



IIC2115 – Programación como Herramienta para la Ingeniería (II/2023)

Tarea 3

Aspectos generales

- **Objetivo:** aplicar los contenidos de manejo de grandes volúmenes de datos para modelar entidades y sus relaciones, poblarlas con datos y realizar consultas sobre ellas.
- **Lugar de entrega:** domingo 10 de diciembre hasta las 23:59 hrs. en el repositorio privado.
- **Formato de entrega:** ÚNICAMENTE el archivo Python Notebook (**T3.ipynb**) con la solución de la tarea. El archivo debe estar ubicado en la carpeta **T3**. Es requerimiento de formato el utilizar múltiples celdas de texto y código para la construcción de la solución, además de dejar todas estas ejecutadas en el archivo entregado. Entregas que no cumplan el formato de entrega tendrán un descuento de 0,5 pts.
- **Entregas atrasadas:** El descuento por atraso se realizará de acuerdo a lo definido en el programa del curso.
- **Issues:** las discusiones en las *issues* del Syllabus que sean relevantes para el desarrollo de la tarea, serán destacadas y se considerarán como parte de este enunciado.
- **Librerías:** el uso de librerías externas, distintas a las cubiertas en el curso y que solucionen aspectos centrales o extensos de la tarea está prohibido. Solo se podrán usar las que han sido aprobadas explícitamente por el cuerpo docente, previa consulta de los estudiantes, lo que será informado en el Syllabus a través de *issues*.
- **Tarea con errores de sintaxis y/o que generen excepciones en todas las ejecuciones serán calificados con nota 1,0.**

Introducción

En esta tarea deberá manejar un conjunto de datos consistente en información de ciudades y de la huella ecológica de distintos países. Los datos deberán ser primero preprocesados y organizados, con el fin de crear tablas que modelen adecuadamente las entidades presentes. Posteriormente, deberá construir consultas en SQL que permitan responder preguntas relevantes sobre las entidades antes creadas.

Los datos

Los datos 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 almacenada en la tabla **city**, donde 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 ciudad.
19. **IsLanguageOfficial** (TEXT): Indica si el lenguaje es oficial (T) o no (F).
20. **LanguagePercentage** (TEXT): 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** (TEXT): 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

Para la primera misión, puede utilizar la cantidad de consultas de SQL que quiera, mientras que a partir de la segunda, solo podrá realizar **una única consulta** en SQL para cada una, con el fin de obtener la respuesta.

Todas las misiones tienen el mismo puntaje.

M1. Inspeccione las tablas anteriormente descritas y las consultas solicitadas en las siguientes misiones. A partir de esto, genere y complete con datos nuevas tablas, de modo que exista una tabla por entidad. Por ejemplo, si las entidades identificadas con Ciudad, País, Idioma e Información Ecológica, debe crear una tabla por cada una y luego poblarlas con la información correspondiente.

Para cada una de estas nuevas tablas, defina llaves primarias y foráneas cuando corresponda, además de los tipos de datos adecuados para cada columna. 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. ¿En cuántos países se habla cada uno de los idiomas identificados?

M3. ¿Cuántos países tienen un porcentaje mayor a 10% de sus ciudades con más de 1 millón de habitantes?

M4. ¿En cuántos países donde se hable 5 o más lenguajes, el porcentaje de suelo urbano es mayoritario?

M5. Encuentre el país más rico de la región más pobre del mundo.

- M6. Para la región más pobre del mundo, indique el porcentaje de su suelo que está cubierto por bosques.
- M7. Construya un ranking descendente de las regiones del mundo más contaminantes, donde para cada una de ellas se indiquen además los 3 países más contaminantes.
- M8. 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.
- M9. Por cada una de las regiones del mundo, encuentre los 5 países donde las personas generan mayor “riqueza” durante sus vidas.
- M10. 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.

Corrección

Es importante que deje todas las celdas de su trabajo ejecutadas antes de subir el archivo, de lo contrario se le aplicará un descuento de 0,5 pts. al puntaje total.

Para la corrección se revisarán los procedimientos y consultas desarrolladas para responder a las preguntas. No existe una única consulta correcta para cada misión. Si en alguna misión, con excepción de la M1, se realiza más de 1 consulta para responder a la pregunta, el puntaje obtenido en esa misión será de 0 puntos.

Política de Integridad Académica

Los/as estudiantes de la Escuela de Ingeniería de la Pontificia Universidad Católica de Chile deben mantener un comportamiento acorde a la Declaración de Principios de la Universidad. En particular, se espera que mantengan altos estándares de honestidad académica. Cualquier acto deshonesto o fraude académico está prohibido; los/as estudiantes que incurran en este tipo de acciones se exponen a un Procedimiento Sumario. Es responsabilidad de cada estudiante conocer y respetar el documento sobre Integridad Académica publicado por la Dirección de Docencia de la Escuela de Ingeniería.

Específicamente, para los cursos del Departamento de Ciencia de la Computación, rige obligatoriamente la siguiente política de integridad académica. Todo trabajo presentado por un/a estudiante para los efectos de la evaluación de un curso debe ser hecho **individualmente** por el/la estudiante, **sin apoyo en material**

de terceros. Por “trabajo” se entiende en general las interrogaciones escritas, las tareas de programación u otras, los trabajos de laboratorio, los proyectos, el examen, entre otros.

En particular, si un/a estudiante copia un trabajo, o si a un/a estudiante se le prueba que compró o intentó comprar un trabajo, **obtendrá nota final 1.1 en el curso** y se solicitará a la Dirección de Pregrado de la Escuela de Ingeniería que no le permita retirar el curso de la carga académica semestral.

Por “copia” se entiende incluir en el trabajo presentado como propio, partes hechas por otra persona. En caso que corresponda a “copia” a otros estudiantes, la sanción anterior se aplicará a todos los involucrados. En todos los casos, se informará a la Dirección de Pregrado de la Escuela de Ingeniería para que tome sanciones adicionales si lo estima conveniente.

También se entiende por copia extraer contenido sin modificarlo sustancialmente desde fuentes digitales como Wikipedia o mediante el uso de asistentes inteligentes como ChatGPT o Copilot. Se entiende que una modificación sustancial involucra el análisis crítico de la información extraída y en consecuencia todas las modificaciones y mejoras que de este análisis se desprendan. Cualquiera sea el caso, el uso de fuentes bibliográficas, digitales o asistentes debe declararse de forma explícita, y debe indicarse cómo el/la estudiante mejoró la información extraída para cumplir con los objetivos de la actividad evaluativa.

Obviamente, está permitido usar material disponible públicamente, por ejemplo, libros o contenidos tomados de Internet, **siempre y cuando se incluya la referencia correspondiente.**

Lo anterior se entiende como complemento al Reglamento del Estudiante de la Pontificia Universidad Católica de Chile (<https://registrosacademicos.uc.cl/reglamentos/estudiantiles/>). Por ello, es posible pedir a la Universidad la aplicación de sanciones adicionales especificadas en dicho reglamento.

Compromiso del Código de Honor

Este curso suscribe el Código de Honor establecido por la Universidad, el que es vinculante. Todo trabajo evaluado en este curso debe ser propio. En caso que exista colaboración permitida con otros/as estudiantes, el trabajo deberá referenciar y atribuir correctamente dicha contribución a quien corresponda. Como estudiante es un deber conocer el Código de Honor (<https://www.uc.cl/codigo-de-honor/>).