

Escuela Superior Politécnica del Litoral

Introducción a la Ciencia de Datos

Prof: Carmen Vaca, Ph.D.

2020 -1

Detección de redes de corrupción en Ecuador usando fuentes heterogéneas (datos públicos)

Allison Barrezueta Miranda

Felipe González Collantes

Juan Francisco Laso

## Detección de redes de corrupción en Ecuador usando fuentes heterogéneas (datos públicos)

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 (TIO, 2020). 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 (SERCOP, 2020). 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.

## **Recolección de Datos:**

El primer dataset obtenido corresponde a los datos provistos por el [Servicio Nacional de Contratación](#) Pública acerca de las contrataciones hechas debido a la emergencia sanitaria. Dicho dataset fue elegido debido a que nos provee un insight a diferentes aspectos de estas contrataciones. En este caso no fue necesaria la utilización de ningún scrapper debido a que los datos son provistos en formato xls.

A partir de estos datos, con el propósito de obtener una mayor información sobre los proveedores de las diferentes entidades públicas, creamos dos scrappers diferentes. El primero ([obtenido de la búsqueda de empresas en la Superintendencia de Compañías](#)), para obtener datos sobre las empresas contratadas. De aquí deseamos acceder a la información de los accionistas de las empresas. Por otro lado, el segundo scrapper ([obtenido de la búsqueda de personas en la Superintendencia de Compañías](#)) está enfocado en la obtención de datos acerca de las relaciones con empresas de las personas proveedoras de servicios. Es decir, si dicha persona es accionista o administrador en alguna empresa.

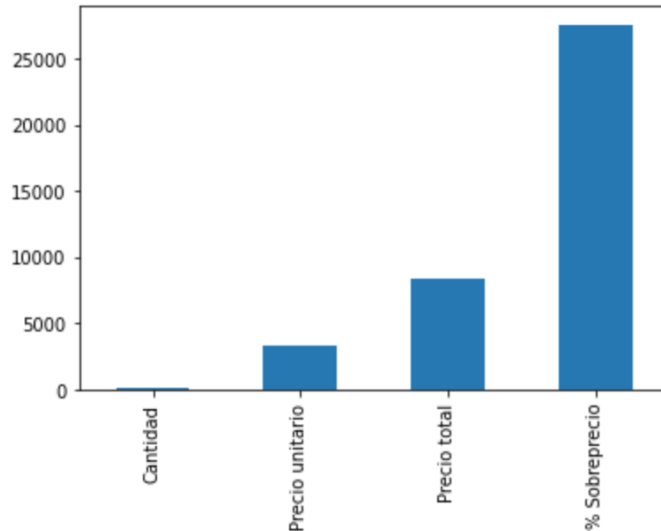
En general, se busca utilizar el primer dataset para el análisis de los contratos y la búsqueda de irregularidades en los mismos, mientras los obtenidos de nuestros dos scrappers funcionan como un apoyo para enriquecer la data.

Con respecto a la investigación, se realizaron preguntas previas para la guía de la investigación:

### **¿Cuál es la brecha entre el precio ofertado y el precio real del bien comprado en los registros de fundas?**

Por medio de validaciones y filtros por productos que inciden en la pandemia, podemos observar que en el producto de fundas de cadáver el promedio del valor real es de \$ 12, un valor obtenido de manera externa, a través de la investigación secundaria. Obtenemos que la brecha o el sobreprecio promedio fue de 275.56 veces el precio real promedio.

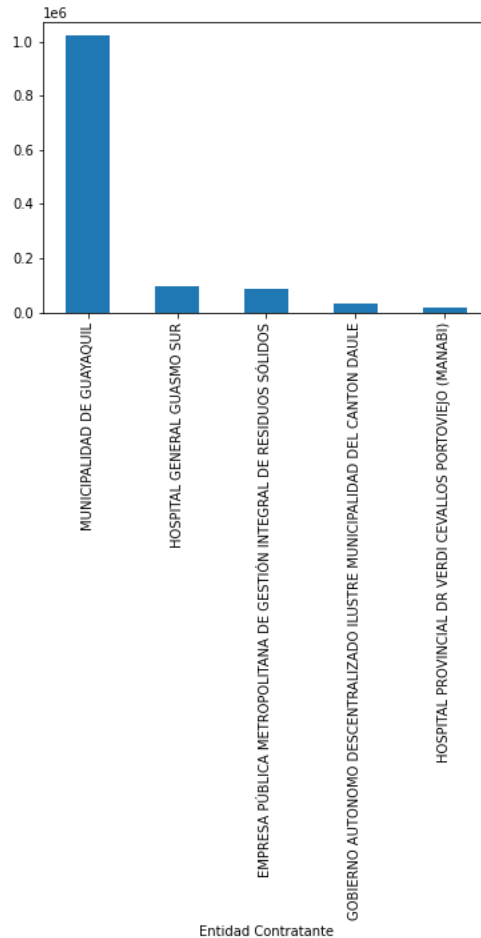
Cantidad 59.827586  
 Precio unitario 3306.836552  
 Precio total 8365.519310  
 % Sobreprecio 27556.971264  
 dtype: float64



**¿Cuáles son las entidades que mas han gastado en esta pandemia en ítems relacionados con el COVID-19?**

Podemos observar que las entidades que más han gastado en la pandemia corresponden a la Municipalidad de Guayaquil con un Precio total de 1.02 millones de dólares en solo 3 proveedores durante los meses de marzo hasta julio. Dado que Guayaquil fue la primera ciudad en ser centro epidemiológico no es de sorprender que el precio total de los contratos obtenido sea el mayor. Aunque en realidad debería pertenecer a entidades de tipo centro de salud como es la segunda entidad del Hospital del Guasmo.

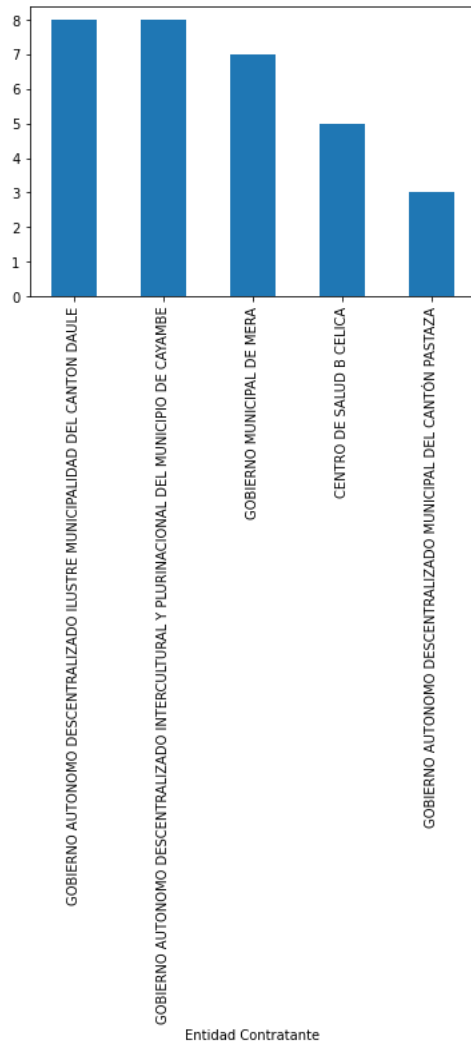
Entidad Contratante	Proveedor		
	Precio unitario	Precio total	
<b>MUNICIPALIDAD DE GUAYAQUIL</b>	3	340000.00000	1020000.0
<b>HOSPITAL GENERAL GUASMO SUR</b>	2	135.60000	95520.0
<b>EMPRESA PÚBLICA METROPOLITANA DE GESTIÓN INTEGRAL DE RESIDUOS SÓLIDOS</b>	1	88800.00000	88800.0
<b>GOBIERNO AUTONOMO DESCENTRALIZADO ILUSTRE MUNICIPALIDAD DEL CANTON DAULE</b>	8	10.76875	32575.0
<b>HOSPITAL PROVINCIAL DR VERDI CEVALLOS PORTOVIEJO (MANABI)</b>	1	16500.00000	16500.0



## ¿Cuáles son las entidades que han tenido más contratos con proveedores en ítems relacionados con el COVID-19?

Podemos observar que las entidades que más han contratado en este estado de emergencia corresponden al GAD del cantón Daule, seguido con GAD del Municipio de Cayambe. Un insight que podemos obtener es que aunque la Municipalidad de Guayaquil ha gastado más en estos productos de emergencia, no ha contratado muchos proveedores.

	Proveedor	Precio unitario	Precio total
Entidad Contratante			
GOBIERNO AUTONOMO DESCENTRALIZADO ILUSTRE MUNICIPALIDAD DEL CANTON DAULE	8	10.768750	32575.00
GOBIERNO AUTONOMO DESCENTRALIZADO INTERCULTURAL Y PLURINACIONAL DEL MUNICIPIO DE CAYAMBE	8	15.525000	6295.60
GOBIERNO MUNICIPAL DE MERA	7	21.350000	5145.10
CENTRO DE SALUD B CELICA	5	47.740000	4695.60
GOBIERNO AUTONOMO DESCENTRALIZADO MUNICIPAL DEL CANTÓN PASTAZA	3	1114.843333	8976.03



¿Qué tipo de relaciones existen entre las entidades que gastan con las personas más nombradas al momento de realizar o designar algún tipo de contrato/compra?

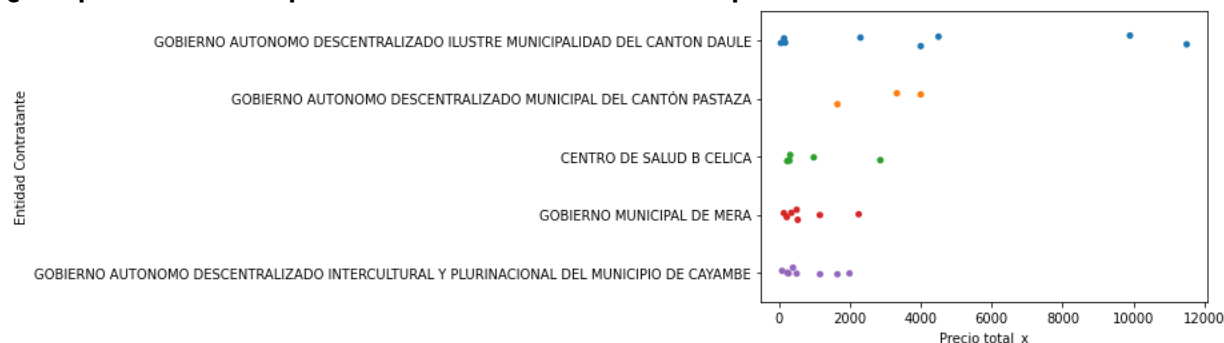
	veces_contratada	Precio total
veces_contratada	1.000000	-0.349134
Precio total	-0.349134	1.000000

Existe una correlación baja negativa entre las veces que una empresa o persona fue contratada y el precio total del contrato. Esto quiere decir que existe una tendencia débil que indica que a mayor precio del contrato recibido, dicha empresa o persona contratada tendrá menos contratos.

**¿Se puede encontrar algún tipo de incidencia/patrón en la contratación de proveedores por parte de una entidad pública?**

Durante el análisis se concluyó que no se buscan patrones que inciden, se buscan anomalías o datos que sean irregulares, a partir de estos casos se detectan patrones de casos irregulares.

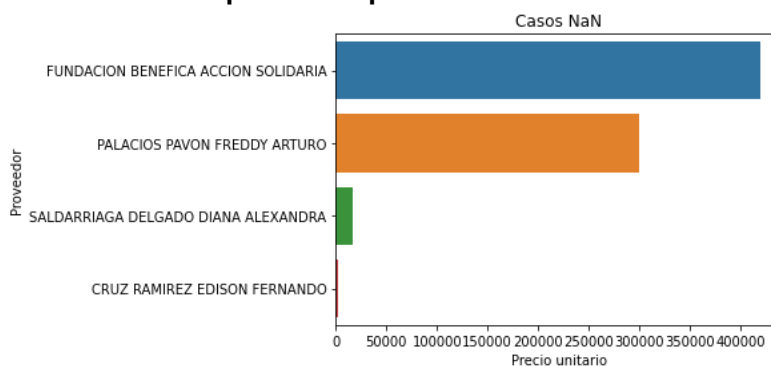
**¿En qué entidades públicas se frecuentan este tipo de acontecimientos?**



**¿Qué casos son relevantes para la investigación de corrupción de productos importantes en la pandemia?**

Hemos identificado dos casos a estudiarse que son casos con ítems no especificados dentro de la plataforma y casos de cantidades unitarias de productos que claramente no son unitarios.

**¿Qué resultados se obtuvo después de aplicar las características?**



Obtuvimos los siguientes proveedores:

- FUNDACION BENEFICA ACCION SOLIDARIA
- PALACIOS PAVON FREDDY ARTURO
- SALDARRIAGA DELGADO DIANA ALEXANDRA
- RUZ RAMIREZ EDISO FERNANDO

**¿Existen ítems para los cuales se den casos irregulares con más frecuencia?**

De acuerdo con el análisis realizado, los ítems con más casos irregulares son aquellos de necesidad inmediata y de gran volumen.

Específicamente, las variables provienen de licitaciones, anuncios de adjudicación de contratos, avisos de modificación de contratos, anuncios de finalización de contratos y avisos de corrección administrativa. En primer lugar, se puede recopilar una base de datos cuantitativa capturando los archivos de texto de los anuncios de la fuente oficial en línea, como un boletín de contratación pública; y en segundo lugar, aplicar estrategias de minería de texto tanto automáticas como manuales, dando lugar a variables con un significado claro y categorías bien definidas. Las variables de interés incluyen, entre otras, el nombre del proveedor, el valor del contrato y la fecha límite para presentar ofertas (M. Fazekas, I. János, 2014)).

### **Insights:**

- Al tener outliers de tanta diferencia se presenta una gran dificultad para observar los datos graficados.
- Aunque la Municipalidad de Guayaquil ha gastado más en estos productos de emergencia, no ha contratado muchos proveedores.
- Encontramos NaNs que en este caso aparentemente tenían un significado: esconder información.
- En general, los contratos de insumos médicos de montos muy altos no tienen toda la información.
- No es posible establecer los márgenes de ganancia de los proveedores en los contratos a partir de datos heterogéneos.