



**Desarrollo de algoritmos en Python para optimización del Plan Integral de Gestión de
Terremotos en Seguros Generales Suramericana S.A.**

Nicolas Patiño Monsalve

Informe final de practica – semestre de industria (Ingeniería civil)

Asesor

Álvaro José Mattos Olivella, Doctor (PhD)

Universidad de Antioquia

Facultad de Ingeniería

Ingeniería Civil

Medellín, Antioquia, Colombia

2024

Cita

(Patiño, 2024)

Referencia

(Patiño, 2024). *Desarrollo de algoritmos en Python para optimización del Plan Integral de gestión de Terremotos en Seguros Generales Suramericana S.A.* [Informe de practica]. Universidad de Antioquia, Medellín, Colombia.

Estilo APA 7 (2020)



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Agradecimientos

Agradezco al equipo de analítica de Seguros Generales Suramericana S.A conformado por Isabel Cristina Franco, Nicolás Gómez, Sergio Acevedo, Gloria Zapata, y Julián Lastra por impactar de manera positiva en mi futuro profesional, cada uno de ellos me ha enseñado y ha servido como guía en mi proceso de prácticas académicas.

Nuevamente, un especial agradecimiento a Sergio Acevedo Maya, el cual ha sido la persona que más ha impactado en mi transición de academia a empresa, introduciéndome el mundo de la ciencia de datos e inteligencia artificial, y siempre preocupándose para transmitir su conocimiento y experiencia.

A mi asesor interno Álvaro José Mattos Olivella por prestarme un excelente servicio, brindándome ideas, orientación y herramientas para cumplir con mis deberes académicos.

A los docentes que más han impactado en mi formación académica, Marcela Aldana, Juan Carlos Obando y Carlos Alberto Riveros.

Agradezco también a aquellas personas que me ayudan de manera simbólica, con su apoyo, como es el caso de mi familia, haciendo énfasis en mi madre, hermano y primo paterno, finalmente agradezco a mi actual pareja Valentina Muñoz, la cual además de brindarme su apoyo se ha preocupado por entender, escucharme y ayudarme en general con estos escritos académico.

Tabla de contenido

Resumen	8
Abstract	9
Introducción	10
1 Planteamiento del problema	12
2 Justificación.....	13
3 Objetivos	14
3.1 Objetivo general	14
3.2 Objetivos específicos.....	14
4 Marco teórico	15
4.1 Terremoto	14
4.2 Seguro.....	14
4.3 PIGT	14
4.4 Georreferenciación	14
4.5 Python	14
4.6 Librería Pandas.....	14
4.7 Visual Basic for Applications	14
4.8 Expresiones regulares (Regex)	14
5 Metodología	20
6 Resultados	25
7 Discusión.....	27
8 Conclusiones	30
Referencias	32

Lista de tablas

Tabla 1 <i>Ventajas y desventajas del algoritmo de estandarizar direcciones.</i>	21
Tabla 2 <i>Ventajas y desventajas del algoritmo de base de datos a partir de inspecciones.</i>	23
Tabla 3 <i>Ventajas y desventajas del algoritmo reporte trimestral automatizado.</i>	25

Lista de figuras

Figura 1 <i>Geocoficiar con Google Sheet.</i>	21
Figura 2 <i>Flujo de trabajo, estandarizador de direcciones y geo codificación.</i>	21
Figura 3 <i>Flujo de trabajo, base de datos (inspecciones).</i>	23
Figura 4 <i>Flujo de trabajo, informe automatizado.</i>	24
Figura 5 <i>Comparativa disponibilidad de datos mes</i>	26
Figura 6 <i>Ejemplificación del estandarizador de direcciones.</i>	26

Siglas, acrónimos y abreviaturas

PIGT	Plan Integral de Gestión de Terremoto
SFC	Superintendencia Financiera de Colombia
NSR	Norma Sismo Resistente
VBA	Visual Basic for Applications
MW	Magnitud de momento

Resumen

El presente trabajo aborda la implementación de algoritmos informáticos en Python para mejorar la eficiencia y la calidad del proceso de gestión de información del Plan Integral de Gestión de Terremoto (PIGT) en Seguros Generales Suramericana S.A. Se han establecido tres objetivos específicos: la implementación de un algoritmo de estandarización de direcciones y georreferenciación automatizada, la creación de una base de datos en Excel extrayendo información de archivos PDF y el desarrollo de un algoritmo de automatización de informes trimestrales; los algoritmos fueron diseñados en Python. Los resultados obtenidos han demostrado un avance significativo en la optimización de los procesos de gestión de información, mejorando la eficiencia en el proceso de georreferenciación en un 3077% y de completitud de campos PIGT en un 30800%, los cuales son relevantes para la toma de decisiones en la empresa.

Palabras clave: Algoritmos informáticos, Python, Gestión de información, Plan Integral de Gestión de Terremoto (PIGT), Seguros Generales Suramericana S.A., Georreferenciación automatizada, Extracción de información, Automatización de informes trimestrales.

Abstract

This work addresses the implementation of computer algorithms in Python to enhance the efficiency and quality of information management processes within the Earthquake Management Plan (PIGT) at Seguros Generales Suramericana S.A. Three specific objectives were established: implementing an address standardization and automated georeferencing algorithm, creating an Excel database by extracting information from PDF files, and developing a quarterly report automation algorithm. These algorithms were designed in Python. The results demonstrate significant improvements in optimizing information management processes, with a 3077% increase in georeferencing efficiency and a 30800% improvement in PIGT field completeness. These enhancements are crucial for informed decision-making within the company.

Keywords: Computer algorithms, Python, Information management, Earthquake Management Plan (PIGT), Seguros Generales Suramericana S.A., Automated georeferencing, Information extraction, Quarterly reports automation.

Introducción

Seguros Generales Suramericana S.A es una compañía que ofrece soluciones de aseguramiento y similares en América Latina, entre estos se encuentra el seguro contra terremoto. Desde el año 2018 la Superintendencia Financiera de Colombia (SFC) ordenó a las empresas aseguradoras con cobertura de terremoto emplear el Plan Integral de Gestión de Terremoto con la finalidad de tener una estimación cercana de la magnitud de las pérdidas en caso de un siniestro y así evitar escenarios de sobre estimación de reservas y tarifas elevadas, esto debido a que las pérdidas por un siniestro sísmico son importantes y la empresa es responsable de su modelación.

En 2019 Seguros Sura acoge el PIGT (Plan Integral de Gestión de Terremoto), y a su vez, uno de los requerimientos más importantes es contar con los campos PIGT completos en cada uno de las propiedades aseguradas, estos campos son, Longitud, Latitud, Numero de Pisos, Uso del inmueble, Sistema estructural, y Año de construcción, estos campos están relacionados con la norma sismo resistente colombiana NSR-10 (AIS 2014). Con estos campos se puede saber la edición de la norma que se usó, la respuesta del sistema estructural y su ubicación específica. Sin embargo, el porcentaje de completitud de los predios es baja, debido a que las personas naturales desconocen información y esta no es solicitada por los asesores comerciales, en otras palabras, es difícil que un cliente conozca el sistema estructural de su vivienda, o negocio, y su año de construcción.

En consecuencia, Seguros Generales Suramericana S.A ha contratado generaciones de practicantes de Ingeniería Civil o Construcciones Civiles que por medio de herramientas como Google Maps, Google Earth y registro fotográfico es posible determinar el sistema estructural y estimar un intervalo de año de construcción. Sin embargo, este proceso manual, si bien ha sido de utilidad resulta muy ineficiente cuando se enfrenta a grandes volúmenes de datos, con el objetivo de mejorar la productividad se optó por desarrollar tres códigos en Python los cuales están contrarrestando tres problemas fundamentales relacionados con el tiempo.

Los asesores comerciales cuentan con dos canales llamados Cliente Servidor (CS) y Guide Wire (GW) para expedir pólizas, estos dos canales presentan diferencias en el formato en que se presenta la dirección, sin mencionar la multiplicidad de maneras en las que el usuario puede ingresar dicho dato, por tanto, es necesario desarrollar un sistema de estandarización de direcciones que asegure que toda la información esté en homogeneizada antes de proceder con el proceso de georreferenciación.

Seguros Sura cuenta con una serie de inspecciones, los cuales son documentos con la información completa del inmueble. Estos documentos pueden representar un avance importante para completar información en la cartera de terremoto, sin embargo, revisar estas inspecciones en formato PDF resulta tedioso, por lo que se propone que un software que cree una base de datos a partir de dichas inspecciones.

Por último, una vez finalizado el PIGT, el proceso subsecuente consiste en enviar informes trimestrales con la modelación y estimación de reservas a la Superintendencia Financiera de Colombia (SFC), este informe cuenta con una plantilla en el que se deben calcular y actualizar los estados de riesgo de cada póliza, dado a que el volumen de registros es grande (cerca de 260 mil registros) puede ser susceptible a omisiones y errores humanos, en función a esta problemática este último código está destinado a reducir estos errores.

1 Planteamiento del problema

Cuando se expide una póliza con cobertura de terremoto, la información se envía a la base de datos de Sura en diferentes formatos según el canal comercial utilizado por los asesores ("CS" o "GW"), estos canales funcionan como bases de datos transaccionales las cuales registran la información en la Data Warehouse. Al extraer los datos de la base de datos relacional con ayuda de SQL, el campo de dirección del inmueble asegurado se ve altamente afectado debido a que cada canal tiene un formato distinto y a las múltiples maneras en que los usuarios pueden diligenciar la información, lo que impide la homogeneidad de los datos. Esta situación presenta retos como la identificación de negocios asegurados en el mismo punto geográfico (cúmulo), la completitud y disponibilidad de la información y la automatización del proceso de georreferenciación de direcciones.

Seguros Generales Suramericana S.A cuenta con levantamientos de información documentada de los inmuebles, las cuales contienen información relevante para completar campos como año de construcción, sistema estructural, e inclusive si ha tenido repotenciones a lo largo del tiempo. Esta información es almacenada en documentos PDF en el software de uso exclusivo de Seguros Sura, sin embargo, debido a que son documentos extensos y lo único que necesario de la inspección es conocer campos PIGT, resulta tedioso la extracción de datos, por esto, actualmente se opta por adquirir la información a través de registro fotográfico e intuir la información por medio de los conocimientos de la carrera o herramientas externas.

Uno de los pasos del PIGT es el reporte trimestral que se debe entregar a la Superintendencia Financiera de Colombia luego de la modelación. Para esto Seguros Sura cuenta con una plantilla en la cual se reemplazan una gran cantidad de valores, los cuales deben ser calculados trimestralmente, estos incluyen cantidades monetarias, porcentajes y fechas. El proceso de calcular valores a mano para luego ser actualizados en la plantilla está sujeto a tener errores humanos y omisiones.

2 Justificación

La solución planteada consiste en tres algoritmos los cuales atienden a necesidades diferentes.

- 1. Estandarizador de direcciones:** Para abordar el primer problema descrito anteriormente, se decidió desarrollar un algoritmo que transforme todas las direcciones a un formato unificado, basado en el formato utilizado por Google Maps para la lectura de direcciones. Al estandarizar las direcciones de esta manera, se facilita el proceso de georreferenciación automática y la detección de cúmulos de manera más eficiente.
- 2. Base de datos (Inspecciones):** El segundo código tiene como objetivo crear una base de datos a partir de la información recopilada en las inspecciones de Seguros Sura. Utilizando palabras clave identificadas, esta información se almacenará en columnas de Excel. Disponer de una base de datos de inspecciones representará un avance significativo en la completitud de la cartera general de seguros contra terremotos.
- 3. Informe automático:** El tercer código está diseñado para realizar diversos cálculos y reemplazos en la plantilla que se debe entregar trimestralmente a la Superintendencia Financiera de Colombia. Utilizando Visual Basic for Applications (VBA) y Python, este código automatiza el proceso, evitando errores y emisiones que pueden ocurrir durante el procesamiento manual.

3 Objetivos

3.1 Objetivo general

Desarrollar y aplicar algoritmos informáticos en Python para mejorar la eficiencia y la calidad de los datos en el proceso de gestión de información del PIGT en Seguros Generales Suramericana S.A.

3.2 Objetivos específico

- Implementar un algoritmo de georreferenciación automatizada que optimice el proceso de asignación de coordenadas a los diferentes riesgos asegurados por la empresa, reduciendo así el tiempo y los errores asociados con la geolocalización manual.
- Diseñar y ejecutar un algoritmo para la extracción y agregación de información relevante de inspecciones de propiedades en una hoja de cálculo Excel, con el fin de enriquecer la base de datos del plan integral de gestión de terremotos (PIGT) y mejorar la calidad de los datos disponibles.
- Desarrollar un algoritmo de automatización de informes trimestrales que actualice de manera automática la plantilla de informes, garantizando la integridad de los datos y reduciendo la probabilidad de errores en el proceso de generación y presentación de informes.

4 Marco teórico

4.1 Terremoto

Un terremoto es un fenómeno natural caracterizado por el movimiento repentino y violento de la superficie terrestre, generalmente causado por la liberación de energía acumulada en las placas tectónicas (SGC, 2020). Esta liberación de energía puede provocar vibraciones o sacudidas en la Tierra, lo que a menudo resulta en daños a edificaciones, infraestructuras y, en casos extremos, pérdidas humanas.

Principales terremotos en Colombia

- **Terremoto de Cúcuta de 1875:** Este fue uno de los terremotos más destructivos en la historia de Colombia, con una magnitud estimada de 6.8 MW (magnitud de momento). Causó graves daños en Cúcuta y otras áreas cercanas como Villa del Rosario. Las pérdidas se estimaron en 10 millones de pesos de la época y se fallecieron alrededor de 3000 personas. (SGC, 2020).
- **Terremoto de Popayán de 1983:** Con una magnitud de 5.6 MW, este terremoto causó daños significativos en la ciudad de Popayán y áreas circundantes, dejando a 250 fallecidos y alrededor de 1500 personas heridas. Aproximadamente 13800 viviendas presentaron daños muy graves y 4964 fueron destruidas (SGC, 2023).
- **Terremoto de Armenia de 1999:** Este devastador terremoto, con una magnitud de 6.2 MW, golpeó la región de Armenia y el Eje Cafetero, causando más de mil muertos y 200 mil personas afectadas, se estima que las partidas materiales costaron alrededor de 1,6 billones de pesos (El Espectador, 2014).

4.2 Seguros

Un seguro es un contrato mediante el cual una entidad aseguradora se compromete, a cambio de una prima, a indemnizar al asegurado o beneficiario en caso de que ocurra un evento predeterminado, como pérdida, daño o riesgo, cubierto por el contrato (SSS, 2021).

La cobertura de terremoto es un tipo de póliza de seguro que ofrece protección específica contra los daños causados por terremotos. Esta cobertura puede incluir la reparación o reconstrucción de edificaciones dañadas, así como la compensación por pérdidas materiales asociadas con un terremoto. Dado el riesgo sísmico en áreas propensas a terremotos, esta cobertura proporciona tranquilidad financiera a los asegurados en caso de que ocurra un evento sísmico.

4.3 Plan Integral de Gestión de Terremoto.

El Plan Integral de Gestión de Terremoto (PIGT) establece una serie de buenas prácticas para gestionar el riesgo catastrófico de la Compañía, estimando adecuadamente las reservas y la compra de reaseguro para responder en un evento sísmico, este se crea para garantizar las reservas técnicas necesarias de acuerdo con el nivel y la naturaleza de los riesgos asegurados, así como con adecuados niveles patrimoniales que salvaguarden la solvencia de la Compañía y garanticen los intereses de tomadores y asegurados. El PIGT se originó por orden Superintendencia Financiera de Colombia (SFC) la cual, expidió la Circular externa 006 de 2018, estableciendo la necesidad de que las compañías aseguradoras diseñen y ejecuten un Plan Integral de Gestión del Ramo de Terremoto. (Seguros Generales Suramericana S.A., 2018)

4.4 Georreferenciación.

La georreferenciación es un proceso fundamental en los Sistemas de Información Geográfica (SIG), que consiste en asignar coordenadas geográficas (como latitud y longitud) a objetos o entidades en la Tierra (Goodchild, 1997). Este proceso permite la ubicación precisa de

elementos geográficos en un mapa digital y es esencial para diversas aplicaciones, desde la navegación hasta la planificación urbana.

El sistema de coordenadas geográficas utilizado para la creación del algoritmo fue WGS84, debido a que es el sistema empleado por Google Maps. Este sistema clasifica la Tierra con tres ejes: latitud, longitud y altura desde el elipsoide hasta un punto superficial de la Tierra. En la práctica, solo se utilizó la latitud y la longitud.

- **Latitud:** Es el ángulo medido desde el ecuador hacia el norte o sur. Por ejemplo, Latitud: 40.748817° (positivo hacia el norte).
- **Longitud:** Es el ángulo medido desde el meridiano de Greenwich hacia el este u oeste. Por ejemplo, Longitud: -73.985428° (negativo hacia el oeste)

4.5 Python

Python es un lenguaje de programación, conocido por su sintaxis clara y legible (Python Software Foundation, 2022). Es ampliamente utilizado en una variedad de aplicaciones, incluyendo desarrollo web, análisis de datos, inteligencia artificial, aprendizaje automático, automatización de tareas, entre otros. En el campo de la ciencia de datos Python se destaca por varias razones, por ejemplo:

- **Gratuito y de código abierto:** Python es un lenguaje gratuito y de código abierto, lo que significa que cualquier persona puede usarlo y contribuir a su desarrollo. Además, esta razón lo hace atractivo para microempresas que no tengan los recursos para licencias de otros lenguajes.
- **Diversidad de módulos y bibliotecas:** Python cuenta con una vasta colección de bibliotecas y módulos especializados en ciencia de datos, como Pandas para manipulación y análisis de datos, NumPy para cálculos numéricos, entre otros.
- **Integración con otras tecnologías:** Python se integra fácilmente con otras tecnologías y lenguajes de programación. Esto es especialmente útil en ciencia de datos, donde es necesario interactuar con bases de datos, herramientas de visualización etc.

Sin embargo, algunas desventajas de Python podrían ser:

- **Velocidad de procesamiento:** Python es conocido por ser más lento en su procesamiento en comparación con otros lenguajes de programación como C++ o Java. Esto se debe en parte a su naturaleza interpretada, lo que significa que el código se ejecuta línea por línea en lugar de compilarse a código máquina antes de la ejecución. (Grinberg, 2017)
- **Consumo energético:** De acuerdo con un estudio realizado por Pereira et al. (2017), Python tiende a consumir más energía en comparación con lenguajes compilados debido a la cantidad de recursos que requiere durante su ejecución.

4.6 Librería Pandas

Para ejecución de los códigos es relevante el uso de la librería Pandas, la cual es una biblioteca de código abierto en Python diseñada para la manipulación y análisis de datos (McKinney, 2010). Su estructura principal, el DataFrame, proporciona una forma flexible y eficiente de trabajar con datos tabulares. Desarrollado por McKinney (2010) y mantenido por colaboradores como Reback y Van den Bossche (2020). Autores como VanderPlas (2016) han destacado su el uso de la librería en el libro "Python Data Science Handbook", donde se exploran técnicas de análisis de datos utilizando Pandas y otras herramientas de Python.

4.7 Visual Basic for Applications (VBA)

Visual Basic for Applications (VBA) es un lenguaje de programación desarrollado por Microsoft que se utiliza para automatizar tareas y crear aplicaciones en entornos de Microsoft Office, como Excel, Word, Access y PowerPoint (Microsoft, 2023).

VBA puede integrarse eficazmente con Python, permitiendo automatizar tareas en Excel y Word mientras se aprovechan las capacidades avanzadas de análisis de datos y manipulación de Python. Por ejemplo, un script en Python puede realizar cálculos complejos y luego exportar los resultados a Excel, donde VBA puede tomar esos datos y generar gráficos. Estos gráficos, a su vez, pueden ser insertados automáticamente en documentos de Word mediante VBA.

Comandos importantes en VBA para Word incluyen *Documents.Open* para abrir documentos, *Selection.TypeText* para insertar texto, y *Table.Add* para crear tablas, facilitando así la automatización y mejora de flujos de trabajo en la creación y manejo de documentos.

4.8 Expresiones regulares (regex)

Las expresiones regulares, también conocidas como regex, son secuencias de caracteres que forman un patrón, los cuales son utilizados principalmente para la búsqueda y manipulación de texto. Regex permite especificar reglas precisas para la identificación de cadenas de caracteres en direcciones, correos electrónicos, números telefónicos, entre otros; que cumplan con ciertos criterios sintácticos (Friedl, 2006).

El proceso de estandarización de direcciones se benefició enormemente de la herramienta regex, capaz de manejar una amplia variedad de patrones encontrados en las direcciones para realizar los reemplazos necesarios. Desde expresiones regulares simples como $r'(calle/cll/kl/cl\./alle/cle): 'cl'$, que identifican y reemplazan palabras como "calle, cll, kl, cl., alle, cle" por "cl", hasta expresiones más complejas como $r'(\w+\.\s\d+\s?[a-zA-Z]*)\s([a-zA-Z\s]*)\s*(\d+\s*-\s*\d+)' : r^I\2 \# \3'$, que capturan direcciones como "Tv. 53 A 20 – 15" y las reformulan añadiendo un "#" en la posición adecuada (Tv. 53 A # 20 – 15).

5 Metodología

1. Estandarizador de direcciones

Este algoritmo presenta el siguiente flujo de trabajo:

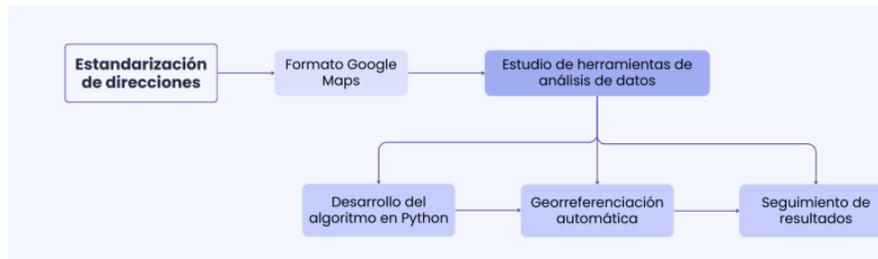
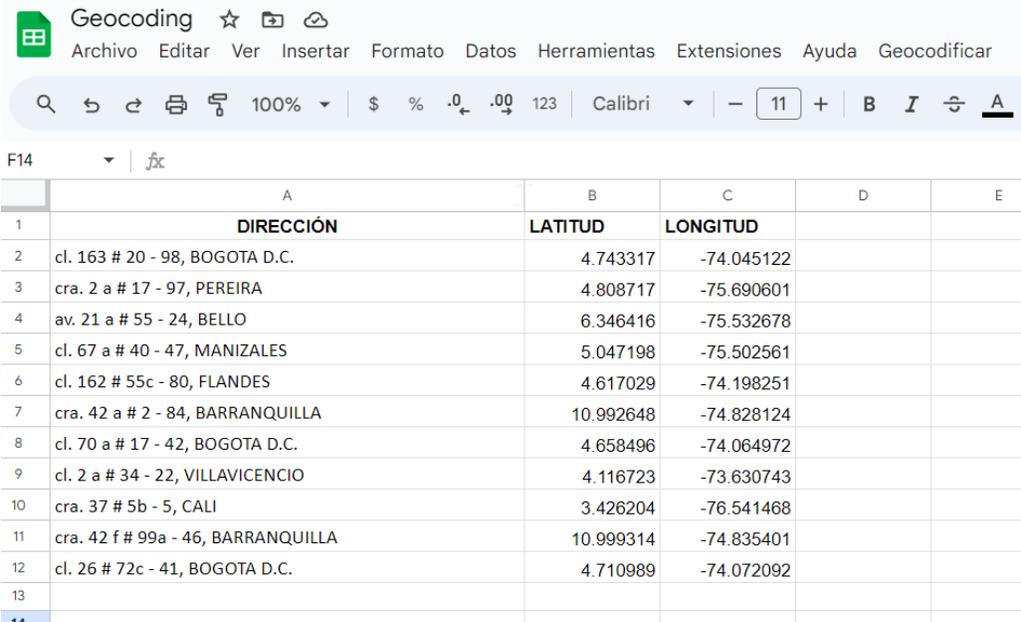


Figura 1. Flujo de trabajo, estandarizador de direcciones y geo codificación **Fuente:** Propia

- **Formato Google Maps:** Durante muchos años, Google Maps ha sido la aplicación preferida en el equipo de analítica debido a su fiabilidad en la entrega de resultados. Por esta razón, se exploró el formato de dirección más adecuado para asegurar que la aplicación interprete correctamente las ubicaciones en territorio colombiano.
- **Estudio de herramientas de análisis de datos:** Se dedicó un tiempo considerable al estudio del módulo Pandas, el cual permite convertir archivos de Excel (xlsx o xlsxm) en DataFrames en Python. Esta herramienta facilitó la transformación, selección y visualización de datos mediante funciones predefinidas. Además, se estudió arduamente el uso de expresiones regulares (Regex) las cuales se convirtieron en una herramienta clave para el desarrollo del algoritmo.
- **Desarrollo del algoritmo en Python:** Se creó una función llamada *estandarizar_direcciones* y como se mencionó en el marco teórico, se implementaron paso a paso diversas expresiones regulares con diferentes niveles de complejidad en un diccionario para abordar todos los patrones identificados en los datos de entrada hasta la fecha, asegurando que se ajustaran al formato definido en el primer paso. Este proceso se apoyó en Pandas, el cual vinculó el archivo correspondiente a la cartera de terremoto y seleccionó la columna “Dirección del

inmueble” convirtiendo estos registros en los datos de entrada de la función, los cuales fueron transformados con el diccionario definido anteriormente.

- **Georreferenciación Automática:** Debido a que Seguros Sura no dispone de una API key de Google Maps, se utilizó la versión gratuita que no permite la georreferenciación directamente desde Python y solamente permite georreferenciar 2000 direcciones cada 24 horas. Sin embargo, fue posible realizar este proceso utilizando una hoja de Google Sheet, activando la opción de "Geocodificar" desde las extensiones de la hoja de cálculo de Google; esto permite ingresar las direcciones estandarizadas y obtener las longitudes y latitudes correspondiente. En la Figura 2 se puede visualizar lo descrito anteriormente.



	A	B	C	D	E
1	DIRECCIÓN	LATITUD	LONGITUD		
2	cl. 163 # 20 - 98, BOGOTA D.C.	4.743317	-74.045122		
3	cra. 2 a # 17 - 97, PEREIRA	4.808717	-75.690601		
4	av. 21 a # 55 - 24, BELLO	6.346416	-75.532678		
5	cl. 67 a # 40 - 47, MANIZALES	5.047198	-75.502561		
6	cl. 162 # 55c - 80, FLANDES	4.617029	-74.198251		
7	cra. 42 a # 2 - 84, BARRANQUILLA	10.992648	-74.828124		
8	cl. 70 a # 17 - 42, BOGOTA D.C.	4.658496	-74.064972		
9	cl. 2 a # 34 - 22, VILLAVICENCIO	4.116723	-73.630743		
10	cra. 37 # 5b - 5, CALI	3.426204	-76.541468		
11	cra. 42 f # 99a - 46, BARRANQUILLA	10.999314	-74.835401		
12	cl. 26 # 72c - 41, BOGOTA D.C.	4.710989	-74.072092		
13					

Figura 2. Geocodificar con Google Sheet. Fuente: Propia

- **Seguimiento de resultados:** Una vez completado el algoritmo, se procedió a depurar errores y realizar el mantenimiento necesario. Inmediatamente después de que el algoritmo empezó a funcionar correctamente, se implementó en el equipo para abordar diversas problemáticas, como la completitud de datos, la detección de cúmulos y la disponibilidad de información.

En la tabla 1 se ven reflejadas las ventajas y desventajas de la implementación de este algoritmo.

Tabla 1 Ventajas y desventajas del algoritmo de estandarización de direcciones.

Ventajas	Desventajas
<p>La creación del algoritmo de estandarización de direcciones ha permitido homogeneizar la información proveniente de diferentes canales de venta.</p> <p>El algoritmo permite de manera sencilla el proceso de georreferenciación con Google Maps.</p>	<p>El proceso de estandarización puede requerir ajustes periódicos para adaptarse a cambios en la forma en que se ingresan las direcciones, lo que implica un mantenimiento continuo del algoritmo.</p> <p>Los resultados de longitud y latitud pueden variar con el real en cuestión de cuadras, sin embargo, se consideran errores admisibles.</p>

2. Base de datos (Inspecciones)

Para este algoritmo se tuvo el siguiente flujo de trabajo durante el proyecto.



Figura 3. Flujo de trabajo, base de datos (inspecciones) **Fuente:** Propia

- Búsqueda de módulos:** Para abordar esta problemática, inicialmente se buscó un módulo de Python que permitiera leer PDF y lo transforme en texto, para esto, se encontró que la mejor opción fue la librería llamada “PdfPlumber”.
- Detección de patrones en inspecciones:** En este paso se estudió el formato de las inspecciones donde se identificó que cuentan con un formato único, facilitando el proceso, además, la información que se requería se disminuía a 4 páginas que no varían en orden, es decir, en todas las inspecciones la información se encontraba en el mismo número de página.
- Desarrollo de algoritmo en Python:** El algoritmo usa el módulo PDF plumber para leer y convertir el PDF en cadenas de texto, por temas de optimización de código, solo se realiza con las 4 hojas necesarias, y luego se procedió a guardar en variables

la información relevante, las cuales son “Numero de póliza”, “identificación o NIT”, “Nombre del inmueble”, “Dirección”, “Municipio”, “Departamento”, “Longitud”, “Latitud”, “Numero de pisos”, “Uso del inmueble”, “Sistema estructural”, “Año de construcción” y “Forma de la estructura” y con ayuda de Pandas la información extraída de cada inspección se diligenciaba en la fila correspondiente a cada inmueble en Excel.

- **Desarrollo masivo:** Una vez el prototipo descrito en el paso anterior funcionó correctamente se usa el módulo OS de Python que permitió leer una carpeta donde se alojaron todas las inspecciones y cíclicamente los campos fueron registrados en Excel, logrando construir una base de datos a partir de estos PDFs.

Este algoritmo tiene las siguientes ventajas y desventajas.

Tabla 2 Ventajas y desventajas del algoritmo de base de datos a partir de inspecciones.

Ventajas	Desventajas
La generación automática de una base de datos a partir de las inspecciones ha simplificado el proceso de recopilación y organización de información. El uso de este software ha permitido reducir el tiempo dedicado a revisar manualmente las inspecciones, liberando recursos para otras actividades. Algoritmo que permite ser reciclado para otros fines de la compañía.	Si cambia el formato de las inspecciones el código no funcionará bajo la lógica que se implementó.

3. Informe automático

Para este algoritmo se implemento el siguiente flujo de trabajo.



Figura 4. Flujo de trabajo, informe automático **Fuente:** Propia

- **Programación en VBA:** Los datos con los que se realizan cálculos que se efectúan trimestralmente se encuentran ubicados en 4 archivos de Excel diferentes, además, hay que realizar filtros y otros procesos para obtener los resultados necesarios, para optimizar este proceso se optó por realizar una macro que lea estos archivos de Excel y pegue las hojas necesarias en el mismo libro de la macro para posteriormente hacer los cálculos necesarios y tener cada valor de manera organizada.
- **Creación de plantilla genérica:** En el documento de Word, que funciona como plantilla, se actualizaron los valores que son diligenciados trimestralmente por el nombre de la variable encerrados en dos llaves para facilitar su detección, por ejemplo: {{Total_Riesgos}} garantizando que no faltaron campos para efectuar esta acción.
- **Desarrollo de algoritmo en Python:** En Python se crea un diccionario en el cual la clave son las variables que se agregaron en el documento de Word previamente y el valor, la celda correspondiente de los cálculos que efectuó la macro, el documento de Word es leído con la librería “Docx” y el Excel con “Pandas”, finalmente se enlazo el código de Python con el de VBA y se creó un ejecutable que entrega el documento con toda la información para ser enviado, y actualizando la plantilla para un nuevo uso.

Las ventajas y desventajas que presentó este algoritmo se reflejan en la tabla 3.

Tabla 3 *Ventajas y desventajas del algoritmo reporte trimestral automatizado.*

Ventajas	Desventajas
La automatización del proceso de relleno de la plantilla de Word ha reducido significativamente el tiempo y los errores asociados con esta tarea manual. Se ha mejorado la precisión en la generación de informes trimestrales, evitando omisiones y fallos en la inserción de datos.	La configuración inicial del programa requirió tiempo y recursos adicionales, especialmente en la etapa de desarrollo y pruebas, ya que es un documento legal y se debía asegurar su correcto funcionamiento.

6 Resultados

- **Estandarización de Direcciones:** La georreferenciación automática comenzó cuando más de 2000 direcciones carecían de coordenadas de longitud y latitud. Inicialmente, los practicantes recopilaban esta información manualmente a un ritmo de aproximadamente 65 direcciones por día. Sin embargo, tras automatizar este proceso, se lograron georreferenciar 2000 predios en un día, mejorando la eficiencia en un 3077 %.

$$Eficiencia/dia = \frac{rendimiento\ código}{rendimiento\ manual} \cdot 100\% = 3077\%$$

El uso del algoritmo no se limita solo a este caso debido a que cada mes se actualiza la cartera de terremoto, lo que implica que nuevos registros lleguen con datos de georreferenciación incompletos. Gracias al código implementado, es posible completar estos registros de manera mucho más eficiente.

A pesar de que la estandarización de direcciones fue creada para dar solución al problema de georreferenciación manual ha contribuido en diversas problemáticas como:

- **Detección de cúmulos:** por temas de contratación, un mismo punto geográfico no puede superar cierto valor asegurado, y por el momento Seguros Sura no tiene control de esto, como consecuencia podría verse afectado el cálculo de deducibles. Actualmente se pueden encontrar estos cúmulos en una semana.
- **Compleitud de campos:** Es usual encontrar varias pólizas correspondientes a un mismo inmueble, por ejemplo: grandes empresas, centros comerciales, entre otros. Anteriormente no era posible tener control de los inmuebles levantados por los practicantes, ya que la información solamente es cruzada por la identificación de la póliza, ahora con la dirección en un mismo formato es posible levantar datos con información que han recolectado los practicantes o que se encuentra en bases externas de los edificios, llegando a completar toda la

información PIGT de 15400 riesgos en un día, comparando con el rendimiento manual que se estima de 50 día.

$$EficienciaCom/dia = \frac{rendimiento\ código}{rendimiento\ manual} \cdot 100\% = 30800\%$$

Sin embargo, para alcanzar esta meta fue necesario evolucionar el código mediante detección y depuración de errores, a continuación, veremos una comparativa entre el método manual y el automático para la completitud de campos en un mes.

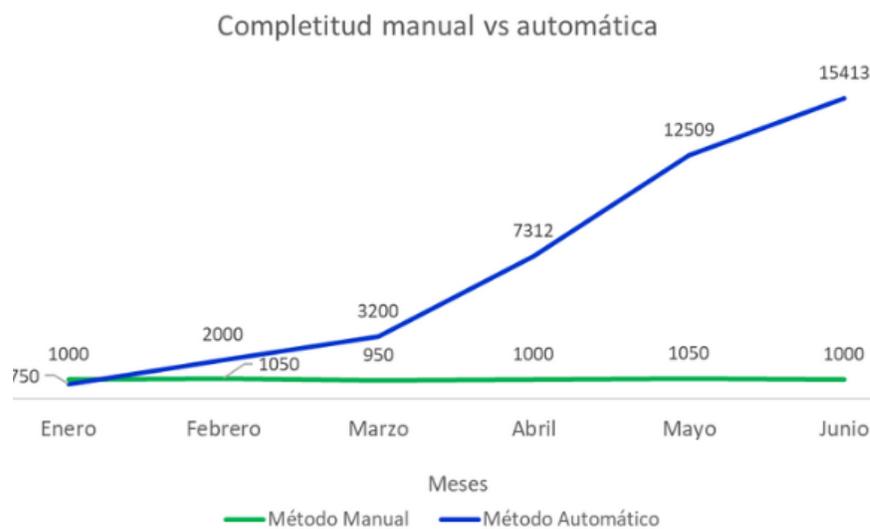


Figura 5. Comparativa disponibilidad de datos mes **Fuente:** Sura

El presentar la información con el mismo formato ha resultado ser de gran utilidad para el equipo, a continuación, se dará un ejemplo del estandarizador de direcciones, este caso particular cuenta con 140 registros que equivalen a la misma dirección escrita de formas diferentes, los cuales no se hubiesen podido identificar a no ser por este código. (Figura 5)



Estandarizador de direcciones

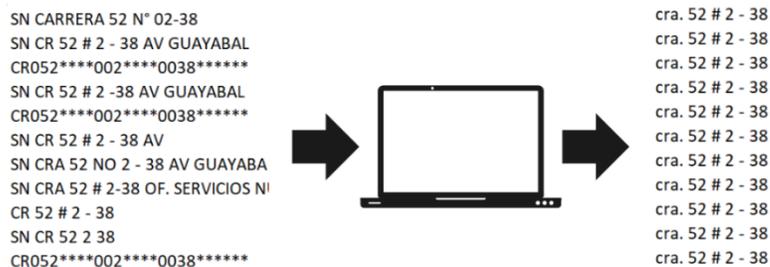


Figura 6. Ejemplificación del estandarizador de direcciones Fuente: Sura

- **Creación de Base de Datos a partir de Inspecciones:** A pesar de que solo 2% del número de pólizas totales de la cartera cuentan con registros (5200), estos documentos brindan información confiable y ayudaron a aumentar el porcentaje general de completitud y el porcentaje de completitud según el valor asegurado.
- **Informe automático:** Trimestralmente se ocupaban aproximadamente de 5 a 6 horas haciendo los cálculos y los reemplazos necesarios debido a la extensión de páginas del documento, luego con el código este proceso se redujo a 20 -25 minutos, es decir, la eficiencia aumento un 15%.

7 Discusión

Estandarización de direcciones:

- El algoritmo de estandarización de direcciones desarrollado presenta varias ventajas significativas, como la homogeneización de la información proveniente de diferentes canales de venta y la posibilidad de georreferenciación con Google Maps. Sin embargo, también es importante considerar las limitaciones y áreas de mejora.

- Uno de los principales desafíos con la estandarización de direcciones es la necesidad de mantener y actualizar constantemente el algoritmo para adaptarse a cambios en la forma en que se ingresan las direcciones. Las direcciones pueden variar considerablemente en su formato y estructura, lo que implica que el algoritmo debe ser lo suficientemente flexible y robusto para manejar estos cambios sin perder precisión. Esto puede requerir un esfuerzo continuo en términos de mantenimiento y ajustes periódicos.
- Además, aunque el algoritmo permite la georreferenciación automática, los resultados de longitud y latitud pueden no ser completamente precisos, pudiendo variar en cuestión de cuerdas. Estos errores, aunque admisibles, pueden afectar la exactitud de los análisis posteriores, especialmente en aplicaciones que requieren alta precisión geográfica.
- A pesar de las limitaciones, la implementación del algoritmo ha tenido un impacto positivo considerable. La automatización de la georreferenciación ha mejorado la eficiencia en un 3077%, permitiendo la georreferenciación de 2000 direcciones en un día en comparación con las 65 que se realizaban manualmente. Este aumento en eficiencia no solo ha reducido el tiempo necesario para completar las tareas, sino que también ha permitido abordar problemáticas adicionales, como la detección de cúmulos y la completitud de campos, mejorando la calidad y utilidad de los datos.
- Para un futuro no muy lejano se puede buscar adaptar el código de georreferenciación de direcciones a la API de GeoSura el cual es la herramienta SIG de uso exclusivo de la empresa, evitando usar la hoja de Google Sheet y aumentando la cantidad de registros por día.

Creación de Base de Datos a partir de Inspecciones:

-
- La automatización de este proceso ha simplificado la recopilación y organización de la información, reduciendo el tiempo dedicado a revisar manualmente las inspecciones y liberando recursos para otras actividades.
 - Una de las principales desventajas de este enfoque es que el algoritmo está diseñado específicamente para un formato de inspección particular. Si el formato de las inspecciones cambia, el algoritmo no funcionará correctamente bajo la lógica implementada, lo que nuevamente subraya la necesidad de mantenimiento y adaptabilidad del código.
 - Además, aunque solo el 2% del número total de pólizas cuenta con registros de inspección, estos documentos brindan información confiable obtenida de ingenieros que han visitado el lugar, ayudando a aumentar el porcentaje general de completitud de los datos.

Informe Automático:

- La automatización del proceso de generación de informes trimestrales mediante el uso de VBA y Python ha reducido significativamente el tiempo y los errores asociados con esta tarea. La precisión en la generación de informes ha mejorado, evitando omisiones y fallos en la inserción de datos.
- La configuración inicial del programa requirió tiempo y recursos adicionales, especialmente en la etapa de desarrollo y pruebas. Dado que se trata de un documento legal, fue esencial asegurar su correcto funcionamiento, lo que implicó un esfuerzo considerable en la etapa de validación.
- En Python se deben especificar todos los formatos que se van a usar ya que no existe la herramienta “copiar formato” como en las herramientas de office, lo cual representó un reto y líneas de código que pueden ser simplificadas.

-
- El código puede ser mejorado en cuestiones de tiempo, ya que una de las desventajas de vincular VBA y Python es el tiempo de procesado, si el código se hubiese implementado en su totalidad en VBA hubiese sido más eficiente.

8 Conclusiones

- Los resultados derivados de los problemas planteados en el informe reflejan un avance significativo en la mejora de la eficiencia y la calidad de los datos en el proceso de gestión de información del Plan Integral de Gestión de Terremoto (PIGT) en Seguros Generales Suramericana S.A.
- Con los algoritmos implementados en el equipo de analítica, cada mes se pudo contar con estados de completitud mucho más altos que en los días en que el proceso se hacía completamente manual.
- El desarrollo e implementación del algoritmo de georreferenciación automatizada ha demostrado ser una herramienta efectiva para optimizar el proceso de asignación de coordenadas a los riesgos asegurados, reduciendo considerablemente el tiempo y los errores asociados con la geolocalización manual. Esto ha mejorado la precisión de la ubicación de los riesgos asegurados y ha facilitado el envío de información a diferentes áreas de la empresa que lo requieran.
- La creación del algoritmo de inspecciones de propiedades ha enriquecido la base de datos del PIGT. Al automatizar este proceso, se ha mejorado la calidad y la disponibilidad de datos relevantes sobre las propiedades aseguradas, lo que ha facilitado el análisis de riesgos y la gestión de información para la empresa.
- El desarrollo del algoritmo de automatización de informes trimestrales ha tenido un impacto para garantizar la integridad de los datos y reducir la probabilidad de errores en el proceso de generación y presentación de informes. La actualización automática de la plantilla de informes ha agilizado este proceso, permitiendo a la empresa cumplir con los requisitos regulatorios de manera más eficiente y precisa.

-
- En conjunto, estos resultados evidencian el impacto positivo de la implementación de algoritmos informáticos en Python en la optimización de los procesos de gestión de información del PIGT en Seguros Generales Suramericana S.A., Sin embargo, se reconoce la importancia de continuar evaluando y mejorando estos algoritmos para mantener su eficacia y relevancia en un entorno empresarial en constante evolución.
 - La flexibilidad de los algoritmos permite su adaptación a otros problemas y áreas dentro de la empresa, lo que potencialmente puede generar más mejoras en la eficiencia de otros procesos.
 - La estandarización de direcciones ha generado un gran impacto en la empresa, se está buscando integrar este algoritmo a la base de datos de la compañía, lo cual permitiría abordar problemas que pueden tener otros equipos de trabajo.

Referencias

- DeMers, M. N. (2008). *Fundamentals of geographic information systems*. Wiley.
- Goodchild, M. F. (1997). *Geographical information systems and science*. *Projections: Quarterly newsletter of the National Center for Geographic Information and Analysis*, 17(1), 1-6.
- Longley, P. A., & Batty, M. (Eds.). (2003). *Advanced spatial analysis: The CASA book of GIS*. ESRI Press.
- Longley, P. A., et al. (2005). *Geographic information systems and science*. John Wiley & Sons.
- Asociación Colombiana de Ingeniería Sísmica (AIS). (2014). *Reglamento Colombiano de Construcción Sismo Resistente NSR-10*. Bogotá, Colombia: AIS.
- McKinney, W. (2010). *Data structures for statistical computing in Python*. *Proceedings of the 9th Python in Science Conference*, 51-56.
- Microsoft. (2023). *Visual Basic for Applications (VBA) Documentation*. Microsoft.
- Mitchell, A. (1999). *The ESRI guide to GIS analysis, Volume 1: Geographic patterns & relationships*. ESRI Press.
- Python Software Foundation. (2022). *Python programming language - official documentation*. Python Software Foundation.
- Reback, J., & Van den Bossche, J. (2020). *pandas-dev/pandas: Pandas 1.0.5*. Zenodo. <https://doi.org/10.5281/zenodo.3509134>

-
- Seguros Generales Suramericana S.A. (2018). Plan integral de gestión de terremoto (PIGT). Seguros Generales Suramericana S.A.
 - Servicio Geológico Colombiano (SGC). (2020). Información sobre sismos y terremotos en Colombia. Servicio Geológico Colombiano.
 - El Espectador. (2014, Enero 25). Se conmemoran 15 años del terremoto de Armenia. El
 - Friedl, J. E. F. (2006). *Mastering Regular Expressions* (3rd ed.). O'Reilly Media
 - Grinberg, M. (2017). *Flask Web Development: Developing Web Applications with Python*. O'Reilly Media.
 - Servicio Geológico Colombiano. (SGC). (2023) Aniversario No. 40 del terremoto de Popayán: Un momento para pensar en la importancia de construir bien en Colombia. Servicio Geológico Colombiano.
 - Superintendencia de Sociedades de Colombia (SSS). (2021). Guía sobre seguros y coberturas. Superintendencia de Sociedades de Colombia.
 - Tomlinson, R. F. (2003). *Thinking about GIS: Geographic information system planning for managers*. ESRI Press.
 - VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. O'Reilly Media.
 - Pereira, R., Paulino, H., Couto, M., Ribeiro, R., Rua, R., Cunha, J., & Saraiva, J. (2017). *Energy Efficiency Across Programming Languages: How Do Energy, Time, and Memory Relate?* Proceedings of the 10th ACM SIGPLAN International Conference on Software Language Engineering.