

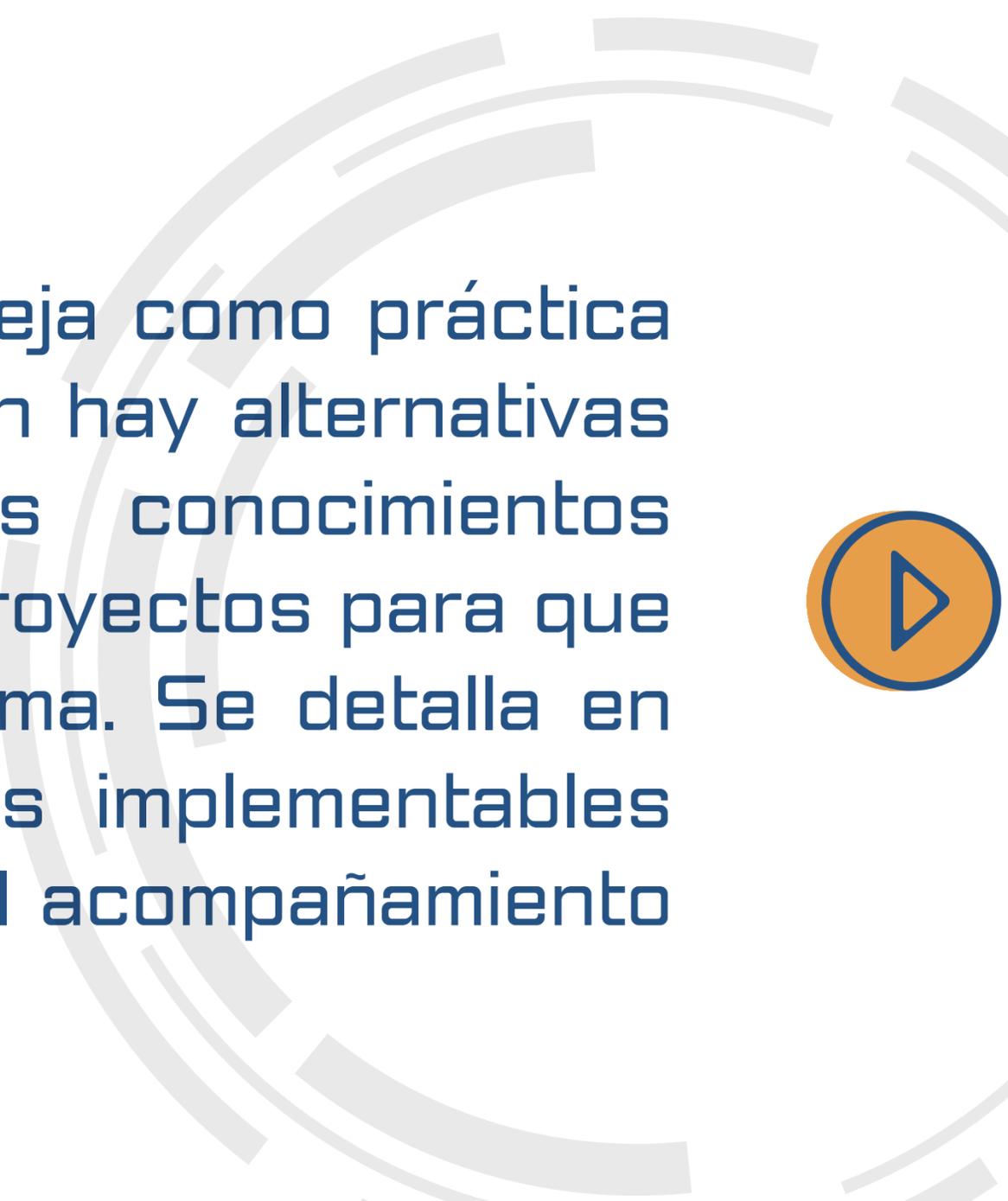


MÓDULO 3

NIVEL

EXPLORADOR





En el último bloque de 10 horas se deja como práctica asistida para el proyecto final. Si bien hay alternativas que se pueden planear con los conocimientos adquiridos, es necesario acotar los proyectos para que sean alcanzables según el cronograma. Se detalla en este documento algunas alternativas implementables con los conocimientos adquiridos y el acompañamiento de los docentes, tutores y monitores.





Planteamiento del proyecto El objetivo del proyecto es poner en práctica los conocimientos adquiridos durante las sesiones del bootcamp. Para esto se propone como objetivo del proyecto el análisis exploratorio y la visualización de algún conjunto de datos reales. En general, estos conjuntos de datos tienen problemas como valores faltantes, valores nulos, columnas que no aportan información, filas que no aportan información, errores, entre muchos otros problemas comunes en el análisis de datos. El proyecto debe cumplir los siguientes requisitos:



Se deben emplear bases de datos con información real. Para esto se sugieren los conjuntos de datos abiertos gubernamentales (disponibles en <https://www.datos.gov.co/browse?sortBy=newest&utf8=%E2%9C%93>). Utilizando los filtros del panel izquierdo se puede localizar algún tema o conjunto de datos que sea de interés, descargar y usar. También se pueden emplear datos de otros países o generados por otras entidades, siempre y cuando se puedan exportar en formato .CSV (por ejemplo, también se pueden buscar datos en Kaggle). Debe contener una fase exploratoria que permita saber qué hay en el conjunto de datos, que variables son útiles y qué acciones de limpieza necesitan ser realizadas



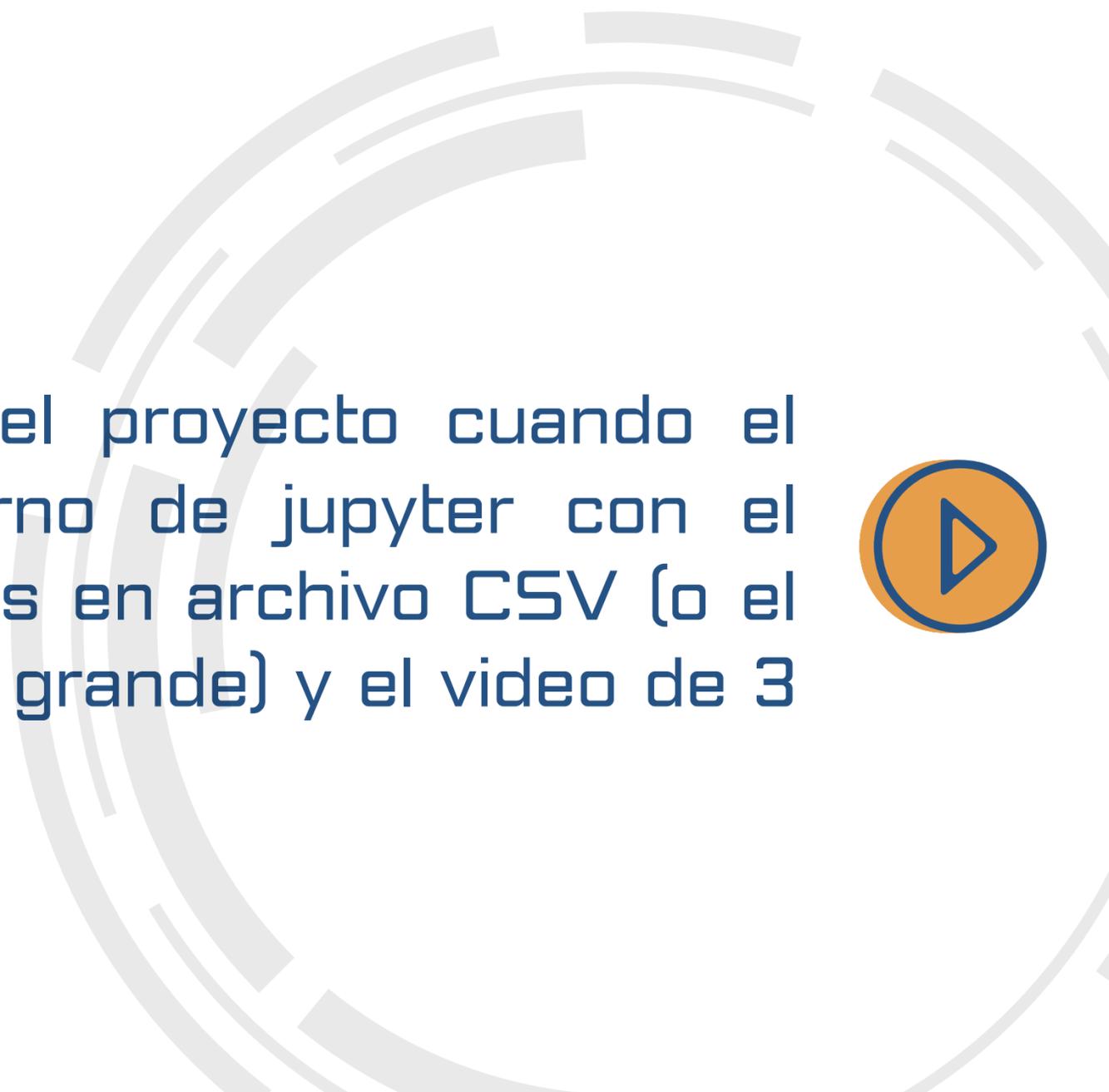


Debe contener una fase de tratamiento y limpieza de los datos. En esta fase se arreglarán los problemas de faltantes, nulos, errores, entre otros vistos en el curso. Debe estar en formato de Jupyter notebook. Para la evaluación del proyecto se debe grabar un video de máximo 3 minutos en donde se muestren los resultados (pueden ser los gráficos y textos del cuaderno de jupyter notebook) y las conclusiones que se pueden sacar del análisis exploratorio de los datos para ese conjunto de datos seleccionado.



Metodología de trabajo del proyecto Para el desarrollo del proyecto los estudiantes contarán con 10 horas de acompañamiento en el módulo tres (las 10 últimas horas del bootcamp). En esas horas los estudiantes podrán recibir asesoría de los docentes, resolver dudas, preguntar lo que necesiten para afianzar los conocimientos, entender lo explicado en el bootcamp y afianzar sus conocimientos. El proyecto puede realizarse de forma individual o en grupos de máximo dos personas. Una vez se seleccione el conjunto de datos, se debe mostrar al tutor, este revisará el proyecto y aprobará o sugerirá cambios en el conjunto de datos. Para validar el avance, semanalmente cada estudiante presentará su progreso en una sesión corta de 3 a 5 minutos en donde brevemente mostrará qué ha logrado y en qué se enfocará en trabajar en la presente semana. Esto con el fin de detectar si hay grupos o estudiantes que tienen problemas y poder recibir ayuda oportunamente.





Entregables Se considera completo el proyecto cuando el estudiante haya entregado el cuaderno de jupyter con el conjunto de datos, el conjunto de datos en archivo CSV (o el link en caso de que el archivo sea muy grande) y el video de 3 minutos con la explicación del proyecto.

