

# Detección de Redes de Corrupción en Ecuador utilizando fuentes heterogéneas.

Allison Barrezueta  
Escuela Superior Politécnica  
del Litoral, ESPOL  
Faculty of Electrical and Computer  
Engineering  
Campus Gustavo Galindo  
Km 30.5 Vía Perimetral,  
P.O. Box 09-01-5863  
Guayaquil, Ecuador  
abarrezu@espol.edu.ec

Felipe Gonzales  
Escuela Superior Politécnica  
del Litoral, ESPOL  
Faculty of Electrical and Computer  
Engineering  
Campus Gustavo Galindo  
Km 30.5 Vía Perimetral,  
P.O. Box 09-01-5863  
Guayaquil, Ecuador  
felgcoll@espol.edu.ec

Juan Laso  
Escuela Superior Politécnica  
del Litoral, ESPOL  
Faculty of Electrical and Computer  
Engineering  
Campus Gustavo Galindo  
Km 30.5 Vía Perimetral,  
P.O. Box 09-01-5863  
Guayaquil, Ecuador  
jflaso@espol.edu.ec

**Resumen**—En este documento se investigan los contratos dados por el SERCOP durante la emergencia sanitaria. En particular, se centra en encontrar la manera de identificar aquellos contratos en los que se haya dado casos de corrupción. El propósito es encontrar la viabilidad de aplicar una combinación de Clustering, Natural Language Processing y Social Network Analysis para encontrar dichos casos.

**Index Terms**—Corrupción, Clustering, SNA, NLP.

## I. INTRODUCCIÓN

Según la Organización Internacional de Transparencia, la corrupción se la define como un abuso del poder confiado para beneficio privado, esto genera una disminución en la confianza, debilita la democracia e impide el desarrollo socioeconómico. Dentro del marco de la corrupción existen diferentes formas de corrupción y comportamientos, entre estos nos centramos en la corrupción definida como el comportamiento que tienen los servidores públicos (o entidades públicas) en exigir o tomar dinero o favores a cambio de servicios.

El sistema contratación pública en los últimos 10 años ha representado en promedio alrededor de 26.62 % del presupuesto general del Estado. Este sistema tiene falencias en su ejecución y recientemente han salido a la luz casos de corrupción a través de este servicio de compras públicas. Casos como sobreprecios, coimas y comisiones que involucran a las dos partes aportan a la generalización de que Ecuador es un país con servidores públicos corruptos.

En relación con esto, no es de sorprender que este tipo de actividades dentro del marco público sea uno de los más vulnerables a la corrupción. Si bien ha habido muchos indicadores cualitativos de corrupción de alto nivel en la contratación pública, es reciente que se han puesto a disposición indicadores cuantitativos.

Un indicador es una representación de un factor o cantidad asociado. Por ejemplo, el índice de precios al consumidor sirve como un indicador del costo de vida general que consiste en muchos factores que algunos de los cuales no están incluidos en la computación del IPC. Los indicadores son dispositivos

estadísticos comunes empleados en la economía y otras ramas de la ciencia.

Con todo esto el problema que queremos abarcar es como podríamos ayudar a identificar posibles entidades o empresas prospectos que tiendan a comercializar de manera corrupta dentro del SERCOP para disminuir la corrupción.

Es necesario priorizar acerca de nuestro objetivo para identificar patrones en el Servicio Nacional de Contratación Pública, si bien no se abarcaron todos los casos, nos enfocamos en obtener datos e información desde el mes de marzo hasta julio, que corresponden a los meses en que la pandemia afectó al Ecuador. Nos enfocamos en productos pertenecientes a la emergencia sanitaria, específicamente nos centramos en las mascarillas y las fundas de cadáveres. Nuestro objetivo fue analizar las contrataciones del estado ecuatoriano para obtener casos irregulares que ayuden de manera inductiva a descubrir más casos anómalos.

## II. TRABAJOS RELACIONADOS

Existen diversos trabajos relacionados en los que se utilizan técnicas de clustering para la identificación de fraude o corrupción [1] [2]. Esto se debe a dos características principales de dicha técnica: la agrupación de registros similares y es un algoritmo de aprendizaje no supervisado. En este caso en específico, es de especial interés dividir los datos en dos grupos. Estos dos grupos serán los contratos comunes y los contratos que tengan alguna característica fuera de lo común.

### II-A. Economía de Países desarrollados

En los últimos años se ha observado una notable expansión de la capacidad de los economistas para medir la corrupción. Esto, a su vez, ha dado lugar a una nueva generación de estudios microeconómicos bien identificados. Debido a ello se han examinado las pruebas sobre la corrupción en los países en desarrollo a la luz de estos recientes avances, centrándonos en tres preguntas: cuánta corrupción hay, cuáles son las consecuencias de la eficiencia de la corrupción y qué determina el nivel de corrupción [1]. Encontrando como

resultado pruebas sólidas de que la corrupción responde a incentivos económicos, pero también que los efectos de las políticas anticorrupción a menudo se atenúan a medida que los funcionarios encuentran estrategias alternativas para perseguir mayores rentas.

## II-B. Banco Internacional

Los bancos internacionales de desarrollo otorgan préstamos de bajo interés a los países en desarrollo en un esfuerzo por estimular el desarrollo social y económico. Estos préstamos apoyan proyectos clave de infraestructura, incluyendo la construcción de carreteras, escuelas y hospitales. Sin embargo, a pesar de los mejores esfuerzos de los bancos de desarrollo, estos fondos de préstamos a menudo se pierden por fraude, corrupción y colusión. En un esfuerzo por sancionar y detener esta irregularidad y para asegurar el uso adecuado de los fondos, los bancos de desarrollo llevan a cabo investigaciones extensas y costosas que pueden tardar más de un año en completarse. Por ello se desarrolló un sistema en conjunto con el Grupo del Banco Mundial para mejorar el tiempo y la eficiencia de los costos de su proceso de investigación. Utilizando datos históricos de premios monetarios y resultados de investigación anteriores, el clasificador asigna una "puntuación de riesgo," a los contratos del Banco Mundial. Esta puntuación de riesgo está diseñada para permitir a los investigadores del Banco Mundial identificar los contratos con más probabilidades de conducir a una investigación fundamentada [2]. Si se implementa, se prevé que nuestro sistema automatizado identifique con éxito el fraude, la corrupción y la colusión en el 70 % de los casos.

## III. DATASET

Para el experimento se recurrieron a datasets obtenidos del Sistema Nacional de Compras Públicas (SERCOP) y de la Superintendencia de Compañías y Seguros (SUPERCIAS) teniendo así un dataset principal de 19618 registros con 10 atributos cada uno, que datan de contrataciones realizadas durante emergencia sanitaria. El periodo de tiempo del dataset tiene un inicio el 17 de Marzo de 2020, finalizando el 26 de Julio de 2020. Este dataset fue obtenido directamente de la página web del Servicio Nacional de Compras Públicas.

Se recopiló una base de datos cuantitativa y cualitativa descargando los archivos de la fuente oficial en línea, contenidos en un archivo EXCEL. Se aplicaron estrategias de minería de texto de manera automática (web scraper), dando lugar a un nuevo registro. Las variables de interés que se tomó fueron nombre del proveedor, valor del contrato, nombre de la entidad contratante, nombre de los accionistas relacionados.

## IV. METODOLOGÍA

Para el análisis se utilizaron métodos univariados y multivariados para encontrar una correlación entre las variables de estudio y encontrar datos atípicos para la determinación de corrupción en dicho periodo de estudio.

En primer lugar, se realizó un análisis exploratorio del dataset, con el propósito de poder realizar decisiones acertadas

sobre las técnicas a utilizar para la identificación de los casos de corrupción.

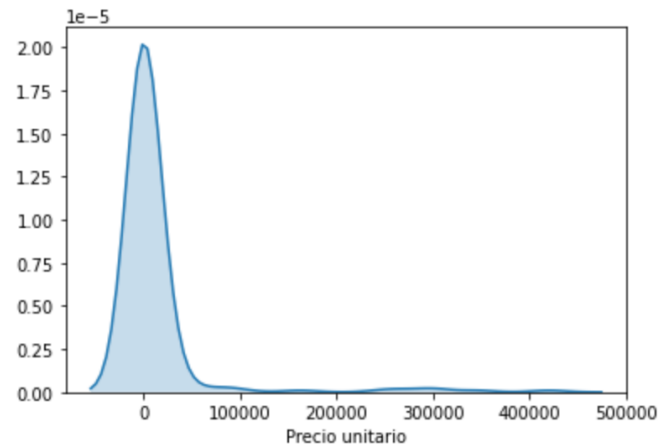


Figura 1. Distribución del Precio Unitario

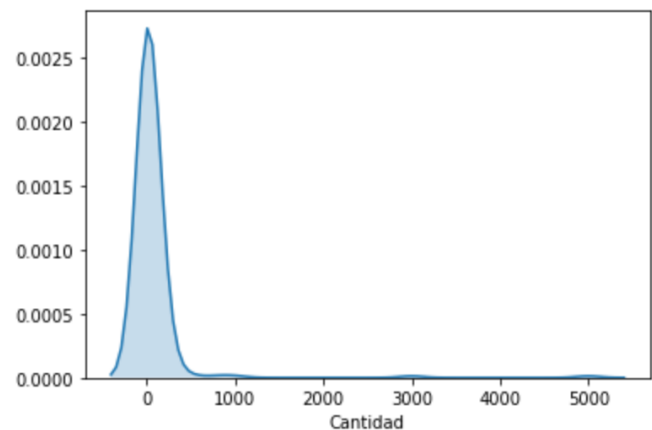


Figura 2. Distribución del Precio Unitario

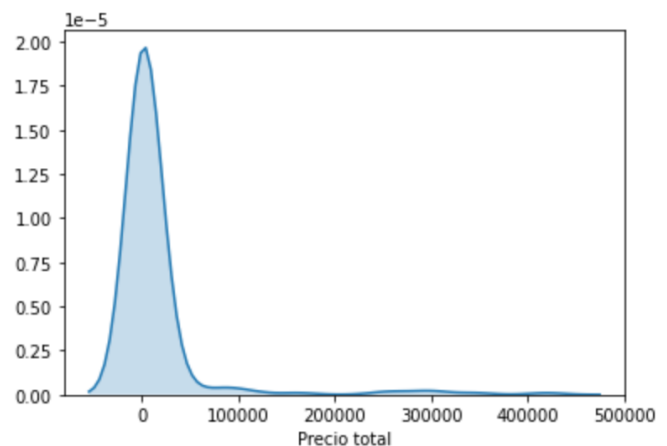


Figura 3. Distribución del Precio Unitario

Se puede ver que en general, los atributos contables de las Fig. 1, 2 y 3 tienen forma acampanada asimétrica con sesgo hacia la derecha. Esto nos indica que la gran mayoría de valores se encuentran en un rango relativamente bajo, pero que existen datos muy aberrantes.

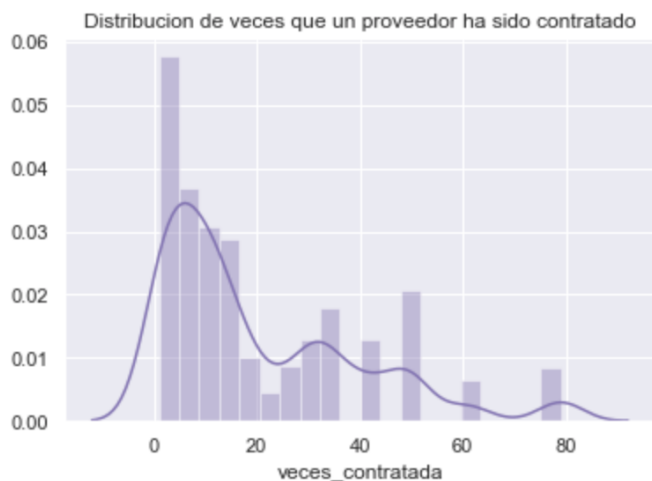


Figura 4. Distribucion de veces que un proveedor ha sido contratado

Observando la distribución de veces que un proveedor ha sido contratado (Fig. 4), resaltan ciertos aspectos. El primero y el más claro, es que existe en esta distribución también cuenta con un sesgo hacia la derecha, siendo lo más común que una empresa sea contratada de 1 a 5 veces (6%). Sin embargo, dicha distribución tiene ciertos picos en la cola.

Al analizar registros con datos aberrantes (Fig. 5), es posible observar que muchos de ellos contienen datos nulos. En muchos casos estos datos se deben limpiar. Sin embargo en esta ocasión estos registros podrían indicar una omisión intencional de información. Es decir, considerar los registros nulos podría ser útil para la detección de corrupción.

Objeto contractual	Unidad de medida	Cantidad	Descripción compra por ítem	Precio unitario	Precio total
ADQUISICIÓN DE MASCARILLAS DESCARTABLES DE PRO...	Unidad	1.0	NaN	420000.0	420000.0
ADQUISICIÓN DE GEL ANTISEPTICO, MASCARILLA N95...	Unidad	1.0	ADQUISICIÓN DE GEL ANTISEPTICO, MASCARILLA N95...	345361.0	345361.0
ADQUISICIÓN DE MASCARILLAS N95, ANTE EL COVID-...	Unidad	1.0	NaN	300000.0	300000.0
ADQUISICIÓN DE MASCARILLAS KN95, ANTE EL COVID...	Unidad	1.0	NaN	300000.0	300000.0
ADQUISICIÓN DE BATAS, CUBRE ZAPATOS, MASCARILL...	Unidad	1.0	SEGÚN CONTRATO	267248.0	267248.0

Figura 5. Registros con datos aberrantes

Siguiendo con el análisis de datos aberrantes y tomando como referencia un caso mencionado en noticieros nacionales,

se observó el sobreprecio de las fundas para cadáveres. En este caso, como se puede ver en la Fig. 6, la media de sobreprecio llega a un 25000 %. Esta es una confirmación clara de la existencia de corrupción dentro de dichos contratos

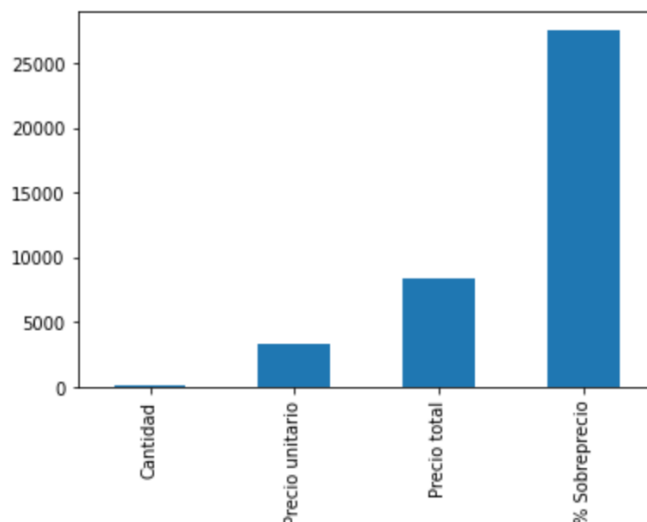


Figura 6. Media de Cantidad, Precio Unitario, Precio total y Porcentaje de sobreprecio para contratos de adquisición de compra de bolsas para cadáveres.

Se puede observar en la Fig. 7 que múltiples entidades incurrieron en estos contratos. Sin embargo, la más notable (Ministerio de Salud Pública Matriz) sobrepasa al resto por varias veces. Se realizó también un análisis sobre los accionistas y administradores de las empresas proveedoras. Se quería conocer si es que estas personas (naturales o jurídicas) estén involucradas con otras empresas, y de esta manera formar redes que se extiendan más allá de los contratos por emergencia sanitaria.

Como se puede observar en la Fig. 8, el 36% de los proveedores que son personas naturales tienen 5 empresas más a su cargo. Es decir, sería factible realizar redes de corrupción que se extiendan más allá de los contratos por emergencia sanitaria. Los años de nombramiento de los proveedores como administradores de las empresas mencionadas en la Fig. 8 pueden ser observados en la Fig. 9. En este caso, La mayor cantidad de nombramientos fueron hechas en el año 2015 (10%).

## V. MÉTODOS

Debido a que la corrupción es un término que puede tomar muchas formas al ser aplicada, para ser detectada es necesario la utilización de diversas técnicas que ayuden a enriquecer los datos disponibles.

### V-A. Categorización por objeto contractual

Al realizar el análisis exploratorio de nuestro dataset, se hizo aparente la necesidad de obtener categorías a partir del objeto contractual de cada registro. Sin embargo, este campo presenta un problema: no existe un formato definido para llenarlo. Por

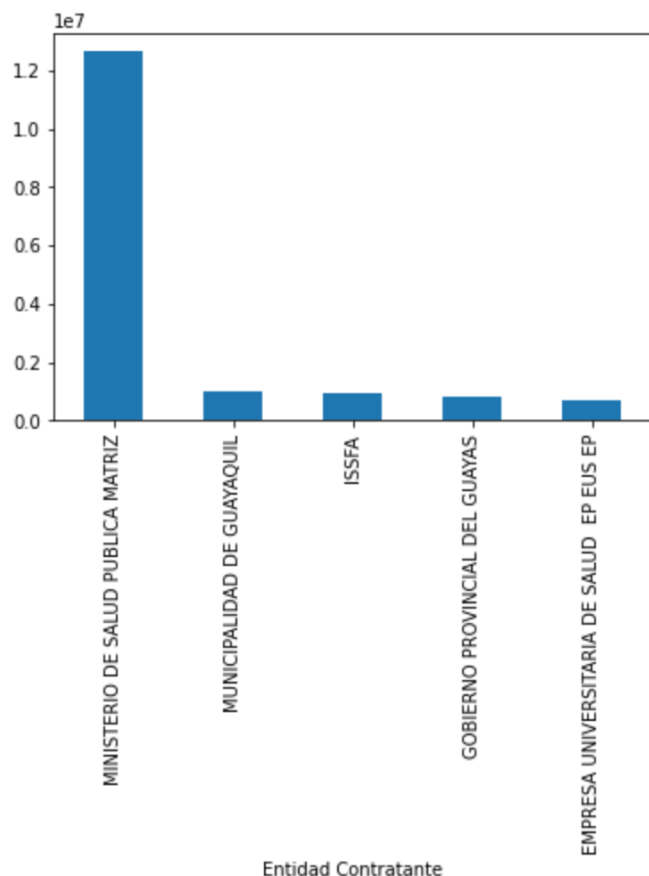


Figura 7. Top 5 de entidades con mayor gasto durante la pandemia en fundas y kits de salud

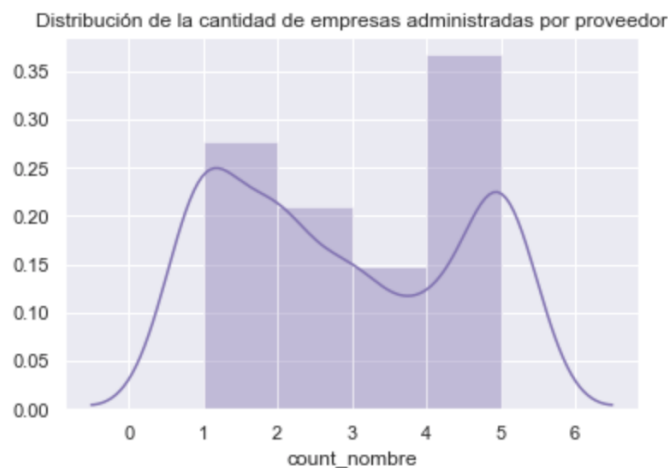


Figura 8. Cantidad de empresas administradas por proveedor en promedio

esta razón se realizó la categorización utilizando técnicas de Natural Language Processing (NLP).

En primer lugar, se tokenizan las palabras. Este proceso incluye la stemización de las palabras, así como la determinación y remoción de las stopwords del corpus. Al terminar el primer paso, se obtiene un vocabulario de 4563 tokens. El siguiente

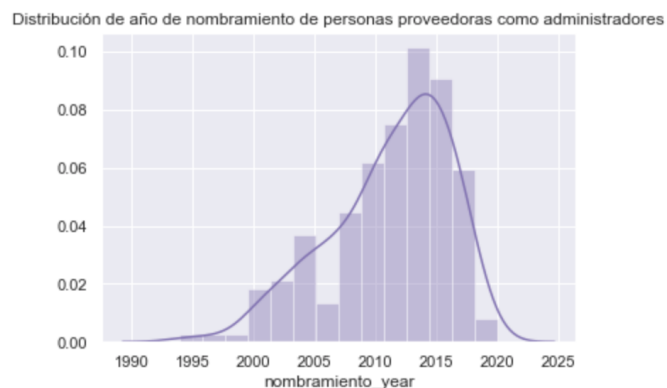


Figura 9. Distribución de año de nombramiento de personas proveedoras como administradores

paso es realizar un análisis de componentes principales (PCA), para reducir la dimensionalidad y obtener mejores resultados en la detección de temas. Se obtuvieron 131 componentes que explican el 70 % de la varianza. Para la categorización de los tópicos se utiliza K-means, junto con el Elbow Method para determinar el número de clusters.

#### V-B. Detección de casos aberrantes

Para la detección de casos de corrupción es necesario determinar todas las contrataciones fuera de lo común. Se utilizó un proceso similar al anterior, sin embargo, se utilizó un algoritmo de clustering diferente. En este caso, se eligió DBSCAN debido a que puede detectar outliers sin tener que agruparlos necesariamente en un cluster.

Los atributos utilizados para la clasificación son los valores contables del contrato (cantidad, precio unitario, precio total) junto con la unidad de medida de lo adquirido (metros, kilogramos, etc) y la categoría obtenida en el punto anterior. Se agregó también un atributo, la relación entre el precio unitario y la cantidad adquirida.

Como fue mencionado anteriormente, es necesario enriquecer los datos. Con este propósito, los atributos contables fueron sometidos a un preprocesamiento particular: se hizo estandarización de acuerdo a las categorías obtenidas. Es decir, se calcularon las métricas de estandarización (media y varianza) en cada grupo y se aplicó Standard Scaler en cada uno de ellos. El razonamiento detrás de la estandarización descrita es que las características de un outlier cambian entre grupos, por lo que para hacer una detección apropiada de los mismos, es necesario hacer una estandarización segmentada.

En este caso también fue necesaria la utilización de PCA. Junto con la aplicación de DBSCAN resultó en la formación de un cluster con 191 outliers. Es decir, se identifican un 0.96 % de casos irregulares.

## VI. RESULTADOS

#### VI-A. Social Network Analysis

En el primer grafo observado en la Fig. 10 se puede observar una revelación importante: casi todos los proveedores poseen un solo accionista.

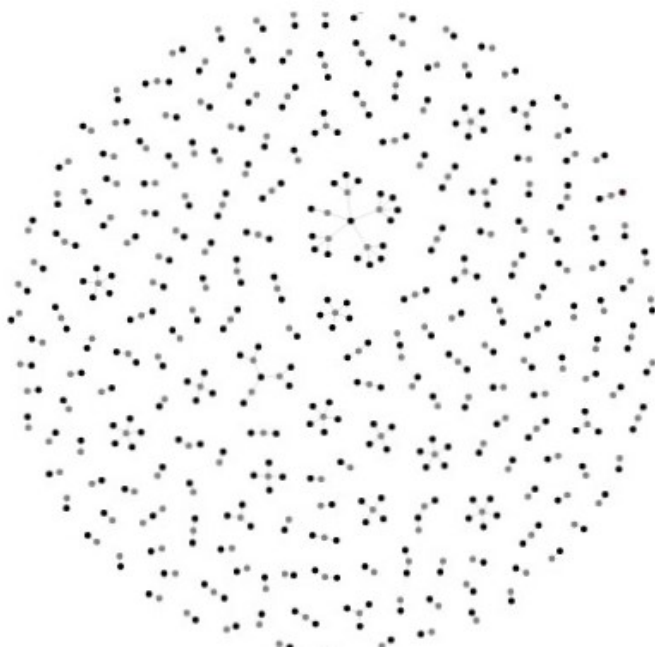


Figura 10. Grafo Proveedores (negro) en relación a los accionistas (gris).

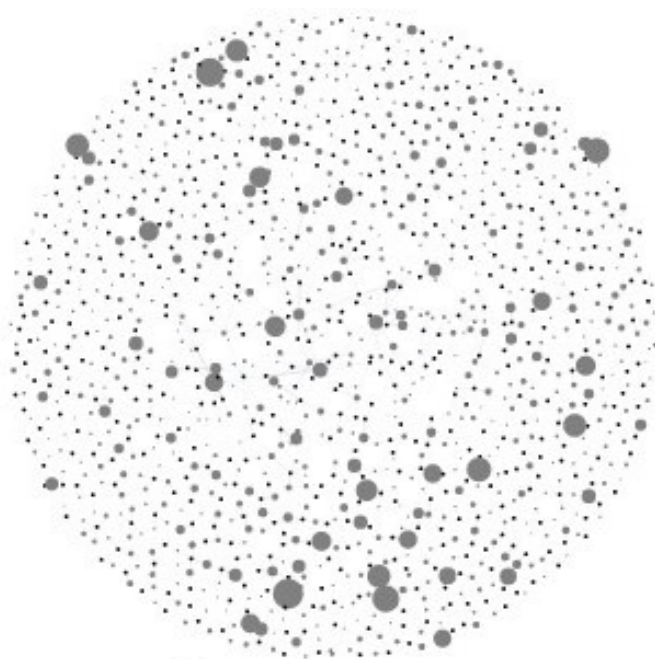


Figura 11. Grafo Recuento de proveedor y entidad contratante.

Mientras que en la Fig. 11 podemos decir que se encuentra una relación geográfica, en las cuales los puntos que se encuentre conectados, comparten un lugar geográfico.

#### VI-B. Detección de tópicos mediante NLP

De la categorización realizada se obtuvieron los siguientes grupos:

1. Kits alimenticios y productos de primera necesidad.
2. Medicamentos e insumos médicos a hospitales.

3. Compra de mascarillas, guantes y elementos de protección de personal.
4. Insumos médicos y de protección y desinfección (agrupados por la palabra emergencia e Insumos, esto puede indicar que fueron compras hechas al inicio).
5. Kits alimenticios e insumos. Diferenciados del primer grupo por nombres largos, parecidos y específicos.
6. Compra de prendas y accesorios de protección para personal no hospitalario.
7. Compra insumos de protección, medicinas y bioseguridad. (Agrupados por tener nombres cortos y específicos. Ej. “ADQUISICION DE KIT DE PROTECCION”).
8. Compra insumos de protección, medicinas, bioseguridad hecho a nivel cantonal, provincial o de parroquia. (Muchos outliers).
9. Compras para emergencia sanitaria no limitadas a insumos del grupo 4. (No contienen la palabra “Insumo”, agrupados por la palabra emergencia, esto puede indicar que fueron compras hechas al inicio)
10. Compras catalogadas como dispositivos médicos (respiradores, jeringas, etc). En general, el objeto contractual incluye las palabras “dispositivos médicos”.
11. Compras catalogadas como equipos de protección de personal. Esto puede incluir equipos de desinfección, mascarillas, trajes, etc. (Se diferencian por tener las palabras “EQUIPOS DE PROTECCIÓN”)

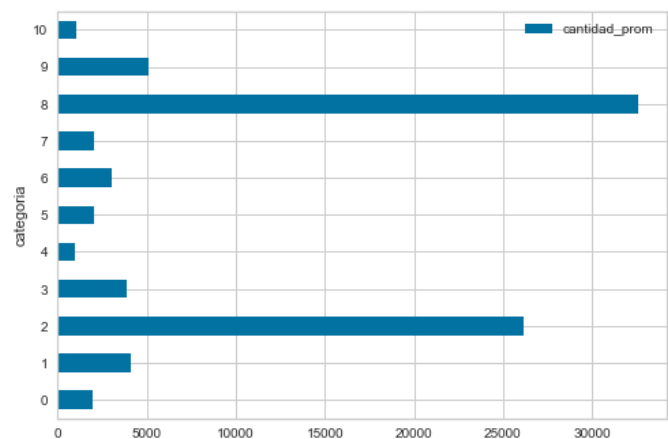


Figura 12. Promedio de cantidades requeridas por categoría durante la pandemia.

Del análisis se pudo observar que las cantidades requeridas en promedio que saltan a la vista fue el grupo 8 y el grupo 2, compra de mascarillas, guantes y elementos de protección de personal y compras para emergencia sanitaria no limitadas a insumos del grupo 3, respectivamente

Por otro lado, si comparamos el total que se gastó por grupos podemos observar en la Fig. 13 que los principales gastos fueron en la compra de prendas y accesorios de protección para personal no hospitalario, compra de insumos de protección y compras de insumos de protección a nivel cantonal o provincial. Cabe recalcar que el grupo 4 que corresponde

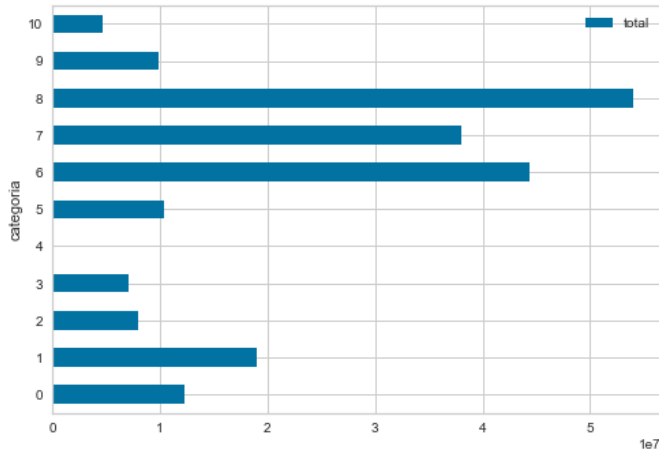


Figura 13. Total en dólares por categoría.

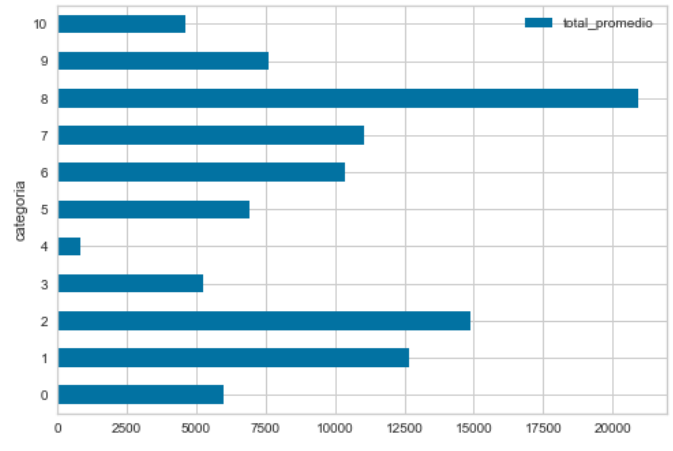


Figura 15. Promedio precio total por categoría.

a alimentos no aparece un total significativo. Al realizar un análisis de la categoría 4, que trata de kits alimenticios se puede observar que en la figura 14 los únicos proveedores durante la cuarentena de kits fueron 5 personas naturales o jurídicas.

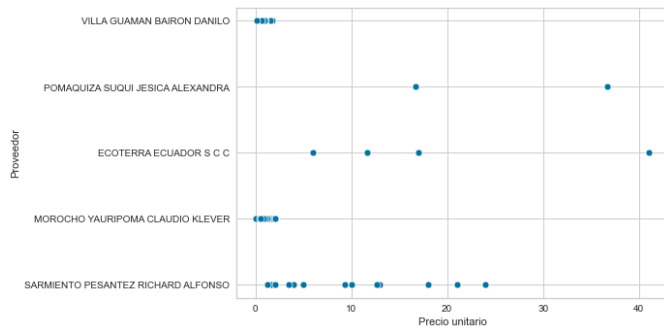


Figura 14. Proveedores kits de alimentos.

Aunque solo existieron 5 proveedores durante la pandemia no se registran datos fuera del margen promedio del precio unitario de \$ 15.4.

A partir de la figura 15 podemos observar que los grupos mas relevante son las categorías 2 y 8, en promedio el gasto de estas categorías ha sido el mayor. En este caso los datos atípicos encontrados no son aportan a la investigación puesto que se encuentran a los extremos de la dispersión de la demanda de cada grupo, a medidas que aumentan las cantidades el precio unitario es bajo y viceversa.

#### VI-C. Detección de casos aberrantes

Para la detección de casos aberrantes existen solamente dos grupos: Aquellos marcados como comunes y aquellos que no pudieron ser encasillados dentro de esta clasificación.

Como se puede observar en las figuras 16 y 17, existe una clara diferenciación entre estos dos grupos. El grupo de los casos no comunes contiene en general un mayor precio unitario, precio total y relación entre precio unitario con

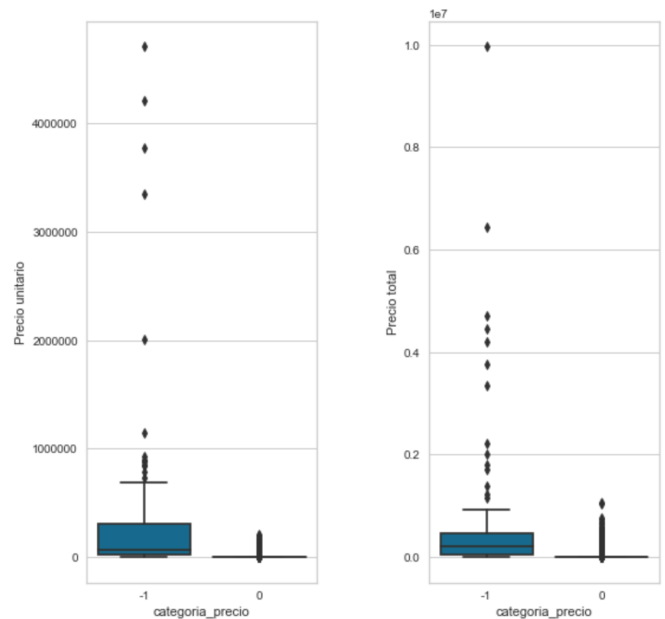


Figura 16. Box plot de Precio unitario y Precio total por grupos.

la cantidad adquirida. Este grupo corresponde solamente a un 0.96 % del total de contratos, sin embargo representa un 44.68 % del total gastado durante la emergencia sanitaria, y está conformado por 190 contratos de los cuales existen 169 proveedores diferentes.

## VII. DISCUSIÓN

Con los resultados obtenidos dentro de la investigación se podrían analizar patrones para diferentes tipos de corrupción en otras áreas correspondientes al estado ecuatoriano, por ejemplo comparar los resultados obtenidos con lavados de dinero, evasión de impuestos, etc. También se debe recalcar que es una de las pocas investigaciones que existen en el país sobre este tema, por lo que abre paso a muchos más estudios de diferentes tipos de corrupción y como emplear métodos para sobrellevar casos relevantes de corrupción.

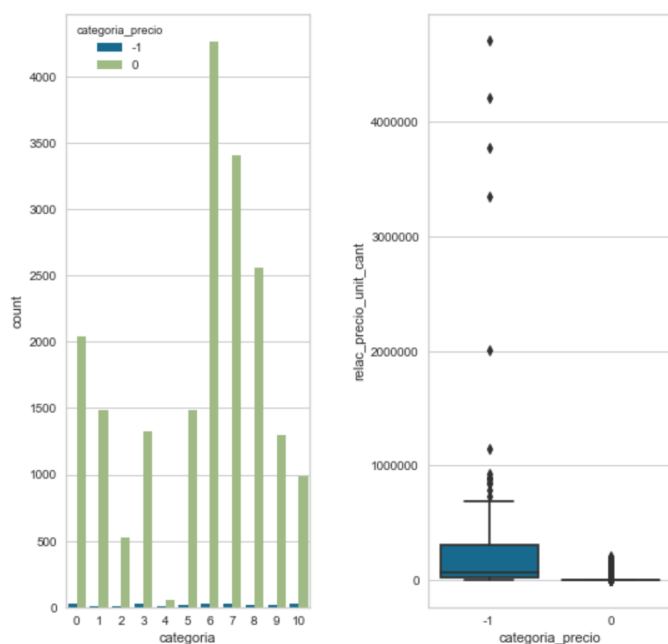


Figura 17. Bar plot de la cantidad de ocurrencias por cada categoría de objeto contractual y Box plot de relacion del precio unitario con la cantidad adquirida.

## VIII. CONCLUSIÓN

Con respecto al analisis de redes que se realizó a los accionas con las empresas se concluyó que en general los proveedores poseen un solo accionista.

De la investigación se obtuvo indicios clave de posibles casos de corrupción, sin embargo es necesario recalcar que la metodología del SERCOP abre paso para que los contratantes y contratistas puedan manipular esta herramienta a su favor puesto que existen valores sumamente aberrantes que son inconclusos por la falta información concluyente a partir de la base de datos obtenida de dicha institución.

Este estudio ofrece una base sobre lo que puede llegar a ser un auténtico sistema de detección de corrupción. Sin embargo, como ya fue mencionado anteriormente, es necesario contar con información concluyente que ayude a marcar un caso sospechoso como corrupción.

## REFERENCIAS

- [1] C. Alexandre and J. Balsa, "Client profiling for an anti-money laundering system," *CoRR*, vol. abs/1510.00878, 2015. [Online]. Available: <http://arxiv.org/abs/1510.00878>
- [2] S. P. Bhosale, "A survey : Outlier detection in streaming data using clustering approached," 2014.