



IIC2115 – Programación como Herramienta para la Ingeniería (I/2024)

Laboratorio 2

Aspectos generales

- **Objetivo:** evaluar individualmente el aprendizaje sobre análisis de datos tabulares con Pandas, en un problema práctico con datos deportivos.
- **Lugar de entrega:** Parte 1 lunes 08/04 a las 17:30, Parte 2 domingo 14 de abril a las 23:59 hrs, ambas en el repositorio privado.
- **Formato de entrega:** archivos Python Notebook L2.1.ipynb y L2.2.ipynb con las soluciones de las partes 1 y 2 del laboratorio. Los archivos deben estar ubicados en la carpeta L2. Utilice múltiples celdas de texto y código para facilitar el trabajo del cuerpo docente. Entregas que no cumplan el formato tendrán un descuento de 0,5 pts.
- **Entregas atrasadas:** el descuento por atraso para la Parte 1 es de 1 punto cada 10 minutos o fracción. El descuento por atraso para la Parte 2 es de 1 punto por cada hora o fracción.
- **Issues:** Las discusiones en las *issues* del Syllabus que sean relevantes para el desarrollo de la evaluación, serán destacadas y se considerarán como parte de este enunciado. Así mismo, el uso de librerías externas que solucionen aspectos fundamentales del problema no podrán ser utilizadas. Solo se podrán utilizar las que han sido aprobadas en las *issues*, 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.**

- exploración y descripción - completas columna 3pts - generar un único df - consultas

Descripción del problema

Considere la información contenida en los archivos `stats.csv` y `salaries.csv`, que describen estadísticas y sueldos, respectivamente, para cada jugador de la NBA (National Basketball Association) durante la temporada 2017-2018. En base a esto, y asumiendo que las tablas ya están pobladas, ud. deberá completar las siguientes misiones, utilizando una única consulta en SQL para cada una.

Parte 1: análisis exploratorio

Explore y procese los datos a través de las siguientes misiones:

1. Abra ambos archivos y describa el contenido de cada uno, indicando qué modela cada columna, su tipo de dato y sus características principales. Para esto, puede valerse de análisis numérico y/o visual.
2. Exceptuando las columnas relacionadas con los puntos convertidos, identifique aquellas columnas con datos extremos y/o faltantes y corríjelas o elimínelas, de acuerdo a algún criterio definido por usted.
3. Consolide la información en un único **DataFrame**, cuidando de combinar adecuadamente los sueldos con las estadísticas
4. Complete los valores faltantes para la columna `3Points`, cuando sea posible inferirlos a partir del resto de las columnas. Para aquellos que no es posible, impute el valor faltante de acuerdo a algún criterio definido por ud.

Parte 2: consultas

Responda a las siguientes consultas, en base a los datos generados en la parte anterior:

1. Encuentre los 5 equipos de la NBA para los que sus 5 jugadores más veteranos cometieron más *fouls*. Para cada equipo, obtenga además la cantidad de *fouls* agregada que realizaron estos jugadores. Ordene los resultados de manera ascendente en base a los *fouls* e imprímalos en pantalla.
2. Encuentre los 10 jugadores más costosos de la NBA de acuerdo a la cantidad de dinero que les pagaron por cada punto que convirtieron, solo considerando aquellos jugadores que convirtieron puntos. Para cada jugador, obtenga además su sueldo total en la temporada. Ordene los resultados de manera descendente, de acuerdo al sueldo total, e imprima los resultados en pantalla.

3. Para los jugadores que fueron parte de 2 o más equipos distintos durante la temporada, encuentre aquellos que tuvieron la mejora más significativa en cuanto a puntos convertidos por partido. Ordene los resultados de manera descendente en base a la métrica de mejora e imprímalos en pantalla.

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 ptos. al puntaje total.

Para la corrección se revisarán los procedimientos desarrollados para responder las diferentes misiones y consultas propuestas y la estructura de cómo utiliza las librerías en ellas. 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, Gemini 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/>).