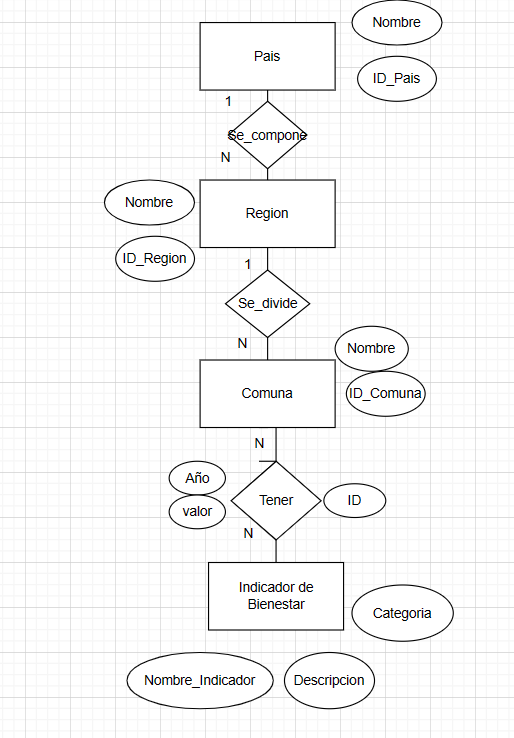
**Parte B: Información territorial basada en las comunas de Chile**

**Integrantes:** Cristóbal Rebolledo

Benjamín Parra

Carolina Obreque

Modelo Entidad-Relación:



Modelo Relacional:

Pais(PK\_id\_pais,Nombre)

Region(PK\_ID\_Region, Nombre, FK\_ID\_Pais)

Comuna(PK\_ID\_Comuna, Nombre, FK\_ID\_region)

Tener(PK\_ID, Año, Valor, FK\_ID\_comuna, FK\_nombre\_indicador)

Indicador de Bienestar(PK\_Nombre\_indicador, Descripcion, Categoria)

Diccionario de Datos:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nombre de tabla | | Pais |  |  |  |
| **pk** | **Atributo** | **Tipo de dato** | **Tamaño** | **fk** | **tabla** |
| x | ID\_Pais | SMALL INT | 6 |  |  |
|  | Nombre | CHAR | 30 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nombre de tabla | | Region |  |  |  |
| **pk** | **Atributo** | **Tipo de dato** | **Tamaño** | **fk** | **tabla** |
| x | ID\_Region | SMALL INT | 6 |  |  |
|  | Nombre | VARCHAR | 100 |  |  |
|  | ID\_Pais | SMALL INT | 6 | x | Pais |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nombre de tabla | | Comuna |  |  |  |
| **pk** | **Atributo** | **Tipo de dato** | **Tamaño** | **fk** | **tabla** |
| x | ID\_Comuna | SMALL INT | 6 |  |  |
|  | Nombre | VARCHAR | 100 |  |  |
|  | ID\_Region | SMALL INT | 6 | x | Region |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nombre de tabla | | Tener |  |  |  |
| **pk** | **Atributo** | **Tipo de dato** | **Tamaño** | **fk** | **tabla** |
| **x** | **ID** | **SMALL INT** | **6** |  |  |
|  | Año | YEAR | 4 |  |  |
|  | Valor | SMALL INT | 1000000 |  |  |
|  | ID\_Comuna | SMALL INT | 6 | x | Comuna |
|  | Nombre\_indicador | VARCHAR | 100 | x | Indicador de bienestar |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nombre de tabla | | Indicador de Bienestar | |  |  |
| **pk** | **Atributo** | **Tipo de dato** | **Tamaño** | **fk** | **tabla** |
| x | Nombre\_Indicador | VARCHAR | 100 |  |  |
|  | Descripcion | VARCHAR | 300 |  |  |
|  | Categoria | VARCHAR | 100 |  |  |

Variables a utilizar

Las variables que vamos a utilizar en nuestra base de datos son entretención, trabajo, salud, educación y seguridad. De cada una de estas variables guardaremos su nombre, descripción del lugar, como ha cambiado al pasar los años y su valor. Las elegimos porque son importantes para cada comuna ya que si un usuario desea informarse sobre una región específica querrá buscar información en un sitio simple y que abarque la información más importante, como son las variables que ocupamos, en donde sus datos específicos será la cantidad de locales, sus direcciones, disponibilidad de trabajo, la seguridad, educación para la familia, etc. Entonces al saber todos estos datos relevantes de cada comuna, podrá decidir fácilmente sobre qué comuna le conviene más vivir gracias a la ayuda de nuestra base de datos.

Para métrica de entretención, nos dispusimos a calcular la cantidad de estadios por comuna, para obtener los datos accedimos a el siguiente link: [Anexo:Estadios de fútbol de Chile - Wikipedia, la enciclopedia libre](https://es.wikipedia.org/wiki/Anexo:Estadios_de_f%C3%BAtbol_de_Chile). Gracias a que los datos ya estaban correctamente tabulados, basto con una simple importación con Excel para obtener la tablas, que fueron Estadios y Otros estadios. Ahora con los datos en la planilla, el mismo Excel nos permitió eliminar las columnas de mas y anexar todo, además de eliminar los paréntesis para poder tener el nombre de la comuna limpio en cada fila. Guardamos el archivo como un csv y usamos un pequeño programa de Python que llamamos quita.py que le da un formato correcto al texto. El código es el siguiente:

import unicodedata

import codecs

def remove\_accents(text):

    return ''.join(c for c in unicodedata.normalize('NFD', text) if unicodedata.category(c) != 'Mn').replace('"',"").replace("'","")

file = codecs.open("estadios\_Salida.csv","r","utf-8-sig")

text = "".join(file.readlines())

output = codecs.open("estadios\_Salida\_SIN\_TILDES.csv","w","utf-8-sig")

output.write(remove\_accents(text))

file.close()

output.close()

Este código normaliza los caracteres UTF-8 y elimina las comillas. Guardamos el archivo de salida como estadios\_Salida\_SIN\_TILDES.csv.

Como métrica de seguridad, quisimos contabilizar los delitos DMCS (Delitos de mayor connotación social) tales como homicidio, lesiones, violación, robo con violencia, etc. Para poder hacer una correcta comparación tomaremos la tasa de incidencia cada 100 000 habitantes, de esta manera podemos comparar varias comunas sin importar su cantidad de habitantes.

Extrajimos la información de la plataforma [pazciudadana.cl](https://datoscomunales.pazciudadana.cl/). Desde la pestaña de Ranking comunal se nos proporciona un acceso directo a estas estadísticas a través de un documento Excel. Lo descargamos directamente y hicimos un proceso similar al anterior:

1. Eliminamos las columnas que no aportaban a la investigación
2. Guardamos el archivo usando una extensión csv
3. Ejecutamos quita.py para normalizar el texto, eliminando tildes y comillas
4. Ahora tenemos un archivo csv listo para ser leído.