

# Evaluación Parcial 2

## Análisis exploratorio y transformación de datos

Sigla	Nombre Asignatura	Tiempo Asignado	% Ponderación
FMY0100	Fundamentos de Machine Learning	2 horas	35%

### 01. Agente evaluativo

X	Heteroevaluación		Coevaluación		Autoevaluación
---	------------------	--	--------------	--	----------------

### 02. Tabla de Especificaciones

Resultado de Aprendizaje	Indicador de Logro (IL)	Indicador de Evaluación (IE)*	Ponderación Indicador Logro	Ponderación Indicador de Evaluación
RA1 Aplica estadística descriptiva e inferencial a los datos recolectados, considerando problemáticas, objetivos y tipo de negocio, para descubrir aspectos relevantes del negocio.	IL1.1 Aplica elementos básicos de estadística descriptiva e inferencial a los datos entregados	Utiliza al menos 4 funciones necesarias para extraer elementos básicos de estadística descriptiva.	20%	10%
		Interpreta los elementos estadísticos obtenidos a partir de los datos.		10%
	IL1.2 Manipula datos mediante la aplicación de conceptos y elementos de álgebra lineal para descubrir patrones de datos.	Obtiene coeficiente de correlación por sí solo entre las características o bien la matriz de Correlación de Pearson.	20%	10%
		Interpreta los valores obtenidos identificando valores de correlación directa e inversa.		10%
	IL1.3 Analiza el contexto a partir de los datos entregados	Con el uso de estadísticos obtiene KPI relevantes como cantidad de derrotas, cantidad de triunfos, cantidad de	10%	10%

	utilizando KPI para hacer descubrimientos del negocio.	jugadores, promedio de tiempo de partida, etc.		
RA2 Realiza lectura de datos considerando técnicas de limpieza y uso de lenguajes de programación, orientados al análisis de datos, para descubrir aspectos relevantes del negocio.	IL2.1 Diferencia los tipos de datos considerando la naturaleza de las variables involucradas en una problemática dada.	Identifica cuando el dato es numérico continuo, discreto, o variable categórica nominal u ordinal.	10%	10%
	IL2.2 Evalúa la relevancia de los datos de un problema, verificando aquellos datos faltantes y seleccionando herramientas estadísticas y gráficas para realizar acciones de limpieza y transformación mediante el uso de Python.	Trabaja los missing values (valores NaN) y los datos outliers, considerando las mejores prácticas con respecto al uso de los datos.	20%	10%
		Identifica cuales son características significativas para el contexto del objetivo que el equipo se planteó de negocio.		10%
	IL2.3 Identifica si los datos son apropiados para el modelo, según el orden de magnitud presente, seleccionando la rutina de Python adecuada para estandarizarlos.	Utiliza técnicas de transformación de datos adecuada de acuerdo a la naturaleza de estos.	20%	10%
		Utiliza las técnicas de transformación necesarias, ya sea de scaling, normalización o encoding para dejar listos los datos para la fase de Modeling.		10%
Total			100%	100%

### 03. Instrucciones para el/la estudiante

Esta es una evaluación sumativa que corresponde a la Evaluación Parcial 2. Consiste en una entrega parcial del caso semestral entregado, que equivale al **35%** de ponderación sobre la nota final de la asignatura.

En esta entrega debes considerar lo necesario basado en la metodología CRISP-DM a las 3 primeras fases: Business Understanding, Data Understanding y Data Transformation.

Recuerda que estas fases no son completamente secuenciales, y que puedes iterar entre las primeras dos, hasta tener objetivos definidos del negocio basado en los datos.

El **tiempo** para desarrollar esta evaluación es de 2 semanas y se realiza de manera **grupal** de acuerdo con los equipos definidos.

**La evaluación consiste en:**

- Business Understanding: Detallar y describir en el formato habilitado, todo el conocimiento y aspectos relevantes que tengan sobre el caso de estudio. Identificar aspectos claves del uso de Machine Learning en el área de los videojuegos, analizar si los datos entregados son relevantes o si servirán para poder realizar predicciones en tareas de regresión y de clasificación, qué otros datos no presentes en el dataset podrían servir para un análisis más integral. Definir objetivos para trabajar en las siguientes fases.
- Data Understanding: Analizar y definir las características que contiene el dataset entregado, considerando tipos de datos, naturaleza de los datos, estadísticos, distribución, correlación, etc. Todo lo necesario para poder entender como se compone la fuente de información entregada, y con ello apoyar el entendimiento del negocio, identificando cuales son las transformaciones que se deben realizar en la fase posterior.
- Data Preparation: Transformar y ejecutar las rutinas de limpieza necesarias para que los objetivos planteados inicialmente puedan cumplirse. Esto considera al menos tratamiento de NaN, outliers, scaling, normalización, estandarización, encoder, ingeniería de características, etc. siempre considerando los objetivos planteados y la selección de features significativos de acuerdo al target seleccionado, utilizando técnicas como la correlación de pearson.



# Pauta de Evaluación

## Pauta tipo: Escala de valoración

Categoría	% logro	Descripción niveles de logro
<b>Muy buen desempeño</b>	<b>100%</b>	Demuestra un desempeño destacado, evidenciando el logro de todos los aspectos evaluados en el indicador.
<b>Desempeño aceptable</b>	<b>60%</b>	Demuestra un desempeño competente, evidenciando el logro de los elementos básicos del indicador, pero con omisiones, dificultades o errores.
<b>Desempeño incipiente</b>	<b>30%</b>	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.
<b>Desempeño no logrado</b>	<b>0%</b>	Presenta ausencia o incorrecto desempeño.

Indicador de Evaluación	Categorías de Respuesta				Ponderación del Indicador de Evaluación
	Muy buen desempeño 100%	Desempeño aceptable 60%	Desempeño incipiente 30%	Desempeño no logrado 0%	
Utiliza al menos 4 funciones necesarias para extraer elementos básicos de estadística descriptiva. Interpreta correctamente los elementos estadísticos obtenidos a partir de los datos.	10 pts.	6 pts.	3 pts.	0 pts.	10%
Interpreta los elementos estadísticos obtenidos a partir de los datos.	10 pts.	6 pts.	3 pts.	0 pts.	10%
Obtiene coeficiente de correlación por sí solo entre las características o bien la matriz de Correlación de Pearson.	10 pts.	6 pts.	3 pts.	0 pts.	10%
Interpreta los valores obtenidos identificando valores de correlación directa e inversa.	10 pts.	6 pts.	3 pts.	0 pts.	10%
Con el uso de estadísticos obtiene KPI relevantes como cantidad de derrotas, cantidad de triunfos, cantidad de jugadores, promedio de tiempo de partida, etc.	10 pts.	6 pts.	3 pts.	0 pts.	10%

Identifica cuando el dato es numérico continuo, discreto, o variable categórica nominal u ordinal.	10 pts.	6 pts.	3 pts.	0 pts.	<b>10%</b>
Trabaja los missing values (valores NaN) y los datos outliers, considerando las mejores prácticas con respecto al uso de los datos.	10 pts.	6 pts.	3 pts.	0 pts.	<b>10%</b>
Identifica cuales son características significativas para el contexto del objetivo que el equipo se planteó de negocio.	10 pts.	6 pts.	3 pts.	0 pts.	<b>10%</b>
Utiliza técnicas de transformación de datos adecuada de acuerdo a la naturaleza de estos.	10 pts.	6 pts.	3 pts.	0 pts.	<b>10%</b>
Utiliza las técnicas de transformación necesarias, ya sea de scaling, normalización o encoding dejando listos los datos para la fase de Modeling.	10 pts.	6 pts.	3 pts.	0 pts.	<b>10%</b>
<b>Total</b>					<b>100%</b>