



Inteligencia de fuentes abiertas y ciencia de datos:

Un análisis exploratorio para **detectar patrones de corrupción** en el Ministerio de Transporte e Infraestructura y el Consejo Supremo Electoral en el **periodo 2007-2022**

J U N I O 2 0 2 3

Inteligencia de fuentes abiertas y ciencia de datos

Siglas y acrónimos	3
Resumen Ejecutivo	5
1. Introducción	6
2. Marco de palabras probabilístico	8
3. Definición del problema y objetivos	9
3.1. Definición del problema: Corrupción en Nicaragua	9
3.2. Objetivo general	10
3.3. Objetivos específicos	10
4. Marco de parámetros de irregularidad	11
5. Metodología	13
5.1. Recopilación de datos	13
5.1.1. Revisión bibliográfica de Web scraping	13
5.1.2. Revisión bibliográfica sobre machine learning en los procesos de licitaciones públicas	16
5.2. Metodología de Web scraping	20
5.2.1. Modelo de relación entre entidades	20
5.3. Arquitectura de datos	22
5.4. Ejecución del Web scraping	22
5.4.1. Flujo de datos	22
5.4.2. Flujo de códigos	24
6. Análisis de datos	30
7. Discusión de los resultados	35
7.1. Ministerio de Transporte e Infraestructura (MTI)	35
7.2. Consejo Supremo Electoral (CSE)	41
8. Conclusiones y desafíos	47
Referencias	49
Anexos	52

SIGLAS Y ACRÓNIMOS

AED	Análisis Exploratorio de Datos
ALMA	Alcaldía de Managua
API	Application Programming Interface
BCIE	Banco Centroamericano de Integración Económica
BCN	Banco Central de Nicaragua
BI	Business Intelligence
CGR	Contraloría General de la República
CIDH	Comisión Interamericana de Derechos Humanos
CONAVI	Consejo Nacional de Vialidad
CSE	Consejo Supremo Electoral
DAG	Datos Abiertos Gubernamentales
DOM	Document Object Model
ERD	Entity Relation Diagram
FOMAV	Fondo de Mantenimiento Vial
GIEI	Grupo Interdisciplinario de Expertos Independientes
HTTP	Hypertext Transfer Protocol
IA	Inteligencia Artificial
IE	Information Engineering
IPC	Índice de Percepción de la Corrupción
ML	Machine Learning
MTI	Ministerio de Transporte e Infraestructura
NIC	National Intelligence Council
OACNUDH	Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos
OCDE	Organización para la Cooperación y el Desarrollo Económico
ODS	Objetivos de Desarrollo Sostenible
OIJ	Organismo de Investigación Judicial
OSC	Organizaciones de la Sociedad Civil
PAC	Planes Anuales de Contratación
PCA	Principal Component Analysis

Inteligencia de fuentes abiertas y ciencia de datos

PGR	Presupuesto General de la República
PIP	Programa de Inversión Pública
RUC	Registro Único de Contribuyente
SABLE	Scraping Assisted by Learning
TI	Transparencia Internacional
UE	Unión Europea
V-DEM	Varieties of Democracy
WEP	Words of Estimative Probability

RESUMEN EJECUTIVO

Nicaragua ha sido calificada negativamente en varios índices de corrupción debido a la presencia de un gobierno considerado como “dictatorial”. Investigaciones realizadas por medios de comunicación disidentes han revelado la utilización del sector público como medio para obtener beneficios privados. Con el objetivo de profundizar en este problema, la presente investigación realiza un análisis empírico exploratorio basado en evidencias con el fin de identificar patrones de corrupción en el Ministerio de Transporte e Infraestructura (MTI) y en el Consejo Supremo Electoral (CSE) desde 2007 hasta 2022.

En el caso del MTI, se destacan los procedimientos de pocos días y un alto volumen agregado, como la “Compra por Cotización”. Estos procesos tienen un promedio de 4 días desde la fecha de competencia hasta el cierre, con contratos promedio de C\$2.3 millones. Resulta interesante observar la asignación de estos montos en tan poco tiempo. Por otro lado, la “Contratación Directa” tiene un monto de C\$1.5 billones, pero el tiempo promedio no supera un día. Cada contrato tiene un valor promedio de C\$13.2 millones, lo cual es elevado para ser deliberado en un solo día.

En cuanto a las empresas ejecutoras de proyectos para dicha institución, los resultados indican que las empresas MECO, Astaldi y Eterna presentan valores atípicos en el volumen de contratos y montos adjudicados. Lo anterior podría estar relacionado a la cercanía entre dichas empresas y el gobierno nicaragüense. Llama la atención que las tres empresas antes mencionadas han estado implicadas en casos de corrupción en diferentes países de América Latina. Por ejemplo, la empresa MECO ha estado involucrada en casos de corrupción en Panamá y Costa Rica, la empresa Astaldi en El Salvador y Perú, y la empresa Eterna en Honduras.

Por su parte, el CSE muestra diferencias significativas en comparación con el MTI en términos de montos asignados y distribución de datos. El total adjudicado por el CSE es de C\$404.0 millones, en contraste con los C\$57 billones del MTI, lo que implica una diferencia de aproximadamente el 99.3%. Las contrataciones simplificadas representan el 74.2% del monto total adjudicado por el CSE, lo que indicaría tiempos de

licitación cortos. Esto indica que el CSE asignó un 40.8% más de fondos que el MTI a través de este tipo de contrataciones. Estas prácticas plantean dudas sobre su adecuada justificación y cumplimiento de la ley.

Algunos de los proveedores destacados son Interlector Nicaragua S.A, Mega Comunicaciones S.A, Consultores Especialistas en Sistema S.A y Suministro de Informática Quiñonez. Sin embargo, se han planteado críticas y sospechas sobre las relaciones entre estos proveedores y el CSE debido a prácticas poco transparentes. Además, el CSE ha realizado gastos innecesarios y poco claros en diversas áreas, por ejemplo, la compra de vehículos de lujo. También se ha observado una subdivisión de contratos para evadir controles y límites de gasto público. Considerando los hallazgos tanto para el MTI como para el CSE, resulta fundamental abordar estas prácticas y promover la transparencia en el uso de los recursos públicos.

Palabras clave

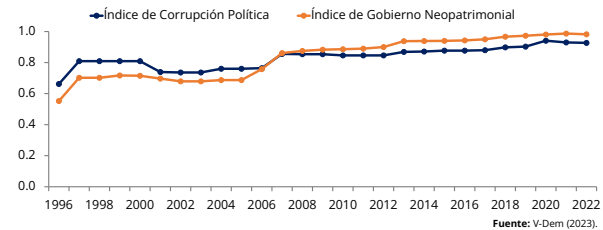
Ciencia de datos, Corrupción, Nicaragua, Web Scraping, Unsupervised Machine Learning, Clustering.

INTRODUCCION

Diversos indicadores encargados de evaluar los niveles de transparencia o corrupción de un país evalúan negativamente a Nicaragua. Entre estos está el Índice de Percepción de la Corrupción (IPC) elaborado por Transparencia Internacional (TI), el cual se centra en analizar la corrupción en el sector público definiéndola como el abuso del cargo público en beneficio privado. Las encuestas utilizadas para compilar el IPC suelen formular preguntas relacionadas con el mal uso del poder público para obtener beneficios privados, enfocándose, por ejemplo, en el soborno de funcionarios del Estado en contrataciones públicas.

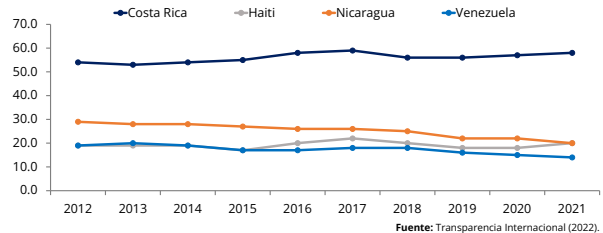
También se destaca el Índice de Gobierno Neopatrimonial y el Índice de Corrupción Política, ambos elaborados por Variedades de la Democracia (V-Dem). El Índice de Gobierno Neopatrimonial tiene como objetivo evaluar la dependencia de un régimen hacia un individuo, considerando aspectos como el nepotismo y la utilización de recursos públicos para legitimar su posición política. Por otro lado, el Índice de Corrupción Política se centra en la prevalencia de la corrupción en diversas esferas políticas, como el poder ejecutivo, legislativo y judicial; además, engloba distintas formas de corrupción, que van desde el frecuente abuso de poder por parte de funcionarios hasta prácticas elitistas que perjudican a la sociedad, el soborno y la manipulación de la legislación. En el Gráfico 1 se observa la evolución del país en los índices antes mencionados elaborados desde 1996 hasta la fecha¹.

GRÁFICO 1.
Evolución de índices en Nicaragua
Puntaje de 0 a 1



Con respecto al Índice de Gobierno Neopatrimonial y al Índice de Corrupción Política se puede observar que el ascenso hacia el punto máximo empieza en 2007, año en el que da inicio la administración del presidente Ortega; es decir, desde ese año se observa una mayor incidencia de la corrupción. Por su parte, en el IPC, Nicaragua obtiene una calificación de 19 puntos (en un máximo de 100²), mostrando un descenso de 09 puntos entre 2012 y 2021³. Lo anterior sitúa al país en el tercer peor puesto en América Latina, sólo superado por Haití y Venezuela (Gráfico 2). A nivel global, Nicaragua ocupa el puesto 167 de 180.

GRÁFICO 2.
Evolución de Índice de Percepción de la Corrupción
Puntaje de 0-100



Ante la presencia de un Estado cooptado que se niega a investigar y sancionar los casos de corrupción y que, por el contrario, es instrumentalizado para operacionalizar la corrupción, diversos actores de la sociedad civil, principalmente medios de comunicación, han intentado investigar dicho fenómeno. Algunos medios como Nicaragua Investiga han descubierto

1 Se destaca que en estos gráficos la direccionalidad es opuesta a la de otros índices V-Dem, que generalmente van de normativamente peor a mejor. Es decir, en estos índices un valor más cercano a 01 indica un peor desempeño.

2 En donde el valor más cercano a 0 indica mayor percepción de corrupción y el valor más cercano a 100 indica menor percepción.

3 Se incluye Costa Rica como referencia.

Un análisis exploratorio para detectar patrones de corrupción

que comunicadores que hacen apología del delito son financiados con las arcas del Estado (Garmendia, 2020). Por su parte, Divergentes realizó una investigación en la que descubrieron patrones de adquisiciones frecuentes por parte de la Alcaldía de Managua (ALMA) (Salazar, 2020). En este ejercicio pudieron nombrar a los principales beneficiarios y sus vínculos con agentes del Gobierno.

Considerando las iniciativas anteriores, en el presente informe se pretende estudiar la corrupción que se produce en el ámbito de la contratación pública del Estado. Para lograr dicho objetivo, se aplican técnicas innovadoras de extracción, minería, ingeniería, análisis y visualización de datos que permiten detectar posibles patrones de corrupción.

El informe se encuentra estructurado en 08 secciones. La primera sección corresponde a la presente introducción. En la segunda sección se aborda el marco de palabras probabilístico. Posteriormente, en la tercera sección se define el problema de estudio y los objetivos. En la cuarta sección se aborda el marco de parámetros de irregularidad. La metodología utilizada se aborda en la quinta sección. En la sexta sección se realiza el análisis de datos, y luego en la séptima sección se presenta una discusión sobre los resultados del estudio. Finalmente, en la octava sección se abordan las conclusiones y desafíos que se desprenden del estudio. Posteriormente se presentan las referencias y anexos.

A lo largo del texto se utilizan Palabras de Probabilidad Estimativa (WEP, por sus siglas en inglés), cuyo uso es convencional en la comunidad de inteligencia de datos. En este estudio, las WEP consisten en un marco de palabras para estimar con cierto grado de probabilidad lo que se considera como pautas de corrupción, es decir, algo que debe investigarse a mayor profundidad por generar sospechas. Kent (1993) introdujo por primera vez los WEP para cuantificar los juicios cualitativos¹ presentando las probabilidades matemáticas equivalentes a las expresiones verbales de probabilidad.

El Consejo Nacional de Inteligencia de Estados Unidos (NIC, por sus siglas en inglés) establece que, para las expresiones de verosimilitud o probabilidad, un producto analítico debe utilizar uno de los siguientes conjuntos de términos (Office of the Director of National Intelligence, 2015, p. 3):

Tabla 1. Palabras de probabilidad estimativa

Casi ninguna posibilidad	Muy improbable	Insólito	Posibilidades más o menos iguales	Probable	Muy probable	Casi seguro
Remoto	Altamente improbable	Improbable	Probabilidades casi iguales	Probable	Muy probable	Casi seguro
01-05%	05-20%	20-45%	45-55%	55-80%	80-95%	95-99%

Fuente: Elaboración propia con información de "Analytic Standards" (2015) por Office of the Director of National Intelligence.

El uso de este marco probabilístico debe entenderse como una interpretación de los autores del presente informe. La narrativa de este texto, sus hechos y análisis están sustentados en entrevistas publicadas en medios de comunicación masiva, así como documentos, informes y libros disponibles al público. En el texto se mencionan nombres de personas naturales y jurídicas que han estado involucradas por error o a propósito, acusadas pero no condenadas, o condenadas y sobreesidas en casos de corrupción. En vista de lo anterior se les preserva la presunción de inocencia en observancia de los derechos individuales mientras las cortes judiciales correspondientes no demuestren lo contrario.

¹ Un juicio cualitativo es una evaluación o valoración subjetiva basada en características no cuantificables o en juicios de calidad.

3.1 Definición del problema: Corrupción en Nicaragua

Según TI (2022), Nicaragua experimentó un descenso de 09 puntos en el IPC desde el 2012 hasta el 2021. Actualmente es el tercer país con peor calificación del continente americano en materia de corrupción, sólo por delante de Venezuela y Haití. La clasificación de Nicaragua se ha reducido debido a la falta de libertad de expresión, asociación, y el limitado acceso a la justicia. El presidente Daniel Ortega ha respondido a las acusaciones de corrupción con medidas enérgicas contra los medios de comunicación, con la reducción del espacio cívico y con la captura de las instituciones de control y fiscalización de los recursos públicos, entre ellas la Contraloría General de la República (CGR).

Los hechos anteriores generan un círculo vicioso en donde el deterioro de la vida democrática facilita la corrupción y esto, a su vez, facilita el uso de las herramientas anticorrupción en la criminalización de la oposición política. El Observatorio Ciudadano de Corrupción (2022, p. 22) señala que la presión internacional y los informes de organizaciones como el Grupo Interdisciplinario de Expertos Independientes (GIEI), Comisión Interamericana de Derechos Humanos (CIDH), Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (OACNUDH) y Amnistía Internacional no son suficientes para detener la estrategia del gobierno de Nicaragua de reprimir a la oposición política. A pesar de las regulaciones, políticas y procedimientos dirigidos a prevenir la corrupción en las obras públicas y la cooperación jurídica internacional, persiste el secretismo en su aplicación y se utilizan para criminalizar a la oposición. Algunos ejemplos de ello es la cancelación masiva de la personería jurídica de Organizaciones de la Sociedad Civil (OSC), y las acusaciones de corrupción y lavado de dinero para justificar el encarcelamiento de líderes de oposición, exaspirantes a la presidencia, y periodistas independientes (AFP, 2021).

Esta tendencia de deterioro de la transparencia también se refleja en la opinión de la ciudadanía. De acuerdo con la Corporación Latinobarómetro (2021, p. 84), el presidente Ortega y sus colaboradores fueron percibidos como el principal grupo social y político involucrado en la corrupción en Nicaragua, con un 58% de las personas

encuestadas. Le sigue la Policía con un 46%, mientras que los jueces, los funcionarios públicos y el gobierno local reciben un 32%, 27% y 24%, respectivamente. Otros grupos percibidos como implicados en la corrupción son los empresarios, los diputados y el personal de Hacienda, con un 24%, 24% y 22%, respectivamente. El informe Pulse of Democracy 2021 publicado por la Universidad de Vanderbilt en mayo de 2022 mostró que el 22% de las personas encuestadas en Nicaragua informó que un funcionario público les había pedido o habían tenido que pagar un soborno en los últimos 12 meses, mientras que el 19% informó de lo mismo por parte de un agente de policía (Rodríguez & Zechmeister, 2021).

Más concretamente, el medio independiente Expediente Público encontró falta de transparencia en el proceso de contratación pública de Nicaragua entre 2017 y 2019 (Valle, 2020). Durante este tiempo, hubo una disminución en el número de proyectos realizados por el Fondo de Mantenimiento Vial (FOMAV) y un aumento en el número de proyectos ejecutados por el MTI financiados con recursos externos¹, a pesar de que los datos del Presupuesto General de la República (PGR) muestran que hubo fondos internos para estos proyectos.

Expediente Público en su estudio determinó que la falta de transparencia viola lo establecido en la Ley No. 621, Ley de Acceso a la Información Pública. Además, entre sus hallazgos destaca prácticas que rompen los principios de libre competencia, igualdad, transparencia, publicidad e integridad del sistema de adquisiciones. En particular, señala que la subdivisión de proyectos para licitar en cuantías menores no es legal porque se puede estar eludiendo o se pretende eludir los procedimientos de contrataciones establecidos en la Ley No. 737, Ley de Contrataciones Administrativas del Sector Público, lo que va en contra de los principios de legalidad, transparencia y publicidad. Estas divisiones reducen el poder de negociación del Estado, generan más procedimientos, pero con menos rigor y crean una sobrecarga administrativa

1 Los Planes Anuales de Contratación (PAC) de 2017 a 2019 indican que todos los proyectos se realizaron con fuentes externas.

Inteligencia de fuentes abiertas y ciencia de datos

en las áreas de contratación. Asimismo, aumenta el abuso de la contratación simplificada y la existencia de redes de corrupción para adjudicar contratos de obras públicas.

Además, la subdivisión de contrataciones puede estar direccionando la compra hacia un proveedor favorecido en relación con otros, lo que viola el principio de libre competencia o igualdad de condiciones en el proceso de búsqueda de adquisiciones de bienes, servicios u obras por parte del Estado. El estudio también encontró discrepancias entre los importes previstos para estos proyectos, reflejados en sus PAC de 2017 a 2019 y los importes realmente ejecutados, reflejados en los informes del Programa de Inversión Pública (PIP) del MTI. Por ejemplo, los importes reportados en el PIP son inferiores a los del PAC, especialmente en 2017 y 2019. Las diferencias en los montos pueden indicar falta de transparencia y posible corrupción, así como problemas con la evaluación técnica. El estudio también encontró que los problemas pueden provenir de la falta de capacidad profesional del personal de las instituciones, que no son elegidos por mérito sino por favoritismo político.

Este informe también muestra que el uso de metodologías como las encuestas de percepción, la revisión manual de contratos, la consolidación de índices de terceros y la evaluación de indicadores cualitativos, si bien son sumamente valiosas para evaluar la situación general de la corrupción en Nicaragua, no son capaces de medirla ni de encontrar sistemáticamente nuevos comportamientos sospechosos. Es en este punto donde este proyecto hace una nueva contribución, al utilizar metodologías de Inteligencia Artificial (IA) para analizar información a una escala que investigadores humanos individuales no podrían lograr, utilizando datos duros que no dependen de casos de corrupción ya reportados o descubriendo vínculos desconocidos por los expertos actuales. Esta es también una forma de procesar los datos para hacerlos más accesibles al público e ir más allá de las afirmaciones generales sobre la propagación de la corrupción o el nepotismo en un país con autoridades poco cooperativas, sino que también puede generar información que pueda utilizarse en eventuales procesos de rendición de cuentas, transparencia y democratización.

3.2 Objetivo general

Realizar una investigación exploratoria utilizando técnicas de minería de datos y machine learning para detectar patrones de corrupción en los procesos de licitación pública del MTI y del CSE a partir de una base de datos propia de procesos de licitación que comprende el periodo entre 2007 y 2022.

3.3 Objetivos específicos

- Realizar un web scraping del sitio web de licitaciones públicas www.gestion.nicaraguacompra.gob.ni.
- Estructurar los datos extraídos en una base de datos.
- Aplicar técnicas de minería de datos y machine learning a los datos extraídos para detectar posibles patrones de corrupción en las licitaciones de estas instituciones.

La Ley No.737, Ley de Contrataciones del Sector Público¹ tiene por objeto regular y uniformar el proceso de contratación en el sector público para garantizar la transparencia, competencia y equidad. Esta ley establece normas y procedimientos para la adquisición de bienes, servicios y obras por parte de las entidades públicas, incluyendo el uso de plataformas electrónicas, la selección de proveedores y la gestión de contratos. El Arto. No. 06 de dicha ley esboza los principios que rigen la contratación pública que incluyen:

- Eficiencia: los organismos y entidades del sector público deben planificar, programar, organizar, desarrollar y supervisar sus actividades de contratación en tiempo y forma y en las mejores condiciones de coste y calidad.
- Publicidad: las organizaciones y entidades del sector público deben garantizar la transparencia de sus actividades administrativas, dando a conocer los procedimientos de contratación y permitiendo a los ciudadanos el acceso a la información relacionada con estos procesos.
- Transparencia: todas las actividades de contratación deben basarse en criterios objetivos y evaluaciones de las ofertas, tal y como se define en la definición de "mejor oferta" del artículo 2 de esta ley. El acceso a la información sólo puede restringirse cuando la información solicitada pueda dar ventaja a una oferta sobre otras o cuando los documentos hayan sido definidos como confidenciales en los respectivos términos y condiciones.
- Igualdad y libre competencia: los procedimientos para las actividades de contratación deben incluir regulaciones o tratamientos que promuevan la más amplia, objetiva e imparcial competencia, pluralismo y participación de los potenciales proveedores. Cualquier interesado que reúna

los requisitos legales y reglamentarios podrá participar en las actividades de contratación en igualdad de condiciones y sin restricción alguna no derivada de especificaciones técnicas y objetivos inherentes al objeto del concurso.

- Viabilidad Tecnológica: los bienes, servicios u obras deberán reunir las condiciones de calidad y modernidad tecnológica necesarias para satisfacer eficazmente los fines para los que se requieran, desde el momento en que se adquieran o contraten, y por un plazo razonable, con la posibilidad de ser adaptados, integrados y actualizados, con los avances científicos y tecnológicos.
- Control: las actividades de contratación reguladas por esta ley serán fiscalizadas por la CGR de conformidad con la Ley No. 681, "Ley Orgánica de la Contraloría General de la República y del Sistema de Control de la Administración Pública y de Regulación de los Bienes y Recursos del Estado"² y el control judicial será realizado por el Poder Judicial.
- Principio del Debido Proceso: todas las personas, naturales o jurídicas, que participen en el proceso de contratación lo harán en igualdad de condiciones, contarán con el tiempo y los medios necesarios para su defensa y podrán presentar los recursos y solicitudes que establece esta ley.
- Principio de Integridad: los actos relacionados con las actividades de contratación pública deben caracterizarse por la honestidad, veracidad, inviolabilidad, justicia, imparcialidad y probidad. El sector público y los proveedores deben observar normas éticas y evitar prácticas corruptas y fraudulentas en las actividades de contratación.

1 Fue aprobada el 19 de octubre de 2010 y publicada en el diario oficial los días 8 y 9 de noviembre de 2010.

2 Fue publicada en La Gaceta, Diario Oficial No. 113 del 18 de junio de 2009.

Tabla 2. Operatividad de principios

Principio	Variable operativa	Método de medición
Eficiencia	Plazo de adjudicación de los contratos	Número de días desde la publicación del anuncio de licitación hasta la adjudicación del contrato
Transparencia	Publicación de información sobre contratación pública	Número de anuncios de licitación y resultados publicados en el sitio web oficial
Igualdad y libre competencia	Número de licitadores por proceso de contratación	Número de licitadores por proceso de contratación y relación entre licitadores y contratos adjudicados

Fuente: Elaboración propia con base a la Ley 737

Además de estos principios, hay otros artículos de la ley que son fundamentales en este análisis. En concreto el Art. No. 25. sobre la Prohibición de Subdividir los Contratos establece que:

El objeto de la contratación o la ejecución de un proyecto no podrán ser subdivididos en cuantías menores, de forma que, mediante la celebración de varios contratos, se eludan o se pretenda eludir los procedimientos de contrataciones establecidos en esta Ley para el caso.

Asimismo, el Art. No. 27 sobre Procedimientos de contratación, plantea que en función de la cuantía del contrato o de las circunstancias especiales de carácter contractual previstas en la ley, los contratos del sector público se celebrarán mediante uno de los siguientes procedimientos: i) licitación pública (para contratos superiores a C\$3,000,000); ii) licitación selectiva (para contratos entre C\$500,000 y C\$3,000,000); iii) contratación simplificada (para situaciones específicas descritas en la ley); y iv) contratos menores (para contratos inferiores a C\$500,000). Además, existe un procedimiento independiente para seleccionar empresas de consultoría, independientemente del valor del contrato. El artículo también establece que cualquier desviación de estos procedimientos dará lugar a la nulidad del contrato y que debe darse especial consideración a las pequeñas y medianas empresas, de conformidad con los acuerdos comerciales de Nicaragua. Los elementos más relevantes de los distintos procedimientos pueden resumirse en la siguiente tabla.

Tabla 3. Elementos de análisis

Procedimiento	Cuantía del contrato	Circunstancias especiales
Licitación pública	Más de C\$3,000,000	N/A
Licitación selectiva	Más de C\$500,000 y hasta C\$3,000,000.	N/A
Contratación simplificada	Independientemente de la cantidad	Situaciones taxativamente señaladas por la Ley
Adquisiciones menores	Por un importe no superior a C\$500,000	N/A
Concurso para la selección de consultores	Independientemente de la cantidad	N/A

Fuente: Elaboración propia con base a la Ley 737

5.1 Recopilación de datos

5.1.1 Revisión bibliográfica de Web scraping

El *web scraping* puede definirse como un proceso muy específico independiente de otros como el rastreo web, o como un término paraguas que contiene esas otras definiciones. Matta et al (2020), siguiendo a Diouf et al (2019), definen el objetivo principal del *scraping* como "extraer información de uno o muchos sitios web y procesarla en estructuras simples como hojas de cálculo, bases de datos o archivos CSV" (p. 6041). Además, para Matta et al (2020), el *web scraping* puede realizarse "tanto físicamente como a través de software que provoca la navegación web humana para recopilar información específica" (p. 8206).

Otros autores lo circunscriben a sitios web individuales, presumiblemente describiendo la tarea de rastreo web como un proceso distinto, a saber, "el web scraping es el conjunto de técnicas utilizadas para obtener automáticamente alguna información de un *sitio web* [énfasis agregado] en lugar de copiarla manualmente" (Vargiu & Urru, 2013, p. 44). Al incluir la automatización como parte de la definición, estos autores excluyen la posibilidad del *scraping* físico sugerido por Matta et al.

En una línea similar, Landers et al. (2016) sostienen que "el web scraping (...) es el proceso de recuperar o raspar datos de un *sitio web* [énfasis agregado]" (2016, p. 475), también separan el "proceso mundano y aburrido de extraer datos manualmente", que "utiliza la *automatización inteligente* [énfasis agregado] para recuperar cientos, millones o incluso miles de millones de puntos de datos" (2016, p. 475). En particular, Mitchell (2018) postula que el *web scraping* elimina las barreras entre los datos comprensibles por humanos y los legibles por máquinas, creando una infinidad de oportunidades para la investigación, al extraer automáticamente datos legibles por humanos que ningún humano podría extraer y transformarlos en un lenguaje que las máquinas puedan leer; por lo tanto, Mitchell excluye las Interfaces de Programación de Aplicaciones (API, por sus siglas en inglés) y la extracción manual de su definición de *scraping*, incluyendo el rastreo (o *crawling*) como un elemento anexo inseparable (2018, pp. x–xi).

En esta línea, Khder separa la extracción de datos mediante API del *web scraping*, pero también considera que "el web scraping o web crawling se refiere al procedimiento de extracción automática de datos de sitios web mediante software" (2021, p. 1), por lo que confunde ambos conceptos. Siguiendo a Plamen Milev, Khder (2021, p. 1) sostienen que "el análisis de sitios web, el rastreo de sitios web y la organización de datos son los tres procesos principales y entrelazados".

Los desarrolladores de herramientas de *web scraping* a menudo consideran el rastreo como parte de sus herramientas de *scraping* o al menos incluyen capacidades de rastreo, sin las cuales su software sería menos práctico. Por ejemplo, Barbaresi (2021) desarrolló la Trafilatura, la cual es "una herramienta de web scraping para el descubrimiento y la recuperación de textos que descarga, analiza y raspa sin problemas datos de páginas web. Puede rastrear y descubrir textos dentro de un sitio web y procesarlos en consecuencia" (p. 125). En el caso de la herramienta UzunExt, desarrollada por Uzun (2020), la parte de rastreo es fundamental para la propia extracción, porque "consta de dos componentes principales: rastrear páginas web y extraer datos de ellas. El desarrollador utiliza la información adicional obtenida de las páginas web durante el proceso de rastreo para aumentar la eficiencia del tiempo de extracción" (p. 61726). Por último, Dumbacher y Diamond forman parte de un equipo que desarrolla "Scraping Assisted by Learning" (SABLE), un conjunto de "herramientas para el rastreo web y el scraping web que están asistidas por el aprendizaje automático" (2018, p. 1). Por lo tanto, destaca cómo el scraping, crawling y la clasificación de textos pueden ser fases distintas en un mismo proceso. En otras palabras, el crawling no se limita a ser una parte del scraping; en este contexto, la definición de scraping resulta ser mucho más precisa y restringida.

En términos de categorización de las distintas técnicas de *web scraping* existen diferentes enfoques, pero incluso las definiciones más restringidas constan de varias categorías. Khder (2021) utiliza siete técnicas a efectos de clasificación: técnica tradicional de copiar y pegar; captura de texto y uso de expresiones regulares; programación con HTTP; análisis sintáctico de HTML; análisis sintáctico

Inteligencia de fuentes abiertas y ciencia de datos

de Modelo de Objetos del Documento (DOM, por sus siglas en inglés); software de *Web scraping*; y analizadores de páginas web basados en visión por ordenador. Además, este autor emplea el concepto de enfoques, pero circunscrito al universo de Python, concretamente con tres bibliotecas: Regular expressions, BeautifulSoup y Lxml.

Diouf et al. (2019) también utilizan el enfoque como categoría organizativa, pero de forma más exhaustiva. Entre los enfoques se encuentran:

- Enfoque de mimetismo: cuando un conjunto de reglas que utilizan selectores DOM se aplica a diferentes instancias de la misma web o a diferentes webs con diseños idénticos, no funcionará en webs dinámicas o si se modifica el diseño figura.
- Enfoque de medición del peso: analizando el árbol DOM de una página web, la herramienta deduce, mediante un algoritmo, el punto de inicio del texto principal y sus posteriores nodos hijos. No requiere entrenamiento, pero los resultados tienden a ser ruidosos.
- Enfoque diferencial: asumiendo que dos páginas del mismo sitio web sólo difieren en el contenido, la herramienta sólo extrae las diferencias, asumiendo que éste es el texto principal de cada página.
- Enfoque de aprendizaje automático: con el uso de un conjunto de entrenamiento de páginas web que proporcionan al algoritmo una matriz de posiciones de bloques de texto, el modelo puede deducir por sí mismo la ubicación del texto a extraer.

Diouf et al. (2019) también clasifican las herramientas en listas o bibliotecas. En las primeras incluyen extensiones de navegador (como *Spider*, *Agenty*, *Data Miner*, *Dexi.io*) y softwares y plataformas (*Import.io*, *Mozenda web*, *PySpider*, *Weboob*, o *Screen Scraper*). En estas últimas las dividen por lenguaje de programación, ya sea python (Beautiful Soup, Newspaper, Lxml), Java (Jsoup, Jaunt, StormCrawler, Norconex, Collector) o NodeJs (Cheerio, Apify). Con el fin de proponer UzunExt como un enfoque nuevo y más rápido para el *web scraping*, Uzun (2020) divide los enfoques de scraping entre supervisados, no supervisados e híbridos.

En el primer caso, “los patrones de extracción se aprenden a partir de datos etiquetados” (Uzun, 2020, p. 61727). Los primeros métodos utilizaban “inducciones envueltas”, similares al enfoque Mimicry ya descrito, mientras que las técnicas supervisadas más recientes “se centran en predicciones basadas en un modelo de aprendizaje obtenido a partir del conjunto de datos

de entrenamiento”. Los enfoques no supervisados pueden dividirse entre técnicas basadas en estadísticas, detección de plantillas y suposición. El método basado en estadísticas utiliza información como la proporción de contenido, la proporción de etiquetas, la densidad de enlaces, el número de nodos padre y otros para identificar el contenido legible por humanos que debe extraerse. Los hallazgos de detección de plantillas utilizan una página web maestra para identificar los principales bloques de texto de las nuevas páginas web. Por último, las técnicas de suposición utilizan “la estructura del documento web, los nodos de contenido, el bloque de contenido principal, algunos elementos, información de estilo, etc. para extraer contenido relevante” (2020, pp. 61727–61728).

Las distintas aplicaciones del *web scraping* también pueden clasificarse en diferentes escalas. Por ejemplo, Matta et al (2020, p. 8204) describen usos específicos como la investigación de mercados, el análisis de sentimientos y el seguimiento de precios. Khder (2021) agrupa estas actividades en una categoría más amplia: *Business Intelligence* (BI). Dado el enfoque más amplio, puede ser útil para examinar el abanico de posibilidades sin dejar de lado casos concretos. Volviendo a su lista de usos en inteligencia de mercados, enumera el seguimiento de precios, los estudios de mercado y el análisis de sentimientos.

El seguimiento de precios es especialmente útil en el comercio electrónico, dado que la información de los competidores debe ser pública, pero también es más dinámica, por lo que la definición de precios debe basarse en un seguimiento permanente de las actividades de la competencia. La investigación de mercados se beneficia de muchas maneras del *web scraping*, como “fijación de precios de mercado, análisis de tendencias de mercado, optimización de puntos de entrada, investigación y desarrollo... seguimiento de la competencia” (Khder, 2021, p. 152), y seguimiento de precios en bolsa. El análisis del sentimiento de los medios sociales también es útil para determinar la percepción dinámica de productos, líderes de opinión e incluso candidatos políticos. Khder afirma que el *web scraping* constituye una poderosa ventaja competitiva para las empresas que aprovechan su potencial, sobre todo en escenarios en los que puede explotarse la optimización de motores de búsqueda para atraer a clientes sensibles al precio. Por último, Khder (2021) también describe los usos de la IA, la Ciencia de Datos, el Big Data, la Computación en Nube y la Ciberseguridad.

Otros autores, como Diouf et al (2019) también describen el papel que tiene en los motores de búsqueda, que es probablemente el primer uso que tuvo el web

Un análisis exploratorio para detectar patrones de corrupción

scraping, pero destacando usos específicos como Teambuilder¹ para la búsqueda de empleo o sistemas colaborativos como Grundy, GroupLens, Video Recommender y Ringo o PHOAKS². Lo más importante es que se mencionan herramientas generales de *scraping* como bitLy, Blekko o ScaperWiki (estas dos últimas descatalogadas) que pueden ofrecer nuevas vías para el periodismo de datos hasta motores más centrados en el campo como NewsPaper, FactExtract y NewsOne. El periodismo no es el único uso relacionado con la transparencia pública, precisamente porque el *web scraping* puede ayudar a extraer y visualizar patrones que de otro modo resultan imposibles por medios manuales, aunque los periodistas, como explican Diouf et al (2019, p. 6041) no han aprovechado todo el potencial del scraping. No obstante, el *web scraping* abre nuevas posibilidades más allá del vertiginoso ciclo de noticias.

Ulbricht (2020) desarrolla el concepto de "*demos scraping*", que "engloba las prácticas de obtención de información sobre los ciudadanos mediante el análisis automatizado de datos de rastreo digital que se reutilizan con fines políticos" (p. 426). Para ella, el enfoque tecnocrático del *demos scraping* pretende "convertirse en una práctica democrática que pretende generar una verdad superior y ser de utilidad en la elaboración de políticas" donde "las élites científicas, económicas, políticas y burocráticas hacen afirmaciones representativas sobre las poblaciones basándose en estas prácticas", por lo tanto, al escuchar "más voces" pueden producir "mejores políticas" (Ulbricht, 2020, p. 431). Esto, según Ulbricht, es defectuoso en muchos sentidos, pero lo más significativo es que sitúa a los ciudadanos como "consumidores" pasivos de la democracia.

Lo anterior reduce claramente el alcance de la participación política en las democracias que a menudo ya se limita al acto de votar de vez en cuando, dado que "uno de los elementos centrales de la democracia es hacer visibles las disputas políticas; otro es su naturaleza de proyecto inacabado" (Ulbricht, 2020, p. 431). En otras palabras, la democracia necesita un grado de conflicto regulado y permanente que no puede alcanzar una solución perfecta diseñada por unas élites "expertas". Uno de los principales problemas es que el *demos scraping* reconfigura el espacio político de representación, dado que, la mayoría de los datos raspados no fueron generados como un comportamiento político per se, por lo tanto, pesa de manera desproporcionada la dimensión consumista de los ciudadanos. En este sentido, "la hiperreactividad se ha extendido de las ventas online a los partidos políticos y prospera en un contexto de ideologías políticas en declive" que se alza como "un peligro para la confianza política, la política estable y las políticas sostenibles" (Ulbricht, 2020, pp. 433 y 434)

Ulbricht (2020) reconoce que, a pesar de sus limitaciones y dudosas promesas, el *demos scraping* está en proceso de establecerse como una tecnología de representación y una herramienta de gobernanza. Luego crítica que las iniciativas de "tecnología cívica", que pretenden escudriñar la elaboración de políticas a través del *web scraping*, no son realmente empoderamiento político, sino activación política por parte de unos pocos "ciudadanos dotados de habilidades tecnológicas avanzadas, inclinación y tiempo" (p. 432). El presente estudio podría incluirse en esta categoría, aunque no comparta el optimismo a ultranza de los entusiastas de la tecnología cívica. Sin embargo, este trabajo es consciente de las limitaciones del *web scraping* para escudriñar la formulación de políticas, pero también propone que Ulbricht (2020) pasa por alto el potencial que tiene la extracción automática de datos públicos en sociedades con espacios de representación política gravemente obstaculizados como en Nicaragua.

En este sentido, se considera necesario trascender el "*demos scraping*" como una práctica *top-down* que pretende aplanar el comportamiento democrático de los ciudadanos y abogar por un despliegue más multilateral y colaborativo de las técnicas de *scraping* que permita el escrutinio y la participación, y al mismo tiempo reduzca las múltiples barreras de acceso a las herramientas de *scraping*. En este sentido, incluso Ulbricht (2020) reconoce que el *web scraping* seguirá avanzando como tecnología de representación. Tal afirmación significa que el desarrollo del *web scraping* debe de ser considerado en la discusión democrática; por lo tanto, deberían seguirse explorando y compartiendo más casos de estudio como éste.

1 TeamBuilder es una agencia de reclutamiento de liderazgo en marketing con enfoque digital.

2 Estos autores también mencionan el uso en publicidad, al "establecer automáticamente algunas previsiones sobre los intereses del usuario, recogiendo preferencias o gustos similares" (p. 6041), lo denominan "colaborativo" porque el *web scraping* aquí se utiliza para "pretender predecir la importancia de elementos para un usuario concreto en función de elementos previos valorados por otros usuarios" (Diouf et al, 2019, p. 6041). Este proceso no siempre es claramente consentido o entendido por el usuario, al que tampoco compensa necesariamente ceder sus rastros de datos. También describen usos en el sector de la salud, como "WhichGenes y PathJam...dos meta servidores que utilizan el *web scraping* como medio para afrontar el análisis del enriquecimiento de un conjunto de genes" u otras herramientas de biología molecular, medicina, genética y otros campos que se basan en el *web scraping* como Protein Information Crawler, DrugBank, ChemSpider, BioSpider, OReFil y MEDPIE (Diouf et al 2019, p. 6041).

En el siguiente apartado se profundizará en la contratación pública y las herramientas digitales, pero cabe destacar que tras extensas revisiones bibliográficas sólo en el 3% de los casos utilizaron *web scraping*, sólo el 4% minería de textos y sólo el 3% se desarrollaron en América Latina (Modrušan, Rabuzin, et al., 2021a; Torres Berru et al., 2020). El ejemplo más notorio de *web scraping* en contratación pública en América Latina fue llevado a cabo por Torres-Berru y su equipo en Ecuador (Torres-Berru & López Batista, 2022). El presente estudio es el primero en utilizar el *web scraping* para analizar la contratación pública en Nicaragua.

5.1.2 Revisión bibliográfica sobre machine learning en los procesos de licitaciones públicas

La contratación pública puede definirse como "el proceso mediante el cual el sector público adquiere servicios, bienes y obras de construcción a terceras empresas" (Lyra et al., 2022, p. 1), pero lo más importante es que los sistemas de contratación permiten a los gobiernos ejecutar sus presupuestos para proporcionar bienes, servicios e infraestructuras a la comunidad (Volosín, 2015, p. 3).

La magnitud de las adquisiciones públicas varía según las fuentes. Por ejemplo, el Banco Mundial estimó en 2009 que las adquisiciones públicas en América Latina y el Caribe pasaron del 10% al 15% del PIB (Volosín, 2015, p. 7). Más recientemente, la Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2020) destacó una disminución de las compras públicas del 6,7% en 2009 al 6% del PIB en 2017 en la misma región, con variaciones entre países. Por ejemplo, alcanza el 3,6% en México y asciende a 9,9% en Perú. En este último país, la contratación pública representa el 46,4% del gasto público total, mientras que alcanza el 34% en Colombia. Las proporciones más bajas de contratación pública contra el gasto público total se encuentran en las mayores economías de la región, Brasil con un 13,5% y México con un 16%, reduciendo en ambos casos la media regional al 17,4%.

Volosín (2015) define tres objetivos principales para los sistemas de contratación pública: el desarrollo humano, la eficiencia y la transparencia. Estos objetivos pretenden garantizar el mayor valor posible en bienes, servicios y obras; el alcance al mayor número posible de ciudadanos (teniendo en cuenta consideraciones distributivas); el menor coste posible; y un proceso transparente, libre de la influencia de la corrupción y con un control permanente por parte de los organismos públicos pertinentes, los participantes en el proceso de contratación y la sociedad civil en general (p. 6). Los tres objetivos están estrechamente relacionados y se afectan mutuamente. En el caso de la transparencia, su falta provocaría ineficiencias

económicas en la ayuda internacional y la inversión pública, el deterioro de la confianza democrática en el gasto público y, además, obstáculos al cumplimiento de los Objetivos de Desarrollo Sostenible (ODS), afectando especialmente a la distribución social de la riqueza.

Según Transparencia Internacional (2022), entre el 20% y el 25% del valor de los contratos públicos se pierde debido a la corrupción. Esta corriente de pérdidas es una de las principales razones para la implementación de sistemas electrónicos de contratación pública, que ha sido adoptada por más de dos tercios de los países de la región (OCDE, 2020). Estos sistemas permiten una mayor eficiencia, menor burocracia, entre otros beneficios. La presente investigación se centra en los efectos potenciales sobre la transparencia creados por los datos digitales generados por estos sistemas.

Los gobiernos se benefician de esta nueva información, ya que, al tener digitalizadas las ofertas, pueden hacer un análisis más exhaustivo y detallado de los licitadores y de las ofertas en sí mismas, mientras se asegura una mayor transparencia en la realización de los concursos. Sin embargo, lo más importante es que las agencias anticorrupción tienen más control y mejores herramientas para identificar comportamientos sospechosos, es decir, estas herramientas permiten la rendición de cuentas horizontal (Volosín, 2015, p. 4). Volosín también plantea que la contratación pública electrónica debe ser capaz de generar información inteligente y agregada a la que también tenga acceso el público, esto es lo que denomina rendición de cuentas social, que, junto con la participación ciudadana y el libre acceso, constituye una poderosa herramienta de transparencia democrática (p. 6).

Los portales de compras públicas generan una gran cantidad de datos e información, y por ende también son un punto de partida para discutir acerca de datos abiertos, que son, tal como los define la OCDE, "datos gubernamentales liberados sin ningún tipo de barreras técnicas y legales y, si es posible, de forma gratuita" con el fin de permitir que "sean utilizados, reutilizados y redistribuidos libremente por cualquier persona, en cualquier momento y en cualquier lugar" (OCDE, 2020, p. 134). La OCDE utiliza tres criterios para evaluar los Datos Abiertos Gubernamentales (DAG): a) la disponibilidad de los datos, que se refiere "a los marcos políticos que regulan la liberación de los datos gubernamentales"; b) la accesibilidad de los datos, que mide la forma de liberar los datos, por ejemplo, si están en formatos legibles por máquina, en archivos editables, y similares; y c) el apoyo gubernamental a la reutilización de datos, que mide "los esfuerzos para animar a las partes interesadas de la Administración y la sociedad civil a utilizar los DGA". Según la Carta de Datos Abiertos (s.f.), los datos abiertos deben estar disponibles libremente

Un análisis exploratorio para detectar patrones de corrupción

para ser utilizados, reutilizados y redistribuidos por cualquiera, con requisitos mínimos como la atribución. Para esto, deben estar disponibles con un marco legal claro, ser legibles por máquina, oportunos y completos.

Es fácil concluir que los datos abiertos son necesarios para prevenir y detectar la corrupción. Sin embargo, antes de profundizar en los usos de las nuevas tecnologías para estos fines, es importante explorar la definición de corrupción, que hasta ahora se ha utilizado como término genérico pero que puede subdividirse. Por ejemplo, Lyra et al. (2022) definen la corrupción como "la actividad relacional entre agentes corruptores y entidades corrompidas, siempre con la participación del organismo público". Además, definen la colusión como "un acuerdo entre terceros para defraudar operaciones de la administración pública, sin su participación" y el fraude como "delitos de falsificación de documentos, propuestas fraudulentas, uso de empresas ficticias, introducción de falsa evaluación de competidor, fraude de identidad y robo de datos".

Por su parte, Torres-Berru et al (2020) siguiendo a Padhi y Mohapatra (2011) siguen utilizando corrupción como un término paraguas que incluye "soborno, colusión, malversación, apropiación indebida, fraude, abuso de discrecionalidad, favoritismo y nepotismo" (p. 253). Modrušan et al (2021a) describen directamente la corrupción en la contratación pública como "el abuso de poder en beneficio propio" (p. 1). En total, Modrušan et al utilizan trece tipos diferentes de corrupción y los clasifican por impacto y probabilidad. Dentro de su ámbito de revisión, los fraudes más comunes en la contratación pública son la manipulación de licitaciones, la colusión entre proveedores y empleados y la colusión entre proveedores; en tanto, los de mayor impacto son el soborno y las comisiones ilegales, el conflicto de intereses, las licitaciones colusorias, la ejecución y las donaciones a partidos políticos.

A continuación, se exponen diferentes tecnologías y casos prácticos utilizados para detectar o prevenir irregularidades y corrupción en la contratación pública mediante el uso de IA. Se destaca que, este ejercicio no es una meta-revisión sistemática, lo cual ya se ha hecho recientemente por otros autores (Khder, 2021; Lyra et al., 2022; Mavidis & Folinas, 2022; Modrušan, Rabuzin, et al., 2021a; Nai et al., 2022; Torres Berru et al., 2020). En cambio, se discuten una diversidad de ejemplos que han informado el proceso de definición de una metodología útil para este estudio, pero que también configura las posibilidades presentes y futuras de este esfuerzo.

Para Serrano (2021), la IA es "el conjunto de todas las tareas en las que un ordenador puede tomar decisiones"; además, "un ordenador toma estas decisiones imitando las

formas en que un humano toma decisiones" pero también "pueden imitar procesos evolutivos, procesos genéticos o procesos físicos" (pp. 3–4). Este autor desmitifica el término hasta el punto de decir que "cada vez que vemos a un ordenador resolver un problema por sí mismo, ya sea conducir un coche, encontrar una ruta entre dos puntos, diagnosticar a un paciente o recomendar una película, estamos ante inteligencia artificial" (Serrano, 2021, p. 4). Además, diferencia el aprendizaje automático (*Machine Learning*, ML) de la IA definiendo el ML como "el conjunto de todas las tareas en las que un ordenador puede tomar decisiones basándose en datos" (Serrano, 2021, p. 4). Así, por ejemplo, hay algunos métodos de IA utilizados en la contratación pública que no utilizan, por definición, el ML. Un ejemplo de lo anterior es la minería de textos y el análisis de redes (aunque las técnicas avanzadas pueden ciertamente incorporar ML).

Serrano (2021) también estudia el funcionamiento de las redes neuronales³, las cuales pueden aprender de los datos y hacer predicciones o tomar decisiones sin ser programadas explícitamente⁴. Son especialmente útiles para tareas en las que la relación entre entradas y salidas es compleja y no lineal⁵ (p. 279).

Modrušan, Rabuzin, et al. (2021a) plantean diferencias entre aprendizaje supervisado y no supervisado. De acuerdo con los autores, se puede decir que "en el aprendizaje supervisado, se tienen variables objetivo, definidas con precisión como salida del modelo, mientras que en el aprendizaje no supervisado no se tienen variables preestablecidas; así, los modelos son adecuados para buscar anomalías" (p. 3). El aprendizaje supervisado se entrena con datos etiquetados, lo que significa que los datos de entrada tienen una salida o etiquetas correspondientes. El modelo aprende a asignar los datos de entrada a las etiquetas de salida y puede hacer predicciones sobre nuevos datos no vistos. Por su parte, el aprendizaje no supervisado, se entrena con datos no etiquetados, ya que no hay salida o etiquetas correspondientes para los datos de entrada.

3 Las redes neuronales son modelos computacionales inspirados en el cerebro humano. Consisten en nodos interconectados que procesan información para resolver problemas específicos, capaces de aprender y adaptarse a partir de los datos de entrada. Se utilizan en reconocimiento de patrones, visión por computadora, entre otros (Serrano 2021, pp. 277-299).

4 Se utilizan para una amplia gama de tareas como el reconocimiento de imágenes, el reconocimiento del habla, el procesamiento del lenguaje natural y la toma de decisiones.

5 Esto no significa que no se puede utilizar una regla o ecuación simple para describirla.

Inteligencia de fuentes abiertas y ciencia de datos

El objetivo del aprendizaje no supervisado es descubrir patrones o estructuras subyacentes en los datos como agrupar puntos de datos similares, detectar anomalías o reducir la dimensionalidad de los datos (Serrano, 2021, pp. 17–22). Esta consideración es útil a tener en cuenta en el presente estudio, dado que se busca identificar potenciales anomalías en el proceso de adquisiciones públicas.

Algunas técnicas habituales de aprendizaje no supervisado son el Análisis de Componentes Principales (PCA por sus siglas en inglés), que se emplea para identificar patrones en los datos y extraer características importantes, la agrupación (Clustering), que agrupa puntos de datos similares, y el bosque de aislamiento (Isolation Forest), que detecta valores atípicos o anomalías en el conjunto de datos.

Modrušan et al. (2021b) que realizaron una revisión diferentes tecnologías y casos prácticos utilizados para detectar o prevenir irregularidades y corrupción en la contratación pública mediante el uso de IA son bastante específicos a la hora de describir las técnicas utilizadas por los artículos que consultaron. De acuerdo con estos autores, los "métodos más utilizados en los estudios son la regresión lineal y logística, las redes neuronales y los algoritmos Naive Bayes, ya que son los más utilizados para la clasificación y la agrupación" (p. 3). Independientemente del método utilizado, les preocupa la calidad de los conjuntos de datos y las banderas rojas utilizadas para definir la corrupción. Entre las banderas rojas encontradas en la literatura destacan las siguientes:

- Plazo inusualmente corto entre el anuncio de la licitación y la fecha límite para la presentación de ofertas.
- El tiempo transcurrido entre el anuncio de la licitación y la firma del contrato.
- Un alto porcentaje de ofertas rechazadas administrativamente en el procedimiento.
- Un número inusualmente reducido de ofertas correctas en el nivel del procedimiento de contratación.
- Una oferta aceptada antes de la fecha límite de presentación de ofertas.
- Elevada proporción del valor de los contratos firmados en condiciones especiales en relación con el valor total de todos los contratos de un cliente individual, etc.

Por su parte, Torres-Berru et al. (2020) encontraron 102 estudios que utilizan técnicas de IA. Dichos estudios

los clasifican en basados en Bayes (12 estudios), Redes Neuronales⁶ (9), Máquinas de Vectores Soporte (9), Árboles de Decisión (7), Bosques Aleatorios (7), regresión logística o lineal (12) y Otros (17). Los autores también agruparon los estudios en otros tipos de clasificaciones, tales como Tipos de Aprendizaje (supervisado, no supervisado, semi supervisado), Tipo de Minería (Predictiva, Descriptiva), Fuente de Datos (Pública, Privada), Tipo de Datos (Datos, Texto, Audio), Instrumento de Recolección (Encuesta, Web Scraping, Base de Datos), Campo de la Ciencia (Informática, Economía, Matemática, Estadística), Sector de Actividad (Público, Privado, Mixto) y otros no tan relevantes. Cabe destacar que, la mayoría de los estudios utilizan bases de datos y sólo unos pocos (2%) utilizan *web scraping*. Además, aunque el estudio se centra en la contratación pública, sólo 7 estudios analizan específicamente las compras públicas y sólo 22 se ocupan de las actividades gubernamentales. En un estudio similar, Torres-Berru y López-Batista (2022) descubrieron que menos del 9% de la bibliografía sobre análisis de compras públicas se había producido en América Latina, principalmente en Brasil (p. 118).

Por su parte, Lyra et al. (2022) encontraron 48 estudios de análisis de compras públicas en total, en los que la técnica más popular era la "ciencia de redes", que aparecía en 34 estudios, seguida del aprendizaje automático, con 15 estudios. Para abordar la clasificación más específica, utilizaron como división primordial el tipo de datos utilizados, ya fueran estudios de datos etiquetados o investigaciones exploratorias (relacionadas con datos no etiquetados), con un 52% y un 48% de representación en la muestra, respectivamente. En el campo de datos etiquetados, los métodos de aprendizaje automático utilizados fueron Redes Neuronales Artificiales (3 estudios), Máquina de Vectores Soporte (4 estudios), Random Forest⁷ (3), Regresión Logística (2), Regresión Lasso (2), Modelo de Clasificación Gradient Boosting (1), Naive Bayes (1), Árboles Boosted (1), clasificación binaria (1).

De ellas, las más exitosas fue Random Forest, que se empleó para "encontrar cárteles de manipulación

6 Las redes neuronales son modelos computacionales inspirados en el cerebro humano. Consisten en nodos interconectados que procesan información para resolver problemas específicos, capaces de aprender y adaptarse a partir de los datos de entrada. Se utilizan en reconocimiento de patrones, visión por computadora, entre otros (Serrano 2021, pp. 277-299).

7 O bosque aleatorio en español.

Un análisis exploratorio para detectar patrones de corrupción

de licitaciones, identificar conexiones estrechas entre el gobierno y las empresas privadas en interacción repetida y dispersión geográfica, y detectar licitaciones únicas y concentración del gasto público". Por su parte, la Regresión Logística "ayudó a encontrar licitaciones sospechosas con técnicas de minería de texto y a detectar licitaciones únicas y concentración del gasto público", mientras que la regresión Lasso "tenía como objetivo detectar cárteles de manipulación de licitaciones y predecir la ineficiencia y la corrupción en la contratación pública" (Lyra et al., 2022, p. 19).

En el caso de la ciencia de redes, las técnicas más utilizadas fueron el análisis de conglomerados (8 estudios), que agrupa nodos con gran similitud -y en comunidades-, y las medidas de centralidad (4), utilizadas para "reflejar la prominencia, el estatus, el prestigio o la visibilidad de un nodo y a menudo explican comportamientos en la red" (Lyra et al., 2022, p. 19). Una conclusión importante a la que llegaron Lyra et al. es que los indicadores de datos etiquetados más importantes para exponer comportamientos fraudulentos eran los gastos contractuales, el título del contrato, los pesos de proximidad y los valores medios de los vectores propios (20).

Aunque los datos etiquetados son muy deseables y tienden a arrojar los mejores resultados, funcionan mejor para descubrir cárteles y colusiones que ya se conocen con el fin de construir un algoritmo. Cuando los datos no están etiquetados, la ciencia de redes es útil para "identificar relaciones ocultas entre los agentes de la red", por lo que el algoritmo agrupa los nodos con diferentes métodos estadísticos (Lyra et al., 2022, p. 22). Los trabajos más exitosos emplean más de una técnica para cubrir las limitaciones de las demás.

Por su parte, Nai et al. (2022) clasifican las técnicas utilizadas en los 14 estudios revisados en tres categorías: a) aprendizaje automático; b) análisis de redes y minería de textos; y c) redes neuronales. Curiosamente, encontraron tres estudios de Brasil y uno de Paraguay, Colombia y Ecuador, lo que significa que hay algunas investigaciones que se están realizando en la contratación pública y la IA en la región. Casi todos los estudios emplearon *Machine Learning*, con un total de 13, con técnicas como *Self-Organizing map*, *Support Vector Machine*, *Principal Component Analysis*, *Random Forest*, *Isolation Forest*, *Lasso classification model*, *binary logistic regression*, *Gradient Boosting Machine*, entre otras. Cuatro estudios utilizan análisis de redes, sobre todo clustering, análisis de regresión y teoría de grafos. Finalmente, seis estudios

utilizan enfoques de Redes Neuronales, tales como Redes Neuronales Bidireccionales y Redes Neuronales Profundas⁸.

Si se examinan las diferentes reseñas bibliográficas, queda claro que existen algunos vacíos que aún no se han abordado. Por ejemplo, si bien la contratación pública digital existe en muchos lugares, no se implementa en el mismo grado; por lo tanto, la cantidad de información generada por el sistema varía. Geográficamente, como se ha discutido anteriormente, muy poca literatura ha sido producida en América Latina y el *web scraping* se utiliza poco como principal método de recolección de información. Adicionalmente, muy pocos estudios tienen un enfoque interdisciplinario incorporando profesionales de las ciencias sociales para proveer un alcance más amplio de herramientas analíticas.

El presente estudio se ubica en la intersección de algunos de estos vacíos, ya que el sistema de compras públicas en Nicaragua, si bien se encuentra digitalizado, no mantiene las subastas en línea (aunque la información se actualiza puntualmente), no existe una base de datos descargable y el etiquetado no es necesariamente uniforme a través de los años. Por lo tanto, se tiene que raspar los datos públicos para crear una base de datos y encontrar una manera de homogeneizar las etiquetas de los datos para que haya identificadores clave únicos para cada entrada de datos a pesar de los diferentes formatos. Además, no se ha podido acceder de forma constante y confiable al sitio web. Nicaragua también es un país que atraviesa una crisis democrática, y cualquier esfuerzo que utilice datos abiertos para identificar posibles comportamientos fraudulentos, en particular con la ayuda internacional, sería muy valioso.

A continuación, se analizan brevemente tres artículos que presentan ejemplos de trabajos que de una forma u otra informan las posibilidades de este estudio. Por ejemplo, Ortiz-Prado et al. utilizaron el *web scraping* para extraer información del portal de contratación pública de Ecuador, con una librería de Python llamada Selenium, que extrae información del código HTML de la página web. Esto les permitió extraer estas variables: "institución gubernamental responsable del proceso de compra (Ministerio, Gobierno Central, Gobiernos Autónomos Descentralizados), ubicación donde se basó el contrato (Municipio y Provincia), tipo de producto o servicio, precio unitario y tipo de contrato (contrato directo o subasta pública)" (Ortiz-Prado et al., 2021, p. 82).

8 En esta revisión, se ve la mayor superposición de técnicas que cualquier otra, incluso cuando cubre menos artículos que otras.

Inteligencia de fuentes abiertas y ciencia de datos

Dichos autores centraron en las compras públicas en el contexto de la pandemia de Covid-19, por lo que utilizaron un análisis estadístico para determinar si los precios pagados se desviaban significativamente de los precios de mercado en ese momento y descubrieron que "algunos precios eran excepcionalmente superiores a su valor de mercado y hasta un 1300% de diferencia con compras similares [...] en comparación con 2019, el precio medio de los guantes de reconocimiento médico aumentó hasta un 1.307%, el de las pastillas de 500 mg de paracetamol, hasta un 796%, y el de los frascos de oxígeno, un 30.8%" (Ortiz-Prado et al., 2021).

Por su parte, Modrušan et al. (2021b) presentan un sistema que utiliza la minería de textos e indicadores equilibrados para supervisar la contratación pública. El sistema utiliza técnicas como el preprocesamiento de textos, la extracción de características y el análisis de sentimientos para extraer información relevante y evaluar el rendimiento de los procesos de contratación en términos de eficiencia, eficacia y transparencia. El sistema se probó con un conjunto de datos y se comprobó que extraía con éxito información relevante y evaluaba el rendimiento de la contratación pública con un alto nivel de precisión.

Finalmente, Rodríguez et al. (2022) utilizaron conjuntos de datos de varios países y aplicaron varios algoritmos de aprendizaje automático (como regresión logística, árboles de decisión y Random Forest) para analizar los datos. Los resultados del estudio mostraron que el sistema propuesto es capaz de detectar la colusión con gran precisión. Entre las técnicas de aprendizaje automático aplicadas, las "tres de mayor rendimiento [...] han sido los Árboles Extra, el Bosque Aleatorio y Ada Boost (métodos ensemble)", especialmente en la identificación de patrones de comportamiento colusorio y la detección de datos sospechosos. Los autores también propusieron una herramienta de visualización para ayudar a los usuarios del sistema a comprender los resultados e identificar los casos más sospechosos.

5.2 Metodología de Web scraping

5.2.1 Modelo de relación entre entidades

Para definir la arquitectura de datos y cómo se desarrolla la estructura para el análisis, se creó un Modelo Entidades-Relación (ERD, por sus siglas en inglés). Un ERD es una representación gráfica de las entidades y sus relaciones entre sí. En este caso, por su simplicidad visual se optó por usar la notación de la Ingeniería de Información (IE, por sus siglas en inglés). Esta notación permite mostrar las entidades como cajas y las relaciones entre ellas como líneas que conectan las cajas.

El ERD que se construyó contiene 12 entidades representadas como cajas:

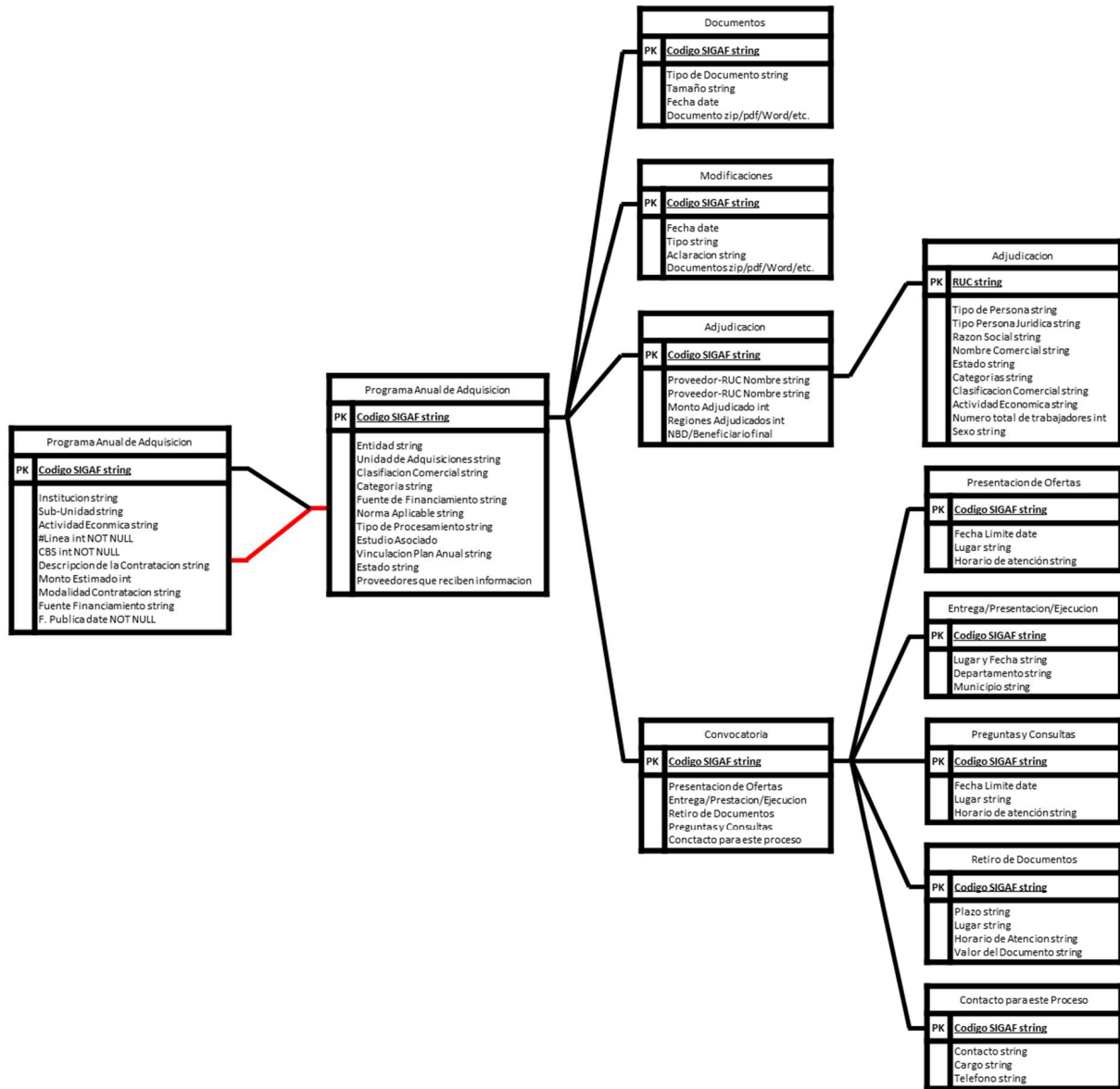
1. Programa anual de adquisiciones
2. Todos los procesos
3. Documentos
4. Modificaciones
5. Adjudicación
6. Convocatoria
7. Proveedores
8. Presentación de ofertas
9. Entrega/Presentación/Ejecución
10. Preguntas y consultas
11. Retiro de documentos
12. Contacto para este proceso

Cada entidad tiene su propio conjunto de atributos, representados como pequeños rectángulos dentro de las cajas de entidad. Por ejemplo, el Programa anual de adquisiciones tiene atributos como "código SIGAF", Institución etc... La relación entre estas entidades se representa como líneas que conectan las cajas de entidad. Esta representación se denomina cardinalidad y representa cómo cada entidad puede relacionarse con otra. Estas pueden ser:

1. Uno a uno (1:1)
2. Uno a muchos (1: N)
3. Muchos a uno (N:1)
4. Muchos a muchos (N: N)

En la siguiente figura, se ha modelado la relación de puntos de datos contenidos en las interfaces relevantes del sitio web www.nicaraguacompra.com.ni. Dicha representación gráfica es la ruta para la construcción de la arquitectura de datos que es explicada en el siguiente apartado.

Figura 1. Diagrama de relación de entidades



Fuente: Elaboración propia

5.3 Arquitectura de datos

Dado que los datos publicados en el sitio web son de acceso público, no es necesario tener en cuenta específicamente los permisos y el acceso. Se extrajeron los datos del sitio web mediante código desarrollado en Python y con la ayuda de las bibliotecas integradas proporcionadas por la plataforma, como se menciona en la metodología de *web scraping*.

Posteriormente, se dividió la información y los datos disponibles en el portal de licitaciones en diferentes cuadros en función de su relación. Cada proceso se bifurcó en función de la función que cumple y de la fase en la que aparece. La estructura de datos comienza con el "PAC", que es el punto inicial para entrar en las tablas relacionadas que contienen detalles sobre el contrato y sus datos. Cada tabla tiene una llave primaria denominada "SIGAF" que se utiliza para identificar la unicidad del proceso o contrato (Fig. Y.1). Mientras que la clave "SIGAF" es excelente para rastrear los datos dentro del MTI, en el caso del CSE hay que utilizar el número "Línea" para rastrear los contratos.

El almacenamiento de estas tablas se realiza después del *scraping*. Las dos etapas de la metodología de *web scraping* mencionadas anteriormente dan lugar a dos formatos diferentes de almacenamiento de datos: CSV⁹ y una hoja de cálculo de Excel. En cada una de las páginas web escaneadas hay información sobre contratos que, utilizando la librería Pandas, se extrae. Entre esta información está el código SIGAF, la cadena de entidades, la unidad de adquisición, la clasificación del contrato y el tipo de producto, etc. Estos datos se procesan a través de la biblioteca y se cargan como un archivo CSV en la carpeta de resultados dentro de la plataforma de procesamiento en la nube. Posteriormente, el CSV se importa en una hoja Excel unificada que contiene diferentes pestañas basadas en el año y el ministerio extraído.

La ficha tiene el formato {año}_{ministerio}. Por ejemplo, 2021_ministerio_de_transportes_e_infraestructuras. Las diferentes pestañas se importan y anexan con el archivo de resultados CSV del código que están interrelacionadas por la clave primaria presente en la pestaña "Resumen".

La tabla dentro de cada pestaña fue estructurada para albergar una serie de puntos de datos como:

Tabla 4. Descripción del contenido de tablas

Punto de datos	Descripción
Código SIGAF	Clave única primaria para identificar el contrato
Clave de entidad	El ministerio identificado (por ejemplo, MTI)
Clasificación comercial	La categoría del contrato adjudicado/tramitado
Fuente de financiación	El origen de los fondos generados
Condición	La fase del contrato (Adjudicado, en curso, rechazado, etc.)
Vinculación Plan anual	El código para vincular el contrato al documento del plan anual
Tipo de intervención	Una licitación o un concurso
Modalidad de procedimiento	Comprobar la modalidad del contrato
Norma aplicable	La norma asociada del contrato

Fuente: Elaboración propia

Para navegar y utilizar los detalles del contrato, basta con analizar la información proporcionada en el Excel unificado, porque contiene toda la información necesaria para cada contrato. Sin embargo, para cotejar algunos detalles y confirmar hipótesis fue importante consultar los documentos contractuales vinculados que están disponibles en las filas.

5.4 Ejecución del Web scraping

5.4.1 Flujo de datos

El flujo de datos del proyecto comienza con la recopilación de datos del sitio web https://www.gestion.nicaraguacompra.gob.ni/siscae/portal/adquisiciones-gestion/busqueda?accion=todos&usr_ua_id=todos. El mayor desafío en este punto fue que todo el sitio, incluyendo su motor de búsqueda y la forma en que muestra los resultados, se brinda como una forma dinámica y no como sitios individuales con URL únicas. No es posible determinar si lo anterior es una programación intencionada o no. En caso de que fuera intencional, el objetivo podría ser dificultar el acceso a la información o aumentar la seguridad del sitio; en caso de lo segundo, sería una mala práctica en términos de ciberseguridad¹⁰.

Independientemente de si hubo intencionalidad o no en la programación web, lo cierto es que esto aumenta la dificultad de crear una base de datos, lo cual constituye un ejemplo de cómo en teoría los datos pueden ser públicos, pero no realmente abiertos o fácilmente accesibles. Para superar esta limitación se utilizó una API capaz de generar consultas (*queries*) al sitio web sin la necesidad de pasar por

9 Comma Separated Values es un formato que permite manejar grandes cantidades de datos en texto sin formato, ahorrando grandes cantidades de espacio a costas de que el formato solo es legible para programas computacionales.

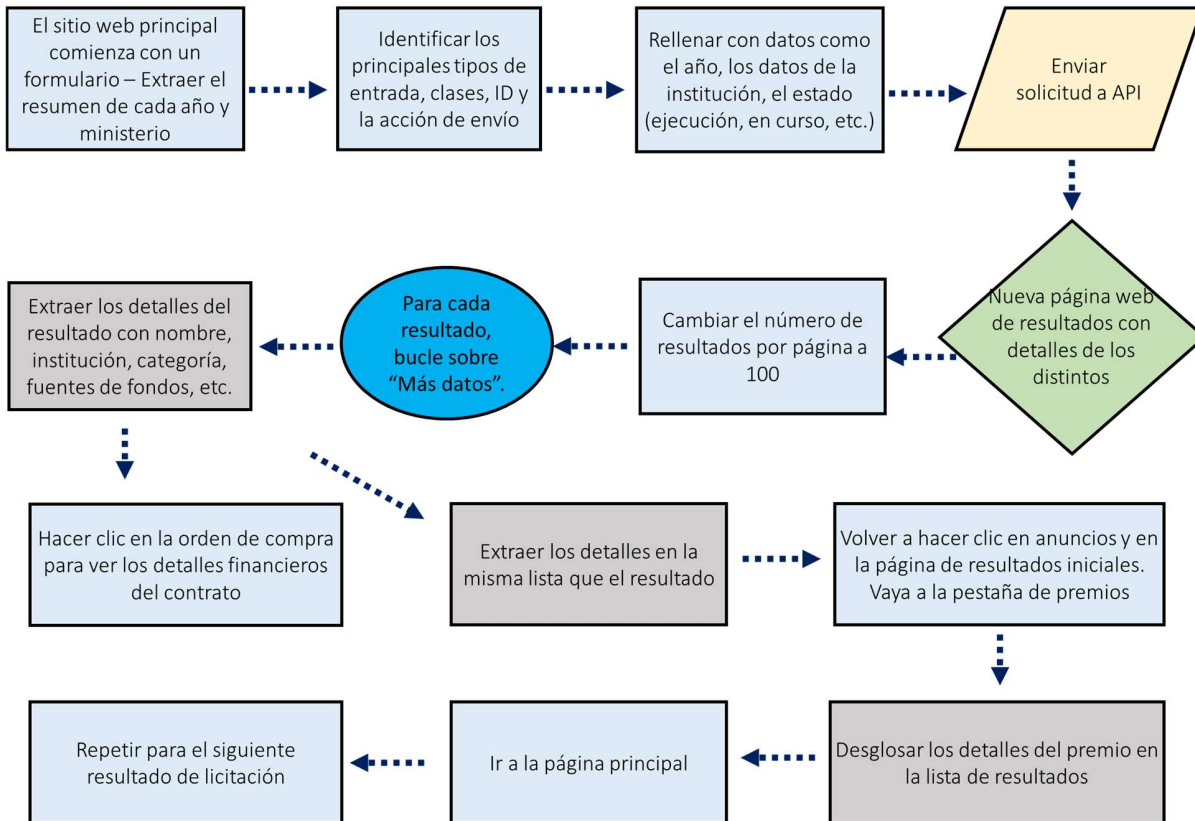
10 Un sitio web seguro suele incluir API bien gestionadas, claves ocultas y estructuras internas. Al acceder al sitio web para extraer los datos, se pudo averiguar todos estos detalles a través del elemento de inspección del navegador web, lo que revela que la seguridad del sitio web no es robusta y no impide que el público acceda a información privada.

Un análisis exploratorio para detectar patrones de corrupción

la interfaz gráfica, pero siguiendo de forma análoga todos los pasos de navegación de la interfaz de usuario tal como se haría en un navegador. Lo anterior permite acceder a la información sin necesidad de utilizar URLs. Para ejecutar las consultas, se identifican los distintos campos

para introducir información en el formulario del sitio web principal, incluidos el año, los datos de la institución y las condiciones (como en ejecución o en curso), incluido el botón "enviar acción". A continuación, se identifican según sus tipos, clases e identificadores HTML.

Figura 2. Flujo de datos



Fuente: Elaboración propia

Inteligencia de fuentes abiertas y ciencia de datos

A partir de la información de los documentos de síntesis, se rellena el formulario con los datos necesarios y la herramienta hace click en el botón de envío. Una vez enviado el formulario, se generan los resultados de la consulta, que muestran detalles de diferentes contratos. El número de resultados por página pasa a 50, por lo que se necesitan menos iteraciones del proceso. Posteriormente, para cada resultado, el programa hace un bucle sobre el botón "Más datos" para extraer los detalles de cada contrato, incluido el nombre, la institución, la categoría, las fuentes de financiación, etc. Posteriormente, la herramienta hace click en la orden de compra para acceder a los detalles financieros de cada contrato; estos detalles se añaden a la misma lista que el resultado al que corresponden. El proceso vuelve a la página inicial de resultados, navega hasta la pestaña de adjudicaciones y repite el proceso. Estos archivos se almacenan en una carpeta, con el nombre de la institución y el nombre del archivo como etiqueta. Por último, la herramienta vuelve a la página principal y repite el proceso para el siguiente resultado.

5.4.2 Flujo de códigos

La implementación de una herramienta de *web scraping* para este proyecto consta de varias fases. Se utilizó Proton VPN (Región: EE.UU.) para la navegación anónima y Google Cloud Notebook como IDE/plataforma. Posteriormente se eligió Python como lenguaje de programación principal y bibliotecas de Python como Scrapy, Requests, BFsoup, Selenium, Http y Pandas por varias razones que las hacen útiles y adecuadas para este proyecto de *web scraping*, las cuales se discuten a continuación:

- Scrapy: es un marco de *web scraping* rápido y poderoso que facilita la extracción de datos de sitios web. Scrapy es especialmente útil para seguir enlaces y extraer datos de varias páginas, lo que lo hace perfecto para proyectos de *web scraping* más grandes y complicados.
- Requests: esta herramienta permite realizar solicitudes HTTP de manera sencilla y eficiente. Permite enviar solicitudes HTTP con varios métodos, como GET y POST, lo que facilita la interacción con sitios web y la extracción de datos. Además, es fácil de usar y se integra bien con otras bibliotecas de Python.
- BeautifulSoup (BFSoup): se utiliza para analizar documentos HTML y XML. Permite navegar, buscar y modificar el árbol de análisis sintáctico de una manera muy sencilla. BeautifulSoup es útil para extraer datos de sitios web con estructuras HTML complicadas y no estructuradas.
- Selenium: es una herramienta de automatización de navegadores web que permite interactuar con sitios web de manera similar a como lo haría un usuario humano. Es especialmente útil para el *web scraping* de sitios web que utilizan JavaScript y AJAX para cargar contenido dinámico, porque puede manejar eventos y ejecutar scripts en el navegador.
- Http (HTTPClient): esta proporciona una interfaz simple y consistente para realizar solicitudes HTTP. Facilita la interacción con sitios web y la gestión de cookies, redireccionamientos y otros aspectos de las solicitudes HTTP.
- Pandas: está habilitada para la manipulación y análisis de datos. Permite trabajar con estructuras de datos como Data Frames y Series, lo que facilita la limpieza, transformación y análisis de los datos extraídos durante el proceso de *web scraping*.

Estas bibliotecas fueron seleccionadas por el equipo investigador debido a su eficiencia, capacidad para abordar sitios web complicados y sus amplias comunidades de usuarios. Estas características facilitan la resolución de problemas y la implementación de soluciones en proyectos de *web scraping*, haciendo que estas bibliotecas sean ideales para el proyecto en cuestión.

Una vez seleccionadas las herramientas, se determinaron las variables necesarias como la URL, los datos de salida y los resultados. Luego, se esbozaron y estructuraron los archivos de las diferentes funciones para garantizar un enfoque claro y organizado de la información. Se definieron las funciones, los parámetros y los resultados que brindaría cada función. Esto incluye la lógica del código, el algoritmo y el desglose de tareas y subtareas¹¹ del proyecto. Antes de iniciar el proceso de codificación, se analizó a fondo la estructura del sitio web, su navegación y los scripts implicados. A continuación, se confirmaron los distintos atributos y resultados que requería el equipo de análisis de datos y se nombraron los archivos, las bibliotecas utilizadas dentro de los archivos y las funciones que contienen.

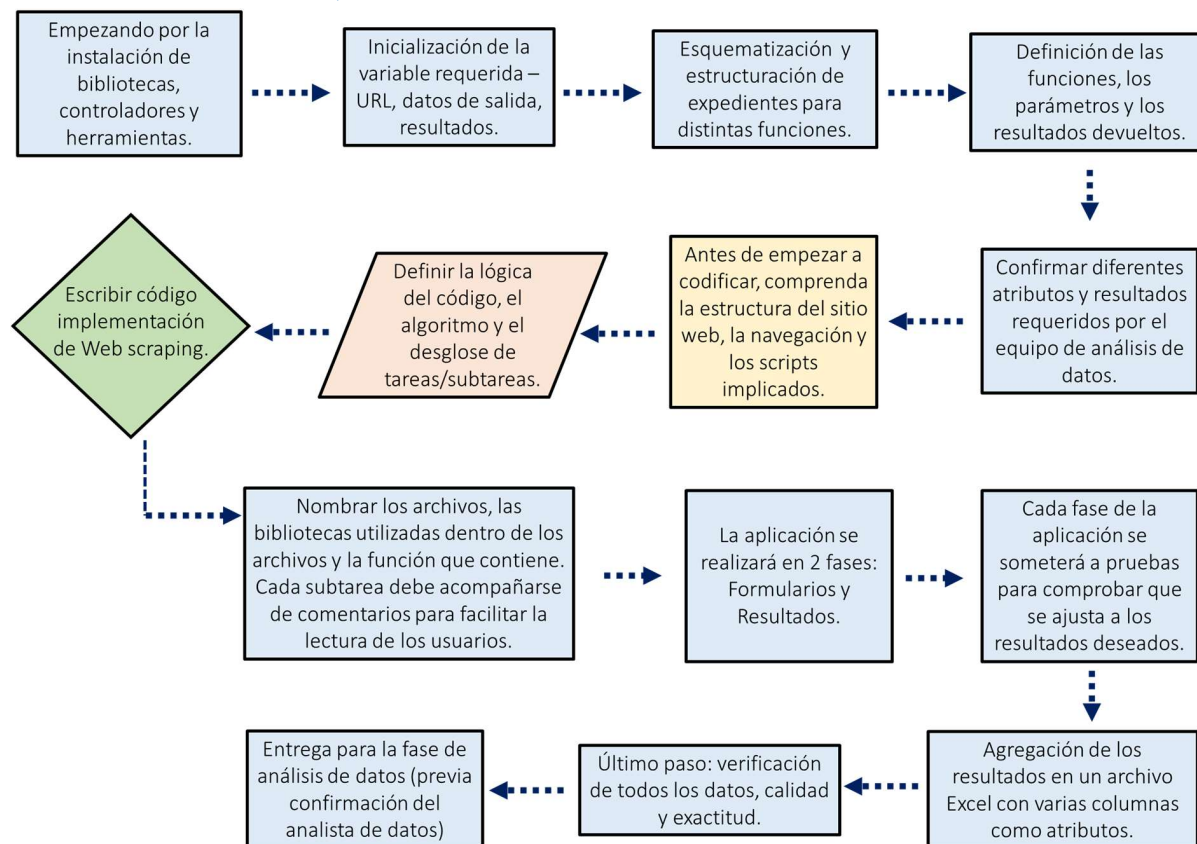
Se realizó la aplicación en tres fases: a) formularios; b) resultados; y c) en formato CSV. Durante cada fase se probó la implementación para que coincidiera con la salida requerida. Una vez confirmada la salida, se entregaron los datos al equipo a cargo del análisis de datos.

11 Cada subtask va acompañada de comentarios para facilitar la lectura a los usuarios.

Un análisis exploratorio para detectar patrones de corrupción

El último paso consistió en verificar todos los datos, su calidad y exactitud. Se agregaron los resultados en un archivo Excel con varias columnas como atributos. Esta metodología garantizó un enfoque claro y eficaz para completar la fase de *web scraping* de este proyecto.

Figura 3. Flujo de código en Python



Fuente: Elaboración propia

Inteligencia de fuentes abiertas y ciencia de datos

Los datos utilizados son accesibles en la página web del portal gubernamental de licitaciones de los diferentes ministerios y en la sección de años. Extraer los datos de todas las subsecciones de las páginas web requiere un enfoque centrado en la recopilación de datos. Para lograrlo se desarrolló un script de web scraping que rastrea las diferentes páginas web del portal con restricciones de criterios como el año y el ministerio. Las ventajas de utilizar un script son la automatización de la extracción de datos y la precisión fiable, lo que da confianza a la hora de analizar los datos extraídos.

La metodología de *web scraping* consta de dos fases: extractor de formularios, resultados de contratos y descarga de archivos. La razón de dividir el scraping en dos fases es mantener la integridad de los datos entre las distintas etapas del flujo de datos dentro del sitio web. Cada fase del *scraping* sirve de entrada para la siguiente y, por tanto, debe introducirse con precisión en el flujo de datos. El propio sitio web, además de tener problemas de accesibilidad, plantea dificultades debido a su compleja estructura HTML. Por tanto, resolver estos problemas fue una prioridad antes de iniciar el proceso de *web scraping*. Asimismo, comprender la canalización de datos permite estructurar adecuadamente el complejo formato del sitio web y ayuda a extraer datos para su análisis.

Extracción de formularios

En esta sección se describe el proceso de extracción de formularios. La primera página del sitio web Nicaragua Compra es un elemento de formulario HTML que recoge las distintas opciones y las envía para recargar la página de resultados (Figura 6). Esto añade complejidad al proceso porque los elementos de formulario no son estructuras que se puedan extraer; además, la recarga y el reenvío de la página a través de la API tampoco se pueden *scrapear* para extraer resultados. Por lo tanto, es necesario establecer una forma diferente de tratar el elemento de formulario.

Para seleccionar los años y el ministerio, se creó un nuevo script en Python que recibe del usuario los datos de los años que se van a rastrear y los detalles del ministerio. Junto con los dos criterios mencionados hay 4 casillas de verificación que también deben tenerse en cuenta en el formulario: a) válido; b) rendimiento; c) en evaluación; y d) adjudicado. Estas opciones corresponden a las distintas fases del contrato que son analizadas. Aunque hay un total de 7 opciones, estas 4 son las más relevantes para el proyecto, porque son las categorías en las que los fondos monetarios están siendo o fueron ejecutados. Tras cargar estos datos en el script extractor de formularios, el código se ejecuta para enviar las opciones y obtener la página de resultados de la API del sitio web. Una vez que la página de resultados carga el año y el ministerio correcto, es posible ver los detalles de cada contrato y sus respectivos documentos.

Figura 4. Pantalla de inicio - búsqueda

La imagen muestra una interfaz de usuario para la búsqueda de licitaciones. En la parte superior, hay un menú desplegable para 'Ordenar por:' con 'Fecha de Cierre' seleccionado, y un campo 'Resultados por página:' con el valor '10'. El formulario principal contiene los siguientes campos:

- Nombre de la Adquisición: [campo de texto]
- Número de Adquisición: [campo de texto]
- Código SIGAF: [campo de texto]
- Ejercicio: [menú desplegable con '2021']
- Procedimiento: [menú desplegable con 'Todos']
- Tipo de Procedimiento: [menú desplegable con 'Todos']
- Clasificación Comercial: [botón 'Seleccionar Rubro'] y [enlace 'Presione aquí para buscar en Catálogo de Bienes y Servicios']
- Institución: [menú desplegable con 'Ministerio de Transporte e Infraestructura']
- Departamento: [menú desplegable con 'Todos']
- Departamento de Lugar de Entrega: [campo oculto]
- Estado: [checkboxes para 'Vigente', 'Ejecución', 'Cancelado', 'En Evaluación', 'Desierto', 'Adjudicado', 'Suspendido']

En la parte inferior del formulario, hay dos botones: 'Limpiar' y 'Buscar', y un botón 'Ir a Inicio' con un icono de casa.

Fuente: Captura de pantalla tomada de nicaraguacompra.gob.ni

Un análisis exploratorio para detectar patrones de corrupción

Resultados de los contratos

El flujo de datos se desplaza a la página de resultados desde el extractor de formularios y muestra los diferentes contratos/licitaciones ganadas por las partes para una duración específica bajo los ministerios nacionales seleccionados. Cada uno de los contratos de licitación contiene detalles como la fecha del contrato ganado, el importe emitido, los documentos relacionados con la licitación, el código único SIGAF/Línea, la fecha de publicación, el estado de la licitación y la fuente de financiación. Estos resultados contractuales son diferentes para cada licitación. Especialmente para los contratos en el MTI es posible identificar las diferentes licitaciones por su código único SIGAF y para el CSE pueden ser identificadas a través de sus números de "Línea".

El proceso para raspar la segunda fase del proyecto es mediante el uso de las librerías proporcionadas por Python para *web scraping*. Se probó el código tanto con Scrapy como con BeautifulSoup, decidiéndose implementar el código completo con este último en integración con Selenium. Se seleccionó BeautifulSoup debido a su popularidad y facilidad de uso, y la biblioteca Selenium debido a que permite automatizar la interacción de la página web. En esta sección específica, se utiliza un script para añadir los resultados de cada contrato a un diccionario y para evitar interferencias manuales o errores humanos, además, se utiliza la lógica de paginación para

navegar por las páginas de resultados. Inmediatamente después de entrar en la página de la lista de resultados, se muestra un resumen de todos los contratos tramitados para el año que depende del ministerio (Figura 5).

Figura 5. Resumen contratos

Procedimientos vigentes:	0
Procedimientos en evaluaciones:	1496
Procedimientos adjudicados:	2338
Procedimientos en ejecución:	37598
Procedimientos de ciertos:	2567
Procedimientos suspendidos:	30

Fuente: Captura de pantalla tomada de nicaraguacompra.gob.ni

Estas cifras dan una estimación de cuántas licitaciones se necesita extraer y en qué estado se encuentran. Además, las estadísticas sirven como punto de referencia para comprobar los resultados del *web scraping*.

A continuación, el sitio web muestra por defecto 10 resultados a partir de los criterios de búsqueda, la herramienta selecciona 100 resultados por página. El siguiente paso del proceso consiste en navegar por cada una de las páginas de detalles del contrato para extraer datos sobre la licitación (Figura 6 y 7). Para ello, se navega hasta el botón "Más datos", que lleva a los detalles de cada contrato. En esta etapa, comienza el raspado de las distintas adjudicaciones, órdenes de compra y documentos.

Figura 6. Datos licitación

Resultados						
Criterio de Búsqueda:		2021 + Ministerio de Transporte e Infraestructura + Vigente + Ejecución + En Evaluación + Adjudicado				
172 resultado/s obtenidos. Para ajustar el criterio de búsqueda, presione Búsqueda Avanzada .						
Ordenar por:		Fecha de Cierre	Resultados por página:		10	
Procedimiento	Detalles					Acciones
Licitación o Concurso Público Internacional BCIE 40811/2021	Estado: Adjudicado	Código SIGAF: #2021-013001-000001	Publicación: 14/01/2021	Cierre: 01/03/2021 09:00:00 AM	Última Actualización: 09/06/2021 04:46:04 PM	Más Datos
Ministerio de Transporte e Infraestructura (MTI) - Division de Adquisición MTI Construcciones y Mejoras EN PROCESO de Bienes de Dominio Público (95901002) Ampliación de la Carretera Empalme Las Flores ¿ Catarina - Guanacaste						
CONTRATACION MENOR 42662/2021	Estado: Ejecución	Código SIGAF: #2021-013001-000032	Publicación: 22/02/2021	Cierre: 24/02/2021 03:00:00 PM	Última Actualización: 09/03/2021 02:33:31 PM	Más Datos
Ministerio de Transporte e Infraestructura (MTI) - Division de Adquisición MTI Mantenimiento y Reparación de A/A de Vehículos (78180109) Servicio de mantenimiento de aires acondicionados para dos vehículos de la DGTA						
CONTRATACION MENOR 42808/2021	Estado: Ejecución	Código SIGAF: #2021-013001-000031	Publicación: 22/02/2021	Cierre: 24/02/2021 02:00:00 PM	Última Actualización: 09/03/2021 02:33:07 PM	Más Datos
Ministerio de Transporte e Infraestructura (MTI) - Division de Adquisición MTI Juntas de cardán (25173810), Soportes o conjuntos del eje (26111532), Zapatas de freno de tambor (25171712), Buje de automóvil (25172009), Sistema de dirección hidráulica (25174204), Sistemas de frenado para automóviles (25171702), Bombas hidráulicas (40151533) Adquisición de repuestos para vehículo asignado a la Dirección General de Transporte Acuático						

Fuente: Captura de pantalla tomada de nicaraguacompra.gob.ni

Figura 7. Cambio número resultados

Resultados						
Criterio de Búsqueda:		2021 + Ministerio de Transporte e Infraestructura + Vigente + Ejecución + En Evaluación + Adjudicado				
172 resultado/s obtenidos. Para ajustar el criterio de búsqueda, presione Búsqueda Avanzada .						
Ordenar por:		Fecha de Cierre		Resultados por página: 100		
Procedimiento	Detalles					Acciones
Licitación o Concurso Público Internacional BCIE 40811/2021	Estado: Adjudicado	Código SIGAF: #2021-013001-000001	Publicación: 14/01/2021	Cierre: 01/03/2021 09:00:00 AM	Última Actualización: 09/06/2021 04:46:04 PM	Más Datos
Ministerio de Transporte e Infraestructura (MTI) - Division de Adquisición MTI Construcciones y Mejoras EN PROCESO de Bienes de Dominio Público (95901002) Ampliación de la Carretera Empalme Las Flores ¿ Catarina - Guanacaste						
CONTRATACION MENOR 42662/2021	Estado: Ejecución	Código SIGAF: #2021-013001-000032	Publicación: 22/02/2021	Cierre: 24/02/2021 03:00:00 PM	Última Actualización: 09/03/2021 02:33:31 PM	Más Datos
Ministerio de Transporte e Infraestructura (MTI) - Division de Adquisición MTI Mantenimiento y Reparación de A/A de Vehículos (78180109) Servicio de mantenimiento de aires acondicionados para dos vehículos de la DGTA						
CONTRATACION MENOR 42808/2021	Estado: Ejecución	Código SIGAF: #2021-013001-000031	Publicación: 22/02/2021	Cierre: 24/02/2021 02:00:00 PM	Última Actualización: 09/03/2021 02:33:07 PM	Más Datos
Ministerio de Transporte e Infraestructura (MTI) - Division de Adquisición MTI Juntas de cardán (25173810), Soportes o conjuntos del eje (26111532), Zapatas de freno de tambor (25171712), Buje de automóvil (25172009), Sistema de dirección hidráulica (25174204), Sistemas de frenado para automóviles (25171702), Bombas hidráulicas (40151533) Adquisición de repuestos para vehículo asignado a la Dirección General de Transporte Acuático						

Fuente: Captura de pantalla tomada de nicaraguacompra.gob.ni

En la siguiente pantalla (Figura 8), la herramienta extrae todos los datos de la licitación como la entidad oferente, la unidad de adquisiciones, la clasificación comercial, la categoría, la fuente de financiamiento, la norma aplicable, el tipo de procedimiento, la modalidad procedimental y el estado de la licitación. Además, se extraen los datos de la adjudicación como el adjudicatario, el monto del contrato y el beneficiario final. Posteriormente, se ingresa a la pestaña donde se ubica la orden de compra (Figura 9) y se pueden observar todas las órdenes de compra relacionadas a la adjudicación y su estado. En este paso, se extrae principalmente el número de ejercicio de cada orden de compra. Finalmente, la herramienta se devuelve a la página de resultados y repite el proceso con la siguiente licitación en lista hasta finalizar con todos los resultados para ese criterio de búsqueda.

Figura 8. Datos adjudicación

Licitación Pública Nacional Bienes Nro 2021-013001-000131 / 2021
Servicios de Reproducción de Materiales de Uso Didáctico

Entidad: MTI - Ministerio de Transporte e Infraestructura
 Unidad de Adquisiciones: MTI - Division de Adquisición MTI
 Clasificación Comercial: Impresión digital
 Categoría: Servicios
 Fuente de Financiamiento: Banco Interamericano de Desarrollo -BID-
 Norma Aplicable: BANCO INTERAMERICANO DE DESARROLLO
 Tipo de Procedimiento: Licitación Pública Nacional Bienes
 Modalidad Procedimental: SIN MODALIDAD
 PrePBC., Estudio Asociado: -
 Vinculación Plan Anual: 2021-013001-000499
 Estado: **EJECUCIÓN**

Proveedores que Reciben Información: 2

Documentos Convocatoria Orden de Compra

RESUMEN DE ADJUDICACIONES

Proveedor-RUC	Representada	Monto Adjudicado	Renglones Adjudicados	NBD / Beneficiario Final
PRODUCTIVE BUSINESS SOLUTIONS NICARAGUA, S.A.- J0310000006202		C\$ 795,548.52	0	3MZHYS / PAUL BARNABY SCOTT

Observaciones: -
 Renglones desiertos: -
 Para mas detalles presione "Ver Documentos".

Fuente: Captura de pantalla tomada de nicaraguacompra.gob.ni

Figura 9. Datos orden de compra

Licitación Pública Nacional Bienes Nro 2021-013001-000131 / 2021
Servicios de Reproducción de Materiales de Uso Didáctico

Entidad: MTI - Ministerio de Transporte e Infraestructura
 Unidad de Adquisiciones: MTI - Division de Adquisición MTI
 Clasificación Comercial: Impresión digital
 Categoría: Servicios
 Fuente de Financiamiento: Banco Interamericano de Desarrollo -BID-
 Norma Aplicable: BANCO INTERAMERICANO DE DESARROLLO
 Tipo de Procedimiento: Licitación Pública Nacional Bienes
 Modalidad Procedimental: SIN MODALIDAD
 PrePBC., Estudio Asociado: -
 Vinculación Plan Anual: 2021-013001-000499
 Estado: **EJECUCIÓN**

Proveedores que Reciben Información: 2

Documentos Convocatoria Adjudicación

Ordenes de Compra

Nº/Ejercicio	Proveedor/RUC	Monto	Estado	Detalles	NBD / Beneficiario Final
2021-013001-000166 / 2021	PRODUCTIVE BUSINESS SOLUTIONS NICARAGUA, S.A.- J0310000006202	C\$ 795,548.52	FINALIZADA	GRP	

Para mas detalles presione "Ver Documentos".

Volver

Fuente: Captura de pantalla tomada de nicaraguacompra.gob.ni

Como resultado del *web scraping* se obtuvo una base de datos por institución. Debido a que la naturaleza de ambas instituciones es distinta y a la diferencia en número de licitaciones y monto ejecutado, se decidió realizar el análisis en bases de datos separadas. Sin embargo, se aplicó el mismo procedimiento y mismas técnicas a cada institución. A continuación, se presenta una breve descripción de la limpieza y procesamiento de los datos¹.

El análisis empieza con la limpieza y preparación de los datos (Figura 10), porque esto puede afectar los resultados. Lo primero fue reducir el número de columnas a través de un proceso denominado *melting o pivoting*. Se observó que había una secuencia de 5 columnas que se repetía 9 veces, lo que resultaba en 45 columnas que podía reducirse² a 5. Estas 45 columnas (en su mayoría NaN) contienen información como monto adjudicado, número de Registro Único de Contribuyente (RUC), nombre comercial del proveedor y nombre de beneficiario final.

Figura 10. Limpieza y preparación de los datos

```

1 MTI ANALYSIS

1.1 Data Loading and Preliminary Processing

In [2]: import pandas as pd
import unicodecode

In [3]: # load the 'MTI' sheet from the combined_data.xlsx file into a dataframe
df = pd.read_excel('combined_data.xlsx', sheet_name='MTI')

In [7]: # Set the options to display all columns
pd.set_option('display.max_columns', None)

In [4]: df.shape
Out[4]: (5641, 68)

In [5]: df.columns
Out[5]: Index(['PageNo', 'P_contract', 'P_no', 'M_Estado', 'M_SIGAF', 'M_Publicacion',
'M_Cierre', 'M_Ultima', 'M_Body1', 'M_Body2', 'M_Body3', 'M_Datos',
'D_EntidadId', 'D_UnidadId', 'D_ClasificacionId', 'D_CategoriaId',
'D_FuenteDeFinanciamientoI', 'D_NormaAplicableId',
'D_TipoProcedimientoId', 'D_ModalidadId', 'D_EstudipBCId',
'D_VinculacionId', 'D_Estado', 'D_ADJ_ProveedorNombre_0',
'D_ADJ_ProveedorRUCColV_0', 'D_ADJ_Monto_0', 'D_ADJ_RenglonColV_0',
'D_ADJ_BeneficiarioFinal_0', 'D_ADJ_ProveedorNombre_1',
'D_ADJ_ProveedorRUCColV_1', 'D_ADJ_Monto_1', 'D_ADJ_RenglonColV_1',
'D_ADJ_BeneficiarioFinal_1', 'D_ADJ_ProveedorNombre_2',
'D_ADJ_ProveedorRUCColV_2', 'D_ADJ_Monto_2', 'D_ADJ_RenglonColV_2',
'D_ADJ_BeneficiarioFinal_2', 'D_ADJ_ProveedorNombre_3',
'D_ADJ_ProveedorRUCColV_3', 'D_ADJ_Monto_3', 'D_ADJ_RenglonColV_3',
'D_ADJ_BeneficiarioFinal_3', 'D_ADJ_ProveedorNombre_4',
'D_ADJ_ProveedorRUCColV_4', 'D_ADJ_Monto_4', 'D_ADJ_RenglonColV_4',
'D_ADJ_BeneficiarioFinal_4', 'D_ADJ_ProveedorNombre_5',
'D_ADJ_ProveedorRUCColV_5', 'D_ADJ_Monto_5', 'D_ADJ_RenglonColV_5',
'D_ADJ_BeneficiarioFinal_5', 'D_ADJ_ProveedorNombre_6',
'D_ADJ_ProveedorRUCColV_6', 'D_ADJ_Monto_6', 'D_ADJ_RenglonColV_6',
'D_ADJ_BeneficiarioFinal_6', 'D_ADJ_ProveedorNombre_7',
'D_ADJ_ProveedorRUCColV_7', 'D_ADJ_Monto_7', 'D_ADJ_RenglonColV_7',
'D_ADJ_BeneficiarioFinal_7', 'D_ADJ_ProveedorNombre_8',
'D_ADJ_ProveedorRUCColV_8', 'D_ADJ_Monto_8', 'D_ADJ_RenglonColV_8',
'D_ADJ_BeneficiarioFinal_8'],
dtype='object')

```

Fuente: Elaboración propia

1 Para un acceso detallado, remitirse a los códigos de Python en el siguiente repositorio Github.

2 Es importante especificar que reducir se refiere a agrupar los valores de estas columnas bajo las cinco columnas principales.

Un análisis exploratorio para detectar patrones de corrupción

Para la reducción de las columnas, se agruparon los datos por columnas no repetidas y se extrajeron los valores de las columnas repetidas para cada grupo (Figura 11). Luego, se crearon nuevas filas con los valores extraídos y se agregaron a un nuevo *data frame*. Así se garantizó que no se perdieran datos y que cada fila tenga información completa sobre cada contratación pública.

Figura 11. Agrupación de datos por columnas

```
In [6]: # Define non-repeated columns
non_repeated_cols = ['Procedimiento', 'P_no', 'Estado_1', 'Codigo_SIGAF', 'Fecha_Publicacion', 'Fecha_Cierre',
                    'Ultima_Actualizacion', 'M_Body1', 'Actividad_Economica', 'Descripcion_Contratacion',
                    'Institucion', 'D_ClasificacionId', 'Categoria', 'Fuente_Financiamiento', 'Norma_Aplicable',
                    'Tipo_Procedimiento', 'Vinculacion_PAC', 'Estado_2']

# Group by non-repeated columns
groups = df.groupby(non_repeated_cols, dropna=False)

In [13]: # Extract values from repeated columns for each group and assign to new rows
new_rows = []
for name, group in groups:
    # Check if any column in the sequence has data
    has_data = False
    for i in range(9):
        if not group[f'D_ADJ_Monto_{i}'].isna().all():
            has_data = True
            break

    # Add new row if any column in the sequence has data
    if has_data:
        for i in range(9):
            proveedor = group[f'D_ADJ_ProveedorNombre_{i}'].iloc[0]
            ruccolv = group[f'D_ADJ_ProveedorRUCColV_{i}'].iloc[0]
            monto = group[f'D_ADJ_Monto_{i}'].iloc[0]
            if pd.isna(monto):
                continue
            renglones = group[f'D_ADJ_ReglonesColV_{i}'].iloc[0]
            beneficiario = group[f'D_ADJ_BeneficiarioFinal_{i}'].iloc[0]
            new_row = {col: group[col].iloc[0] for col in non_repeated_cols}
            new_row['Proveedor'] = proveedor
            new_row['RUCColV'] = ruccolv
            new_row['Monto'] = monto
            new_row['ReglonesColV'] = renglones
            new_row['BeneficiarioFinal'] = beneficiario
            new_rows.append(new_row)

# Convert list of new rows to dataframe
new_df = pd.DataFrame(new_rows)
```

Fuente: Elaboración propia

El resultado del *melting* permite llevar a cabo un proceso adecuado al momento de aplicar un análisis de machine learning. Sin embargo, esto tiene un *trade-off*, porque incrementa la complejidad de la data al crear (en algunos casos) duplicación en información referente a información general única de los contratos. Esta decisión solo afecta al conteo de procedimientos anuales únicos, inflándolos un poco menos del 15% en el caso del MTI y 38% para el CSE. Para obtener la duración de cada proceso, se estandarizó la fecha de publicación y fecha de cierre a formato date time, la cantidad de días entre una fecha y otra se obtuvo utilizando la fórmula *networkdays*, ya que esta calcula días hábiles netos y no días calendarios.

Se observó que en la columna correspondiente a los montos otorgados se presentaban cantidades en córdobas y en dólares. Los montos en dólares necesitaban convertirse a moneda local para un análisis adecuado; para ello, se creó una función en Python capaz de convertir un valor monetario de la moneda original a C\$ utilizando el tipo de cambio para el año dado (Figura 12). La tasa de conversión se obtuvo del sitio web del Banco Central de Nicaragua (BCN). La data disponible consistía en la tasa de cambio diario desde el 2007 hasta el 2022. Se decidió calcular la media anual y utilizar ese monto como tasa de cambio anual.

Figura 12. Función para convertir valores de US\$ a C\$

```
# Create Python dictionary where the keys are the 'year' column of a pandas DataFrame named 'data'
#and the values are the 'rate' column of the same DataFrame
exchange_rates = dict(zip(data['year'], data['rate']))

#Double Check US$ and C$ are the only currency symbols
currencies = df['Monto'].str[:2].unique()
currencies
...

In [4]: # Split the Monto column into currency and amount columns
df[['moneda', 'monto']] = df['Monto'].str.split(n=1, expand=True)

In [5]: # Delete commas from values in 'monto'
df['monto'] = pd.to_numeric(df['monto'].str.replace(',', ''), errors='coerce')

In [6]: # Convert 'Monto' column to numeric type
df['monto'] = pd.to_numeric(df['monto'], errors='coerce')

In [7]: def convert_currency(value, currency, year, exchange_rates):
    """
    Convert a monetary value from the original currency to C$ using the exchange rate
    for the given year.

    Args:
        value (float or int): The monetary value to convert.
        currency (str): The currency symbol for the original currency.
        year (int): The year of the exchange rate to use for the conversion.
        exchange_rates (dict): A dictionary containing exchange rates for each year.

    Returns:
        float: The converted monetary value in C$, or the original value if the currency
        is already in C$.

    Raises:
        ValueError: If the value is not a float or integer, the currency symbol is not
        recognized, or there is no exchange rate available for the given year.
    """
    if not isinstance(value, (int, float)):
        raise ValueError(f"Invalid value: {value}")
    elif currency == 'C$':
        return value
    elif currency == 'US$':
        exchange_rate = exchange_rates.get(year, None)
        if exchange_rate:
            return value * exchange_rate
        else:
            raise ValueError(f"No exchange rate found for year {year}")
    else:
        raise ValueError(f"Invalid currency symbol: {currency}")

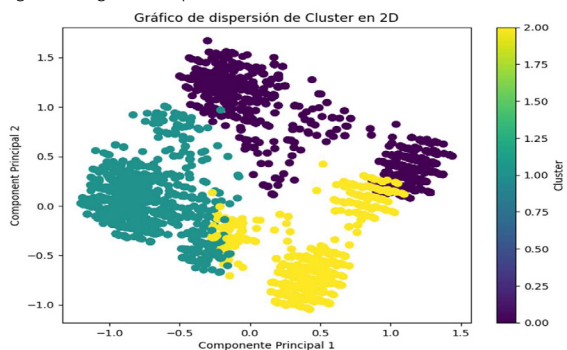
In [10]: # Apply currency conversion to the Amount column
df['monto'] = df.apply(lambda row: convert_currency(row['monto'], row['moneda'],
                                                    row['Fecha_Publicacion'].year, exchange_rates), axis=1)
```

Fuente: Elaboración propia

Tras haber eliminado algunas columnas innecesarias y renombrado otras, se procedió a realizar un *cluster analysis*. El objetivo era generar agrupaciones basadas en el tipo de procedimiento, categoría, proveedor y monto. Se aplicaron 3 modelos de cluster distintos con algunas modificaciones para mejorar su puntuación. En el primer intento se llevó a cabo una codificación binaria para las variables categóricas (procedimiento, categoría, proveedor), se normalizaron las variables numéricas y se combinaron ambas en un solo conjunto de datos. Luego, se utilizó el método de "Elbow" para encontrar el número óptimo de clusters (K) y se aplicó el algoritmo de *clustering K-Means*. Posteriormente, se agregó la columna del cluster encontrado a los datos originales y se visualizaron usando Principal Component Analysis en una figura de dispersión de 2 dimensiones. Finalmente, se calculó el puntaje de Silhouette para evaluar la calidad del *clustering*.

Figura 13. Figura de dispersión

Figura 15. Figura de dispersión



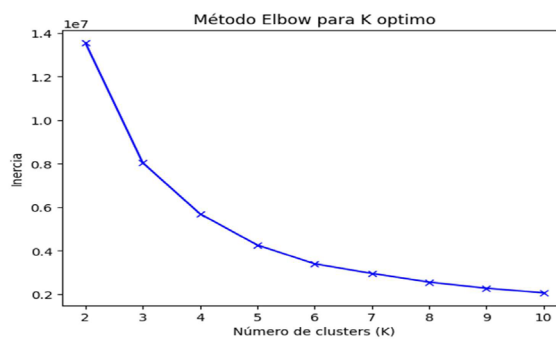
Fuente: Elaboración propia

En la Figura 13 se puede observar que algunos de los *data points* están mal agrupados. Precisamente, se decidió evaluar su desempeño con el puntaje promedio de Silhouette, que va de -1 a 1; entre más cercano a 1 sea una puntuación, indica agrupaciones mejor definidas. Para este modelo el puntaje fue 0.13, este resultado es considerado bajo.

El segundo modelo fue un *cluster* aglomerativo. Para ello, se utilizó LabelEncoder para las características categóricas y StandardScaler para las características numéricas. Luego se calculó la matriz de distancia de Gower³ utilizando la biblioteca Gower. Después de establecer los elementos diagonales de la matriz de distancia en 0, se buscó el número óptimo de clústeres utilizando la puntuación de Silhouette. La figura de puntuación de Silhouette se trazó para diferentes valores de K. Se eligió el valor óptimo de K según la observación de la figura. Luego, se ajustó el modelo final de *clustering* aglomerativo con K óptimo. Se calculó la puntuación de silueta promedio y se creó la matriz de enlace y el dendrograma para visualizar los *clusters*. Este modelo tuvo una puntuación de 0.07, un delta de 46.15% menos que el modelo anterior. Por tanto, este modelo no tenía ningún uso.

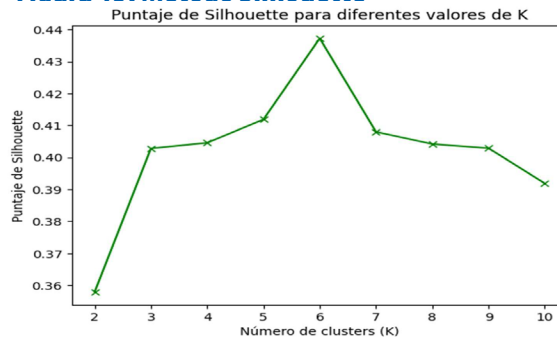
Considerando lo anterior, se decidió iterar una vez más sobre el primer modelo, pero esta vez cambiando algunos parámetros para mejorarlo. Se tomaron las mismas variables categóricas y numéricas. Sin embargo, esta vez se transformaron con OneHotEncoder y StandardScaler, respectivamente. Posteriormente, se realizó una reducción de dimensionalidad de los datos utilizando t-SNE para obtener solo dos dimensiones, lo que permite visualizar los clusters en un plano. Posteriormente, como se observa en la Figura 14 y 15, se utilizó el método Elbow y Silhouette Scores para encontrar el número óptimo de *clusters*.

Figura 14. Método Elbow



Fuente: Elaboración propia

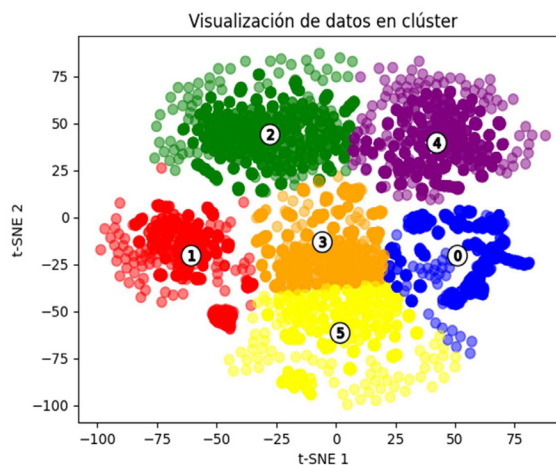
Figura 15. Método Silhouette



Fuente: Elaboración propia

Se tomó el valor más alto del puntaje de Silhouette para decidir el número de clusters. En este caso $k=6$, se aplicó el modelo de KMeans con este número de clusters y se obtuvieron las etiquetas de los clusters. Con estas etiquetas, se graficaron los puntos en el plano t-SNE, coloreándolos de acuerdo con su asignación de cluster.

Figura 16. Visualización de datos en clúster



Fuente: Elaboración propia

Finalmente, tal y como se observa en la Figura 16, se contó el número de puntos en cada cluster y se asignó la etiqueta correspondiente a cada punto en la base de datos original. El puntaje de Silhouette para este modelo fue de 0.43, es decir 30 puntos base sobre el primer modelo. Este resultado es moderado.

³ La distancia de Gower es una medida de similitud que se utiliza para comparar objetos o registros con variables mixtas, es decir, variables que pueden ser tanto numéricas como categóricas.

Inteligencia de fuentes abiertas y ciencia de datos

En articulación con estos resultados, se elaboró una serie de preguntas con el objetivo de identificar algunas banderas rojas sobre los procesos de licitación y adjudicación. Estas fueron:

- ¿Cuáles son los proveedores con el mayor número de licitaciones? ¿Hay algún proveedor con un número inusualmente alto de licitaciones en comparación con otros?
- ¿Cómo se distribuyen las licitaciones por categoría de actividad económica?
- ¿Existe alguna correlación entre el monto adjudicado y la duración del proceso de licitación (desde la fecha de publicación hasta la fecha de cierre)?
- ¿Cuáles son los proveedores más frecuentes en las licitaciones y cuál es el monto total adjudicado a cada uno?
- ¿Cuál es la proporción de licitaciones por tipo de procedimiento (por ejemplo, licitación restringida, abierta, etc.)?
- ¿Cuáles son las categorías de contratación con los montos adjudicados más altos en promedio?
- ¿Existen licitaciones con plazos de cierre muy cortos o largos en comparación con otros en la misma categoría?
- ¿Cuántas licitaciones ha ganado cada proveedor a lo largo del tiempo? ¿Hay algún proveedor que destaque en comparación con otros?
- ¿Cuáles son las principales fuentes de financiamiento para los contratos adjudicados a cada proveedor?
- ¿Existen proveedores que participen en licitaciones de múltiples categorías de contratación?
- ¿Cuál es la duración promedio del proceso de licitación para cada proveedor? ¿Hay proveedores con tiempos de adjudicación notablemente más cortos o largos que otros?

Las preguntas fueron respondidas de forma visual con el uso de Tableau, las visualizaciones se comparten en la siguiente sección.

7.1 Ministerio de Transporte e Infraestructura (MTI)¹

El modelo de *cluster* con mejor calificación logró separar seis clusters y agruparlos moderadamente. Una de las debilidades del modelo puede estar altamente relacionada con que dentro de los valores únicos del tipo de procedimiento había 47 registros, muchos de los cuales tenían el mismo nombre, pero con errores de registro, por ejemplo, Licitación Pública con y sin tilde, en letras minúsculas y mayúsculas. Asimismo, concatenadas con otras palabras como licitación pública internacional, entre otros. Esta super dimensionalidad de la categorización se presenta errada desde la fuente, es decir son malas digitaciones o categorizaciones por parte del equipo encargado de subir la data al sitio Nicaragua Compra.

Se intentó subsanar esta super dimensionalidad innecesaria haciendo uso de las categorías establecidas por la Ley No. 737. Se creó la función `group_procedure`, que toma una fila del *data frame* que contiene información sobre un procedimiento de contratación y categoriza ese procedimiento en una de las siguientes cinco categorías: a) Licitación Pública; b) Licitación Selectiva; c) Contratación Simplificada; d) Contrataciones Menores; o e) Por Concurso. La función utiliza tanto el nombre del procedimiento como el monto del contrato para determinar la categoría. En caso de que el monto del contrato sea mayor o igual a C\$3,000,000, la categoría será Licitación Pública, mientras que, si está entre C\$500,000 y C\$3,000,000, será Licitación Selectiva.

Si el procedimiento no se clasifica en ninguna de las categorías anteriores, se asigna a la categoría "Other". La función devuelve la categoría correspondiente a la fila dada. El resultado eran 4 categorías: Licitación Pública, Licitación Selectiva, Por concurso y Other. El inconveniente radica en que, con alta probabilidad, la categorización original no respeta los montos establecidos por la ley. Como resultado, la recategorización termina asignando aproximadamente el 97% del monto total licitado a la modalidad de licitación pública. Esto iba a significar un enorme sesgo al momento de correr un nuevo modelo, así que finalmente se decidió dejar el modelo en el que el procedimiento tenía una super dimensionalidad redundante. La premisa es que valía la pena dejar la categorización original y correr el modelo sobre esta data para ver la posibilidad de agrupar por similitudes.

En el siguiente *treemap* se puede apreciar la visualización de los *clusters*. Los valores están ordenados de la siguiente forma: tipo de procedimiento, monto en córdobas, número de procedimientos, promedio de días de duración y número de cluster al que pertenece.

¹ Para mejor acceso a visualizaciones remitirse al siguiente repositorio Tableau public.

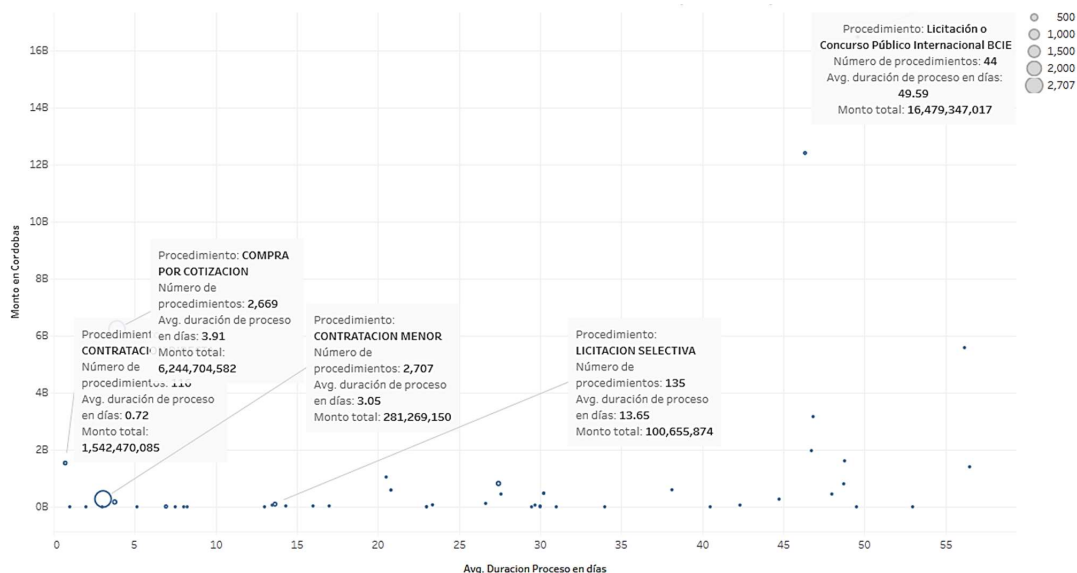
Figura 17. Tipo de procedimiento organizado por cluster²

Compra por cotización 6,244,704,582 2,669 3.91 5	Compra por cotización menor 172,852 2 2	Selección directa - BM 12,906,294 6 3	Concurso 119,674,324 14 26.64	Concurso público internacional 445,203,097 11 48	Consultor individual 34,219,871 25 14.32	Comparación de precios especial obras / BID 65,277,714 4 13.50	Comparación de precios obras 2,702,922 1 13	Concurso firma consultora 484,186,170 40 30.23
	Contratación directa 1,542,470,085 116 0.72	Consultor Individual Banco Mundial 50,182,200 2 16	Licitación pública 35,064,517 3 30	Licitación pública internacional 12,409,008,435 67 46.34	Licitación pública nacional 826,155,058 156 27.44	Licitación restringida 10,356,037 36 23	Licitación selectiva 100,655,874 135 13.65	
Contratación menor 2,707 3.05 4	Comparación de precios bienes y servicios 2,585,871 24 7.50	Licitación o concurso público internacional 5,573,374,497 10 56.20	Licitación pública internacional bienes y servicios 11,901,590 4 49.50	SDOI-Solicitud Oferta InterNacional / BM 14,238,003 2 40.50	SDOI-Solicitud Oferta InterNacional Obras/ BM MINED 816,540,045 4 48.75	SDON- Solicitud Oferta Nacional Obras / BM MTI 286,668,735 4 44.75	Selección basada en el menor costo 1,799,202 2 30	
	Comparación de Precios Especial Bienes/BID 13,381,974 16 8	Licitación o concurso público internacional con oferta financiera 1,608,703,826 14 48.79	Licitación pública internacional obras 3,176,082,718 6 46.83	Selección basada en las calificaciones del consultor 23,782,467 10 17	Solicitud de ofertas internacional 1,424,639,212 2 56.50	Selección de firma consultora BID 454,609,087 26 27.62	Selección de firma consultora BM 74,646,183 12 42.33	Concurso Público Nacional - BCIE MTX 11,378,828 2 31
Contratación simplificada 177,431,010 154 3.77	Licitación pública 606,167,174 19 38.16	Licitación o concurso público internacional limitado 1,971,716,002 4 46.75	Licitación pública nacional bienes 3,621,821 4 29.50	Selección con base en una sola fuente 1,312,170 1 1.00				
Comparación de precios 18,219,635 86 6.93	Materia excluida 44,084 6 5.17	Licitación o concurso público nacional 4,394,874 34 2	Licitación pública nacional obras 67,244,295 3 29.67	Selección basada en calidad y costo 596,429,840 36 20.81	Solicitud de cotización BM 9,862,518 4 8.25	Solicitud de Oferta Nacional / BM 4,105,484 1 53	Licitación estandar - MTX 1,038,527,006 4 20.50	
	Licitación por registro 75,813,712 30 23.37	Licitación o concurso público internacional BCIE 16,479,347,017 44 49.59						

A partir de la visualización anterior, destacan aquellos procedimientos con un bajo promedio en duración de días de proceso y un alto volumen agregado. Véase ‘Compra por Cotización’. En promedio, desde que se publicó la fecha para competir hasta que cerró, cada uno de estos procesos tiene un total de 4 días, con un promedio de C\$2.3 millones cada contrato. Resulta interesante observar estos montos asignados en tan poco tiempo. Asimismo, Contratación Directa tiene un monto de C\$1.5 billones, pero el tiempo promedio no llega ni a un día. El valor promedio de cada contrato sube a un valor de C\$13.2 millones, esto es un monto elevado para ser deliberado en 1 día. De la misma forma, el procedimiento ‘Licitación o Concurso Público Internacional BCIE’ tiene un monto de C\$16 billones. En este caso, lo llamativo es el monto, porque de toda la data disponible en esta muestra, el Banco Centroamericano de Integración Económica (BCIE) financió un total de C\$26 billones, es decir un 46.7% del monto total ejecutado por el MTI entre 2007 y 2022.

2 Para más detalles ver Anexos.

Figura 18. Distribución proveedor, monto y duración de proceso (2007-2022)³



Se inspeccionó la granularidad⁴ de este financiamiento. Específicamente, se analizaron los 25 proveedores con mayor monto distribuido por fuente. En la siguiente gráfica se puede observar que uno de los principales ejecutores de proyectos financiados por el BCIE es la Constructora MECO S.A. Esta constructora ejecutó un total de C\$8.5 billones, es decir un 34.7% del monto total financiado por el BCIE a lo largo del periodo de análisis.

³ Para más detalles ver Anexos.

⁴ Se refiere al nivel de detalle que posee un dato o una estructura de datos

Figura 19. Top 25 proveedores distribuidos por fuente de financiamiento⁵

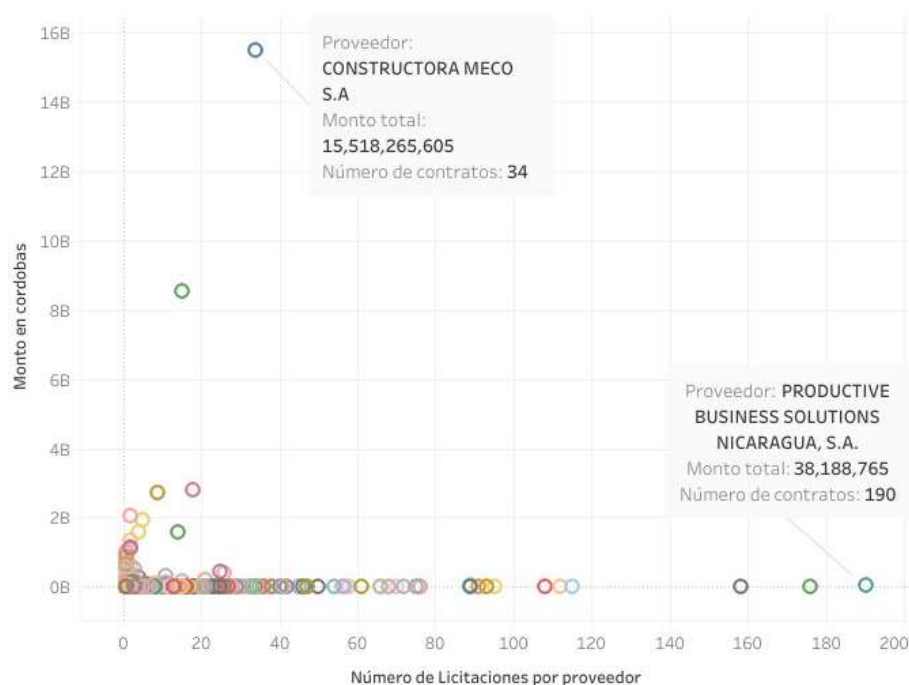
Proveedor	Null	Banco Centroamericano de Integración Económica - BCIE	Banco Interamericano de Desarrollo - BID	Banco Mundial	Gobierno de Nicaragua
CONSTRUCTORA MECO S.A.		8.535.499.127 34.73%	4.057.305.994 29.79%	2.26.716.868 71.09%	417.552.343 12.85%
CONSTRUCTORA SANTA FE LIMITADA SUCURSAL NICARAGUA MEDINA RODRIGUEZ, RENE	2.813.192.73 76.69%	6.030.120.905 24.53%	1.524.014.966 11.19%		991.929.345 30.52%
ASTALDI S.P.A. SUCURSAL NICARAGUA			2.32.203.663 17.05%		271.957.813 8.37%
EMPRESA DE CONSTRUCCIÓN Y TRANSPORTE ETERNA, S.A. DE C.V. - 05019001047930		2.064.504.432 8.40%			
PROFESIONALES DE LA CONSTRUCCION S.A DE C.V		570.393.093 2.32%	1.024.793.332 7.52%	329.172.454 10.51%	
Grupo Pycsa, S.A. DE C.V. - 3-101- 035078-17	393.269.307 10.72%				1.179.807.921 36.30%
B. LANZAS CONTRATISTAS CIVILES SOCIEDAD ANONIMA		741.552.286 3.02%	633.393.531 4.65%		
Constructora Santa fe Limitada -		1.332.449.107 5.42%			
Consortio FCC Américas (FCC Construcción America, S.A.) y Operadora CICSA, s.a. de c.v. -			1.150.700.768 8.45%		
CONSTRUCTORA MAKRO, S.A DE C.V -			1.102.797.564 8.10%		
CONSTRUCTORA MAKRO SOCIEDAD ANÓNIMA DE CAPITAL VARIABLE		1.038.035.122 4.22%			
Llansa Ingerieros S.A -		992.425.885 4.04%			
CONSORCIO MALACATOYA (LLANSA INGENIEROS-D. GUERRERO		875.927.114 3.56%			
Constructora Makro, S.A. de C.V. -		735.437.942 2.99%			
consorcio santa fe-prodemex, s.a. -			620.851.521 4.56%		
CONSORCIO CEMEX-MECO- LLANSA		609.735.908 2.48%			
ASC GROUP INTERNATIONAL CORPORATION SOCIEDAD ANONIMA				576.302.650 18.40%	
Prodecon S.A de C.V -		546.142.002 2.22%			
MIRSA CONASA		506.684.821 2.06%			
Promotora y Desarrolladora Mexicana Sociedad Anónima de Capital Variable			480.939.910 3.53%		
Terraceria y Cimentaciones del Sur, S.A. DE C.V. - 91020	462.025.967 12.59%				
FCC CONSTRUCCION DE CENTROAMERICA, SOCIEDAD ANONIMA			411.297.540 3.02%		
CONSTRUCTORA LATINO Y CIA. LTDA.					389.172.948 11.97%
Profesionales de la Construcción S,A de CV (PRODECOM) -			291.163.339 2.14%		

5 Para más detalles ver Anexos.

Un análisis exploratorio para detectar patrones de corrupción

Llama la atención que un porcentaje tan elevado del volumen adjudicado esté concentrado en una sola empresa. Por lo anterior, se identificó una posible relación entre proveedor, número de licitaciones y monto licitado. Si bien no hay una relación entre dichas variables, lo que queda claro es que hay empresas con una gran cantidad de contratos cuyo monto ejecutado no supera los C\$100 millones y empresas con menos de 35 contratos que superan los C\$15 billones. En este caso, Constructora MECO S.A es nuevamente un valor atípico.

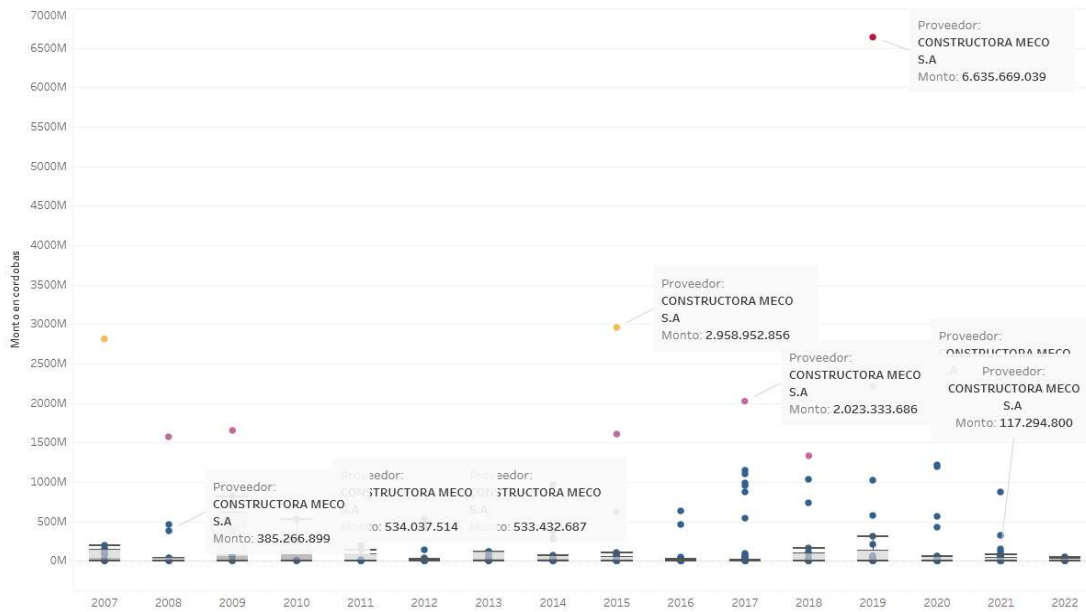
Figura 20. Distribución del Monto en relación con el Número de Procedimientos por Proveedor⁶



En búsqueda de valores atípicos, se graficó un *box plot* para mostrar la mediana, los cuartiles y los valores extremos. En este caso, el monto licitado se observa en el eje Y, y la fecha en el eje X. Se puede identificar que Constructora MECO S.A ha sido un valor atípico por 8 años en total. Los años en los que más dinero ejecutó fueron en 2015 (C\$2.9 billones), 2017 (C\$2 billones), 2018 (C\$6.6 billones) y 2020 (C\$2.3 billones).

⁶ Para más detalles ver Anexos.

Figura 21. Box plot del monto de licitaciones por año (2007-2022)⁷



El hecho de que la constructora MECO sea un valor atípico no necesariamente está relacionado con corrupción, pero es un comportamiento que levanta alertas. Por lo anterior, se hizo una búsqueda complementaria para apoyar el análisis con bibliografía secundaria. Esta empresa estuvo involucrada en escándalos de corrupción en Panamá durante el gobierno de Ricardo Martinelli (2009-2014). De acuerdo con el Centro Latinoamericano de Investigación Periodística, la compañía pagó aproximadamente US\$9,4 millones en sobornos, equivalentes al 5% del costo total de seis obras de infraestructura que realizó en ese país por un monto de al menos US\$1,400 millones (López, 2022). Según el Ministerio Público panameño, estos pagos ilegales buscaban obtener contratos de obra pública o agilizar adelantos oficiales de obra. MECO fue una de las 62 empresas constructoras locales y extranjeras investigadas en el caso Blue Apple, relacionado con blanqueo de capitales y realización de sobornos (López, 2022). A pesar de admitir su participación en el esquema de corrupción, el dueño y fundador de MECO, Carlos Cerdas, llegó a un acuerdo de colaboración con la Fiscalía panameña, lo que permitió a la empresa evitar cargos y seguir participando en concursos de obra pública.

MECO también está involucrada en varios escándalos de corrupción en Costa Rica. En el "Caso Diamante",

el Organismo de Investigación Judicial (OIJ) y la Fiscalía Anticorrupción investigan una posible red de corrupción liderada por esta empresa para sobornar a personeros de ocho municipalidades del país y así ganar contratos de obra pública. Cinco alcaldes han sido detenidos y se investigan 15 contratos de licitación obtenidos a raíz de presuntos sobornos (Meléndez, 2021). La empresa también está involucrada en el "Caso Cochinilla", que investiga una extensa red de supuestos sobornos a funcionarios, que involucran a otras constructoras y al Consejo Nacional de Vialidad (CONAVI). El presidente de MECO, Carlos Cerdas, fue detenido como uno de los principales sospechosos. La trama ha malversado unos \$127 millones y ha sido catalogada como "crimen organizado" (Murillo, 2021).

El Gobierno de Nicaragua ha sido el cliente estatal que más recursos anticipó a la constructora MECO en los últimos cinco años, con un total de \$120,6 millones pagados por adelantado, según los estados financieros de la empresa (Rodríguez, 2021). Entre 2016 y 2020, MECO recibió anticipos de largo plazo por \$303,9 millones en los cinco países⁸ donde opera, Nicaragua aportó el 40% de esos anticipos. La mayoría

7 Para más detalles ver Anexos.

8 Además de Nicaragua, MECO también opera en Costa Rica, Colombia, El Salvador y Panamá.

Un análisis exploratorio para detectar patrones de corrupción

de los proyectos ejecutados por MECO en Nicaragua sus obras viales, siendo común que las instituciones públicas del país estipulan anticipos en sus contratos, posiblemente para recibir ofertas más competitivas.

Ante las denuncias en Costa Rica, MECO intentó sacar parte de su maquinaria hacia Nicaragua, un país en el que ha tenido una fuerte presencia y ha realizado numerosos proyectos⁹. Sin embargo, tras darse a conocer públicamente su intención, MECO desistió de hacerlo y anuló en el sistema las declaraciones de exportación de la maquinaria bajo supervisión de la Aduana de Peñas Blancas en Costa Rica. Esta situación sugiere la existencia de una relación estrecha entre MECO y el gobierno nicaragüense, y cómo la empresa podría haber buscado aprovechar esa relación en momentos de dificultad en otros países (Chacón, 2021).

Astaldi es también otra empresa con contratos adjudicados (C\$2.7 billones). Esta constructora italiana ha estado involucrada en casos de corrupción y comportamientos sospechosos en América Latina. En El Salvador, Astaldi fue parte de un esquema de corrupción en la construcción de la planta hidroeléctrica El Chaparral, lo que llevó al gobierno a decretar un embargo por \$227.6 millones contra la compañía (BN Americas, 2020). El expresidente salvadoreño Mauricio Funes enfrenta cargos por soborno, lavado de dinero y conspiración criminal relacionados con el contrato y pagos excesivos a Astaldi. Además, en Perú, el expresidente Martín Vizcarra fue acusado de recibir sobornos del consorcio Obraisa-Astaldi mientras era gobernador de la región de Moquegua. Astaldi experimentó dificultades financieras y tuvo que entrar en un proceso de protección judicial en 2017. La compañía italiana Webuild S.p.A. adquirió el 65% de Astaldi, rescatándola de la protección judicial (Reina, 2020).

Otra empresa a la que el MTI ha adjudicado contratos y ha estado relacionada con prácticas cuestionables en otros países es Eterna (C\$2.6 billones), compañía dirigida por Alberto Díaz Lobo. En Honduras Eterna ganó un contrato de INVEST-H para desarrollar obras adicionales en el corredor agrícola al ofertar un monto inicial de 144 millones de lempiras, el más bajo entre los competidores. No obstante, durante la etapa final de la firma del contrato,

el monto se incrementó inexplicablemente a 157 millones de lempiras (El Perro Amarillo, 2020a). Adicionalmente, en junio de 2019, Díaz Lobo obtuvo un contrato para la rehabilitación de carreteras en Neteapa y Danlí, valorado en más de 750 millones de lempiras, sin cumplir con la legislación hondureña sobre contrataciones estatales. Estas situaciones reflejan las prácticas dudosas en las que Eterna y su propietario se han visto involucrados en proyectos de infraestructura en el país. Asimismo, esta empresa forma parte de círculos más amplios vinculados con el nepotismo y la corrupción en Honduras (El Perro Amarillo, 2020b).

7.2 Consejo Supremo Electoral (CSE)¹⁰

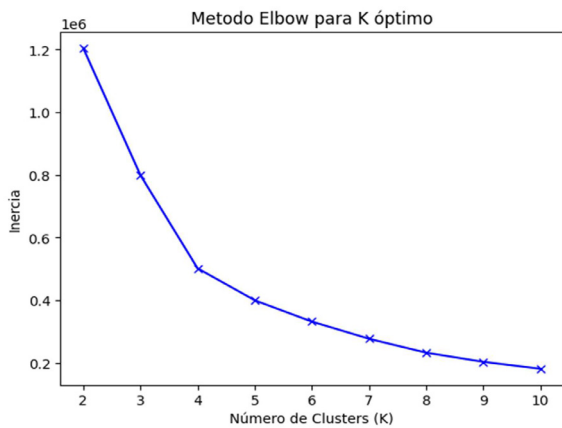
El CSE presenta diferencias significativas con respecto al MTI, tanto en el monto adjudicado como en la distribución de los datos. En relación con los datos, las licitaciones están adecuadamente categorizadas en cuatro grupos: Contratación Simplificada (C\$300 millones), Contratación Menor (C\$96.7 millones), Licitación Selectiva (C\$4.3 millones) y Compra por Cotización (C\$3.5 millones). La suma de estos montos asciende a C\$404 millones para todo el período analizado, en contraste con el MTI, cuyo monto total llega a C\$57 billones. Esto implica que la diferencia porcentual entre ambas cifras es aproximadamente del 99.29%. No obstante, el hecho de que el monto sea menor no implica que no exista la posibilidad de identificar señales de alerta o banderas rojas.

Para analizar al CSE, se aplicó la misma metodología que al MTI. En este caso, el modelo de *clusters* con mejor puntuación fue el que se aplicó en la última iteración al MTI. Se utilizaron 3 variables categóricas (Procedimiento, Categoría y Proveedor) y 2 numéricas (monto, duración). Se utilizó el OneHotEncoder para normalizar la data categórica y el StandarScaler para la data numérica. El método Elbow muestra que el número óptimo de clusters es 4.

9 La constructora ha estado presente en suelo nicaragüense desde 1992, participando en proyectos como la Planta Hidroeléctrica Centroamérica, la Planta Hidroeléctrica de Tumarín y la construcción del Estadio Nacional de Béisbol Dennis Martínez, así como en la construcción y rehabilitación de carreteras y la colocación de capa asfáltica en aeropuertos.

10 Para mejor acceso a visualizaciones remitirse al siguiente repositorio Tableau public

Figura 22. Método Elbow para K óptimo



Fuente: Elaboración propia

Con una puntuación Silhouette de 0.46, se decidió dejar el modelo con estos resultados. El modelo fue capaz de crear cuatro grupos con la siguiente distribución puntual:

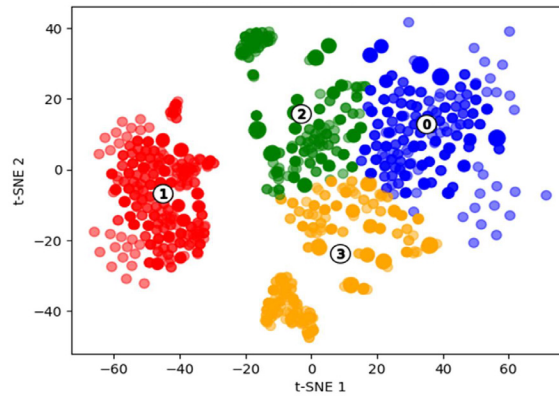
Cluster 0: 485 data points

Cluster 1: 467 data points

Cluster 2: 425 data points

Cluster 3: 432 data points

Figura 23. Visualización de datos en clúster



Fuente: Elaboración propia

La data se exportó en un archivo .xlsx para su visualización en Tableau. Al inspeccionar los datos, se utilizó un treemap para visualizar el tipo de licitación, monto, porcentaje del total, número de procesos y el cluster al que pertenece.

En la siguiente Figura, se observa que las contrataciones simplificadas representan el 74.15% del monto total adjudicado por el CSE. Sin embargo, lo más notable es que esta categoría no solo tiene un bajo número de procesos (131 procesos), sino que el tiempo promedio entre la apertura y el cierre de la licitación es de apenas 3 días. Aunque es lógico que el plazo para este tipo de licitación sea breve, la ley establece condiciones específicas bajo las cuales se deben llevar a cabo estos procedimientos. Al examinar la granularidad de los datos¹¹, los bienes y servicios adquiridos mediante este tipo de procedimiento, a simple vista, no parecen cumplir con la lógica de emergencia y calamidad que establece la No. Ley 737 en el artículo 58. Además, el CSE asignó un 40.8% más de fondos que el MTI a través de la modalidad de contratación simplificada.

11 Dentro de la lista de bienes y servicios se puede observar escáneres, grapadoras, insecticidas, impresoras, barreras de control de infección para operaciones de uso odontológico, softwares y equipo informático.

Figura 24. Tipo de procedimiento, monto, duración promedio (2009-2021)¹²

Tipo de procedimiento	Promedio de duración del proceso	Monto total	Número de procedimientos
Contratación simplificada	3	300.062.695	131
Contratación menor	3	1.219	96.735.140
Licitación selectiva	8	4.329.062	8
Compra por cotización	4	3.520.642	451

Según el Art. 26 de la Ley No. 737, las contrataciones simplificadas deben estar justificadas por situaciones descritas en la ley, por ejemplo: contrataciones en situaciones de emergencia o calamidad pública, cuando hubiere un proveedor único en el mercado y el bien o servicio no pudiese ser sustituido por otro, entre otros. Sin embargo, se puede observar que esta práctica se empezó a implementar a partir de 2018.

Figura 25. Figura 27. Monto por procedimiento (2009-2021)¹³

Procedimiento	Compra por cotización	Contratación menor	Contratación simplificada	Licitación selectiva
2009	760.643			
2010	2.319.258			
2011	440.74	1.721.963		
2012		448.131		
2013		2.481.304		
2014		3.602.707		
2015		1.418.274		
2016		7.317.819		
2017		10.017.997		
2018		12.376.245	117.648.250	
2019		5.428.150	35.050.992	1.303.288
2020		5.092.404	14.254.981	
2021		46.830.144	133.108.542	3.025.775

Es importante señalar que por cuatro años consecutivos esta ha sido una de las formas en las que más se canaliza el dinero de esta institución. ¿Cuáles son los bienes y servicios que se han adquirido en esta calidad, y quiénes son los principales proveedores? Para responder a esta pregunta, el siguiente *tree map* visualiza un desglose de los bienes y servicios adquiridos por contratación simplificada en 2018-2022.

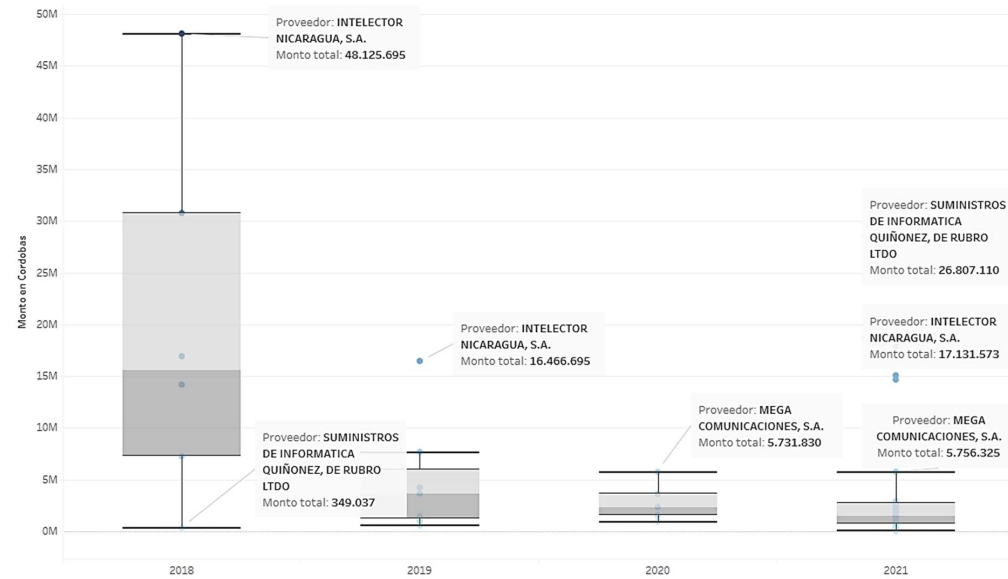
12 Para más detalles ver Anexos.

13 Para más detalles ver Anexos.

Figura 26. Box plot de Montos de Proveedores por año en córdobas¹⁴

Equipo informatico y accesorios 2018 79.067.830	Software 2018 21.652.876		Servidores de computador 2019 13.219.846		Unidades de fotocopiadoras o imágenes 2021 13.219.846		Computadoras de escritorio 2021 7.980.231		
							Computadoras de escritorio 2019 4.310.649		
	Maquinaria, suministros y 2020 12.276.029	Maquinaria, suministros y 2021 4.838.355	Impresoras láser 2021 8.170.194	Software de sistema operativo 2021 7.803.988		Escáneres 2019 7.592.933		Impresión digital 2021 6.128.232	
	Vehículos de motor 2018 16.928.544						Impresión digital 2019		
Servicios de renovación y reparación de edificios 2021 11.987.394			Cable de fibra óptica 2021 7.210.106	Insecticidas 2021 5.667.262	Mantenimiento de impresoras 2021 5.448.166		Máquinas trituradoras de papel o accesorios 2021 5.140.819		
Equipos Informaticos y 2021 6.139.201	Equipo informatico y 2020 1.978.952		Camiones ligeros o vehiculos utilitarios deportivos 2021 5.941.691		Grapadoras 2021 5.030.394	Cinta de enmascarar 2021 2.425.580	Llantas para automoviles o camioneta 2021 2.249.278		
Interruptores de red 2021 21.899.508			Enrutadoras (routers) de red 2019 5.890.544		Barreras de control de infeccion para operaciones de uso 2021 3.788.526		Software de servidor de aplicaciones 2019 1.178.097	Vehiculos todoterrenos o de tracción 2019 998.408	

Figura 27. Box plot de Montos de Proveedores por año en córdobas¹⁵



14 Para más detalles ver Anexos.

15 Para más detalles ver Anexos.

Un análisis exploratorio para detectar patrones de corrupción

Los tres principales bienes adquiridos fueron: Equipos informáticos y accesorios (C\$79 millones), interruptores de red (C\$21.8 millones) y software (C\$21 millones). Dentro de los principales proveedores está Interlector Nicaragua S.A, Mega Comunicaciones S.A, Consultores Especialistas en Sistema S.A y Suministro de Informática Quiñonez.

De acuerdo con *Artículo 66*, la relación entre oferentes como Intellector y el CSE han sido objeto de críticas y sospechas por parte de opositores, esto debido a prácticas poco transparentes en la forma en que el CSE maneja sus contratos (Navarro, 2020). En particular, ha sido cuestionada la necesidad de utilizar intermediarios locales (como TELSSA e Intellector) en la adquisición de productos y servicios de seguridad informática, cuando ese tipo de compras se pueden hacer directamente a los fabricantes y en línea, posiblemente a precios más bajos. La urgencia y rapidez en la compra de estos productos, como el Fortinet 300D, el software antivirus Kaspersky y las licencias de software Sophos, podrían estar justificando sobreprecios que benefician a los propietarios de estas empresas intermediarias en Nicaragua (Navarro, 2020).

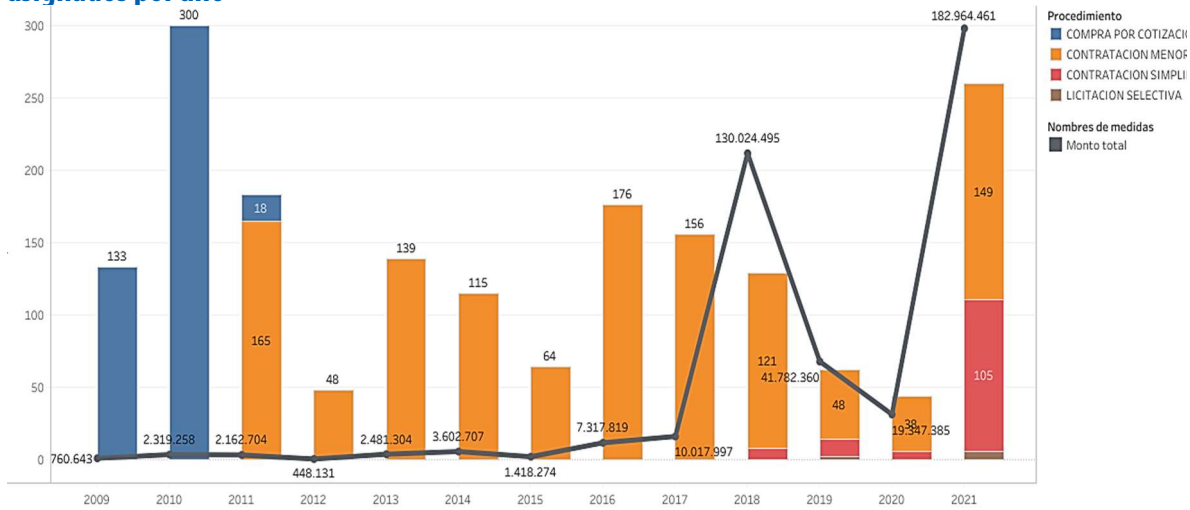
Además, el CSE ha realizado gastos que parecen ser innecesarios y poco claros en diferentes áreas. Por ejemplo, el CSE gastó C\$17 millones en 19 camionetas 4x4, C\$1.2 millones en 14 motos y C\$465 mil en un vehículo marítimo en 2018. Estas compras se llevaron a cabo mediante procesos ilegales, utilizando contrataciones simplificadas bajo la justificación de "situación de urgencia" por las elecciones regionales de la Costa Caribe en 2019. Hay otras compras que llaman la atención en años posteriores. Por ejemplo, la compra de una camioneta de lujo en junio de 2021 bajo la modalidad de contratación simplificada, la cual debe ser utilizada para adquisiciones urgentes en situaciones de emergencia según lo contemplado en la Ley No. 737. El costo de la camioneta fue de 74,000 dólares. Según la documentación consultada por *Confidencial*, la justificación para la compra fue mejorar la flota vehicular del CSE de cara a las elecciones y "asegurar la fluidez del envío y recepción del material electoral". Sin embargo, un experto en contrataciones cuestionó la necesidad de comprar un vehículo de lujo para transportar material electoral y señaló que se habían comprado vehículos en 2018 (Artículo 66, 2020; Enríquez, 2021)

Además, el CSE gastó 20 millones de córdobas en telefonía e internet en 2020, a pesar de que su página web no funciona correctamente, lo que limita el acceso a información pública. El mantenimiento de esta web no funcional representa un gasto considerable sin un beneficio aparente para la ciudadanía y cuyos detalles no están desglosados en los informes oficiales.

La subdivisión de compras es otra práctica preocupante del CSE. Ha sido señalado con anterioridad la existencia de una tendencia a subdividir ilegalmente contratos para que no sean sujetos a los controles de gasto público, lo cual se refleja en la cantidad decreciente de adjudicaciones anuales, tal como lo revela una investigación de Nicaragua Investiga (Orozco, 2020). Por ejemplo, los únicos 2 contratos publicados por el CSE hasta agosto del 2020 corresponden a C\$3.2 millones en uniformes para el personal, que además no fueron adjudicados. Estos contratos, al tratarse de una misma compra de un mismo bien, según la Ley de Contrataciones Administrativas del Sector Público, deberían hacerse en una misma licitación; sin embargo, al subdividir el contrato en dos, se eluden controles pues cada contrato no sobrepasa el umbral de los C\$3 millones de córdobas. No obstante, más allá de eso resulta difícil concebir que el CSE haya podido realizar el grueso de sus actividades de enero a agosto del 2020 sin adjudicar ni un solo contrato, levantando sospechas de que se puede estar incurriendo en subdivisión de contratos a tal punto que nada del gasto de esa institución necesitó ser agregado al portal de compras (Orozco, 2020).

En la Figura 28 se puede observar que las contrataciones menores habían sido la modalidad de compra histórica en el CSE y que a partir de 2018 se empezaron a implementar las licitaciones simplificadas. Adicionalmente, vale la pena resaltar es que el CSE nunca gastó tanto como lo hizo a través de contrataciones simplificadas. Desde 2018 hasta 2021, el monto adjudicado por contratación simplificada fue de C\$300 millones, de un monto global de C\$404 millones. Es decir, 74% del gasto histórico en contrataciones, se llevó a cabo a través de contrataciones simplificadas.

Figura 28. Tipo de procedimiento y montos asignados por año¹⁶



Estos gastos poco claros, junto con la utilización de intermediarios locales y las contrataciones irregulares, ponen en evidencia la falta de transparencia y posibles actos de corrupción en el manejo de los recursos del CSE.

¹⁶ Para más detalles ver Anexos.

El *web scraping* como herramienta para recopilar datos demuestra ser una alternativa valiosa que facilita la supervisión ciudadana, especialmente en contextos como el nicaragüense, donde el acceso a la información pública no solo es limitado, sino que en gran medida no cumple con los estándares de datos abiertos. Al realizar el presente estudio se encontraron desafíos significativos en términos de accesibilidad para realizar el *web scraping*. Durante el mes de diciembre, el equipo intentó acceder al sitio web más de 30-40 veces, pero aproximadamente el 40% de las veces, el sitio web no funcionaba o se bloqueaba. Estos problemas ocurrieron tanto al usar una VPN como al no usarla.

Cuando no se usaba una VPN, el sitio web no cargaba correctamente, ya que no podía identificar la dirección IP extranjera y tardaba demasiado en responder. Esto causaba un error de tiempo de espera y, a su vez, el error "El sitio no puede ser alcanzado". Al probar con una VPN activada, se observó que el sitio web era accesible para navegar y realizar *web scraping* si se conectaban a la región correcta de VPN. Las VPNs de los Países Bajos y Berlín resultaron inaccesibles, mientras que la de EE. UU. permitió el acceso al sitio web. Entre las VPNs utilizadas para las pruebas, Turbo VPN no pudo acceder al sitio web cuando se conectó a diferentes regiones de la Unión Europea (UE) y la conexión era débil en la región americana. En cambio, aunque Proton VPN no logró acceder al sitio web desde la región de la Unión Europea, sí estableció una conexión sólida en la región de Estados Unidos. Esto permitió al equipo empezar a desarrollar el prototipo para el raspado de datos web.

Para el usuario promedio el sitio web puede parecer sencillo y fácil de navegar. Sin embargo, desde la perspectiva de un desarrollador que intenta realizar *web scraping*, la situación es diferente. En primer lugar, se encontró una estructura de página web compleja en la sección de "Todos los Procesos", donde un cambio en la opción de año desencadena una actualización automática de la página. Este comportamiento dificulta enormemente la programación del formulario de búsqueda, porque la página se actualiza al ejecutar el código, eliminando los estados previamente guardados en las variables.

Además, el sitio web emplea llamadas a la API web de una manera innecesariamente complicada. Aunque la mayoría de los sitios web utilizan API internas en el backend para comunicarse con los servidores y obtener información de bases de datos centralizadas, este sitio web en particular implementa un enfoque poco convencional. En lugar de una comunicación directa entre dos API, se establece un proceso complicado que involucra una ubicación temporal, lo que ralentiza el funcionamiento del sitio y genera confusión entre los desarrolladores y programadores que intentan realizar *web scraping*. Como se indicó previamente, este enfoque no es el método convencional para diseñar la arquitectura de un sitio web. Podría tratarse de un error debido a la inexperiencia o de un intento deliberado de obstaculizar el acceso al diseño web. Teniendo en cuenta el contexto actual que enfrenta el país y la política establecida de dificultar el acceso a la información, es muy probable que esta estructura haya sido construida de forma intencionada.

Otro aspecto notable es la falta de seguridad en la red del sitio web. Aunque las API normalmente están protegidas con conexiones seguras y configuraciones de red que las ocultan de los usuarios públicos, en este caso, las API están expuestas. Esto podría interpretarse como un intento deliberado de dificultar el acceso a la información, ya que las partes restantes del sitio web están construidas de manera cuidadosa y segura. Aunque las API expuestas pueden ser accedidas por usuarios externos, el mecanismo complicado para navegar a través de las páginas web del sitio representa un obstáculo adicional. En resumen, el proceso de *web scraping* en este sitio de compras públicas resultó ser considerablemente más complicado de lo esperado, debido a la complejidad de la estructura de la página web, las intrincadas llamadas a la API y la falta de seguridad en la red. Estos desafíos llevaron al equipo de desarrolladores a emplear enfoques más sofisticados y adaptar sus técnicas de programación para superar las barreras impuestas y acceder a la información que, en teoría, debería ser de dominio público y de fácil acceso.

El uso de Python y técnicas de ciencias de datos permite correr distintos análisis de la información disponible. Una gran ventaja de este tipo de análisis con lenguaje de programación es la automatización. En buena medida se

Inteligencia de fuentes abiertas y ciencia de datos

puede utilizar como regla para otras instituciones y luego adaptar con pequeñas variaciones. En este caso, el modelo de cluster mostró con mayor claridad su utilidad en el CSE. La ventaja de estos algoritmos es procesar grandes cantidades de información y aplicar métodos, estadísticas en pocas líneas de código, que de otra manera consumiría mucho tiempo llevarlas a cabo. En este caso, las técnicas aplicadas fueron clave para revelar una serie de patrones que sugieren la posibilidad de falta de transparencia y prácticas irregulares en la gestión de recursos públicos.

Al analizar los datos obtenidos mediante las técnicas de programación empleadas se observa que, el MTI presenta una asignación significativa de contratos en períodos de tiempo muy cortos. En específico, se observa que los montos asignados en los procesos de compra por cotización y contratación directa son considerablemente altos en relación con el tiempo que se les dedica. Asimismo, se identifica que las empresas MECO S.A, Astaldi, y Eterna, han recibido una asignación desproporcionada de contratos y fondos por parte del MTI. Estas empresas han estado involucradas en escándalos de corrupción en diversos países de América Latina, lo que plantea serias preocupaciones sobre la integridad de sus prácticas comerciales. Además, se evidencia una posible relación estrecha entre dichas empresas y el gobierno nicaragüense, lo que sugiere una influencia política en la asignación de contratos. Estos hallazgos resaltan la necesidad de un escrutinio más riguroso y medidas de control para garantizar la equidad y la eficiencia en la asignación de contratos de infraestructura en Nicaragua.

Por su parte, el análisis revela que las contrataciones simplificadas representan la mayor parte del gasto del CSE, pero solo involucran un bajo número de procesos y un tiempo promedio de licitación extremadamente corto. Esto plantea dudas sobre si se cumplen las condiciones de emergencia y calamidad que justificarían este tipo de procedimiento. Los principales proveedores de las compras realizadas por el CSE son Interlector Nicaragua S.A, Mega Comunicaciones S.A, Consultores Especialistas en Sistema S.A, y Suministro de Informática Quiñonez. Sin embargo, la relación entre algunos de estos proveedores y el CSE ha sido objeto de críticas y sospechas debido a prácticas poco transparentes y posibles sobrepagos en la adquisición de productos y servicios de seguridad informática. Además, se evidencian gastos innecesarios y poco claros, como la adquisición de vehículos de lujo y el mantenimiento de una página web disfuncional. También se señala la subdivisión ilegal de contratos para evitar controles de gasto público, lo cual plantea preocupaciones sobre la transparencia en las actividades

del CSE. Es fundamental abordar estas preocupaciones y promover la transparencia y la rendición de cuentas en el manejo de los recursos institucionales para garantizar la confianza y la legitimidad de las instituciones electorales.

Este estudio, enfocado en dos productos principales, el scraping y análisis de datos, ha demostrado un gran potencial. A pesar de que el análisis cuantitativo puede estar limitado por las preguntas de investigación y los métodos empleados, los datos raspados y modelados permiten su exploración desde diversas perspectivas, ampliando su alcance. Con base en esta potencialidad, se propone un scraping integral del sitio web y la creación de una base de datos relacional que contenga información sobre la adjudicación de contratos, proveedores y planes anuales de adquisiciones. Enfocando esta propuesta, se busca ampliar la actual infraestructura de recopilación y análisis de datos para incluir todos los ministerios y alcaldías. Para complementar esto, se sugiere desarrollar una interfaz de usuario intuitiva que facilite el acceso rápido y el estudio de los datos de licitaciones, incorporando funciones de búsqueda y filtrado avanzadas, visualizaciones dinámicas y opciones de informes personalizados.

Finalmente, este tipo de investigación es invaluable para formular preguntas de investigación que pueden abordarse desde múltiples disciplinas, especialmente en el ámbito periodístico, pero es imprescindible el acceso a los datos. Estos estudios pueden alcanzar una mayor profundidad cuando se complementan con fuentes adicionales, como se hizo en este caso, destacando prácticas cuestionables como la adjudicación de contratos sin justificación adecuada, el uso de intermediarios locales y la subdivisión de compras para eludir controles.

BIBLIOGRAFIA

- AFP. (2021, agosto 25). Tres hijos de la expresidenta Violeta Chamorro acusados de corrupción en Nicaragua. El Universo. <https://www.eluniverso.com/noticias/internacional/tres-hijos-de-la-expresidenta-violeta-chamorro-acusados-de-corrupcion-en-nicaragua-nota/>
- Meléndez, A. (2021, noviembre 15). Investigan posibles sobornos de MECO en ocho municipalidades para ganar contratos: Cinco alcaldes permanecen detenidos. Monumental. <https://www.monumental.co.cr/2021/11/15/investigan-posibles-sobornos-de-meco-en-ocho-municipalidades-para-ganar-contratos-cinco-alcaldes-permanecen-detenidos/>
- Murillo, A. (2021, junio 17). “Cochinilla”, la mega trama de corrupción que sacude Costa Rica. Divergentes. <https://www.divergentes.com/cochinilla-la-mega-trama-de-corrupcion-que-sacude-costa-rica/>
- Barbaresi, A. (2021). Trafilatura: A web scraping library and command-line tool for text discovery and extraction. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations, 122–131.
- BNamericas—Government applauds judicial decision to rec... (s/f). BNamericas. Com. <https://www.bnamericas.com/en/news/government-applauds-judicial-decision-to-recover-2276-million-from-the-construction-of-el-chaparral>
- Transparencia Internacional. (2022). Corruption Perceptions Index 2021. Transparency International. https://images.transparencycdn.org/images/CPI2021_Report_EN-web.pdf
- Artículo 66. (2020, agosto 20). CSE compró millonario lote de camionetas mediante procesos ilegales. Artículo 66. <https://www.articulo66.com/2020/08/20/cse-corrupcion-compras-sin-control/>
- Orozco, D. (2020, agosto 5). CSE gasta 20 millones de córdobas en telefonía y mantenimiento de una web inservible. Nicaragua Investiga. <https://nicaraguainvestiga.com/nacion/23968-cse-gasta-20-millones-de-cordobas-en-telefonía-y-mantenimiento-de-una-web-inservible/>
- Valle, D. (2020). Obras públicas de Nicaragua: Un banquete de corrupción. Expediente Público. <https://www.expedientepublico.org/especial/obras-publicas-de-nicaragua-un-banquete-de-corrupcion/>
- Diouf, R., Sarr, E. N., Sall, O., Birregah, B., Bouso, M., & Mbaye, S. N. (2019). Web Scraping: State-of-the-Art and Areas of Application. 2019 IEEE International Conference on Big Data (Big Data), 6040–6042. <https://doi.org/10.1109/BigData47090.2019.9005594>
- Dumbacher, B., & Diamond, L. K. (2018). SABLE: Tools for web crawling, web scraping, and text classification. Federal committee on statistical methodology research conference.
- Garmendia, E. (2020, julio 14). Las Piedrecitas-7 Sur: Una gran obra empuñada por la corrupción. Nicaragua Investiga. <https://nicaraguainvestiga.com/reportajes/21824-piedrecitas-7-sur-obra-corrupcion-nicaragua/>
- El Perro Amarillo. (2020a). ELVIN SANTOS Y ALBERTO DÍAZ LOBO, SOCIOS EN MÁS ESTAFAS DE INVEST-H. <https://elperroamarillo.news/ned/elvin-santos-y-alberto-diaz-lobo-socios-en-mas-estafas-de-invest-h/>
- Enríquez, O. (2021, septiembre 3). Poder Electoral oculta el 68% de sus ‘contratos electorales’. Confidencial. <https://confidencial.digital/politica/poder-electoral-oculta-el-68-de-sus-contratos-electorales/>

Inteligencia de fuentes abiertas y ciencia de datos

- Observatorio de corrupción. (2022). Informe Nicaragua: Seguimiento al Compromiso de Lima Observatorio Ciudadano de Corrupción. https://occ-america.com/wp-content/uploads/2021/10/Informe-Nicaragua_Espanol.pdf
- Khder, M. A. (2021). Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. *International Journal of Advances in Soft Computing & Its Applications*, 13(3).
- Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research. *Psychological methods*, 21(4), 475.
- Corporación Latinobarómetro. (2021). Latinobarómetro 2021. <https://www.latinobarometro.org/latContents.jsp>
- López, L. (2022, octubre 13). La corrupción sí le pagó a Meco en Panamá. CLIP. <https://www.elclip.org/la-corrupcion-si-le-pago-a-meco-en-panama-2/>
- Rodríguez, L. & Zechmeister, E. (2021). Pulse of Democracy. Vanderbilt University. https://www.vanderbilt.edu/lapop/ab2021/2021_LAPOP_AmericasBarometer_2021_Pulse_of_Democracy.pdf
- Lyra, M. S., Damásio, B., Pinheiro, F. L., & Bacao, F. (2022). Fraud, corruption, and collusion in public procurement activities, a systematic literature review on data-driven methods. *Applied Network Science*, 7(1), 83.
- El Perro Amarillo. (2020b). Marco Bográn entregó al tío supervisión de millonarios contratos. <https://elperroamarillo.news/secciones-especiales/secretos/marco-bogran-entrego-al-tio-supervision-de-millonarios-contratos-invest-h/>
- Matta, P., Sharma, N., Sharma, D., Pant, B., & Sharma, S. (2020). Web scraping: Applications and scraping tools. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(5), 8202–8206.
- Mavidis, A., & Folinas, D. (2022). From Public E-Procurement 3.0 to E-Procurement 4.0; A Critical Literature Review. *Sustainability*, 14(18), 11252.
- Salazar, M. (2020, septiembre 8). El gran contratista de la Alcaldía de Managua. *Divergentes*. <https://www.divergentes.com/el-super-contratista-de-la-alcaldia-de-managua/>
- Reina, P. (2021). Webuild Acquires Majority of Astaldi. <https://www.enr.com/articles/50583-webuild-acquires-majority-of-astaldi>
- Mitchell, R. (2018). *Web scraping with Python: Collecting more data from the modern web*. O'Reilly Media, Inc.
- Modrušan, N., Mršić, L., & Rabuzin, K. (2021). Intelligent Public Procurement Monitoring System Powered by Text Mining and Balanced Indicators. *Data Management Technologies and Applications: 9th International Conference, DATA 2020, Virtual Event, July 7–9, 2020, Revised Selected Papers 9*, 115–133.
- Modrušan, N., Rabuzin, K., & Mršić, L. (2021a). Review of Public Procurement Fraud Detection Techniques Powered by Emerging Technologies. *International Journal of Advanced Computer Science and Applications*, 12(2).
- Nai, R., Sulis, E., & Meo, R. (2022). Public Procurement Fraud Detection and Artificial Intelligence Techniques: A Literature Review.
- Organización para la Cooperación y el Desarrollo Económicos. (2020). *Panorama de las Administraciones Públicas América Latina y el Caribe 2020*. OECD. <https://doi.org/10.1787/1256b68d-es>
- Office of the Director of National Intelligence. (2015). *Analytic Standards*. Office of the Director of National Intelligence. <https://www.dni.gov/files/documents/ICD/ICD%20203%20Analytic%20Standards.pdf>
- Carta de Datos Abiertos . (2021). CARTA INTERNACIONAL DE DATOS ABIERTOS. <https://opendatacharter.net/principles-es/>

Un análisis exploratorio para detectar patrones de corrupción

- Ortiz-Prado, E., Fernandez-Naranjo, R., Torres-Berru, Y., Lowe, R., & Torres, I. (2021). Exceptional prices of medical and other supplies during the COVID-19 pandemic in Ecuador. *The American journal of tropical medicine and hygiene*, 105(1), 81.
- Rodríguez, O. (2021, julio 19). Gobierno de Nicaragua adelantó a MECO millones en pagos de obras públicas. *La Nación*. <https://www.nacion.com/economia/finanzas/gobierno-de-nicaragua-adelanto-a-meco-millones-en/OK24EKG3FRHONDUT44OEKTPL3U/story/>
- Navarro, A. (2020). Presidencia de Nicaragua gasta 916 mil córdobas para blindar sistema informático tras ataques de Anonymous. <https://www.articulo66.com/2020/04/28/presidencia-de-nicaragua-gasta-916-mil-cordobas-blindar-sistema-informatico-ataques-anonymous/>
- Rodríguez, M. J. G., Rodríguez-Montequín, V., Ballesteros-Pérez, P., Love, P. E., & Signor, R. (2022). Collusion detection in public procurement auctions with machine learning algorithms. *Automation in Construction*, 133, 104047.
- Serrano, L. (2021). *Grokking Machine Learning*. Simon and Schuster.
- Kent, S. (1993). *Words of Estimative Probability*. CIA. <https://www.cia.gov/static/0aae8f84700a256abf63f7aad73b0a7d/Words-of-Estimative-Probability.pdf>
- Torres-Berru, Y., López Batista, V. F., Torres-Carrión, P., & Jimenez, M. G. (2020). Artificial Intelligence techniques to detect and prevent corruption in procurement: A systematic literature review. *Applied Technologies: First International Conference, ICAT 2019, Quito, Ecuador, December 3–5, 2019, Proceedings, Part II 1*, 254–268.
- Torres-Berru, Y., & López Batista, V. F. (2022). Data and Text Mining for the Detection of Fraud in Public Contracts: A Case Study of Ecuador's Official Public Procurement System. *Doctoral Symposium on Information and Communication Technologies-DSICT*, 116–127.
- Ulbricht, L. (2020). Scraping the demos. *Digitalization, web scraping and the democratic project*. *Democratization*, 27(3), 426–442.
- Uzun, E. (2020). A novel web scraping approach using the additional information obtained from web pages. *IEEE Access*, 8, 61726–61740.
- Vargiu, E., & Urru, M. (2013). Exploiting web scraping in a collaborative filtering-based approach to web advertising. *Artif. Intell. Res.*, 2(1), 44–54.
- Chacón, V. (2021, julio 16). MECO lo intentó, pero desistió de sacar maquinaria hacia Nicaragua. *Semanario Universidad*. <https://semanariouniversidad.com/pais/meco-lo-intento-pero-desistio-de-sacar-maquinaria-hacia-nicaragua/>
- Volosín, N. (2015). *Datos abiertos, corrupción y compras públicas*. Santiago de Chile:(ILDA). *Iniciativa Latinoamericana por los Datos Abiertos*.

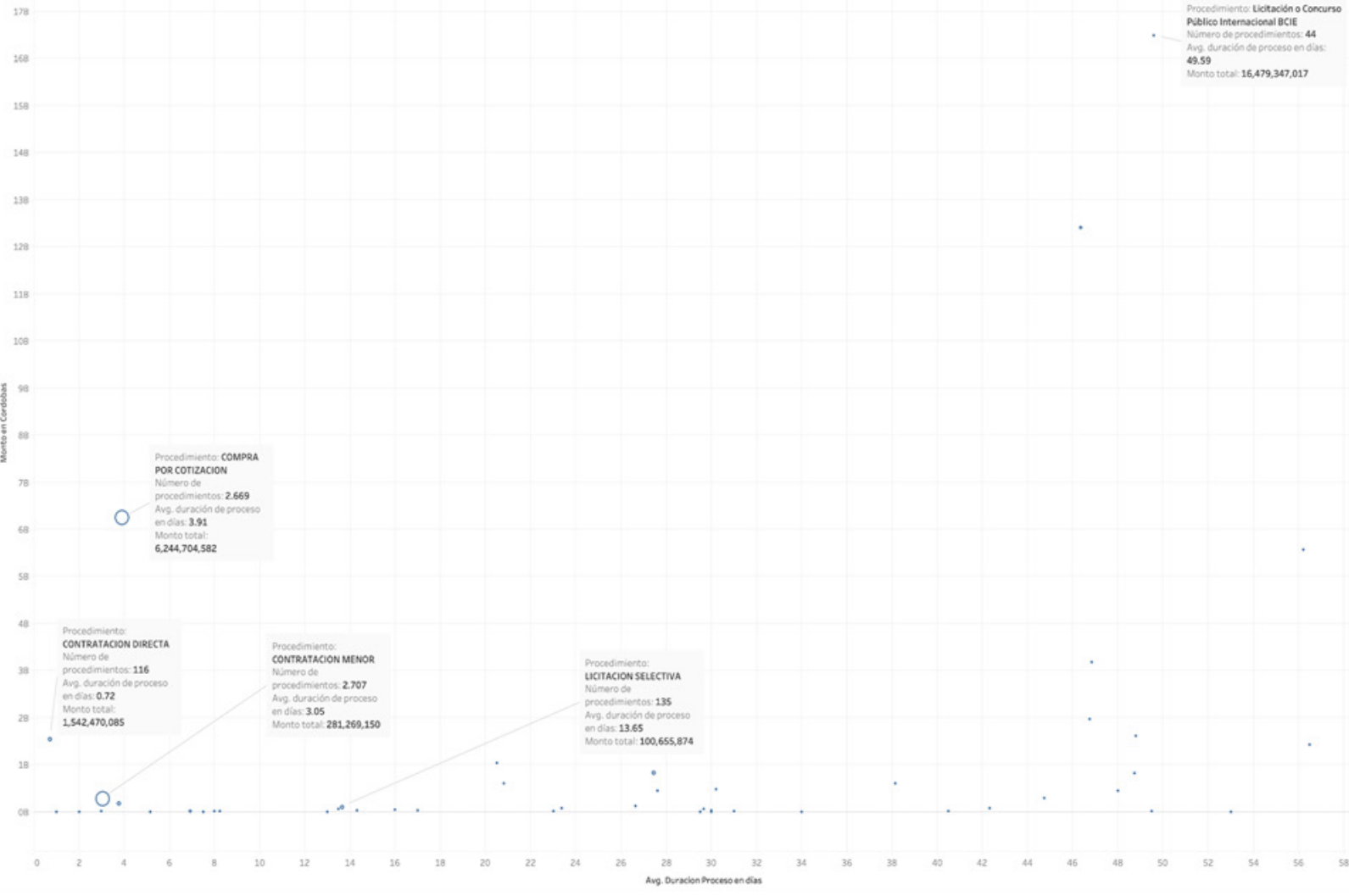
Anexos

Tipo de procedimiento organizado por cluster

COMPRA POR COTIZACION 6,244,704,582 2,669 3,91 5	COMPRA POR COTIZACION MENOR 172,892 2 2,00 2	Selección Directa - BM 12,906,294 6 3,00 2	CONCURSO 113,674,324 14 26,64 1	CONCURSO PUBLICO INTERNACIONAL 445,203,097 11 48,00 1	CONSULTOR INDIVIDUAL 34,219,871 25 14,32 1	Comparación de Precios Especial Obras / BID 65,277,714 4 13,50 1	Comparación de Precios Obras 2,702,922 1 13,00 1	Concurso Firma Consultora 484,186,170 40 30,23 1	Concurso Público Nacional - BCIe MEX 11,378,828 2 31,00 1
	CONTRATACION DIRECTA 1,542,470,085 118 0,72 2	Consultor Individual Banca Mundial 30,182,200 2 16,00 1	LICITACION PUBLICA INTERNACIONAL 35,044,517 3 30,00 1	LICITACION PUBLICA INTERNACIONAL 32,409,008,435 67 46,34 1	LICITACION PUBLICA NACIONAL 826,155,058 156 27,44 1	LICITACION RESTRICTIVA 10,354,037 38 23,00 1	LICITACION SELECTIVA 100,655,874 135 13,65 1	LICITACION Estandar - MEX 1,038,527,006 4 20,50 1	
CONTRATACION MENOR 281,269,150 2,707 3,05 4	Comparación de Precios Bienes y Servicios 2,585,871 24 7,50 2	LICITACION O CONCURSO PUBLICO INTERNACIONAL 5,573,374,497 30 56,20 1	Licitación Pública Internacional Bienes y Servicios 11,901,590 4 49,50 1	SDOI- Solicitud Oferta InterNacional / BM 14,238,093 2 40,50 1	SDOI- Solicitud Oferta InterNacional Obras/ BM MINED 816,540,045 4 48,75 1	SDON- Solicitud Oferta Nacional Obras/ BM MFI 286,668,735 4 44,75 1	SELECCION BASADA EN EL MENOR COSTO 1,799,202 2 30,00 1		
	Comparación de Precios Especial Bienes / BID FINANCIERA 13,361,974 16 8,00 2	LICITACION O CONCURSO PUBLICO INTERNACIONAL CON OFERTA FINANCIERA 1,608,703,826 34 48,79 1	LICITACION O CONCURSO PUBLICO INTERNACIONAL LIMITADO 3,973,736,002 4 46,75 1	Licitación Pública Internacional Obras 3,176,082,718 6 46,83 1	LICITACION PUBLICA NACIONAL Bienes 3,621,821 4 29,50 1	SELECCION BASADA EN LAS CALIFICACIONES DEL CONSULTOR 23,782,467 10 17,00 1	SOLICITUD DE OFERTAS INTERNACIONAL 1,424,639,212 2 56,50 1	Selección de Firma Consultora BID 454,609,087 26 27,62 1	Selección de Firma Consultora BM 74,649,183 12 42,33 1
CONTRATACION SIMPLIFICADA 177,431,010 154 3,77 4	LICITACION PUBLICA 606,167,174 19 38,16 2	LICITACION O CONCURSO PUBLICO NACIONAL 4,394,874 2 34,00 1	Licitación Pública Nacional Obras 67,244,295 3 29,67 1	SELECCION CON BASE EN UNA SOLA FUENTE 1,312,170 1 1,00 1	SOLICITUD DE OFERTAS 9,862,518 4 8,25 1	Solicitud de Cotización BM 4,105,484 1 53,00 1	Solicitud de Oferta Nacional/ BM 4,105,484 1 53,00 1		
	MATERIA EXCLUIDA 44,084 6 5,17 2	LICITACION POR REGISTRO 75,813,712 30 23,37 1	LICITACION O CONCURSO PUBLICO INTERNACIONAL BCIe 16,479,347,017 44 49,59 1	SELECCION BASADA EN CALIDAD Y COSTO 594,429,840 36 20,81 1					
COMPARACION DE PRECIOS 18,219,635 86 4,93 2									

Un análisis exploratorio para detectar patrones de corrupción

Distribución proveedor, monto y duración de procesos (2007-2022)

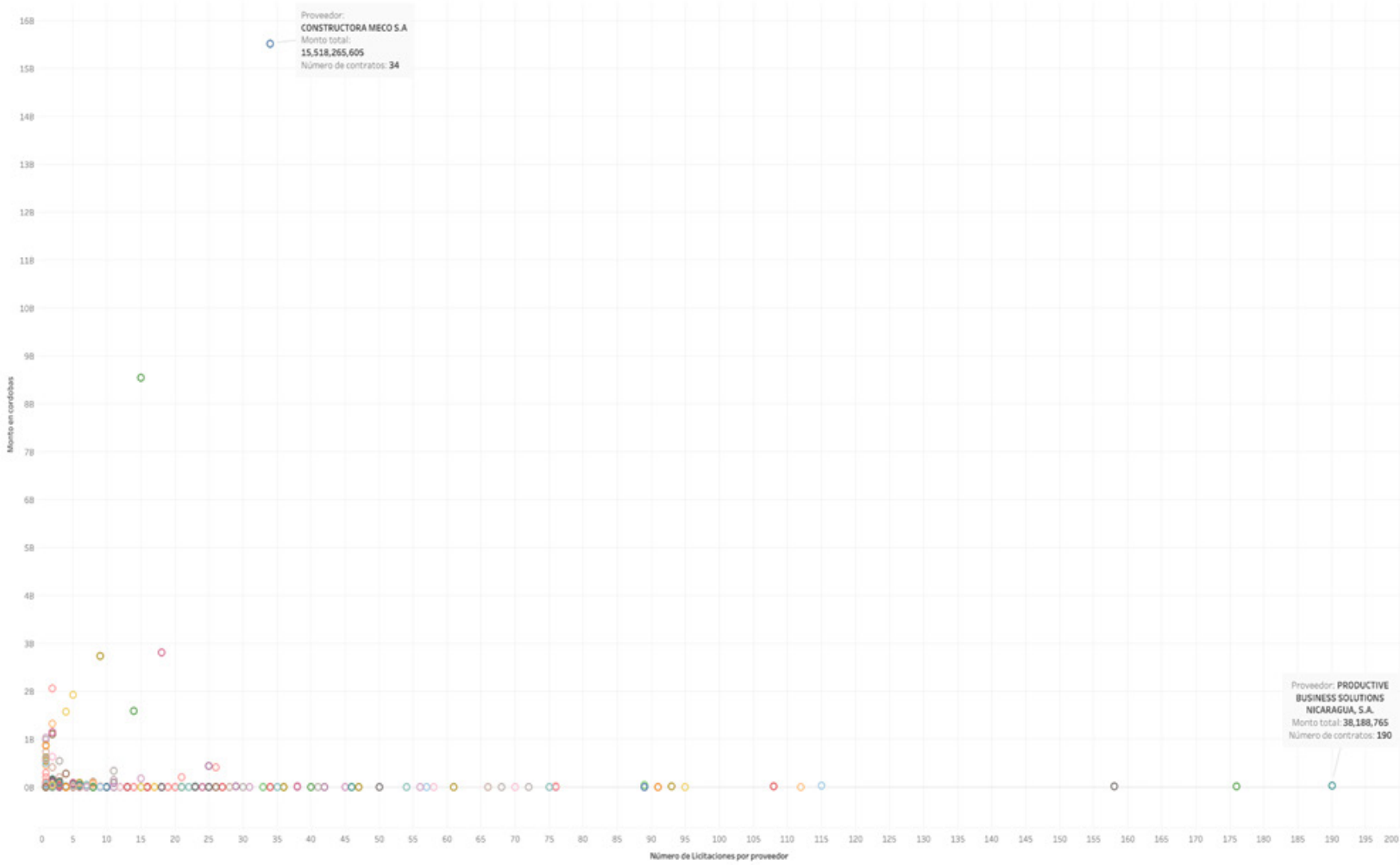


Top 25 proveedores distribuidos por fuente de financiamiento

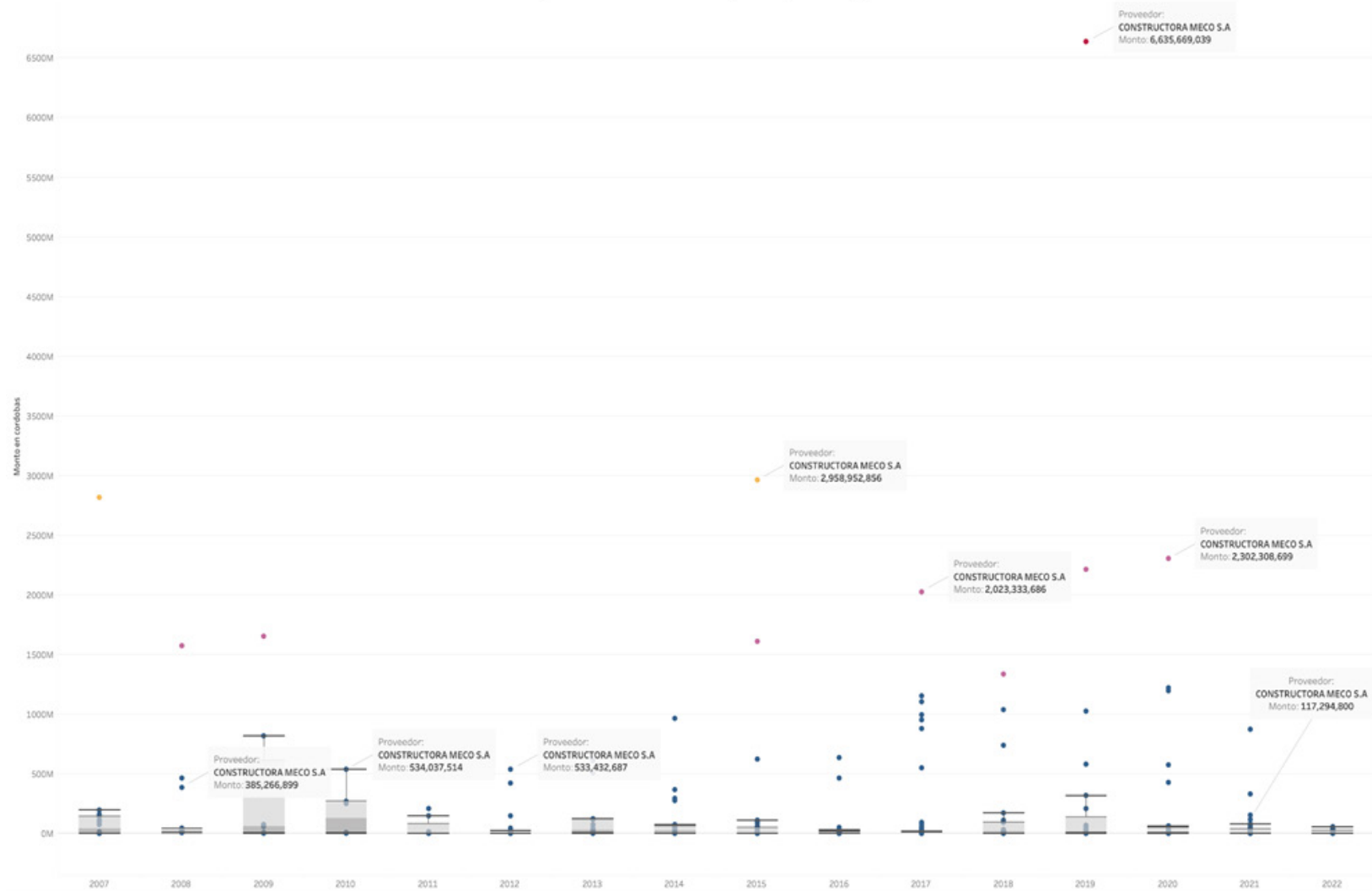
Proveedor	Null	Banco Centroamericano de Integración Económica -BCIE-	Banco Interamericano de Desarrollo -BID-	Banco Mundial	Gobierno de Nicaragua
CONSTRUCTORA MECO S.A		8,535,499,127 34,73%	4,057,305,994 29,79%	2,226,716,858 71,09%	417,552,343 12,85%
CONSTRUCTORA SANTA FE LIMITADA SUCURSAL NICARAGUA		6,030,120,905 24,53%	1,524,014,966 11,19%		991,929,345 30,52%
MEDINA RODRIGUEZ, RENE	2,813,192,730 76,89%				
ASTALDI S.P.A. SUCURSAL NICARAGUA			2,322,201,653 17,05%		271,957,813 8,37%
EMPRESA DE CONSTRUCCIÓN Y TRANSPORTE ETERNA, S.A. DE C.V. - 0503...		2,064,504,432 8,40%			
PROFESIONALES DE LA CONSTRUCCION S.A DE C.V		570,393,093 2,32%	1,024,793,332 7,52%	329,172,454 10,51%	
Grupo Pycsa, S.A. DE C.V. - 3-101-035078-17	393,269,307 10,72%				1,179,807,921 36,30%
B.LANZAS CONTRATISTAS CIVILES SOCIEDAD ANONIMA		741,552,286 3,02%	633,393,531 4,65%		
Constructora Santa fe Limitada -		1,332,449,107 5,42%			
Consorcio FCC Américas (FCC Construcción America, S.A.) y Operadora CICSa, s.a. de c...			1,150,700,768 8,45%		
CONSTRUCTORA MAKRO, S.A DE C.V.-			1,102,797,564 8,10%		
CONSTRUCTORA MAKRO SOCIEDAD ANÓNIMA DE CAPITAL VARIABLE SUCURS...		1,038,035,122 4,22%			
Llansa Ingerieros S.A -		992,425,885 4,04%			
CONSORCIO MALACATOYA (LLANSA INGENIEROS-D. GUERRERO INGENIEROS, ...		875,927,114 3,56%			
Constructora Makro, S.A. de C.V. -		735,437,942 2,99%			
consorcio santa fe prodemex, s.a. -			620,851,521 4,56%		
CONSORCIO CEMEX-MECO-LLANSA		609,735,908 2,48%			
ASC GROUP INTERNATIONAL CORPORATION SOCIEDAD ANONIMA				576,302,650 18,40%	
Prodecon S.A de C.V -		546,142,002 2,22%			
MIRSA CONASA		506,684,821 2,06%			
Promotora y Desarrolladora Mexicana Sociedad Anónima de Capital Variable (PR...			480,939,910 3,53%		
Terracería y Cimentaciones del Sur, S.A. DE C.V. - 91020	462,025,967 12,59%				
FCC CONSTRUCCIÓN DE CENTROAMERICA, SOCIEDAD ANONIMA			411,297,540 3,02%		
CONSTRUCTORA LATINO Y CIA. LTDA.					389,172,948 11,97%
Profesionales de la Construcción S.A de CV (PRODECOM) -			291,163,339 2,14%		

Un análisis exploratorio para detectar patrones de corrupción

Distribución del Monto en relación al Número de Procedimientos por Proveedor

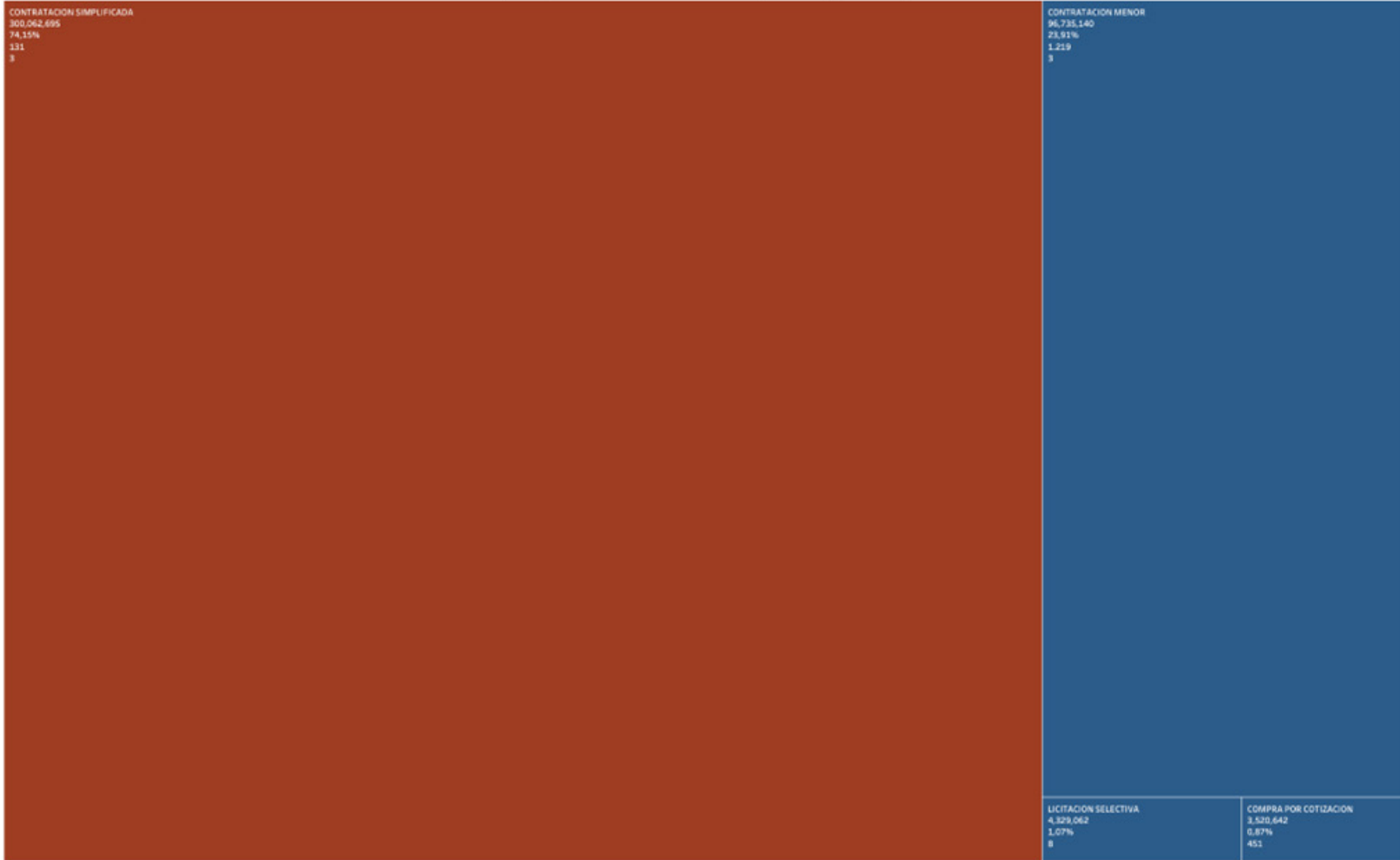


Box plot del monto de licitaciones por año (2007-2022)



Un análisis exploratorio para detectar patrones de corrupción

Tipo de procedimiento, monto, duración promedio (2009-2021)

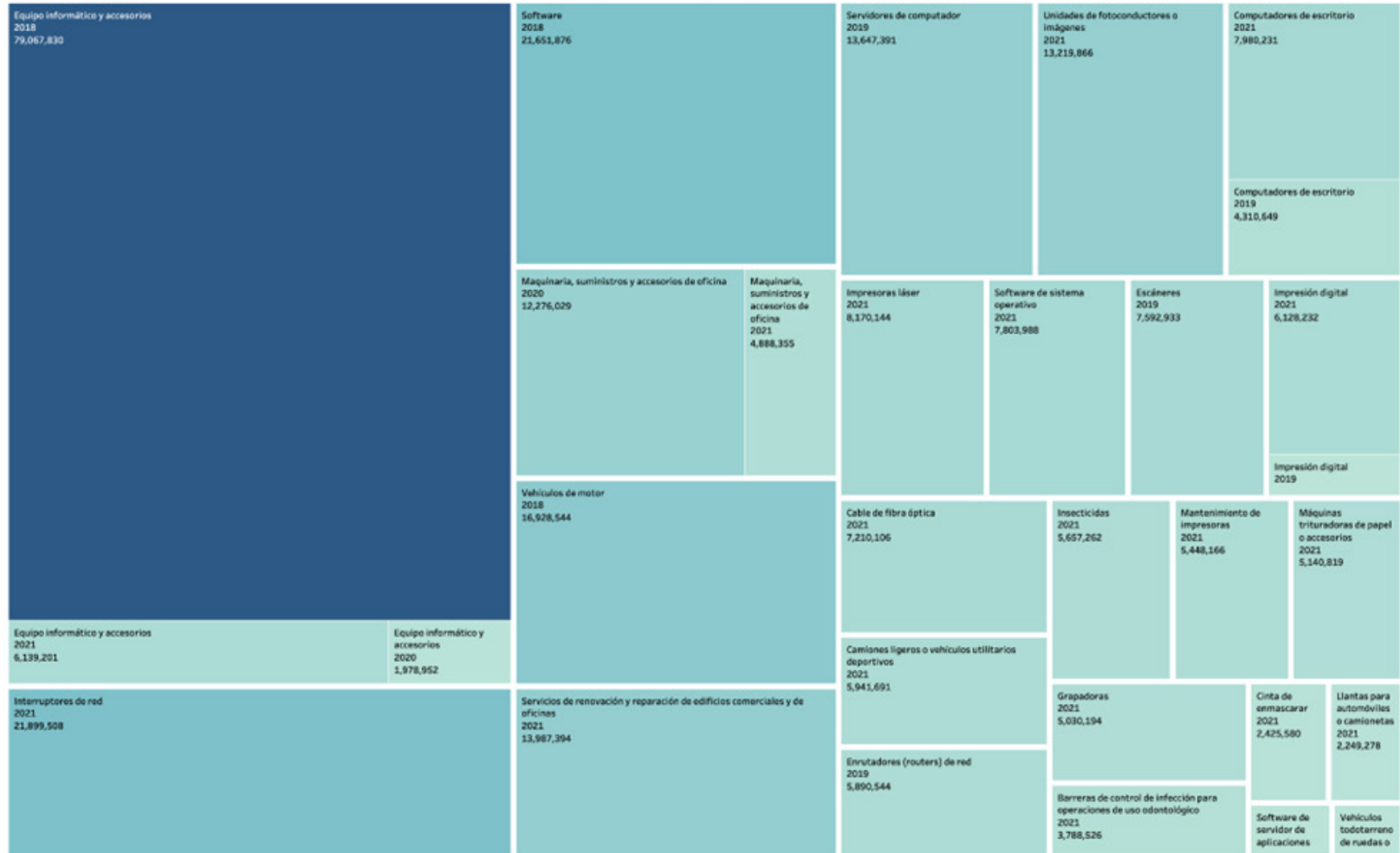


Inteligencia de fuentes abiertas y ciencia de datos

Monto por procedimiento (2009-2021)

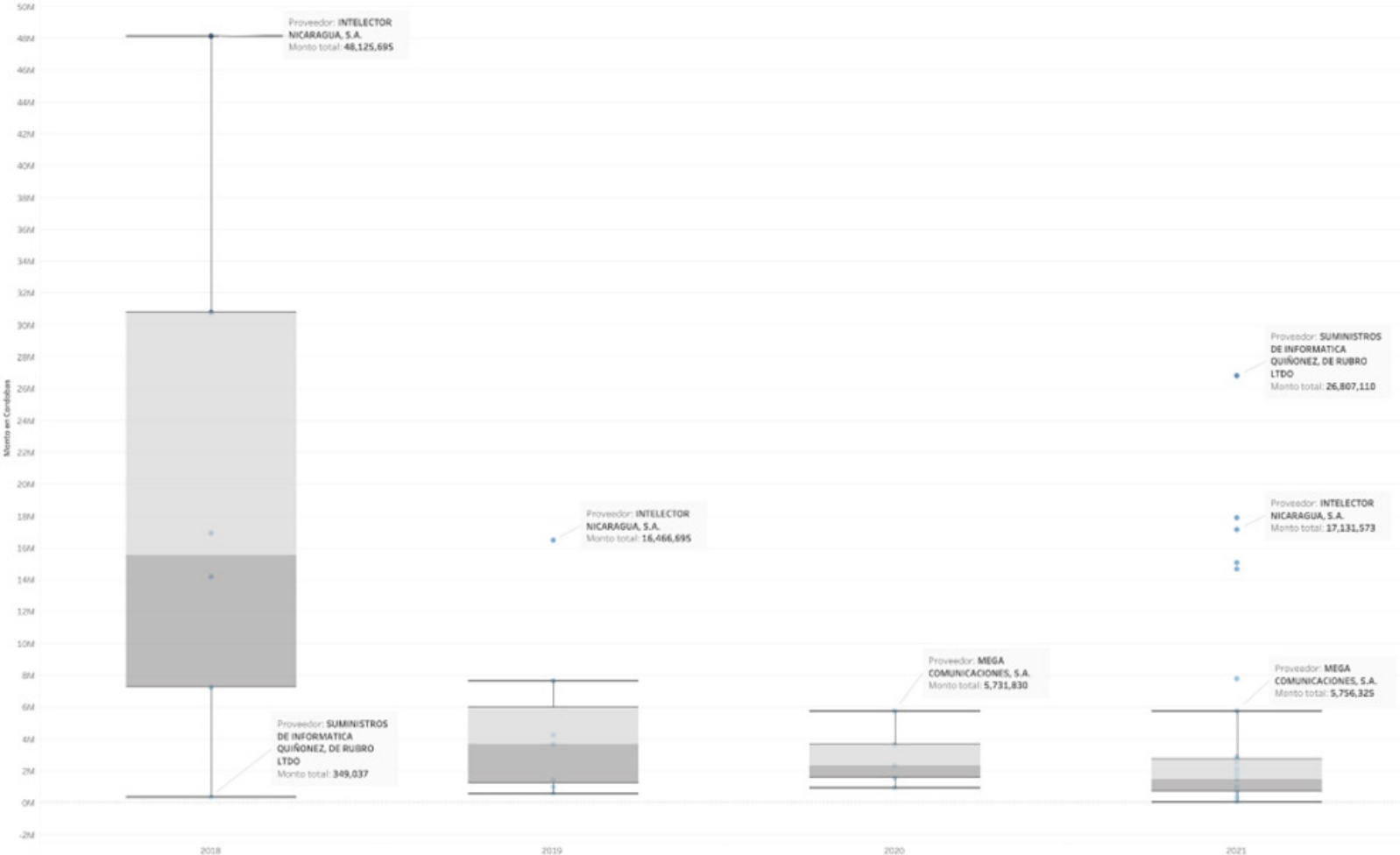
Procedimiento	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
COMPRA POR COTIZACION	760,643	2,319,258	440,740										
CONTRATACION MENOR			1,721,963	448,131	2,481,304	3,602,707	1,418,274	7,317,819	10,017,997	12,376,245	5,428,150	5,092,404	46,830,144
CONTRATACION SIMPLIFICADA										117,648,250	35,050,922	14,254,981	133,108,542
LICITACION SELECTIVA											1,303,288		3,025,775

Descripción de bienes y servicios adquiridos por contratación simplificada

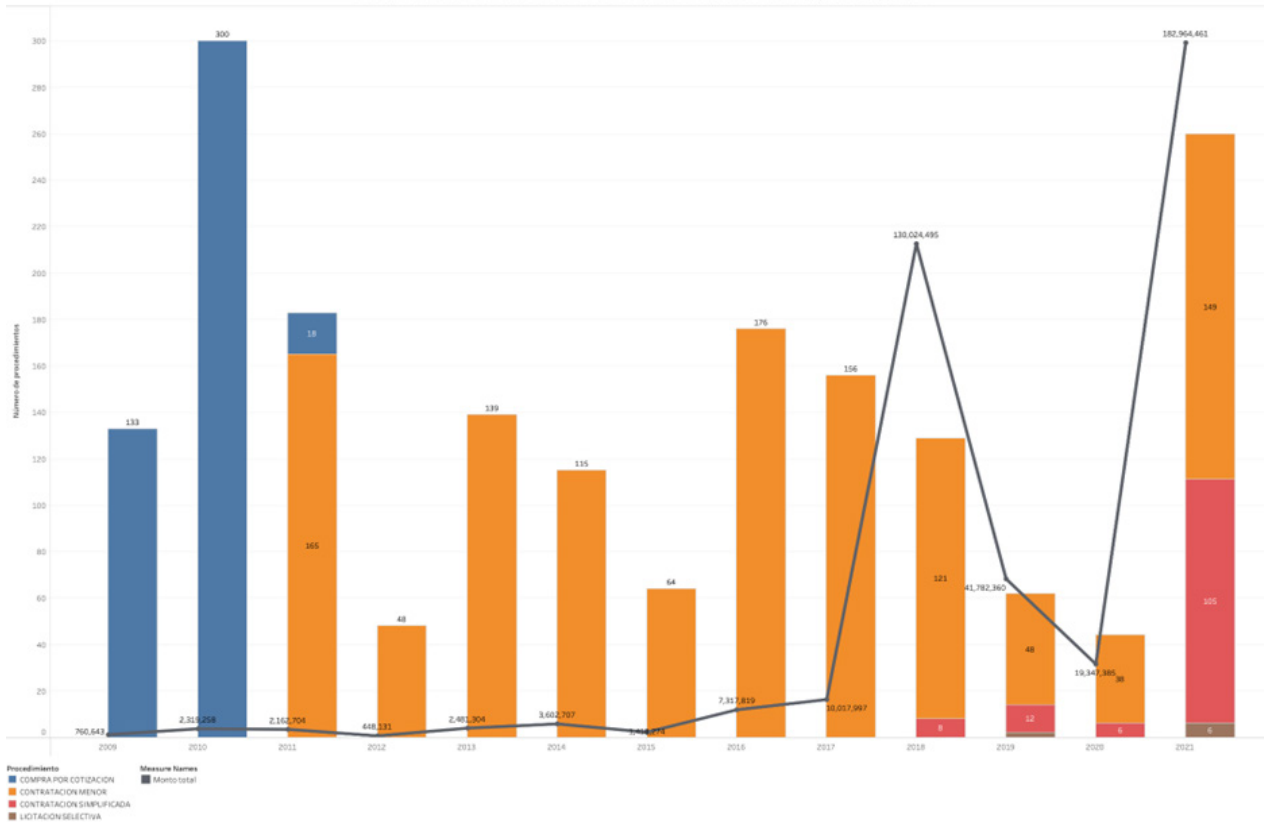


Un análisis exploratorio para detectar patrones de corrupción

Box Plot de Montos de Proveedores por Año en Cordobas



Comparación del Número de Procedimientos y Montos asignado por Año



gestion.nicaraguacompra.gob.ni/siscae/portal/adquisiciones-gestion/busqueda?accion=todos&usr_ua_id=todos



This site can't be reached

www.gestion.nicaraguacompra.gob.ni took too long to respond.

Try:

- Checking the connection
- [Checking the proxy and the firewall](#)

ERR_CONNECTION_TIMED_OUT

Details

Reload

Un análisis exploratorio para detectar patrones de corrupción

