



IIC2115 – Programación como Herramienta para la Ingeniería (I/2020)

Actividad de Participación 5 y 6 - Análisis de datos y manejo de GIS

Objetivos

- Utilizar correctamente las librerías Pandas y Geopandas para realizar análisis de los datos con componentes geográficas.

Instrucciones

Esta es una evaluación que implica realizar la programación de las soluciones de manera estrictamente individual. Así, si usted utiliza código de terceros para la realización de esta evaluación, habrá cometido una falta que el profesor, de saberlo, deberá reportar a la dirección de pregrado de la Escuela. Teniendo esto presente, ud. podrá intercambiar información con terceros acerca de esta evaluación y también buscar la información en otros recursos. También está permitido utilizar código previamente publicado en internet, siempre y cuando sea debidamente referenciado, sin que esto sea considerado una falta a la ética académica.

Luego de entregar la actividad, deberá responder en el Siding un cuestionario relativo a estas instrucciones. El no responderlo en los tiempos establecidos implicará que su solución no será corregida.

Entrega del código

- **Lenguaje a utilizar:** Python 3.6
- **Entrega:** lunes 15 de junio a las 23:59 hrs.
- **Lugar:** repositorio privado en GitHub.

- **Formato de entrega:** Ambas actividades deben entregarse en carpetas separadas, o sea, las misiones de la P05 deben entregarse en la carpeta P05, mientras que las de la P06 debe hacerse en la carpeta P06. Para cada una de las actividades, debe entregar un archivo python notebook (P05.ipynb o P06.ipynb) y un archivo python (P05.py o P06.py) con la solución a las misiones. No se debe subir ningún otro archivo a las carpetas. Utilice múltiples celdas de texto y código para facilitar la revisión de su programa.
- **Descuentos:** Entregas que no cumplan el formato de entrega tendrán un descuento de 0.5 pts.
- Entregas atrasadas, con errores de sintaxis y/o que generen excepciones serán calificadas con nota 1.0.
- Todos los comandos e instrucciones importantes de cada solución deben ser implementados usando funciones de *pandas*, *matplotlib* y/o *sklearn*. Esto significa que no está permitido iterar sobre los DataFrame, calculando con funciones básicas de Python lo requerido.
- Está prohibido el uso de librerías externas distintas a las mencionadas anteriormente, que sean estructurales en la solución de los problemas.

Actividad P05

Considere el archivo de datos **distritosRM.csv** disponible en el link (<https://bit.ly/2C9IrdH>). El archivo csv contiene los resultados de la encuesta ADIMARK que busca identificar los distintos gupos socio-económicos distribuidos en la ciudad. posee las siguientes columnas de información:

- **NOM_REG:** Nombre de la región
- **COD_PROV:** Código de la provincia
- **NOM_PROV:** Nombre de la provincia
- **COD_COM:** Código de la comuna

- **NOM_COM**: Nombre de la comuna
- **NOM_DIS**: Nombre del distrito
- **CODIGO**: Código del distrito
- **E**: Número de personas del grupo socio-económico E
- **D**: Número de personas del grupo socio-económico D
- **C3**: Número de personas del grupo socio-económico C3
- **C2**: Número de personas del grupo socio-económico C2
- **ABC1**: Número de personas del grupo socio-económico ABC1
- **ADIMARK_T**: Número total de personas consideradas
- **Area**: Área del distrito
- **Densidad**: Densidad del distrito
- **1_UF_M2**: Valor del m^2 del suelo en UF.

Importe el archivo mediante **pandas** para crear un **DataFrame** y realice las siguientes misiones:

- M1. Analice la existencia de valores por defecto, nulos, incompletos y outliers en todas las columnas numéricas. Para cada uno de estos casos, mantenga, complete o modifique la información de acuerdo a algún criterio debidamente justificado. **(2.0 ptos.)**
- M2. Construya un **DataFrame** que capture la información agregada a nivel de comuna. Elija para cada columna numérica la función de agregación adecuada, considerando que para las columnas que indican sector socioeconómico, se esperan indicadores porcentuales con respecto a los encuestados de cada comuna. **(2.0 ptos.)**
- M3. Construya un predictor para el valor del m^2 por distrito. Utilice una regresión lineal como modelo y reporte el error en la estimación como cifra promedio, e identifique los distritos con el mayor y el menor error en la estimación. **(2.0 ptos.)**

Actividad P06

Considere los archivos geográficos **geoDistritosRM.shp** disponible en el link (<https://bit.ly/2MVRck6>) y **paradas.shp** disponibles en el link (<http://www.dtpm.cl/descargas/pops20/P0%202002-03-2020%20Paradas%20SerFrec.zip>).

Importe los archivos mediante **geopandas** para crear **GeoDataFrames** y realice las siguientes misiones:

M1. Integre la información geográfica de su **GeoDataFrame** de distritos a su **DataFrame** de la actividad **P05**. Luego convierta el **DataFrame** resultante a un **GeoDataFrame**. **(2.0 ptos.)**

M2. Determine la densidad de paraderos de buses para cada distrito y la densidad de paraderos respecto al número de personas. Realice dos visualizaciones de ambos indicadores en mapas diferentes, mostrando solo geometrías que tienen valor de m^2 en UF mayor a 2. **(4.0 ptos.)**

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.