

Desafíos Técnicos Identificados:

Detección de Bots: TripAdvisor puede implementar medidas de seguridad para detectar bots y bloquear solicitudes automáticas. Esto podría obstaculizar el raspado.

- Solución: Usar el encabezado "User-Agent" para simular un navegador real y evitar ser detectado como un bot.

Estructura de la Página Cambiante: TripAdvisor puede modificar la estructura de su sitio web, lo que rompería las selecciones y búsquedas de elementos.

- Solución: Utilizar técnicas de scraping dinámico para adaptarse a cambios en la estructura.

Paginación y Límites: TripAdvisor puede limitar la cantidad de resultados por página y las páginas accesibles, limitando la cantidad de datos que se pueden extraer.

- Solución: Implementar paginación y bucles para recopilar resultados de múltiples páginas.

Duplicación de Datos: Algunas páginas web pueden mostrar resultados duplicados, lo que lleva a datos duplicados en la extracción.

- Solución: Implementar lógicas para eliminar duplicados durante la extracción y el almacenamiento de datos.

Calificaciones de Viajeros: La obtención de calificaciones de viajeros puede ser desafiante debido a la falta de una estructura numérica.

- Solución: Implementar un mapeo de las etiquetas de calificación a valores numéricos para capturar la calificación de viajeros (bubble o star)

Monitoreo Continuo: Es fundamental monitorear el proyecto para detectar problemas a medida que surgen y ajustar el scraping en consecuencia.

- Solución: Establecer un sistema de monitoreo continuo para supervisar la ejecución del proyecto y responder a problemas en tiempo real.