

Análisis de mercado **¿Cuántos de los jóvenes en acción decidieron continuar con su proceso formativo (Técnico > Tecnológico > Pregrado)?**

El principal programa de intervención que realiza prosperidad social en los últimos años es jóvenes en acción. En base a ese enunciado inicia el desarrollo de la investigación

De Miguel Gutierrez, Universidad del Tolima, julio de 2023.



Fase ADM-1 Business Question.

Descripción general de las fases del análisis.

Para la realización de este análisis de mercado se establecen las siguientes 5 grandes etapas:

- **Fase ADM-1 Business Question.**
 - ADM-1.1 Definición del tema central.
 - ADM-1.2 Establecer la pregunta sistémica por resolver.
 - ADM-1.3 Proyectar los objetivos a lograr con la investigación.
 - ADM-1.4 Crear el plan de trabajo.
 - ADM-1.5 Investigar los recursos técnicos disponibles para trabajar.
 - ADM-1.6 Asignar los roles de los participantes de la investigación.
- **Fase ADM-2 Data collection.**
 - ADM-2.1 Definir los medios y fuentes para la recolección y obtención de los datos.
 - ADM-2.2 Iniciar la recolección de los datos.
 - ADM-2.3 Realizar la elección de los datos a utilizar.
 - ADM-2.4 Analizar la calidad de los datos.
- **Fase ADM-3 Prepare and explore data.**
 - ADM-3.1 Eliminación de los datos duplicados.
 - ADM-3.2 Imputación o eliminación de los datos nulos o vacíos.
 - ADM-3.3 Tratamiento de los datos atípicos.
 - ADM-3.4 Normalización y estandarización del conjunto de datos.
 - ADM-3.5 Codificación de variables categóricas a cuantitativas.
 - ADM-3.6 Selección de categorías de mayor importancia para el análisis.
- **Fase ADM-4 Data analysis.**
 - ADM-4.1 Determinar el nivel de correlación entre variables.
 - ADM-4.2 Explorar las variables estadísticas descriptivas.
 - ADM-4.3 Realizar visualizaciones de los datos.
 - ADM-4.4 Evolución de las variables en series de tiempo.
- **Fase ADM-5 Present Findings.**
 - ADM-5.1 Creación del contexto para el lector de la información.
 - ADM-5.2 Enseñar los hallazgos claves.
 - ADM-5.2 Enseñar los hallazgos incidentales.

Fase ADM-1 Business Question.

Descripción:

Primera etapa del análisis de mercado, que se estructura y se definirá los siguientes objetivos:

- ADM-1.1 Definición del tema central.
- ADM-1.2 Establecer la pregunta sistémica por resolver.
- ADM-1.3 Proyectar los objetivos a lograr con la investigación.
- ADM-1.4 Crear el plan de trabajo.
- ADM-1.5 Investigar los recursos técnicos disponibles para trabajar:
 - Selección de las tecnologías a usar.
 - Fuentes de datos previas para utilizar.
 - Qué nivel de acceso a la información se puede obtener.
 - El volumen de datos a manejar.
- ADM-1.6 Asignar los roles de los participantes de la investigación.
 - Stakeholder (Cliente).
 - Data analyst (Administrador 1)
 - Product owner (Administrador 2).

ADM-1.1 Definición del tema central.

Descripción:

Establecer una temática principal, permite crear de forma practica la pregunta causal de la investigación, por ende, se realiza la preselección de los siguientes temas como posibles opciones para elegir:

- Inclusión social y reconciliación.
- Tecnología e información.
- Ambiente y desarrollo sostenible
- Finanzas y economía.

Para realizar la elección se tiene en cuenta los parámetros, requerimientos, obligaciones y restricciones que se deben cumplir en la realización de esta actividad académica por lo cual se define como el tema central:

Inclusión social y reconciliación.

ADM-1.2 Establecer la pregunta sistémica por resolver.

Descripción:

Teniendo la inclusión social y reconciliación como tema central se puede empezar a definir las bases de la pregunta sistémica, entendiendo que el tema elegido tiene gran amplitud de frentes, se decide generar la siguiente lista de subcategorías a elegir:

- Víctimas del conflicto armado.
- Adultos mayores.
- Niños.
- Adolescentes.
- Equidad de género.

Se decide escoger **Adolescentes**.

El principal programa de intervención que realiza prosperidad social en los últimos años es jóvenes en acción. En base a es enunciado se genera la siguiente pregunta:

¿Cuántos de los beneficiarios decidieron continuar con su proceso formativo (Técnico > Tecnológico > Pregrado)?

ADM-1.3 Proyectar los objetivos a lograr con la investigación.

Ficha Técnica.

1. Título de la Investigación:

¿Cuántos de los beneficiarios decidieron continuar con su proceso formativo (técnico > tecnológico > Pregrado)?

2. Objetivos de la Investigación:

- Evaluar el impacto de jóvenes en acción en la continuidad académica de sus beneficiarios.
- Identificar los factores que influyen en la decisión de continuar sus estudios.

ADM-1.3 Proyectar los objetivos a lograr con la investigación.

2. Objetivos de la Investigación:

- Determinar las percepciones y experiencias de los beneficiarios con respecto al subsidio.
- Que tanto conocimiento tienen los beneficiarios sobre los alcances del programa.

3. Metodología:

- **Diseño de la investigación:** Estudio descriptivo y correlacional.
- **Población objetivo:** Beneficiarios de jóvenes en acción.
- **Muestra:** Selección aleatoria de beneficiarios.
- **Técnicas de recolección de datos:** Encuestas y entrevistas.
- **Variables para medir:** Continuidad académica, percepciones y experiencias del subsidio, factores de influencia.
- **Análisis de datos:** Estadístico descriptivo, correlacional y cualitativo de las bases de datos públicos de prosperidad social y de las obtenidas en las encuestas y entrevistas.

4. Procedimiento:

- Obtención de permisos y aprobaciones éticas necesarias.
- Diseño de los instrumentos de recolección de datos (cuestionario y guía de entrevistas).
- Selección y reclutamiento de la muestra de beneficiarios.
- Aplicación de los cuestionarios y realización de las entrevistas.
- Registro y codificación de los datos recolectados.
- Análisis de los datos utilizando técnicas estadísticas y cualitativas.
- Interpretación de los resultados y formulación de conclusiones.
- Elaboración del informe final de la investigación.

ADM-1.3 Proyectar los objetivos a lograr con la investigación.

5. Cronograma:

- **Fase 1 - Business Question** (2 semanas).
- **Fase 2 - Data collection** (2 semanas).
- **Fase 3 - Prepare and explore data** (2 semanas).
- **Fase 4 - Data análisis** (2 semanas).
- **Fase 5 - Present Findings** (2 semanas).

6. Recursos:

- **Personal:** Analista de datos, encuestador.
- **Materiales:** Encuesta digital, bases de datos públicas.
- **Tecnología:** Software de análisis de datos (Python y Power BI), herramientas de presentación (Power Point).

7. Conceptualización:

Contextualizar a los encuestados de los que es jóvenes en acción, cual es el objetivo del programa y cuáles son los alcances ya sea en tiempo o en recursos económicos.

8. Limitaciones y Consideraciones Éticas:

- Posibles sesgos de respuesta por parte de los participantes.
- Respetar la privacidad y confidencialidad de los participantes.
- Garantizar el consentimiento informado de los participantes.
- Mantener la confidencialidad de los datos recolectados.

ADM-1.4 Crear el plan de trabajo.

Descripción:

Establecer el objetivo del programa y cuáles son los alcances en el tiempo dispuesto (1 semestre).

Enlace del plan de trabajo a ejecutar: [Cronograma de trabajo.](#)

ADM-1.4 Crear el plan de trabajo.

Elementos	ABR	MAYO	JUN
<div><div>ADM-1 Business Question</div><div><div>ADM-2 Definición del tema central.</div><div>ADM-7 Establecer la pregunta sistémica por resolver.</div><div>ADM-8 Proyectar los objetivos a lograr con la investigación.</div><div>ADM-9 Crear el plan de trabajo.</div><div>ADM-10 Investigar los recursos técnicos disponibles para trabajar.</div><div>ADM-11 Asignar los roles de los participantes de la investigación.</div></div></div>			
<div><div>ADM-3 Data collection</div><div><div>ADM-13 Definir los medios y fuentes para la recolección y obtención de los datos.</div><div>ADM-14 Iniciar la recolección de los datos.</div><div>ADM-15 Realizar la elección de los datos a utilizar.</div><div>ADM-16 Analizar la calidad de los datos.</div></div></div>			
<div><div>ADM-4 Prepare and explore data</div><div><div>ADM-17 Eliminación de los datos duplicados.</div><div>ADM-18 Imputación o eliminación de los datos nulos o vacíos.</div><div>ADM-19 Tratamiento de los datos atípicos.</div><div>ADM-20 Normalización y estandarización del conjunto de datos.</div><div>ADM-21 Codificación de variables categóricas a cualitativas.</div><div>ADM-22 Selección de categorías de mayor importancia para el análisis.</div></div></div>			
<div><div>ADM-5 Data analysis</div><div><div>ADM-23 Determinar el nivel de correlación entre variables.</div><div>ADM-24 Explorar las variables estadísticas descriptivas.</div><div>ADM-25 Realizar visualizaciones de los datos.</div><div>ADM-26 Evolución de las variables en series de tiempo.</div></div></div>			
<div><div>ADM-6 Present Findings</div><div><div>ADM-27 Creación del contexto para el lector de la información.</div><div>ADM-28 Enseñar los hallazgos claves.</div><div>ADM-29 Establecer los resultados logrados.</div><div>ADM-30 Definir la conclusiones y recomendaciones.</div></div></div>			

ADM-1.5 Investigar los recursos técnicos disponibles para trabajar.

Selección de las tecnologías a usar:

- Python.
- Power bi.

Fuentes de datos previas para utilizar:

- Datos públicos de la página de datos.gov.co.

Qué nivel de acceso a los datos se puede obtener:

- Fuentes abiertas (Directo).
- Encuestas (Directo, Virtual).
- Entrevistas (Directo, Virtual).

El volumen de datos a manejar:

- Fuentes abiertas (Base de datos con 260.000 registros).
- Encuestas (Por definir).
- Entrevistas (Por definir).

ADM-1.6 Asignar los roles de los participantes de la investigación.

Descripción:

Se establecerán los siguientes roles a las siguientes personas:

- Stakeholders (Julio Loaiza).
- Data analyst (Miguel Gutierrez).
- Product owner (Mardubil Rojas).



Fase ADM-2

Data collection

Fase ADM-2 Data collection

Descripción:

Primera etapa del análisis de mercado, que se estructura y se definirá los siguientes objetivos:

- ADM-2.1 Definir los medios y fuentes para la recolección y obtención de los datos.
- ADM-2.2 Iniciar la recolección de los datos.
- ADM-2.3 Realizar la elección de los datos a utilizar.
- ADM-2.4 Analizar la calidad de los datos.

ADM-2.1 Definir los medios y fuentes para la recolección y obtención de los datos.

Fuentes cerradas (Directas):

- [Encuesta.](#)

Sección 1 de 3

Análisis: Evolución del nivel académico

Somos estudiantes de la universidad del Tolima realizando una investigación con el objetivo de encontrar si los beneficiarios que reciben o recibieron **jóvenes en acción** han continuado con su formación académica (**técnica > tecnológica > pregrado**), cuales son sus motivaciones o sus limitaciones para continuar con el proceso formativo.

1. Correo: *

Texto de respuesta corta

2. Nombre:

Texto de respuesta corta

3. Apellido:

Texto de respuesta corta

ADM-2.1 Definir los medios y fuentes para la recolección y obtención de los datos.

4. ¿En que rango de edad se encuentra? *

- 1. 16 - 18
- 2. 19 - 21
- 3. 22 - 25
- 4. 26 - 28

5. Departamento de residencia:

Texto de respuesta corta

Sección 2 de 3

Datos relacionados a Jóvenes en acción



Descripción (opcional)

1. ¿Es o alguna vez fue usted beneficiario de **jóvenes en acción**? *

- ☐ Sí. Actualmente soy beneficiario
- ☐ Sí. Alguna vez fui beneficiario
- ☐ No. Nunca he sido beneficiario

2. ¿Durante cuantos años a recibido jóvenes en acción? *

- ☐ 1
- ☐ 2
- ☐ 4
- ☐ 5

Estado "en transición".

En jóvenes en acción existen múltiples "estados" que puede tener un participante, los mas comunes son el "inscrito" o "beneficiario", en este caso el estado "en transición" es un periodo de gracia de 1 año en el cual un participante que culmine su formación académica pueda inscribirse en un nuevo programa académico de nivel superior al anterior (ejemplo graduarse de un técnico e iniciar un pregrado) y seguir recibiendo el beneficio de jóvenes en acción.

ADM-2.1 Definir los medios y fuentes para la recolección y obtención de los datos.

3. ¿Tenia conocimiento sobre la información del enunciado anterior? *

☐ Sí

☐ No

4. ¿Logro usar este beneficio para continuar con su formación académica? *

☐ Sí

☐ No

5. ¿En caso de haber logrado usar este beneficio cual fue el transito académico que realizo? *

☐ Técnico > Tecnología

☐ Técnico > Pregrado

☐ Tecnólogo > Pregrado

☐ No hice uso de este beneficio

Sección 3 de 3

Información académica

Descripción (opcional)

1. ¿Qué nivel académico estaba cursando cuando empezó a recibir jóvenes en acción? *

☐ Técnico

☐ Tecnología

☐ Pregrado

2. ¿De cual de los siguientes niveles académicos se a graduado? *

☐ Técnico

☐ Tecnología

☐ Pregrado

ADM-2.1 Definir los medios y fuentes para la recolección y obtención de los datos.

3. ¿En este momento estudia, trabaja o ambas? *

- ☐ Estudiante
- ☐ Trabajador
- ☐ Ambas

4. ¿Cómo realizo su primer formación académica? *

- ☐ Presencial
- ☐ A distancia
- ☐ Virtual

5. Si ya trabaja, ¿Cómo continua ahora su formación académica? *

- ☐ Presencial
- ☐ A distancia
- ☐ Virtual
- ☐ Todavía no he tenido mi primer experiencia laboral relacionada con mis estudios

ADM-2.2 Iniciar la recolección de los datos.

Descripción:

Obtenemos la muestra objetivo a lograr en base al total de los datos obtenidos de la base de datos de jóvenes en acción.

Calculadora de muestra

Nivel de confianza: ?

☒ 95% ☐ 99%

Margen de Error: ?

5

Población: ?

260112

Limpiar

Calcular Muestra

Tamaño de Muestra:

385

ADM-2.3 Realizar la elección de los datos a utilizar.

Descripción:

Para determinar que categorías debemos utilizar se realizara un análisis previo de que datos son los que se disponen de las fuentes abiertas, ya que los datos recolectados de manera manual se han definido de forma previa para que cada categoría sea útil y así optimizar el tiempo necesario para su obtención.

- CantidadDeBeneficiarios
- CantidadBeneficioConsolidado
- CantidadUltimobeneficio
- CodigoDepartamentoAtencion
- CodigoMunicipioAtencion
- CondicionSexual
- Discapacidad
- EstadoBeneficiario
- EstadoCivil
- Etnia
- FechaInscripcionBeneficiario
- FechaUltimoBeneficioAsignado
- Genero
- NivelEscolaridad
- NombreDepartamentoAtencion
- NombreMunicipioAtencion
- País
- RangoEdad
- TipoAsignacionBeneficio
- TipoBeneficio
- TipoDocumento
- TipoPoblacion
- ValorBeneficioConsolidadoAsignado
- ValorUltimoBeneficioAsignado

En la base de datos publica podemos encontrar que existe una gran cantidad de categorías, 24 en total, por ende, se decide analizar los estadísticos de cada columna para observar que tan útil seria la información que nos brindaría cada una de estas.

	CondicionSexual	Discapacidad	EstadoBeneficiario	EstadoCivil
count	260112	260112	260112	260112
unique	10	3	2	7
top	ND	NO	NO ACTIVO	Sin Información
freq	258915	174578	158945	249807

	Etnia	FechaInscripcionBeneficiario	FechaUltimoBeneficioAsignado	Genero	NivelEscolaridad
count	260112	260112	260112	260112	260112
unique	9	50	1	3	7
top	ND	2014-07-01	1900-01-01	Mujer	ND
freq	242782	29570	260112	144076	138197

ADM-2.3 Realizar la elección de los datos a utilizar.

	NombreDepartamentoAtencion	NombreMunicipioAtencion	Pais	RangoEdad
count	260112	260112	260112	260112
unique	33	1018	5	5
top	ANTIOQUIA	BOGOTA D.C.	ND	18-29
freq	25672	19476	135841	213376
	TipoAsignacionBeneficio	TipoBeneficio	TipoDocumento	TipoPoblacion
count	260112	260112	260112	260112
unique	1	1	5	5
top	INCENTIVO MONETARIO CONDICIONADO A LA FORMACION	MONETARIO	CC	SISBEN
freq	260112	260112	169793	206694
	ValorBeneficioConsolidadoAsignado		ValorUltimoBeneficioAsignado	
count	260112		260112	
unique	5		4	
top	1.500.001 - 3.000.000		0 - 1.300.000	
freq	77971		257220	

Se puede observar que varias de las variables categóricas tienen como dato top o moda ND (no está disponible) o que por su propia descripción no son relevantes para la investigación.

También datos cuantitativos que contiene valores como códigos de municipio y departamento cuando ya hay columnas con los nombres explícitos, por esto se realiza una nueva base de datos con estas categorías eliminadas y ajustando los nombres de algunas columnas para hacerlos mucho más legibles

Con esto se pasa de tener 24 columnas a 16.

ADM-2.4 Analizar la calidad de los datos.

Descripción:

Se encuentra que la base de datos no contiene registros vacíos, todas las categorías de los tipos de datos son de tipo objeto, se deberá realizar cambios de categorías como en las fechas y los datos cuantitativos para facilitar el análisis posterior.

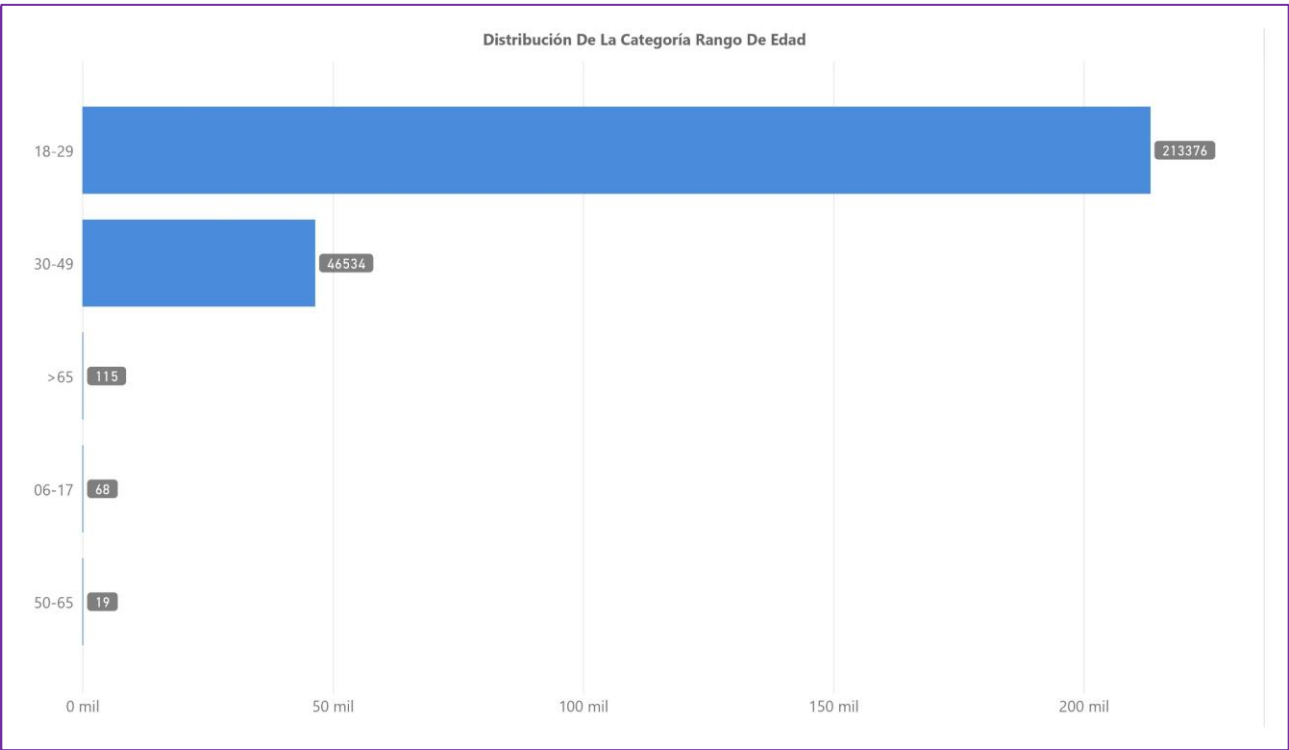
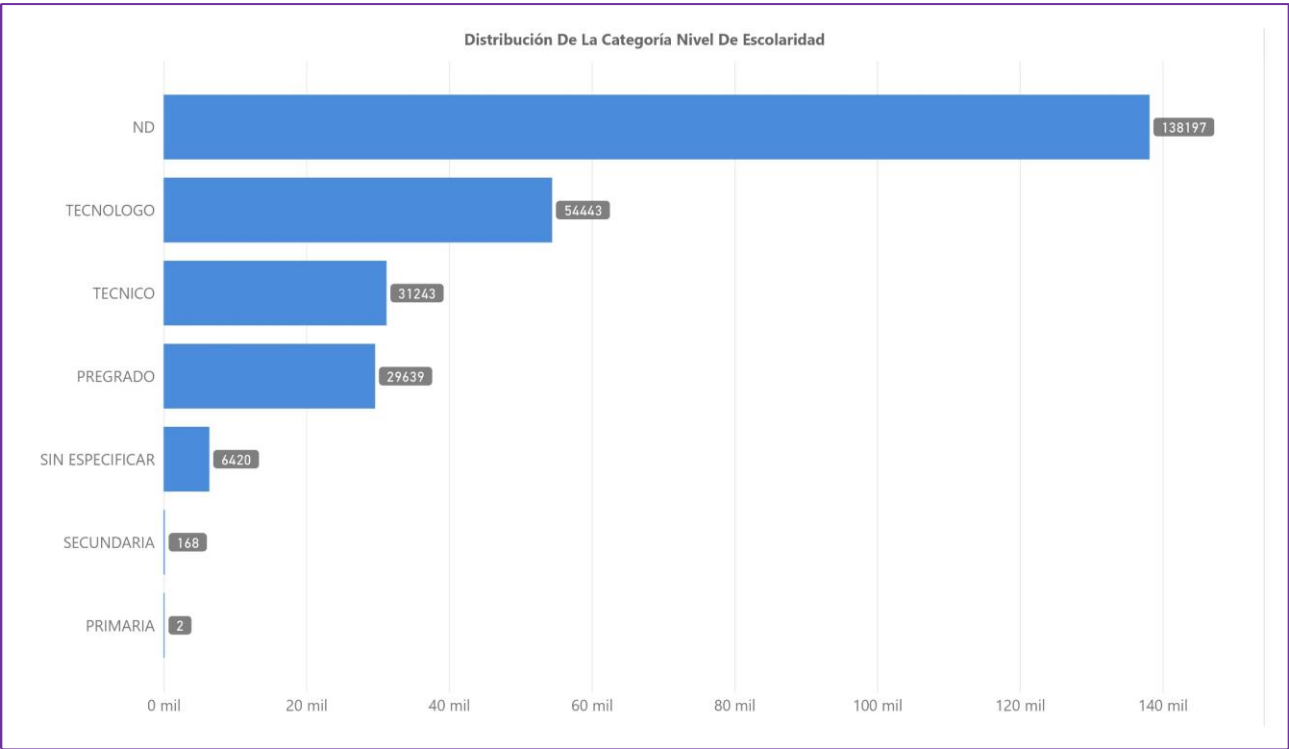
#	Column	Non-Null Count	Dtype
0	Discapacidad	260112 non-null	object
1	Estado	260112 non-null	object
2	Estado Civil	260112 non-null	object
3	Etnia	260112 non-null	object
4	Fecha Inicial	260112 non-null	object
5	Fecha Final	260112 non-null	object
6	Genero	260112 non-null	object
7	N. Escolaridad	260112 non-null	object
8	Departamento	260112 non-null	object
9	Municipio	260112 non-null	object
10	Rango Edad	260112 non-null	object
11	T. Asignacion	260112 non-null	object
12	T. Beneficio	260112 non-null	object
13	T. Documento	260112 non-null	object
14	T. Poblacion	260112 non-null	object
15	V. Total Asignado	260112 non-null	object
16	V. Ult. Asignado	260112 non-null	object

A estas columnas se agregarán 3 nuevas categorías relacionadas con la fecha que serán el año el número de mes y el nombre de mes.

Continuando con el análisis de los datos, se analizará la distribución de cada categoría (columna) para saber cómo se clasifican cada uno de los registros que estas contienen.

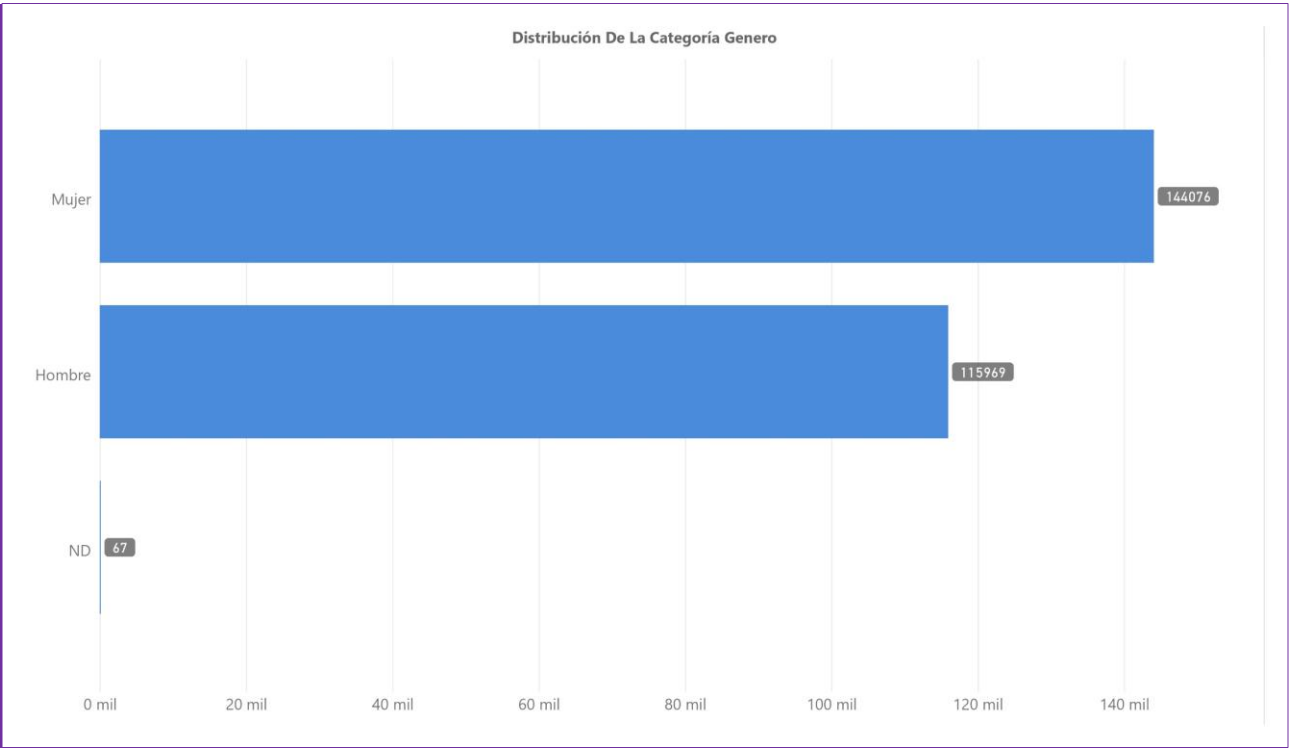
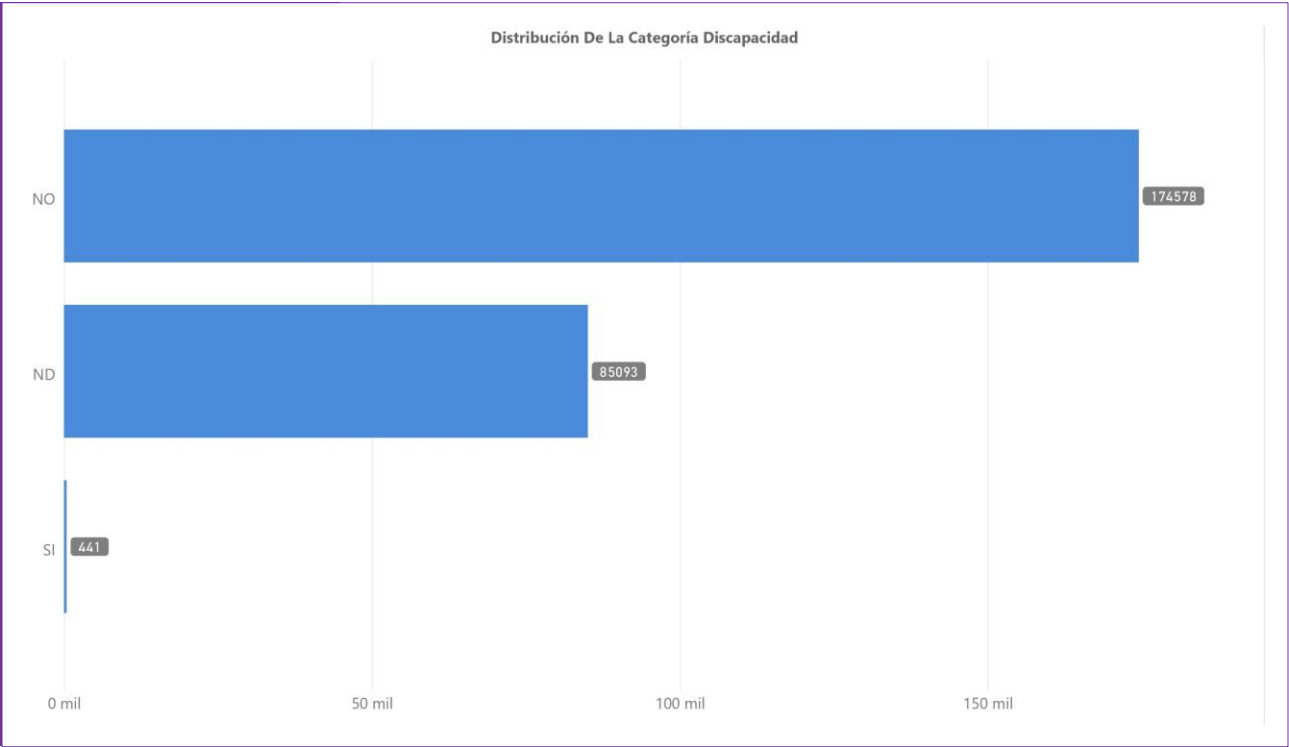
ADM-2.4 Analizar la calidad de los datos.

Se puede la distribución de los datos del nivel de escolaridad y el rango de edad, resaltando la gran cantidad de datos ND encontrados en el nivel de escolaridad.



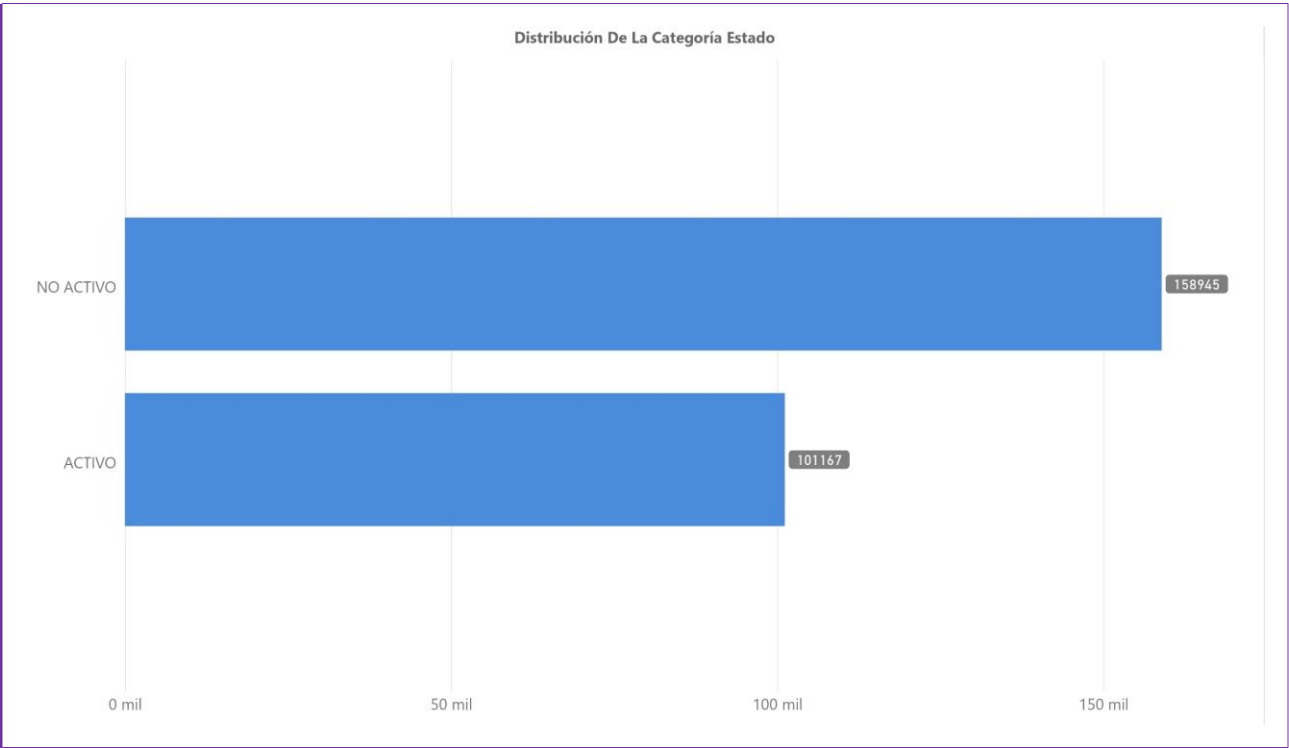
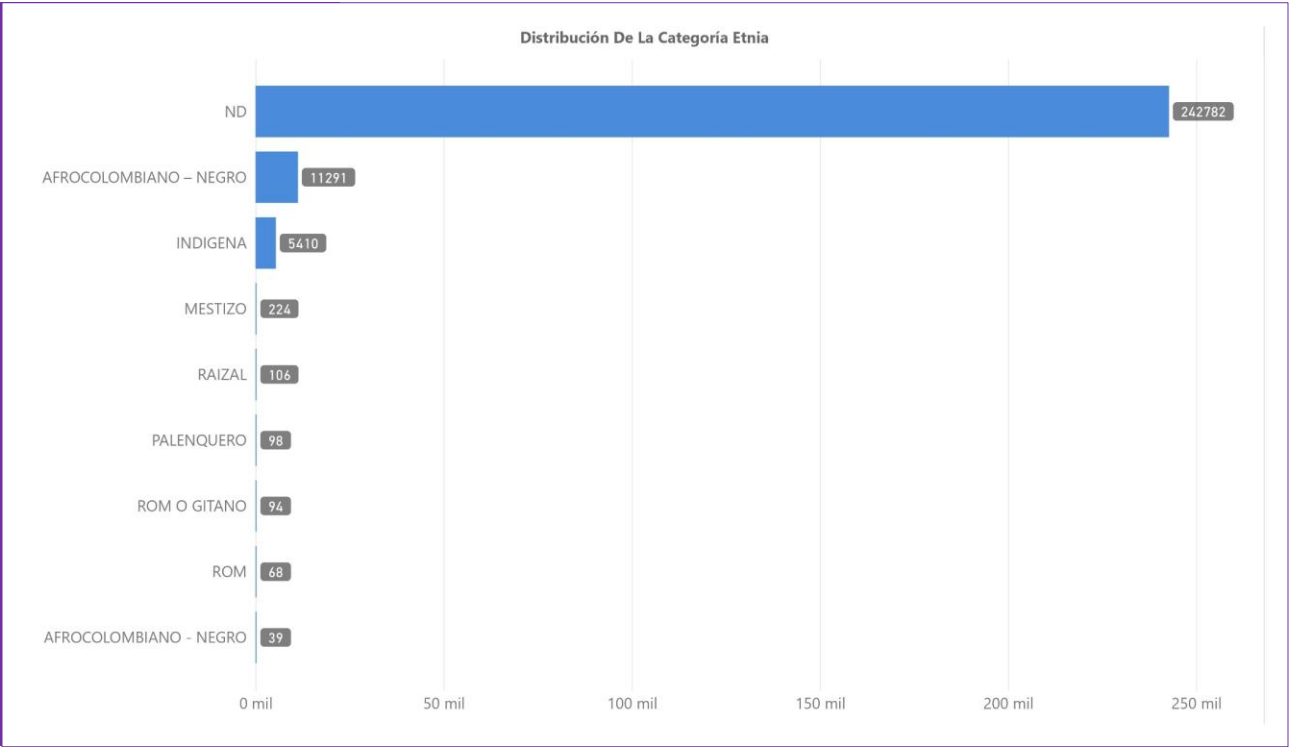
ADM-2.4 Analizar la calidad de los datos.

Nuevamente el tipo de registros ND toma una parte significativa en algunas de las categorías como lo es el caso de “Discapacidad”, esto probablemente producto de una mala recolección.



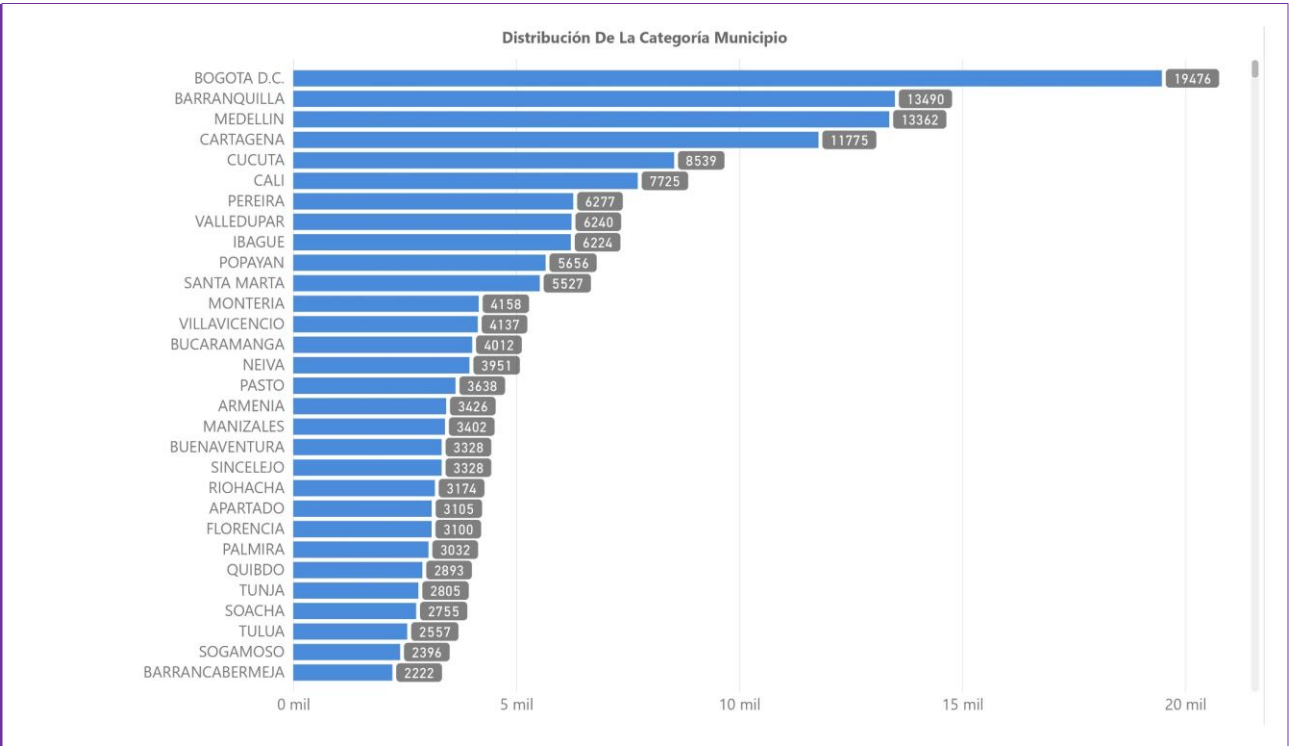
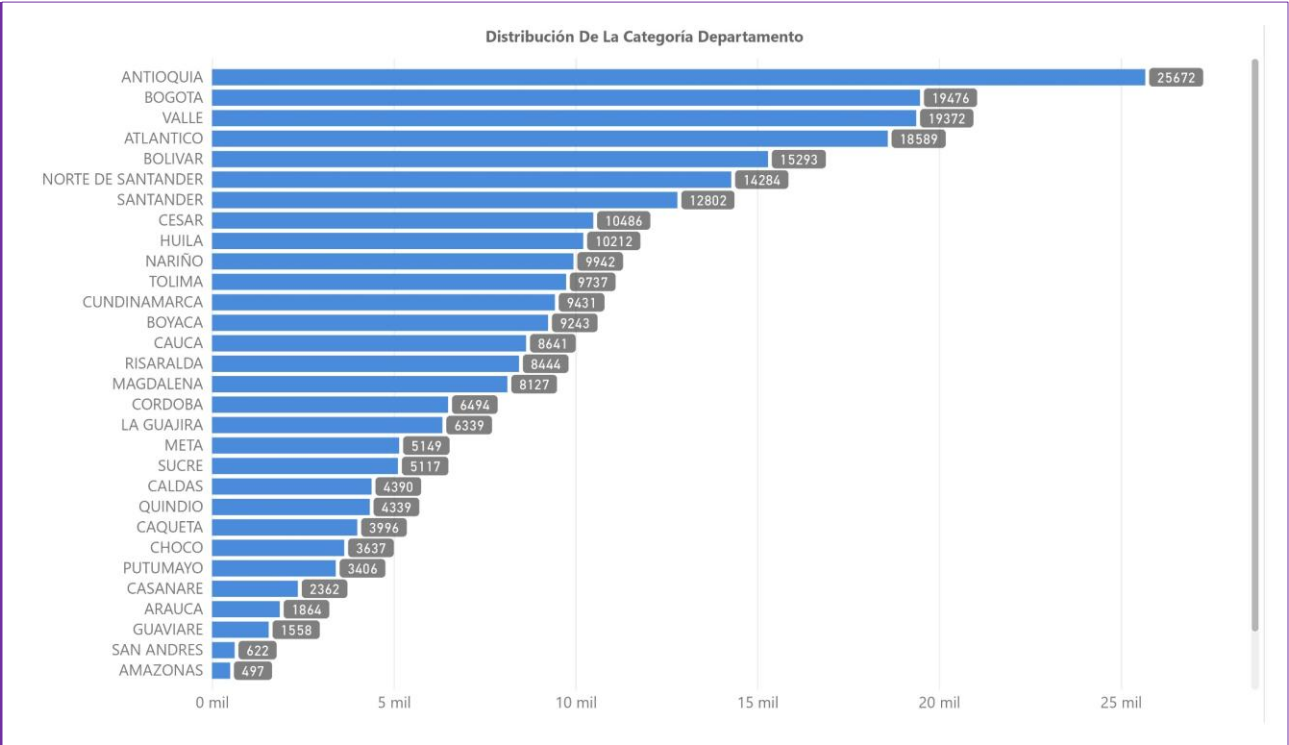
ADM-2.4 Analizar la calidad de los datos.

Parece un patrón recurrente que en cierto tipo de categorías el dato ND se hace presente, se puede partir del supuesto que la recolección de datos en años anteriores no fue muy buena



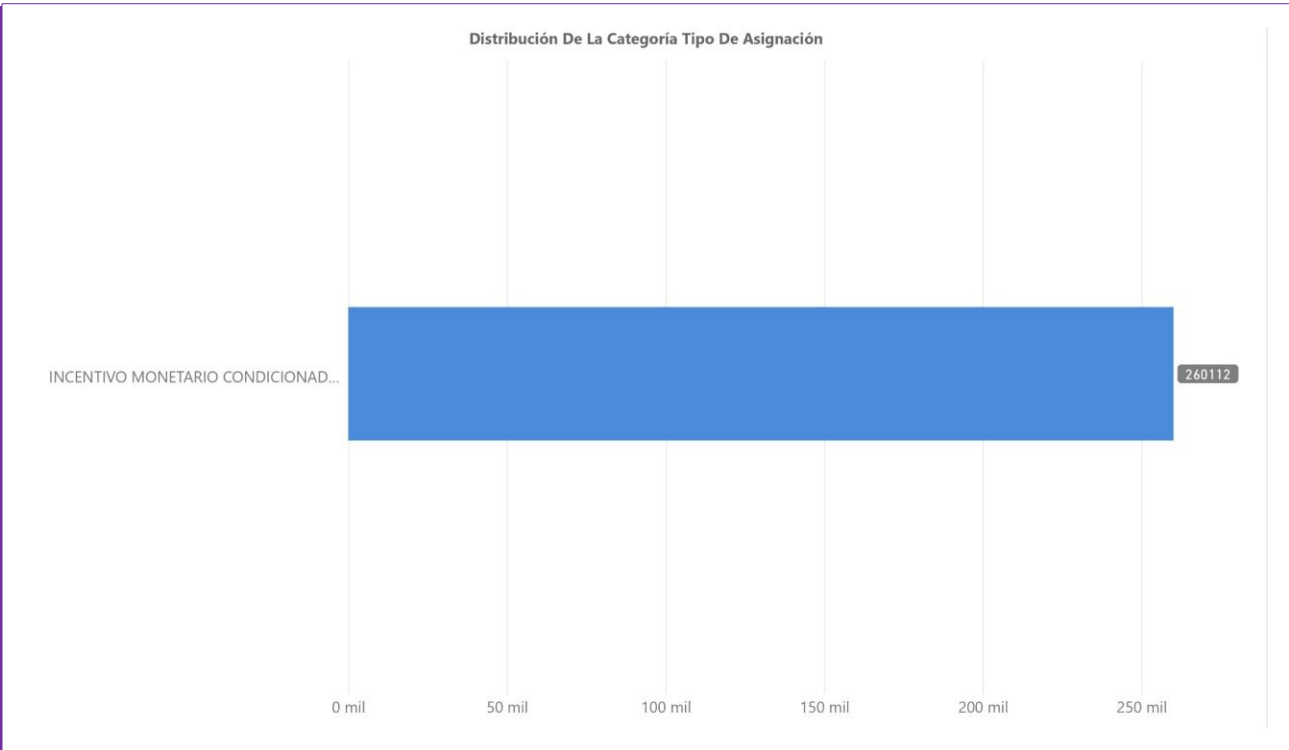
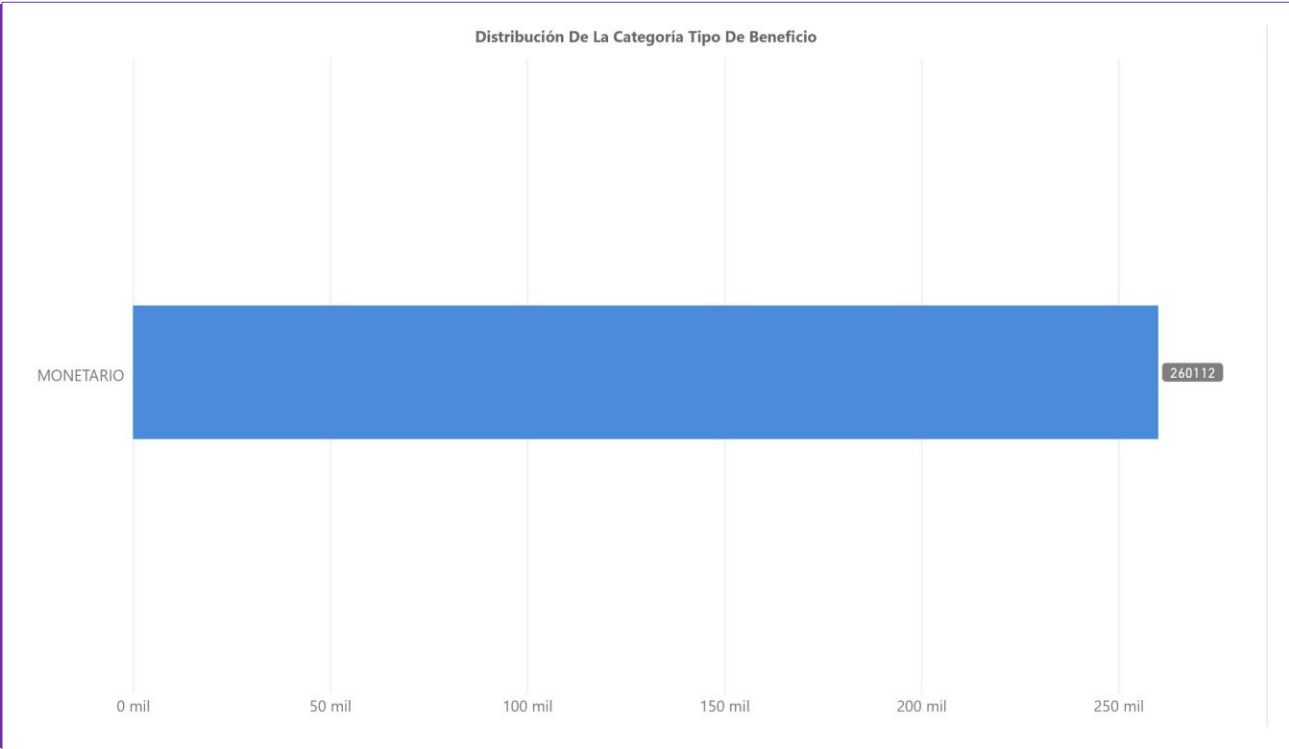
ADM-2.4 Analizar la calidad de los datos.

A nivel geográfico no se mantiene la misma proporción entre los departamentos con mayor número de beneficiarios y sus municipios.



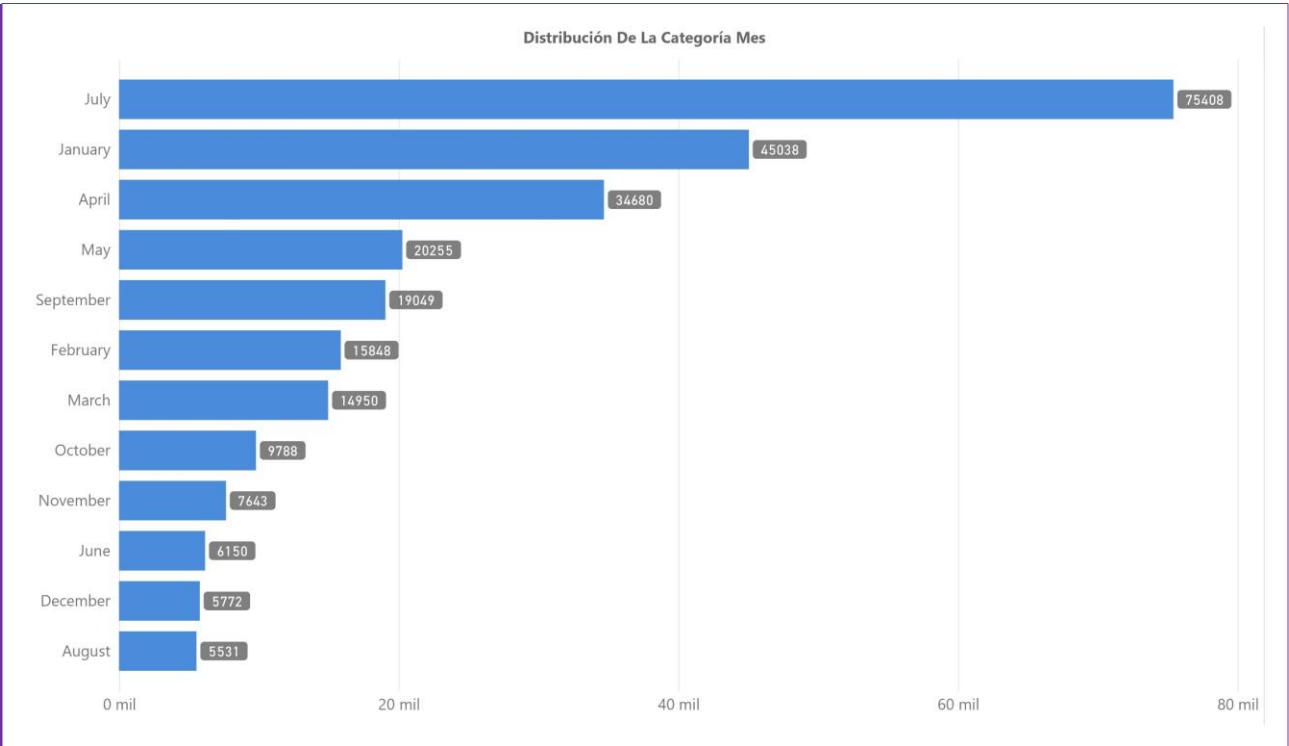
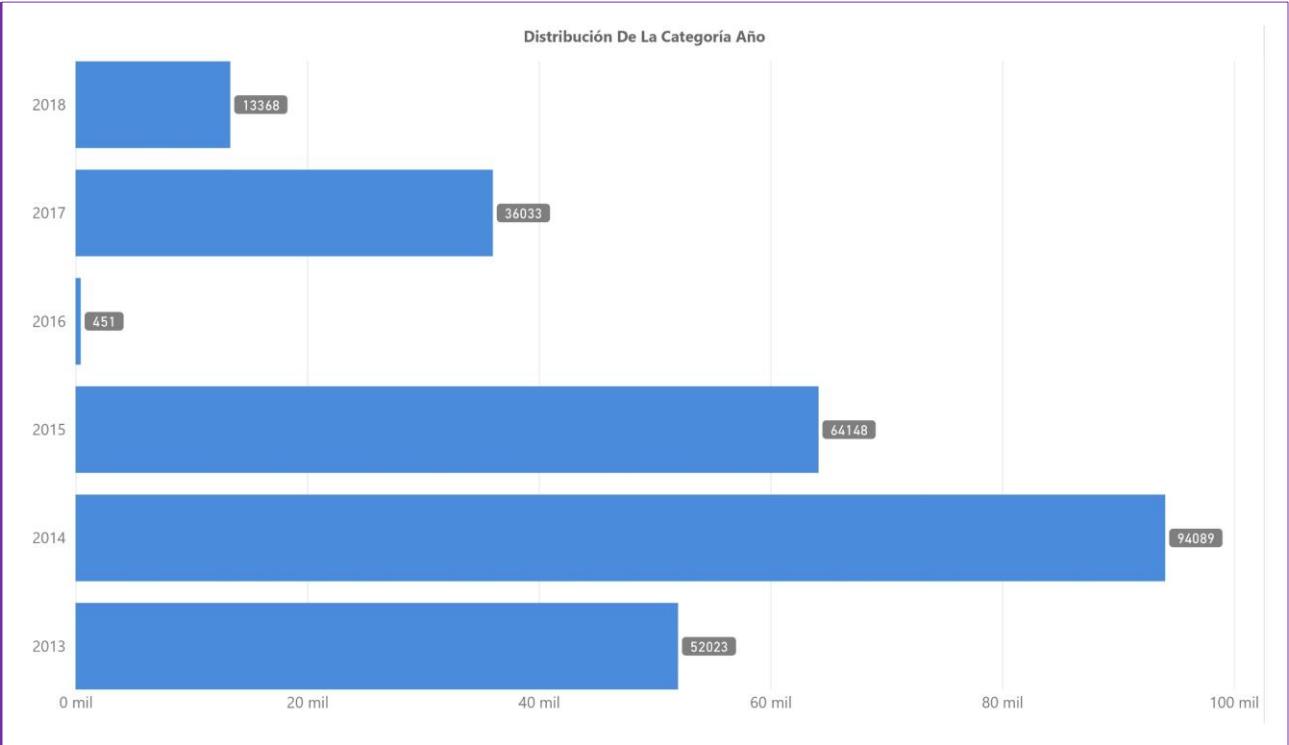
ADM-2.4 Analizar la calidad de los datos.

Estas dos categorías resultan poco útiles para el análisis debido que solo tienen un tipo de registro por lo cual deberán ser eliminadas de la base de datos.



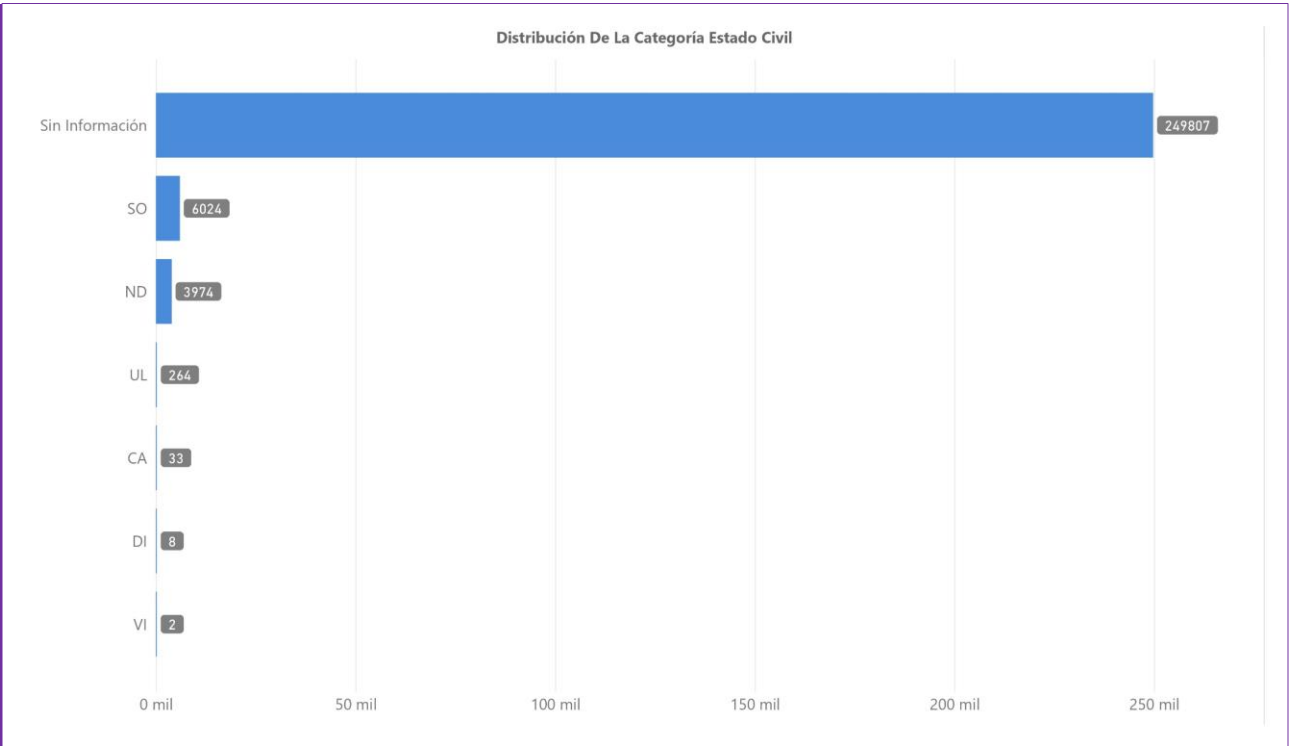
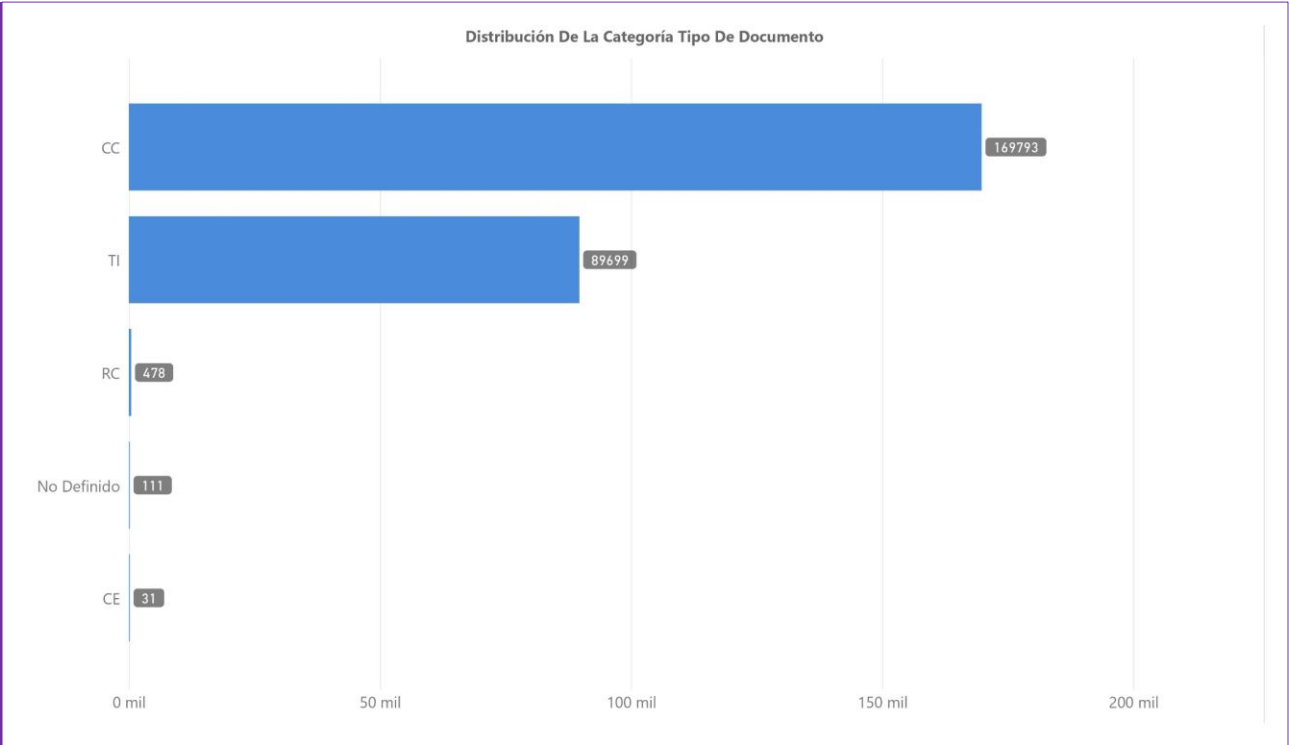
ADM-2.4 Analizar la calidad de los datos.

En la base de datos apenas si hay registros del año 2016, no se sabe cuál sea la justificación de esta ausencia, también se ve que julio es el mes donde más se inscriben beneficiarios.



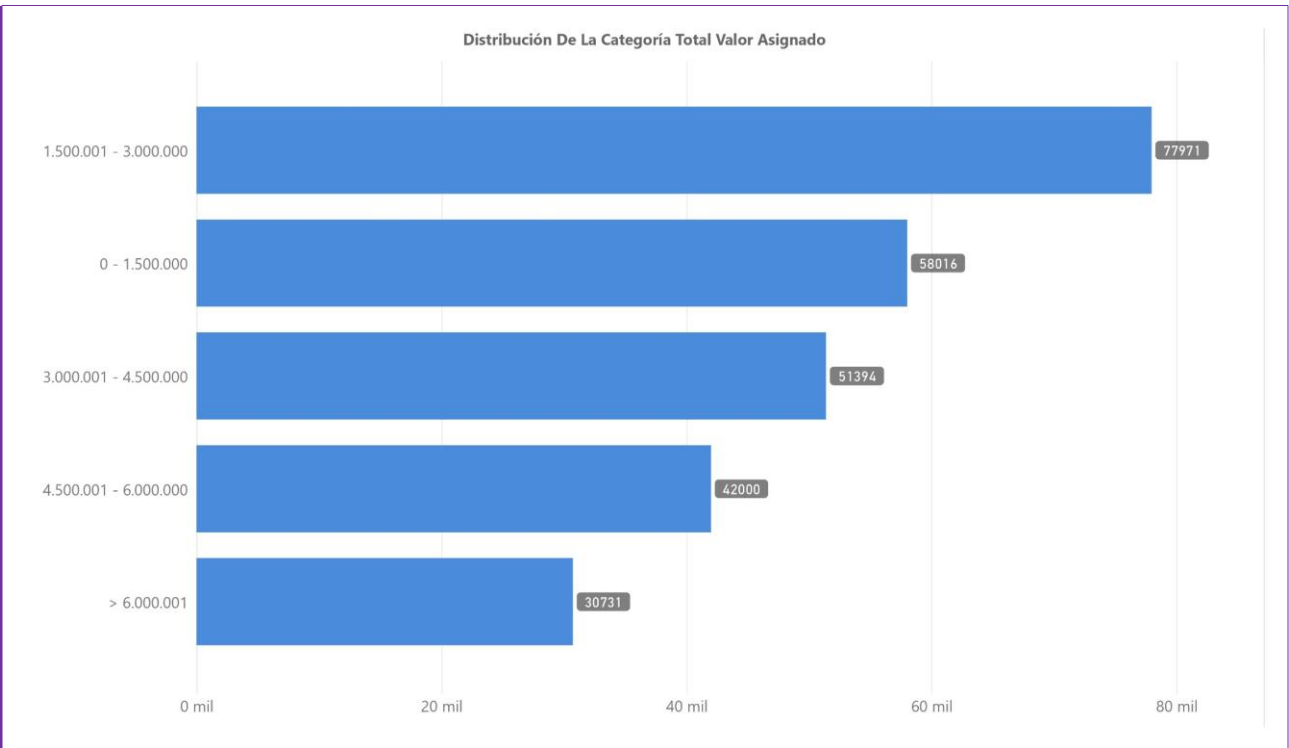
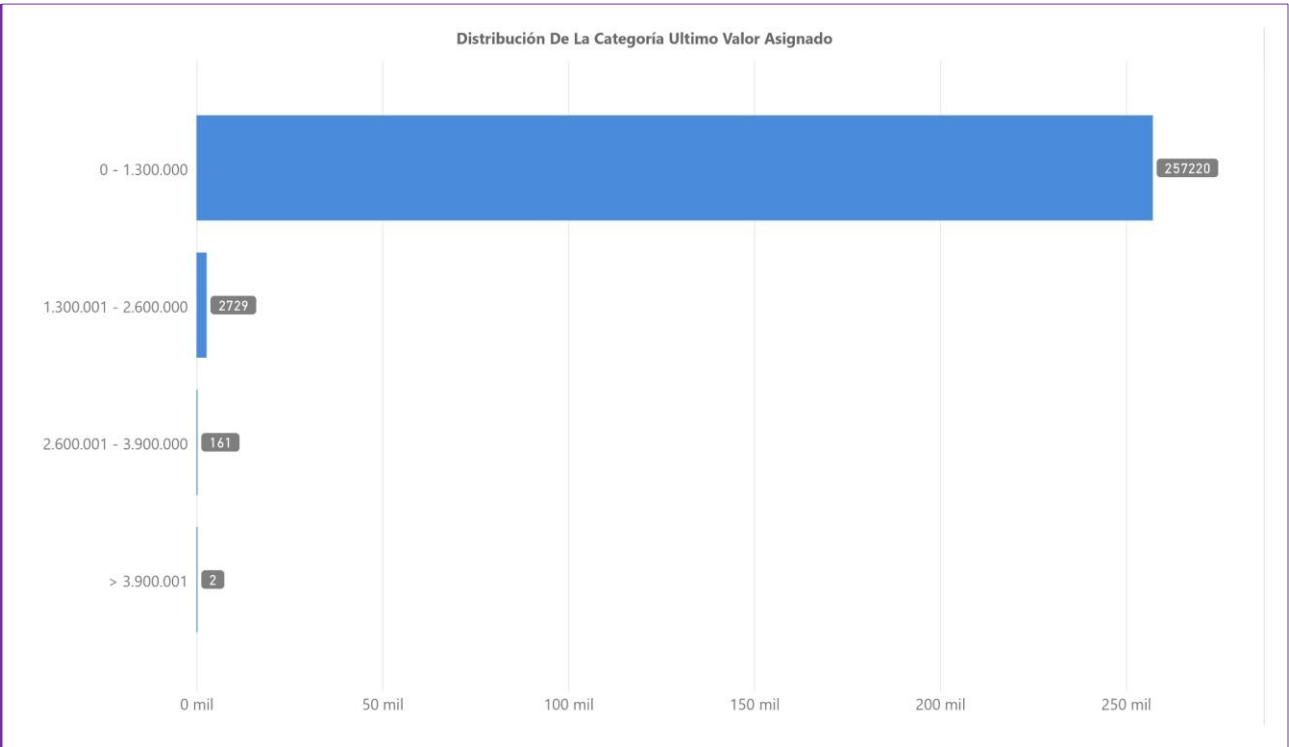
ADM-2.4 Analizar la calidad de los datos.

Se interpreta que muchos de los beneficiarios al momento de terminar su bachillerato ingresan rápidamente a un nivel superior de formación debido al número de tarjetas de identidad.



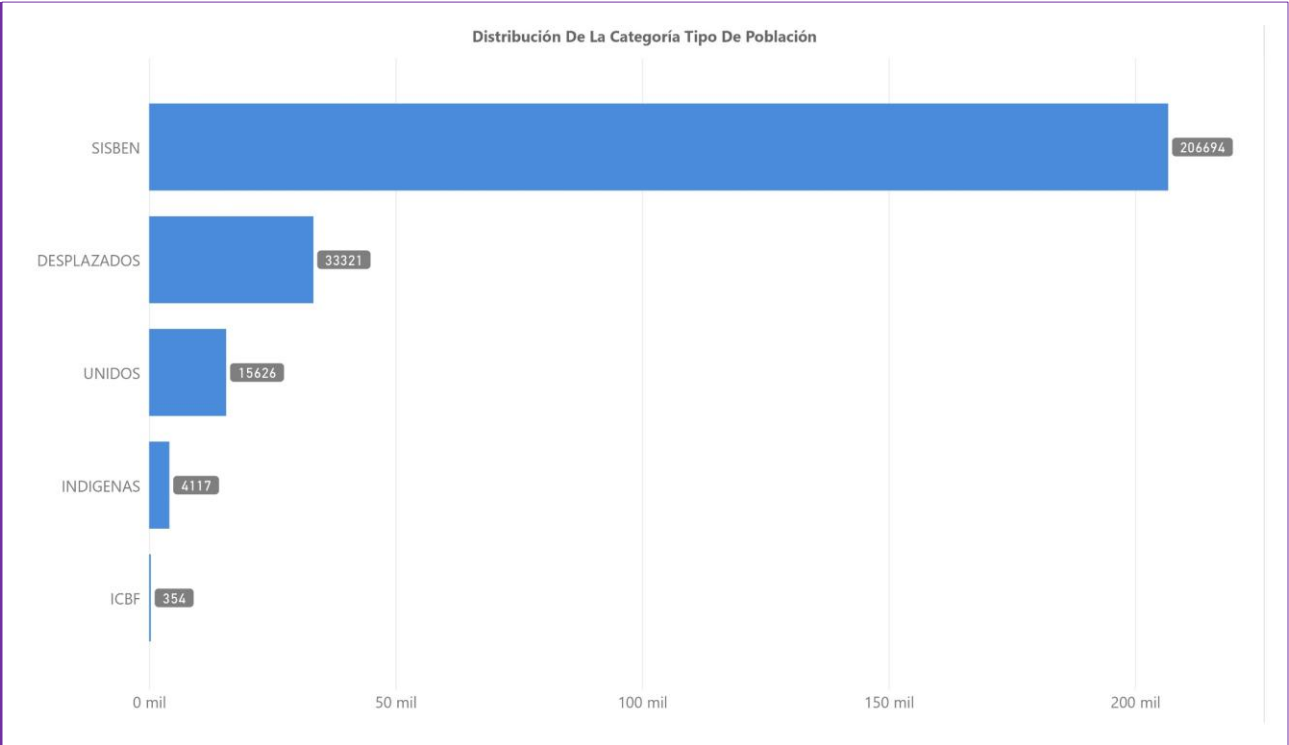
ADM-2.4 Analizar la calidad de los datos.

La categoría “Ultimo valor asignado” deberá ser eliminada debido a que su información es poco relevante, sin embargo, el valor total asignado si podría ser útil para determinar cuánto tiempo se usó estudiando



ADM-2.4 Analizar la calidad de los datos.

Se encuentra que a mayoría de la población esa sisbenizada, además del gran número de personas desplazadas, no contiene valores ND.



The background of the slide features a blurred image of a person with dark hair, seen from the side, working at a desk. A semi-transparent blue overlay covers the left and bottom portions of the image, while a solid orange vertical bar runs along the right edge. Centered within the blue area is a white rectangular box containing the text.

Fase ADM-3
Prepare and
explore data.

Fase ADM-3 Prepare and explore data.

Descripción:

Se realizarán los siguientes pasos con el fin de obtener los registros idóneos para la posterior exploración de los datos:

- ADM-3.1 Eliminación de los datos duplicados.
- ADM-3.2 Imputación o eliminación de los datos nulos o vacíos.
- ADM-3.3 Tratamiento de los datos atípicos.
- ADM-3.4 Normalización y estandarización del conjunto de datos.
- ADM-3.5 Codificación de variables categóricas a cuantitativas.
- ADM-3.6 Selección de categorías de mayor importancia para el análisis.

ADM-3.1 Eliminación de los datos duplicados.

En este caso se encontró que la columna de "fecha final", "T. Beneficio" y "T. Asignación" solo tienen un único registro que se repite en todos los registros, las columnas "Estado Civil" y "Discapacidad" tiene mayoritariamente registros ND por lo cual se procederá a eliminar completamente todas estas columnas.

Con esta modificación se pasa de 16 a 14 columnas.

- | | |
|------------------|---------------------|
| • Estado | • Rango Edad |
| • Etnia | • T. Población |
| • Fecha Inicial | • V. Total Asignado |
| • Genero | • V. Ult. Asignado |
| • N. Escolaridad | • Año |
| • Departamento | • # Mes |
| • Municipio | • Mes |

ADM-3.2 Imputación o eliminación de los datos nulos o vacíos.

Descripción:

Con las categorías ya filtradas la mayor parte de los registros no tienen datos nulos exceptuando ciertas columnas como “Etnia” o “N. Educativo”, estas categorías tienen la característica de que sus registros son del tipo categórico por lo cual realizar una imputación de datos no asegura una distribución muy precisa de los datos por lo cual se opta por omitir estos registros ND al momento de analizar dichas categorías.

ADM-3.3 Tratamiento de los datos atípicos.

Descripción:

Debido al tipo a que los datos en su mayoría son del tipo categórico no se logran encontrar datos atípicos como sería usual en una categoría con registros numéricos.

ADM-3.4 Normalización y estandarización del conjunto de datos.

Descripción:

Para el caso de la base de datos públicas los datos ya fueron previamente normalizados por lo cual no es necesario realizar nuevamente el proceso.

ADM-3.5 Codificación de variables categóricas a cuantitativas.

Descripción:

No se requiere para el avance del análisis la cuantificación de ninguna de las categorías en la base de datos ni en la encuesta.

ADM-3.6 Selección de categorías de mayor importancia para el análisis.

Descripción:

Se seleccionaron ya las categorías con las cuales se piensa dar inicio al análisis, cabe destacar la reducción significativa del número de categorías inicial a lo que se tiene ahora, debido a lo irrelevante o la mala recolección de los datos para dicha categoría.

- **Estado**
- **Etnia**
- **Fecha Inicial**
- **Genero**
- **N. Escolaridad**
- **Departamento**
- **Municipio**
- **Rango Edad**
- **T. Población**
- **V. Total Asignado**
- **V. Ult. Asignado**
- **Año**
- **# Mes**
- **Mes**



Fase ADM-4 Data analysis.

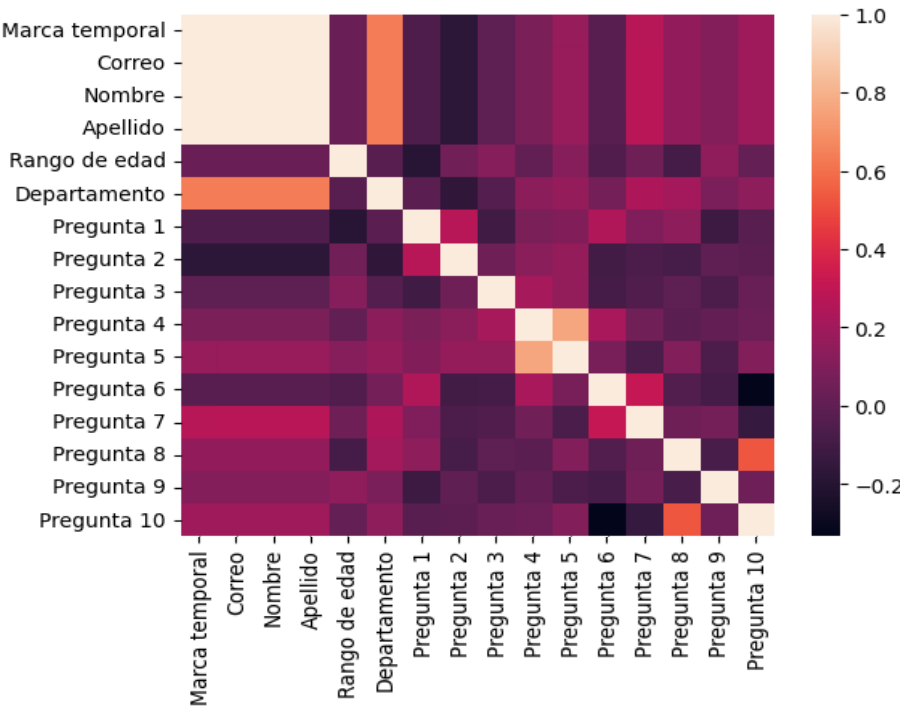
- ADM-4.1 Determinar el nivel de correlación entre variables.
- ADM-4.2 Explorar las variables estadísticas descriptivas.
- ADM-4.3 Realizar visualizaciones de los datos.
- ADM-4.4 Evolución de las variables en series de tiempo.

Heatmap showing the correlation matrix for 14 variables. The color scale ranges from -0.2 (dark purple) to 1.0 (light yellow). The diagonal is all 1.0. Notable correlations include Estado and Fecha Inicial (approx. 0.8), and # Mes and Mes (approx. 0.9).

ADM-4.1 Determinar el nivel de correlación entre variables.

Descripción:

Para el caso de la encuesta también se mantiene el mismo patrón de correlación donde las categorías de tipo fecha y las de datos personales son las únicas que tienen una correlación positiva.



ADM-4.2 Explorar las variables estadísticas descriptivas.

Descripción:

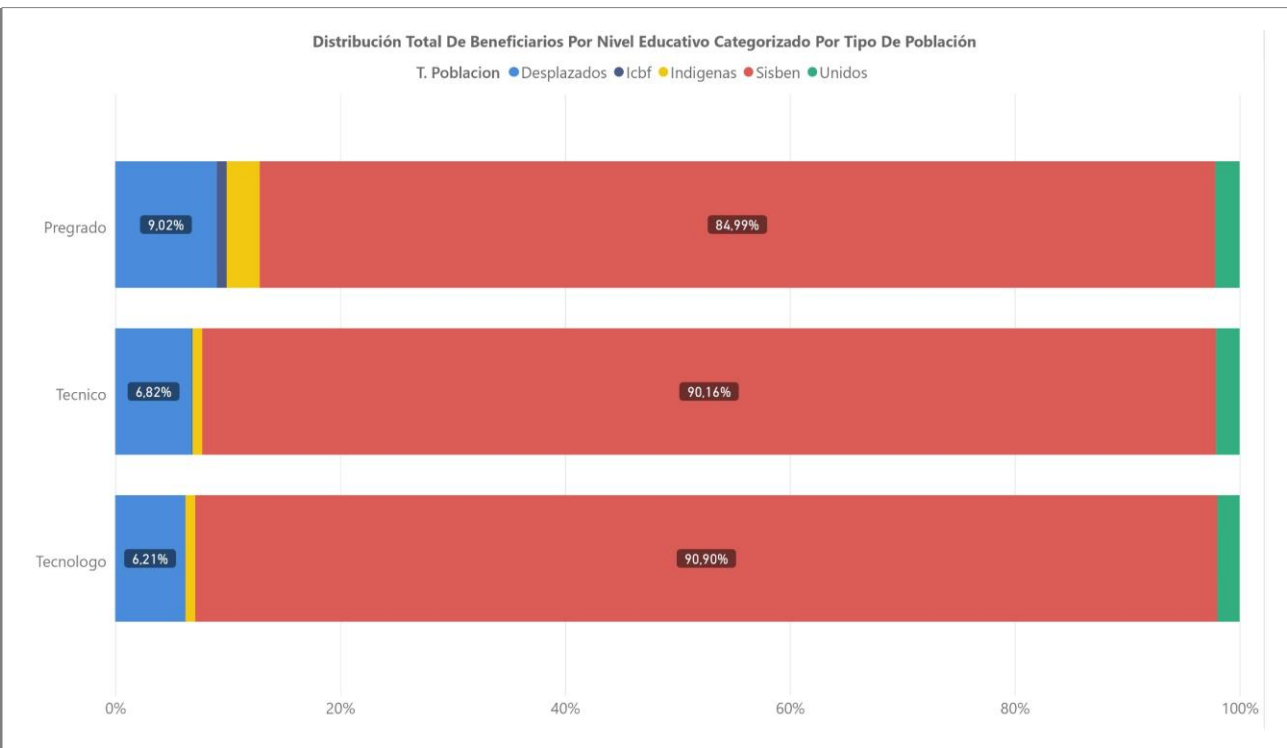
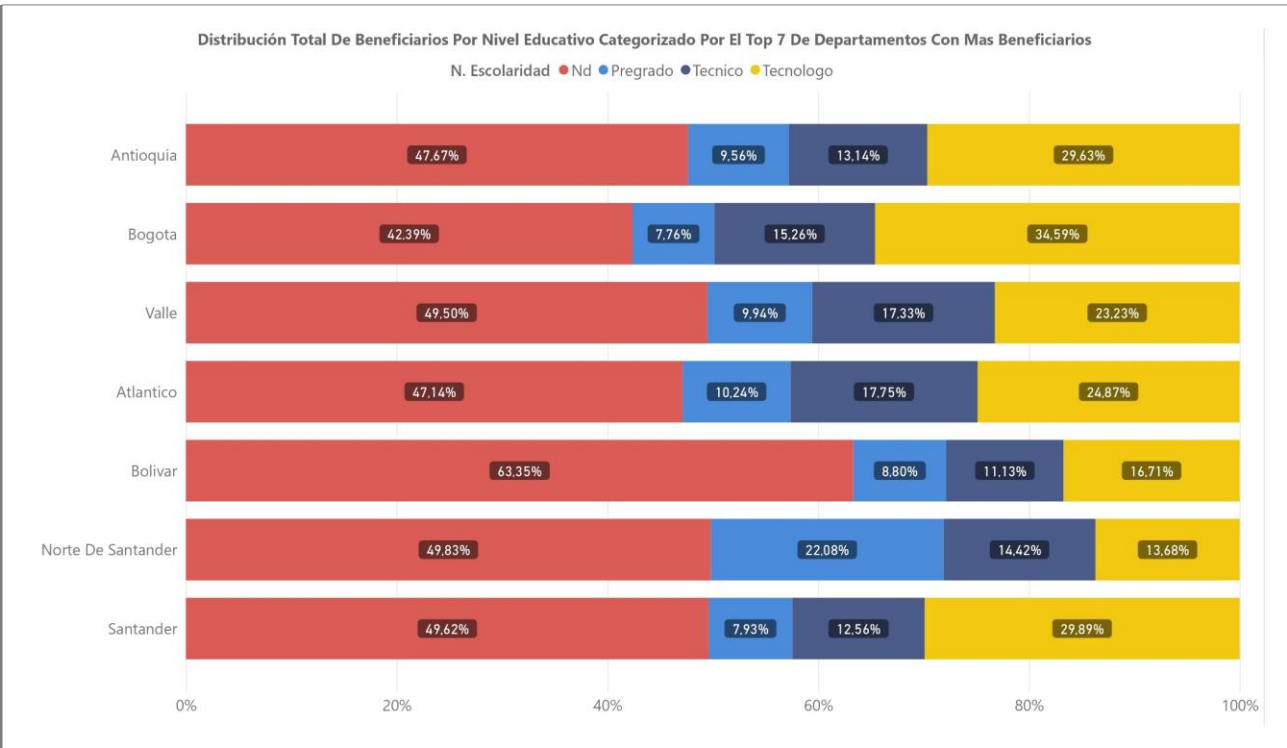
Estas serían las estadísticas descriptivas de las categorías escogida para realizar el análisis, en el siguiente paso se apreciarán de forma más grafica

	Estado	Etnia	Fecha Inicial	Genero	N. Escolaridad	Departamento
count	260112	260112	260112	260112	260112	260112
unique	2	9	50	3	7	33
top	NO ACTIVO	ND	2014-07-01 00:00:00	Mujer	ND	ANTIOQUIA
freq	158945	242782	29570	144076	138197	25672

	Municipio	Rango Edad	T. Documento	T. Poblacion	V. Total Asignado	Mes
count	260112	260112	260112	260112	260112	260112
unique	1018	5	5	5	5	12
top	BOGOTA D.C.	18-29	CC	SISBEN	1.500.001 - 3.000.000	July
freq	19476	213376	169793	206694	77971	75408

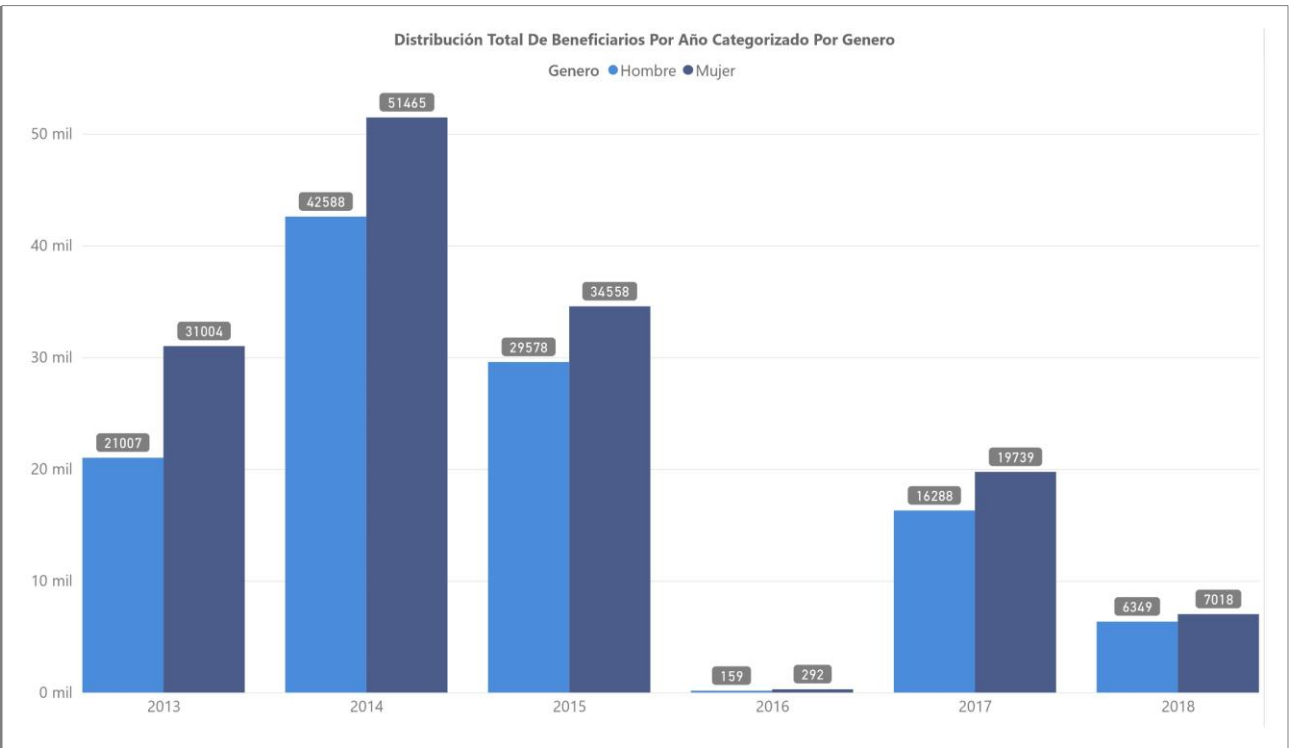
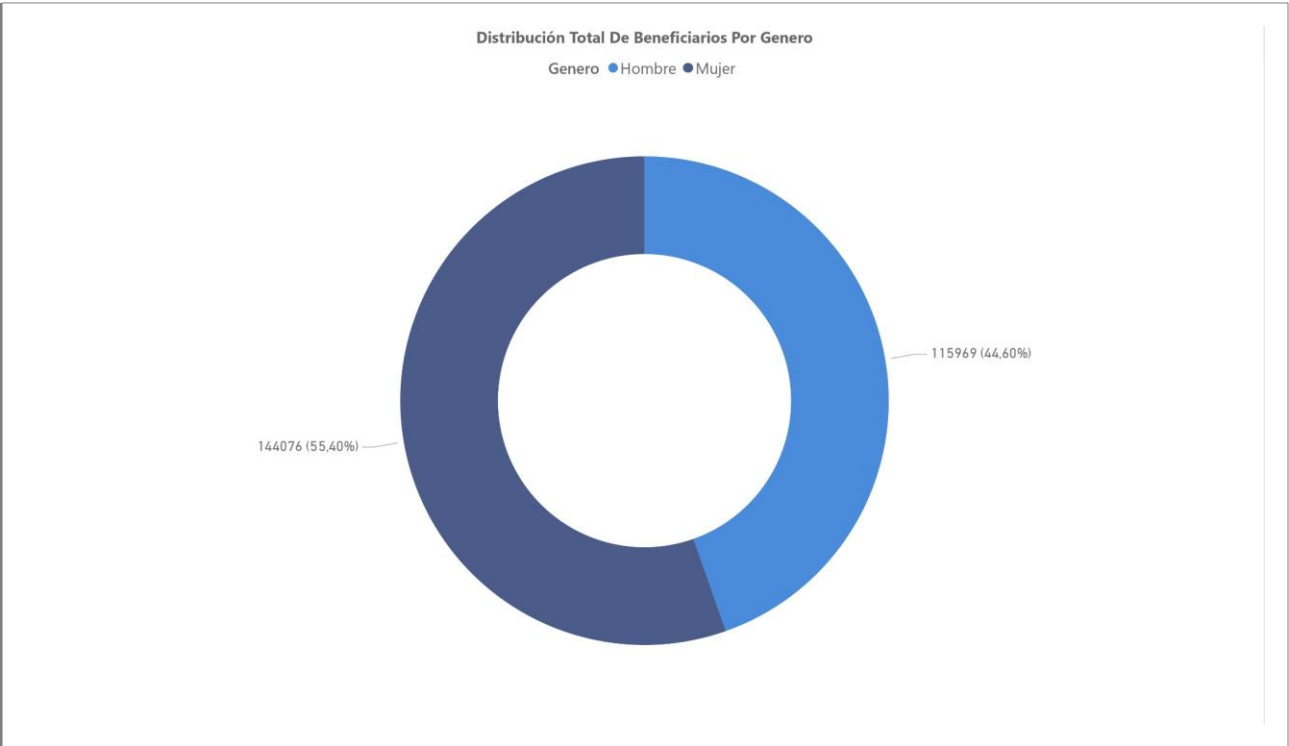
ADM-4.3 Realizar visualizaciones de los datos.

Exceptuando los registros ND, se encuentra que en el top 7 dptos. el nivel académico que predomina es el tecnólogo, la población desplazada destaca en todos los niveles académicos.



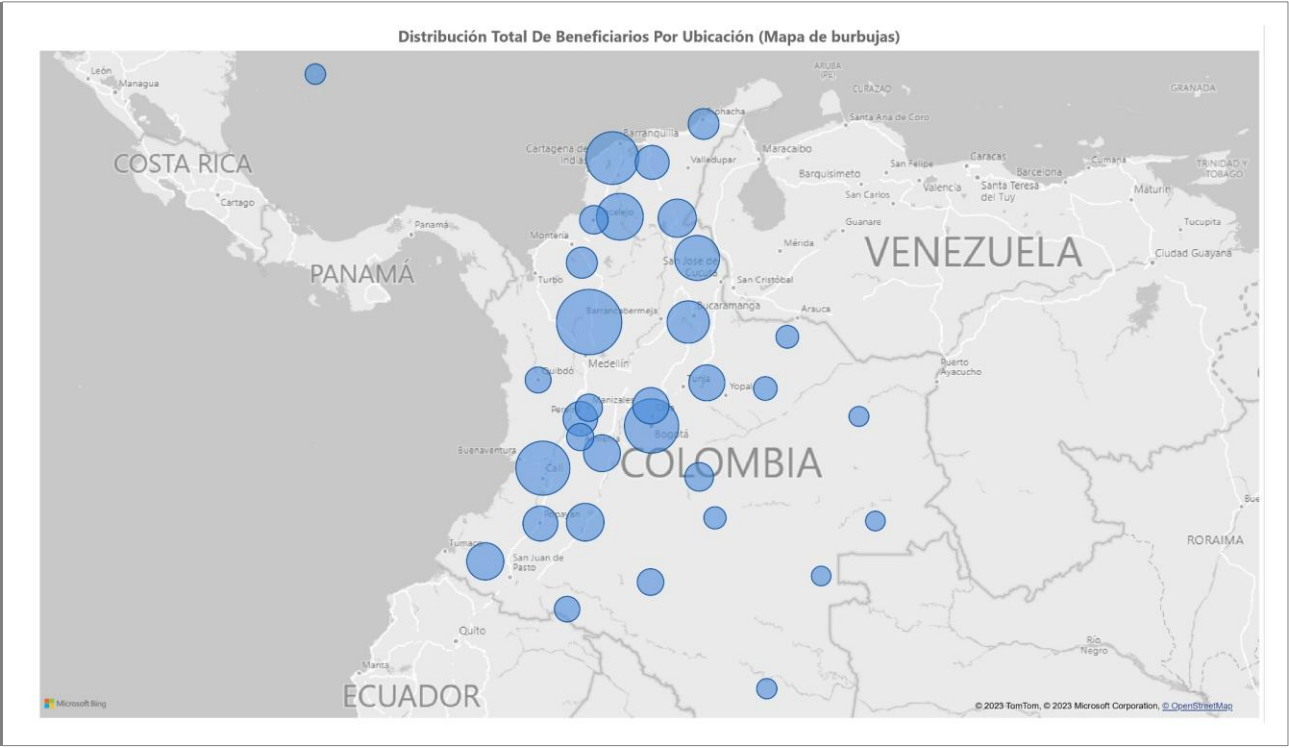
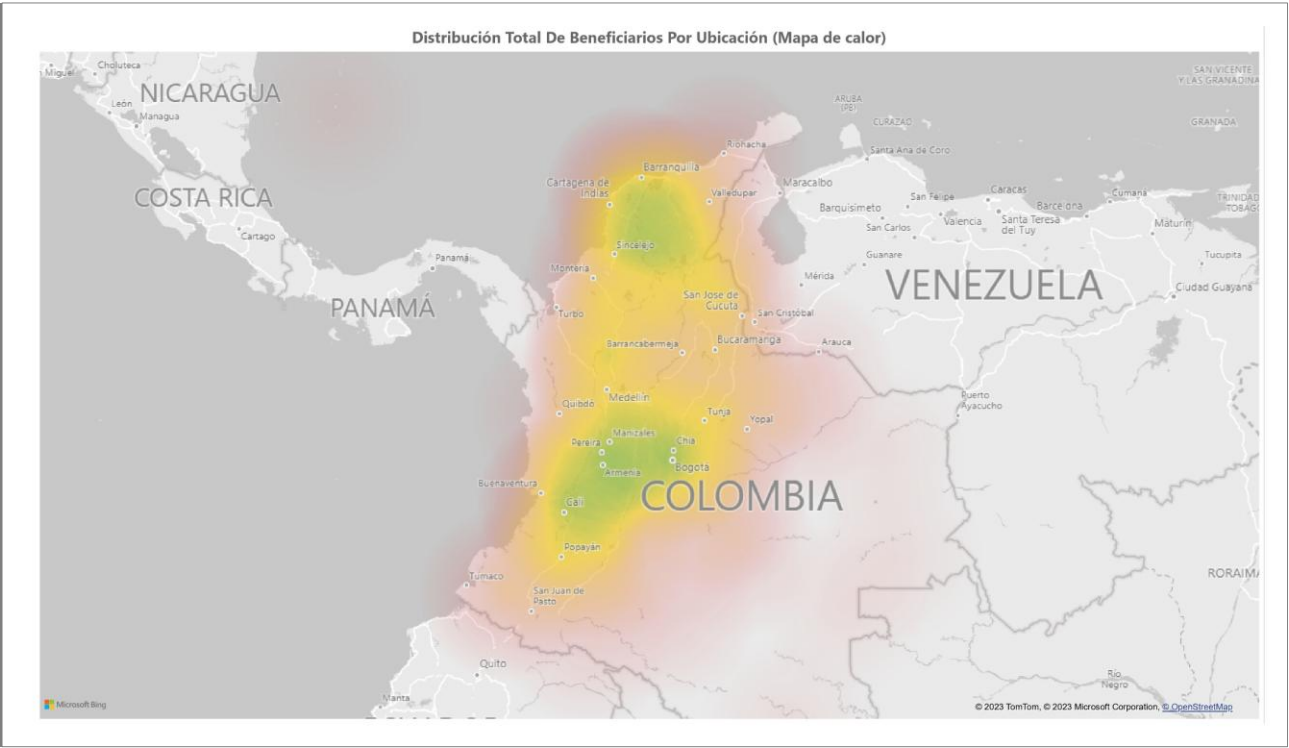
ADM-4.3 Realizar visualizaciones de los datos.

Ahora observando la evolución de la categoría género a través del tiempo, se aprecia la tendencia de que las mujeres tienen mayor número de beneficiarias.



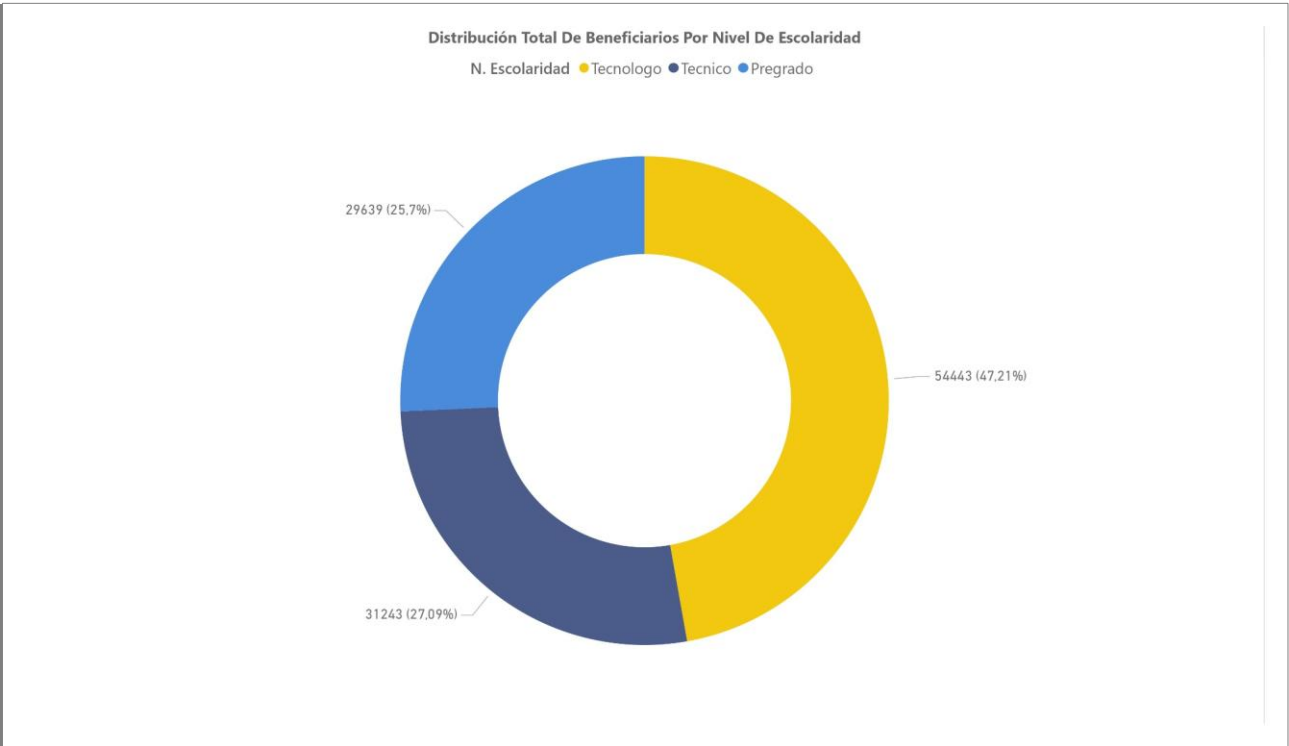
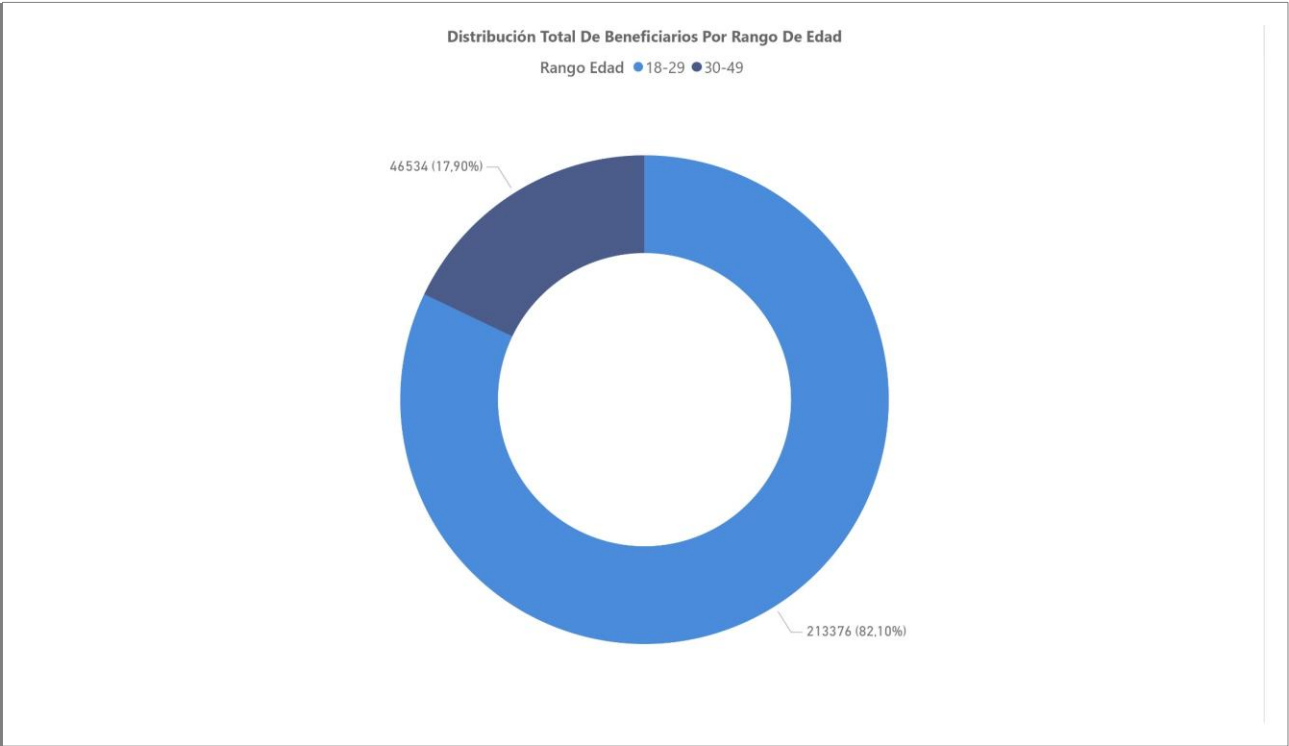
ADM-4.3 Realizar visualizaciones de los datos.

La distribución de los beneficiarios en los mapas indica que una mayoría significativa de los beneficiarios se encuentra reunida alrededor de las grandes urbes del país.



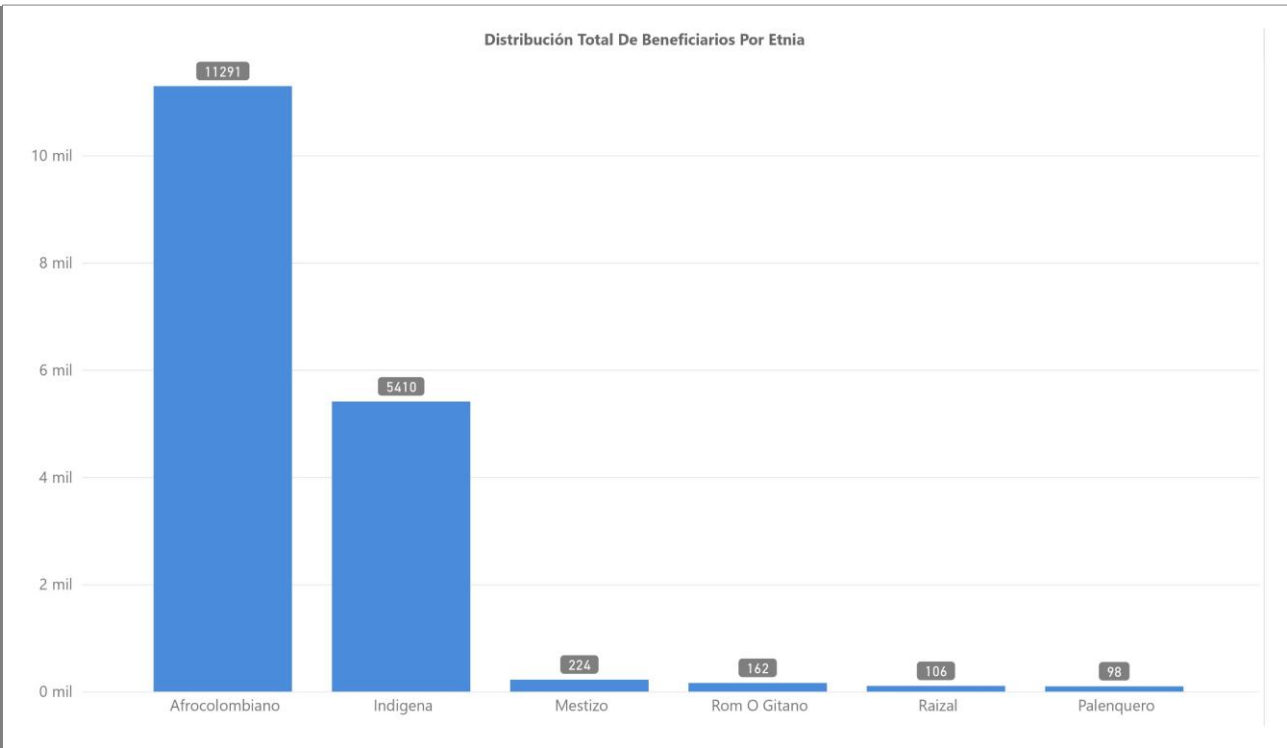
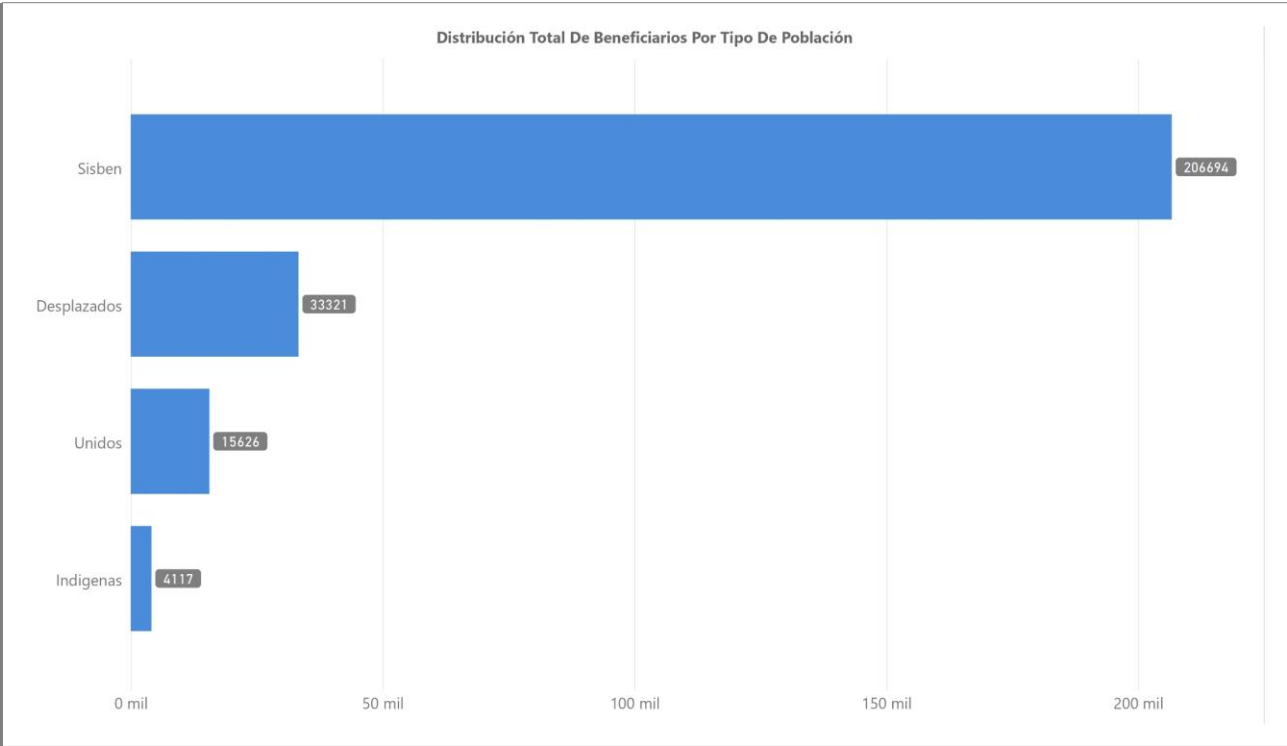
ADM-4.3 Realizar visualizaciones de los datos.

Se puede suponer que el rango de 30-49 sea de los inicios del programa o errores de ingreso ya que la edad focal del beneficio no permite las edades contenidas en dicho rango.



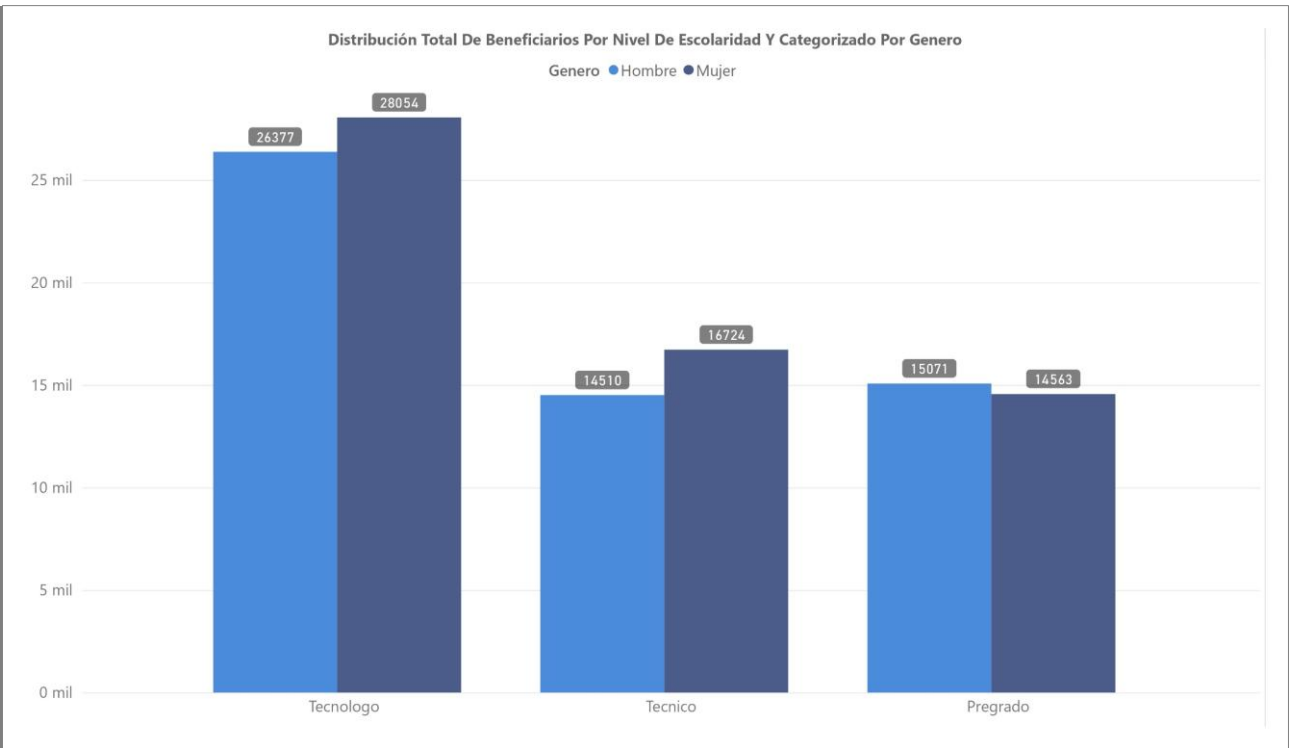
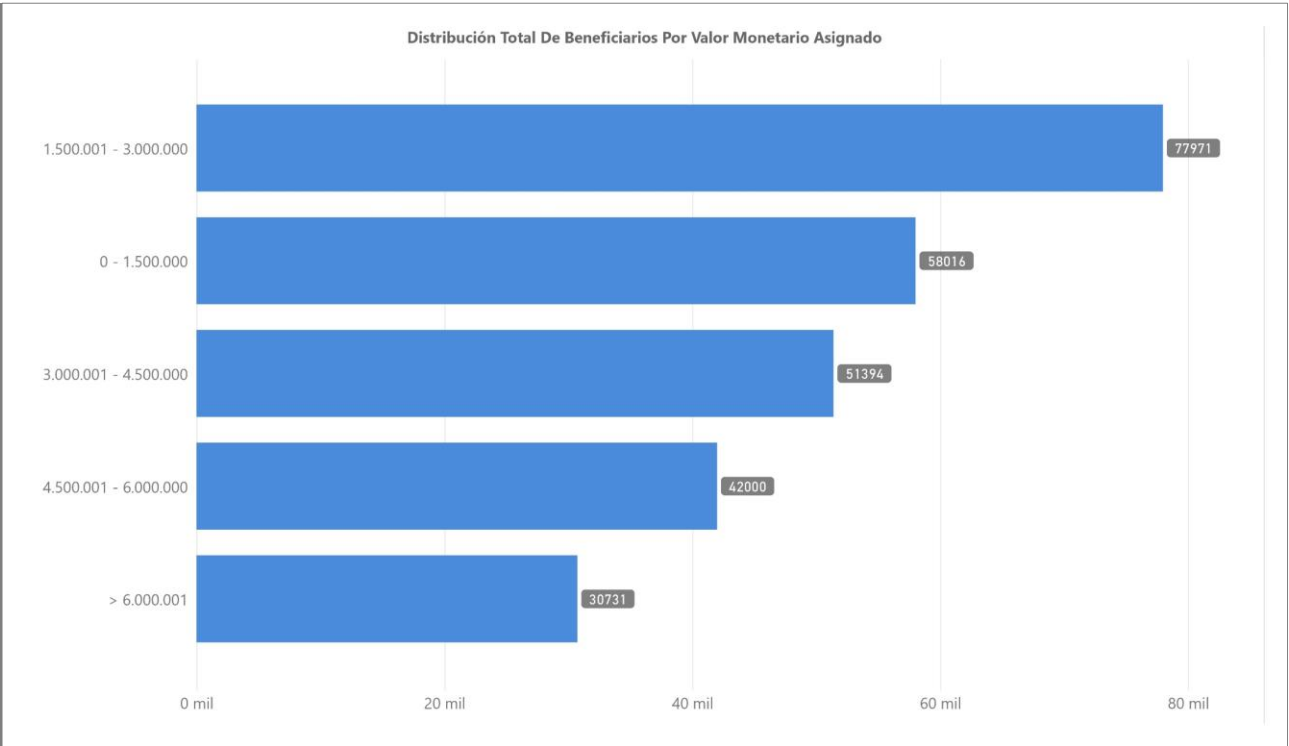
ADM-4.3 Realizar visualizaciones de los datos.

El Sisbén y los afrocolombianos son el tipo de población y etnia más predominantes entre los beneficiarios.

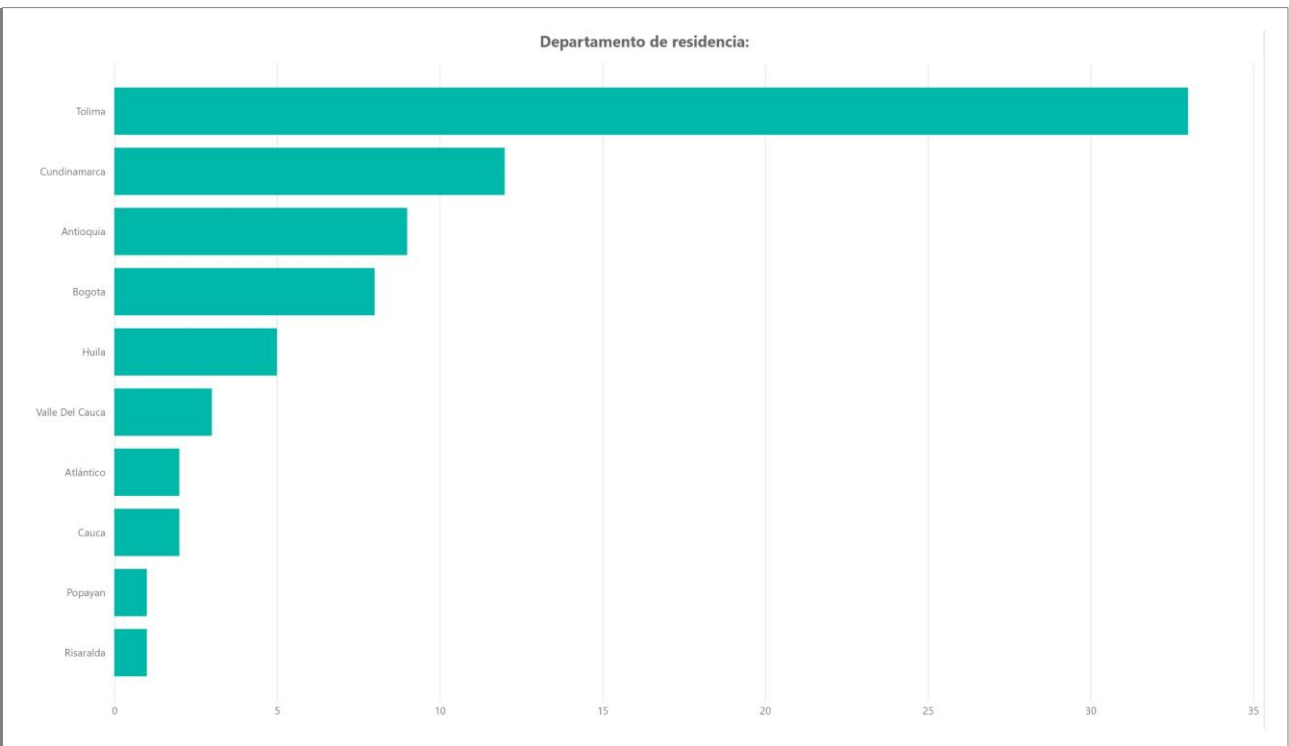
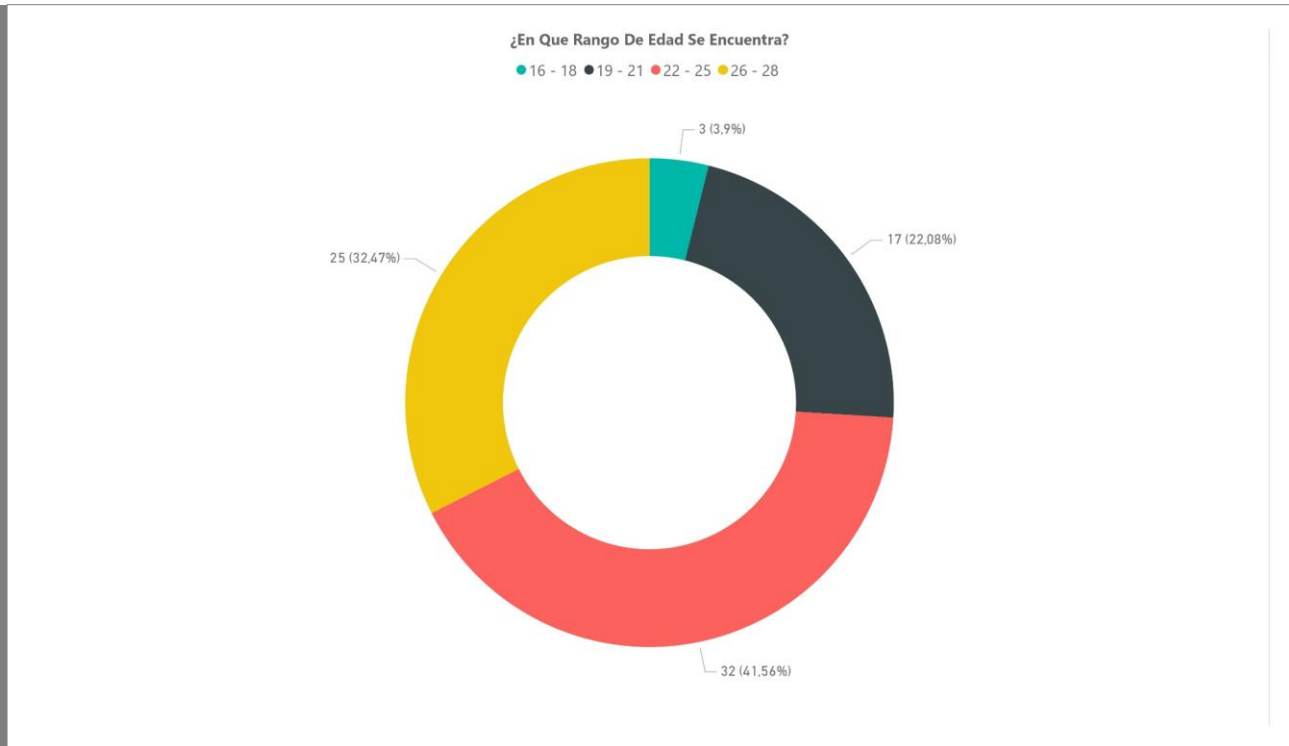


ADM-4.3 Realizar visualizaciones de los datos.

Los valores asignados concuerdan con que el nivel educativo mayoritario sea el tecnólogo, ya que por el tiempo en que se reciben las transferencias coincide con el posible estimado recibido.



ADM-4.3 Realizar visualizaciones de los datos (encuesta).



ADM-4.3 Realizar visualizaciones de los datos (encuesta).

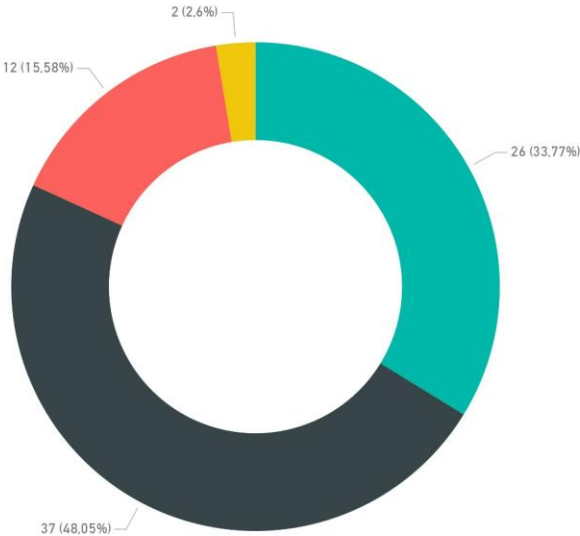
¿Es O Alguna Vez Fue Usted Beneficiario De Jóvenes En Acción?

No. Nunca he sido beneficiario Sí. Actualmente soy beneficiario Sí. Alguna vez fui beneficiario



¿Durante Cuantos Años A Recibido Jóvenes En Acción?

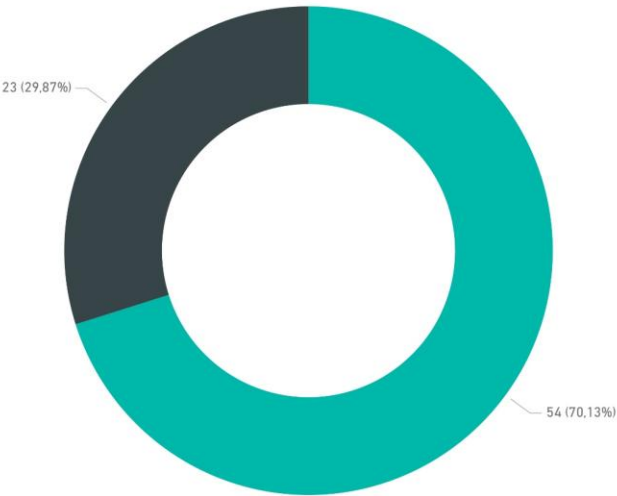
1 2 4 5



ADM-4.3 Realizar visualizaciones de los datos (encuesta).

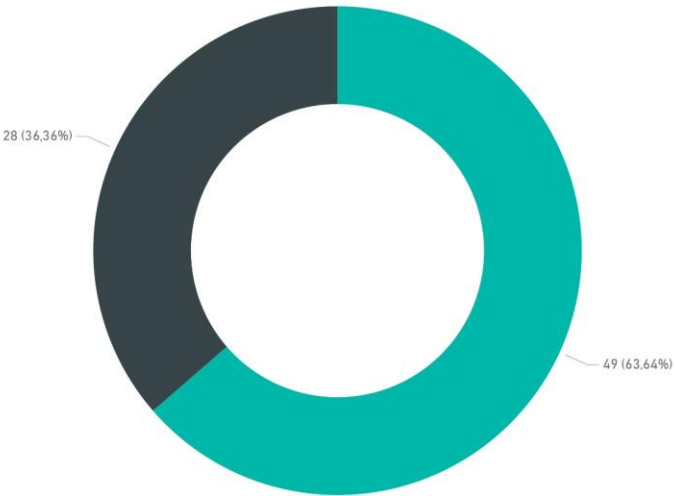
¿Tenia Conocimiento Sobre La Información Sobre El Estado De "En Transición"?

● No ● Si

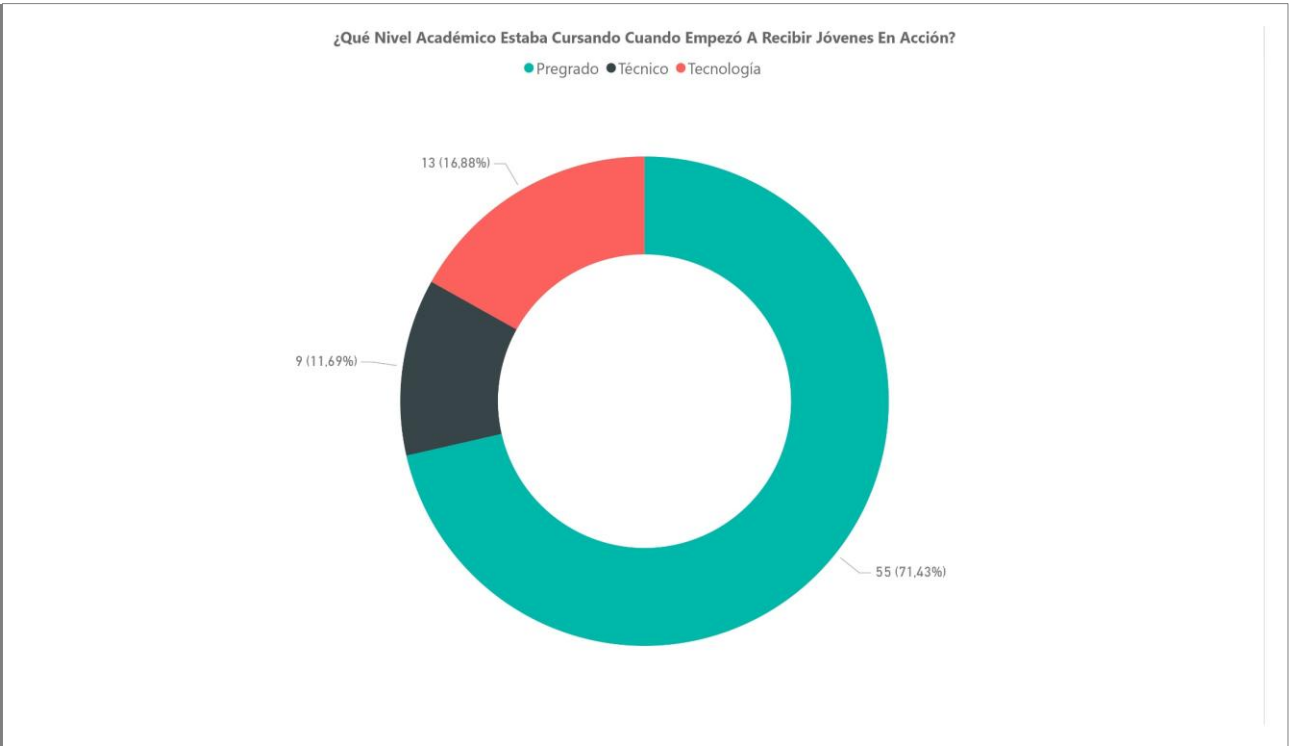
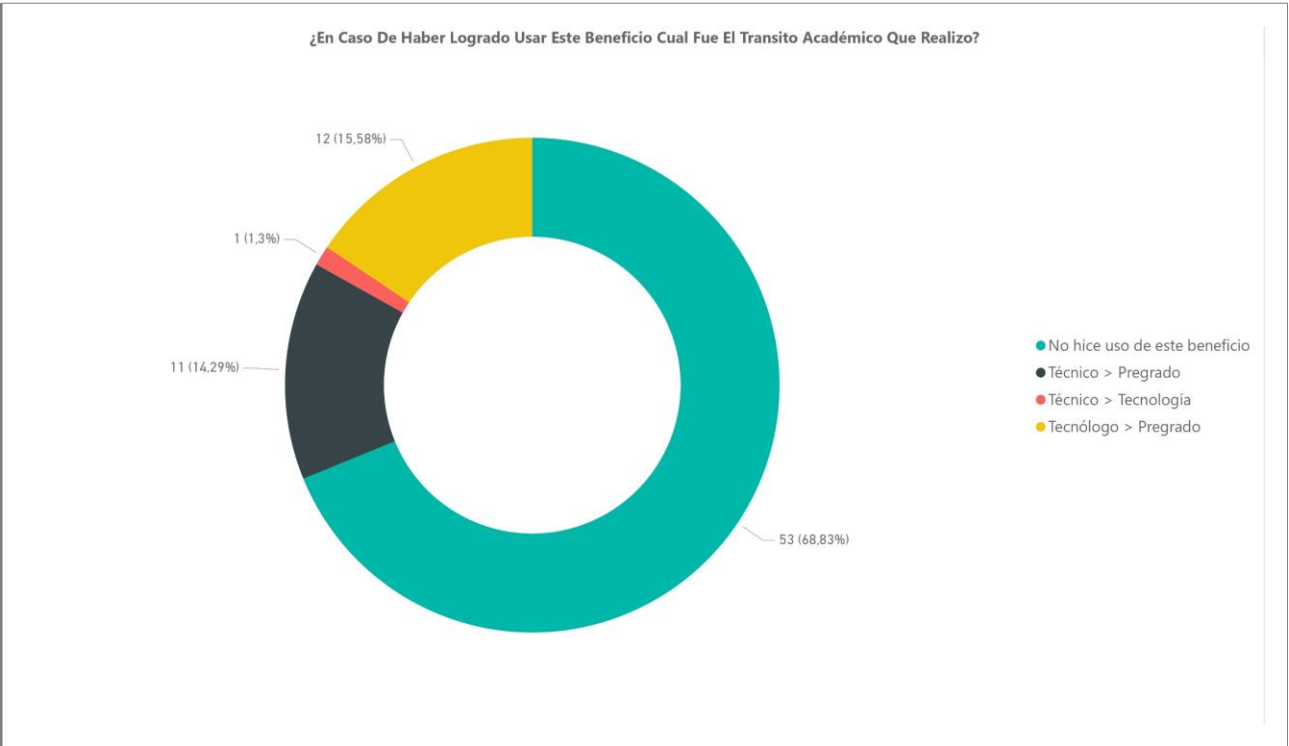


¿Logro Usar Este Beneficio (En Transición) Para Continuar Con Su Formación Académica?

● No ● Si

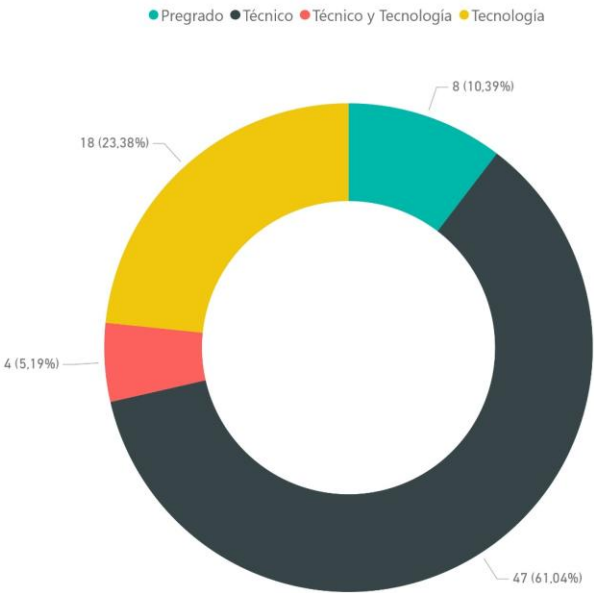


ADM-4.3 Realizar visualizaciones de los datos (encuesta).

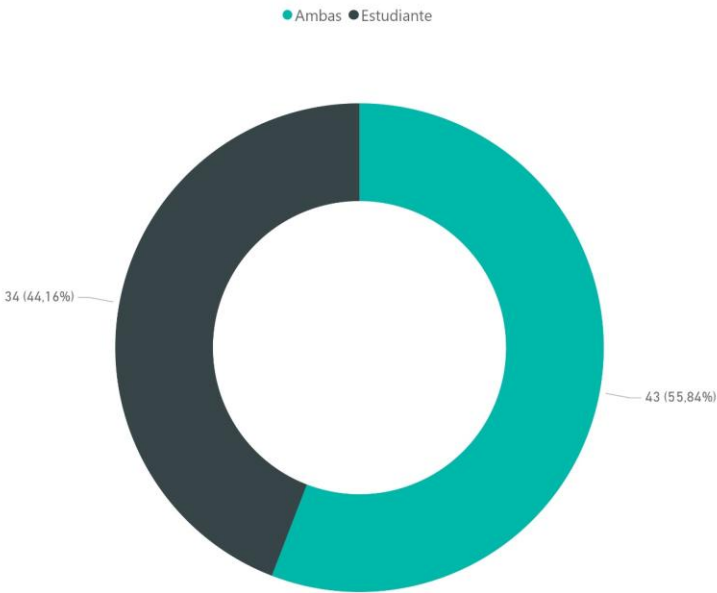


ADM-4.3 Realizar visualizaciones de los datos (encuesta).

¿De Cual De Los Siguietes Niveles Académicos Se A Graduado?

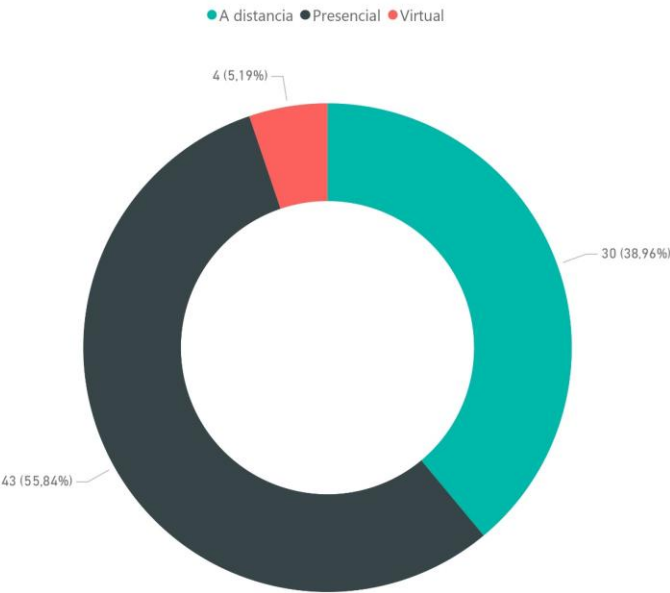


¿En Este Momento Estudia, Trabaja O Ambas?

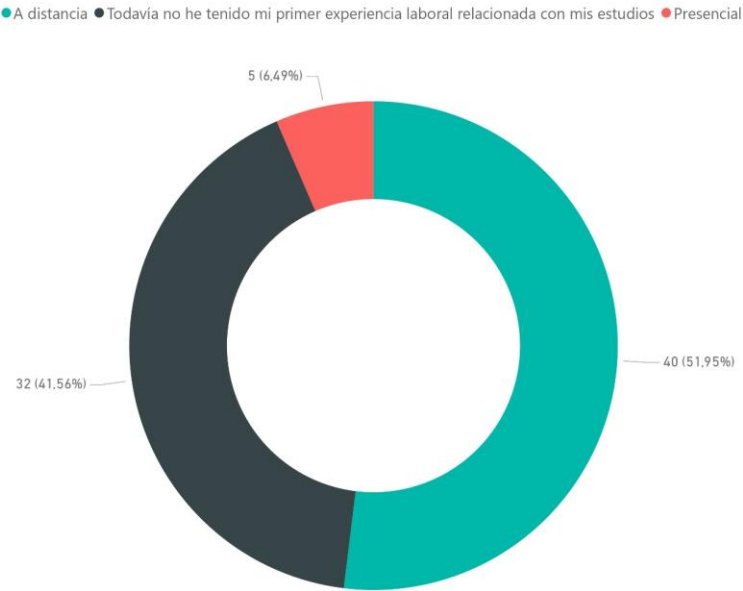


ADM-4.3 Realizar visualizaciones de los datos (encuesta).

¿Cómo Realizo Su Primer Formación Académica?

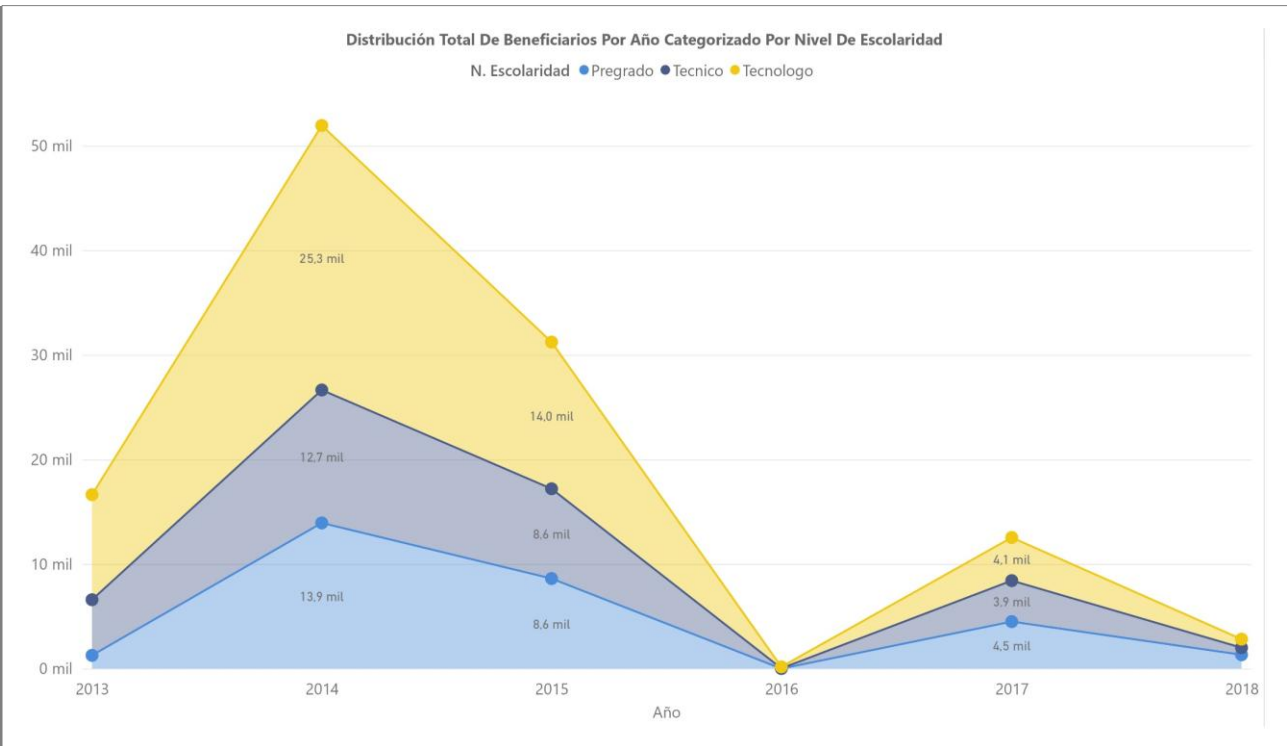
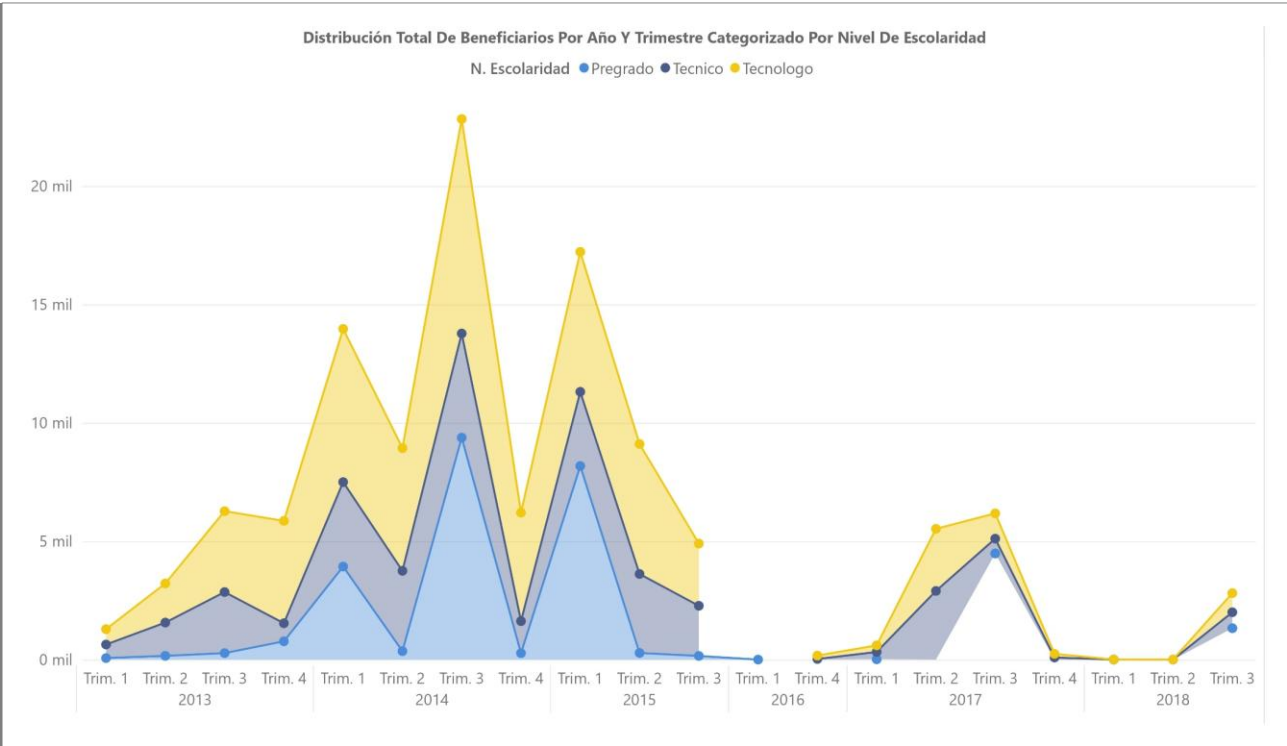


Si ya trabaja, ¿Cómo continua ahora su formación académica?



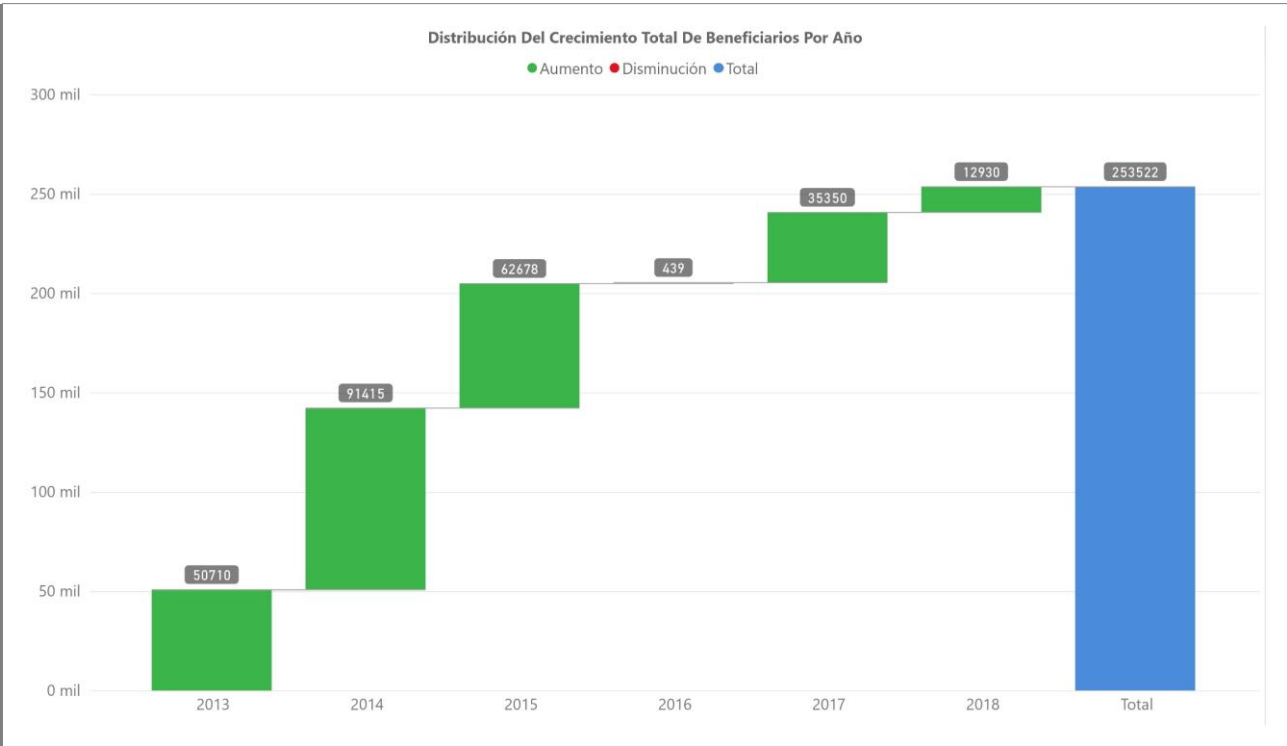
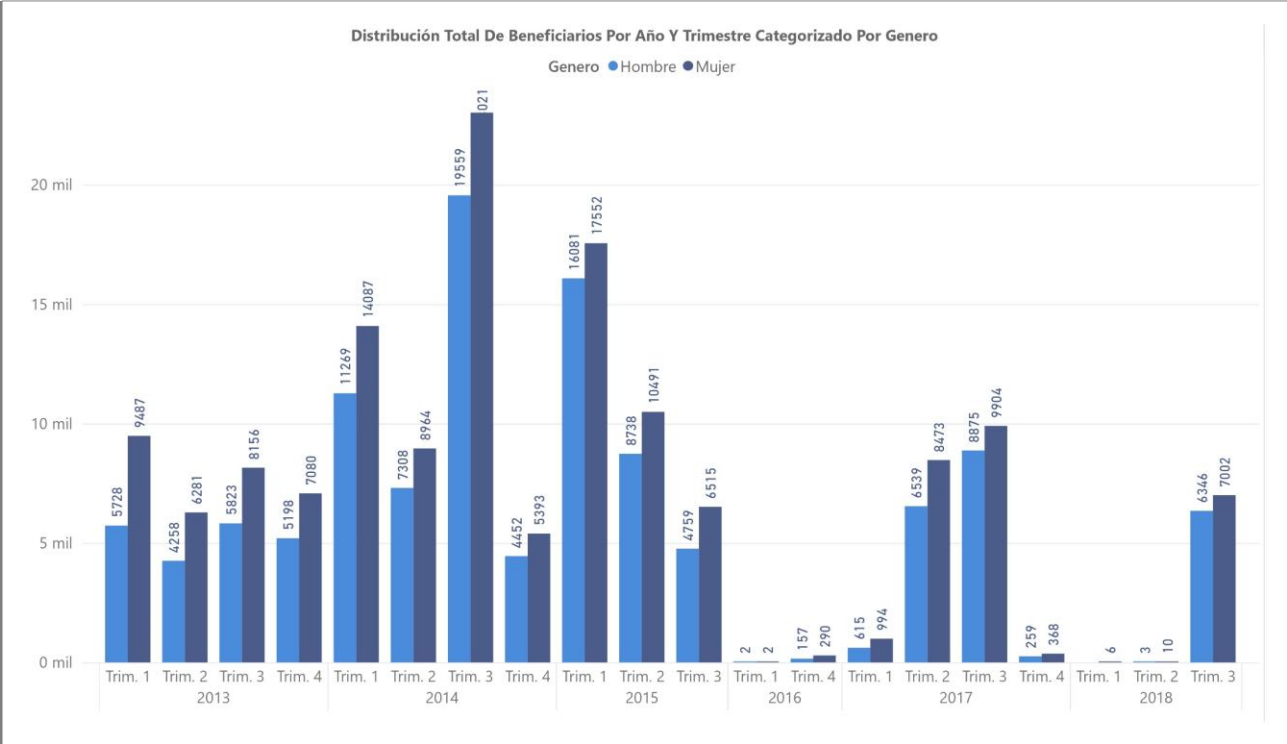
ADM-4.4 Evolución de las variables en series de tiempo.

La disminución de beneficiarios a través del tiempo puede suponer que los registros estén incompletos o que los jóvenes desconozcan cada vez más el programa.



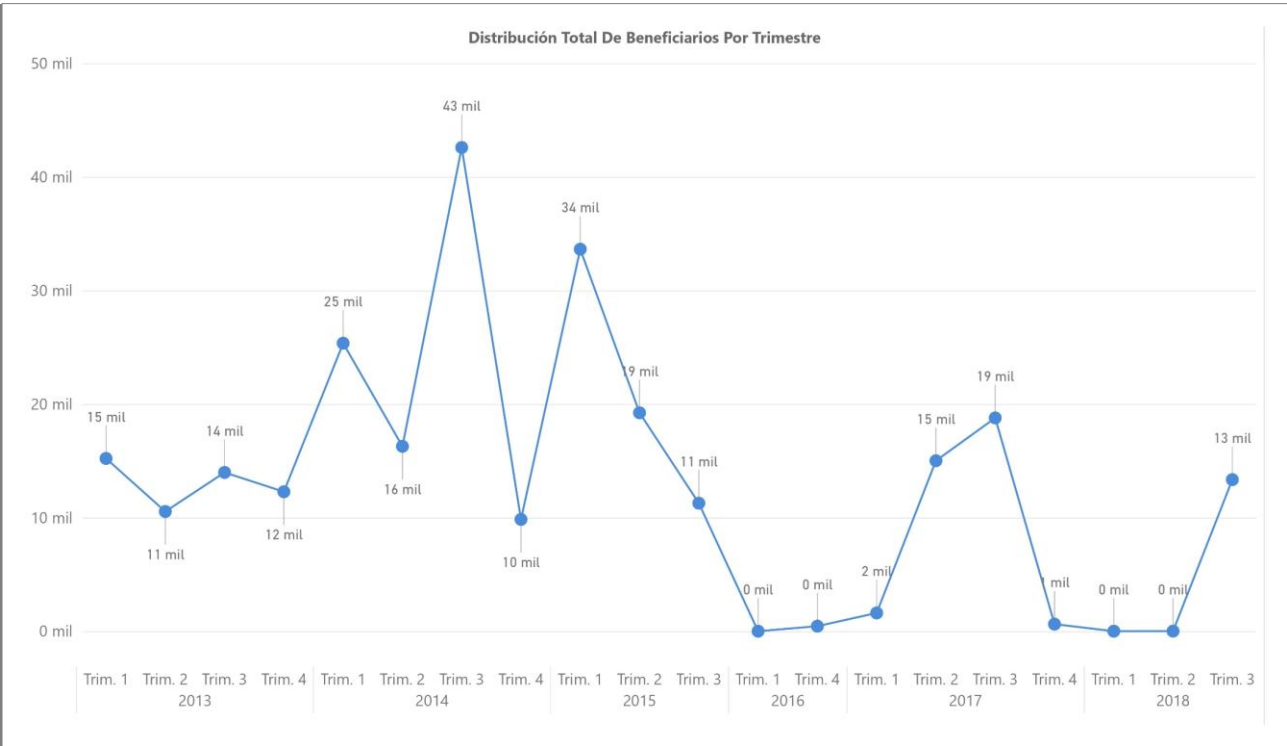
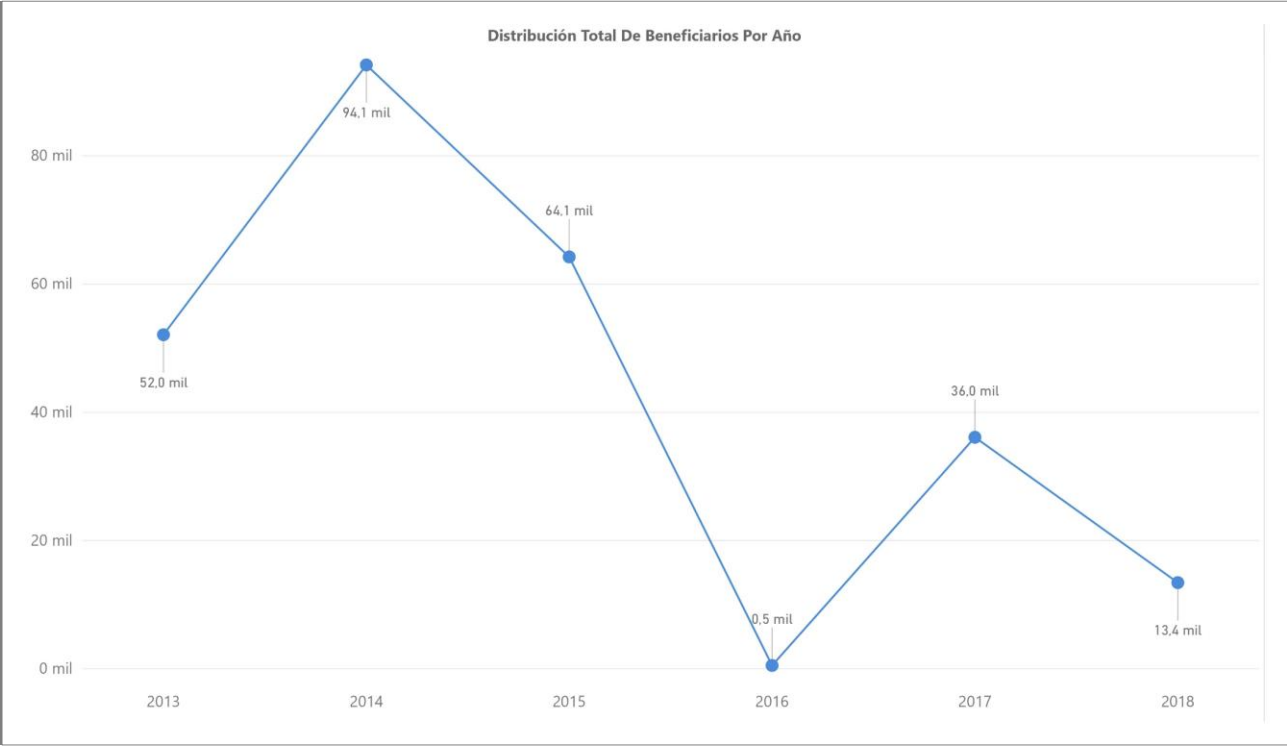
ADM-4.4 Evolución de las variables en series de tiempo.

Es motivo de una posterior investigación el saber porque el número de nuevos beneficiarios disminuye en los años posteriores.



ADM-4.4 Evolución de las variables en series de tiempo.

Es necesario indagar en los procedimientos de registro de los programas de prosperidad social ya que resulta extraño la poca cantidad de registros en ciertos periodos de tiempo.





Fase ADM-5 Present Findings.

Fase ADM-5 Present Findings.

Descripción:

Se realizarán los siguientes pasos con el fin de obtener los registros idóneos para la posterior exploración de los datos:

- ADM-5.1 Creación del contexto para el lector de la información.
- ADM-5.2 Enseñar los hallazgos claves.
- ADM-5.3 Establecer los resultados logrados.
- ADM-5.4 Definir la conclusiones y recomendaciones.

ADM-5.1 Creación del contexto para el lector de la información.

Descripción:

Se parte de la premisa de el nivel de implicación que tiene el programa de jóvenes en acción del departamento de prosperidad social en la continuación de su proceso formativo a un nivel superior de educación, debido a que el programa cuenta con una continuidad en un periodo de tiempo más largo si el estudiante decide seguir estudiando, por ende se quiere identificar cuantos de los beneficiarios tenían conocimiento de este beneficio, si hicieron uso del mismo y que tanto influyo en la decisión de seguir formándose.

La investigación se encontró con múltiples altibajos, debido a que la base de datos públicas que es nuestro mayor volumen de datos se encontraba significativamente llena de registros “ND = No Data”, por lo cual no resultan muy útiles para el análisis, por otra parte, una de las categorías más importantes como lo era “Fecha de finalización”, resulto ser una categoría con un único registro, por lo cual resulto imposible calcular cuánto tiempo permaneció en el programa cada uno de los registros.

Con estas implicaciones el objetivo principal de la investigación era prácticamente imposible de conseguir, debido a esto nos vimos en la necesidad de basarnos en las proporciones obtenidas en la encuesta realizada, para de esta forma intentar extrapolar una información sobre la otra.

ADM-5.1 Creación del contexto para el lector de la información.

Descripción:

De esta forma se debe hacer claridad en el hecho de que el objetivo principal de la investigación se puede considerar parcialmente alcanzada o directamente no alcanzado. Sin embargo, se encontraron hallazgos importantes que serán compartidos en el siguiente punto.

ADM-5.2 Enseñar los hallazgos claves.

Descripción:

Los siguientes datos son los hallazgos obtenidos de las bases de datos públicas y de la encuesta realizada.

1. El nivel académico con más beneficiarios es tecnólogo con un total de 54443.
2. Los niveles académicos técnico y tecnólogo tienen un número de beneficiarios bastante próximo 31243 y 29639 respectivamente.
3. La mayoría de los beneficiarios son mayores de edad con un total de 169793.
4. Son pocos los beneficiarios identificados como discapacitados con un total de 441.
5. Las mujeres son el género con mayor número de beneficiarias con un total de 144074
6. La mayoría de los beneficiarios no se identifican con ningún tipo de etnia con un total de 242782
7. La etnia con más beneficiarios es afrocolombiana con un total de 11291.
8. El top 5 de departamentos con mayor número de beneficiarios es Antioquia con 25672, Bogotá con 19476, Valle del Cauca con 19372, Atlántico con 18589 y Bolívar con 15293, mostrando una concentración en el centro y en la costa atlántica del país.
9. El top 5 municipios con mayor número de empresarios es Bogotá con 19476, Barranquilla con 13490, Medellín con 13362, Cartagena con 11775 y Cúcuta con 8539.
10. El año con mayor número de ingreso de beneficiarios fue el 2014.
11. El mes con mayor número de ingresos de beneficiarios es julio.

ADM-5.2 Enseñar los hallazgos claves.

Descripción:

Los siguientes datos son los hallazgos obtenidos de las bases de datos públicas y de la encuesta realizada.

12. La mayoría de los beneficiarios se clasifican como sisbenizados con un total de 206604, seguido por los desplazados con un total de 33321.
13. El rango mayoritario de edad entre los beneficiarios encuestados es 22 a 25 años con 41,56%.
14. De los beneficiarios encuestados el 48,05% han recibido por 2 años el subsidio de JeA.
15. De los beneficiarios encuestados el 70,13% no tenían conocimiento del estado de tránsito dispuesto en JeA.
16. De los beneficiarios encuestados el 63,64% no usaron el beneficio de en tránsito para su formación académica.
17. De los beneficiarios encuestados el 15,58% realizó el tránsito de tecnólogo > pregrado, el 14,29% realizó el tránsito de técnico > pregrado y el 1,3% realizó el tránsito de técnico > tecnólogo.
18. De los beneficiarios encuestados el 71,43 % inicio a recibir JeA en el nivel académico pregrado.
19. De los beneficiarios encuestados el 61,04% se graduaron de una tecnología.
20. De los beneficiarios encuestados el 55,84% estudia y trabajan al mismo tiempo.
21. De los beneficiarios encuestados el 55,84% realizaron su primera formación en formato presencial.
22. De los beneficiarios encuestados el 51,95% que trabaja continua su formación en formato a distancia.
23. La tendencia del número de beneficiarios en base a los datos obtenidos tiende a ir a la baja.

ADM-5.3 Enseñar los hallazgos incidentales.

Descripción:

Los siguientes hallazgos se obtienen producto del análisis y el planteamiento de hipótesis para obtener una posible causa de origen de alguna irregularidad o anomalía en los datos

1. El número de registros ND tan significativamente alto en múltiples categorías de alta importancia como por ejemplo el N. de escolaridad, se supone un incorrecto diligenciamiento o perdida de los datos.
2. Categorías que solo cuentan con un tipo de registro como por ejemplo fecha ultima beneficio asignado, se supone un incorrecto diligenciamiento o el no uso de la categoría.
3. La ausencia casi total de registros en ciertos intervalos de tiempo como por ejemplo el año 2016, se supone una pérdida total de los datos.
4. Un bajo nivel de segmentación en la edad de los beneficiarios ya que solo le logra encontrar dos tipos de rangos etarios, de las cuales una de ellas no concuerda con los lineamientos establecidos por el programa de JeA.
5. Poca o nula información pública de los formatos para el registro de información que alimentan la base de datos, se supone esta como una de las principales causales del alto número de registros de datos ND.
6. Poca o nula información pública del ciclo del procesamiento y aseguramiento de la calidad del dato para el programa JeA, se supone esta otra razón para comprender el porqué del estado actual en que se encuentra la base de datos.
7. Poca o nula información pública de las aptitudes y requerimientos formativos de las personas encargadas de realizar los registros de los datos, se supone en otra posible causal del alto número de registros ND.