



IIC2115 – Programación como Herramienta para la Ingeniería (II/2017)

Actividad 13

Objetivos

- Introducir diferentes herramientas de análisis de datos en Python.

Entrega

- **Lenguaje a utilizar:** Python
- **Lugar:** GitHub
- **Hora:** 17:30

Instrucciones

Laika está muy preocupada por el medio ambiente y sospecha que, así como vamos contaminando, pronto acabaremos con nuestro planeta. Para poder demostrarle este hecho al mundo y comenzar a concientizarlos sobre la importancia de cuidar nuestro planeta, Laika decide analizar datos obtenidos a lo largo de los años sobre los niveles de Ozono (O_3) y material particulado de 2.5 micrómetros ($PM_{2.5}$).

Laika sabe que debe comenzar esta campaña por algún lado, por lo que decide delegarles esta labor a los alumnos del curso IIC2115 para que sientan en su propio pellejo los cambios que están ocurriendo en el mundo. Ella no sabe realmente si los datos obtenidos demuestran su suposición, por lo que te toca a ti obtener todo lo necesario para poder entregarle a Laika un análisis descriptivo de los datos, de manera que ella sea capaz de aplicar sus vastos conocimientos para llegar a una conclusión justificada.

Paso 1: Limpieza y depuración de Datos

El *Dataset* obtenido, llamado `classified_info.csv` contiene datos extremos que seguramente aparecen por causas ajenas a nuestro análisis, así como también presenta datos incompletos en algunos días. Es por esto que, con el fin de realizar un buen análisis, deberás eliminar los datos extremos y ajustar los datos faltantes de 2 formas distintas, con el objetivo de comparar el comportamiento de la distribución de los datos al trabajar de distintas formas la información incompleta. Para esto último, creen 2 nuevos *Datasets*. En uno de ellos rellenen los datos faltantes con la media, y en el otro borren los datos inexistentes. Considera utilizar la librería `pandas`.

Paso 2: Rellenar los datos categóricos

Se logró descubrir que la variable categórica *Environmental_risk* – que puede ser nulo, bajo, medio, alto o extremo – depende de los niveles de Ozono y PM2.5 de cada mes. Bajo este supuesto, debes obtener dicha variable a partir de un Decision Tree para el *Dataset* correspondiente.

Paso 3: Describir y comparar los datos obtenidos

Laika espera obtener una descripción de los datos a partir de variables descriptivas. Es por eso que tu programa debe imprimir la media, desviación estándar, máximo, mínimo, Kurtosis y correlación entre Ozono y PM2.5 en pantalla de manera **ordenada**, al fin y al cabo, Laika no se merece tener que descifrar la información que le entreguen a causa del desorden¹.

Paso 4: Visualizar datos

Finalmente, Laika desea poder visualizar algunos datos sobre ambos *dataset*. En particular quiere:

- Histograma de probabilidad de PM2.5.
- Boxplot de PM2.5 por mes.
- Evolución por año del O_3 y PM2.5

¹De hecho, Laika se merece el cielo, ¡Aguante Laika!

Política de Integridad Académica

Los alumnos de la Escuela de Ingeniería deben mantener un comportamiento acorde al Código de Honor de la Universidad:

“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.