



82.04 Analítica Descriptiva

Segunda Entrega

Azul de los Ángeles Makk - 61.589

Francisca Sulzberger - 61.636

Fecha de entrega: 23.05.2023

Índice

Introducción	3
Marco teórico	3
Metodología	4
Análisis exploratorio	5
Missings	6
Outliers	7
Preguntas a la base de datos	8
1. Dependencia	9
2. Tipo de solicitante	10
3. Dependencia de ministerio	11
4. Ministerio	12
5. Calidad de respuesta	13
6. Categoría tema	14
7. Categoría tema específico	16
Hipótesis establecidas	17
Bibliografía	23

Introducción

El acceso a la información pública tiene como principal objetivo el generar mayor transparencia, reducir la corrupción, mejorar la calidad de las instituciones, entre otras cosas. El derecho de acceso a la información pública que se nos concede habilita la búsqueda, pedido, recepción, análisis, procesamiento y distribución.

Marco teórico

La información pública son todos los datos que generan, obtienen, transforman, controlan o cuidan los organismos del Estado y empresas indicados en la ley 27.275. La información debe estar de forma completa, en la mayor cantidad de medios posibles, en el mayor nivel de precisión posible y en formatos digitales abiertos (excepto que sea imposible o muy costoso entregar la información de esta manera). Los datos tienen que estar contenidos en documentos de cualquier formato o soporte.

La fuente que se utilizará para responder las preguntas que serán planteadas a continuación es una base de datos que da cuenta de los pedidos de acceso a la información pública solicitados al Gobierno de la Ciudad de Buenos Aires conforme a Ley 104, mediante su portal de datos públicos generados por el mismo.

Asimismo, la Ley 104 de la Ciudad de Buenos Aires declara como principio de aplicación (artículo 2):

“El Derecho de Acceso a la Información Pública se interpretará conforme a la Constitución de la Nación, Constitución de la Ciudad Autónoma de Buenos Aires y a la Declaración Universal de los Derechos Humanos, al Pacto Internacional de Derechos Civiles y Políticos, a la Convención Americana sobre Derechos Humanos y a los instrumentos jurídicos internacionales sobre derechos humanos suscritos y ratificados por la República Argentina. Para la interpretación de esta ley se aplicarán los siguientes principios: de máxima premura, presunción de publicidad y accesibilidad; informalismo, no discriminación, eficiencia,

completitud, disociación, transparencia, formatos abiertos, alcance limitado de las excepciones, in dubio pro petitor, buena fe y gratuidad.”

Los pedidos analizados abarcan desde el año 2013 hasta la actualidad. Su actualización es realizada trimestralmente y su última modificación es el 29 de marzo de 2023 -siendo su fecha de publicación el 10 de mayo de 2021-.

Objetivos

El propósito fundamental de la presente investigación radica en asegurar la plena garantía del acceso a la información pública, procurando alcanzar niveles óptimos de exhaustividad, celeridad y satisfacción.

En definitiva, el análisis de esta base de datos buscar la respuesta a las siguientes preguntas:

1. ¿Cuáles son los pedidos a dependencias, ministerios y dependencias ministeriales que son más propensos a ser respondidos a término?
2. ¿Qué categorías de temas son respondidas mayormente de manera completa?
3. ¿Hay algunos ministerios, dependencias, tipo de solicitante o categorías que son más eficientes -respuestas completas y a término-?

Metodología

Para evaluar la eficiencia del sistema de información, consideramos que en primer lugar esta debe ser dentro del plazo legal estipulado y que esta sea de manera completa, a excepción de los casos en los que la consulta excede los límites de lo estipulado por la Ley de Acceso a la Información Pública.

Por un lado, el primer elemento que integra nuestro estimador de referencia es la proporción de pedidos que no requirieron una prórroga. De esta manera, consideramos aquellos que a mayor incumplimiento, menor será el puntaje final de la eficiencia. Dicha proporción varía entre el 0 y el 1.

Luego, por otro lado, evaluamos que aquellos que cumplan con una respuesta correspondiente, favorecen positivamente a nuestro indicador, A mayor respuestas completas, mayor será la proporción de completitud. Para este caso, no analizamos aquellos pedidos que no pueden ser respondidos ya que no se encuentran dentro del marco de la Ley de Acceso a la Información Pública.

Por ende, la fórmula que utilizaremos se articula de la siguiente manera:

$$\text{Eficiencia} = \text{Compleitud} * \text{Cumplimiento a termino}$$

A modo de aclaración, se indica que en la variable ministerio se incluyen todos aquellos fuera de nivel del GCBA, tengan o no un rango ministerial. A partir de julio 2017 se modificó el criterio de carga, incluyendo solamente los ministerios propiamente dichos y las reparticiones que son fuera de nivel con rango ministerial. Desde el segundo semestre del 2017 se cuenta con la información de tiempo de respuesta.

Análisis exploratorio

La base de datos contiene un total de 24.895 observaciones, así también como un total de 22 columnas.

Las variables de tipo texto son:

- tipo_solicitante: categorización del solicitante que realizó el pedido.
- dependencia: dependencia de ingreso de pedido de información.
- tema: descripción del pedido.
- categoria_tema: categorización del tipo de pedido de información.
- categoria_tema_especifico: categorización más detallada del tipo de pedido de información
- dependencia_ministerio: dependencia ministerial corresponde el pedido de información
- ministerio: ministerio al cual corresponde el pedido de información¹.

¹ A modo de aclaración, se indica que en la variable ministerio se incluyen todos aquellos fuera de nivel del GCBA, tengan o no un rango ministerial. A partir de julio 2017 se modificó el criterio de carga, incluyendo solamente los ministerios propiamente dichos y las

- estado: estado del trámite de pedido de información
- termino: termino en el cual fue entregada la información
- observaciones: observaciones.
- calidad_respuesta: categorización de la respuesta brindada al solicitante.
- prorroga: fecha de prórroga de la entrega de la información solicitada.
- contenido_respuesta: resumen de la respuesta brindada al solicitante.

Las variables de tipo numéricas son:

- periodo: periodo en el que se realiza el pedido de información
- id_solicitante: número de identificación del solicitante de información
- expediente: número de expediente asignado
- tiempo en responder: cantidad de días corridos desde que ingresa la solicitud hasta que se responde

Las variables de tipo fecha son:

- vencimiento_1: fecha del vencimiento del reclamo de la información
- vencimiento_2: fecha del vencimiento con prórroga del reclamo de la información
- fecha_de_respuesta: fecha de entrega de la respuesta al pedido de información
- fecha: fecha de ingreso del pedido de información

Missings

En primer lugar, previo a realizar un análisis de los datos, es importante conocer cómo se compone cada campo. Para esto realizamos un análisis exploratorio donde se identifican valores faltantes o "Missings".

En el análisis exploratorio de datos realizado, se identificaron valores faltantes en la variable "tiempo_de_respuesta", la cual representa la diferencia en días entre la fecha del pedido y la fecha de respuesta.

reparticiones que son fuera de nivel con rango ministerial. Desde el segundo semestre del 2017 se cuenta con la información de tiempo de respuesta.

A su vez, se realizó un test de chi cuadrado para entender cómo era la relación entre "estado" de la solicitud, campo que representa la situación de la solicitud, ya sea "Finalizado" o "Pendiente". En este caso, como hipótesis nula se consideró que es independiente el estado del tiempo de respuesta.

Sin embargo, al realizar el test, obtuvimos como resultado un test significativo, lo que indica que hay evidencia para rechazar la hipótesis nula. Por lo tanto, la variable "estado" y "tiempo_de_respuesta" no son independientes entre sí y existe una relación significativa entre ellas.

Consiguientemente, con el objetivo de abordar esta ausencia de datos, se aplicó una estrategia de imputación específica. En este caso, se completaron los valores faltantes únicamente para aquellos registros en los que el campo "estado" indicaba que la solicitud estaba "Finalizada". Para estos casos, se utilizó la diferencia de fechas correspondiente a cada pedido para reemplazar los valores faltantes en "tiempo_de_respuesta".

No fueron completados los valores faltantes de los registros que contenían en el campo "estado" una solicitud "Pendiente".

De esta manera, se pudo obtener una estimación más precisa y completa de los tiempos de respuesta para los pedidos finalizados en el conjunto de datos.

Outliers

A fines de realizar un trabajo exploratorio de la base de datos, se llevará a cabo un análisis de la variable tiempo_de_respuesta que configura un eje principal para el desarrollo de los indicadores enunciados previamente. Con el objetivo de delimitar los outliers se llevarán a cabo dos métodos: por rango intercuartílico y por percentil 0.99.

1. Rango intercuartílico: En primer lugar obtuvimos los cuantiles 0.25 y 0.75 siendo ellos 4 y 23 respectivamente. Luego, obtuvimos el RIC siendo el mismo la diferencia entre ambos y con él se estimaron las cotas superiores e interiores, restándole al cuantil 0.25 y sumándose al cuantil

0.75. Sacando los valores nulos, se obtuvo un total de 1726 outliers.

2. Por percentil 0.99: En este caso nuevamente se estimó mediante métodos no paramétricos el cuantil 0.99. Luego, obtuvimos la cantidad de registros que exceden tal cifra de 140 días, siendo tal 245.

La diferencia encontrada entre el primer método y el segundo corresponde a las observaciones que se ubican entre la cota superior (correspondiente al cuantil no paramétrico 0.75 adicionando el ric) y el percentil 0.99.

Preguntas a la base de datos

Las preguntas a la base de datos corresponden a conocer principalmente los rangos generales de la información presente. Para conocer ello implica en primer lugar agrupar función a diferentes criterios (dependencia, ministerio, dependencia de ministerio, categoría, categoría tema específico, tipo de solicitante y calidad de respuesta). Una vez realizado el agrupamiento, buscamos conocer la cantidad de registros por respuesta y el promedio en el tiempo de respuesta, ambos ordenados de mayor a menor.

Antes de comenzar, y en pos de estipular nuestro objetivo, se obtendrán dos ratios que resultan de vital importancia calcular, ya que de tal depende el resultado final de eficiencia previamente estipulado:

- Incumplimiento temporal: Corresponde a los registros que acorde a su designación y plazo temporal han requerido una prórroga. Utilizamos el atributo 'prórroga' que responde a si el tiempo de respuesta tuvo que ser extendido a fines de brindar una respuesta. El cálculo del incumplimiento temporal se articula de la siguiente manera:

$$\text{Incumplimeinto temporal} = \frac{\text{Cantidad de pedidos que requirieron prórroga}}{\text{Cantidad de pedidos total}}$$

- Incumplimiento de completitud: Corresponde a los registros que no han sido contestados de manera completa, tal siendo obtenido del atributo 'calidad_respuesta'. El cálculo del incumplimiento de completitud se articula de la siguiente manera:

$$\text{Incumplimiento de completitud} = \frac{\text{Cantidad de pedidos respondidos de manera incompleta}}{\text{Cantidad de pedidos total}}$$

Es importante tener en cuenta que tanto el Incumplimiento temporal como el Incumplimiento de Completitud arrojan como resultado ratios. Es por esto que en las tablas se verán reflejados como un cero seguido de números decimales.

1. Dependencia

La información contenida en la columna dependencia hace referencia a la dependencia de ingreso del pedido de la información. En la base de datos se enlistan un total de 79 dependencias. A modo de ejemplo, se presenta en la Tabla I una tabla con 10 dependencias.

Procedemos en primer lugar a analizar la cantidad de registros por dependencia, ordenado de mayor a menor. La dependencia que presenta es DGSOCAI (Dirección General de Seguimiento de Organismos de Control y Acceso a la Información) con 17.933 registros, siendo que la cantidad de registros que la misma presenta es mayor a cuatro veces más que la dependencia que se ubica en segundo puesto (Dirección General Mesa de Entradas, Salidas y Archivo con 4.116 registros)

Luego, analizamos el promedio de días de respuesta por dependencia, ordenado de mayor a menor. Véase que existe una gran tardanza concentrada para dos dependencias: DGSOCAI con un promedio de 160 días y MGEYA (Mesa General de Entradas, Salidas y Archivo) con un promedio de 112 días.

Además, como fue mencionado anteriormente, se ha calculado el incumplimiento tanto temporal como de completitud. Es así como en la Tabla I se pueden observar los resultados de ambos cálculos para las dependencias.

Dependencia	Cantidad	Promedio	Incumplimiento temporal	Incumplimiento de completitud
DGSOCAI	17933	160	0,26	0,0250
MGEYA	4116	112	0,32	0,0352
UAC1	1114	63	0,62	0,0458
COMUNA1	926	48.4	0,78	0,0605

AJG	285	47	0,50	0,0175
UAC9	131	45	0,01	0,0200
COMUNA10	39	36	0,62	0,0769
UAC14	39	31	0,33	0,0344
COMUNA13	37	30.6	0,19	0,0458
AGC	30	30.4	0,47	0,0701

Tabla I: Tabla de pedidos por dependencia (con 10 ejemplos)

2. Tipo de solicitante

El campo 'tipo de solicitante' almacena la categorización del solicitante que realizó el pedido, el cual se enmarca dentro de cualquiera de las siguientes 5 categorías: particular, actor político, ONG, periodista u otros.

Para realizar los cálculos pertinentes, fue necesario convertir los datos a mayúscula, debido a que la base contenía cierta inconsistencia al registrar ambos "PARTICULAR" y "Particular" o "PERIODISTA" y "Periodista" como diferentes tipos de solicitantes.

En el primer puesto se ubica la categoría 'PARTICULAR', el cual presenta un total de 22.180 pedidos, casi 17 veces más que 'ACTOR POLÍTICO', ubicado en el segundo puesto.

Por otro lado, al calcular el promedio de días por tipo de solicitante, es posible observar que la categoría que concentra el mayor promedio es la de "Particular" con 52 días, mientras que el menor promedio es para la categoría "Otros", con un promedio de 21,4 días.

Asimismo, fue calculado el incumplimiento temporal, el cual se presenta con el mayor valor para la categoría "Otros", y con el menor valor para la categoría "Particular", y el incumplimiento de completitud, en el cual se presenta la misma distribución.

Tipo de solicitante	Cantidad de pedidos	Promedio	Incumplimiento temporal	Incumplimiento de completitud
Particular	22180	52	0,274	0,0257
Actor	1308	29.6	0,631	0,0612

político				
ONG	845	27.6	0,462	0,0462
Periodista	404	23.4	0,564	0,0619
Otros	156	21.4	0,654	0,0833

Tabla II: Tabla de pedidos por tipo de solicitante

3. Dependencia de ministerio

En tercer lugar, procedemos con el análisis del campo de dependencia de ministerio, que como lo indica su nombre detalla la dependencia ministerial del pedido. Existe un total de 383 dependencias ministeriales. A modo de ejemplo, en la Tabla III hemos decidido tomar 10 dependencias ministeriales.

Para su análisis, primero estudiamos la cantidad de cada dependencia del ministerio. Es observable que la Agencia Gubernamental de Control se ubica en el primer puesto dentro de todas las dependencias con 9284 pedidos.

Luego, calculamos el promedio de días de demora en la respuesta. De esta manera, observamos que la D.G. Legal y Técnica MSGC presenta el mayor promedio con 607,02 días.

Además, se realizaron los cálculos de incumplimiento temporal e incumplimiento de completitud.

Dependencia Ministerio	Cantidad Total	Promedio	Incumplimiento temporal	Incumplimiento de completitud
Agencia Gubernamental de Control	9284	177,32	0,019	0,0043
D.G Registro de Obras y Catastro	1839	5,00	0,003	0,0027
D.G. Coord.Legal e Institucional	1147	14,60	0,586	0,1240
Subsec. Gestión Comunal	1084	15,5	0,006	0,0055
D.G. Legal y Técnica MSGC	804	607,02	0,755	0,0485

D.G. Tec. Adm y Legal MJYSGC	551	37,90	0,056	0,0563
Agencia de Protección Ambiental	503	24,60	0,557	0,0318
Secretaría Transporte y Obras Públicas	377	7,17	0,369	0,0186
Sec. Atención y Gestión Ciudadana	370	8,50	0,816	0,0243
D.G. Tec. Adm. y Legal MAYEPGC	343	24,80	0,848	0,0262

Tabla III: Tabla de pedidos por dependencia ministerial (con 10 ejemplos)

4. Ministerio

En cuarto lugar analizamos el campo ministerio, al cual le corresponde la información que se solicita. Enlistados, existen un total de 37 ministerios. Sin embargo, a modo de ejemplo en la Tabla IV se muestran 10 ministerios.

En principio, ordenamos a los ministerios en función de la cantidad de pedidos. De esta manera, vemos que es el Ministerio de Salud el que presenta la mayor cantidad de pedidos, con un total de 847, siendo la Secretaría de Comunicación, Contenidos y Participación Ciudadana la que cuenta con el menor número de pedidos, 17.

Finalmente se realizaron los cálculos de incumplimiento temporal e incumplimiento de completitud. En este último, es posible notar una gran diferencia en el resultado obtenido para el Ministerio de Educación, el cual tiene el mayor ratio de incumplimiento de completitud, con 0,213.

Ministerio	Cantidad Total	Promedio	Incumplimiento temporal	Incumplimiento de completitud
Ministerio de Salud	847	29,7	0,795	0,0531
Ministerio de Ambiente y Espacio Público	828	22,9	0,748	0,0278
Ministerio de Educación	512	23,1	0,924	0,2130
Ministerio de Espacio Público E Higiene Urbana	388	22,9	0,784	0,0412

Secretaría de Ambiente	367	19,9	0,670	0,0409
Ministerio de Desarrollo Económico y Producción	309	25,4	0,896	0,0485
Secretaría de Medios	81	23,1	0,864	0,0617
Secretaría General y Relaciones Internacionales	64	21,6	0,656	0,0312
Consejo Derechos Niñas, Niños y Adolescentes CABA	58	23,4	0,741	0,0517
Secretaría de Comunicación Contenidos y Participación Ciudadana	27	19,7	0,630	0,0606

Tabla IV: Tabla de pedidos a ministerios (con 10 ejemplos)

5. Calidad de respuesta

El campo 'calidad de respuesta' categoriza la respuesta que fue brindada al solicitante. Las posibles respuestas varían en 6 tipos de calidades: 'COMPLETA', 'DENEGATORIA INFUNDADA', 'PARCIAL', 'DENEGATORIA FUNDADA', 'DENEGADA' y 'LEY 104 ART. 4'. En cuanto a la última categoría, se cita a tal artículo de la Ley 104 de la Ciudad Autónoma de Buenos Aires:

“Alcances: Deberá proveerse la información contenida en documentos escritos, fotográficos, grabaciones, soporte magnético, digital o en cualquier otro formato, incluyendo bases de datos, acuerdos, directivas, reportes, estudios, oficios, proyectos de ley, disposiciones, resoluciones, providencias, expedientes, informes, actas, circulares, contratos, convenios, estadísticas, instructivos, dictámenes, boletines o cualquier otra información registrada en cualquier fecha, forma y soporte; que haya sido creada u obtenida por el órgano requerido, y que se encuentre en su posesión y bajo su control.”

Teniendo esto en consideración, procedemos a enlistar de mayor a menor cantidad de pedidos, las calidades de respuesta.

Es posible observar que las respuestas de tipo completa predominan ampliamente por sobre las otras, con un total de 23.022 pedidos (más de 34 veces más que la categoría de tipo parcial, que se ubica en segundo puesto con 669 pedidos).

Para profundizar nuestro análisis sobre la calidad de respuesta, enlistamos las calidades de respuesta en función al promedio de días.

Calidad de respuesta	Cantidad Total	Promedio	Incumplimiento temporal
Completa	23022	18,500	0,297
Parcial	669	31,300	0,534
Denegatoria Fundada	556	15,800	0,293
Denegada	218	14,700	0,156
Denegatoria Infundada	57	95,100	0,544
Ley 104 Art. 4	1	20,000	-

Tabla V: Tabla de calidad de respuesta de los pedidos

6. Categoría tema

Procedemos luego al análisis de categoría tema, campo que indica la categorización del tipo de pedido de información. Existen un total de 15 categorías-tema en total, las cuales se enlistan a continuación en la Tabla VI.

Así como fue realizado previamente, procedemos a enlistar de mayor a menor en función a la cantidad de observaciones por categoría tema. La categoría tema con mayor pedidos es "Habilitaciones, Inspecciones e Infraestructura" con 11.351 pedidos y aquella con la menor cantidad, "Cultura" con 175 pedidos.

Por consiguiente, analizamos luego el promedio de días en la respuesta por categoría tema. Observamos que aquella con mayor promedio en tiempo de respuesta es "Temática Habitacional o Situación De Calle" con 55,5 días mientras que aquella con menor promedio es la categoría tema "Habilitaciones, Inspecciones e Infraestructura" con 7,08 días.

De igual manera que para el resto de los campos, se realizó el cálculo de incumplimiento temporal, en el cual es posible notar que “Servicios de Seguridad” es el tema para el cual hay un mayor incumplimiento temporal, con un resultado de 0,903. Mientras que el más bajo pertenece al tema “Habilitaciones, Inspecciones e Infraestructura” con un resultado de 0,047.

Además, se calculó el incumplimiento de completitud, el cual arrojó como resultado que el tema que tiene un mayor ratio es “Pedidos de Informes y Datos Personales” con 0,380. Mientras que el de menor ratio fue nuevamente “Habilitaciones, Inspecciones e Infraestructura” con un resultado de 0,0052.

Categoría Tema	Cantidad Total	Promedio	Incumplimiento temporal	Incumplimiento de completitud
Habilitaciones, Inspecciones e Infraestructura	11351	7,08	0,047	0,0052
Gestión Pública	2539	46,2	0,650	0,3030
Pedidos de Informes y Datos Personales	2501	21,7	0,423	0,0380
Espacio Público	1632	26,3	0,471	0,0466
Información Financiero Contable	1351	22,4	0,472	0,0370
Servicios Educativos	972	22,1	0,573	0,1310
Aspectos Laborales	958	32,9	0,451	0,0825
Ambiente e Higiene	735	21,9	0,624	0,0354
Servicios de Salud	715	27,0	0,690	0,0490
Servicios en General	488	13,8	0,230	0,0225
Servicios de Seguridad	452	32,2	0,903	0,0752
Temática Habitacional o Situación de calle	432	55,5	0,576	0,0694
Derechos Personales	312	16,8	0,330	0,0303
Información Político Institucional	282	28,1	0,394	0,0567
Cultura	175	19,2	0,314	0,0057

Tabla VI: Tabla de categoría tema de los pedidos

7. Categoría tema específico

Para finalizar el análisis de los campos, decidimos indagar sobre la categoría tema, así analizando la categoría tema específico, el cual responde a una categorización más detallada del campo anteriormente analizado. Existen un total de 75 categoría tema específicos, de los cuales enlistamos 10 a modo de ejemplo de mayor a menor. Tal criterio, nos devolvió que el tema específico con la mayor cantidad de pedidos es 'HABILITACIONES (PEDIDOS DE DATOS, INSPECCIONES, ACTIVIDAD ILEGAL O NO HABILITADA) / LICENCIAS / FUNCIONAMIENTO IRREGULAR.' con un total de 8914 registros, tres veces más que 'OTROS' -que corresponde a las cuales no concuerden con las otras 73 categorías de tema específico.

Por otro lado, también fueron calculados los promedios de días en responder. Tales cálculos nos han arrojado que la categoría tema específico con mayor promedio de días es "Asentamientos y Villas/ Urbanizacion/ Relocalizaciones, Títulos de Propiedad/ Normas Y Reglamentos Internos" con 60. Mientras que la categoría tema específico con menor cantidad de días es "Habilitaciones (Pedidos De Datos, Inspecciones, Actividad Ilegal o No Habilitada) / Licencias / Funcionamiento Irregular (Pedidos De Datos, Inspecciones)" con 5,2.

Luego, al calcular el incumplimiento temporal, obtuvimos que el que mayor ratio tiene es "Presupuestos / Ejecución" con 1. Mientras que el de menor incumplimiento temporal pertenece a "Habilitaciones / Licencias / Funcionamiento Irregular".

Por último, de la misma manera fue calculado el incumplimiento de completitud, el cual demuestra en este caso que el que tiene el mayor resultado fue "Normativa /Incumplimiento de Normas" con 0,5. Mientras que el de menor incumplimiento de completitud fue de "Habilitaciones / Licencias / Funcionamiento Irregular" con 0,0035.

Es menester tener en cuenta que como contamos con muchas categorías distintas, solamente fueron listadas 10 de ellas con el fin de ilustrar y ejemplificar los cálculos hechos para cada categoría.

Las conclusiones que fueron realizadas son en base a la tabla, no al total de categorías que posee la base de datos.

Calidad de respuesta	Cantidad Total	Promedio	Incumplimiento temporal	Incumplimiento de completitud
Habilitaciones / Licencias / Funcionamiento Irregular	8914	5,2	0,022	0,0035
Presupuestos/Ejecución	1060	60,0	1,000	0,0358
Protocolos de seguridad	137	24,0	0,854	0,1170
Servicio de recolección de basura	133	24,3	0,887	0,0526
Actuación policial	83	30,8	0,867	0,0602
Alumbrado	45	29,0	0,911	0,0222
Venta ambulante	44	20,3	0,841	0,0227
Situación de calle	27	23,9	0,815	0,1850
Informes sobre Personal / Sumarios Internos	23	41,4	0,957	0,0435
Normativa /Incumplimiento de Normas	2	46,3	1,000	0,5000

Tabla VII: Tabla de categoría tema específico de los pedidos (con 10 ejemplos)

Hipótesis establecidas

En el siguiente apartado se establecerán tres hipótesis principales, las cuales serán inicialmente exploradas mediante un análisis gráfico y luego probadas mediante un test de hipótesis.

En primer lugar, buscamos evaluar la dependencia entre el tema al que pertenece el pedido y la cantidad de días que se demora la respuesta. Para esto, realizamos la prueba de hipótesis para las variables “Tiempo de respuesta” y “Categoría Tema”, donde se plantean las siguientes hipótesis:

H_0 = El tiempo de demora es independiente de la categoría del tema de la solicitud.

H_1 = El tiempo de demora no es independiente de la categoría del tema de la solicitud.

En términos coloquiales, se busca probar que es indistinta la categoría tema del tiempo de respuesta de la solicitud. Por ende, se procede a realizar un test de chi cuadrado.

Sin embargo, en principio, decidimos realizar una categorización de la variable tiempo de respuesta en cuatro grupos: "Óptimo", "Normal", "Demorado", "Muy Demorado".

Estos grupos fueron hechos en correspondencia con los cuartiles de la muestra respectivamente, es decir: si el tiempo de espera fue menor al primer cuartil, entonces es categorizado como "Óptimo", si se encuentra entre el primer y segundo cuartil, es "Normal", si se encuentra entre el segundo y tercer cuartil, es "Demorado", y en caso de que el tiempo de respuesta sea mayor al tercer cuartil, entonces es categorizado como "Muy demorado".

Luego, procedemos a la creación de una tabla de contingencia que contiene la categoría tema y tiempo de demora categorizado en los cuatro grupos. De esta manera es posible realizar un test de chi cuadrado, ya que esta prueba se realiza para dos variables categóricas.

Como resultado del test, obtuvimos un valor $p < 2.2e-16$, lo que significa que el test resultó significativo. Lo cual nos indica que existe evidencia suficiente para rechazar la hipótesis nula (H_0), ya que existe una dependencia entre las variables categoría tema y tiempo de demora.

Para visualizar la información que nos proporciona el test, hemos realizado un gráfico de caja y bigote - Figura I- donde cada caja representa una categoría tema, y es posible observar la distribución que presenta el tiempo de respuesta en días para cada una de ellas.

A través del gráfico, es posible notar una dependencia entre las variables, tal y como fue indicado por las conclusiones obtenidas con el test de chi cuadrado. También es posible observar outliers o valores atípicos por fuera de todas las cajas. Además, podemos distinguir diferentes tamaños de cajas, debido a que cada una posee distintas dispersiones.

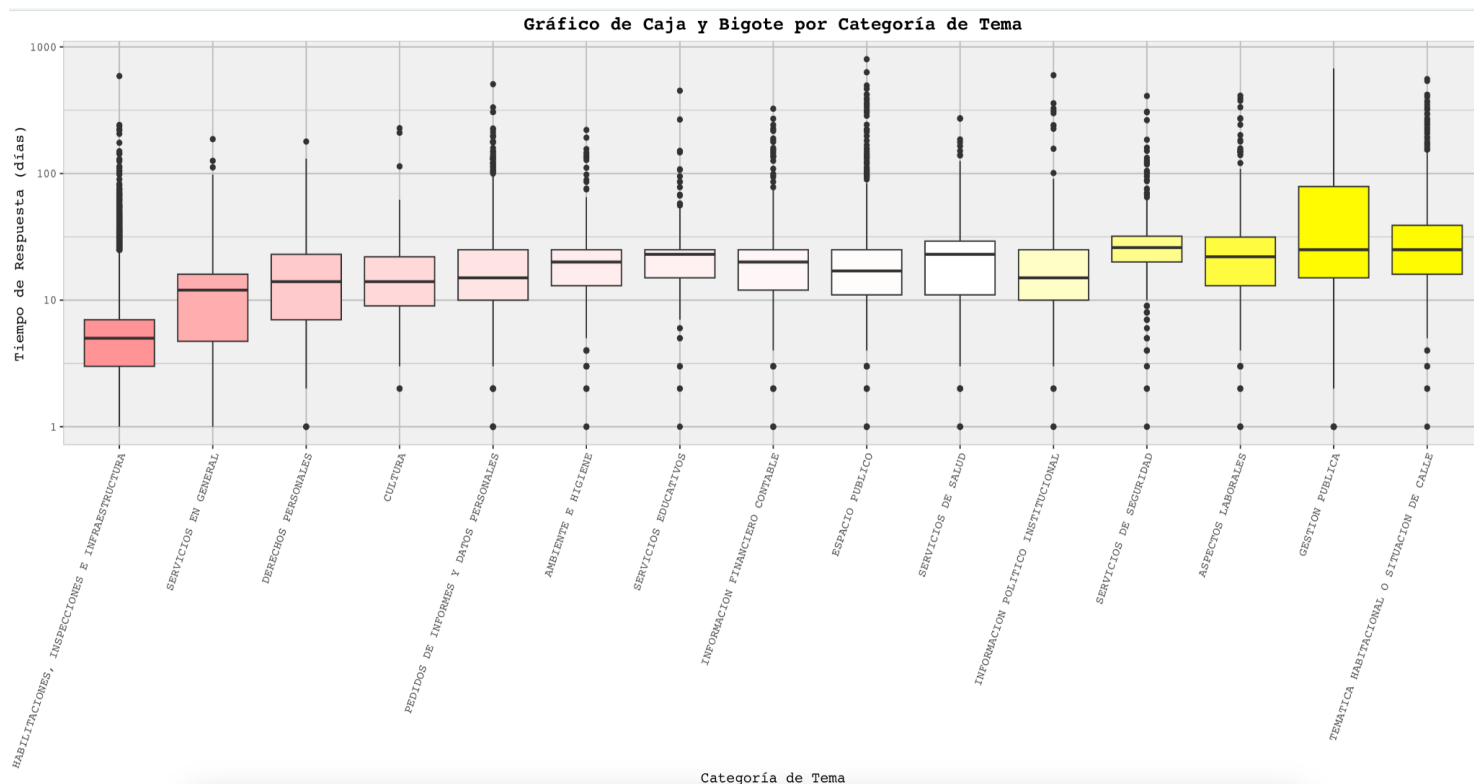


Figura I: Gráfico de caja y bigote del tipo de respuesta en días por categoría de tema

En segundo lugar, buscamos evaluar la relación entre el tiempo de respuesta y la calidad de la misma. Por ende, establecemos la siguiente hipótesis:

H_0 = Existe una correlación positiva entre la media del tiempo de demora a mayor completitud en la respuesta

H_1 = No existe una correlación positiva entre la media del tiempo de demora a mayor completitud en la respuesta

En términos coloquiales, lo que se busca probar con tal hipótesis es que a mayor cantidad de respuestas completas, mayor es el tiempo de demora. Para ello, en primer lugar graficamos las densidades superpuestas de las respuestas completas e incompletas (Figura II). El eje X, siendo en este caso tiempo de respuesta, se ha graficado con una escala logarítmica de base 10.

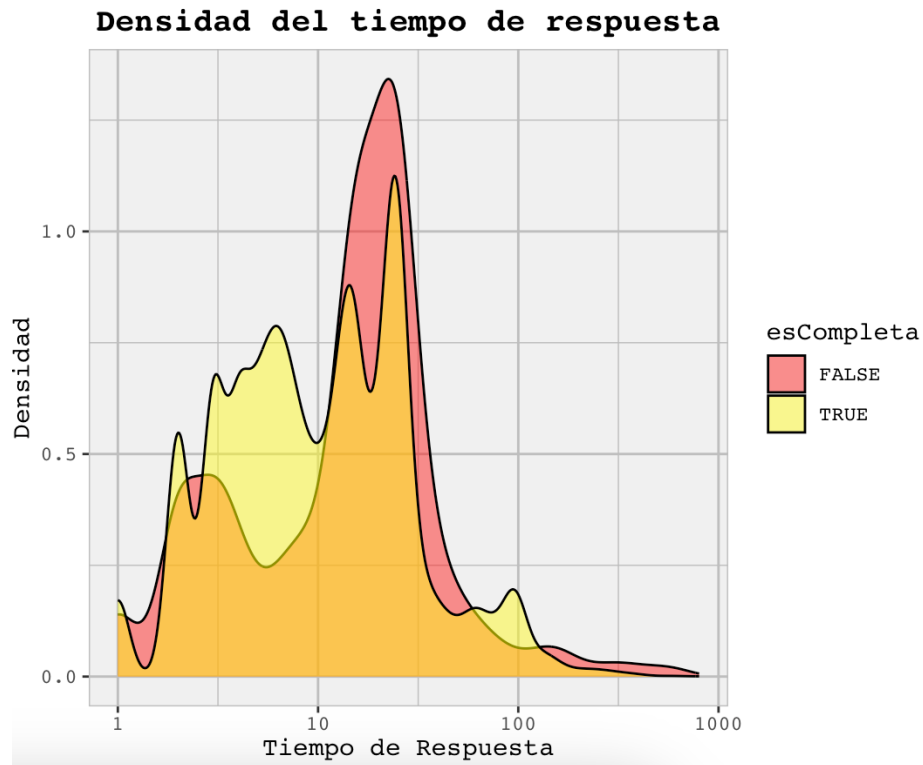


Figura II: Densidad del tiempo de respuesta en función a completitud

Interpretativamente, es posible afirmar que es viable que para las respuestas completas el promedio de tiempo de respuesta sea mayor, dada su distribución.

Luego, con el objetivo de evaluar la diferencia de medias realizamos un t-student, de los cuales partimos de los siguientes supuestos:

1. Los datos son independientes entre sí: En este caso, al tratarse de grupos cuya longitud es distinta, no realizamos un test de correlación.
2. Los datos en cada grupo siguen una distribución normal: Los grupos poseen un total de 23.394 para las respuestas completas y 1.873 para las respuestas incompletas. Para el caso de las completas, al ser ellas mayor a un total de 5.000 no se ha realizado un test de shapiro. Sin embargo, en el caso de las incompletas podemos afirmar que se ha obtenido un p-valor menor al nivel de significancia, lo cual por lo que se puede concluir que existe una relación significativa entre las variables.

3. Existe una homogeneidad en las varianzas: En este caso, realizamos un test para probar que las varianzas dadas que el valor p es menor que el nivel de significancia dado, se concluye que hay evidencia estadística para afirmar que las varianzas entre los grupos son diferentes. Por esta razón, el t-test realizado será de Welch, el que se indica que las varianzas no son iguales.

Finalmente, realizamos el test de Welch, el cual otorga un resultado no significativo. Ello implica que no hay suficiente evidencia para rechazar la hipótesis nula de que indica que la media para las respuestas completas es mayor que para las incompletas. Concluyentemente, no hay diferencias significativas entre las medias de los grupos comparados.

Luego, decidimos analizar el incumplimiento total por ministerio. Para ello, establecimos un nuevo indicador para el incumplimiento, el cual se articula de la siguiente manera.

$$\text{Incumplimiento Total} = \text{Incumplimiento temporal} * \text{Incumplimiento de completitud}$$

De esta manera, acorde a nuestro indicador implica que a mayor cumplimiento -pedidos respondidos a término y de manera completa-, menor será el incumplimiento total calculado. El incumplimiento total fue calculado por el ministerio. De esta manera, establecemos las siguientes hipótesis:

H_0 = Existe una correlación positiva entre la cantidad de pedidos y la completitud.

H_1 = No existe una correlación positiva entre la cantidad de pedidos y la completitud.

A fines de proveer un análisis gráfico de cómo se dispersan estas variables, realizamos un scatterplot de las variables Incumplimiento Total y Cantidad de Pedidos (Figura III).

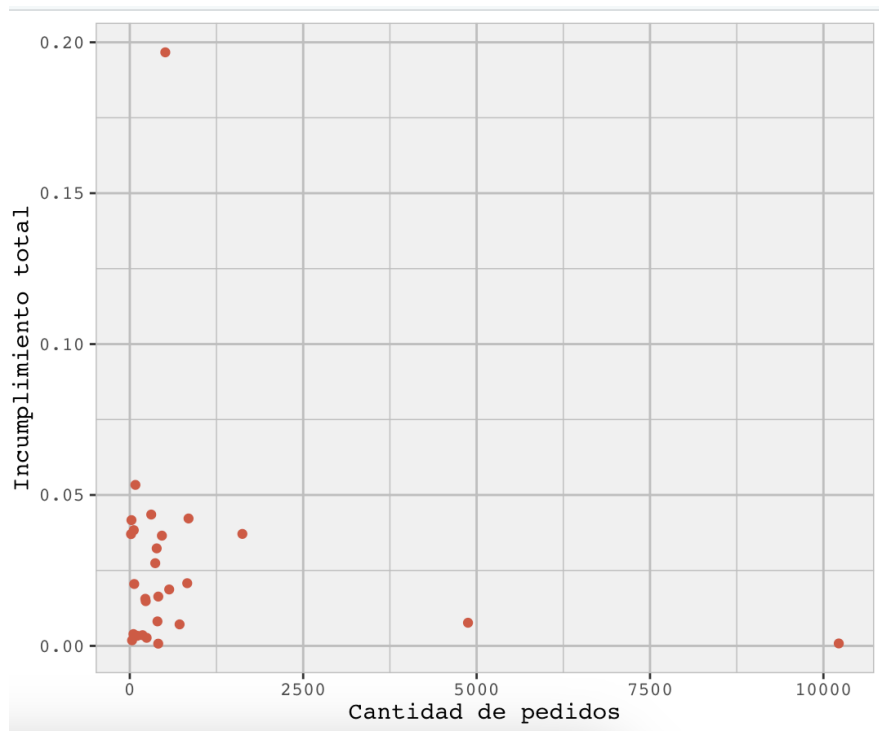


Figura III: Scatterplot de Incumplimiento total y Cantidad de pedidos

Interpretativamente, es posible delimitar que existe una gran cantidad de de datos que tiende a presentar una menor cantidad de datos y a su vez una proporción baja de incumplimiento.

De esta manera, llevamos a cabo un test de correlación de Pearson, a fines de evaluar su relación lineal. El mismo ha arrojado un p-valor mayor al nivel de significancia, lo que implica que no hay evidencia estadística suficiente para afirmar una correlación significativa entre las variables

Bibliografía

- Acceso a la Información Pública. Ministerio de Gobierno. SS de Asuntos Políticos. DG Seguimiento de Organismos de Control y Acceso a la Información. Obtenido en <https://data.buenosaires.gob.ar/dataset/acceso-informacion-publica>
- Acceso a la Información Pública. argentina.gob.ar. Obtenido de <https://www.argentina.gob.ar/justicia/derechofacil/leysimple/acceso-la-informacion-publica>
- Ley A-104 de acceso a la información Pública de la Ciudad Autónoma de Buenos Aires.