

Grupo 1:

Juana Mejía 202021512

Daniela Espinosa 202022615

Pablo Ortega 202021700

## **AVANCE TÉCNICO PROYECTO MULTIDISCIPLINARIO**

### **1. Reporte de entendimiento de los datos**

#### Tipos de categorías:

COD\_DPTO (Código del departamento): categórica

COD\_MUNIC (Código del municipio): categórica

AREANAC (Área de nacimiento): categórica

SIT\_PARTO (Sitio de parto): categórica

OTRO\_SIT (Otro sitio del parto, ¿cuál?): categórica

SEXO (Sexo del nacido): categórica

PESO\_NAC (Rango de peso del nacido): categórica

TALLA\_NAC (Rango de talla del nacido): categórica

ANO (Año nacimiento): numérica

MES (Mes nacimiento): numérica

T\_GES (Rango de tiempo gestación): categórica

NUMCONSUL (Número consultas prenatales): numérica

TIPO\_PARTO (Tipo de parto): categórica

MUL\_PARTO (Multiplicidad del parto): categórica

IDHEMOCLAS (Grupo sanguíneo del nacido): categórica

IDFACTORRH (RH del nacido): categórica

EDAD\_MADRE (Rango de edad de la madre): categórica

EST\_CIVM (Estado civil de la madre): categórica

NIV\_EDUM (Nivel de educación de la madre): categórica

CODPTORE (Departamento de residencia habitual de la madre): categórica

CODMUNRE (Municipio de residencia habitual de la madre): categórica

AREA\_RES (Área de residencia habitual de la madre): categórica

FECHA\_NACM (Fecha de nacimiento del anterior hijo): fecha (numérica)

N\_EMB (Número de embarazos): numérica

SEG\_SOCIAL (Tipo de seguridad social): categórica

APGAR1 (APGAR al minuto de nacido): categórica

APGAR2 (APGAR a los cinco minutos de nacido): categórica

Registros duplicados: 9994

Número y porcentaje de valores faltantes:

OTRO\_SIT (Otro sitio del parto, ¿cuál?): 2490984 (99.78%)

CODPTORE (Departamento de residencia habitual de la madre): 20586 (0.824%)

CODMUNRE (Municipio de residencia habitual de la madre): 20587 (0.824%)

AREA\_RES (Área de residencia habitual de la madre): 20409 (0.818%)

FECHA\_NACM (Fecha de nacimiento del anterior hijo): 1152929 (46.184%)

Número y porcentaje de outliers para variables numéricas:

PESO\_NAC (Peso del nacido): 157743 | 6.32%

TALLA\_NAC (Talla del nacido): 22620 | 0.91%

T\_GES (Tiempo de gestación): 579376 | 23.21%

NUMCONSUL (Número consultas prenatales): 134386 | 5.38%

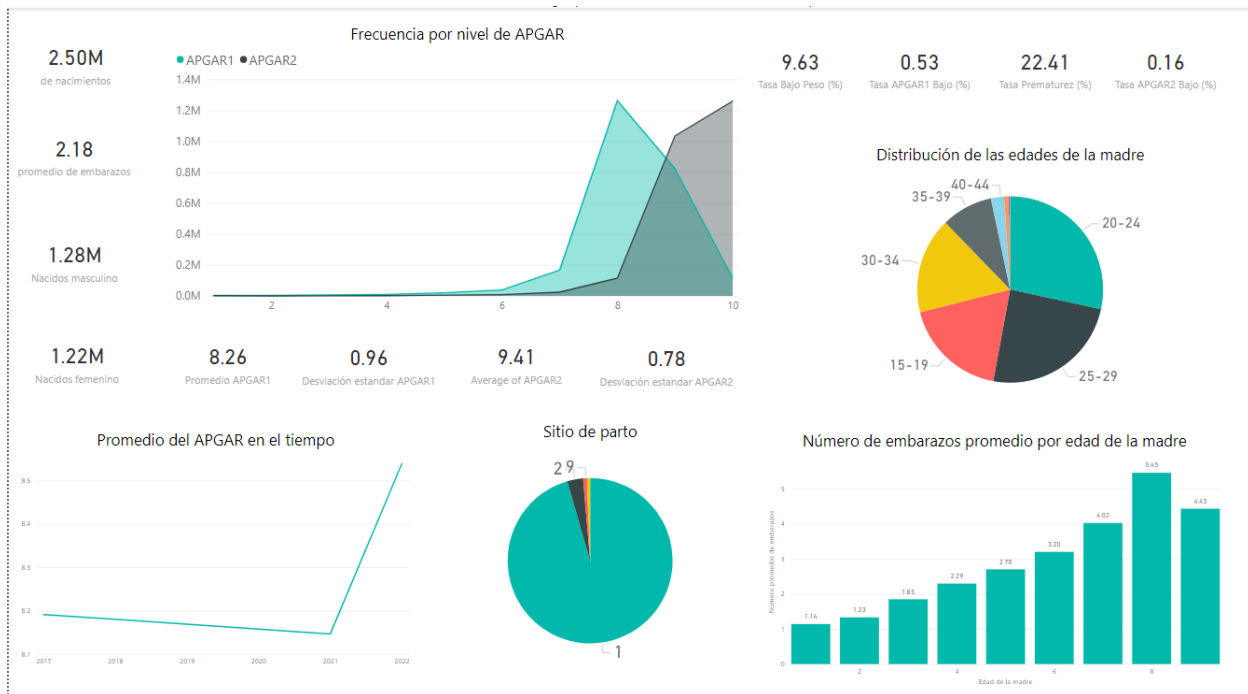
MUL\_PARTO (Multiplicidad del parto): 46308 | 1.86%

EDAD\_MADRE (Rango de edad de la madre): 1791 | 0.07%

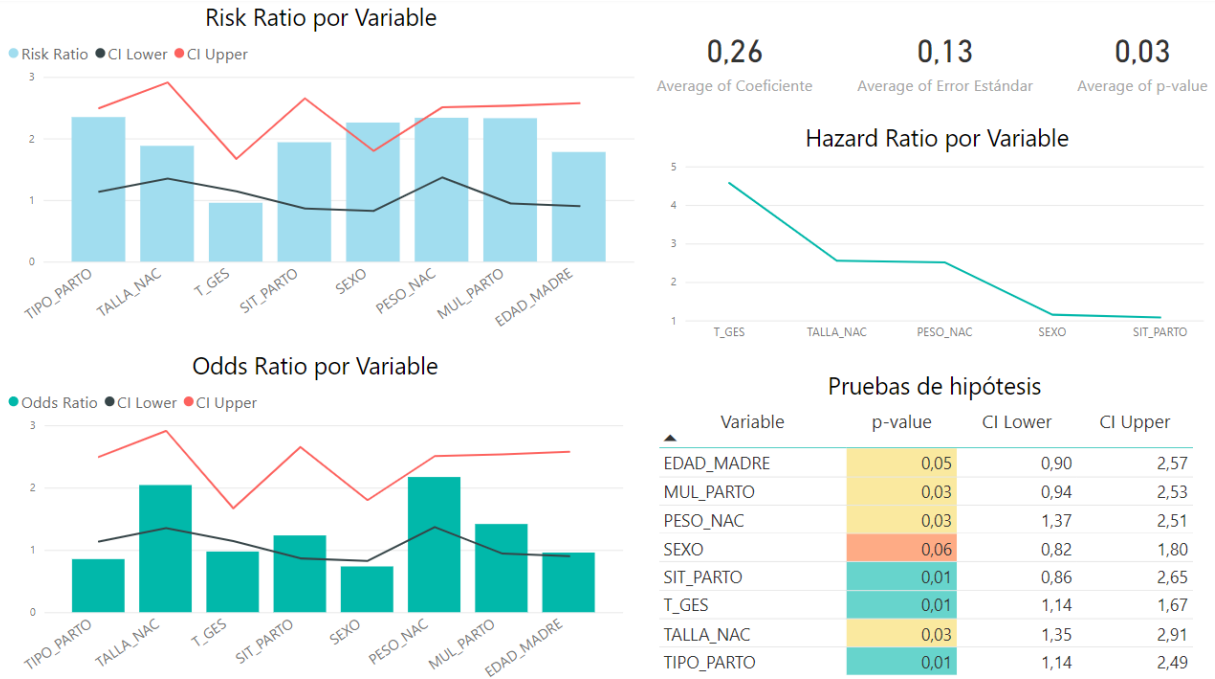
N\_EMB (Número de embarazos): 34556 | 1.38%

## 2. Propuesta de los bocetos asociados a los tableros de control que van a desarrollar:

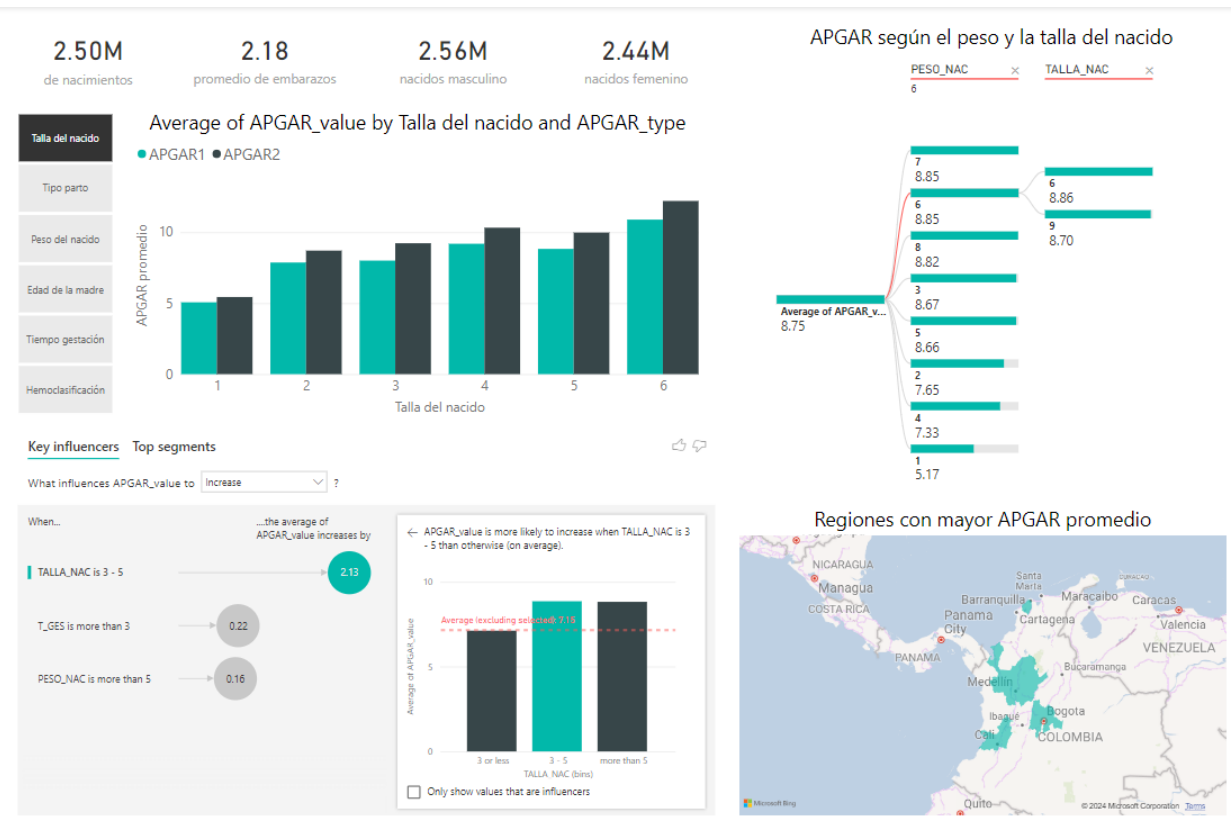
### 1. Análisis descriptivo



## 2. Inferencia estadística



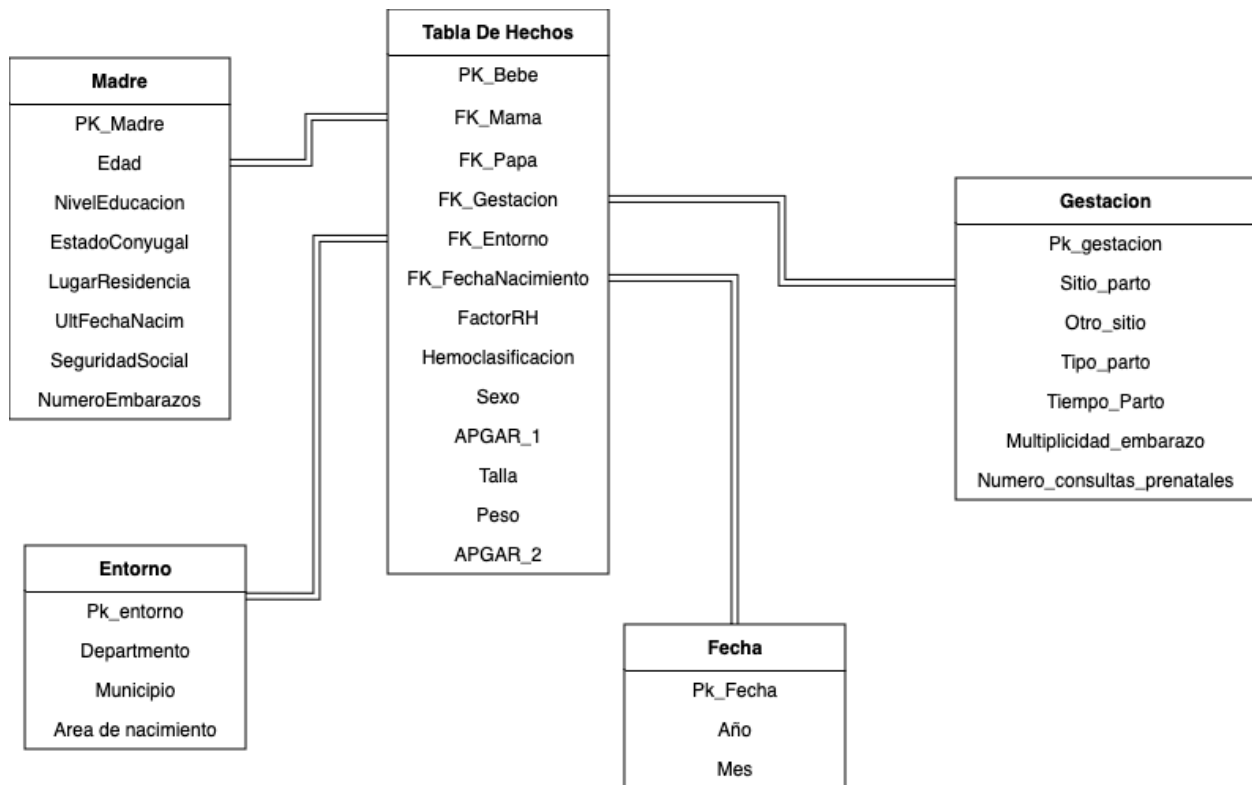
## 3. Análisis clínico



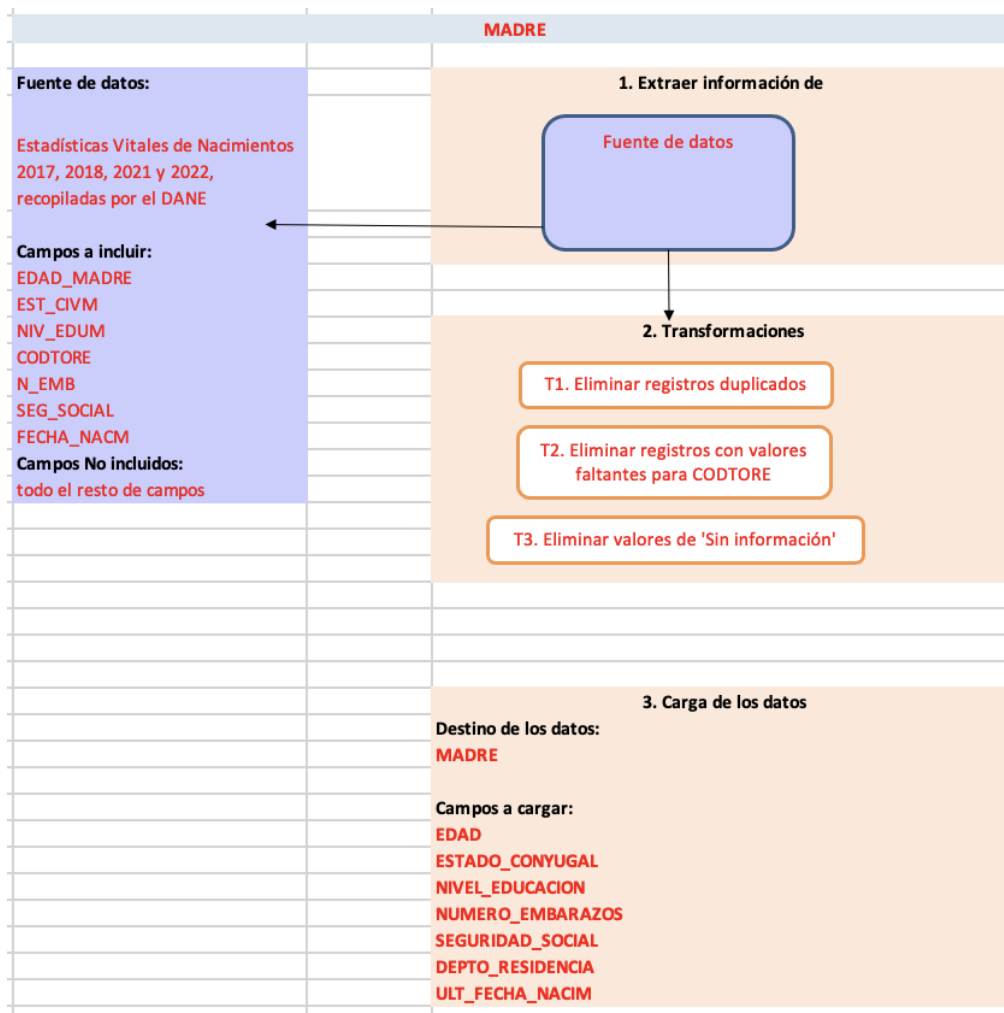
### 3. Modelo multidimensional

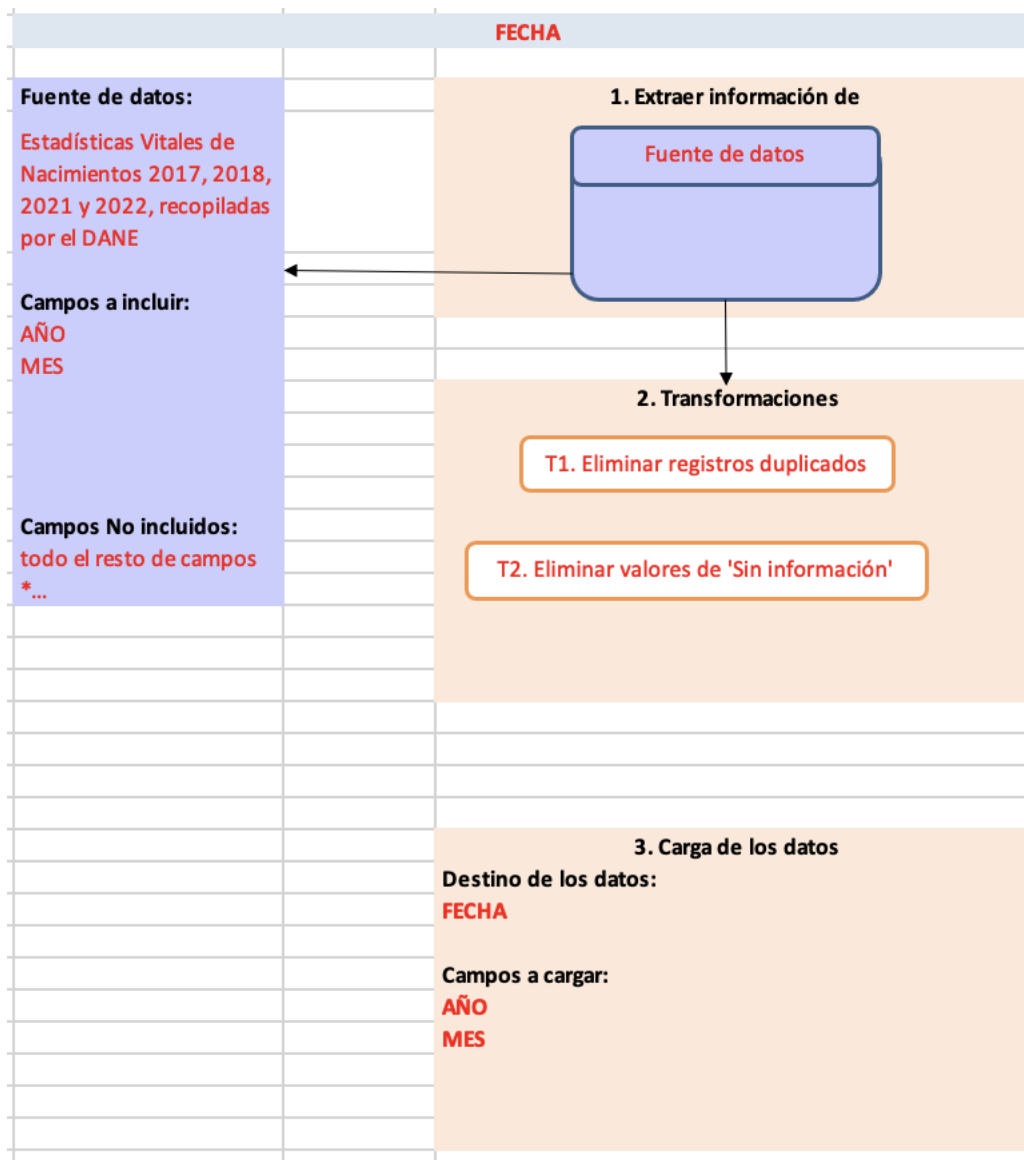
1. Proceso a modelar: Nacimiento de un bebe
2. Grano: El nacimiento de un bebe en una fecha, en un departamento.
3. Dimensiones: Madre, padre, gestación, entorno, fecha, bebe
4. Hechos: Talla, peso, apgar\_1, apgar\_2

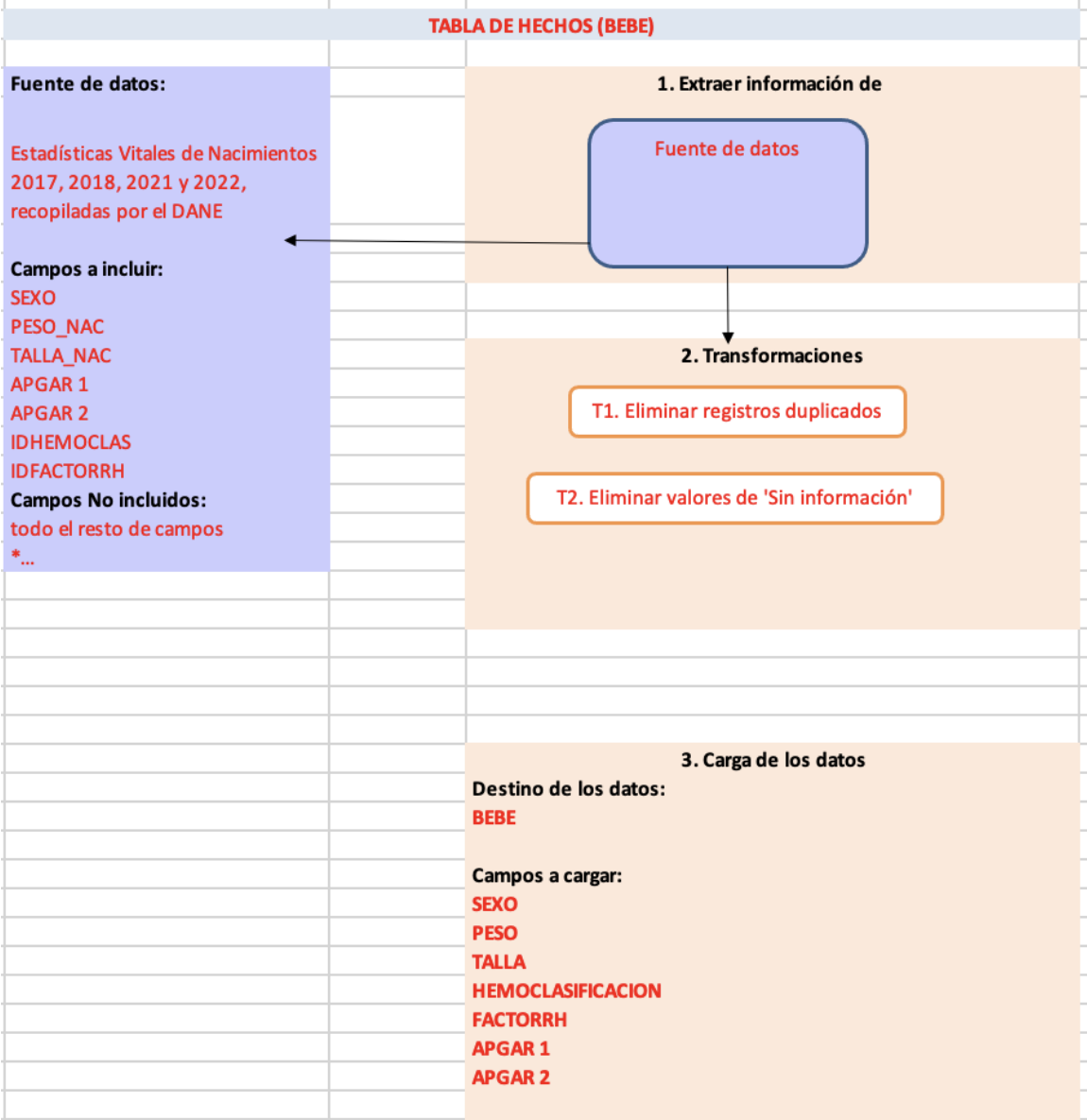
Para poder cumplir con los requerimientos que se obtienen de los objetivos se desarrolló el siguiente modelo. Se decide tener talla, peso y las calificaciones de APGAR debido a que se quieren encontrar cifras sobre como cada dimensión afecta al bebe, sean promedios, porcentajes u otras operaciones.



#### 4. Diseño del proceso ETL







## GESTACION

**Fuente de datos:**

Estadísticas Vitales de Nacimientos  
2017, 2018, 2021 y 2022,  
recopiladas por el DANE

**Campos a incluir:**

SIT\_PARTO  
OTRO\_SITIO  
TIPO\_PARTO  
T\_GES  
NUMCONSUL  
MUL\_PARTO

**Campos No incluidos:**

todo el resto de campos

\*  
...

### 1. Extraer información de

Fuente de datos

## 2. Transformaciones

### T1. Eliminar registros duplicados

## T2. Eliminar registros con valores faltantes para OTRO\_SIT

### T3. Eliminar valores de 'Sin información'

### 3. Carga de los datos

**Destino de los datos:**

## GESTACIÓN

**Campos a cargar:**

\* NombreCampo1

SITIO PARTO

**OTRO SITIO**

TIPO\_PARTO

## TIEMPO PARTO

NUMERO\_CONSULTAS\_PRENATALES

MULTIPLICIDAD\_EMBARAZO



