**Logotipo

Descripción generada automáticamente**

**Sura Digital CLV: Valor del Cliente con Inteligencia de Datos**

Erik Estiven Arias Paniagua

Anderson Guisao Padrón

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor  
Nombres completos, Título académico más alto

Universidad de Antioquia  
Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2025

|  |  |
| --- | --- |
| **Cita** | (Muñoz Zapata & Martínez Naranjo, 2023) |
| **Referencia**  **Estilo APA 7 (2020)** | Muñoz Zapata, L., & Martínez Naranjo, J. A. (2023). *Título del trabajo* Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia. |

**** 

Especialización en Analítica y Ciencia de Datos, CohorteIX.

Centro de Investigación Ambientales y de Ingeniería (CIA).

|  |  |
| --- | --- |
|  | Diagrama  Descripción generada automáticamente con confianza media |

Centro de Documentación Ingeniería (CENDOI)

**Repositorio Institucional:** http://bibliotecadigital.udea.edu.co

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Diego José Luis Botia Valderrama

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.

**Dedicatoria**

Texto de dedicatoria centrado.

**Agradecimientos**

Texto de agradecimientos centrado.

**Tabla de contenido**

**No se encontraron entradas de tabla de contenido.**

**Lista de tablas**

[**Tabla 1** Datos originales del dataset.…………………………………………………………...19](#_heading=h.2u6wntf)

[**Tabla 2**](#_heading=h.19c6y18) Datos finales del Dataset………………………………………………………………24

[**Tabla 3**](#_heading=h.3tbugp1) Cálculo del CLV Clientes………………………………………………………………28

**Lista de figuras**

[**Figura 1**](#_heading=h.2r0uhxc) Variables Categóricas………………………………………………………………...25

[**Figura 2** Distribucion de variables numericas………………………………………………….27](#_heading=h.1664s55)

**Figura 3** Formula CLV…………………………………………………………………………28

**Figura 4** Análisis de correlación numérica…………………………………………………….29

**Siglas, acrónimos y abreviaturas**

**APA** American Psychological Association

**Cms.** Centímetros

**ERIC** Education Resources Information Center

**Esp.** Especialista

**MP** Magistrado Ponente

**MSc** Magister Scientiae

**Párr.** Párrafo

**PhD** Philosophiae Doctor

**PBQ-SF** Personality Belief Questionnaire Short Form

**PostDoc** PostDoctor

**UdeA** Universidad de Antioquia

# Resumen

El resumen permite identificar la esencia del escrito, es un abstract. Realiza una descripción general de tu proyecto: qué se persigue, qué datos se tiene, qué estrategia se siguió para las iteraciones, que obstáculos hubo, qué resultados se obtuvieron, etc. La longitud es mínimo 150 y máximo 250 palabras.

Incluye al final de dicha página la dirección de los repositorios GitHub.

*Palabras clave*: palabra 1, palabra 2, palabra 3, palabra 4.

# Abstract

El abstract es el mismo resumen pero en idioma inglés. Conserva la misma extensión o aproximada, es decir, mínimo 150 y máximo 250 palabras.

*Keywords***:** …..

# Descripción del problema

En América Latina, el mercado de seguros de vehículos enfrenta desafíos estructurales en términos de atracción y fidelización de clientes. A pesar de que la región muestra un crecimiento sostenido en el volumen de primas, la penetración de seguros voluntarios sigue siendo baja en comparación con otras regiones del mundo. Esto se debe, en gran parte, a factores como la informalidad económica, la falta de cultura aseguradora, la percepción de altos costos y la complejidad en los procesos de adquisición [1].

Adicionalmente, la retención de clientes continúa siendo un reto para las aseguradoras de la región. Muchas compañías interactúan de forma esporádica con sus asegurados, lo que limita la construcción de relaciones duraderas. Esta desconexión suele traducirse en bajos niveles de lealtad, rotación elevada y una fuerte presión sobre los costos de adquisición de nuevos clientes [2].

A pesar del avance en transformación digital, gran parte del sector asegurador latinoamericano no ha implementado de manera efectiva soluciones tecnológicas que permitan ofrecer experiencias más personalizadas, ágiles y predictivas. Herramientas como la analítica avanzada, los modelos de predicción de abandono (churn), y las estrategias basadas en Customer Lifetime Value (CLV) siguen siendo poco utilizadas o están en etapas tempranas de adopción [3].

Por otro lado, las expectativas de los consumidores han cambiado radicalmente: buscan experiencias híbridas, accesibles desde canales digitales, pero con atención humana cuando lo requieren. Muchas aseguradoras no logran satisfacer esta demanda, generando fricciones en la experiencia del cliente y debilitando su fidelización [4].

## Problema de negocio

En Colombia, el mercado de seguros voluntarios para vehículos enfrenta una penetración considerablemente baja. Según la Federación de Aseguradores Colombianos (Fasecolda), solo el 12,2% del parque automotor del país cuenta con algún tipo de cobertura voluntaria, y en el caso de las motocicletas, esta cifra se reduce a apenas el 3% [5]. Esta baja cobertura refleja tanto una oportunidad de expansión para las aseguradoras como un llamado urgente a mejorar los mecanismos de atracción y fidelización de clientes.

Según Hwang et al. (2004), la deserción de clientes tiene un impacto directo en el Customer Lifetime Value (CLV), ya que reduce la frecuencia y duración de las transacciones, afectando negativamente la proyección de ingresos [6]. Esta situación es especialmente crítica en los seguros de movilidad, donde el ciclo de vida del cliente y su comportamiento digital ofrecen señales valiosas que, si se aprovechan adecuadamente, pueden optimizar tanto la retención como la rentabilidad del portafolio.

En este contexto, los canales digitales representan una vía estratégica para transformar la relación con los usuarios, permitiendo ampliar el alcance comercial y personalizar las ofertas de valor. No obstante, las aseguradoras colombianas enfrentan dificultades para aprovechar estos canales debido a la limitada implementación de modelos analíticos que permitan segmentar con precisión, anticipar comportamientos de deserción y actuar proactivamente para mejorar la experiencia del cliente [7].

## Aproximación desde la analítica de datos

En el contexto de la analítica de negocios, la identificación de clientes de alto valor se ha consolidado como una estrategia clave para la optimización de procesos comerciales y de retención. En esta monografía, se propone el uso de modelos de Machine Learning (ML) para predecir el Customer Lifetime Value (CLV) de los usuarios del ecosistema digital de Sura, enfocados en el segmento de movilidad. Este enfoque busca orientar las decisiones de adquisición y fidelización de clientes, maximizando la rentabilidad y eficiencia de las campañas de marketing y atención personalizada [8].

Los modelos predictivos permiten encontrar patrones en grandes volúmenes de datos históricos de comportamiento digital, uso de servicios, frecuencia de interacción, perfil demográfico y canales preferidos. Técnicas como *regresión lineal, XGBoost, Random Forest*, así como enfoques más avanzados como *modelos de redes neuronales profundas o survival analysis,* han demostrado ser eficaces en la estimación del CLV [9].

Un ejemplo concreto de aplicación sería la identificación temprana de clientes con alta propensión a abandono, pero con un CLV alto proyectado, lo que permitiría activar campañas de retención específicas, evitando pérdidas significativas. De igual forma, se podrían ajustar las estrategias de adquisición digital para enfocar recursos en perfiles similares a los de clientes con alto CLV, mejorando el ROI de las campañas [10].

Este enfoque data-driven tiene una relevancia directa en la optimización de la estrategia comercial de Sura, permitiendo tomar decisiones basadas en valor futuro estimado en lugar de criterios generales como volumen o frecuencia de compra. Además, fortalece la personalización en la experiencia del cliente, construyendo relaciones más duraderas y rentables [11]

**Limitaciones de la solución propuesta**

Pese a su potencial, la solución analítica propuesta presenta algunas limitaciones:

* No se proporciona una metodología estandarizada para la recolección y consolidación de datos históricos completos del cliente, lo cual puede afectar la precisión de los modelos [9].
* No se abordan criterios de calidad de datos, como integridad, consistencia o actualizaciones en tiempo real, que son fundamentales para estimaciones fiables del CLV.
* No contempla un plan de implementación del modelo entrenado en sistemas productivos o plataformas digitales existentes de Sura (por ejemplo, CRM, apps móviles).

**Riesgos asociados al desarrollo del proyecto**  
Se identifican diversos riesgos que podrían afectar la viabilidad del proyecto:

* Disponibilidad y representatividad de los datos: los modelos de ML requieren grandes volúmenes de datos históricos. Si los datos no están consolidados o no representan adecuadamente a los clientes actuales, el modelo será poco fiable [8].
* Limitaciones computacionales: el entrenamiento de ciertos modelos avanzados puede requerir infraestructura de cómputo especializada (GPU, clusters), que podría no estar disponible.
* Sobrecarga por complejidad del modelo: si el modelo es demasiado complejo, podría generar resultados poco interpretables o difícilmente explicables a usuarios no técnicos, lo que podría comprometer su adopción [12].
* Riesgo de sesgo algorítmico: si los datos de entrenamiento están sesgados, el modelo podría reproducir o amplificar esos sesgos, afectando la equidad en la toma de decisiones[13].

## Origen de los datos

El conjunto de datos utilizado en este proyecto, denominado Data\_Analisis\_Movilidad\_U2Y.csv, fue suministrado por la compañía Suramericana de Seguros. Este dataset contiene información de clientes del canal digital recopilada durante los últimos dos años. Debido a la naturaleza sensible de los datos personales involucrados, se han aplicado medidas de protección desde la fase de diseño, en cumplimiento con los lineamientos establecidos por la Ley de Habeas Data.

El conjunto de datos se encuentra almacenado en la nube corporativa de Suramericana, desde donde es extraído mediante consultas a la base de datos Teradata, específicamente solicitando los registros asociados al canal digital.

El *dataset* final contiene 45 columnas y un total de 22,904 registros, que abarcan información demográfica de los clientes (edad, sexo, estado civil, ubicación), detalles de pólizas de seguros, características de los vehículos asegurados, así como métricas financieras relacionadas con primas y siniestros

## Métricas de desempeño

Dado que el modelo propuesto es de tipo supervisado y busca predecir una variable continua (CLV), se seleccionan las siguientes métricas de evaluación del desempeño técnico:

• **Error Absoluto Medio (MAE):** Esta métrica refleja la magnitud media del error entre los valores predichos por el modelo y los valores reales observados. Es ampliamente utilizada por su facilidad de interpretación y robustez ante la presencia de valores atípicos [14]. Se establece como umbral deseable un MAE inferior al 10 % del promedio del CLV en el dataset, lo que garantizaría una predicción lo suficientemente precisa para su uso en campañas de marketing y retención.

• **Raíz del Error Cuadrático Medio (RMSE):** A diferencia del MAE, el RMSE penaliza más fuertemente los errores de gran magnitud, lo que lo hace particularmente útil en contextos donde los errores extremos pueden generar impactos significativos en decisiones comerciales [15]. Un RMSE inferior al 15 % del CLV promedio es considerado aceptable, dado el comportamiento de las variables financieras analizadas.

• **Coeficiente de Determinación (R²):** Esta métrica representa la proporción de la variabilidad de la variable dependiente (CLV) que puede ser explicada por el modelo. Según Chicco, Warrens y Jurman (2021), un valor de R² igual o superior a 0.70 se considera adecuado para modelos aplicados a problemas de predicción en entornos reales, permitiendo una interpretación clara de la efectividad del modelo en la captura de patrones relevantes [16].

Además del rendimiento técnico, se consideran métricas orientadas a valorar el impacto del modelo en los procesos clave del negocio: la identificación, adquisición y retención de clientes valiosos.

• **Segmentación efectiva por valor:** Se evaluará la capacidad del modelo para separar a los clientes de alto valor de los de bajo valor, mediante análisis de distribución del CLV estimado y su comparación con valores históricos. Se espera una segmentación clara entre los cuartiles superiores e inferiores, lo cual permitirá direccionar estrategias diferenciales de fidelización y marketing.

• **Precisión en la predicción agregada:** Se contrastará la suma total del CLV predicho por el modelo con la suma real del CLV histórico, esperándose una diferencia máxima del ±5 %. Esta métrica global permitirá identificar sesgos acumulativos que podrían afectar decisiones estratégicas.

• **Cobertura de segmentación aplicable:** Se medirá el porcentaje de clientes del canal digital que pueden ser clasificados de forma confiable en grupos según su CLV estimado. Se espera lograr una cobertura superior al 85%, lo cual garantizaría la aplicabilidad real del modelo en entornos comerciales.

La definición de estos umbrales está respaldada por evidencia empírica. Según diversos estudios, un R² entre 0.70 y 0.80 suele considerarse adecuado para modelos de predicción de CLV, siempre que el MAE y el RMSE estén por debajo de 0.20 [17]. Asimismo, técnicas de ensamblado como *Random Forest*, *XGBoost* y *LightGBM* han demostrado mejor rendimiento frente a modelos lineales simples en problemas similares, especialmente por su capacidad para manejar relaciones no lineales y estructuras complejas de datos [18].

# Objetivos

## Objetivo general

Identificar clientes de alto valor en el ecosistema digital de Sura mediante modelos de Machine Learning, con el fin de optimizar la adquisición y retención de clientes a través de la exploración del Customer Lifetime Value (CLV) en el segmento de movilidad.

## Objetivos específicos

* Obtener y consolidar datos relevantes del ecosistema digital de Sura asociados al comportamiento de los clientes del segmento de movilidad, incluyendo transacciones, interacciones digitales, historial de compras, renovaciones, uso de servicios y características demográficas.
* Preprocesar, limpiar y transformar los datos recolectados, manejando valores nulos, outliers y normalización de variables, así como codificación de atributos categóricos relevantes para el modelo.
* Calcular el Customer Lifetime Value (CLV) histórico de los clientes del canal digital de Sura correspondientes a los últimos dos años, utilizando los datos consolidados de primas devengadas, siniestros incurridos y costos de asistencias, con el fin de establecer una métrica base confiable para análisis predictivo y segmentación.
* Construir modelos predictivos supervisados para estimar el CLV futuro de los clientes, utilizando algoritmos como regresión lineal, árboles de decisión, random forest, XGBoost o redes neuronales, evaluando su desempeño con métricas como MAE, RMSE y R².
* Entregar al gerente del canal digital de Sura un informe ejecutivo con los resultados del modelo de CLV, con el fin de respaldar la toma de decisiones estratégicas relacionadas con la inversión en pauta digital y las acciones de fidelización de clientes

# Datos

## Datos originales

Este es un conjunto de datos crudos suministrado por la compañía Suramericana de Seguros, correspondiente a clientes del canal digital en el segmento de movilidad. El dataset fue extraído directamente desde la nube corporativa de la organización mediante consultas estructuradas al sistema de almacenamiento Teradata, empleando filtros específicos para seleccionar los registros de clientes activos en los últimos dos años.

El archivo principal, denominado *Data\_Analisis\_Movilidad\_U2Y.csv*, se encuentra en formato delimitado por comas (CSV), compatible con herramientas analíticas como Python, R y SQL. El archivo cuenta con un tamaño aproximado de 15.6 MB y contiene 22,904 registros distribuidos en 45 columnas. Estos datos reflejan múltiples dimensiones de los clientes y sus pólizas, lo que permite un análisis detallado tanto del comportamiento individual como de tendencias globales.

Tabla 1

Datos originales del *Dataset.*



Nota. Elaboración propia, código disponible en el repositorio asociado al proyecto (Guisao & Arias, 2025) [19].

**Estructura y descripción de columnas**

Las columnas del archivo abarcan información demográfica, financiera, comercial, vehicular y de uso de servicios, todas relacionadas con las pólizas contratadas por los clientes. A continuación, se describen algunas categorías principales de datos:

• **Datos demográficos:** *Incluyen Edad\_Cliente, Sexo\_Cd, Departamento\_Cliente, Ciudad\_Cliente y Estado\_Civil\_Desc*. Estas variables permiten analizar el perfil del cliente desde una perspectiva sociodemográfica.

• **Datos financieros:** Contienen información como *Prima\_Total\_Devengada\_Certificado, Siniestros\_Total\_Incurridos\_Certificado, Costo\_Total\_Asistencias\_Certificado y Margen\_Bruto\_Total\_Certificado*. Estos valores han sido clave para el cálculo del CLV y presentan distribuciones sesgadas, con valores extremos en algunos casos.

• **Datos de pólizas:** Variables como *Numero\_Poliza\_Anonymized, Subproducto\_Desc y Dias\_Vigencia\_Certificado* indican los productos adquiridos, su duración y tipo de cobertura.

* **Datos de vehículo:** Entre estas se encuentran *Valor\_Comercial\_Vehiculo, Marca\_Vehiculo, Modelo\_Vehiculo* y variables accesorias como *Valor\_Accesorios\_Vehiculo\_Sia.*

• **Variables binarizadas:** Se incluyen columnas como *Cliente\_Tiene\_Soat\_Activo, Cliente\_Tiene\_Hogar\_Activo*, entre otras, que fueron convertidas a formato categórico (‘Sí’/‘No’) para facilitar la interpretación

Asimismo, se incorporaron dos variables nuevas durante la fase de preparación: CAC\_Promedio, con valores diferenciados según el tipo de vehículo (autos: $300.000; motos: $70.000), y CLV, que representa la variable objetivo para el modelado predictivo.

**Distribución y calidad de los datos**

El dataset se encuentra completo y sin valores nulos en la mayoría de sus columnas. Sin embargo, se identificaron algunas inconsistencias, como edades no válidas *(Edad\_Cliente con valores <18 años)*, las cuales fueron imputadas con la media de la variable. En campos como *Sexo\_Cd*, se eliminaron registros con valores no válidos (-1), y en columnas como Estado\_Civil\_Desc o Rango\_Ingresos\_Cliente, se excluyeron registros con alta proporción de datos faltantes (>80%).

La variable Num\_Total\_Asistencias\_Certificado presentó una distribución altamente sesgada, con valores atípicos en algunos clientes que utilizaron el servicio de forma intensiva. Esto se consideró importante para segmentaciones futuras. Las variables cuantitativas se transformaron al formato numérico estándar para su análisis en Python, corrigiendo errores de separación decimal y símbolos.

**Acceso y restricciones**

El acceso a los datos está restringido y controlado internamente por Sura. Para la extracción del dataset fue necesario el uso de credenciales corporativas y permisos de acceso al entorno de consultas Teradata. El archivo original no se encuentra disponible públicamente y su uso está limitado a fines analíticos internos, en cumplimiento con la Ley de Habeas Data en Colombia.

Para proteger la confidencialidad de los clientes, se aplicaron técnicas de anonimización en campos sensibles tales como *Dni\_Aseg\_Anonymized, Asegurado\_Id\_Anonymized, Nombre\_Agente\_Anonymized* y *Numero\_Poliza\_Anonymized*. Estos campos fueron excluidos de los análisis predictivos, evitando cualquier intento de reidentificación de usuarios.

**Limitaciones del conjunto de datos**

Una de las principales limitaciones es que los registros solo cubren un período de dos años recientes, lo cual podría subestimar el valor histórico real de clientes con relaciones previas. Además, la eliminación de algunas categorías específicas con baja representación (como productos descontinuados) podría sesgar la interpretación de comportamiento de ciertos nichos de clientes. Finalmente, la alta proporción de registros sin información en ciertas variables sociodemográficas limita el uso de dichas variables en modelos explicativos más complejos.

Describe también el modo de acceso a los datos y restricciones que pudiera haber sobre los mismos.

**Limpieza**

En esta etapa se aplicaron acciones clave de limpieza y transformación de datos con el objetivo de asegurar la calidad, coherencia y relevancia del conjunto de datos para análisis posteriores. Las operaciones realizadas fueron:

* Originalmente contenían valores numéricos representados como texto con formato regional, es decir, usando puntos como separadores de miles y comas como separadores decimales. Para estandarizar estos datos al formato numérico requerido por Python (donde el punto es el separador decimal).

Este procedimiento asegura que variables como Valor\_Comercial\_Vehiculo, Margen\_Tecnico\_Total\_Certificado o Ultima\_Prima\_Anualizada\_Cert, entre otras, puedan ser tratadas correctamente como valores cuantitativos, eliminando inconsistencias derivadas del formato de entrada y habilitando análisis posteriores precisos y confiables.

* En las columnas Valor\_Accesorios\_Vehiculo\_Sia y Valor\_Accesorio\_Especial\_Veh

contenían valores no numéricos representados por el carácter "?", se sustituyeron todos por 0, asumiendo que la ausencia de un valor representa la inexistencia de un accesorio o valor monetario asociado.

**Tratamiento de la data**

* En la columna *‘Edad\_Cliente’* presenta 415 registros con valores ‘-1’ ó inferiores a 18 años, por la cantidad de registros, los cuales son alrededor del 1.8% del total de la data se toma la decisión de reemplazarlos por la media de la columna.
* La Columna *‘segmento\_edad’* presenta 417 registros con valores *‘?’* los cuales se reemplazan por la moda de la columna.
* En la columna *‘Sexo\_Cd’* se encontraron 382 registros con valores *‘-1’* los cuales fueron eliminados ya que no tenían mayor representación en el conjunto de datos.
* En las columnas *“Ciudad\_Cliente”, “Departamento\_Cliente”, “Sector\_Economico\_Cliente”*, se eliminan los registros *“Sin\_Información”*, estos son registros que representan menos del 1% de la data.
* Se excluyeron planes específicos de la variable *“Subproducto\_Desc”* que no aplicaban al análisis o representaban menos del 1% del total de registros, como “PLAN BICISURA”, *“No Aplica”* y *“Plan Seguro Muévete Libre”,* con el fin de evitar sesgos en el modelo.
* Se eliminan algunas variables que no serán utilizadas en el dataset, los cuales son los valores anonimizados y la columna *“Rango\_Ingresos\_Cliente”* *y “Estado\_Civil\_Desc”*, estas últimas ya que había 18.744 registros sin información.
* En las columnas *“Valor\_Accesorios\_Vehiculo\_Sia”* y *“Valor\_Accesorio\_Especial\_Veh”* se reemplazan los registros con valores *“?”* y se reemplazan por *“0”*

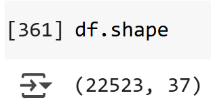
**Conversión de variables binarias numéricas a categóricas legibles:**

* Las variables relacionadas con productos activos del cliente *“Cliente\_Tiene\_Soat\_Activo”, “Cliente\_Tiene\_Hogar\_Activo”,”Cliente\_Tiene\_MasVida\_Activo”,”Cliente\_Tiene\_Salud\_Activo”,”Cliente\_Tiene\_Vida\_Activo”* fueron transformadas de valores numéricos (0 y 1) a etiquetas categóricas *('No' y 'Sí')*. Esto mejora la interpretabilidad y utilidad en visualizaciones y modelos de clasificación.

Este conjunto de transformaciones permitió refinar el data set, eliminando ruido y garantizando que los datos restantes sean representativos, consistentes y aptos para análisis exploratorios, segmentaciones o modelado predictivo.

**El conjunto de datos final quedó con 37 columnas y 22523 registros**

Tabla 2 datos finales del *Dataset*



Nota. Elaboración propia, código disponible en el repositorio asociado al proyecto (Guisao & Arias, 2025) [19].

## Datsets

Con el dataset ya preprocesado, se procedió a la división de los datos en subconjuntos para entrenamiento y validación. Este proceso siguió el enfoque clásico de partición aleatoria estratificada, comúnmente utilizado en aprendizaje supervisado, para asegurar una representación equitativa de las clases y características relevantes en ambos subconjuntos.

* **Dataset de entrenamiento (train):** corresponde al 80 % del total de los registros. Se utilizó para ajustar los parámetros de los modelos de regresión, incluyendo regresión lineal, *Random Forest*, *XGBoost* y *LightGBM*.
* **Dataset de validación (test):** corresponde al 20 % restante de los datos. Se empleó exclusivamente para evaluar la capacidad predictiva de los modelos construidos, permitiendo obtener métricas como MAE, RMSE y R² sin riesgo de sobreajuste (*overfitting*).

**Consideraciones adicionales**

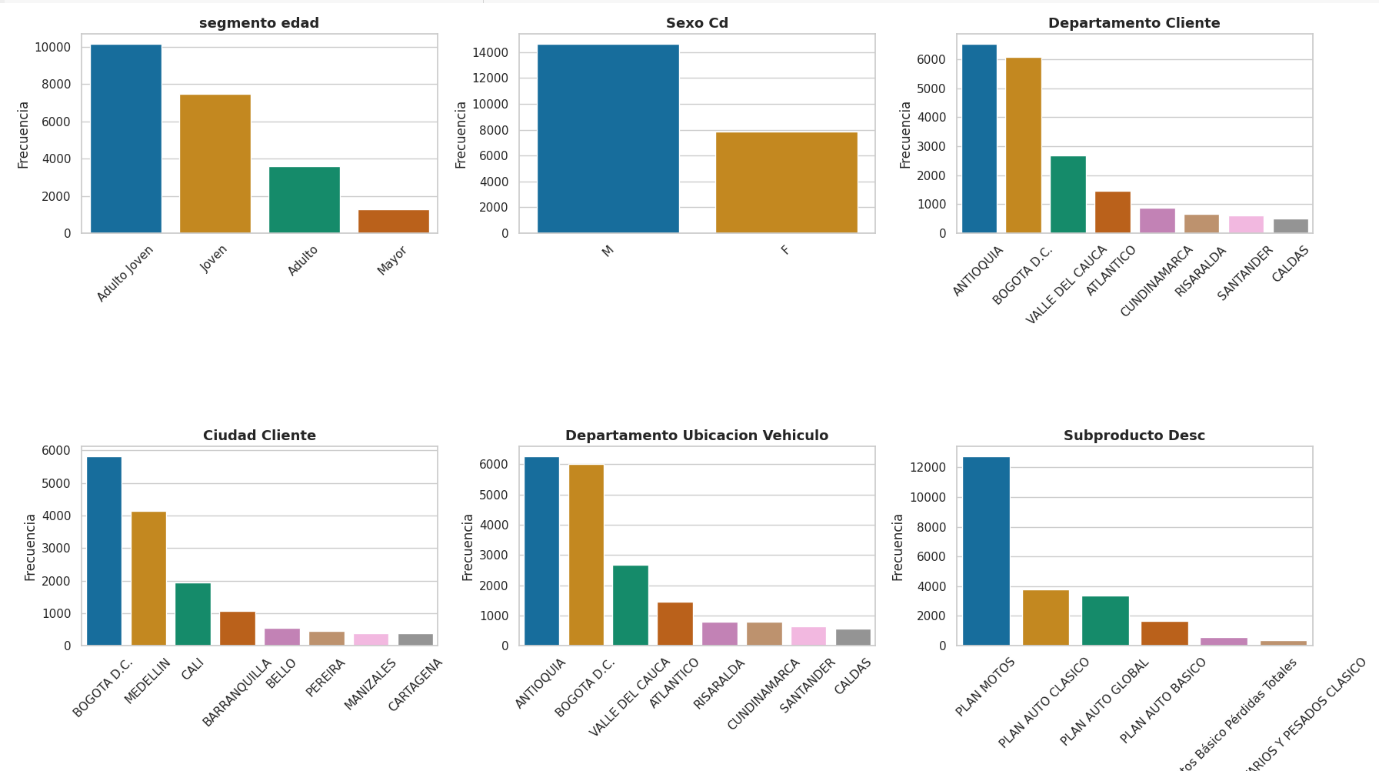
Se descartó el uso de validación cruzada (*k-fold cross-validation*) debido a que el tamaño del dataset y la representatividad en la partición 80/20 ya ofrecían un buen equilibrio entre eficiencia computacional y confiabilidad estadística. No obstante, en futuras iteraciones se contempla la implementación de validación cruzada con *grid search* para optimización de hiperparámetros.  
Además, se mantuvo la variable objetivo CLV fuera de cualquier transformación durante la división, garantizando que no haya fuga de información entre los subconjuntos. Los datasets obtenidos fueron almacenados como archivos independientes en formato .csv, asegurando trazabilidad y facilidad para posteriores experimentaciones.

## Analítica descriptiva

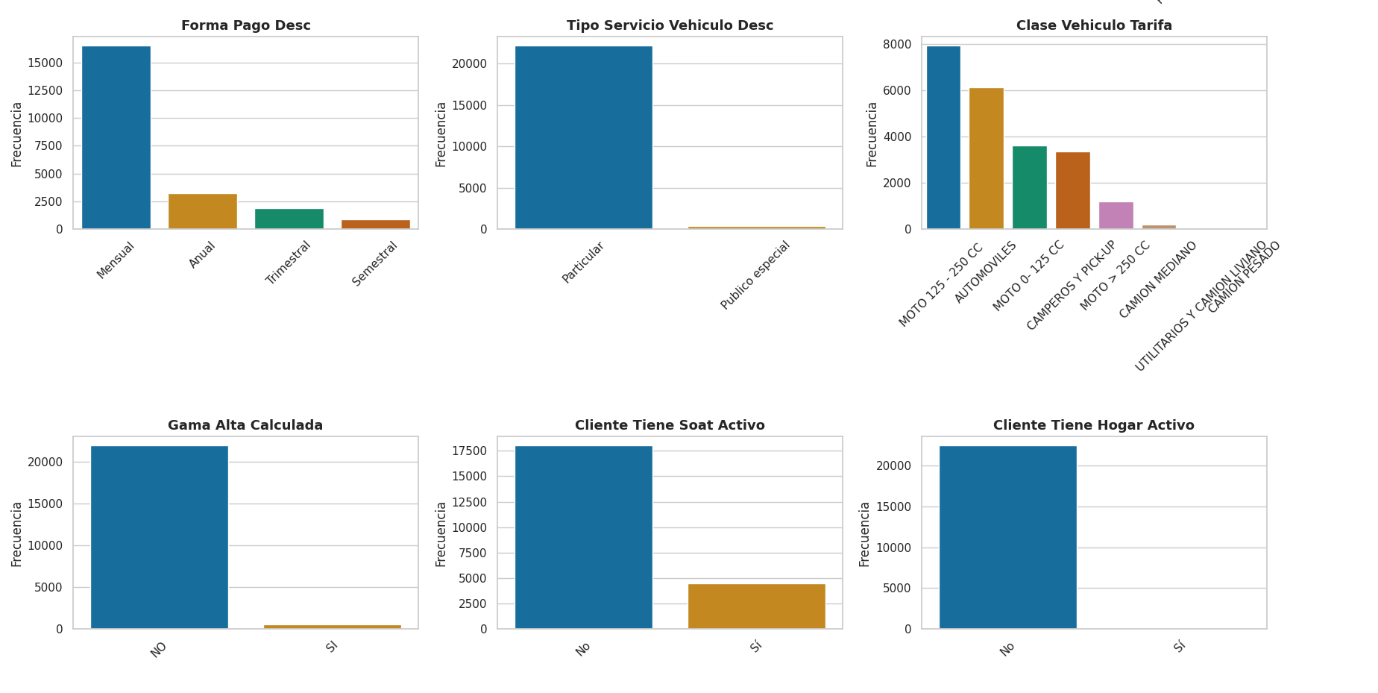
**Validación de Variables Categóricas.**

La mayoría de los clientes pertenecen al segmento "Adulto joven", son de sexo masculino y en su mayoría están ubicados en Antioquia y Bogotá, lo que refleja una alta concentración geográfica y un perfil demográfico definido. La gran mayoría de los vehículos son de uso particular, y se observa un bajo nivel de productos complementarios activos como SOAT o seguros de hogar. Adicionalmente, se identifican problemas de calidad de datos en la variable de estado civil, donde hay alta proporción de registros ‘sin información’, lo que podría afectar la precisión de futuros análisis si vamos a tener en cuenta esta variable.

Figura 1 Variables categóricas



# 

****

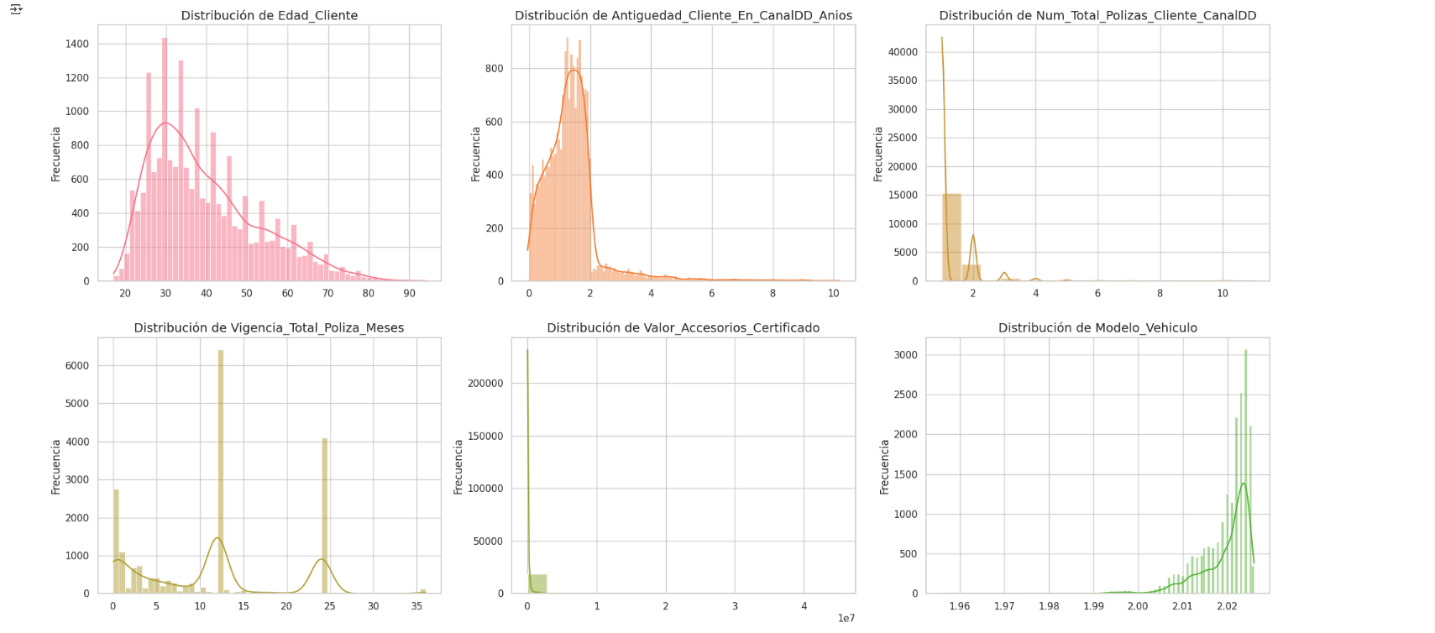
Nota. Elaboración propia, código disponible en el repositorio asociado al proyecto (Guisao & Arias, 2025) [19]

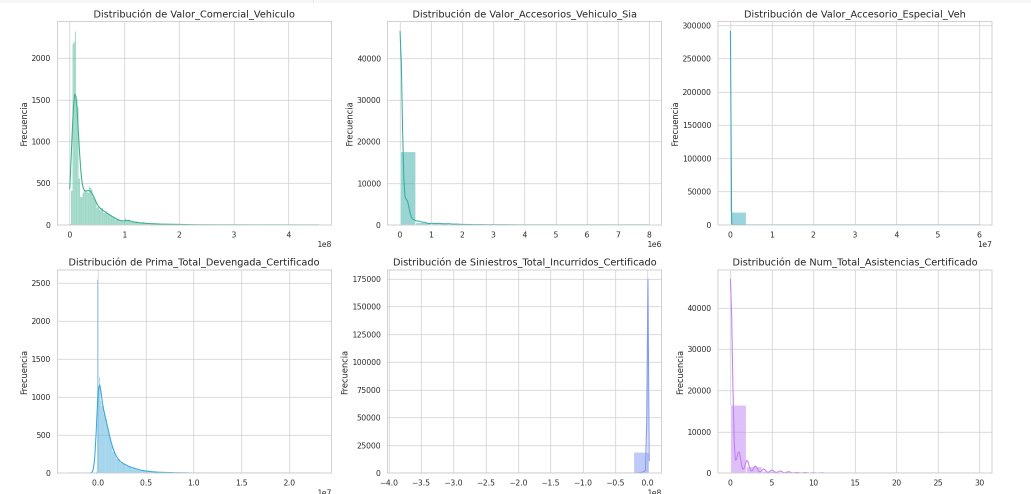
**Validación de variables numéricas**

Las variables numéricas del data set reflejan características clave del perfil del cliente y el estado de sus pólizas. La **Edad del Cliente** y la **Antigüedad en el canal** indican que la mayoría de los clientes son jóvenes, lo que puede influir en su comportamiento y necesidades. El **Número total de pólizas** revela que la mayoría de los clientes tienen pocas pólizas activas, lo cual puede indicar oportunidades para ampliar su portafolio de seguros.

En cuanto a las variables financieras, como el **Valor Comercial del Vehículo**, **Valor de Accesorios**, y la **Prima Total Devengada**, muestran una amplia variabilidad con predominancia de valores bajos y algunos casos de valores muy altos, sugiriendo una heterogeneidad significativa en el tipo de clientes y vehículos asegurados. Esto implica que existen clientes con vehículos de bajo y alto valor.

Las variables relacionadas con siniestros y asistencias indican que la mayoría de los clientes reportan pocas reclamaciones o asistencias, lo cual podría reflejar buen comportamiento o baja frecuencia de eventos adversos, aunque también podría indicar subregistro o clientes con poca experiencia en el uso de sus pólizas.

****Figura 2 Distribución de variables numéricas.

****

Nota. Elaboración propia, código disponible en el repositorio asociado al proyecto (Guisao & Arias, 2025) [19]

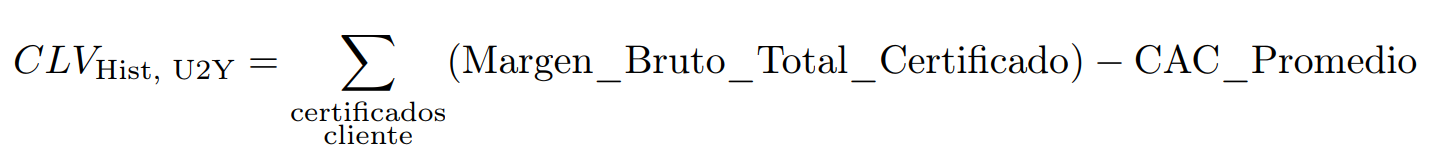
**Inclusión de variables en el Data set**

Dentro de la preparación de los datos debimos incluir dos variables numéricas importantes las cuales nos ayudarán con los diferentes modelos que se van a trabajar, teniendo en cuenta la formula del cálculo del **CLV** propuesto anteriormente.

El primero es *“CAC\_Promedio”* el cual es un dato suministrado por parte de un analista del canal digital, el cual nos informa del *Costo de adquisición promedio* para motos $70.000 y para autos $300.000

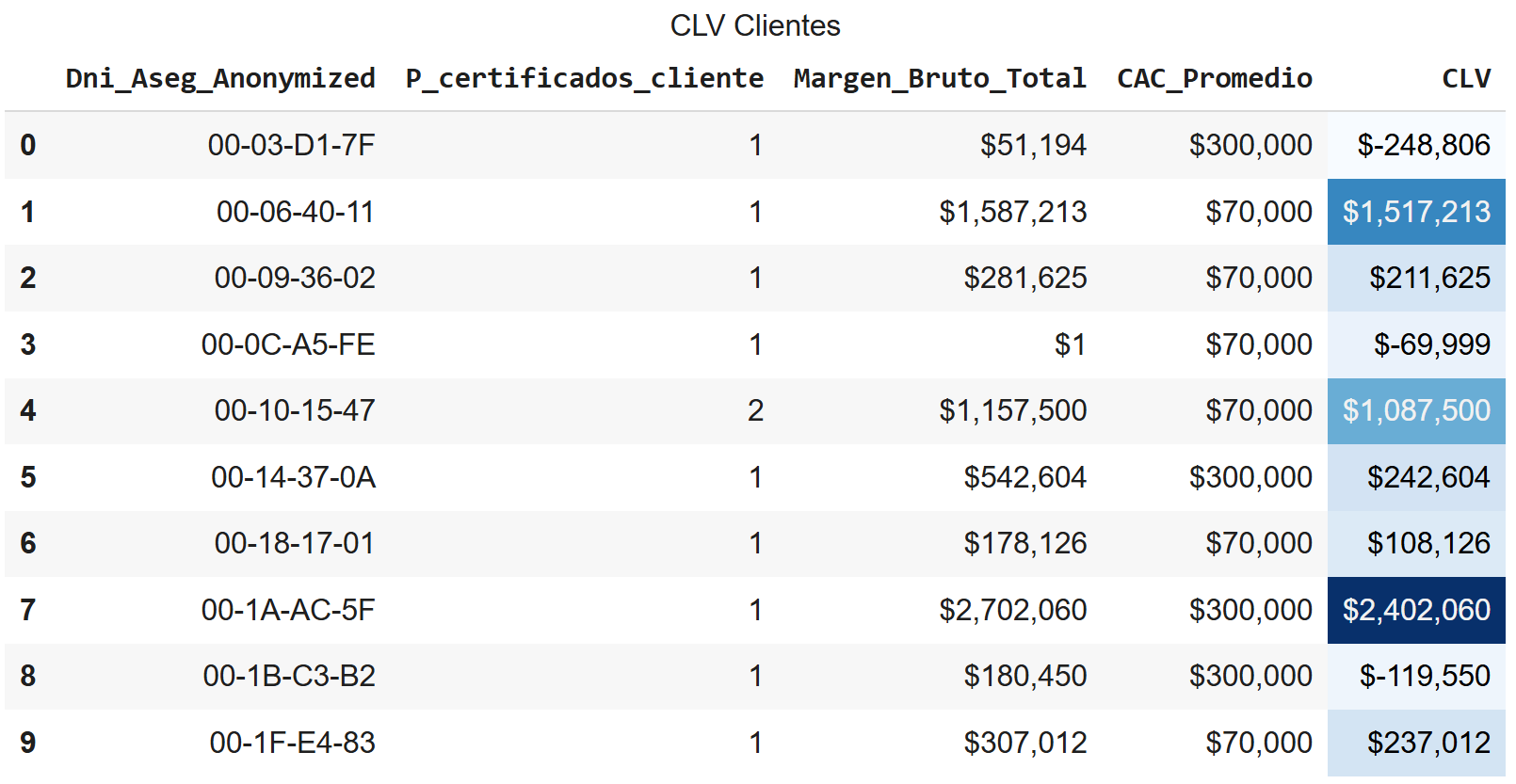
También se ingresa la columna llamada “*CLV”* el cual contiene la información calculada mediante a formula descrita anteriormente para el hallazgo del *Customer Lifetime Value*.

Es importante resaltar que esta va a ser nuestra Target, importante para la implementación de los modelos.

Figura 3. Formula del CLV

Nota. Reproducida de Hwang, H., Jung, T., & Suh, E. (2004). Expert Systems with Applications, 26(2), 181–188. <https://doi.org/10.1016/S0957-4174(03)00135-X> [6]

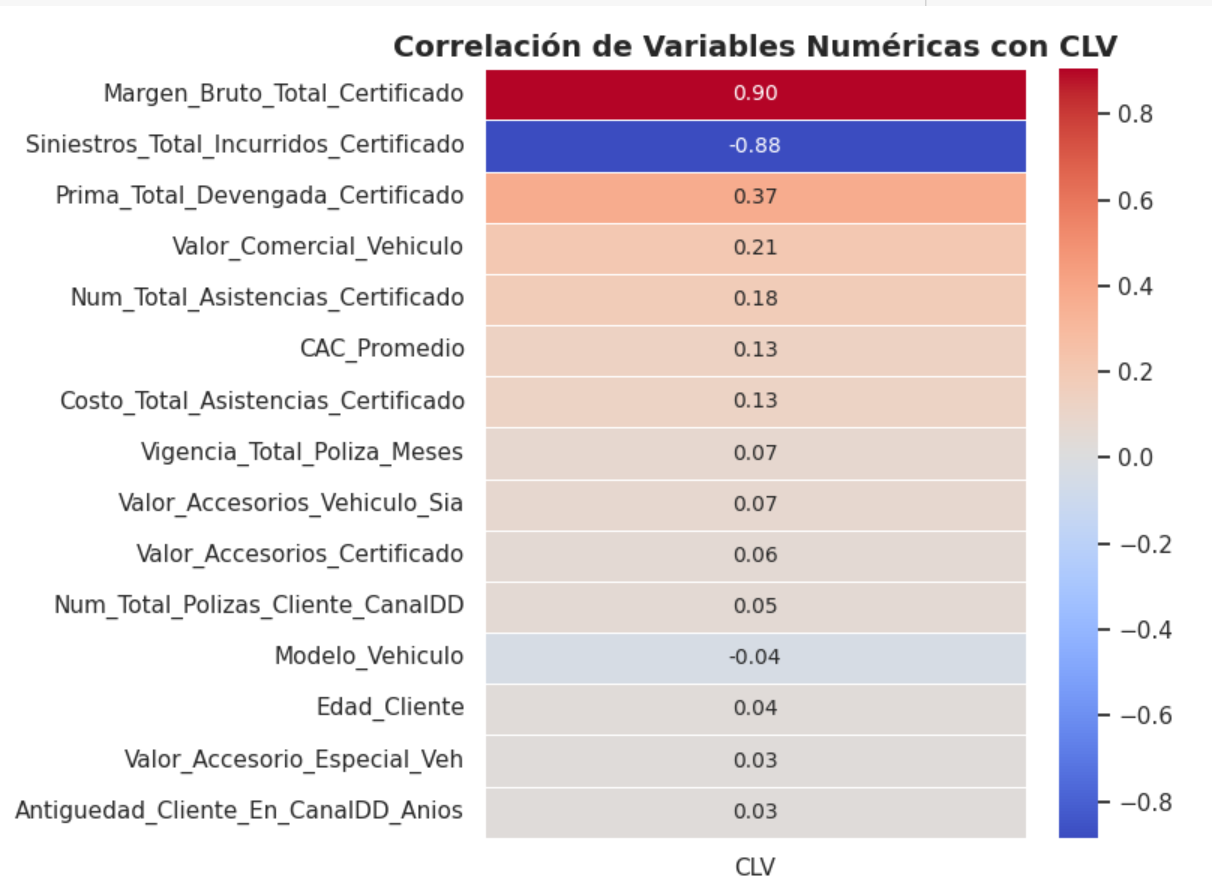
Tabla 3. Calculo de CLV Clientes



Nota. Elaboración propia, código disponible en el repositorio asociado al proyecto (Guisao & Arias, 2025) [19].

Una vez validadas todas las variables numéricas y categóricas, y luego de incluir las nuevas 2 columnas (*‘CAC\_Promedio’ y ‘CLV’*) se procedió a realizar de nuevo la matriz de correlación teniendo en cuenta que ya tenemos la variable predictora.

Figura 4 Análisis de correlación Numérica



Nota. Elaboración propia, código disponible en el repositorio asociado al proyecto (Guisao & Arias, 2025) [19].

El análisis de correlación muestra que las variables numéricas con mayor relación con el (*Costumer lifetime value*) son el ‘*Margen\_Bruto\_Total\_Certificado’* (correlación positiva de 0.90) y los ‘*Siniestros\_Total\_Incurridos\_Certificado’* (correlación negativa de -0.88). Esto sugiere que un mayor margen bruto está fuertemente asociado con un mayor CLV, mientras que un mayor monto en siniestros incurridos reduce significativamente el CLV. Otras variables como la Prima\_Total\_Devengada\_Certificado (0.37) y el Valor\_Comercial\_Vehículo (0.21) tienen una relación positiva más moderada. Por el contrario, variables como el Modelo\_Vehículo, Edad\_Cliente o Antigüedad\_Cliente\_En\_CanalDD\_Anios muestran una correlación cercana a cero, lo que indica una relación muy débil o inexistente con el CLV.

# Proceso de analítica

## Pipeline principal

Describe con un gráfico el flujo de trabajo general de los datos en tu proyecto. Añade o modifica secciones si lo consideras necesario

## Preprocesamiento

Describe las alternativas de preprocesamiento de datos que consideraste, aumentación de datos, etc.

## Modelos

Describe los distintos modelos que consideraste, sus configuraciones, etc.

## Métricas

Describe cómo calculas las métricas de desempeño ML y de negocio (Por ejemplo: con qué funciones de sklearn, tensorflow, etc.)

# Metodología

## Baseline

Describe tu primera iteración, sus resultados y los problemas técnicos que tuviste que resolver

## Validación

Describe tu proceso de validación, particiones de train/test/validación, etc.

## Iteraciones y evolución

Describe en qué te enfocaste en las iteraciones subsiguientes. No es necesario hacer un listado exhaustivo de las iteraciones, pero sí has de describir qué perseguías y qué obtuviste con cada grupo de iteraciones. P.ej. primero hicimos un conjunto de iteraciones para definir las opciones de preprocesamiento, luego hicimos unas iteraciones probando distintos modelos de tal tipo, etc.

## 5.4 Herramientas

Menciona las herramientas que usaste para tu proyecto

# Resultados y discusión

En los resultados se comunican los hallazgos y descubrimientos del estudio. Se incluyen tablas, figuras, diagramas y demás material demostrativo. Al narrar descriptivamente una figura, tabla, etc., en un párrafo, puedes insertar una referencia cruzada, es decir, un hipervínculo al elemento mencionado dentro o fuera de paréntesis, ejemplos: estos resultados se muestran en la **Tabla 1**. Igualmente, los datos son validados con otros instrumentos (**Tabla 2**, **Tabla 3**). Lineamientos que se establecen en la nueva versión de las Normas APA séptima edición (**Figura 1**). La producción intelectual institucional se publica en el Repositorio (**Figura 2**).

La discusión es la interpretación crítica y el análisis de los resultados, que surgen de las preguntas de investigación.

## Métricas

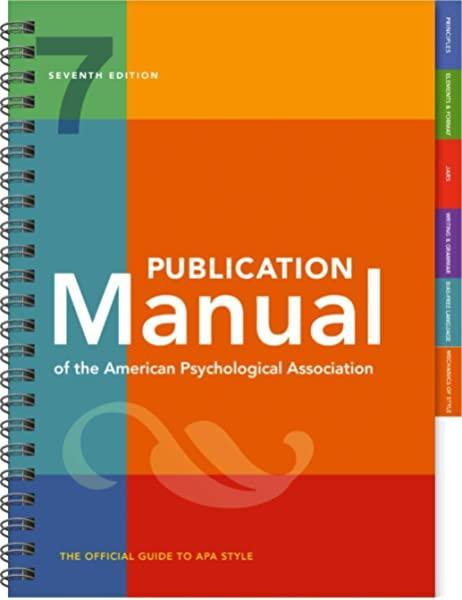
Describe los resultados numéricos de las métricas de las iteraciones que consideres más relevantes, junto con las configuraciones. Usa tablas o gráficas siguiendo el formato mostrado a continuación.

## Evaluación cualitativa

Realiza una evaluación cualitativa de los resultados, indicando casos de overfitting, underfitting, utilidad de los resultados, relación obtenida entre la métrica de ML y la de negocio, etc.

## Consideraciones de producción

Menciona consideraciones técnicas de una posible puesta en producción. Por ejemplo: condiciones de monitoreo del desempeño de los modelos, integración con streams de datos, servicios en la nube, etc.



*Nota.* Fuente https://bit.ly/2IyrZao (American Psychological Association, 2020).

**Figura 2** *Logo Universidad de Antioquia*

****

*Nota.* Fuente http:/www.udea.edu.co

# Conclusiones

Son las interpretaciones finales que recopilan los datos de la investigación, describe lo que se obtuvo, qué se logró y cuáles son los resultados. Guardan relación directa con lo que se mencionó en el planteamiento del problema y en los objetivos. Haz un breve juicio crítico de tu proyecto.

# Recomendaciones

Las recomendaciones son las futuras y posibles líneas de investigación que llevarán a resolver problemas relacionados con la presente investigación.

# Referencias

[1] McKinsey & Company. (2023). *Informe global de seguros 2025: En busca del crecimiento en América Latina*. <https://www.mckinsey.com/locations/south-america/latam/hispanoamerica-en-potencia/informe-global-de-seguros-2025-en-busca-del-crecimiento-en-america-latina>

[2] InsurMarket Latam. (2023). *Retención de clientes: el reto para las aseguradoras*. <https://insurmarketlatam.com/retencion-de-clientes-el-reto-para-las-aseguradoras>

[3] Álvarez, J. E., Arango Serna, M. D., & Valencia Ríos, G. (2017). Medición del valor de vida del cliente en empresas de servicios. *Revista Universidad y Empresa*, 19(32), 9–34. <https://revistas.urosario.edu.co/index.php/empresa/article/view/4609/8473>

[4] RGA Knowledge Center. (2023). *La emergente oportunidad de seguros en América Latina*. <https://rgaconocimiento.com/articulos/la-emergente-oportunidad-de-seguros-en-america-latina>

[5] Federación de Aseguradores Colombianos (Fasecolda). (2024, febrero 13). *Más colombianos protegidos, la meta de 2024*. <https://www.fasecolda.com/cms/wp-content/uploads/2024/02/Comunicado-Cifras-2024-F.pdf>

[6] Hwang, H., Jung, T., & Suh, E. (2004). An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry. *Expert Systems with Applications*, 26(2), 181–188. <https://doi.org/10.1016/S0957-4174(03)00135-X>

[7] Murillo, M. P. (2024, octubre 23). *Los 5 mayores retos del sector asegurador colombiano en 2025*. Bdeo. <https://bdeo.io/blog/los-5-mayores-retos-del-sector-asegurador-colombiano-en-2025/>

[8] Gupta, S., & Lehmann, D. R. (2005). *Managing customers as investments: The strategic value of customers in the long run*. Wharton School Publishing.

[9] Zhang, Y., Wei, Y., & Zhou, J. (2017). *Customer lifetime value prediction based on RFM model and LSTM*. *Procedia Computer Science*, 122, 384–391.

[10] Venkatesan, R., Kumar, V., & Bohling, T. (2007). *Optimal customer relationship management using Bayesian decision theory: An application for customer selection*. *Journal of Marketing Research*, 44(4), 579–594.

[11] Gómez-Uribe, C. A., & Hunt, N. (2015). *The Netflix Recommender System: Algorithms, Business Value, and Innovation*. *ACM Transactions on Management Information Systems (TMIS)*, 6(4), 1–19

[12] Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2012). *Building comprehensible customer churn prediction models with advanced rule induction techniques*. *Expert Systems with Applications*, 39(17), 13078–13086.

[13] Venkatesan, R., Kumar, V., & Bohling, T. (2007). *Optimal customer relationship management using Bayesian decision theory: An application for customer selection*. *Journal of Marketing Research*, 44(4), 579–594.

[14] Han, J., Pei, J., & Kamber, M. (2022). *Data mining: concepts and techniques* (4th ed.). Morgan Kaufmann.

[15] Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O’Reilly Media

[16] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, e623. <https://doi.org/10.7717/peerj-cs.623>

[17] Kaggle Customer Analytics. (2020). *Customer Lifetime Value Prediction*. Kaggle. <https://www.kaggle.com/code/bhavikjikadara/customer-lifetime-value-clv>

[18] KDnuggets. (2020). *Churn prediction with Machine Learning*. <https://www.kdnuggets.com/2020/03/customer-churn-prediction.html>

[19] AnderGP512. (2025). *Proyecto-de-Grado* [Repositorio]. GitHub. <https://github.com/AnderGP512/Proyecto-de-Grado>

# Anexos

En los anexos se incluye material complementario que apoya la documentación investigativa, tales como consentimientos informados, entrevistas, material fotográfico, etc. Evite incluir material que puede estar protegido por derechos de autor, tales como pruebas psicológicas, fragmentos de libros, artículos de revistas, patentes, etc. Recuerda no incluir en tu documento datos de personas o entidades objetos de la investigación, tales como nombres, apellidos, cédulas, números telefónicos, consentimientos informados con datos personales (Resolución 8430 de 1993), nombres de empresas sin el consentimiento escrito del representante legal, fotografías en primer plano de personas (especialmente de menores de edad) y demás información que pueda contravenir los principios emitidos en la Ley Estatutaria 1581 de 2012 (Ley de protección de datos personales).

Los siguientes anexos contienen documentos de interés para el proceso de trabajo de grado, así como trucos y recomendaciones que surgen constantemente en la elaboración de un documento en Word.

## Anexo 1. Autoarchivo en Repositorio y documentos de interés

Al terminar todos los aspectos metodológicos, de redacción, de estructura y diagramación de tu tesis en Word, y con previo aval de la unidad académica, exporta el documento a versión PDF. Prepara también los anexos, si los tiene. Posteriormente, realiza la gestión de autoarchivo en el Repositorio Institucional <http://bibliotecadigital.udea.edu.co>, procedimiento que puedes consultar en video o versión PDF:

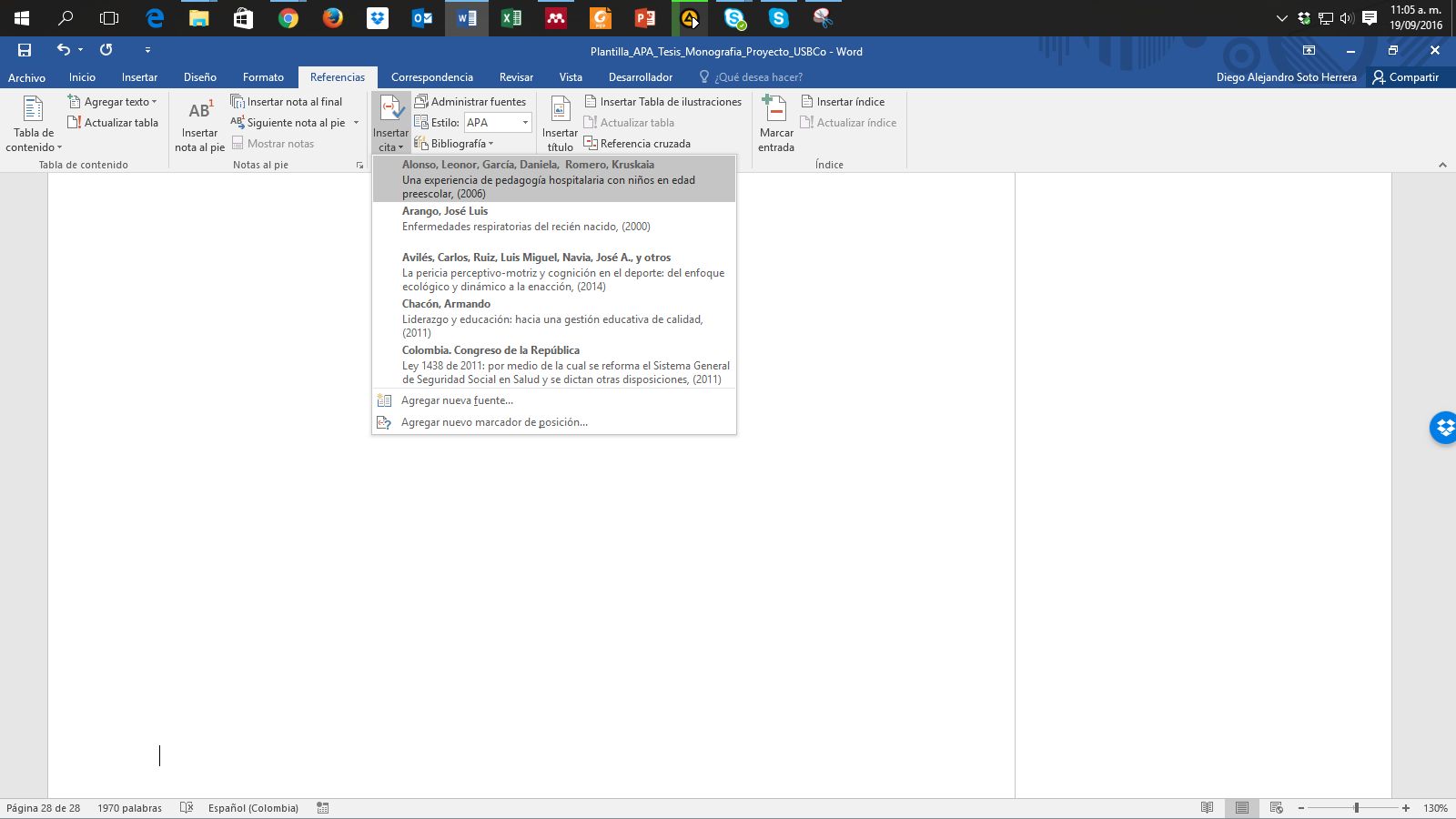
* Gestión de autoarchivo trabajos de grado (video): <https://bit.ly/3wx9U0E>
* Instructivo para el autoarchivo de trabajos de grado en el Repositorio Institucional Universidad de Antioquia (PDF): <https://bit.ly/3fOWbfB>

Recuerda que ya no se entregan trabajos de grado en CD-ROM, únicamente mediante formato digital a través del Repositorio Institucional. Otros documentos de interés para el proceso de entrega de trabajos de grado:

* Formulario institucional de entrega y autorización de trabajos de grado en la Universidad de Antioquia (diligenciar solo para 2 autores o más): <https://bit.ly/2Q0sc9P>
* Resolución Rectoral 47233 (21 de agosto de 2020): por la cual se establecen los lineamientos para la entrega de la producción académica de pregrado y posgrado en sus diferentes formatos y presentaciones al Repositorio Institucional del Departamento de Bibliotecas: <https://bit.ly/2R629hP>
* Políticas del Repositorio Institucional de la Universidad de Antioquia: <https://bit.ly/3t6dcG9>

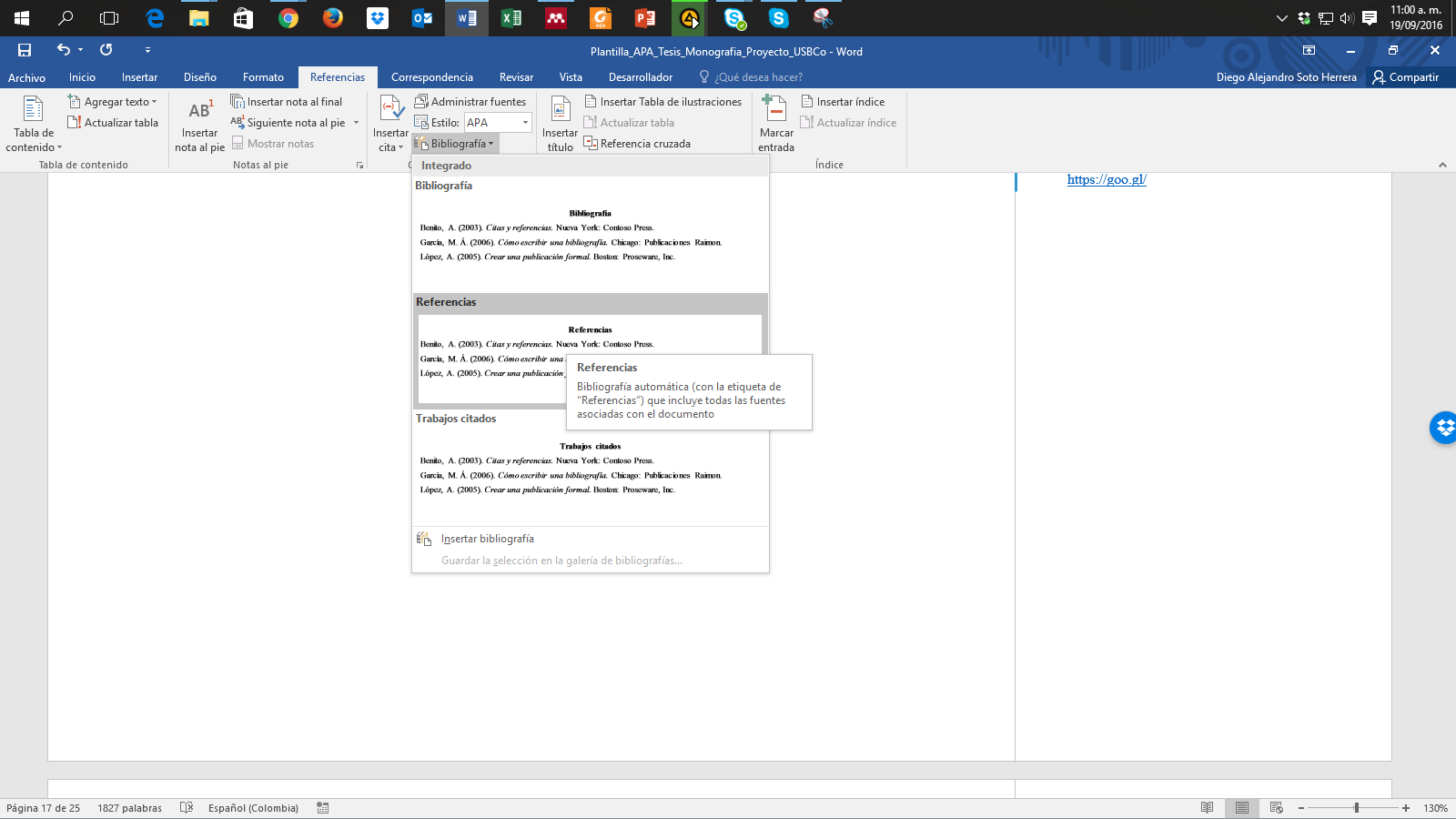
## Anexo 2. Gestor de citas y referencias de Microsoft Word Microsoft Word - Wikipedia

Ingresar las fuentes: Referencias > Administrar fuentes > Nuevo:

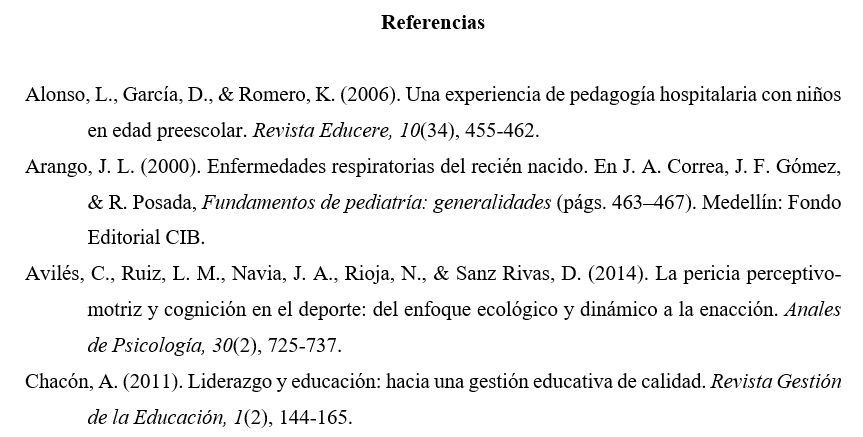
Insertar cita en el texto: Referencias > Insertar cita > Clic en fuente seleccionada:

Cita insertada dentro del texto:

Algunas experiencias significativas se han descrito mediante la pedagogía en hospitales con niños en edad preescolar (Alonso et al., 2006).

Insertar referencias (bibliografía): Referencias > Bibliografía > Referencias

Sección Referencias insertada:



## Anexo 3. Citas y referencias de material legal (leyes, decretos, sentencias, etc.)

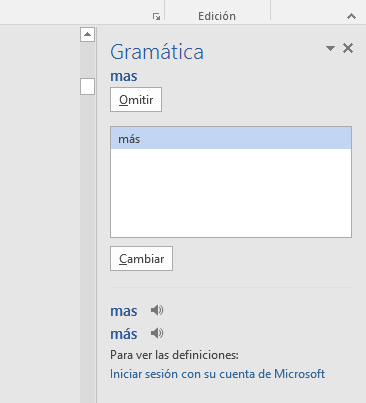
Uno de los aspectos que más puede causar confusión en Normas APA es lo referente a la citación de material legal y jurídico; de hecho, la misma American Psychological Association refiere al uso del manual internacional “Bluebook: A Uniform System of Citation” (este estilo está incluido en Zotero, Mendeley y otros gestores bibliográficos), pues estos dos estilos difieren en su formato de cita y referencia, pues las publicaciones legales citan las referencias al pie de página, en tanto que en el estilo APA se ubican todas las fuentes bibliográficas, incluyendo aquellas de materiales legales, en la lista de referencias. Si deseas conocer y adaptar los lineamientos del Bluebook, puedes consultarlos en <https://www.legalbluebook.com/>; asimismo, algunos ejemplos del manual de la APA están basados en el sistema jurídico estadounidense, lo que sin duda podría causar cierto conflicto con el entorno legal colombiano; ambos serán aceptados en los trabajos de grado y tesis de la Universidad de Antioquia. Sin embargo, para facilitar y adaptar las citas y referencias al sistema legal y jurídico colombiano, recomendamos los siguientes lineamientos basados en Normas APA como primera alternativa de citación y referenciación de los materiales más comunes en Colombia, a saber, leyes, decretos, sentencias, resoluciones, códigos, constitución política, entre otros. La primera recomendación está basada en el prefijo Colombia. como autor corporativo estatal, luego la subentidad y el año. Ejemplos de citas y referencias (se incluye un ejemplo internacional de España).

|  |  |
| --- | --- |
| **Cita (al interior del texto)** | **Referencias** |
| (Colombia. Presidencia de la República, 1991) | Colombia. Presidencia de la República. (1991). *Constitución Política de Colombia.* Presidencia de la República. |
| (Colombia. Congreso de la República, 1994) | Colombia. Congreso de la República. (1994). *Ley 133 de 1994 (mayo 23): por la cual se desarrolla el Decreto de Libertad Religiosa y de Cultos, reconocido en el artículo 19 de la Constitución Política*. Diario Oficial. |
| (Colombia. Presidencia de la República, 1998) | Colombia. Presidencia de la República. (1998). *Decreto 1504 de 1998: por el cual se Reglamenta el Manejo del Espacio Público en los Planes de Ordenamiento Territorial*. Diario Oficial. |
| (Colombia. Congreso de la República, 2014) | Colombia. Congreso de la República. (2014). *Ley 1733 de 2014: Ley Consuelo Devis Saavedra, mediante la cual se regulan los servicios de cuidados paliativos para el manejo integral de pacientes con enfermedades terminales, crónicas, degenerativas e irreversibles en cualquier fase de la enfermedad de alto impacto en la calidad de vida*. Diario Oficial. |
| (Colombia. Corte Constitucional, 2003a) | Colombia. Corte Constitucional. (2003a). *Sentencia SU.805 de 2003: vía de hecho en proceso de lanzamiento por ocupacion de hecho / debido proceso de querellados - vulneración por actuaciones arbitrarias. M. P. Jaime Córdoba Triviño*. Corte Constitucional. |
| (Colombia. Corte Constitucional, 2003b) | Colombia. Corte Constitucional. (2003b). *Sentencia T-361 de 2003: acción de tutela instaurada por Elkis Patricia Jiménez Castro contra la Universidad Cooperativa de Colombia – Seccional Santa Marta. M. P. Manuel José Cepeda Espinosa*. Corte Constitucional. |
| (Colombia. Corte Constitucional, 2006) | Colombia. Corte Constitucional. (2006). *Sentencia T-264 de 2006: acción de tutela instaurada por Fanny Stella Lesmes Galarza, en representación de su menor hijo Paul Andrés Rodríguez Lesmes contra la Universidad de los Andes. M. P. Jaime Araújo Rentería*. Corte Constitucional. |
| (Colombia. Ministerio de Salud y Protección Social, 2012) | Colombia. Ministerio de Salud y Protección Social. (2012). *Resolución 4331 de 2012 (diciembre 19): por medio de la cual se adiciona y modifica parcialmente la Resolución 3047 de 2008 modificada por la resolución 416 de 2009*. Diario Oficial. |
| (Colombia. Ministerio de Hacienda y Crédito Público. Superintendencia Financiera, 2006) | Colombia. Ministerio de Hacienda y Crédito Público. Superintendencia Financiera. (2006). *Circular Externa 048 de 2006 (diciembre 22)*. Superfinanciera. |
| (Colombia. Ministerio de Minas y Energía, 2010) | Colombia. Ministerio de Minas y Energía. (2010). *Reglamento técnico de iluminación y alumbrado público*. Ministerio de Minas y Energía. |
| (España. Ministerio de Trabajo y Asuntos Sociales. Instituto Nacional de Seguridad e Higiene en el Trabajo, 1996) | España. Ministerio de Trabajo y Asuntos Sociales. Instituto Nacional de Seguridad e Higiene en el Trabajo. (1996). *Evaluación de riesgos laborales*. INSHT. |
| (Colombia. Ministerio de Ambiente, Vivienda y Desarrollo Territorial, 2007) | Colombia. Ministerio de Ambiente, Vivienda y Desarrollo Territorial. (2007). *Decreto 3600 de 2007: por el cual se Reglamentan las Disposiciones de las Leyes 99 de 1993 y 388 de 1997 Relativas a las Determinantes de Ordenamiento del Suelo Rural y al Desarrollo de Actuaciones Urbanísticas de Parcelación y Edificación en este tipo de suelo y se adoptan otras disposiciones.* Diario Oficial. |
| (Colombia. Ministerio de Comunicaciones, 2001) | Colombia. Ministerio de Comunicaciones. (2001). *Resolución 000797 DE 2001 (junio 8): por la cual se atribuyen unas bandas de frecuencias radioeléctricas para su libre utilización dentro del territorio nacional*. Diario Oficial. |
| (Colombia. Ministerio de Educación Nacional, 2006) | Colombia. Ministerio de Educación Nacional. (2006). *Plan Decenal de Educación 2006-2016: Pacto Social por la Educación*. Ministerio de Educación Nacional. |
| (Colombia. Congreso de la República, 2010) | Colombia. Congreso de la República. (2010). *Código penal y de procedimiento penal anotado*. Leyer. |
| (Colombia. Congreso de la República, 2006) | Colombia. Congreso de la República. (2006). *Ley 1098 de 2006 (noviembre 8): por la cual se expide el Código de la Infancia y la Adolescencia en Colombia*. Diario Oficial. |
| (Colombia. Contraloría General de la Nación, 2003) | Colombia. Contraloría General de la Nación. (2003). *La deserción escolar en la educación básica media*. Contraloría General de la Nación. |
| (Colombia. Unidad Nacional para la Gestión del Riesgo de Desastres, 2012) | Colombia. Unidad Nacional para la Gestión del Riesgo de Desastres. (2012). *Guía para la formulación del Plan Municipal de Gestión del Riesgo de Desastres*. Sistema Nacional de Gestión del Riesgo de Desastres. |
| (Colombia. Departamento Administrativo Nacional de Estadística, 2011) | Colombia. Departamento Administrativo Nacional de Estadística. (2011). *Encuesta de convivencia escolar y circunstancias que la afectan - ECECA, para estudiantes de 5o a 11o de Bogotá*. DANE. |
| (Colombia. Departamento Administrativo de Ciencia, Tecnología e Innovación, 2015) | Colombia. Departamento Administrativo de Ciencia Tecnología e Innovación. (2015). *Resultados finales de la Convocatoria Nacional para el Reconocimiento y Medición de Grupos de Investigación, Desarrollo Tecnológico o de Innovación y para el Reconocimiento de Investigadores del SNCTeI*. Colciencias. |
| (Colombia. Procuraduría General de la Nación, 2012) | Colombia. Procuraduría General de la Nación. (2012). *Financiamiento del Sistema General de Seguridad Social en Salud: seguimiento y control preventivo a las políticas públicas*. Procuraduría General de la Nación. |

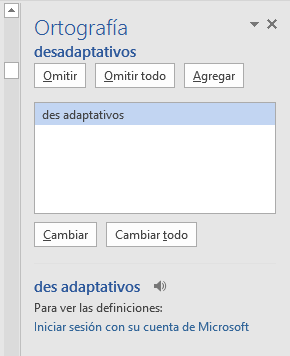
## Anexo 4. Ortografía y gramática

La ortografía y la gramática hacen parte fundamental del trabajo de grado; al finalizar la redacción de tu escrito, realiza una revisión ortográfica de todo el documento. En todo caso, siempre será recomendada y preferible la labor de un corrector de estilo que corrija redacción, ortografía, sintaxis, coherencia, citas, referencias y demás aspectos de estilo. En Microsoft Word, oprime la tecla F7. Tendrá dos tipos de sugerencias: Gramática y Ortografía, donde tendrás la opción de:

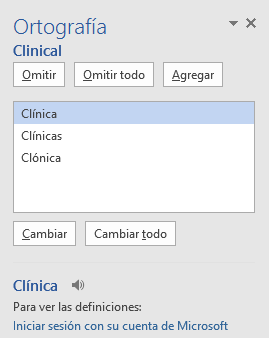
* “Cambiar”, si consideras que efectivamente había un error, ejemplo (mas, sin tilde):



* “Omitir”, si a pesar de la sugerencia consideras que está correctamente, ejemplo (desadaptativos):



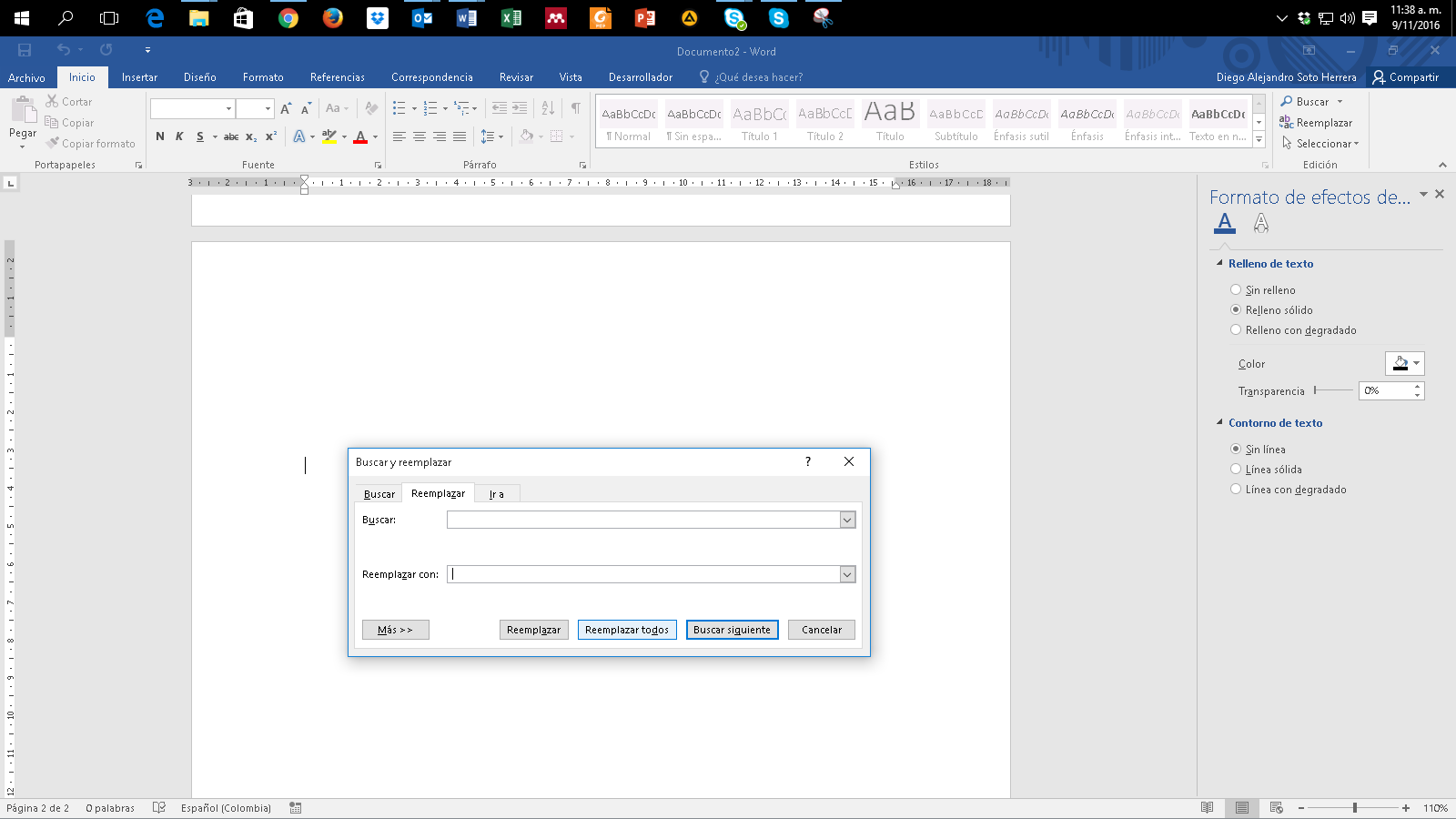
* “Omitir todo” si deseas ignorar la sugerencia de esa palabra u oración en todo el texto.
* “Agregar” si deseas incluir esa palabra en el diccionario en futuras revisiones.



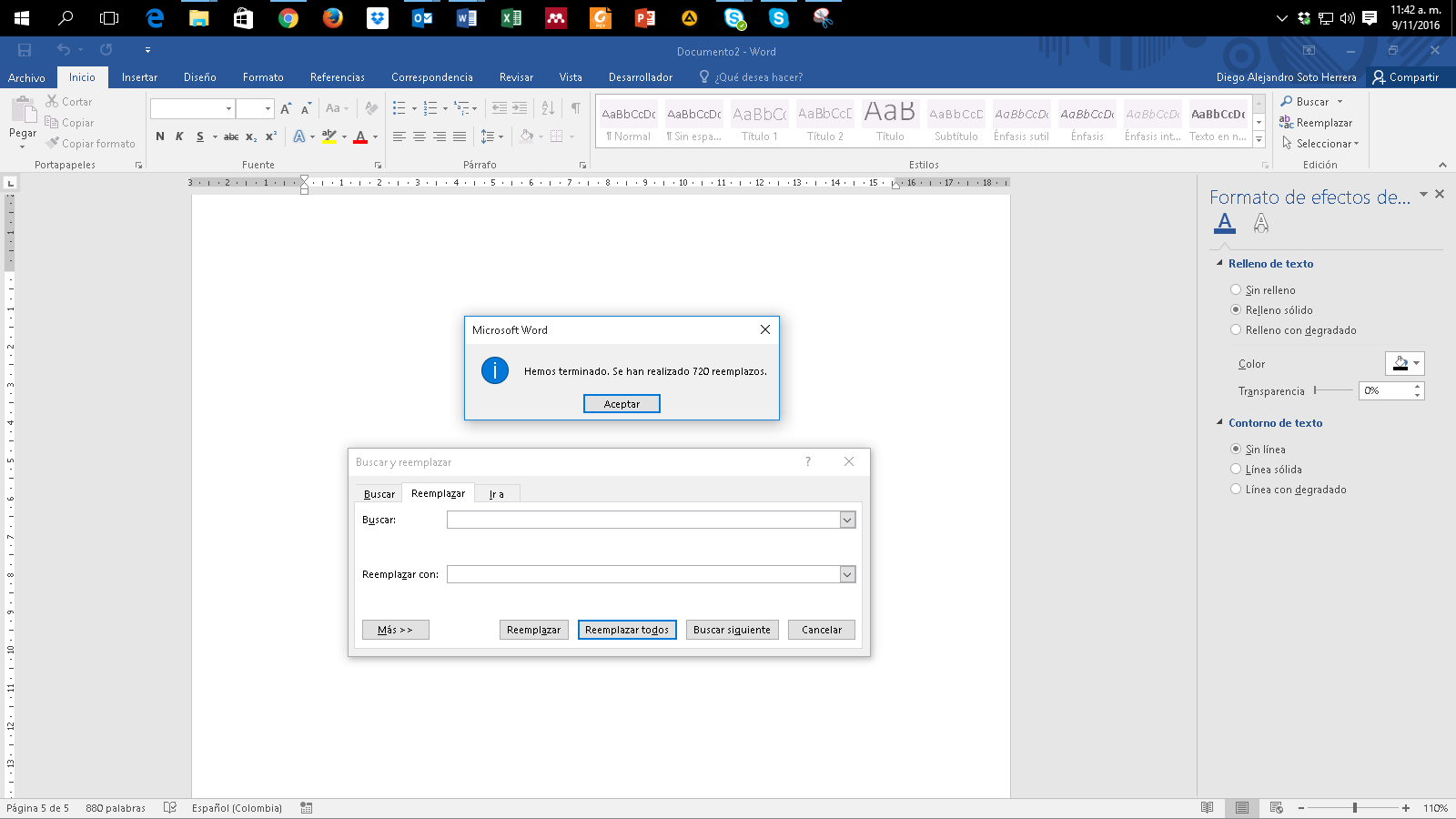
Ten precaución en aceptar cambios sugeridos, pues Microsoft Word no tiene la capacidad completa de interpretar con precisión algunos aspectos de la redacción o gramática en lengua española.

## Anexo 5. Buscar, reemplazar y eliminar espacios (o palabras)

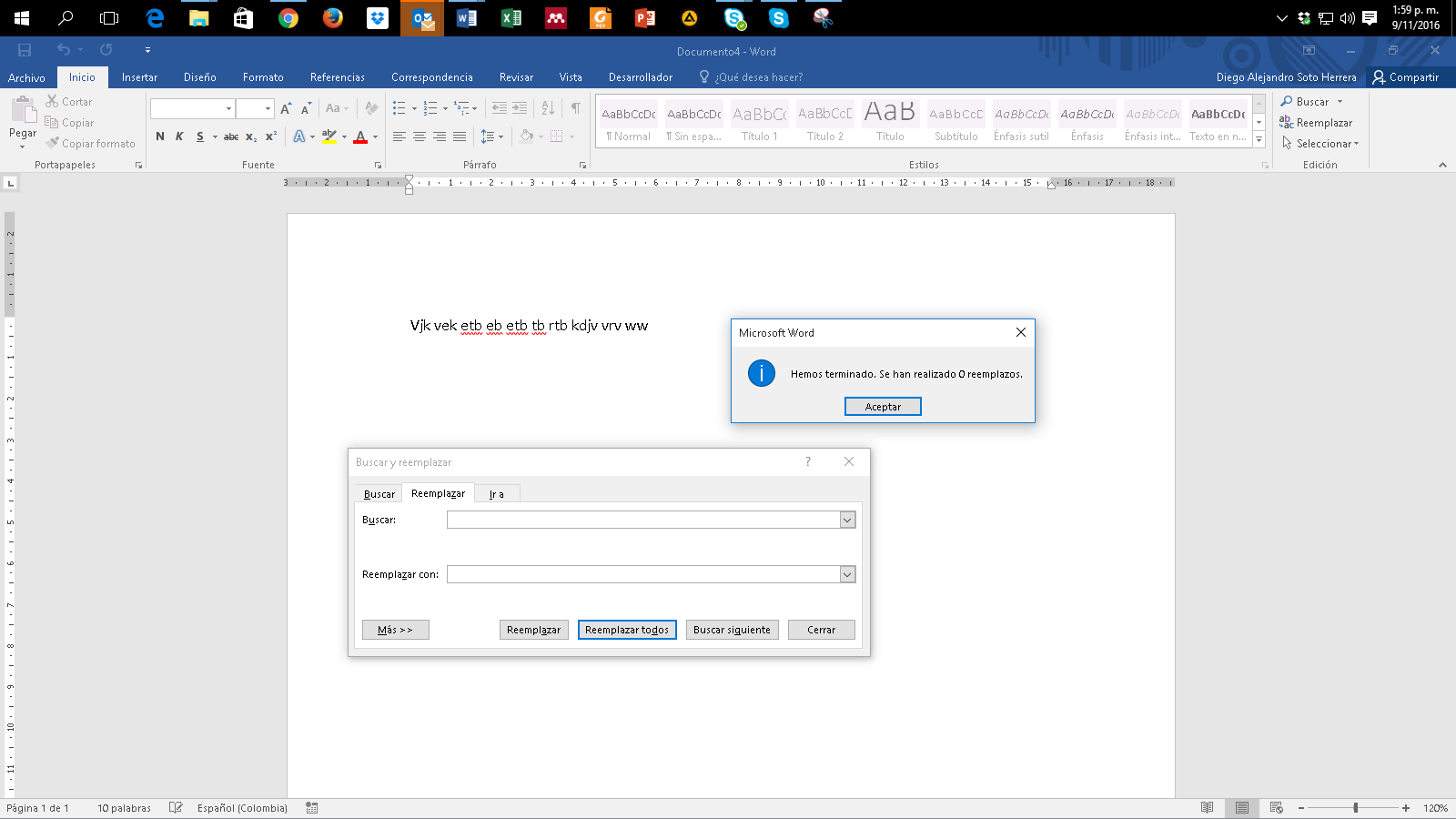
Uno de los errores más comunes al redactar un texto es incluir dobles, triples o cuádruples espacios, que en esencia son casi imperceptibles pero que afectan la distribución del documento. Para eliminar dobles espacios, realiza estos pasos: Ctrl + L (Reemplazar) > Digita 2 espacios en “Buscar” > Digita 1 espacio en “Reemplazar” > Clic en “Reemplazar todos”.



Word notificará cuántos espacios dobles se eliminaron y se han reemplazado por un espacio.



Haz clic en “Reemplazar todos” para eliminar dobles espacios que quedan, por ejemplo, de triples o cuádruples espacios anteriores, hasta que aparezca este mensaje:



## Anexo 6. Atajos de teclado útiles en Microsoft Word

**Ctrl +**

A= Abrir

B= Buscar

C= Copiar

D= Alinear derecha

E= Seleccionar todo

G= Guardar

H= Sangría

I= Ir a página

J= Justificar

K= Cursiva

L= Reemplazar

M= Formato

N= Negrilla

O= Disminuir tamaño

P= Imprimir

Q= Alinear izquierda

R= Cerrar documento

S= Subrayado

T= Centrar

U= Nuevo documento

V= Pegar

X=Cortar

Y= Rehacer

Z= Deshacer

Lista completa de atajos https://bit.ly/3oHliCj

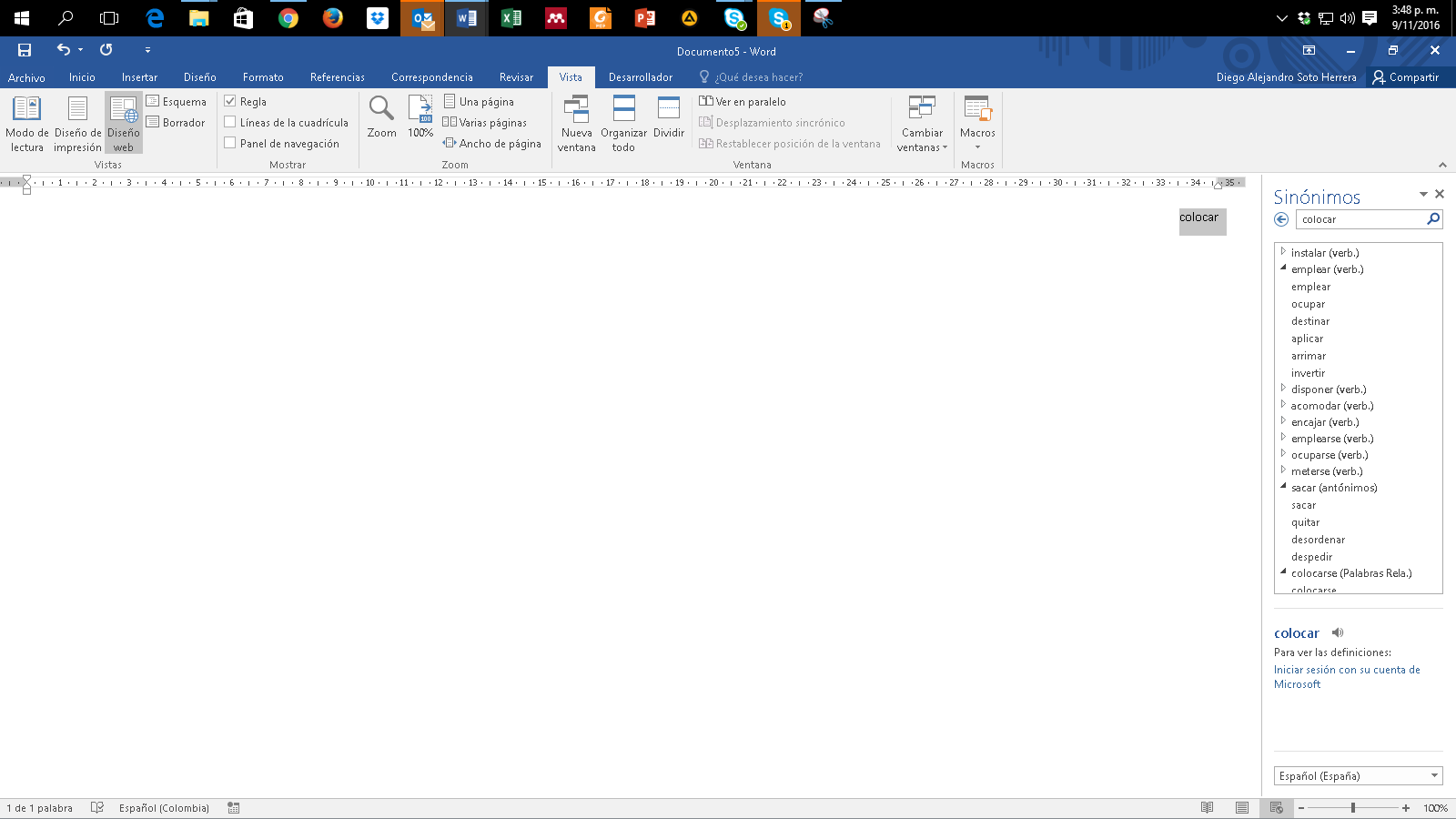
## Anexo 7. Sinónimos y antónimos

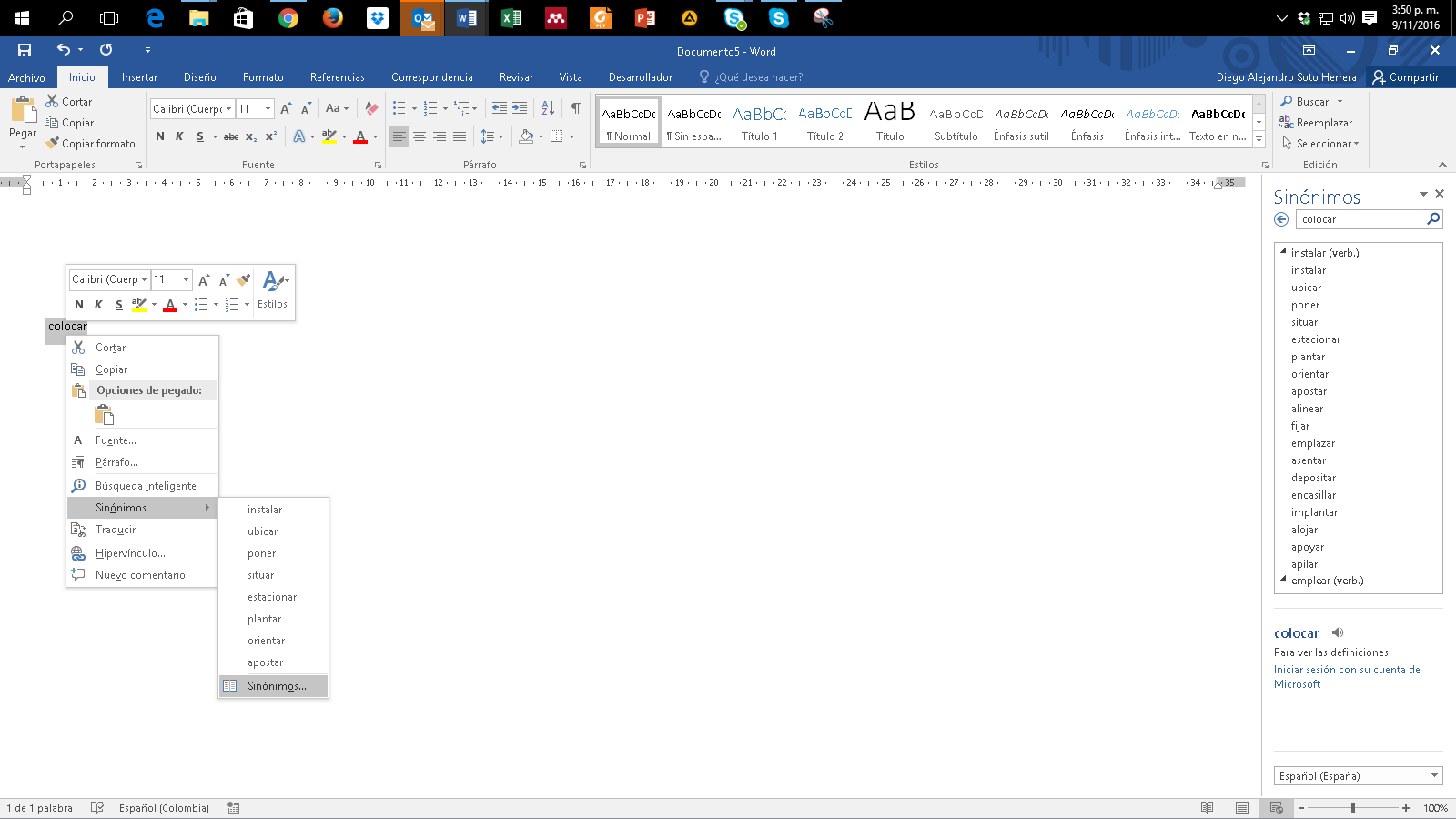
Constantemente surgen inconvenientes al redactar una oración, al no tener la palabra adecuada, un sinónimo o un antónimo. Microsoft Word apoya estas inquietudes, así:

Selecciona la palabra (en este ejemplo “colocar”) > Clic derecho > “Sinónimos”.

Inmediatamente aparecen las sugerencias más usadas. Si no es suficiente y se requieren más alternativas:

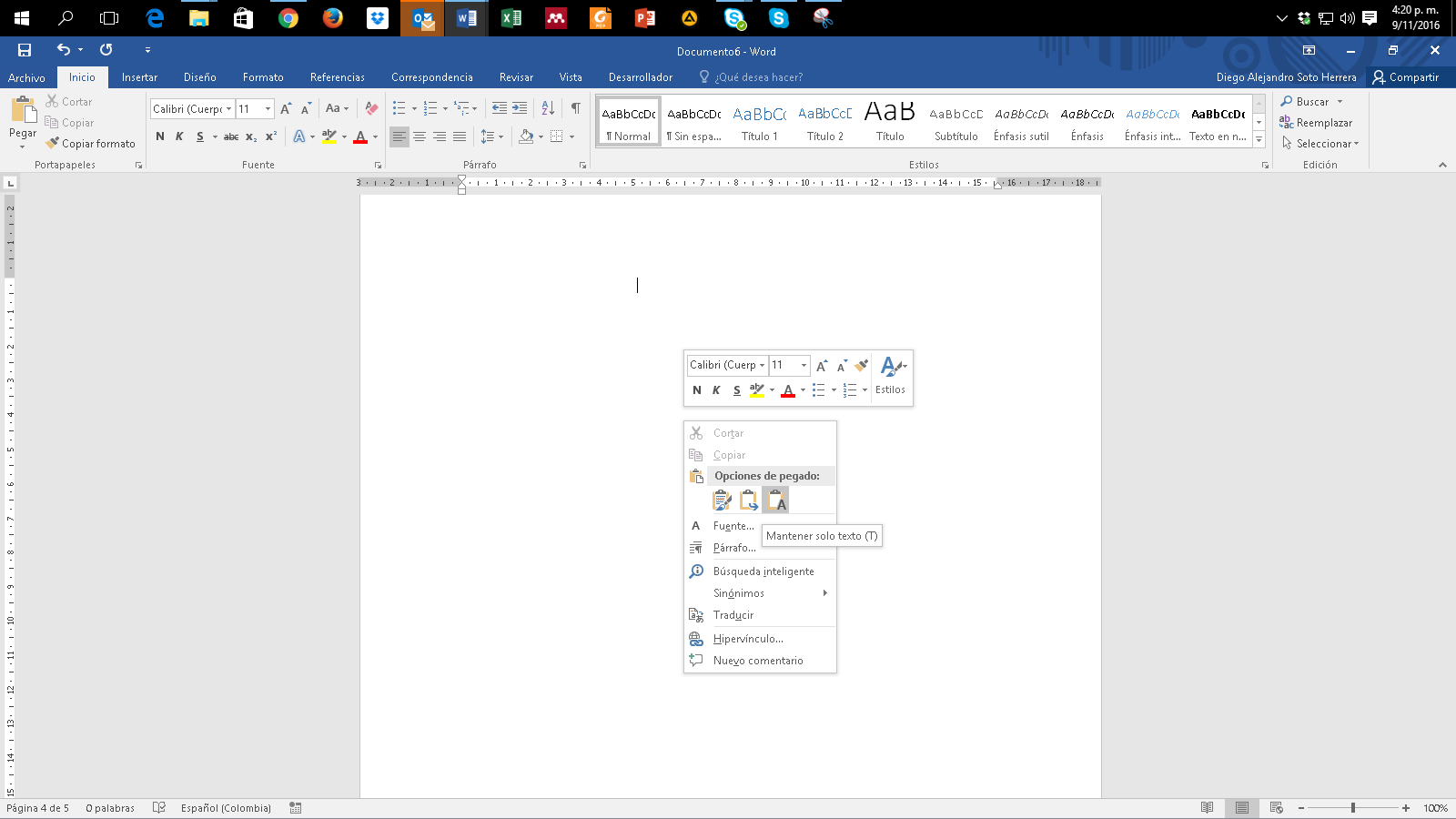
Clic de nuevo en “Sinónimos”, donde aparecen más opciones y los antónimos de esa palabra seleccionada.





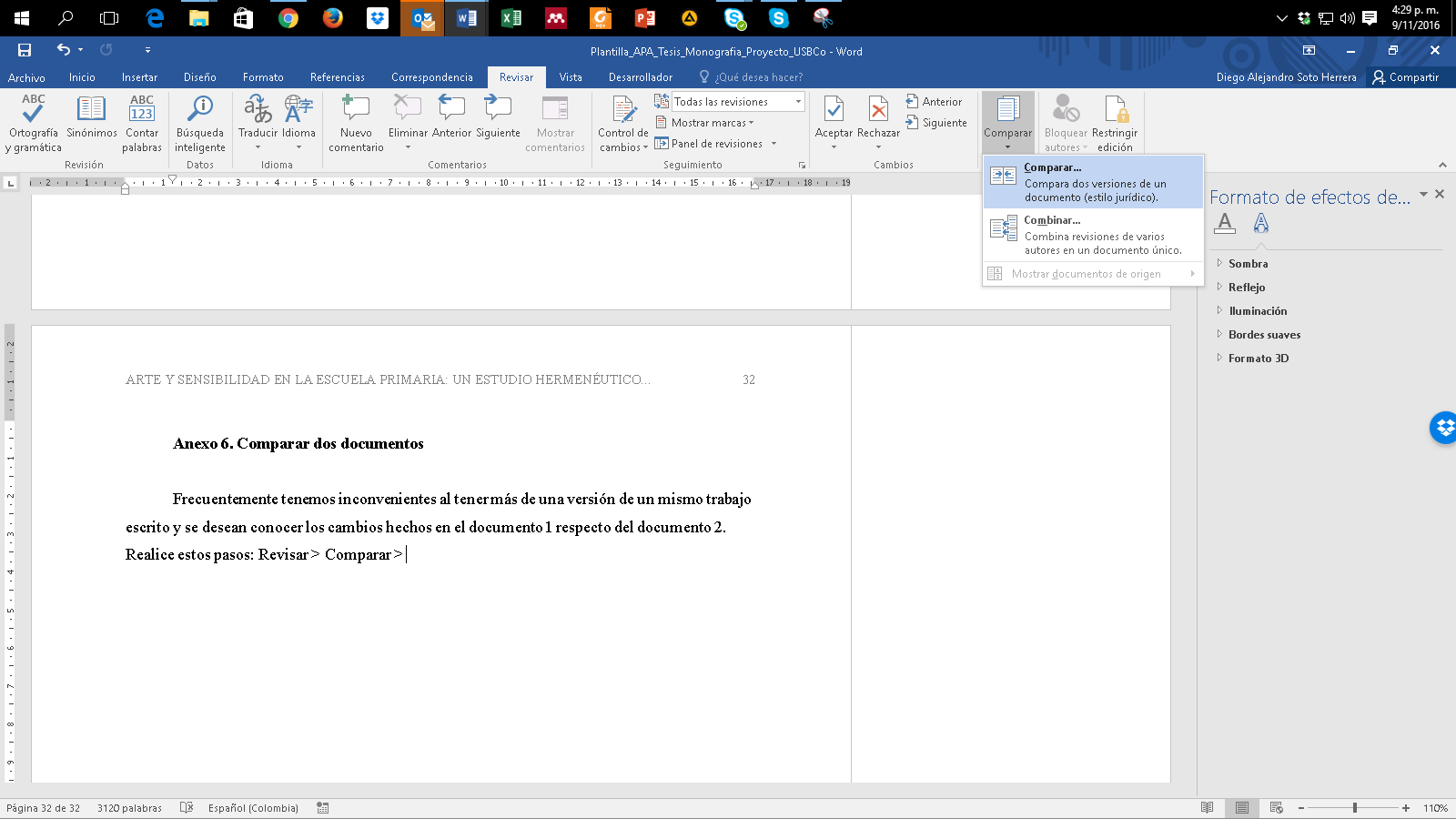
## Anexo 8. Copiar y pegar sin formato

En ocasiones copiamos y pegamos objetos o texto desde páginas web u otras fuentes hacia Word con el conocido Ctrl + C y Ctrl + V; sin embargo, se conservan colores, tipos de letras, tablas, y otros formatos indeseados. Para pegar solo el texto y sin formato alguno, clic derecho > “Mantener solo texto (T)”

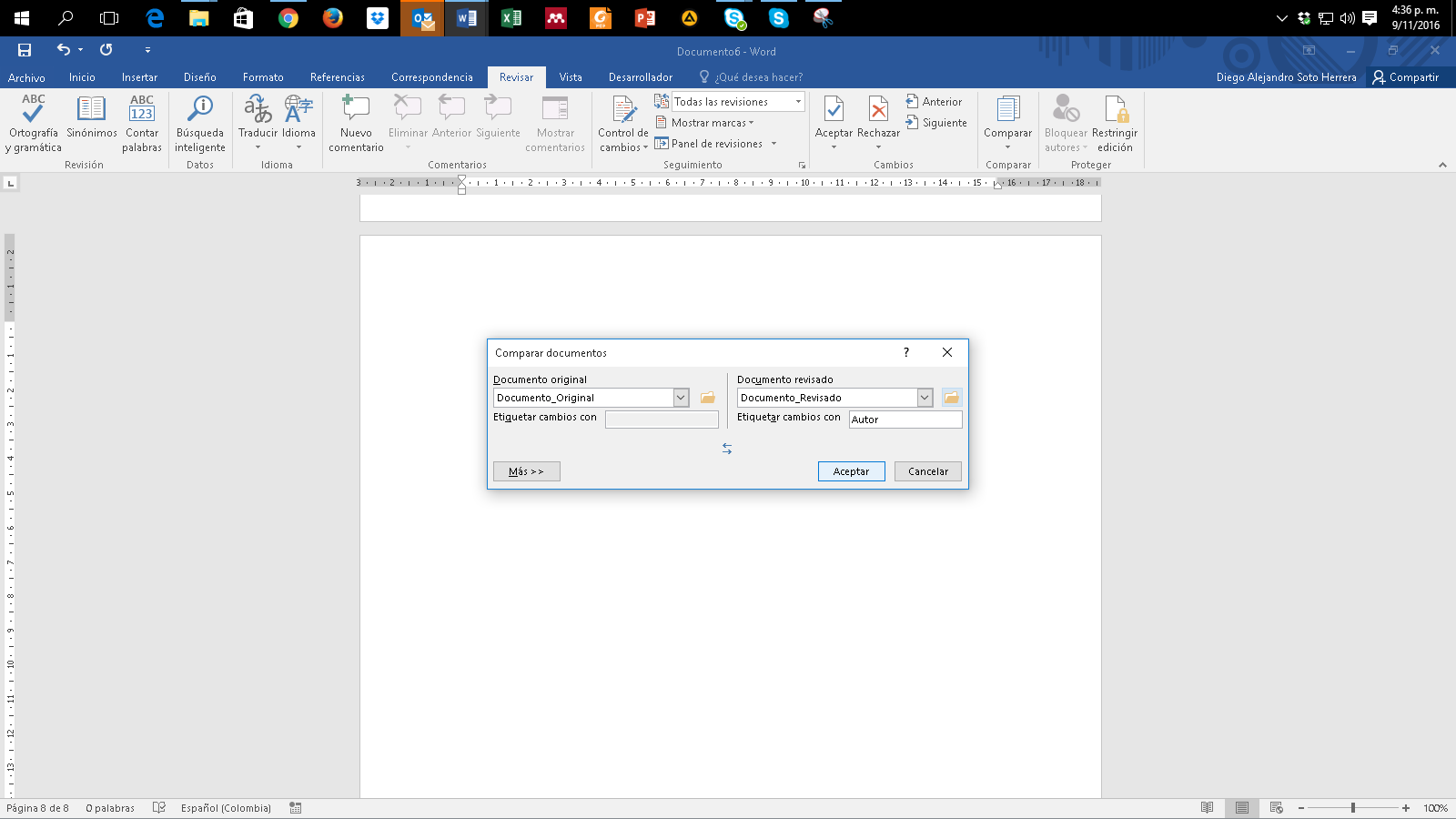


## Anexo 9. Comparar dos documentos

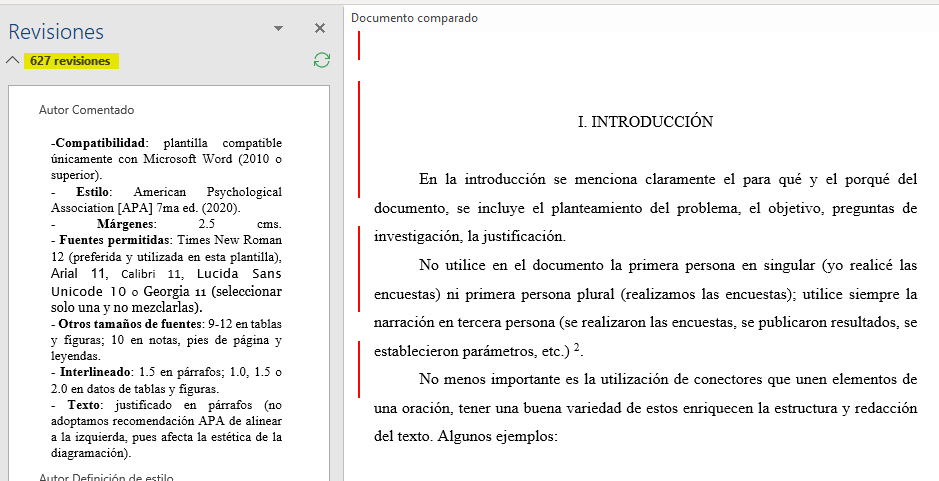
Frecuentemente tenemos inconvenientes al tener más de una versión de un mismo trabajo escrito y se desean conocer los cambios hechos en el documento 1 respecto del documento 2. Realiza estos pasos: Revisar > Comparar > Comparar... Compara dos versiones de un documento (estilo jurídico).



Busca la ruta en tu dispositivo donde se encuentra el documento original (izquierda) y luego el mismo procedimiento con el documento revisado (derecha). > clic en Aceptar.



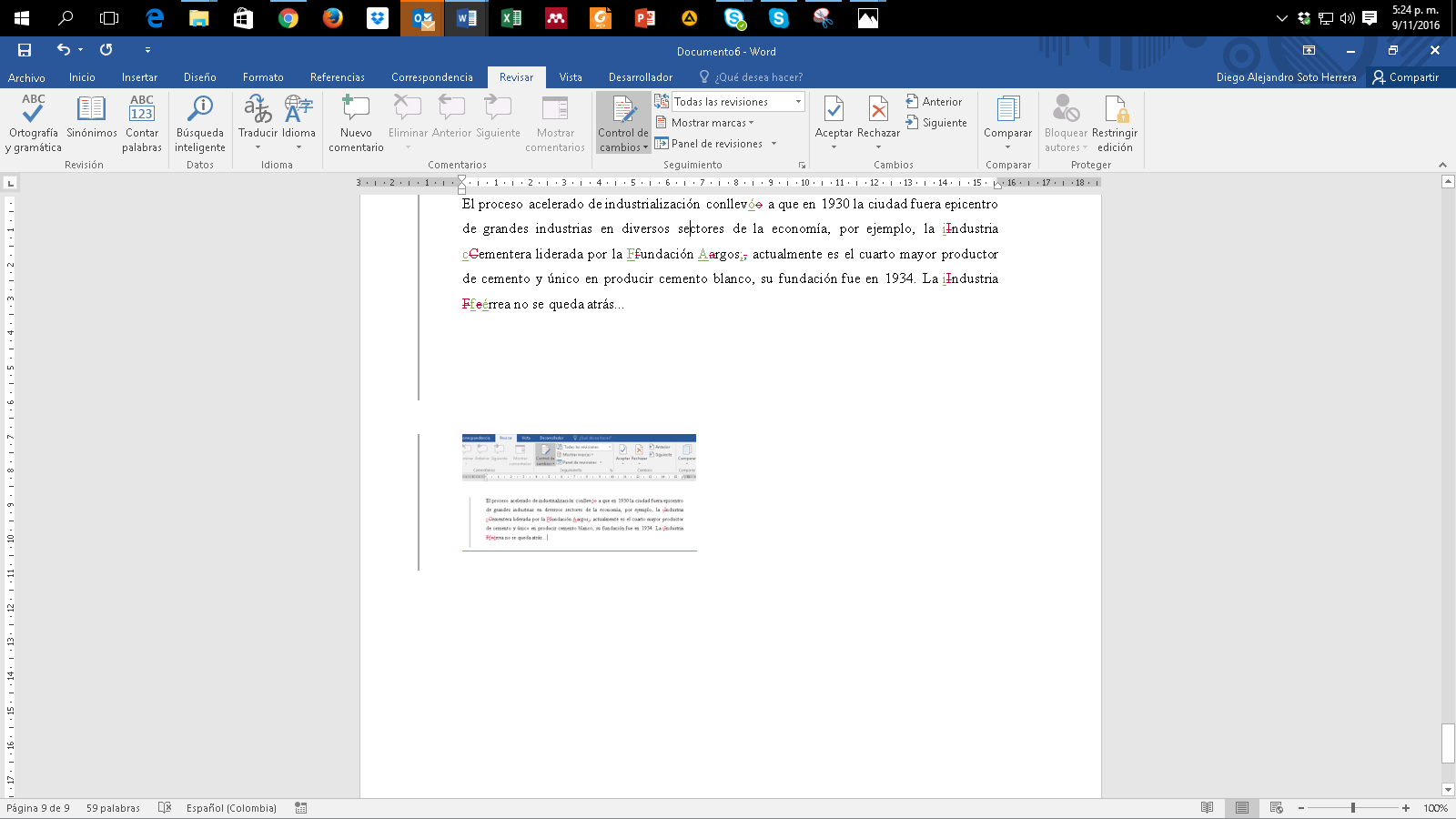
Posteriormente aparece el informe con la cantidad de revisiones hechas en el documento:



## Anexo 10. Control de cambios

Es una de las funciones más útiles, especialmente cuando se desea vigilar, revisar y aceptar cualquier cambio en un documento. Supongamos la interacción entre un estudiante que elabora la tesis y su asesor. El asesor considera que hay que hacer cambios, pero no desea modificar sin que el estudiante se entere y que, por consiguiente, acepte o rechace los cambios y aprenda de las sugerencias. Activa esta opción, así: Revisar > Control de Cambios.

**Modo asesor:** el asesor corrige los errores; es visible lo que se pretende eliminar con tachado guion medio (color rojo) y la sugerencia con guion bajo (color verde):

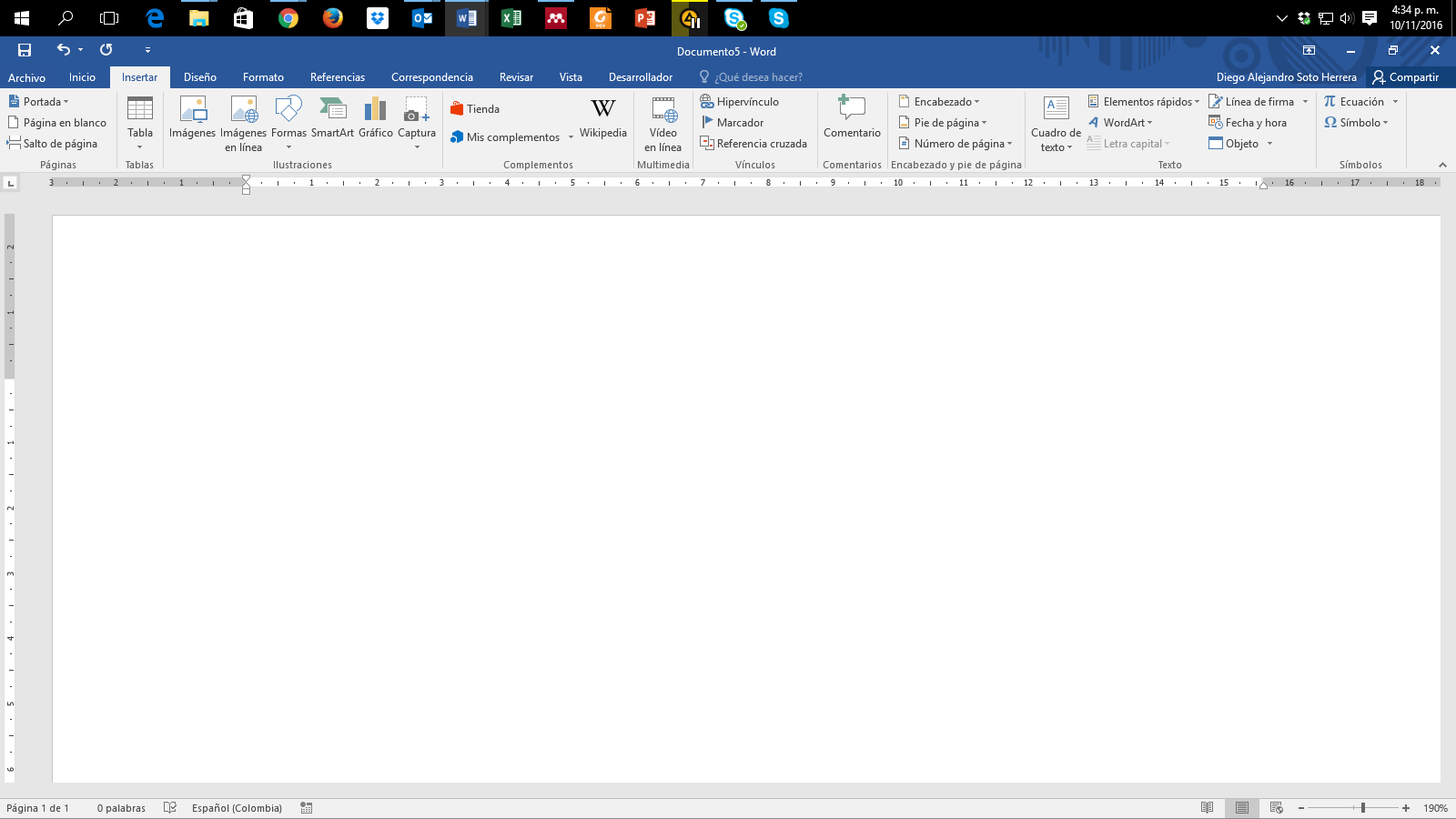


**Modo estudiante:** estudiante recibe archivo con sugerencias (el botón “Control de cambios” debe estar activo), clic en “Siguiente” y tiene la opción de “Aceptar” o “Rechazar” una a una las sugerencias visibles del asesor.



## Anexo 11. Insertar salto de página

Existe una sencilla función llamada “Salto de página” que ahorra tiempo en la estructura del texto, cuando se requiere iniciar en una nueva página en blanco, sin necesidad de insertar “Enter” una y otra vez en cada línea: Insertar > Salto de página. Su método abreviado con el teclado es: Ctrl + Enter.



## Anexo 12. Recortar y abreviar direcciones web largas

Eventualmente utilizamos páginas web, imágenes, documentos en línea, entre otros, y es necesario citarlas o mencionarlas en el texto; sin embargo, esos enlaces son supremamente largos, lo que le resta estética a la presentación del documento, ejemplo:

**Largo**: https://www.youtube.com/watch?reload=9&v=tRH59E1aybE&feature=youtu.be

**Corto**: https://bit.ly/3abhsgE

Utiliza una herramienta en línea para hacer de este enlace mucho más corto. Existe gran variedad de ellos, recomendamos algunos.

https://cutt.ly/ https://bitly.com/ https://tiny.cc/ https://tinyurl.com/

Ejemplo realizado con Bitly https://bitly.com/

Copiar y pega la URL larga en la casilla Shorten your link > Clic en Shorten > Posteriormente aparece la nueva URL corta > Clic en Copy > Pégala en el lugar del texto que la necesites.