



Trabajo sobre información territorial basada en las comunas de Chile

Base de datos, Tarea B

Autores: Martín Suárez, Luis Berrocal

Profesor: Luis Veas

Introducción:

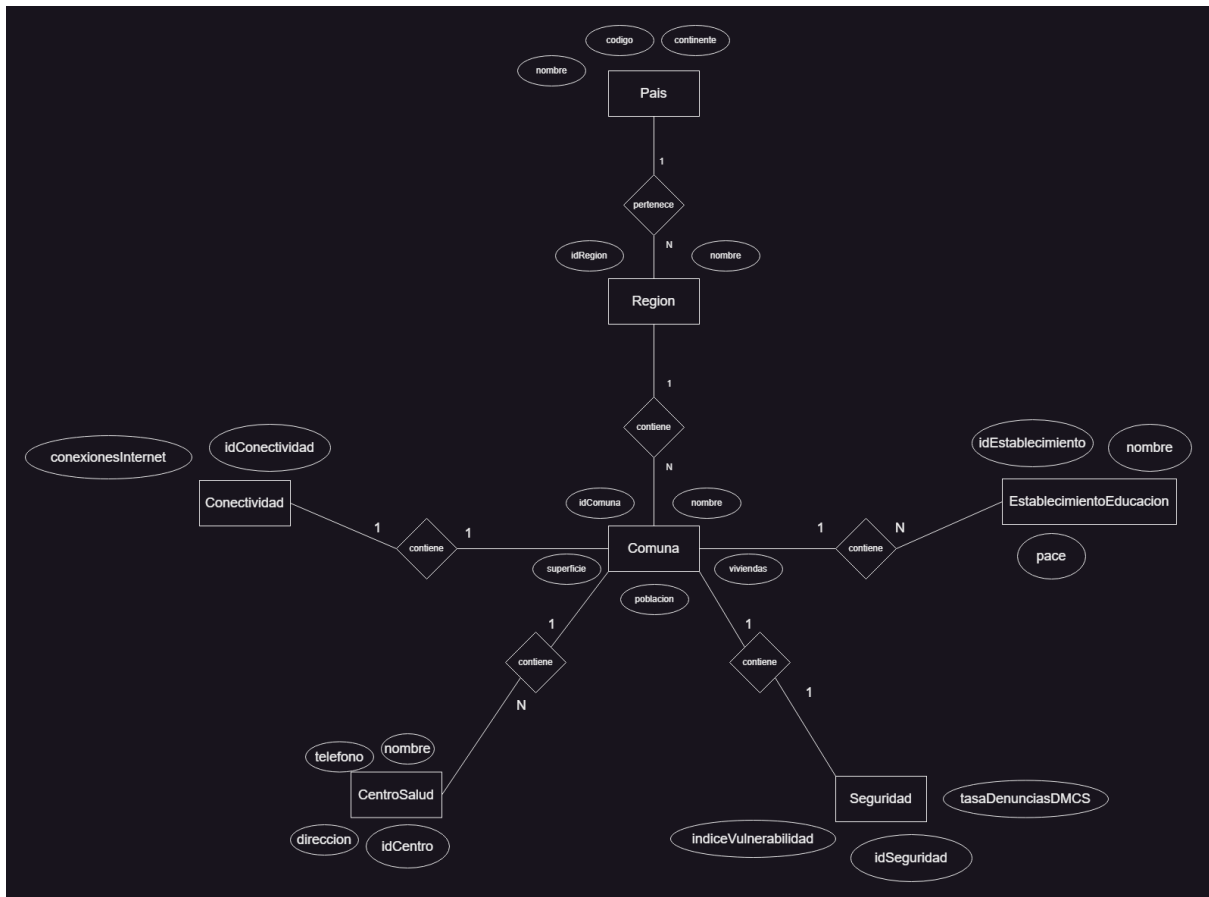
En el marco del estudio y creación de base de datos para el presente trabajo llevaremos a cabo la creación de una base de datos que contenga información valiosa para evaluar el “bienestar”/”calidad” de las diferentes comunas de Chile. El proyecto consiste en seleccionar 4 variables a estudiar por cada comuna y en base a estos aspectos generar la base de datos y posteriormente una serie de consultas SQL para obtener métricas que nos permitan comparar las diferentes comunas en base a un parámetro de “bienestar”/”calidad de vida” propuesta por el grupo. A lo largo del presente informe se irán detallando los diferentes procesos y decisiones que se fueron tomando durante la realización del proyecto.

Metodología:

Para el trabajo se llevó a cabo la creación de un modelo entidad de relación donde modelamos nuestra primera visión de cómo debería ser la base de datos. En este modelo fuimos seleccionando los diferentes aspectos de cada comuna en los que nos basamos para el estudio, y por cada aspecto definiendo uno o varios atributos/datos que como grupo nos interesaban para emplear la evolución de las diferentes comunas.

El principal contratiempo que encontramos fue que cada vez que nos decidimos por una variable de estudio, los pocos datos que encontrábamos en las páginas o tablas siempre estaban caídas o no guardaban los datos por comuna. Múltiples fueron los archivos que logramos encontrar en páginas “oficiales” como por ejemplo “datos.gob.cl” y que los archivos o las redirecciones estuviesen caídas. Tras una ardua búsqueda y reestructuración de nuestro modelo inicial decidimos que los aspectos a estudiar serían Salud, Educación, Conectividad y Seguridad. Por cada uno de estos aspectos seleccionamos variables que corresponden a la información que encontramos por cada comuna de cada. La selección y búsqueda de los datos estará descrita más adelante.

Modelo Entidad-relación:



Modelo relacional:

Pais(PK_codigo, nombre, continente)

Region(PK_idRegion, nombre, FK_codigoPais)

Comuna(PK_idComuna, nombre, poblacion, superficie, viviendas, FK_idRegion, FK_idConectividad, FK_idSeguridad)

EstablecimientoEducacion(PK_idEstablecimiento, nombre, PACE, FK_idComuna)

Seguridad(PK_idSeguridad, tasaDenunciasDMCS, indiceVulnerabilidad, FK_idComuna)

CentroSalud(PK_idCentro, nombre, teléfono, dirección, FK_idComuna)

Conectividad(PK_idConectividad, conexionesInternet, FK_idComuna)

Siguiendo este modelo construimos el siguiente diccionario para la base de datos:

Tabla	Nombre atributo	P K	F K	Descripción	Tipo de dato	Tamaño maximo	Valores de ejemplo	Obligat orio	Tabla origen
Pais									
	codigo	x		codigo del país (iso)	VARCHAR	3	CL,US	si	
	nombre			nombre del pais	VARCHAR	50	Chile	si	
	continente			nombre del continente donde se ubica el pais	VARCHAR	15	America		
Region									
	idRagion	x		identificador de la region	INT		16,4,100	si	
	nombre			nombre de la region	VARCHAR	50	Los rios	si	
	codigo		x	codigo del país (iso)	VARCHAR	3	CL,US	si	Pais
Comuna									
	idComuna	x		identificador unico de la comuna	INT		12,138,202	si	
	nombre			nombre de la comuna	VARCHAR	50	Valdivia,Ñuñoa	si	
	poblacion			numero de habitantes de la comuna	INT		100675	si	
	superficie			superficie de terreno que	INT		1065	si	

				abarca la comuna en km cuadrados					
	viviendas			numero total de viviendas en la comuna	INT		54300	si	
	idRegion		x	identificador de la region	INT		18,17,55	si	Region
	idConectividad		x	identificador de la fila de datos de conectividad	INT		1,53,107	si	Conectividad
	idSeguridad		x	identificador unico de los datos de seguridad	INT		14,56	si	Seguridad
Conectividad									
	idConectividad	x		identificador de la fila de datos de conectividad	INT		1,53,107	si	
	conexioneInternet			cantidad conexiones a internet registradas en diciembre año 2022	INT		276214	si	
	idComuna		x	identificador unico de la comuna	INT		12,138,202	si	Comuna
CentroSalud									
	idCentro	x		identificador unico del centro de salud	INT		12,18,235	si	
	nombre			nombre del centro de salud	VARCHAR	150	Clinica Alemana	si	

	direccion			dirección del centro	VARCHAR	100	Beauchef 765	no	
	telefono			telefono de registrado del centro	NUMERIC	10	532711601	no	
	idComuna		x	identificador unico de la comuna	INT		12,138,202	si	Comuna
Seguridad									
	idSeguridad	x		identificador unico de los datos de seguridad	INT		14,56	si	
	tasaDenunciasDMCS			tasa de denuncia de delitos de mayor connotación social	FLOAT		0,7	si	
	indiceVulnerabilidad			Indice de Vulnerabilidad Socio Delictual o Índice Más Equidad, Más Seguridad	FLOAT		0,3	si	
	idComuna		x	identificador unico de la comuna	INT		12,138,202	si	Comuna
EstablecimientoEducación									
	idEstablecimiento	x		identificador unico del establecimiento	INT		244,52	si	

				educacional					
	nombre			Nombre del establecimiento educacional	VARCHAR	150	Instituto Salesiano	si	
	pace			Indicador si forma parte del programa PACE	BOOL		TRUE	si	
	idComuna		x	identificador unico de la comuna	INT		12,138,202	si	Comuna

Elección y fuentes de variables por aspecto:

Los datos tanto de los códigos de regiones y de las comunas utilizados para sus “ids” son los oficiales correspondientes a código de región y código de comuna. Además los datos tanto de las viviendas, población por comuna y superficie comunal corresponden a los datos del censo 2017 y de la biblioteca nacional del congreso

(<http://www.censo2017.cl/descargue-aqui-resultados-de-comunas/>)

(<https://www.bcn.cl/siit/estadisticasterritoriales/tema?id=140>)

Para salud encontramos en la página del minsal

(<https://reportesdeis.minsal.cl/ListaEstablecimientoWebSite/>) el listado de los centros de salud por comuna, estos los descargamos a modo de excel para modificarlos y ponerles el código de comuna (en lugar del nombre ya que existía conflicto con tildes y demases). Tras estructurar los datos los copiamos en un csv en el orden conveniente para nuestra base de datos y los cargamos apoyados de nuestro programa en python para cargar a nuestra base de datos. Los atributos que nos interesaron y dejamos para el posterior estudio fueron, nombre, dirección y teléfono

Para educación tras varias decepciones logramos encontrar en la siguiente página del

mineduc (<https://datosabiertos.mineduc.cl/directorio-de-establecimientos-educacionales/>)

el listado de centros de educación registrados, y para nuestra suerte incluía un apartado con el código de la comuna. En este caso nos interesó el apartado de nombre (ya que nos indica también el tipo de establecimiento (liceo, jardín, etc)) y si forman parte del programa PACE o no.

En cuanto a los apartados de Conectividad y Seguridad utilizamos la información sacada de la Biblioteca nacional del congreso

(<https://www.bcn.cl/siit/estadisticasterritoriales/tema?id=144> (enlace de ejemplo de uno de los atributos)). Esto debido a que bajo nuestro criterio, al igual que las páginas e informaciones sacadas de páginas de los ministerios, la información es la más fiable y oficial que podemos utilizar. Para estos apartados los atributos a estudiar (incluir en la base de datos) fueron “cantidad de conexiones a internet”, “tasa de denuncia de DMC” e “índice de vulnerabilidad socio delictual”

Para la carga de los archivos nos apoyamos de un programa en python el cual nos facilita la carga masiva de datos desde un csv. En el repositorio de github se encuentra el script con nombre main.py. Además para ciertos datos realizamos carga manual desde la consola de mariDB. Para los datos de comunas y región utilizamos los datos de la página del gobierno y de SII con respecto a los códigos regionales y comunales

Índice de bienestar:

Para el índice de bienestar agregamos la tabla indicadorBienestar la cual almacenara los puntajes de las diferentes comunas con respecto al modelo propuesto a continuación (cambio al diagrama entidad relacion mas tras la explicacion del modelo)

Para nuestro indicador de bienestar analizamos los datos de cada una de las tablas para clasificar a las comunas como malas(0 puntos), medias(1 punto) y altas(2 puntos). Como tenemos 4 categorías el puntaje máximo es de 12. Obviamente este es un modelo muy simplificado de la realidad, pero ayuda a obtener un ranking para tener una idea global de la calidad de vida en cada una de las comunas de Chile.

A continuación la explicación de como y en base a que se eligieron los umbrales de cada una de las 4 categorías:

Salud:

Para esta categoría los umbrales fueron elegidos en base a la densidad de centros de salud (c. de salud/superficie), lo cual indica que tan cerca en promedio vas a tener un centro de salud en esta comuna.

Malo: $X < 0,0625$

Medio: $0,0625 \leq X < 0,1$

Bueno: $0,1 \leq X$

Educación:

Para educación igual nos fijamos en la densidad de centros educacionales, por la misma razón que los centros de salud.

Malo: $X < 0,1$

Medio: $0,1 \leq X < 1$

Bueno: $1 \leq X$

Seguridad:

Para el apartado de seguridad el puntaje estará determinado por su tasa de denuncias de DMCS, la cual indica cuantas denuncias de delitos de mayor connotación social se realizan al año, dividido por cada 100 000 (por lo que ya esta en perspectiva/normalizado). Para este caso entre mas bajo sea mejor.

Malo: $X > 2700$

Medio: $2700 \geq X > 1700$

Bueno: $1700 \geq X$

Conectividad:

En lo que respecta a conexiones a internet, como grupo decidimos calificar las comunas según las conexiones totales, dividido en el número de viviendas de la comuna en cuestión.

Malo: $X < 0,25$

Medio: $0,25 \leq X < 0,75$

Bueno: 0,75=<X

Modelo entidad relación actualizado:

