



**Asignatura:
Minería de Datos
BIY 7121**

Sistema de Evaluación del Semestre

Fechas y porcentajes
tentativos .
Por confirmar.

- Experiencia Aprendizaje 1 - Preparación de Datos
 - Encargo sin presentación – 35% (Semana 04 - Vi 04 de Abril)
- Experiencia Aprendizaje 2 - Técnicas de Minería de Datos
 - Encargo sin presentación – 35% (Semana 10 - Vi 16 de Mayo)
- Experiencia Aprendizaje 3 - Evaluación y Difusión de Modelos
 - Encargo sin presentación – 30% (Semana 15 – Vi 20 de Junio)
- Examen (ET): Presentaciones finales (semanas 16, 17 y 18)
 - Encargo con presentación - 40%

Nota
Presentación ET

Promedio Final = 60% Nota Presentación + 40% Nota Examen Transversal

Evaluación #1

Esta semana tendrán tiempo para el trabajo – entrega primera parte

- Experiencia Aprendizaje 1 - Preparación de Datos
 - Encargo sin presentación – 35% (Semana 04 - Vi 04 de Abril)

Esta es una evaluación que corresponde a una **entrega de encargo sin presentación** y tiene un **35%** de ponderación sobre la nota final de la asignatura.

El **tiempo** para desarrollar esta evaluación es de **2 semanas** y se realiza en grupos de dos.

La evaluación consiste en realizar un informe técnico en formato Jupyter Notebook que contenga las dos primeras fases de la metodología CRISP-DM aplicada a los datos de Starcraft_duoc. Para desarrollar la evaluación utilice los anexos

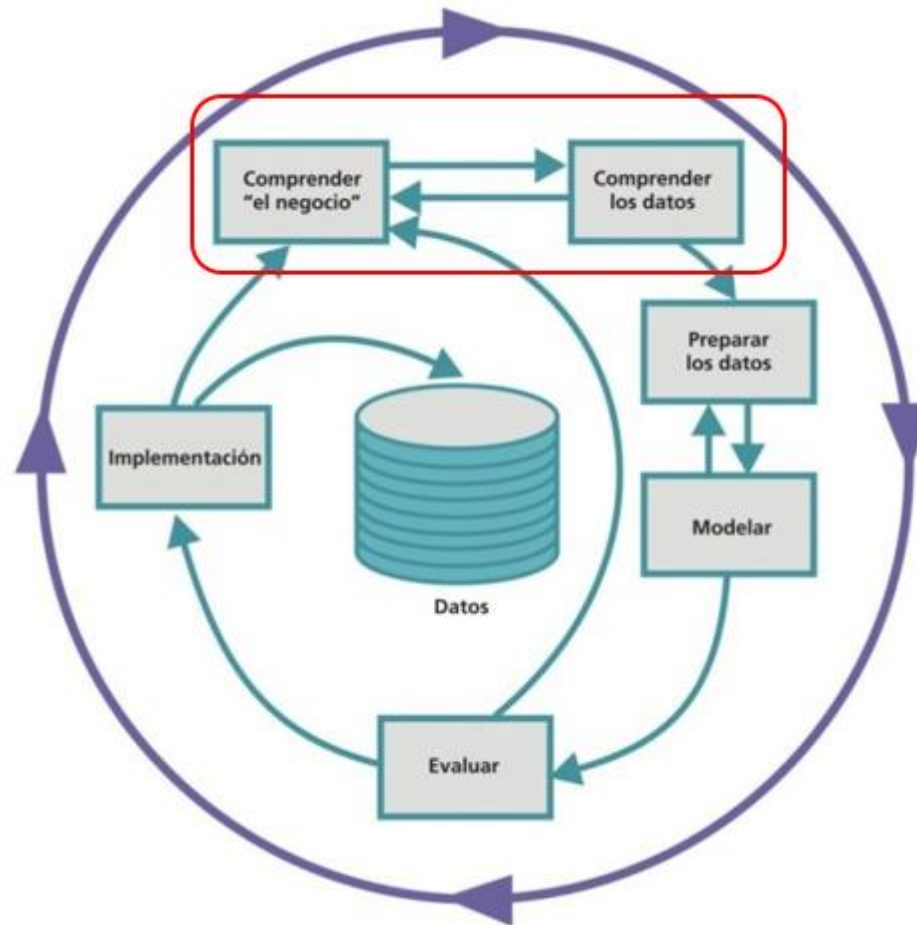
“Evaluación Parcial 1 Anexo 1” y “EV1_Anexo 2_starcraft_duoc”

**Es relevante señalar que debe describir aspectos generales del negocio de videojuego y de los datos. Considere para su elaboración el uso de bloques de códigos debidamente comentados y analizando la información en bloques Markdown.*

Instrucciones:

- Realice un informe técnico que contenga las **dos** primeras fases de la metodología CRISP-DM:
- Identifique los objetivos claves y KPIs relevantes para la compañía coherentes con la fase Business Understanding.
- Calcule estadísticos sobre los datos, identificando insight relevantes, con el fin de analizar la relación entre patrones y datos, definidos en la fase Data Understanding.
- Señale las características y naturaleza claves dentro de los datos para definir los objetivos generales del análisis.
- Identifique la existencia de valores atípicos y valores perdidos, proponiendo rutinas de limpieza que se pueden abordar para generar mayor precisión en los modelos y menor sesgo en el análisis.
- Aplique matriz de correlación, identificando qué correlaciones tienen las características, reconociendo las correlaciones negativas y positivas, e interpretando qué valor generan en el proceso los valores obtenidos.
- Proponga pasos a seguir considerando fases posteriores de acuerdo con los resultados obtenidos en las fases abordadas.

Evaluación #1



Productos a Entregar

- **CUADERNO JUPYTER EN GOOGLE COLAB (considerado como informe)**
 - **Instrucciones generales**
 - Los/as estudiantes deben desarrollar un cuaderno Jupyter en Google Colab, documentando cada paso de la implementación.
 - El cuaderno debe incluir explicaciones en texto (Markdown), código bien estructurado y visualización de resultados (gráficos y tablas).
 - Se debe proporcionar una sección de conclusiones, donde se analicen los resultados obtenidos.
 - **Estructura sugerida del cuaderno**
 - Propuesta de Notebook Jupyter en AVA
 - **Pautas de documentación**
 - Markdown debe utilizarse para explicar cada bloque de código.
 - Comentarios en el código para facilitar la comprensión de la implementación.
 - Uso de gráficos y tablas para visualizar resultados.
 - Códigos bien estructurados y modularizados.

StarCraft II es un juego de estrategia en tiempo real (RTS) desarrollado por Blizzard Entertainment. Lanzado en 2010, es la secuela del icónico StarCraft original. El juego se ambienta en un futuro distante, donde tres facciones, los humanos terran, los enjambres zerg y los misteriosos protoss, luchan por el control del sector Koprulu.

Los jugadores pueden elegir una de las tres razas, cada una con habilidades únicas y estrategias distintas. StarCraft II presenta una campaña envolvente con una narrativa rica y desafiantes misiones para cada raza. Además, destaca en el ámbito multijugador, donde jugadores de todo el mundo compiten en emocionantes batallas.

El juego ha alcanzado un estatus destacado en los deportes electrónicos (eSports), con una comunidad activa y competiciones de alto nivel. La gestión de recursos, la toma de decisiones estratégicas y la velocidad de ejecución son fundamentales en StarCraft II, lo que lo convierte en un desafío constante. Con gráficos impresionantes, una jugabilidad sólida y una escena competitiva vibrante, StarCraft II sigue siendo una referencia en el mundo de los juegos de estrategia en tiempo real.

En este contexto, se presentan los siguientes datos:

- GameID: Unique ID for each game
- LeagueIndex: 1-8 for Bronze, Silver, Gold, Diamond, Master, GrandMaster, Professional leagues
- Age: Age of each player
- HoursPerWeek: Hours spent playing per week
- TotalHours: Total hours spent playing
- APM: Action per minute
- SelectByHotkeys: Number of unit selections made using hotkeys per timestamp
- AssignToHotkeys: Number of units assigned to hotkeys per timestamp
- MinimapAttacks: Number of attack actions on minimal per timestamp
- MinimapRightClicks: Number of right-clicks on minimal per timestamp
- NumberOfPACs: Number of PACs per timestamp
- GapBetweenPACs: Mean duration between PACs (milliseconds)
- ActionLatency: Mean latency from the onset of PACs to their first action (milliseconds)
- ActionsInPAC: Mean number of actions within each PAC
- TotalMapExplored: Number of 24x24 game coordinate grids viewed by player per timestamp
- WorkersMade: Number of SCVs, drones, probes trained per timestamp
- UniqueUnitsMade: Unique units made per timestamp
- ComplexUnitsMade: Number of ghosts, investors, and high templars trained per timestamp
- ComplexAbilityUsed: Abilities requiring specific targeting instructions used per timestamp
- MaxTimeStamp: Time stamp of game's last recorded event

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 Indicador de Evaluación
	Muy buen desempeño 100%	Buen desempeño 80%	Desempeño aceptable 60%	Desempeño incipiente 30%	Desempeño no logrado 0%	
Identifica las fases de la metodología CRISP-DM, reconociendo su valor para apoyar la toma de decisiones	Identifica todas las fases de la metodología CRISP-DM, reconociendo su valor para apoyar la toma de decisiones	Identifica la mayoría de las fases de la metodología CRISP-DM, reconociendo su valor para apoyar la toma de decisiones	Identifica la mayoría de las fases de la metodología CRISP-DM, pero no reconoce su valor para apoyar la toma de decisiones	Identifica la mitad o menos de las fases de la metodología CRISP-DM y no reconoce su valor para apoyar la toma de decisiones.	No identifica las fases de la metodología CRISP-DM.	10%
Identifica objetivos claves y KPIs relevantes para la compañía, coherentes con la fase que investiga el negocio Business Understanding .	Identifica la totalidad de los objetivos claves y KPIs en la etapa que investiga el negocio Business Understanding	Identifica la mayoría de los objetivos claves y KPIs en la etapa que investiga el negocio Business Understanding	Identifica la totalidad de los objetivos claves o de los KPIs en la etapa que investiga el negocio Business Understanding	Identifica menos de la mitad de los objetivos claves y KPIs en la etapa que investiga el negocio Business Understanding	No identifica los objetivos claves y KPIs en la etapa que investiga el negocio Business Understanding	20%
Calcula estadísticos sobre los datos identificando insight relevante y reveladores de estos.	Calcula correctamente los estadísticos sobre los datos, identificando insights relevantes y reveladores para analizar la relación entre patrones y datos.	Calcula con un mínimo de error los estadísticos sobre los datos, identificando insights relevantes y reveladores para analizar la relación entre patrones y datos.	Calcula correctamente los estadísticos sobre los datos, pero no identifica los insights relevantes y reveladores para analizar la relación entre patrones y datos.	Calcula con errores los estadísticos sobre los datos y no identifica insights.	No calcula los estadísticos sobre los datos.	20%
Reconoce las características y naturaleza de los diferentes tipos de datos, con el fin de definir los objetivos generales del análisis	Reconoce la totalidad de los diferentes tipos de datos, así como sus características y naturaleza.	Reconoce la mayoría de los diferentes tipos de datos, así como sus características y naturaleza.	Reconoce la mitad o menos de la mitad de los diferentes tipos de datos, así como sus características y naturaleza.	Reconoce al menos uno de los diferentes tipos de datos, así como sus características y naturaleza.	No reconoce los diferentes tipos de datos, así como sus características y naturaleza.	10%
Identifica Outliers y Missing Values en los datos, con el fin de proponer rutinas de limpieza para ellos.	Identifica la totalidad de los outliers y missing values en los datos, y propone rutinas de limpieza que permitan generar mayor precisión en los modelos y menor sesgo en el análisis.	Identifica la mayoría de los outliers y missing values en los datos, y propone rutinas de limpieza que permitan generar mayor precisión en los modelos y menor sesgo en el análisis.	Identifica la mitad o menos de los outliers y missing values en los datos, y propone rutinas de limpieza que permitan generar mayor precisión en los modelos y menor sesgo en el análisis.	Identifica los outliers y missing values en los datos, pero no propone rutinas de limpieza que permitan generar mayor precisión en los modelos y menor sesgo en el análisis.		10%
Aplica matriz de correlación determinando cómo se relacionan las variables del conjunto de datos	Aplica matriz de correlación, permitiendo un análisis de cómo se relacionan todas las variables del conjunto de datos.	Aplica matriz de correlación, permitiendo un análisis de cómo se relacionan la mayoría de las variables del conjunto de datos.	Aplica matriz de correlación, permitiendo un análisis de cómo se relacionan la mitad o menos de las variables del conjunto de datos.	Aplica matriz de correlación, pero esta no permite el análisis de cómo se relacionan las variables del conjunto de datos.	No aplica matriz de correlación.	15%
Propone pasos a seguir considerando fases posteriores de acuerdo con los resultados obtenidos en las fases abordadas.	Propone una planificación integral para las siguientes fases. La propuesta aborda de manera detallada los pasos específicos para la preparación de datos y el modelado, considerando a fondo los resultados obtenidos.	Propone una planificación efectiva que conecta claramente los resultados de Data Preparation y Modeling con los pasos a seguir.	Propone pasos a seguir de manera básica y general para Data Preparation y Modeling. La propuesta puede carecer de detalles específicos sobre la implementación.	Propone pasos a seguir de manera básica y general solo para la fase de Data Preparation sin considerar mayores aspectos sobre modelos de datos.	No propone de manera efectiva pasos a seguir. La conexión entre los resultados y las acciones planificadas es insuficiente para fases posteriores.	15%
						100%