

ENTREGA FINAL PROYECTO FINAL

CAMILO BECERRA

La base de datos que estamos examinando ofrece una visión detallada de la población estudiantil en una institución educativa específica. Cada registro en la base de datos representa a un estudiante individual, con información que abarca tanto aspectos demográficos como académicos.

En el lado demográfico, los datos incluyen información como la edad, el género, el estrato socioeconómico y si el estudiante pertenece a alguna población vulnerable. Esto permite un análisis detallado de la composición de la población estudiantil y puede ayudar a identificar tendencias o patrones importantes.

En cuanto a los datos académicos, la base de datos proporciona detalles sobre el programa de estudios en el que está inscrito cada estudiante, la facultad a la que pertenece y el tipo de programa académico (ya sea una tecnología o un programa profesional). Estos datos pueden ser útiles para entender las tendencias en la elección de programas académicos y para identificar áreas de estudio populares.

Además, la base de datos también incluye información geográfica, como la ubicación del estudiante y el municipio de origen. Esta información puede ser útil para entender la distribución geográfica de los estudiantes y para identificar áreas con alta concentración de estudiantes.

En cuanto al análisis para las reglas de negocio tenemos las siguientes:

1. **Unicidad del ID del estudiante:** Cada estudiante se identifica por un ID único.
2. **Consistencia del periodo:** El periodo debe seguir el formato de un año seguido de ‘_1’ o ‘_2’, lo que indica el semestre.
3. **Formato del tipo de documento de identidad:** El tipo de documento de identidad debe ser ‘CC’, que representa la Cédula de Ciudadanía.
4. **Validación del código DANE:** El código DANE debe ser un número entero que corresponda a un municipio válido en Colombia.
5. **Correspondencia del municipio de origen:** El municipio de origen debe ser un nombre de municipio válido en Colombia.
6. **Verificación de la fecha de nacimiento:** La fecha de nacimiento debe ser una fecha válida y debe ser coherente con la edad del estudiante.

7. **Consistencia de la edad:** La edad del estudiante debe ser coherente con su fecha de nacimiento.
8. **Correspondencia del grupo etario:** El grupo etario debe ser consistente con la edad del estudiante.
9. **Formato de la vinculación:** La vinculación debe ser 'ANTIGUO' para los estudiantes que ya están matriculados.
10. **Formato del género:** El género debe ser 'M' (Masculino) o 'F' (Femenino).
11. **Rango del estrato:** El estrato debe ser un número entero entre 1 y 6.
12. **Validación del SNIES:** El SNIES debe ser un número entero que corresponda a un programa académico válido en Colombia.
13. **Correspondencia del plan de estudios:** El plan de estudios debe ser un nombre válido de plan de estudios.
14. **Correspondencia del programa académico:** El programa académico debe ser un nombre válido de programa académico.
15. **Correspondencia de la facultad:** La facultad debe ser un nombre válido de facultad.
16. **Correspondencia del lugar:** El lugar debe ser un nombre válido de lugar.
17. **Formato del tipo de programa:** El tipo de programa debe ser 'TECNOLOGÍA' o 'PROFESIONAL', dependiendo del plan de estudios.
18. **Formato de la longitud y la latitud:** La longitud y la latitud deben ser números flotantes que representen coordenadas geográficas válidas.
19. **Formato de la población vulnerable:** La población vulnerable debe ser una de las categorías predefinidas, como 'VICTIMA CONFLICTO', 'DESPLAZADO POR VIOLENCIA', 'EXCOMBATIENTE', 'NINGUNA', etc.

Ahora hablando de identificar y describir las entidades, atributos y relaciones que podemos presentar tenemos lo siguiente:

Entidades:

Persona: Esta entidad representa a un estudiante individual en la institución educativa.

Municipio: Esta entidad representa la ubicación geográfica de cada estudiante.

Programa: Esta entidad representa el programa académico en el que está inscrito cada estudiante.

Atributos:

Para la entidad Persona:

persona_id: Identificador único de la persona.

PERIODO: Periodo académico.

TIPO_DNI: Tipo de documento de identidad.

COD_DANE: Código DANE del municipio de origen.

FECHA_NACIMIENTO: Fecha de nacimiento de la persona.

EDAD: Edad de la persona.

GRUPO_ETARIO: Grupo etario al que pertenece la persona.

VINCULACION: Tipo de vinculación de la persona con la institución.

GENERO: Género de la persona.

ESTRATO: Estrato socioeconómico de la persona.

SNIES: Código SNIES del programa académico en el que está inscrito la persona.

LUGAR: Lugar de residencia de la persona.

TIPO: Tipo de programa académico en el que está inscrito la persona.

LONGITUD: Longitud de la ubicación de la persona.

LATITUD: Latitud de la ubicación de la persona.

POB_VULNERABLE: Indica si la persona pertenece a alguna población vulnerable.

Para la entidad Municipio:

municipio_id: Identificador único del municipio.

COD_DANE: Código DANE del municipio.

MUNICIPIO_ORIGEN: Nombre del municipio de origen.

Para la entidad Programa:

programa_id: Identificador único del programa.

SNIES: Código SNIES del programa.

PLAN_ESTUDIOS: Plan de estudios del programa.

PROGRAMA_ACADEMICO: Nombre del programa académico.

FACULTAD: Facultad a la que pertenece el programa.

Relaciones:

Persona-Municipio: Una persona puede tener un municipio de origen. Esta relación se representa mediante la clave foránea persona_id_fk en la tabla Municipio.

Persona-Programa: Una persona puede estar inscrita en un programa. Esta relación se representa mediante la clave foránea persona_id_fk en la tabla Programa.

Diagrama Entidad Relación

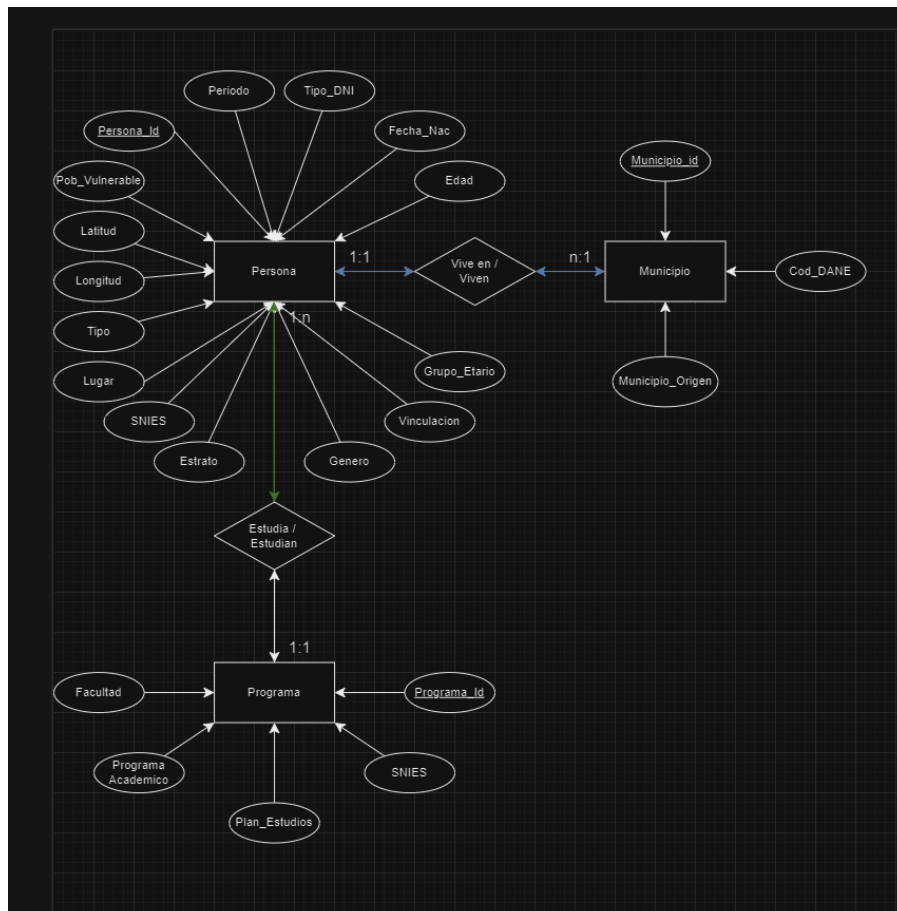
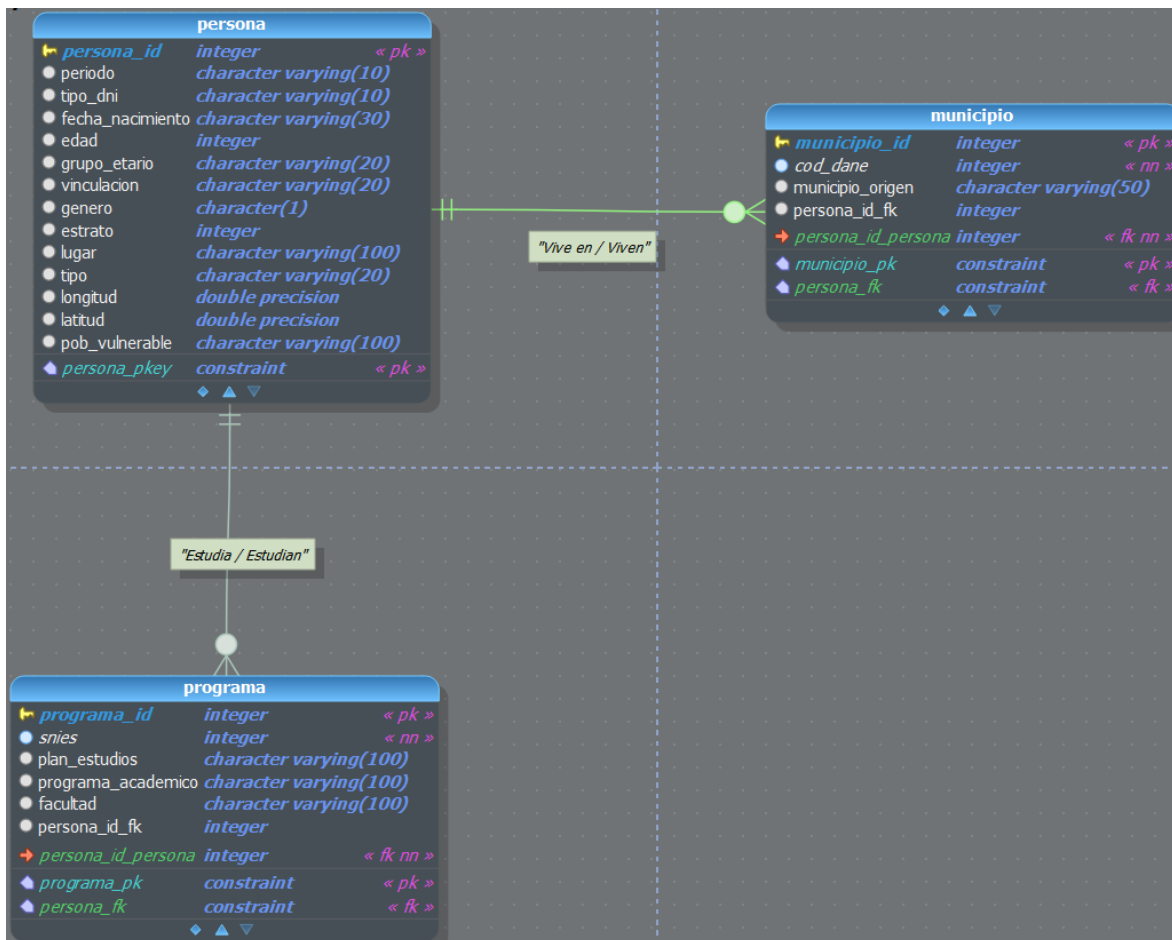


Diagrama Relacional



En ambos diagramas las relaciones de “Vive en / Viven” y “Estudia / Estudian”, corresponden a hablar que así como una persona solo puede vivir en un Municipio, un Municipio puede contener muchas personas que vivan en el según la Base de datos. Y que así como una persona estudia solo un programa, en un mismo programa pueden estudiar varias personas según la Base de datos.

Link a la base de datos:

https://www.datos.gov.co/Educaci-n/Estudiantes-matriculados-del-Polit-cnico-Colombian/uun2-9uuk/about_data

En cuanto al proceso de carga de información a la base de datos, lo que hice con estos nuevos datos fue:

Preparación de los Datos:

Primero, revisé los datos que iba a cargar. Realicé algunas correcciones menores, como eliminar caracteres especiales y asegurarme de que todos los datos estuvieran en el formato correcto CSV delimitado en este caso por “;”. Finalmente, separe los datos por archivos diferentes según los iba a usar para cargarlos en PostgreSQL.

PERSONA	ID	PERIODO	TIPO_DNI	FECHA_NACI	EDAD	GRUPO_ETAI	VINCULACIO	GENERO	ESTRATO	SNIES	LUGAR	TIPO	LONGITUD	LATITUD	POB_VULN	
	1	2023_1	CC	10/29/1980	1	43	39 a 45 a	ANTIGUO	M	2	108246	APARTADÁ	TECNOLOGÍA	-76633	7883	VICTIMA CI
	2	2023_1	CC	#####		22	15 a 26 a	ANTIGUO	M	2	3689	MEDELLÁN	PROFESIONA	-75567	6217	DESPLAZA
	3	2023_1	CC	01/27/2001	1	23	15 a 26 a	ANTIGUO	M	2	108727	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	4	2023_1	CC	02/18/1988	1	35	27 a 38 a	ANTIGUO	F	2	2540	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	5	2023_1	CC	03/19/1978	1	45	39 a 45 a	ANTIGUO	M	2	2937	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	6	2023_1	CC	#####		20	15 a 26 a	ANTIGUO	M	2	2540	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	7	2023_1	CC	03/31/1984	1	39	39 a 45 a	ANTIGUO	M	2	3689	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	8	2023_1	CC	#####		40	39 a 45 a	ANTIGUO	M	2	3689	MEDELLÁN	PROFESIONA	-75567	6217	VICTIMA CI
	9	2023_1	CC	03/27/1988	1	35	27 a 38 a	ANTIGUO	F	3	3347	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	10	2023_1	CC	10/29/1982	1	41	39 a 45 a	ANTIGUO	M	3	108727	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	11	2023_1	CC	07/21/1983	1	40	39 a 45 a	ANTIGUO	M	2	108727	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	12	2023_1	CC	07/14/1980	1	43	39 a 45 a	ANTIGUO	M	2	2937	MEDELLÁN	PROFESIONA	-75567	6217	EXCOMBAT
	13	2023_1	CC	09/20/1986	1	37	27 a 38 a	ANTIGUO	M	2	2937	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	14	2023_1	CC	#####		42	39 a 45 a	ANTIGUO	F	2	2540	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	15	2023_1	CC	05/29/1981	1	42	39 a 45 a	ANTIGUO	F	2	2540	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	16	2023_1	CC	09/20/1979	1	44	39 a 45 a	ANTIGUO	M	2	2859	MEDELLÁN	TECNOLOGÍA	-75567	6217	VICTIMA CI
	17	2023_1	CC	03/22/1982	1	41	39 a 45 a	ANTIGUO	F	3	3347	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	18	2023_1	CC	10/20/1986	1	37	27 a 38 a	ANTIGUO	F	2	1749	MEDELLÁN	TECNOLOGÍA	-75567	6217	NINGUNA
	19	2023_1	CC	#####		38	27 a 38 a	ANTIGUO	F	3	2937	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	20	2023_1	CC	12/14/1978	1	45	39 a 45 a	ANTIGUO	M	3	108270	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	21	2023_1	CC	07/18/1984	1	39	39 a 45 a	ANTIGUO	M	3	2937	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	22	2023_1	CC	09/16/1988	1	35	27 a 38 a	ANTIGUO	F	2	2937	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA
	23	2023_1	CC	#####		43	39 a 45 a	ANTIGUO	F	2	107648	APARTADÁ	PROFESIONA	-76633	7883	NINGUNA
	24	2023_1	CC	06/14/1984	1	39	39 a 45 a	ANTIGUO	M	3	110950	MEDELLÁN	TECNOLOGÍA	-75567	6217	NINGUNA
	25	2023_1	CC	#####		39	39 a 45 a	ANTIGUO	M	3	2937	MEDELLÁN	PROFESIONA	-75567	6217	NINGUNA

PROGRAMA	SNIES	PLAN_ESTUD	PROGRAMA	FACULTAD	PERSONA_ID_FK	
1	108246	6910	TECNOLOGÍA	FACULTAD D	1	
2	3689	9511	INGENIERÍA	FACULTAD D	2	
3	108727	7510	LICENCIATUR	FACULTAD D	3	
4	2540	9011	CONTADURÍA	FACULTAD D	4	
5	2937	9209	INGENIERÍA	FACULTAD D	5	
6	2540	9011	CONTADURÍA	FACULTAD D	6	
7	3689	9511	INGENIERÍA	FACULTAD D	7	
8	3689	9511	INGENIERÍA	FACULTAD D	8	
9	3347	8310	INGENIERÍA	FACULTAD D	9	
10	108727	7510	LICENCIATUR	FACULTAD D	10	
11	108727	7510	LICENCIATUR	FACULTAD D	11	
12	2937	9209	INGENIERÍA	FACULTAD D	12	
13	2937	9209	INGENIERÍA	FACULTAD D	13	
14	2540	9010	CONTADURÍA	FACULTAD D	14	
15	2540	9010	CONTADURÍA	FACULTAD D	15	
16	2859	210	TECNOLOGÍA	FACULTAD D	16	
17	3347	8310	INGENIERÍA	FACULTAD D	17	
18	1749	309	TECNOLOGÍA	FACULTAD D	18	
19	2937	9209	INGENIERÍA	FACULTAD D	19	
20	108270	8710	INGENIERÍA	FACULTAD D	20	
21	2937	9209	INGENIERÍA	FACULTAD D	21	
22	2937	9209	INGENIERÍA	FACULTAD D	22	
23	107648	8210	INGENIERÍA	FACULTAD D	23	
24	110950	7810	TECNOLOGÍA	FACULTAD D	24	
25	2937	9209	INGENIERÍA	FACULTAD D	25	
26	55167	2011	COMUNICAC	FACULTAD D	26	
27	2714	1110	TECNOLOGÍA	FACULTAD D	27	
28	3348	8210	INGENIERÍA	FACULTAD D	28	
29	2937	9209	INGENIERÍA	FACULTAD D	29	

MUNICIPIO	COD_DANE	MUNICIPIO	PERSONA_ID_FK
1	5147	CAREPA	1
2	5088	BELLO	2
3	5125	CAICEDO	3
4	5001	MEDELLIN	4
5	5001	MEDELLIN	5
6	5079	BARBOSA	6
7	5001	MEDELLIN	7
8	5665	SAN PEDRO I	8
9	5001	MEDELLIN	9
10	5266	ENVIGADO	10
11	5360	ITAGUI	11
12	5380	LA ESTRELLA	12
13	5001	MEDELLIN	13
14	5001	MEDELLIN	14
15	27075	BAHIA SOLAI	15
16	5001	MEDELLIN	16
17	5001	MEDELLIN	17
18	5360	ITAGUI	18
19	5001	MEDELLIN	19
20	5380	LA ESTRELLA	20
21	5001	DESCONOCIE	21
22	5001	MEDELLIN	22
23	5147	CAREPA	23
24	5001	MEDELLIN	24
25	5088	BELLO	25
26	5266	ENVIGADO	26
27	5001	MEDELLIN	27
28	5001	MEDELLIN	28
29	5001	MEDELLIN	29

Creación de la Tabla:

Después de preparar los datos, creé tablas en la base de datos para almacenarlos. Definí cada columna de la tabla para que coincidiera con los tipos de datos y la estructura de los datos a cargar.

```
CREATE TABLE Persona (
    persona_id integer,
    PERIODO VARCHAR(10),
    TIPO_DNI VARCHAR(10),
    FECHA_NACIMIENTO VARCHAR(30),
    EDAD INTEGER,
```



```

GRUPO_ETARIO VARCHAR(20),
VINCULACION VARCHAR(20),
GENERO CHAR(1),
ESTRATO INTEGER,
SNIES INTEGER,
LUGAR VARCHAR(100),
TIPO VARCHAR(20),
LONGITUD FLOAT,
LATITUD FLOAT,
POB_VULNERABLE VARCHAR(100),
        PRIMARY KEY (persona_id)
);

```

```

CREATE TABLE Municipio (
    municipio_id INTEGER,
    COD_DANE INTEGER NOT NULL,
    MUNICIPIO_ORIGEN VARCHAR(50),
    persona_id_fk INTEGER,
    CONSTRAINT municipio_pk PRIMARY KEY (municipio_id),
    CONSTRAINT persona_id_fk FOREIGN KEY(persona_id_fk) REFERENCES Persona(persona_id)
);

```

```

CREATE TABLE Programa (
    programa_id INTEGER,
    SNIES INTEGER NOT NULL,
    PLAN_ESTUDIOS VARCHAR(100),
    PROGRAMA_ACADEMICO VARCHAR(100),
    FACULTAD VARCHAR(100),
    persona_id_fk INTEGER,
    CONSTRAINT programa_pk PRIMARY KEY (programa_id),
    CONSTRAINT persona_id_fk FOREIGN KEY(persona_id_fk) REFERENCES Persona(persona_id)
);

```

Carga de Datos:

Una vez que la tabla estuvo lista, procedí a cargar los datos. Utilicé una herramienta de administración de bases de datos para cargar el archivo CSV que contenía los datos. Solamente tuve que especificar la ubicación del archivo y la tabla de destino.

```
-- Cargar datos en la tabla "Persona"
COPY Persona (PERSONA_ID, PERIODO, TIPO_DNI, FECHA_NACIMIENTO, EDAD, GRUPO_ETARIO, VINCULACION, GENERO, ESTRATO)
FROM 'D:\Descargas\Estudiantes_Persona.csv' DELIMITER ';' CSV HEADER;

COPY Municipio (MUNICIPIO_ID, COD_DANE, MUNICIPIO_ORIGEN, PERSONA_ID_FK)
FROM 'D:\Descargas\Estudiantes_Municipio.csv' DELIMITER ';' CSV HEADER;

-- Cargar datos en la tabla "Programa"
COPY Programa (PROGRAMA_ID, SNIES, PLAN_ESTUDIOS, PROGRAMA_ACADEMICO, FACULTAD, PERSONA_ID_FK)
FROM 'D:\Descargas\Estudiantes_Programa.csv' DELIMITER ';' CSV HEADER;
```

Validación de Datos:

Después de la carga, verifiqué que los datos se hubieran cargado correctamente. Ejecuté algunas consultas simples para asegurarme de que todos los datos estuvieran en su lugar y que no hubiera errores obvios.

select * from Persona

	persona_id [PK] integer	periodo character varying (10)	tipo_dni character varying (10)	fecha_nacimiento character varying (30)	edad integer	grupo_etario character varying (20)	vinculacion character varying (20)	genero character	estrato integer	snies integer	lugar character varying (100)
1	1	2023_1	CC	10/29/1980 12:00:00 AM	43	39 a 45 años	ANTIGUO	M	2	108246	APARTADÓ
2	2	2023_1	CC	5/10/2001 0:00	22	15 a 26 años	ANTIGUO	M	2	3689	MEDELLÍN
3	3	2023_1	CC	01/27/2001 12:00:00 AM	23	15 a 26 años	ANTIGUO	M	2	108727	MEDELLÍN
4	4	2023_1	CC	02/18/1988 12:00:00 AM	35	27 a 38 años	ANTIGUO	F	2	2540	MEDELLÍN
5	5	2023_1	CC	03/19/1978 12:00:00 AM	45	39 a 45 años	ANTIGUO	M	2	2937	MEDELLÍN
6	6	2023_1	CC	3/11/2003 0:00	20	15 a 26 años	ANTIGUO	M	2	2540	MEDELLÍN
7	7	2023_1	CC	03/31/1984 12:00:00 AM	39	39 a 45 años	ANTIGUO	M	2	3689	MEDELLÍN
8	8	2023_1	CC	4/07/1983 0:00	40	39 a 45 años	ANTIGUO	M	2	3689	MEDELLÍN
9	9	2023_1	CC	03/27/1988 12:00:00 AM	35	27 a 38 años	ANTIGUO	F	3	3347	MEDELLÍN
10	10	2023_1	CC	10/29/1982 12:00:00 AM	41	39 a 45 años	ANTIGUO	M	3	108727	MEDELLÍN
11	11	2023_1	CC	07/21/1983 12:00:00 AM	40	39 a 45 años	ANTIGUO	M	2	108727	MEDELLÍN
12	12	2023_1	CC	07/14/1980 12:00:00 AM	43	39 a 45 años	ANTIGUO	M	2	2937	MEDELLÍN

select * from Municipio

	municipio_id [PK] integer	cod_dane integer	municipio_origen character varying (50)	persona_id_fk integer
1	1	5147	CAREPA	1
2	2	5088	BELLO	2
3	3	5125	CAICEDO	3
4	4	5001	MEDELLIN	4
5	5	5001	MEDELLIN	5
6	6	5079	BARBOSA	6
7	7	5001	MEDELLIN	7
8	8	5665	SAN PEDRO DE URABA	8
9	9	5001	MEDELLIN	9
10	10	5266	ENVIGADO	10
11	11	5360	ITAGUI	11
12	12	5380	LA ESTRELLA	12
13	13	5001	MEDELLIN	13
14	14	5001	MEDELLIN	14

select * from Programa

	programa_id [FK] integer	snies integer	plan_estudios character varying (100)	programa_academico character varying (100)	facultad character varying (100)	persona_id integer
1	1	108246	6910	TECNOLOGÍA EN DESARROLLO DE SOFTWARE	FACULTAD DE INGENIERÍA	
2	2	3689	9511	INGENIERÍA CIVIL	FACULTAD DE INGENIERÍA	
3	3	108727	7510	LICENCIATURA EN EDUCACIÓN FÍSICA, RECREACIÓN Y DEPORTE	FACULTAD DE EDUCACIÓN FÍSICA, RECREACIÓN Y DEPORTE	
4	4	2540	9011	CONTADURÍA PÚBLICA	FACULTAD DE ADMINISTRACIÓN	
5	5	2937	9209	INGENIERÍA DE PRODUCTIVIDAD Y CALIDAD	FACULTAD DE ADMINISTRACIÓN	
6	6	2540	9011	CONTADURÍA PÚBLICA	FACULTAD DE ADMINISTRACIÓN	
7	7	3689	9511	INGENIERÍA CIVIL	FACULTAD DE INGENIERÍA	
8	8	3689	9511	INGENIERÍA CIVIL	FACULTAD DE INGENIERÍA	
9	9	3347	8310	INGENIERÍA EN HIGIENE Y SEGURIDAD OCUPACIONAL	FACULTAD DE INGENIERÍA	
10	10	108727	7510	LICENCIATURA EN EDUCACIÓN FÍSICA, RECREACIÓN Y DEPORTE	FACULTAD DE EDUCACIÓN FÍSICA, RECREACIÓN Y DEPORTE	
11	11	108727	7510	LICENCIATURA EN EDUCACIÓN FÍSICA, RECREACIÓN Y DEPORTE	FACULTAD DE EDUCACIÓN FÍSICA, RECREACIÓN Y DEPORTE	
12	12	2937	9209	INGENIERÍA DE PRODUCTIVIDAD Y CALIDAD	FACULTAD DE ADMINISTRACIÓN	
13	13	2937	9209	INGENIERÍA DE PRODUCTIVIDAD Y CALIDAD	FACULTAD DE ADMINISTRACIÓN	

✓ Successfully run. Total query runtime: 71 msec. 3008 rows affected. ✕

Análisis identificados

Segmentación demográfica y sociodemográfica: Se sugiere agrupar a los estudiantes según variables como edad, género, estrato socioeconómico y lugar de origen. Esta segmentación permitiría comprender mejor la composición de nuestros estudiantes y adaptar nuestros programas según sus necesidades específicas.

Evaluación de la vulnerabilidad y necesidades especiales: Se propone examinar cómo la condición de "población vulnerable" se relaciona con otras características de los estudiantes, como su lugar de origen o programa académico. Este enfoque permitiría identificar grupos específicos que podrían requerir apoyo adicional o intervenciones específicas.

Análisis de la movilidad y migración estudiantil: Se plantea investigar los patrones de movilidad y migración de los estudiantes, especialmente en relación con su lugar de origen y lugar de estudio. Este análisis nos ayudaría a comprender mejor cómo se desplazan los estudiantes y qué factores influyen en sus decisiones educativas.

<https://github.com/CamiloBecerra1606/Proyecto-Ingenieria-Datos-Camilo-Becerra>