Área de Datos

Fuentes:

* INDEC

Archivo: CBT (Canasta básica) Archivo: XLSX y XLS y Datos calculados en XLSX

IPC (Índice de precios al consumidor) XLS

IPI (IPI Nacion) XLS

* Argentina.gob.ar // Trabajo.gob.ar

Archivo: SIPA (Situación y Evolución del Trabajo Registrado) XLSX

EMAE (Estimador mensual de actividad económica) XLS

* Datos.gob.ar

Archivo: RIPTE (Remuneración imponible promedio de los trabajadores estables) CSV Y HTML

SMVM (Salario mínimo vital y móvil) CSV

* Datos.produccion.gob.ar

Archivo: Puestos de trabajo CSV

Salarios CSV

* Informes de la dirección

Archivo: Censo (Con estimaciones de años venideros) XLSX

IPICORR (Se reciben informes y se calculan) SHEETS WS

* DNRPA

Archivo: Registro de la Propiedad Automotor HTML

* Banco Nacion // Ambito Financiero

Dolar oficial desde 2003 CSV

Dolar blue desde 2003 HTML

Dolar MEP desde 2020 HTML

Dolar CCL desde 2013 HTML

Proceso:

* **Acceso a Sitios Web:** Utilizamos la herramienta "Selenium" en conjunto con Python para emular un navegador Chrome. Esto nos permite acceder a sitios web específicos a través de sus URL.
* **Descarga de Datos:** Una vez que el navegador se inicia, utilizamos Python para acceder directamente a los enlaces deseados y descargar los archivos necesarios desde la web.
* **Limpieza y Transformación de Datos:** Posteriormente, realizamos un proceso de limpieza y transformación de los datos descargados para obtener la información necesaria en un formato legible y estructurado.
* **Uso de DataFrames:** Utilizamos DataFrames para organizar y estructurar los datos. Esto nos permite gestionar y manipular los datos de manera eficiente.
* **Almacenamiento en MySQL:** Cada conjunto de datos obtenidos se almacena en una tabla específica en nuestra base de datos MySQL. Esta base de datos se utiliza para la gestión y almacenamiento centralizado de los datos.
* **Actualización Automatizada:** Para mantener los datos siempre actualizados, hemos implementado un Gateway que automatiza la actualización de los datos en la base de datos MySQL a intervalos regulares.

Presentación de Datos:

* **Visualización en Power BI:** Utilizamos Power BI como una herramienta de visualización y presentación de datos. Power BI nos permite crear informes y gráficos interactivos que facilitan la interpretación de los datos scrapeados.

Fuentes de Datos:

Todos los datos que recopilamos provienen de portales de acceso público proporcionados por el gobierno o entidades pertinentes.

Herramientas que usamos:

* **Selenium:** Se utiliza para emular un navegador web, en particular Google Chrome, lo que permite la interacción con sitios web a través de scripts automatizados.
* **Python:** Python es el lenguaje de programación que se utiliza para escribir scripts y automatizar la adquisición de datos, navegación web y manipulación de los archivos descargados.
* **DataFrames:** Estos se utilizan para organizar y estructurar los datos obtenidos de manera eficiente, lo que sugiere que podrían estar utilizando bibliotecas de Python como Pandas para este propósito
* **MySQL:** Se utiliza como sistema de gestión de bases de datos (DBMS) para almacenar y gestionar los datos recopilados en tablas específicas.
* **Power BI:** Es una herramienta de visualización de datos que se emplea para presentar los datos scrapeados de manera más organizada y en forma de gráficos interactivos.
* **Gateway de Actualización Automatizada:** Se utiliza para mantener los datos en la base de datos MySQL actualizados de manera automática, lo que implica la utilización de alguna solución o script para la actualización programada de datos.
* **Visual Studio Code:** Visual Studio Code, comúnmente abreviado como VS Code, es un entorno de desarrollo integrado (IDE) muy popular. Se utiliza para escribir, depurar y administrar el código fuente en varios lenguajes de programación. Es altamente personalizable y cuenta con una amplia gama de extensiones que pueden mejorar la eficiencia de desarrollo y la productividad.
* **Github:** GitHub es una plataforma de desarrollo colaborativo basada en la nube que se utiliza para gestionar el control de versiones y alojar proyectos de software. Permite a los equipos de desarrollo trabajar juntos en proyectos, realizar seguimiento de cambios en el código y colaborar de manera efectiva en el desarrollo de software. GitHub es ampliamente utilizado en la comunidad de desarrollo de software para alojar y compartir repositorios de código.
* **Excel:** Microsoft Excel es una aplicación de hojas de cálculo ampliamente utilizada para el análisis y la manipulación de datos. Puede ser útil para tareas de análisis de datos, visualización simple y manipulación de datos tabulares.
* **ChatGPT:** Es una inteligencia artificial de lenguaje natural desarrollada por OpenAI que puede responder preguntas, proporcionar información y asistir en una variedad de tareas relacionadas con el lenguaje natural. Puede ser útil para obtener respuestas a preguntas o para asistencia en la comunicación escrita.