

# Evaluación Parcial 3

## Encargo

Sigla	Nombre Asignatura	Tiempo Asignado	% Ponderación
GAA1104	Big Data e Inteligencia de Negocios	2 h	24%

### 01. Agente evaluativo

<input checked="" type="checkbox"/> Heteroevaluación	<input type="checkbox"/> Coevaluación	<input type="checkbox"/> Autoevaluación
--	---------------------------------------	---

### 02. Tabla de especificaciones

Resultado de Aprendizaje	Indicador de Logro (IL)	Indicador de Evaluación (IE)*	Ponderación Indicador Logro	Ponderación Indicador de Evaluación
RA5 Analiza los datos utilizando Business intelligence y técnicas estadísticas descriptivas y modelos para colaborar en la toma de decisiones.	IL 5.1 Aplica las funciones básicas de herramientas tecnológicas adecuadas para elaborar informes sobre grandes volúmenes de datos de acuerdo con los requerimientos de la organización.	5.1.1 Utiliza herramientas tecnológicas adecuadas para el manejo y análisis de grandes volúmenes de datos, como software de Business Intelligence, herramientas de visualización de datos o plataformas de análisis de datos.	15%	15%

	IL 5.2 Comprende los requisitos mínimos que debe contener un informe en términos de contenido, formato, frecuencia y audiencia.	5.2.1 Selecciona la información relevante que debe incluirse en el informe, de acuerdo con los objetivos y necesidades de la organización y los destinatarios del informe.	15%	<b>15%</b>
	IL 5.3 Realiza consultas y extracciones de datos utilizando las herramientas tecnológicas, aplicando filtros, agregaciones para obtener la información relevante para la generación de reportes.	5.3.1 Utiliza herramientas tecnológicas apropiadas para realizar consultas y extracciones de datos, como software de Business Intelligence, bases de datos, lenguajes de consulta o herramientas de visualización de datos.	20%	<b>20%</b>
	IL 5.4 Manipula los datos extraídos utilizando las herramientas tecnológicas, realizando limpieza, normalización y cálculos necesarios para obtener los resultados deseados en los reportes.	5.4.1 Ejecuta limpieza de los datos extraídos, como eliminar valores duplicados, corregir errores, tratar valores faltantes o inconsistentes, y asegurar la integridad de los datos.	30%	<b>30%</b>
	IL 5.5 Genera reportes en el ámbito de la especialidad seleccionando y organizando datos de manera clara, efectiva, estilos y gráficos apropiados para comunicar a la organización.	5.5.1 Organiza los datos de manera lógica y estructurada, siguiendo un orden coherente y facilitando la comprensión del contenido del reporte.	20%	<b>20%</b>
<b>Total</b>			<b>100%</b>	<b>100%</b>

\*Si no se construyeron IE para esta evaluación, eliminar esta columna

## 03. Instrucciones para el/la estudiante

Esta es una evaluación que corresponde a una **entrega de encargo** y tiene un **24%** de ponderación sobre las evaluaciones parciales de la asignatura.

El **tiempo** para desarrollar esta evaluación es de **80 minutos** y se realiza en **equipos** en **laboratorio y desde sus casas**.

**La evaluación consiste en:**

Generar un archivo. ipynb donde redacte de manera clara y ordenada, todas las actividades que están enumerados en esta tarea.

Para esta tarea, usted dispone de los datos de nacimientos en Chile desde 1990 a 2017 que se encuentra en un .zip llamado Nacimientos en Chile

Estos datos fueron obtenidos hace mucho tiempo, pero que, por temas de contingencia, no es posible obtenerlos de nuevo (la página del DEIS de origen de esos datos ya no es pública), por lo que no se cuenta con la documentación completa; sin embargo, los nombres de las columnas, en general, se pueden deducir. Por ejemplo, día\_nac refiere al día en que nació, y mes\_nac al mes correspondiente, edad\_p corresponde a la edad del padre, y edad\_m a la edad de la madre, etc.

1. Juntar todos los archivos de todos los años en un solo dataframe global para efectuar los análisis posteriores.
2. ¿Cuál es el mes más frecuente de nacimientos en Chile? Comentar al respecto.
3. ¿Cuál es el día del año más común en el que la gente en Chile está de cumpleaños?
4. Calcular covarianza y correlación entre peso y talla a nivel general (tomando toda la base). Luego hacerlo por año. ¿Cambia con el paso de los años?
5. Calcular covarianza y correlación entre la edad del padre y la edad de la madre, a nivel general (tomando toda la base). Luego hacerlo por año. ¿Cambia con el paso de los años?
6. Investigue las condiciones para que un bebé cuando nazca se considere “prematuro”, “a término” y “postérmino”. Hacer diagramas de caja para el peso y la talla para estas 3 categorías. Comentar al respecto.
7. Crear una columna llamada “indicador” que valga “1” si el bebé nació en una ambulancia y que valga “2” si el bebé nació en el trayecto (para los datos desde 1996). Caracterice los datos atípicos (outliers) usando el IQR y el primer y tercer cuartil para cada una de estas variables peso, talla, edad del padre y madre, en cada caso con indicador 1 o 2. Comentar al respecto.

# Pauta de Evaluación

## Pauta tipo: Rúbrica

Categoría	% logro	Descripción niveles de logro
Muy buen desempeño	100%	Demuestra un desempeño destacado, evidenciando el logro de todos los aspectos evaluados en el indicador.
Buen desempeño	80%	Demuestra un alto desempeño del indicador, presentando pequeñas omisiones, dificultades y/o errores.
Desempeño aceptable	60%	Demuestra un desempeño competente, evidenciando el logro de los elementos básicos del indicador, pero con omisiones, dificultades o errores.
Desempeño incipiente	30%	Presenta importantes omisiones, dificultades o errores en el desempeño, que no permiten evidenciar los elementos básicos del logro del indicador, por lo que no puede ser considerado competente.
Desempeño no logrado	0%	Presenta ausencia o incorrecto desempeño.

Indicador de Evaluación	Categorías de Respuesta					Ponderación del Indicador de Evaluación
	Muy buen desempeño 100%	Buen desempeño 80%	Desempeño aceptable 60%	Desempeño incipiente 30%	Desempeño no logrado 0%	
5.1.1 Utiliza herramientas tecnológicas adecuadas para el manejo y análisis de grandes volúmenes de datos, como software de Business Intelligence, herramientas de visualización de datos o plataformas de análisis de datos.	Utiliza 1 herramienta tecnológica para el 100% de análisis de grandes volúmenes de datos.	Utiliza 1 herramienta tecnológica para el 80% de análisis de grandes volúmenes de datos.	Utiliza 1 herramienta tecnológica para el 60% de análisis de grandes volúmenes de datos.	Utiliza 1 herramienta tecnológica para el 30% de análisis de grandes volúmenes de datos.	Utiliza 1 herramienta tecnológica para el 0% de análisis de grandes volúmenes de datos.	15%

5.2.1 Selecciona la información relevante que debe incluirse en el informe, de acuerdo con los objetivos y necesidades de la organización y los destinatarios del informe.	Selecciona el 100% de información que le permite generar un informe	Selecciona el 80% de información que le permite generar un informe	Selecciona el 60% de información que le permite generar un informe	Selecciona el 30% de información que le permite generar un informe	Selecciona el 0% de información que le permite generar un informe	<b>15%</b>
5.3.1 Utiliza herramientas tecnológicas apropiadas para realizar consultas y extracciones de datos, como software de Business Intelligence, bases de datos, lenguajes de consulta o herramientas de visualización de datos.	Utiliza herramientas de consultas y extracción obteniendo resultados al 100%	Utiliza herramientas de consultas y extracción obteniendo resultados al 80%	Utiliza herramientas de consultas y extracción obteniendo resultados al 60%	Utiliza herramientas de consultas y extracción obteniendo resultados al 30%	Utiliza herramientas de consultas y extracción obteniendo resultados al 0%	<b>20%</b>
5.4.1 Ejecuta limpieza de los datos extraídos, como eliminar valores duplicados, corregir errores, tratar valores faltantes o inconsistentes, y asegurar la integridad de los datos.	Realiza limpieza de datos, generando una data totalmente procesable.	Realiza limpieza de datos, generando una data procesable, aunque con algún error puntual.	Realiza limpieza de datos, generando una data parcialmente procesable, con algunos errores en los datos.	Realiza limpieza de datos, generando una data procesable, pero que presenta múltiples errores relevantes.	No realiza limpieza de datos o no genera una data procesable	<b>30%</b>
5.5.1 Organiza los datos de manera lógica y estructurada, siguiendo un orden coherente y facilitando la comprensión del contenido del reporte.	Organiza los datos utilizando una estructura totalmente adecuada para facilitar la comprensión del contenido.	Organiza los datos utilizando una estructura en general adecuada para facilitar la comprensión del contenido.	Organiza los datos utilizando una estructura parcialmente adecuada para facilitar la comprensión del contenido.	Organiza los datos utilizando una estructura que resulta poco adecuada para facilitar la comprensión del contenido.	No organiza los datos utilizando una estructura adecuada para facilitar la comprensión del contenido, o la propuesta no es adecuada.	<b>20%</b>
<b>Total</b>						<b>100%</b>