



Pontificia Universidad  
**JAVERIANA**  
Bogotá

Facultad de Ingeniería  
Departamento de  
Ingeniería de Sistemas

Pontificia Universidad Javeriana  
Ingeniería de Sistemas  
**Procesamiento de datos**  
**Primer Parcial**

## Introducción y Reglas

El presente parcial tiene la finalidad de medir las capacidades del estudiante para realizar análisis utilizando las herramientas de procesamiento de datos, con la finalidad de generar valor frente a un problema de negocio dado. Por ende, se espera que este sea capaz, no únicamente de aplicar las herramientas, si no que use estas de la mejor manera posible para poder construir análisis que generen valor.

El parcial **se realizara de manera individual**, por lo que no se espera que se presenten respuestas similares en este. Esta actividad es el componente práctico del parcial que tendrá un valor de 70% sobre la nota general de este, el otro 30% sera una evaluación teórica que se realizara en la clase del 21 de septiembre de 2022.

El 70% del componente práctico se divide a su vez en un informe ejecutivo (40%) y una presentación de negocio (30%). La esencia del ejercicio, mas que la cantidad, es la calidad, por lo que se valora mas la calidad de los análisis generados mas que la cantidad de procesamientos que realicen. El enunciado del problema se presenta a continuación.

## Enunciado

Supongan que ustedes son contactados por un inversor millonario que quiere crear un nuevo equipo de futbol en la liga inglesa, el tiene el interés de conocer un poco mas los equipos y resultados que se obtuvieron en la liga de 17-18, para así poder tener una mejor idea de que tipos de jugadores preferiría contratar y que estilo de juego quiere que tenga su equipo. Para esto ustedes cuenta con la información de los resultados por partido, como también las estadística obtenidas por los equipos y jugadores.

Resultados por partidos: conjunto de datos “results.csv”.

Estadísticas por equipo: conjunto de datos “season\_stats.json”

Estadística de jugadores: conjunto de datos “players.csv”

\*Estos datos se encuentran en la plataforma.

Se espera que ustedes desarrollen 2 entregables, el primero constara de un informe ejecutivo donde se esperan los siguientes componentes:

1. Contexto general del problema.
2. Descripción de los conjuntos de datos y sus variables mas importantes (a consideración de ustedes seleccionar **6**).
3. Reporte de calidad de datos de los conjuntos de datos.
4. Descriptivas generales de las variables consideradas mas importantes (pueden ser de uno o mas conjuntos de datos).
5. Al menos **3** visualizaciones descriptivas de las variables.
6. **5** Hallazgos interesantes del problema de negocio (equipos? Jugadores?...), estos deben ser sustentados por descriptivos y/o gráficas.

Se espera que el documento tenga máximo 6 páginas + 1 portada.

El segundo entregale consta de una presentación ejecutiva de 5 diapositivas + 1 portada, donde presentaran lo que ustedes consideren mas importantes para el inversor. Contaran con 5 minutos para presentar sus conclusiones, estos 5 minutos continuaran con un par de preguntas realizadas por expertos de negocio, donde la respuesta de estas también sera evaluada de manera apreciativa.

Ademas, de estos dos entregales se deben presentar dos anexos, el primero sera el notebook de desarrollo del problema y el segundo un sustento de que el notebook fue ejecutado en databricks.