

DATA
WAITERS

VALIDADOR EPS



INTRODUCCIÓN

Data Agent fue creado con el propósito de ofrecer soluciones de datos, con el objetivo que en la era de la información, cualquier empresa u organización, puede emplear soluciones de software, beneficiarse de ellas, con bajos costos y obteniendo valor de ellos de casi de manera inmediata, haciendo la toma de decisiones de las organizaciones para mantenerse competitivas en el mercado.

EL PROYECTO

El proyecto esta realizado con el objetivo de ofrecer una solución en datos completa en ingestión, transformación, almacenamiento y análisis para un grupo de interés de la entidad prestadora de salud, que en este caso son sus pacientes.

El producto está hecho para demostrar el potencial y beneficio que ofrece el tener un producto de datos, en las siguientes instancias:

- Tiempo: acceso inmediato a la información.
- Productividad: procesos automatizados.
- Confianza: mínima intervención en los procesos.
- Manejo: decisiones basadas en información.
- Compresión: análisis de puntos de mejora y necesidades.

PLANIFICACIÓN

DELIMITACIÓN DEL PROYECTO

El proyecto está hecho con el propósito resolver la manipulación de datos manual, el cual presenta múltiples problemas:

- Ineficiencia: se aplican las mismas formulas una y otra vez, una tarea repetitiva que puede programarse en software.
- Errores: se manejan una cantidad de datos de decenas de miles de datos, que es una cantidad compleja de manejar para un ser humano, pero una muy sencilla para un software.
- Improductividad: una persona puede durar semanas manipulando estos datos, con un software puede realizarse la tarea en segundos.

El objetivo del software de datos es solucionar estos tres puntos en el tratamiento de datos, por lo cual su eficiencia, seguridad y productividad nos permitirá también:

La propuesta permite la ingesta de un set de datos de los pacientes de la organización, el cual procesará, limpiará y almacenará en un repositorio de datos para su posterior:

- Analizar la calidad de los datos, realizar un informe de los errores en la adquisición de los datos y proponer puntos de mejora.
- Construir un modelo de análisis que nos permita visualizar que tipo de datos tenemos y obtener algunas conclusiones de valor para el negocio.

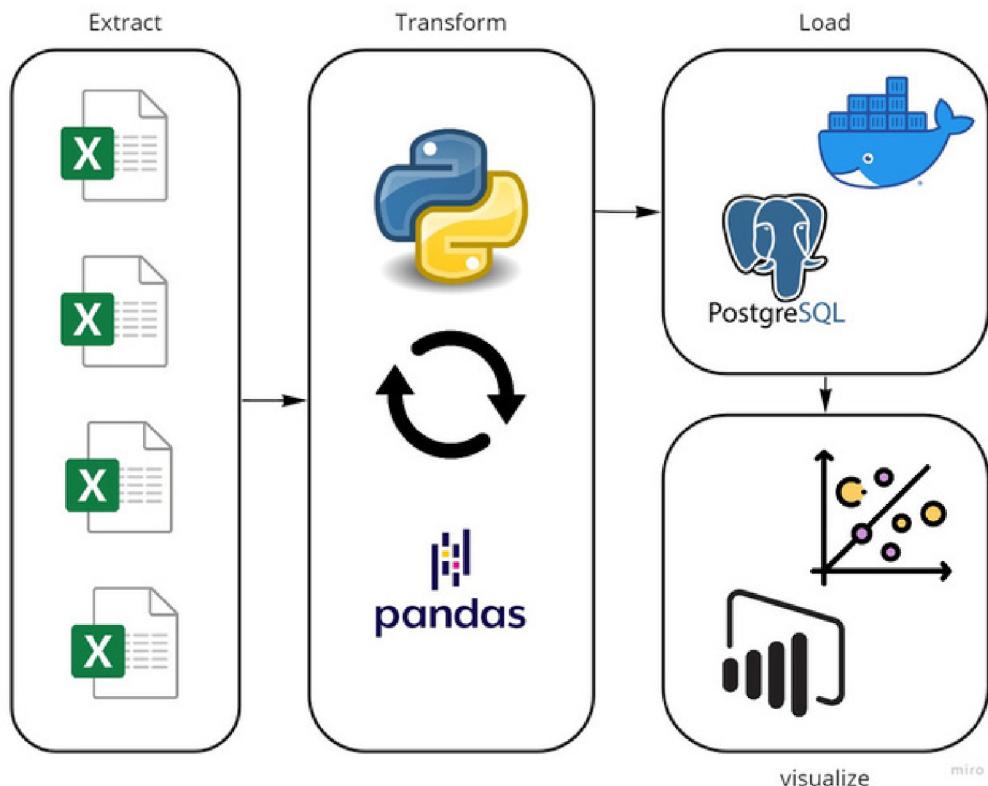
Este proyecto no busca determinar bajo ningún concepto, remplazar el criterio en salud de los profesionales capacitados para ello, tampoco se propone como una herramienta para tomar decisiones médicas, se propone como un modelo informativo y exploratorio para un grupo de interés en particular de una organización.

| | | |
|---|---------------------------------|--|
| 1 | Ingesta y transformación | Programa que detecta anomalías en los datos, duplicados, datos incoherentes de forma automática. |
| 2 | Almacenamiento en OLAP | Base de datos relacional, optimizada para el análisis de datos |
| 3 | Visualización de datos | Tablero informativo con un informe de los datos y un análisis descriptivo de ellos. |

DISEÑO

Modelado sobre solución en datos propuesta. Su enfoque es utilizar tecnologías Open Source, las cuales garantizan bajos costos, seguridad y flexibilidad respaldada por gigantescas comunidades encargadas de mantener el software seguro y accesible a todos.

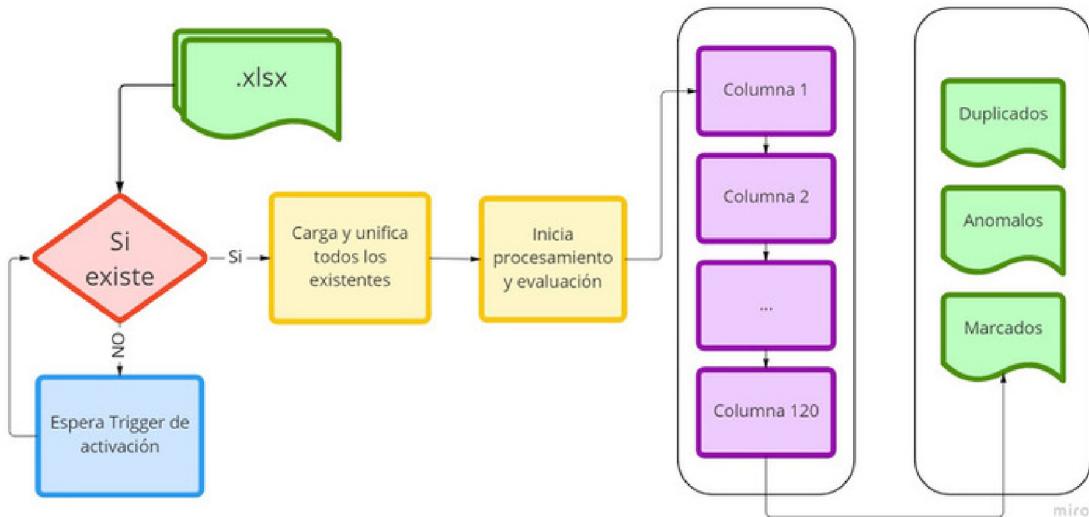
ARQUITECTURA DE LA SOLUCIÓN



- **Extracción:**
 - Se realiza una especie de Storage en donde se llegan los archivos .xlsx para el posterior procesamiento, en este caso utilizaremos un almacenamiento local.
- **Transformación:**
 - Se usa un Script de Python, con su framework para el procesamiento de datos, Pandas, que nos permite la manipulación de datos tabulares.
- **Carga:**
 - Se usa el Sistema de Bases de Datos Relacional PostgreSQL, de código abierto, como un sistema OLAP (Online Analytical Processing) optimizado para el análisis.
- **Visualización:**
 - Utilizaremos la plataforma Microsoft Power BI, que tiene una versión gratuita y completa para realizar un análisis descriptivo de los datos.

CODIFICACIÓN

Aquí se describe un diagrama del plan de codificación del proyecto.



El objetivo de la codificación en las transformaciones es generar los archivos necesarios para generar los reportes de los datos, en donde si existen uno o varios archivos .xlsx puedan analizarse en un solo dataframe unificado, luego realizaremos una evaluación de los valores de cada columna para encontrar anomalías determinadas por los lineamientos técnicos para la evaluación de tales datos, luego de esto se generarán tres archivos, el original con los datos etiquetados según su calidad, los registros con anomalías y tipos de anomalías y un archivo con los duplicados del numero de documento, que es el registro que debe identificar como única a cada persona.

CONCLUSIONES

El objetivo de este proyecto es demostrar que sin importar que tan grande o tan pequeño es el set de datos que podemos tratar, todos podemos sacar beneficio de los modelos de datos que todas las empresas poseen, observándolo desde un punto de vista real, viendo resúmenes e infografías en tiempo real de lo que sucede tanto internamente como externamente