

Trabajo Final

1. Web scrapping y Geocoding (Python script)

Se pide crear un Dataframe de las universidades a nivel nacional (público y privadas) con la información de licenciamiento de la Sunedu. Para ello debe extraer información de la web de Sunedu [Click aquí](#). Usted puede observar que existe dos tablas: uno con información de universidades pública y otra de universidades privadas. El shapefile del Perú a nivel distrital se ubica en la carpeta **shape_file**.

- Usted debe extraer la tabla de universidades públicas y privadas. Luego debe juntar ambas.
- Crear una Dummy que toma el valor de 1 si la universidad es privada, y cero en universidades públicas.
- Crear un Dummy igual a 1 si la universidad recibió el lincencimiento de Sunedu. Notese que la columna Disposición legal de licenciamiento proporciona la información solicitada. Aquellas universidades sin información en esta columna no recibieron el lincenciamiento de la Sunedu.
- Obtener la información de latitud y longitud de todas las universidades (Usar Geocoding).
- Exportar la base de datos en formato excel. Guarde el excel con el nombre de su grupo universidades_grupo(Numero de grupo).
- Ubicar las universidades en un mapa del Perú. Para ello, crear un mapa con las universidades públicas y distinga entre aquellas que recibieron licenciamiento o no. Luego, crear otro mapa con la misma información pero en el caso de universidades privadas ¹

2. Manejo de información geo-referenciada (Python script)

Usted debe trabajar con información cartográfica del Censo 2017. El archivo shapefile respectivo está en la carpeta **5_Informacion_Cartografica_Shape**. La información geográfica de los centros de salud mental para Lima está en el siguiente excel **Salud_mental_peru.xls**.

- Subir la base de datos de las manzanas en formate GeodataFrame. Crear los centroides de las manzanas.
- Crear una variable dummy llamado **Treatment** que toma el valor de 1 si el centroide de la manzana se ubica a menos de 1 km de algun centro de salud mental.

¹Inserte una leyenda a cada gráfico con las siguientes etiquetas (Universidad licenciada, Universidad no licenciada). Inserte como título a cada gráfico: Distribución geográfica de universidades Públicas y Distribución geográfica de universidades privadas respectivamente.

- Crear una dummy llamado **Control** que toma el valor de 1 si la manzana se ubica entre 1 a 2 km del centro de salud mental.
- Grafique las manzanas treatment en color rojo y en azul las manzanas de control. Mientras de color blanco o neutral, las manzanas que no son control o treatment. Usar como fondo de fondo de la imagen a Google maps. Añada una leyenda al gráfico² y el siguiente titulo (Mental Health Center - treatment and control). Use hexbin para sombrear las manzanas tratadas, controles y neutrales.

3. Calificación

3.1. Research assistance Report

- Usted debe explicar las etapas seguidas para resolver cada item.
- En una subsección llamada **Loops and functions** Describir cómo funciona los loops que ha creado. Describir las funciones que haya creado.
- En una subsección llamada **Clean dataframe** debe comentar los filtrados de datos, si detectó duplicados, manejo de missing values, reemplazo de valores, etc.
- En una subsección llamada **Variables Section** debe explicar cómo ha creado nuevas variables.
- Debe añadir imágenes en los item que solicitan crear gráficos.
- El archivo en PDF debe entregarse en Paideia. Está habilitado la herramienta cuestionario.

4. Información complementaria

1. La bases de datos se ubican en la carpeta **Trabajo_final** dentro de la carpeta data que usamos en los laboratorios.
2. Guardar los scripts y archivos de excel en la carpeta **Entrega.TF**
3. Por favor, use el relative path del repositorio del curso en sus scripts finales. Yo simplemente corregí los scripts y espero no encontrar errores. La primera regla de una research assistance es que el Principal investigator (PI) no encuentre errores cuando corra los códigos.
4. Oportunidad de mejorar la nota de participación: Ordene su repositorio con todos los scripts elaborados durante el semestre en su repositorio. Esto puede usarlo para presentarse a trabajos, y vean su códigos.
5. Recuerde comentar sus códigos de manera breve.
6. Se valorará la originalidad en los códigos. NO INTENTE PLAGIAR EL TRABAJO DE SUS COMPAÑEROS.

²Use como etiquetas : Treatment y Control respectivamente

Ejemplo de cómo subir un gráfico en overleaf.

