

R o Python? R y Python! Cómo colaborar en proyectos bilingües (sin sufrir)

Riva Quiroga , Stephanie Orellana , Sara Acevedo

Palabras clave: python - estructuras de datos - gestión de datos

Es habitual ver debates en foros y redes sociales acerca de si es mejor usar R o Python para ciencia de datos. Sin embargo, esta discusión no suele ser muy productiva, ya que en la vida real muchas veces debemos participar en proyectos en los que se utiliza tanto R como Python. A veces, porque las distintas etapas del proceso así lo requieren; otras, porque para algunas personas del equipo puede resultar más fácil resolver determinada tarea con uno de estos lenguajes de programación por sobre el otro. Sea cual sea la razón, es importante tener en cuenta que no siempre tenemos control sobre todo el flujo por el que pasan los datos con los que tenemos que trabajar, ni qué herramientas utilizarán las personas involucradas.

En esta charla relámpago mostraremos algunas consideraciones que es necesario tener cuando participamos en proyectos bilingües en los que se usan ambos lenguajes de programación. En particular, pondremos el foco en problemas que pueden surgir cuando trabajamos en R con datos que antes fueron procesados o analizados en Python, como la presencia de enteros de 64 bits, el uso de listas y diccionarios de Python dentro de un dataframe, las distintas formas de representar valores faltantes, entre otros. A lo largo de la charla, se mostrarán ejemplos concretos de problemas recurrentes, se explicará cómo resolverlos en R y se discutirá cómo prevenir su ocurrencia acordando lineamientos para la gestión de datos dentro del equipo de trabajo. A través de esta presentación, la audiencia aprenderá estrategias concretas que podrá utilizar en su próximo proyecto bilingüe de ciencia de datos para evitar problemas que le pueden hacer perder tiempo y motivación.