

## Prueba de Automatización

### Extracción de Datos de Wikipedia usando Python y Selenium

#### Objetivo:

Demostrar habilidades en la automatización de tareas de extracción de datos web y generación de archivos Excel mediante Python y Selenium.

#### Descripción de la tarea:

El candidato deberá crear un script automatizado con las siguientes acciones:

1. Creación de Cuenta y Login en Wikipedia:
  - a. Automatizar la creación de una cuenta en Wikipedia (Opcional, pero con valor adicional).
  - b. Loguearse en Wikipedia con la cuenta creada.
2. Extracción de Información de Entidades Federativas de México:
  - c. Navegar a la página "[Entidades federativas de México](#)" en Wikipedia.
  - d. Extraer la siguiente información:
    - a) Información de la tabla de Entidades federativas de México por superficie, población y densidad.
    - b) Información de la tabla de Población histórica de México de Censos.
    - c) Información de la tabla de Población histórica de México de Proyecciones de población.
  - e. Para cada estado en la lista, navegar a su página específica y extraer la sección de "Toponimia".
2. Extracción de Información de Estados de los Estados Unidos:
  - a. Navegar a la página "[Estado de los Estados Unidos](#)" en Wikipedia.
  - b. Extraer la siguiente información:
    - i. Listado de Estados.
  - c. Para cada estado en la lista, navegar a su página específica y extraer la sección de "Etimología".
3. Generación de Archivos Excel:
  - a. Crear un archivo Excel para cada extracción anterior realizada tanto para México (a, b y c) como para Estados Unidos (a). Cada archivo debe tener un nombre descriptivo y la información debe estar organizada de manera clara.

## Entregables:

1. Script Automatizado:
  - b. Desarrollado en Python.
  - c. Uso de la librería Selenium para la automatización.
  - d. Debe ser compatible con al menos dos navegadores web (por ejemplo, Chrome y Firefox).
  - e. El código debe estar bien organizado, siguiendo buenas prácticas, ser limpio y reusable.
  - f. Implementar funcionalidades básicas que permitan reutilizar el código para diferentes páginas de Wikipedia o incluso otras páginas web.
2. Estructura de Clases:
  - g. Una clase para México que permita extraer la información requerida para ese país.
  - h. Una clase para Estados Unidos que permita extraer la información requerida para ese país.
3. Archivos Excel Generados:
  - i. Con información extraída, organizada de manera clara.
4. Documentación:
  - j. Comentarios en el código que expliquen las funciones principales y su propósito.
  - k. Un README que detalle cómo ejecutar el script, sus dependencias y cualquier otra consideración relevante.

## Criterios de Evaluación:

1. **Correctitud y Complejidad:** La información extraída debe ser la solicitada y estar correctamente organizada en los archivos Excel.
2. **Robustez y Eficiencia:** El script debe ser capaz de manejar excepciones y errores durante su ejecución sin fallos.
3. **Tiempo de ejecución:** El tiempo de ejecución debe ser razonable y no presentar demoras innecesarias.
4. **Calidad del Código:** El código debe seguir buenas prácticas, ser limpio, organizado y fácil de entender.
5. **Reusabilidad:** Las funcionalidades implementadas deben ser genéricas para permitir su uso en diferentes páginas o contextos.
6. **Modularidad:** Separación clara entre funciones, clases y módulos para facilitar la mantenibilidad.

7. **Compatibilidad:** El script debe ser probado y ser compatible con al menos dos navegadores web.
8. **Documentación:** La claridad y utilidad de los comentarios en el código