



Actividad 2

Aspectos generales

Aspectos generales

- **Objetivo:** evaluar individualmente el aprendizaje sobre análisis y visualización de datos en Python, a través de la construcción de una serie de tareas asociadas a las estadísticas de las ligas de fútbol europeo.
- **Lugar de entrega:** lunes 4 de septiembre a las 17:30 hrs. en el repositorio privado.
- **Formato de entrega:** **UNICAMENTE** el archivo Python Notebook (**A2.ipynb**) con la solución de la Actividad. El archivo debe estar ubicado en la carpeta **A2**. 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.
- **Librerías:** el uso de librerías externas, distintas a las cubiertas en el curso y que solucionen aspectos centrales o extensos de la actividad no podrán ser utilizadas. Solo se podrán usar las que han sido aprobadas explícitamente por el profesor durante la sesión, previa consulta de los estudiantes.
- **Entregas con errores de sintaxis y/o que generen excepciones en todas las ejecuciones** serán calificados con nota 1.0.

Introducción

En esta actividad utilizará un conjunto de datos que recopila información de 12.290 partidos de diversas ligas de fútbol europeo entre los años 2014 y 2020. Estos datos consideran los siguientes indicadores por cada equipo participante en cada partido:

- **league:** liga europea a la que corresponde el partido.
- **season:** año en que comenzó la temporada donde se realizó el partido.
- **date:** fecha en que se jugó el partido.
- **team:** equipo para el cual se indican los datos del partido.
- **h_a:** localía (h) o visita (a) del equipo analizado.
- **result:** resultado del partido para el equipo analizado: triunfo (w), derrota (l), empate (d).
- **pts:** puntos obtenidos en el partido por el equipo analizado.
- **goals_scored:** goles anotados por el equipo analizado.
- **goals_missed:** goles recibidos por el equipo analizado.
- **deep_passes:** pases exitosos realizados por el equipo analizado en los últimos 20 metros de la cancha (pases profundos).
- **deep_passes_allowed:** pases profundos exitosos realizados por el contrincante.
- **ppda:** índice de presión defensiva, calculado a partir de los pases realizados por el equipo contrincante por cada acción defensiva realizada por el equipo analizado (mientras más bajo, más presión).
- **oppda:** índice de presión defensiva del equipo contrincante.

En base a los campos recién descritos y utilizando las librerías presentadas en clases, deberá cumplir una serie de misiones relacionadas con el análisis de datos en Python.

Misiones

Todas las misiones son independientes entre sí y tienen el mismo puntaje, por lo que puede realizarlas en el orden que prefiera:

1. **Agregación:** utilizando los datos, construya un nuevo **DataFrame** con información agregada, que resuma la tabla de posiciones de al menos dos de las ligas, para cada año. El **DataFrame** debe presentar al menos la siguiente información por cada registro:

- **league:** liga europea a la que corresponde el equipo.
- **season:** año en que comenzó la temporada.
- **position:** posición final en la liga.
- **team:** equipo analizado.
- **matches:** número de partidos disputados durante la temporada.
- **pts:** puntos obtenidos durante la temporada.
- **wins:** número de partidos ganados durante la temporada.
- **draws:** número de partidos empatados durante la temporada.
- **defeats:** número de partidos perdidos durante la temporada.
- **goals_scored:** goles anotados durante la temporada.
- **goals_missed:** goles recibidos durante la temporada.
- **deep_passes_avg:** promedio de pases profundos logrados por partido durante la temporada.
- **deep_passes_allowed_avg:** promedio de pases profundos permitidos por partido durante la temporada.
- **ppda_avg:** promedio del índice de presión defensiva durante la temporada.
- **oppda_avg:** promedio del índice de presión defensiva de los equipos contrincantes durante la temporada.

El **DataFrame** debe estar ordenado, respetando el siguiente nivel de importancia: liga, año (ascendente), posición (ascendente). Como referencia, considere la siguiente tabla, que presenta los primeros cinco registros de un **DataFrame** que cumple lo solicitado:

league	season	position	team	matches	pts	wins	draws	defeats	goals_scored	goals_missed	deep_passes_avg	deep_passes_allowed_avg	ppda_avg	oppda_avg
La_liga	2014	1	Barcelona	38	94	30	4	4	110	21	12.868421	3.000000	5.683535	16.367593
La_liga	2014	2	Real Madrid	38	92	30	2	6	118	38	9.236842	4.026316	10.209085	12.929510
La_liga	2014	3	Atletico Madrid	38	78	23	9	6	67	29	5.184211	3.236842	8.982028	9.237091
La_liga	2014	4	Valencia	38	77	22	11	5	70	32	5.342105	4.526316	8.709827	7.870225
La_liga	2014	5	Sevilla	38	76	23	7	8	71	45	8.026316	4.421053	8.276148	9.477805

2. **Visualización:** basándose en el índice `ppda` y en un análisis visual, proponga una metodología que permita identificar los cambios de estilo de juego en los equipos. Esta metodología no tiene que ser infalible ni perfecta, pero debe tener una justificación razonable desde el punto de vista del análisis estadístico realizado. Muestre al menos 1 caso de aplicación de la metodología, indicando además un registro histórico de prensa para justificarlo.
3. **Caracterización de las ligas:** en base a los datos disponibles, realice un análisis que permita caracterizar y diferenciar a las distintas ligas. No es necesario que este análisis genere una métrica única de caracterización.

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 de esta actividad, se revisarán los procedimientos desarrollados para responder las diferentes misiones propuestas y la estructura de como utiliza los módulos `pandas`, `matplotlib` y `numpy` en ellos. Dado lo abierto de las misiones, se espera que las respuestas incluyan análisis y visualizaciones que permitan justificar las decisiones tomadas.

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/>).