



IIC2115 – Programación como Herramienta para la Ingeniería (I/2021)

Taller 4a

Objetivos

- Aplicar los contenidos de análisis exploratorio, limpieza y depuración de datos.

Entrega

- **Lenguaje a utilizar:** Python 3.6 o superior
- **Lugar:** repositorio privado en GitHub. Recuerde incluir todo en una carpeta de nombre **T4a**.
- **Entrega:** lunes 31 de mayo a las 16:50 hrs.
- **Formato de entrega:** archivo python notebook (**T4a.ipynb**) y archivo python (**T4a.py**) con la solución de este enunciado. Los archivos deben estar ubicados en la carpeta **T4a**. No se debe subir ningún otro archivo a la carpeta. Utilice múltiples celdas de texto y código para facilitar la revisión de su programa.
- **NO SE ADMITEN ENTREGAS FUERA DE PLAZO**
- Entregas con errores de sintaxis y/o que generen excepciones serán calificadas con nota 1.0.

Descripción del problema

Considere el conjunto de datos almacenado en el archivo `data.csv`, que contiene datos obtenidos a lo largo de los años sobre los niveles de Ozono (O_3) y material particulado de 2.5 micrómetros ($PM_{2.5}$). Además de esta información, cada registro está categorizado en tres niveles, en base al riesgo ambiental que presentan las mediciones de O_3 y $PM_{2.5}$ para la fecha: bajo, medio y alto. En base a toda esta información, complete las misiones indicadas a continuación.

Misión 0: aspectos básicos

Para cumplir las misiones de este taller, es su responsabilidad explorar inicialmente el contenido del archivo y familiarizarse con el formato en que está almacenada la información. Para eso, utilice los comandos `describe` y `head` de `pandas`.

Misión 1: limpieza y depuración

Tanto para O_3 como para $PM_{2.5}$, el conjunto contiene datos extremos y datos incompletos para algunos días, que fueron generados por motivos desconocidos. Con el fin de facilitar el análisis futuro, deberá **primero** corregir los datos extremos y luego ajustar los datos faltantes de 2 formas distintas. Para esto último, cree 2 nuevos `DataFrame`, en el primero de ellos complete los datos faltantes con la media, y en el segundo elimínelos.

Misión 2: descripción y comparación

A continuación, para ambos `DataFrame` generados en el ítem anterior y de manera independiente, imprima en una tabla ordenada los siguientes indicadores para O_3 y $PM_{2.5}$: media, desviación estándar, máximo, mínimo, Kurtosis. Además, agregue a esta tabla la correlación entre O_3 y $PM_{2.5}$.

Misión 3: visualización

Para ambos `DataFrame` obtenidos en el primer ítem y de manera independiente, genere las siguientes visualizaciones:

- Histograma de $PM_{2.5}$
- Boxplot de O_3 por mes
- Evolución promedio de O_3 y $PM_{2.5}$ por año.

Misión 4: categorización

En base a todos los análisis realizados anteriormente, proponga e implemente en Python un esquema para asignar un nivel de riesgo medioambiental para cada registro que no tiene esta información. Complete esto para ambos `DataFrame` de manera independiente. Comente y analice los resultados.

Corrección

La corrección de este taller se basará en lo adecuado de los mecanismos utilizados para realizar cada misión. En otras palabras, no existe *a priori* un resultado correcto para cada misión, por lo que cada misión se corregirá en base a lo adecuado y justificado que se encuentre el análisis. Cualquier supuesto que se haga para completar las misiones debe quedar claramente indicado.

IMPORTANTE: todas las celdas utilizadas deben estar ejecutadas al momento de entregar el taller, de modo que las salidas generadas sean visibles. En caso de no cumplir esto, su taller no será corregido.

Objetivo de participación

Para verificar la participación durante la clase, debe completar la Misión 1.

Política de Integridad Académica

“Como miembro de la comunidad de la Pontificia Universidad Católica de Chile me comprometo a respetar los principios y normativas que la rigen. Asimismo, prometo actuar con rectitud y honestidad en las relaciones con los demás integrantes de la comunidad y en la realización de todo trabajo, particularmente en aquellas actividades vinculadas a la docencia, el aprendizaje y la creación, difusión y transferencia del conocimiento. Además, velaré por la integridad de las personas y cuidaré los bienes de la Universidad.”

En particular, se espera que mantengan altos estándares de honestidad académica. Cualquier acto deshonesto o fraude académico está prohibido; los alumnos que incurran en este tipo de acciones se exponen a un procedimiento sumario. Ejemplos de actos deshonestos son la copia, el uso de material o equipos no permitidos en las evaluaciones, el plagio, o la falsificación de identidad, entre otros. Específicamente, para los cursos del Departamento de Ciencia de la Computación, rige obligatoriamente la siguiente política de integridad académica en relación a copia y plagio: Todo trabajo presentado por un alumno (grupo) para los efectos de la evaluación de un curso debe ser hecho individualmente por el alumno (grupo), sin apoyo en material de terceros. Si un alumno (grupo) copia un trabajo, se le calificará con nota 1.0 en dicha evaluación y dependiendo de la gravedad de sus acciones podrá tener un 1.0 en todo ese ítem de evaluaciones o un 1.1 en el curso. Además, los antecedentes serán enviados a la Dirección de Docencia de la Escuela de Ingeniería para evaluar posteriores sanciones en conjunto con la Universidad, las que pueden incluir un procedimiento

sumario. Por “copia” o “plagio” se entiende incluir en el trabajo presentado como propio, partes desarrolladas por otra persona. Está permitido usar material disponible públicamente, por ejemplo, libros o contenidos tomados de Internet, siempre y cuando se incluya la cita correspondiente.