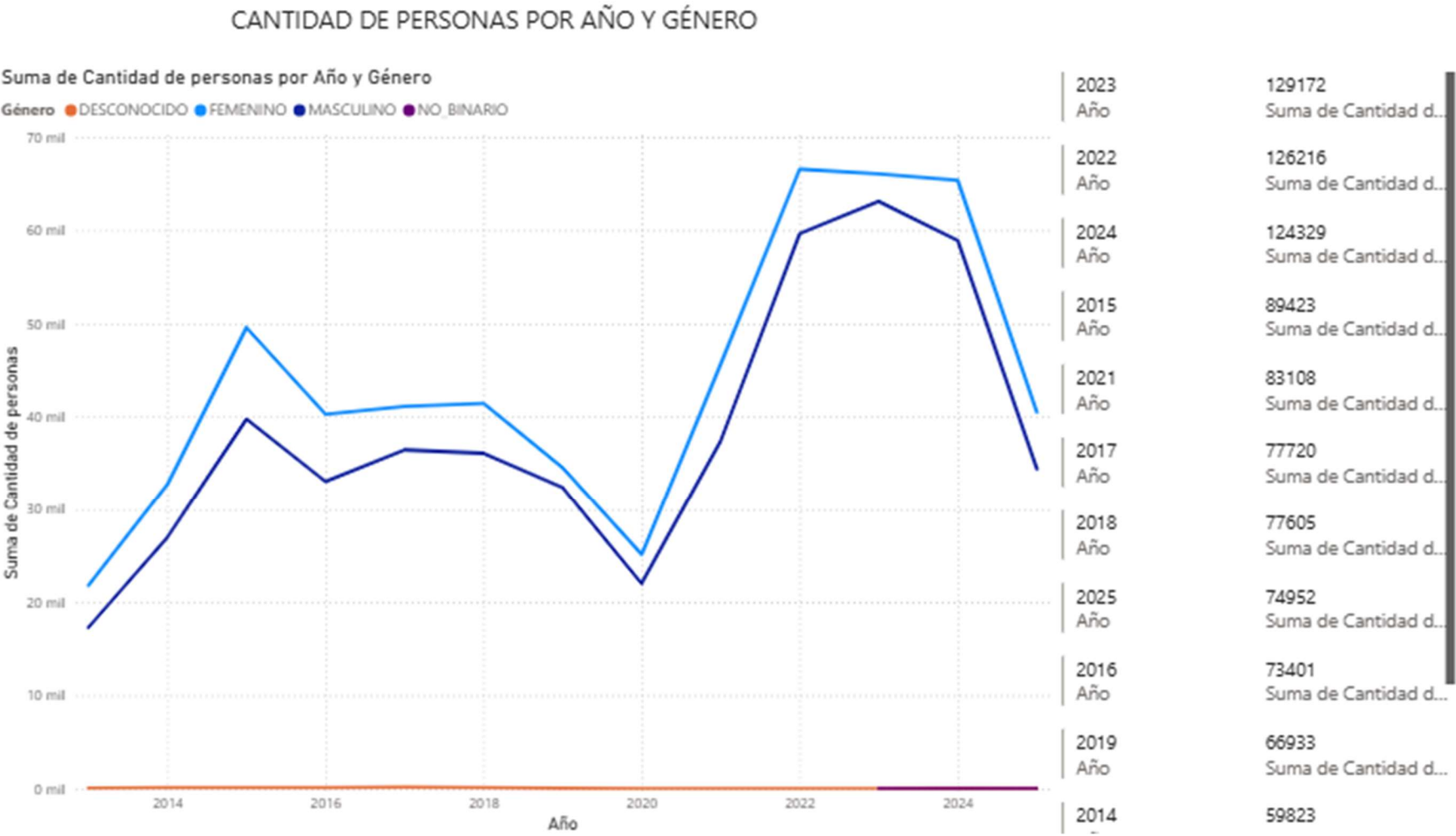
8.3. ¿Existe alguna tendencia insinúe la continuación de emigración de colombianos?



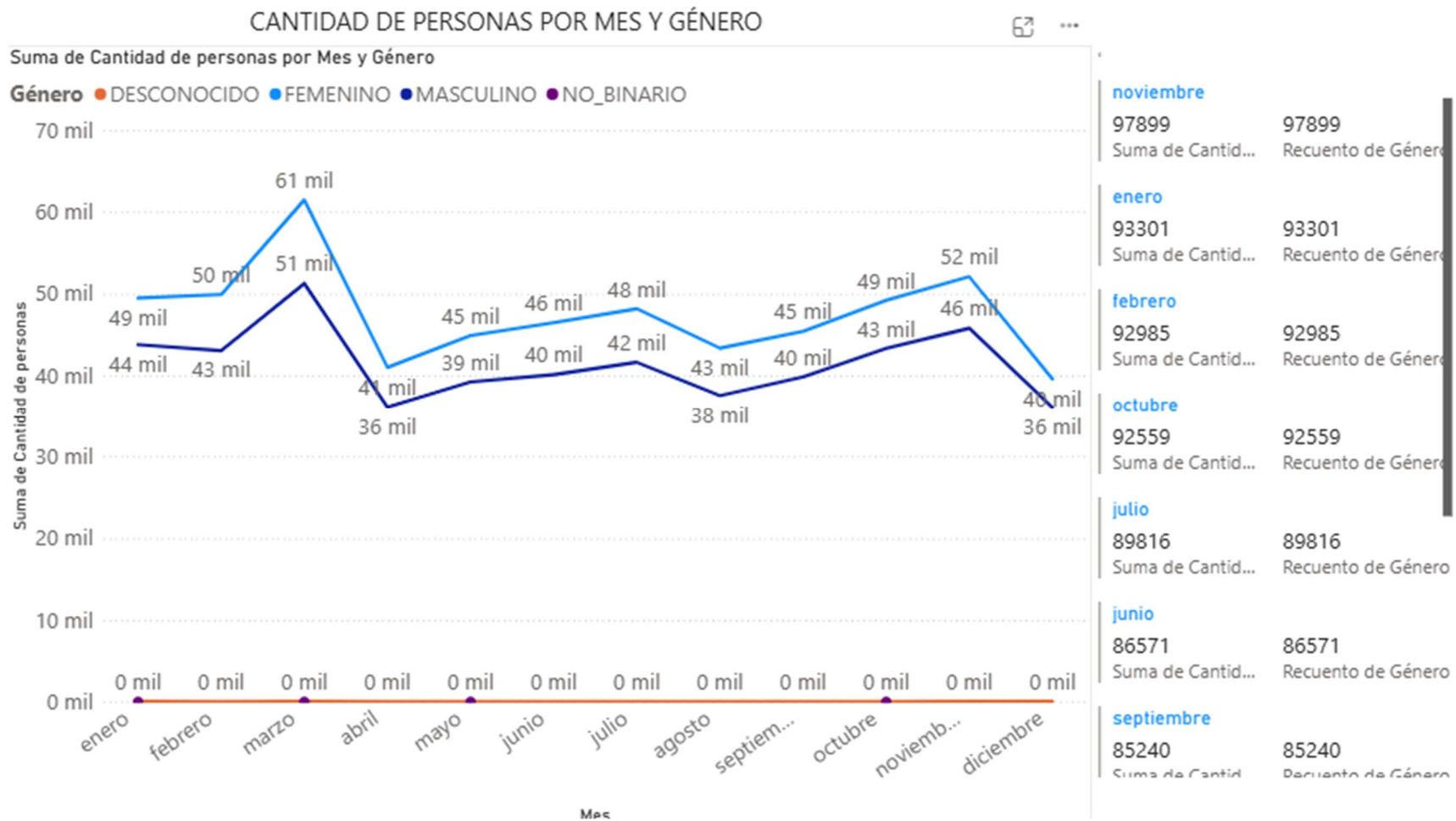
#### 8.4. ¿Qué comportamiento se refleja en función del género de los colombianos que residen en el extranjero?



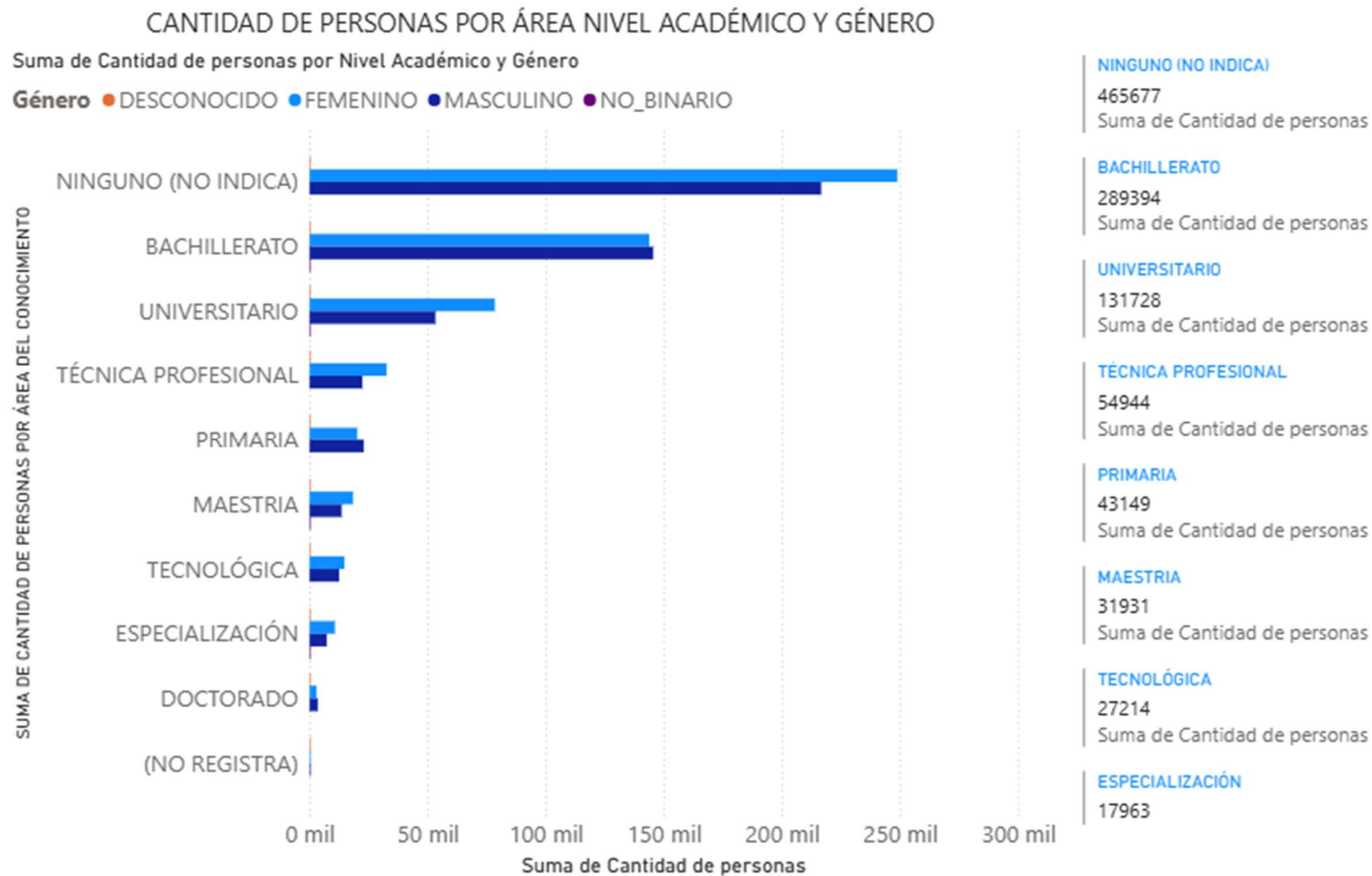
8.5. ¿Cómo se ha cambiado la cantidad de colombianos emigrando, en función del tiempo?



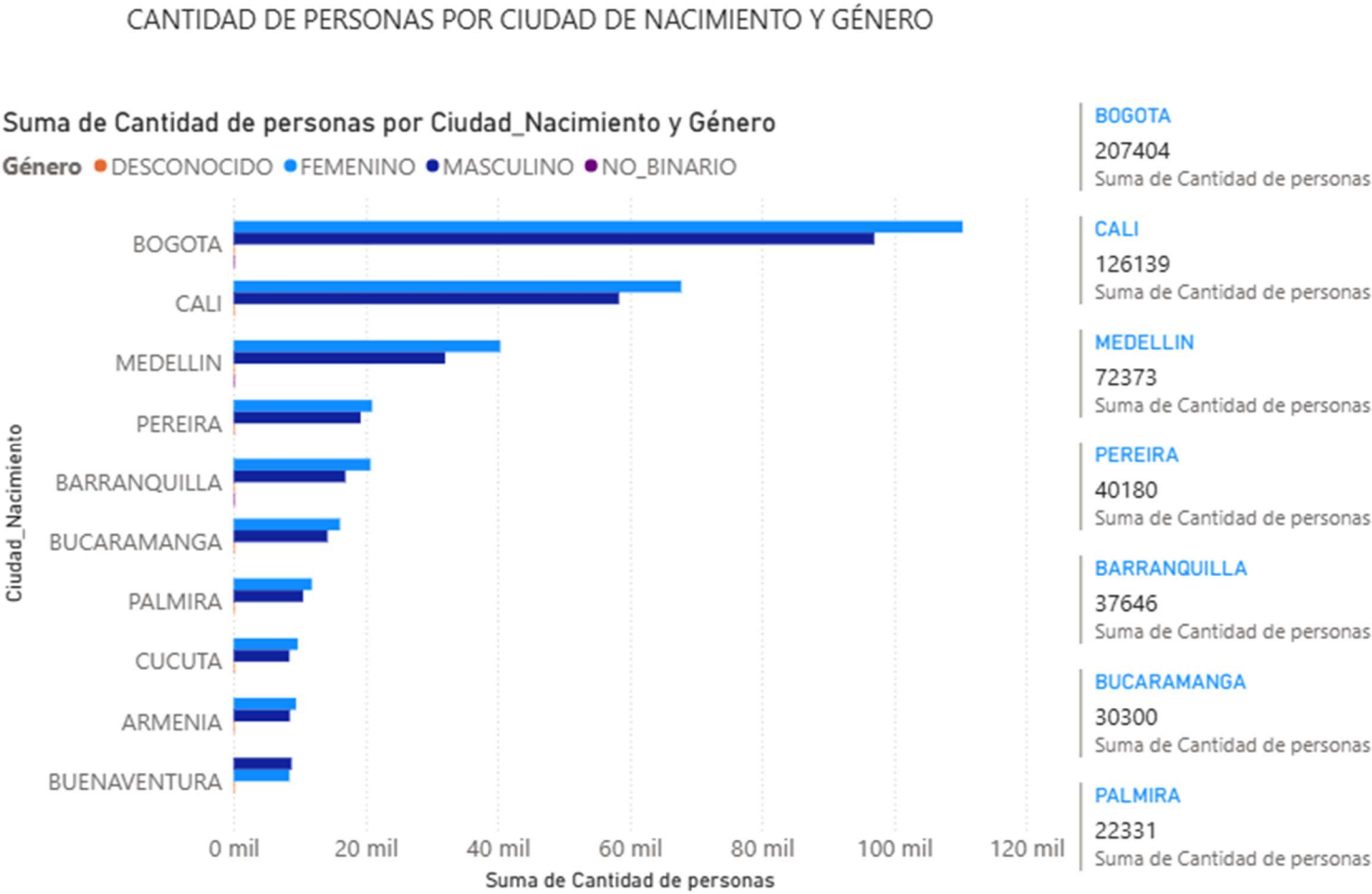
## 8.6. ¿Existe alguna época del año, en donde se evidencie el aumento de emigración?



## 8.7. ¿Existe alguna relación evidente entre el género y su nivel académico?

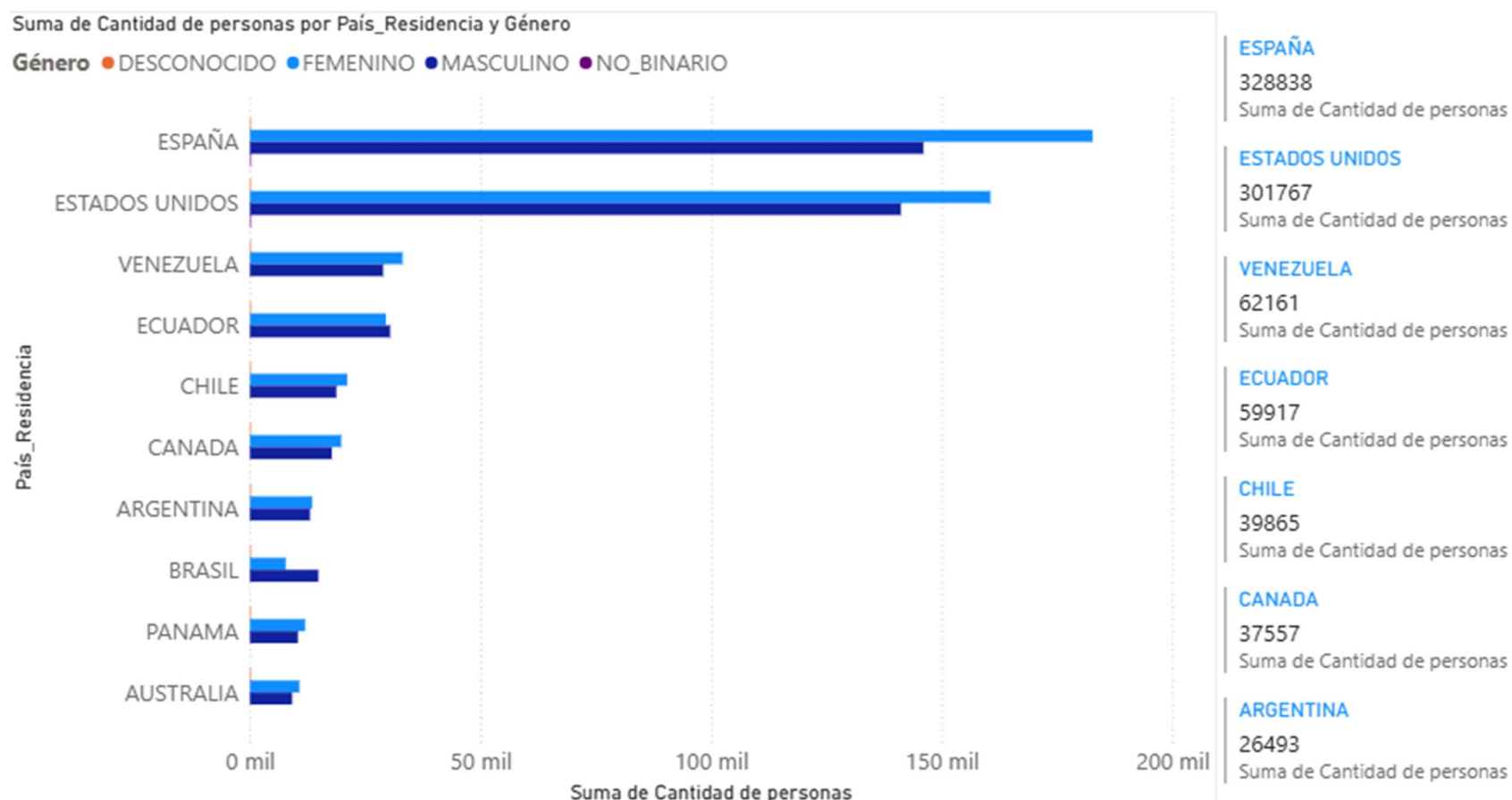


8.8. ¿Cuáles son las 10 ciudades que colombianos emigrantes reportan?

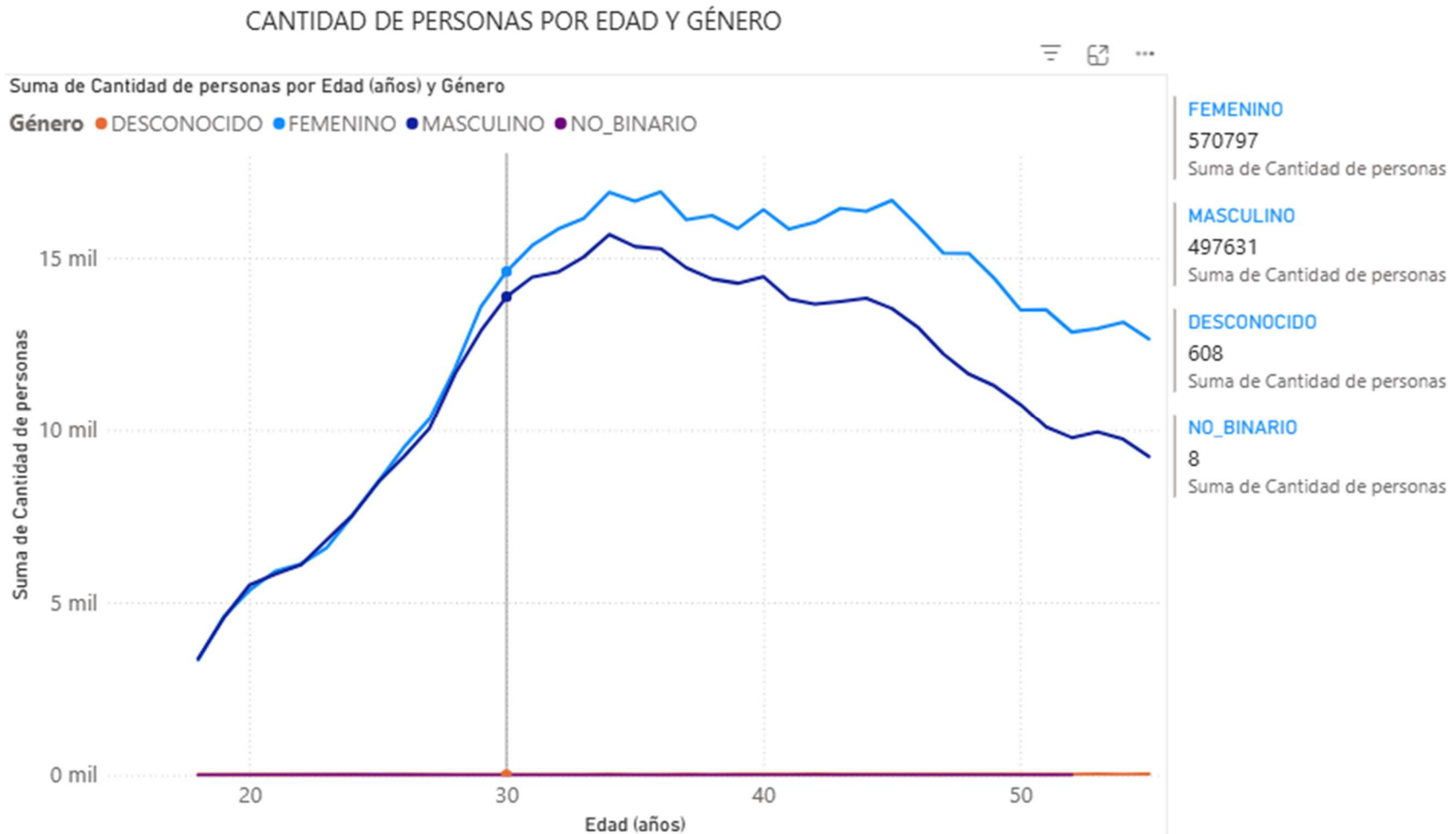


## 8.9. ¿Cuáles son los 10 países que más colombianos emigrantes reportan?

CANTIDAD DE PERSONAS POR PAÍS DE RESIDENCIA Y GÉNERO



## 8.10. ¿Existe alguna relación entre las edades y los géneros?





### 8.11. ¿Existe alguna relación entre el nivel académico y el país de residencia?

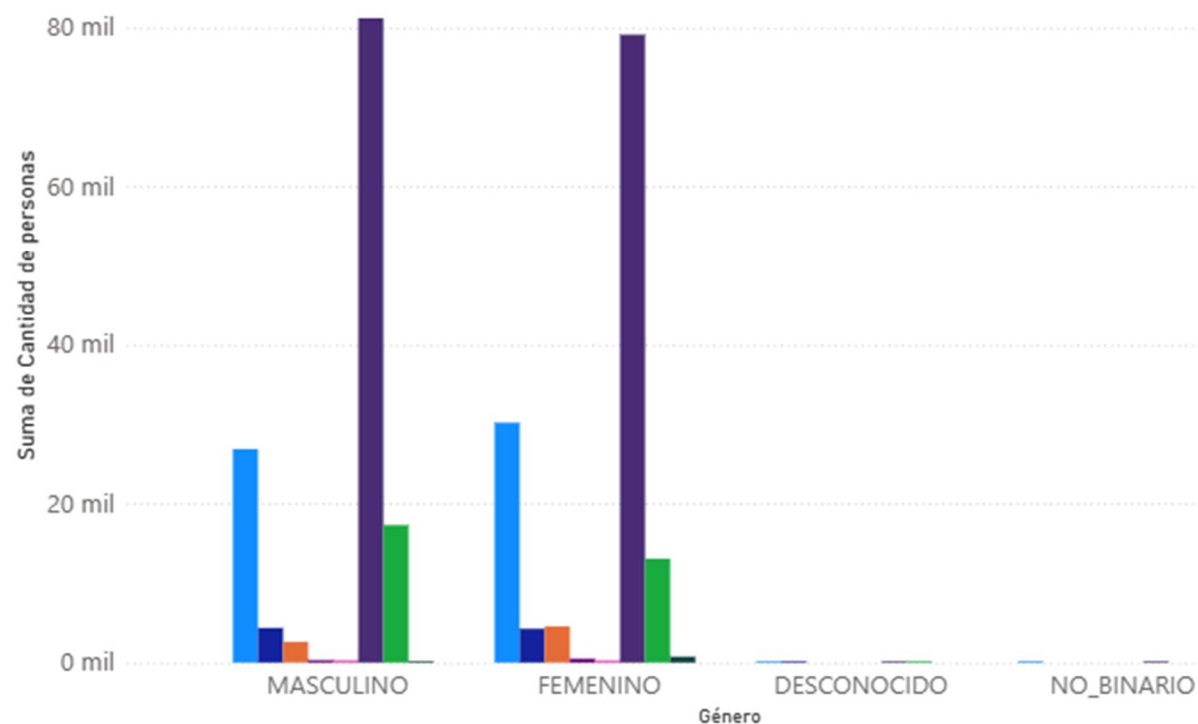


## 8.12. ¿Existen relaciones entre el estado civil y el género?

CANTIDAD DE PERSONAS POR GÉNERO Y ESTADO CIVIL

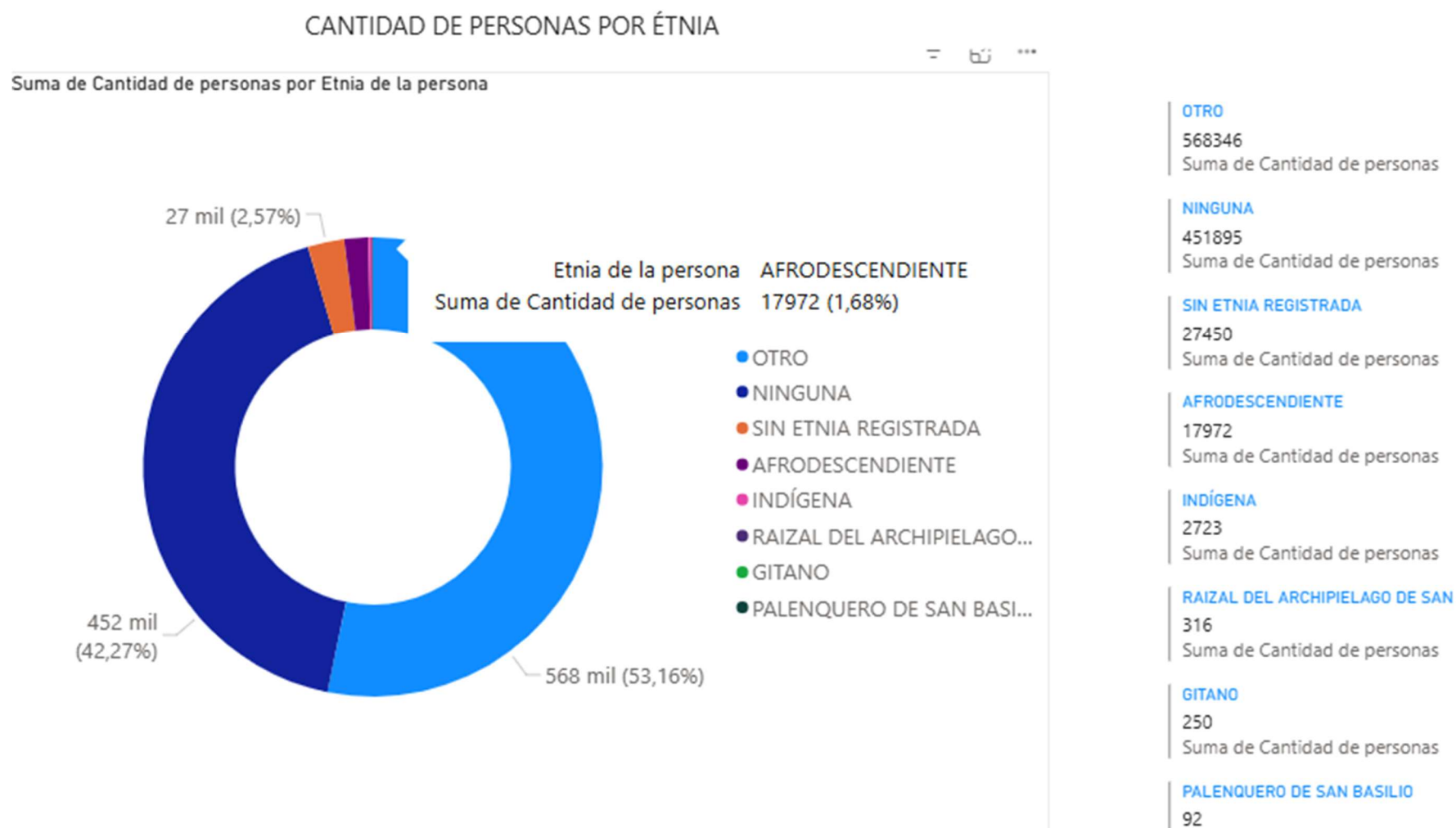
Suma de Cantidad de personas por Género y Estado civil

**Estado civil** CASADO DESCONO... DIVORCIA... SEPARAD... SEPARA... SOLTERO UNION\_LI... VIUDO

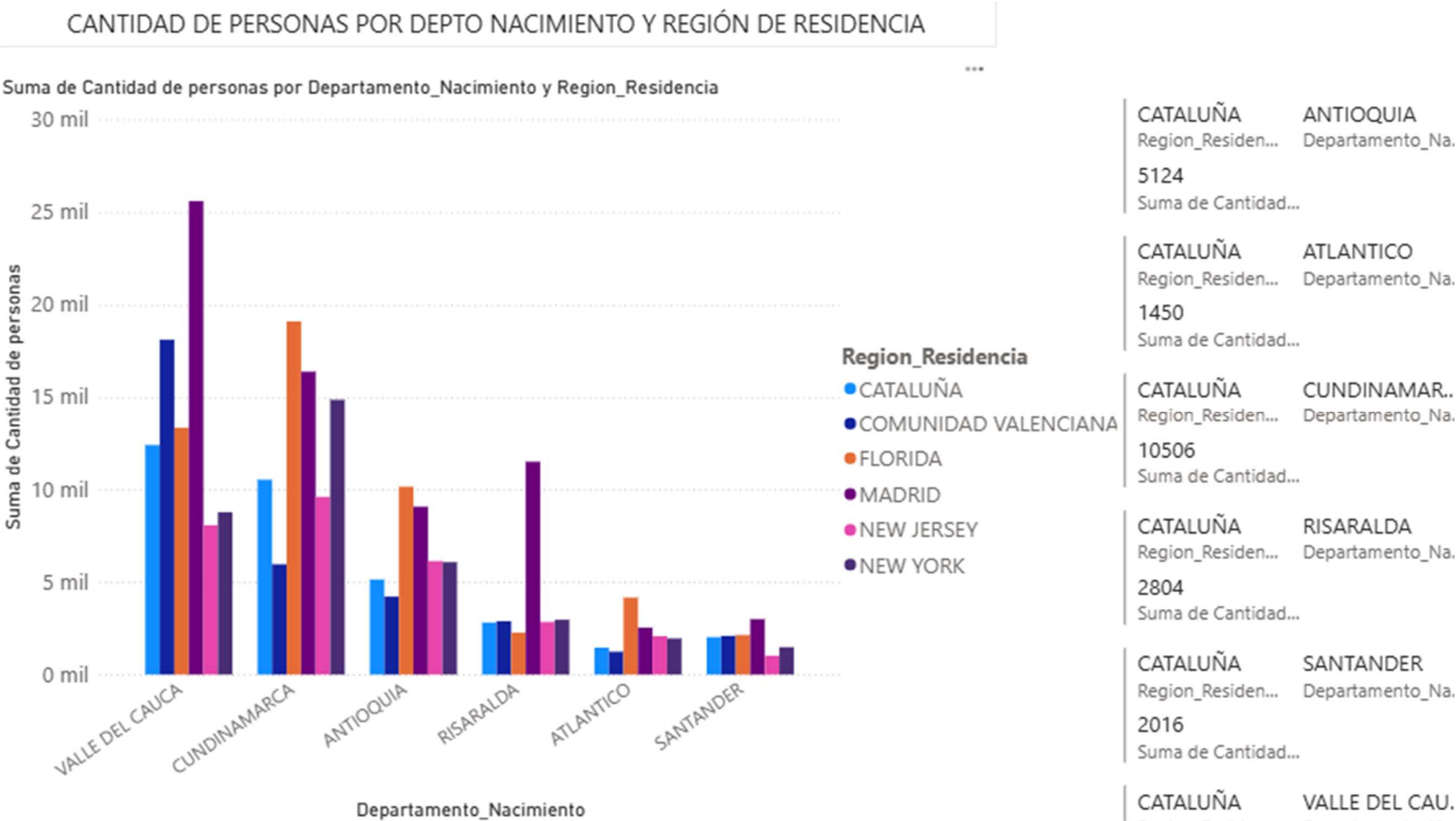


135168	NINGUNO (NO I...
Suma de Canti...	Nivel Académico
ESPAÑA	
País_Residencia	
120933	NINGUNO (NO I...
Suma de Canti...	Nivel Académico
ESTADOS UNIDOS	
País_Residencia	
117003	BACHILLERATO
Suma de Canti...	Nivel Académico
ESPAÑA	
País_Residencia	
81412	BACHILLERATO
Suma de Canti...	Nivel Académico
ESTADOS UNIDOS	
País_Residencia	
50965	UNIVERSITARIO
Suma de Canti...	Nivel Académico
ESTADOS UNIDOS	
País_Residencia	
30298	NINGUNO (NO I...
Suma de Canti...	Nivel Académico
ECUADOR	

### 8.13. ¿Del total de emigrantes, que porcentaje son Afrodescendientes?



8.14. ¿Hay alguna tendencia entre el departamento de nacimiento y la región de residencia?



8.15. ¿De qué ciudades y que cantidad de colombianos fueron registrados en el año 2024 (Top 3)?

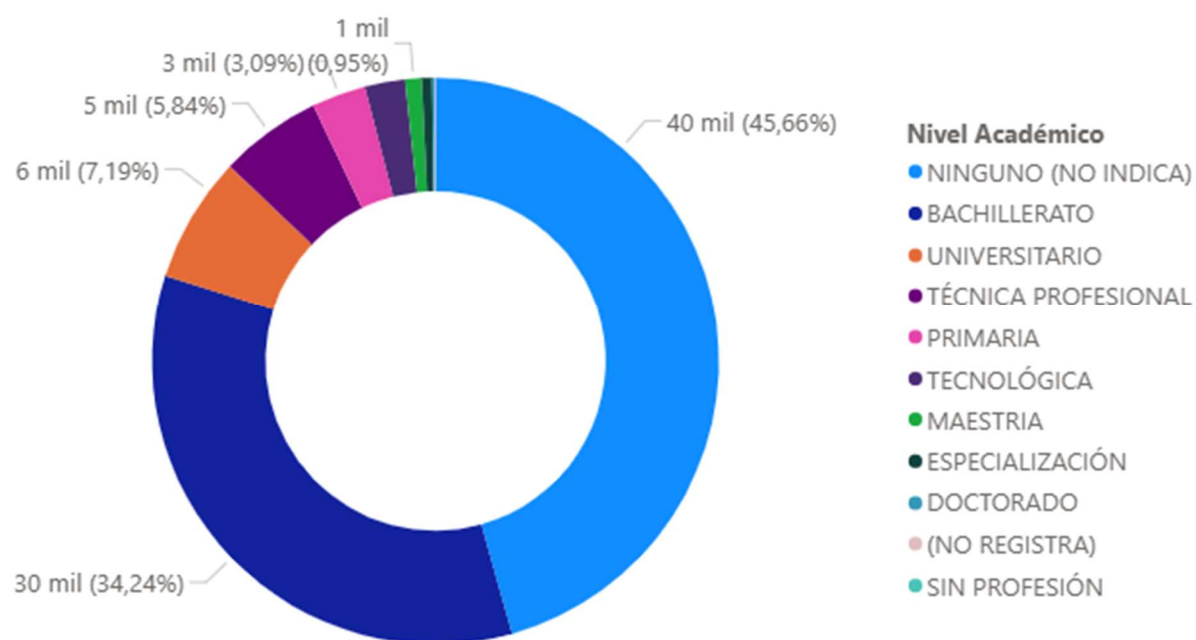


En el cierre del año 2024 se registraron en el exterior, 23.400 personas provenientes de la ciudad de Bogotá, 13.400 provenientes de la ciudad de Cali y 8.400 provenientes de la ciudad de Medellín. Según la página oficial de Telencuestas (<https://telencuestas.com/>), las siguientes cifras se registran según información emitida por el DANE (cifras de proyección). Habitantes de la ciudad de Bogotá entre edades de los 18 y 54 años en el 2025: 4.561.821. Habitantes de la ciudad de Cali entre edades de los 18 y 54 años en el 2025: 1.194.692. Habitantes de la ciudad de Medellín entre edades de los 18 y 54 años en el 2025: 1.465.469.

## 8.16. ¿Cómo se comporta el nivel académico de los emigrantes provenientes del Valle del Cauca?

CANTIDAD DE PERSONAS EN EL VALLE DEL CAUCA Y NIVEL ACADÉMICO

Suma de Cantidad de personas por Nivel Académico

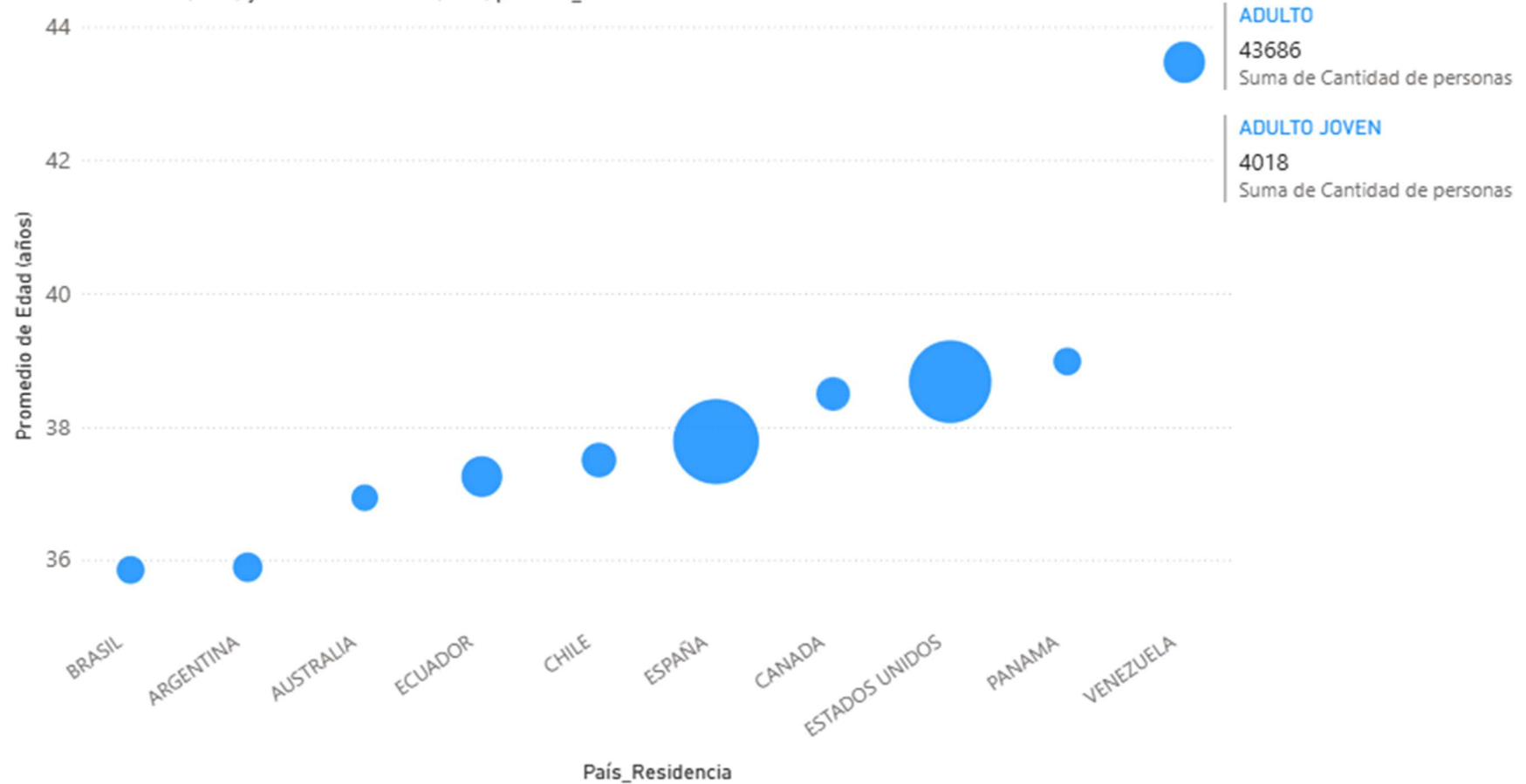


VALLE DEL CAU...	113348
Departamento_Na...	Suma de Cantidad de.
NINGUNO (NO IN...	
Nivel Académico	
VALLE DEL CAU...	76257
Departamento_Na...	Suma de Cantidad de.
BACHILLERATO	
Nivel Académico	
VALLE DEL CAU...	21231
Departamento_Na...	Suma de Cantidad de.
UNIVERSITARIO	
Nivel Académico	
VALLE DEL CAU...	13235
Departamento_Na...	Suma de Cantidad de.
TÉCNICA PROFESI...	
Nivel Académico	
VALLE DEL CAU...	8353
Departamento_Na...	Suma de Cantidad de.
PRIMARIA	
Nivel Académico	
VALLE DEL CAU...	5748
Departamento_Na...	Suma de Cantidad de.
SIN PROFESIÓN	
Nivel Académico	

## 8.17. ¿Cuáles son los países con promedio de edad más bajo?

### CANTIDAD DE PERSONAS POR PAÍS DE RESIDENCIA Y GÉNERO

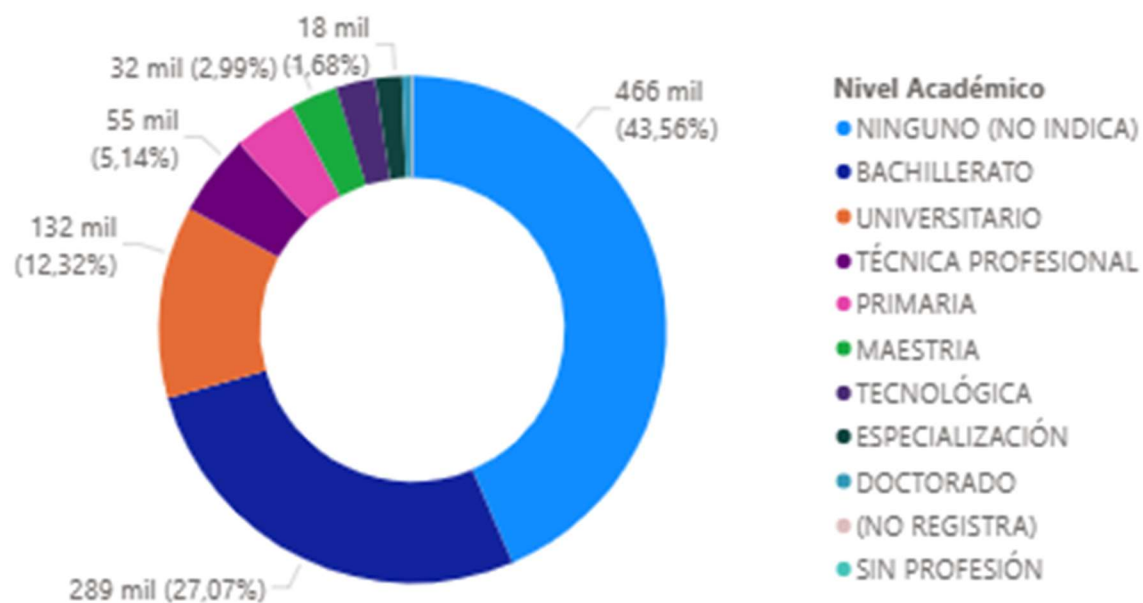
Promedio de Edad (años) y Recuento de Edad (años) por País\_Residencia





## 8.18. ¿Qué relación posee el nivel académico en función de la edad?

Suma de Cantidad de personas por Nivel Académico



Grupo edad

ADULTO	ADULTO JOVEN
--------	--------------

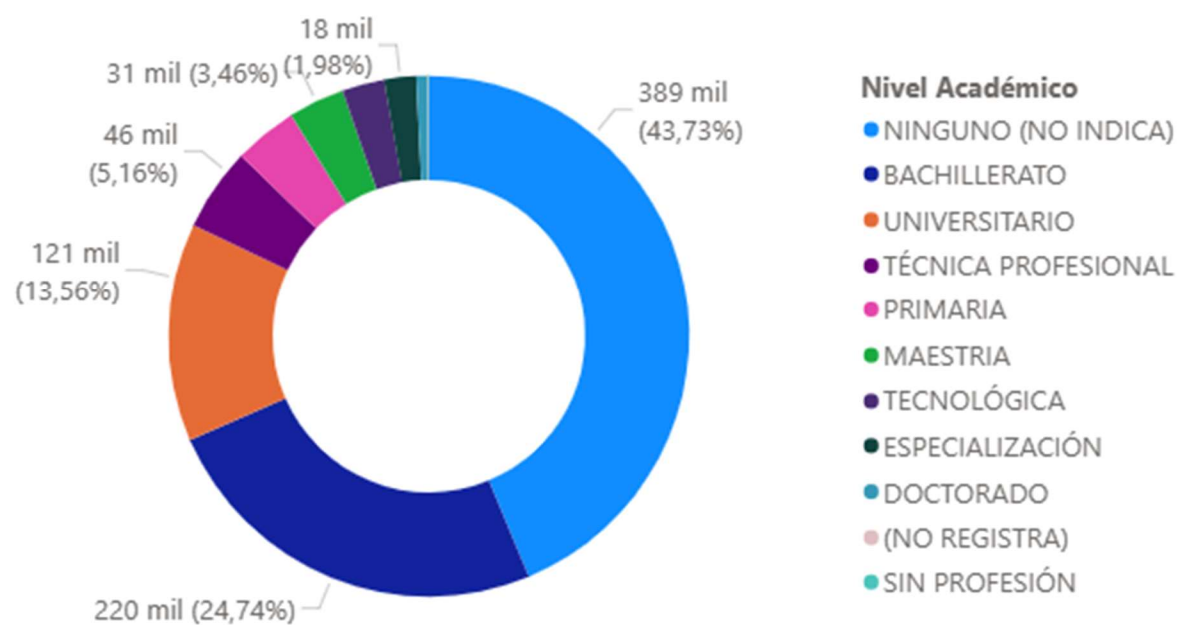
ADULTO JÓVEN : 18 - 28 AÑOS.

ADULTO : 29 - 59 AÑOS.

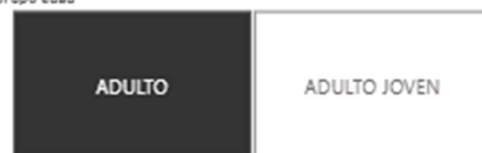
1 mill.

Suma de Cantidad de personas

Suma de Cantidad de personas por Nivel Académico



Grupo edad



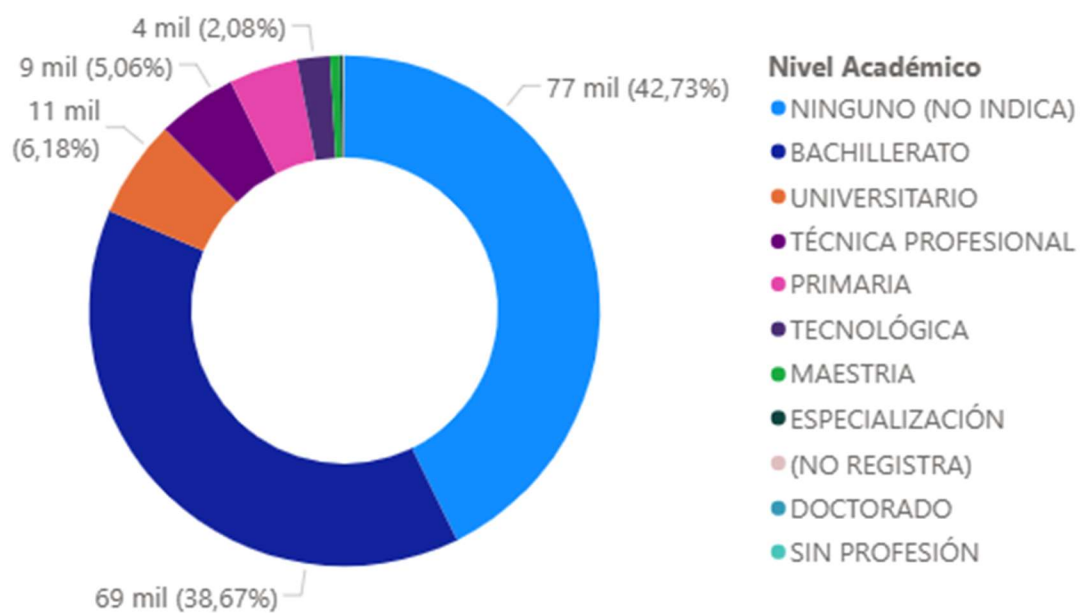
ADULTO JÓVEN : 18 - 28 AÑOS.

ADULTO : 29 - 59 AÑOS.

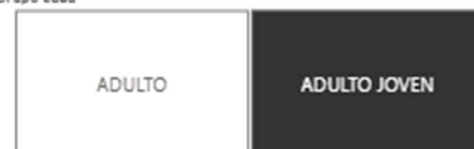
**890 mil**

Suma de Cantidad de personas

Suma de Cantidad de personas por Nivel Académico



Grupo edad



ADULTO JÓVEN : 18 - 28 AÑOS.

ADULTO : 29 - 59 AÑOS.

**179 mil**

Suma de Cantidad de personas

## 9. Conclusiones:

1. Hallazgo relacionado con la pregunta “8.15”.

- **Cali lidera la emigración proporcional (1.12%):**
  - La proporción es más del doble que Bogotá y casi el doble que Medellín.
  - Posibles causas: Crisis laboral, violencia, o mejores oportunidades que las que se brindan en la ciudad o país.
- **Bogotá tiene la menor proporción (0.51%):**
  - A pesar de ser la ciudad más poblada, retiene mejor a su población joven/adulta.
  - Factores: Mayor oferta de empleo, educación superior, o calidad de vida relativa.
- **Medellín se mantiene en un punto intermedio (0.57%):**
  - Equilibrio entre oportunidades locales y emigración.

<b>Ciudad</b>	<b>Emigrantes (18-55)</b>	<b>Población (18-54)</b>	<b>Proporción</b>
Bogotá	23,400	4,561,821	0.51%
Cali	13,400	1,194,692	1.12%
Medellín	8,400	1,465,469	0.57%

Fuente de apoyo para la columna “Población”: <https://telencuestas.com/censos-de-poblacion/colombia/2025/bogota>

2. Uno de nuestros hallazgos se relaciona con la pregunta “8.18”, la cual se relaciona con el nivel académico de los “Adultos Jóvenes” que deciden emigrar. Es un alto porcentaje la cantidad de jóvenes adultos (edades entre los 18 y 28 años) los cuales no tienen títulos académicos.

3. Enfoque en Edades Productivas y Nivel Académico: Dado que el análisis se centró en personas entre 18 y 55 años y se observó la distribución por nivel académico y área de conocimiento, pueden existir políticas o campañas que se dirijan a retener talento o crear oportunidades para profesionales y técnicos en Colombia.

• **¿De qué manera se pueden implementar campañas inclinadas hacia el bienestar social, referente a mitigar la emigración de los colombianos entre los 18 y los 55 años?**

1. Se considera como punto principal, el fortalecer campañas de “empalme” entre los últimos años del bachiller, de manera que los jóvenes tengan una visión más amplia de los entornos académicos relacionados con las instituciones universitarias.

2. Se podrían ofrecer proyectos académicos que incentiven a los adultos jóvenes a adquirir un título universitario para volverse competitivo en el entorno laboral. Las campañas deberían focalizarse (sin excluir a las personas que residan en el extranjero) a las personas que residan en Colombia.
3. Brindar facilidades de acceso a los consulados, e incluir el motivo de emigración de los colombianos que se registren. De esta manera se pueden brindar soluciones enfocadas directamente a la causa del problema o motivo de la emigración.
4. Dado a las altas cifras de emigración, se sugiere de la mano de profesionales en estudios relacionados con lo social, analizar las problemáticas e impactos que puedan afectar al país debido a esta alza de emigrantes colombianos (puede ser en lo social, en lo económico, en lo político, en lo educativo, etc), al igual que las problemáticas que puedan surgir en países como España y Estados Unidos, los cuales son los destinos mas preferidos por los colombianos.
5. Finalmente, y siendo una de las principales conclusiones se necesita de manera urgente implementar, es establecer la adquisición de estos datos, para evitar la gran cantidad de datos atípicos, erróneos y nulos. De esta manera se mitiga la labor de limpieza de los datos, y se puede realizar análisis mucho más confiables.

## Bibliografía:

- Fuente de datos: <https://www.datos.gov.co/Estad-sticas-Nacionales/Colombianos-en-el-exterior-de-25-a-40-a-os/3k8j-4d42>
- Información de habitantes en Colombia (Bogotá, Cali y Medellín): <https://telencuestas.com/>
- Material de apoyo: <https://www.youtube.com/watch?v=fXrTDYhmk9Y>
- Material de apoyo: <https://repositorio.uniremington.edu.co/entities/publication/72640588-a86b-4483-93ba-c68f18050325/full>
- Material de apoyo: <https://repositorio.uniremington.edu.co/server/api/core/bitstreams/6a7c4b4d-a874-4f02-ad56-ffed1f132e06/content>
- Material de apoyo: <https://www.dane.gov.co/files/investigaciones/poblacion/informes-estadisticas-sociodemograficas/presentacion-resultados-estudio-postcensal-migraciones-exteriores-colombia.pdf>
- Material de apoyo: [https://www.dane.gov.co/files/investigaciones/publicaciones/publicaciones/merc\\_lab\\_i\\_ba.pdf](https://www.dane.gov.co/files/investigaciones/publicaciones/publicaciones/merc_lab_i_ba.pdf)