



IIC2115 – Programación como Herramienta para la Ingeniería (II/2020)

Laboratorio 3 - Bases de datos

Objetivos

- Manejar correctamente los conocimientos de SQL para manipular datos globales

Entrega

- **Lenguaje a utilizar:** Python 3.6 o superior
- **Lugar:** Repositorio privado en GitHub. Recuerde incluir todo en una carpeta de nombre **L03**.
- **Entrega:** Domingo 18 de octubre a las 23:59 hrs.
- **Formato de entrega:**
 - Archivo python notebook (**L03.ipynb**) con la solución de las misiones de este laboratorio. Utilice múltiples celdas de texto y código para facilitar la revisión de su laboratorio. **Deje todo ejecutado antes de realizar su commit**, se recomienda utilizar la opción de "restart and run all" disponible en jupyter notebook.
 - Archivo python (**L03.py**) con el mismo código disponible en su archivo ipynb.
 - Todos los archivos deben estar ubicados en la carpeta **L03**. No se debe subir ningún otro archivo a la carpeta. **No suba las bases de datos a GitHub o tendrá un descuento adicional inapelable de 5 décimas.**
- **Descuentos:** El descuento por atraso se realizará de acuerdo a lo definido en el programa del curso. Además de esto, tareas que no cumplan el formato de entrega tendrán un descuento de 0.5 pts.

- **Laboratorios con errores de sintaxis y/o que generen excepciones serán calificadas con nota 1.0.**
- Si su laboratorio es entregado fuera de plazo, tiene hasta el **lunes 19 de octubre a las 11:59AM hrs** para responder el formulario de **entregas fuera de plazo** disponible en el Syllabus.
- Las discusiones en las *issues* del Syllabus en GitHub son parte de este enunciado.
- El uso de **python está prohibido** en el labrotarorio, todo debe ser resuelto mediante consultas SQL.

Introducción

En este laboratorio deberá manejar una serie de datos globales para cumplir una serie de misiones. Cada misión indica el puntaje máximo a obtener por desarrollarla completamente. Aun así, podrá obtener parte del puntaje si la misión no es resulta correctamente.

Los Datos

Los datos globales para resolver este laboratorio se encuentran en una base de datos que posee dos tablas. Una de las tablas alberga información de ciudades y la otra alberga información de la huella ecológica por país. A continuación se detalla cada una de las tablas y la descripción de sus columnas.

Ciudades

La información de ciudades esta alojada en la tabla **city** cada registro corresponde a una ciudad del mundo. Para cada ciudad, considere solo las siguientes columnas (ignore otro tipo de columnas):

1. **ID** (TEXT): Identificador único de una ciudad del mundo
2. **Name** (TEXT): Nombre de la ciudad
3. **CountryCode** (TEXT): Identificador único de un país.
4. **District** (TEXT): Distrito al que pertenece la ciudad.
5. **Population** (TEXT): Población total en la ciudad.
6. **CountryName** (TEXT): Nombre del país al que pertenece la ciudad.
7. **Continent** (TEXT): Continente al que pertenece el país.

8. **Region** (TEXT): Región a la que pertenece el país.
9. **CountrySurface** (TEXT): Superficie del país en Km^2
10. **IndepYear** (TEXT): Año de independencia del país.
11. **CountryPopulation** (TEXT): Población total del país.
12. **CountryLifeExpectancy** (TEXT): Expectativa de vida en años en el país.
13. **GNP** (TEXT): Indicador GNP (Gross National Product) del país.
14. **GNPOld** (TEXT): Indicador anterior GNP (Gross National Product) del país.
15. **CountryLocalName** (TEXT): Nombre local del país.
16. **GovernmentForm** (TEXT): Tipo de gobierno.
17. **HeadOfState** (TEXT): Presidente o primer ministro del país.
18. **Language** (TEXT): Un lenguaje que se habla en la ciuda.
19. **IsLanguageOfficial** (TEXT): Indica si el lenguaje es oficial (T) o no (F).
20. **LanguagePercentage**: Porcentaje de uso de ese lenguaje en el país.

Huella ecológica:

La información de la huella ecológica se encuentra en la tabla **eco**. Cada registro corresponde a un país, considere solo las siguientes columnas (ignore otro tipo de columnas):

1. **Country** (TEXT): Nombre del país.
2. **Region** (TEXT): Región a la que pertenece el país.
3. **Population (millions)** (TEXT): Población total del país en millones de habitantes.
4. **HDI** (TEXT): Índice de desarrollo humano del país.
5. **GDP per Capita** (TEXT): Índice (GPD) Gross Domestic Product per cápita.
6. **Cropland Footprint** (TEXT): Huella producida por el cultivo en el país.
7. **Grazing Footprint** (TEXT): Huella producida por el pastoreo en el país.

8. **Forest Footprint** (TEXT): Huella producida por el área forestal del país.
9. **Carbon Footprint** (TEXT): Huella de carbono producida en el país.
10. **Fish Footprint** (TEXT): Huella producida por la cría de peces en el país.
11. **Total Ecological Footprint** (TEXT): Suma de las huellas enunciadas anteriormente. Medido en gha por persona.
12. **Cropland** (TEXT): Suelo dedicado a cultivo en el país.
13. **Grazing Land** (TEXT): Suelo dedicado al pastoreo en el país.
14. **Forest Land** (TEXT): Suelo correspondiente a bosque en el país.
15. **Fishing Water** (TEXT): Suelo correspondiente a agua de crianza de peces en el país.
16. **Urban Land** (TEXT): Suelo urbano en el país.
17. **Total Biocapacity**: Capacidad biológica total.
18. **Biocapacity Deficit or Reserve** (TEXT): Déficit de la capacidad biológica o de reserva.
19. **Data Quality** (TEXT): Indicador de la calidad de los datos.

Misiones

Usted deberá cumplir una serie de misiones que responden preguntas sobre una base de datos de ciudades y una acerca de la huella ecológica por país. Utilizando los conceptos cubiertos en cátedra y los apuntes del curso, De sus clases de programación, aprendió que es necesario preparar su base de datos, familiarizarse y crear algunas tablas nuevas a partir de la información para no ser redundante. Para ello, siga cuidadosamente las misiones que lo guiarán a lograr su objetivo.

Cada apartado en las misiones M1 a M4 corresponde a 0.5 puntos. Además, existe una misión BONUS donde podrá optar a un máximo de 0.5 puntos (0.25 puntos por consulta). El 7 se obtiene con 6 puntos, el puntaje máximo es 8 puntos, sin embargo la nota máxima en este laboratorio es un 7.5.

M1. Preparando la base de datos: Su primera misión será inspeccionar las tablas anteriormente descritas y generar a partir de ellas nuevas tablas, de manera que no exista información repetida, en la base de datos. Dicho de otra manera, las tablas contienen múltiples **entidades** y su misión es desenredar esto y crear una tabla por cada entidad. Para cada una de estas nuevas tablas, defina llaves primarias,

foráneas cuando corresponda y defina los tipos de datos adecuados para cada columna (actualmente estan todos en formato TEXT). Cualquier supuesto que haga con respecto a la imputación de valores o la decisión de qué valor conservar en caso de ambigüedad debe quedar claramente especificado.

M2. Conteos: En base a las tablas ya separadas, realice **una consulta** en SQL por cada una de las siguientes preguntas relacionas con conteo:

- a. ¿En cuántos países se habla cada uno de los idiomas identificados? Repita esta consulta solo considerando idiomas oficiales.
- b. ¿En cuántos países se habla en un porcentaje mayor a 40% cada uno de los idiomas identificados? Repita esta consulta solo considerando idiomas oficiales.
- c. ¿Cuántos países tienen un porcentaje mayor a 10% de sus ciudades con más de 1 millón de habitantes?
- d. ¿En cuántos países donde se hable 5 o más lenguajes, el porcentaje de suelo urbano es mayoritario?

M3. Encontrando registros llamativos: En base a las tablas ya separadas, realice **una consulta** en SQL por cada una de las siguientes preguntas:

- a. Encuentre el país más rico de la región más pobre del mundo.
- b. Para la región más pobre del mundo, indique el porcentaje de su suelo que está cubierto por bosques.
- c. Para los 5 idiomas más hablados del mundo (número de personas que hablan ese idioma). Identifique para cada idioma, el país donde ese idioma es oficial y predominante que deje la mayor huella ecológica. Ordene los países resultantes en orden decreciente de población.
- d. Asumiendo que el uso de un lenguaje distribuye de manera homogénea en cada país, encuentre la ciudad del mundo donde hay menos personas hablando un lenguaje, con la condición que exista al menos una. Por cada ciudad encontrada (si hay más de una), indique además de la ciudad, el lenguaje, si es oficial y la cantidad de personas que lo hablan.

M4. Rankings: En base a las tablas ya separadas, realice **una consulta** en SQL por cada una de las siguientes preguntas:

- a. Encuentre los 10 distritos con mayor y menor porcentaje de población, en relación al país al que pertenecen.
- b. Construya un ranking descendente de las regiones del mundo más contaminantes, donde para cada una de ellas se indique en orden ascendente los 3 países más contaminantes.

- c. Para las 5 regiones más ricas del mundo, ordenadas de manera ascendente, indique el porcentaje de cada uno de sus usos de suelo, ordenados de acuerdo a algún criterio definido por ud.
- d. Por cada una de las regiones del mundo, encuentre los 5 países donde las personas generan mayor “riqueza” durante sus vidas.

M5. **Bonus:** En base a las tablas ya separadas, realice una consulta en SQL por cada una de las siguientes preguntas:

- a. Considere que las huellas ecológicas se distribuyen homogéneamente entre las ciudades de un país en base a sus poblaciones. Incorpore en **su tabla** de ciudades las huellas ecológicas desagregadas y la total.
- b. Construya un ranking con las ciudades que más huella ecológica dejan por continente (máximo 3 ciudades por continente) solo para ciudades que sean parte de los 20 países que dejan la mayor huella ecológica.

Política de Integridad Académica

“Como miembro de la comunidad de la Pontificia Universidad Católica de Chile me comprometo a respetar los principios y normativas que la rigen. Asimismo, prometo actuar con rectitud y honestidad en las relaciones con los demás integrantes de la comunidad y en la realización de todo trabajo, particularmente en aquellas actividades vinculadas a la docencia, el aprendizaje y la creación, difusión y transferencia del conocimiento. Además, velaré por la integridad de las personas y cuidaré los bienes de la Universidad.”

En particular, se espera que mantengan altos estándares de honestidad académica. Cualquier acto deshonesto o fraude académico está prohibido; los alumnos que incurran en este tipo de acciones se exponen a un procedimiento sumario. Ejemplos de actos deshonestos son la copia, el uso de material o equipos no permitidos en las evaluaciones, el plagio, o la falsificación de identidad, entre otros. Específicamente, para los cursos del Departamento de Ciencia de la Computación, rige obligatoriamente la siguiente política de integridad académica en relación a copia y plagio: Todo trabajo presentado por un alumno (grupo) para los efectos de la evaluación de un curso debe ser hecho individualmente por el alumno (grupo), sin apoyo en material de terceros. Si un alumno (grupo) copia un trabajo, se le calificará con nota 1.0 en dicha evaluación y dependiendo de la gravedad de sus acciones podrá tener un 1.0 en todo ese ítem de evaluaciones o un 1.1 en el curso. Además, los antecedentes serán enviados a la Dirección de Docencia de la Escuela de Ingeniería para evaluar posteriores sanciones en conjunto con la Universidad, las que pueden incluir un procedimiento sumario. Por “copia” o “plagio” se entiende incluir en el trabajo presentado como propio, partes desarrolladas por otra persona. Está permitido usar material disponible públicamente, por ejemplo, libros o contenidos tomados de Internet, siempre y cuando se incluya la cita correspondiente.