



Proyecto 1

Trabajo final

Segundo cuatrimestre de 2021

Docentes:

María Gabriela Bouret

Angélica Peralta Ramos

Integrantes:

Collado, Camila (61487)

Noguera, Abril (61541)

Pettinato, Camila (61050)

Introducción

En este trabajo se utilizó la base de datos sobre los llamados al 147 en relación al COVID-19. La línea 147 es la línea de atención al ciudadano. A través de este medio, se puede acceder telefónicamente a información o realizar quejas y/o denuncias.

Para verificar el funcionamiento línea 147, hicimos distintas pruebas de reclamos desde los 3 medios posibles:

1. Llamando a la línea
2. Desde app BA 147:

Utilizando la app nos dimos cuenta qué la única solicitud relacionada con el covid qué se puede reclamar es el incumplimiento de la cuarentena. Esto sería un reclamo posible a la Línea ya qué sí tuviéramos un reclamo de alguna otra temática no lo podríamos hacer por este medio. Decidimos no seguir con el reclamo de incumplimiento ya qué expondriamos a una persona a un reclamo falso.

Por esta razón, se solicitó limpieza de alumbrado público, el funcionamiento fue correcto y el reclamo se subió con éxito, nos dio la opción de subir fotos del lugar además de comentarios. Por último, enviaron un mail de confirmación y seguimiento del reclamo

The screenshots illustrate the steps to file a request:

- Screenshot 1: Nueva solicitud** - Shows a list of categories: Alumbrado, Arbolado, Calles y veredas, Control edificio, obras y catastro, Educación, Fiscalización actividades comerciales, Limpieza y recolección, Ordenamiento del espacio público, Parques y plazas, Pluviales, and Reciclado y protección ambiental. The "Alumbrado" option is selected.
- Screenshot 2: Alumbrado** - A text input field asks "Comentanos tu problema" (Optional). Below it is a "Comentario" text area. A green "SIGUIENTE" button is at the bottom.
- Screenshot 3: URIARTE 2462** - A map shows the location of the request. A modal window says "¡Bienvenido al mapa!" and "Para continuar cliqueá en alguno de los iconos". It includes icons for "Crear una solicitud" (with a location pin) and "Reforzar una solicitud" (with a flag). A yellow "ENTENDIDO" button is at the bottom right. At the bottom left, it says "Cargando foto..." and "Alumbrado" with the subtext "Limpieza de artefacto de alumbrado". A green "INGRESAR SOLICITUD" button is at the bottom right.

URIARTE 2462

En esta dirección no se registraron solicitudes

Cargando foto... Alumbrado Limpieza de artefacto de alumbrado

INGRESAR SOLICITUD

Alumbrado

Nro. de solicitud: 00514404/21

Uriarte 2462
Limpieza de artefacto de alumbrado

Fecha y hora de ingreso:
11/11/2021 - 10.14h

Comentarios:

Estado: Iniciada

Descripción de estado:
La solicitud fue ingresada con éxito. Estaremos tratando el pedido.

VOLVER AL INICIO



Hola, Camila.
Gracias por ayudarnos a mejorar la Ciudad con la solicitud que ingresaste de Limpieza de artefacto de alumbrado ([00514404/21](#)). La dirección donde vamos a estar trabajando es **URIARTE 2462**.

Queremos informarte que, con motivo de la crisis sanitaria que estamos atravesando, los plazos de resolución de solicitudes podrían verse afectados.

Nuestros esfuerzos están comprometidos en asegurar el normal funcionamiento del alumbrado público, priorizando tareas que nos permitan solucionar emergencias y situaciones de peligro evidente para terceros y/o que afecten la seguridad pública.

La Comuna 14 es quien hará el seguimiento de tu pedido.

La Ciudad cuenta con tecnología LED en sus calles, peatonales y plazas, y supera las 165.000 luminarias. De esta forma, Buenos Aires se convirtió en la primera capital latinoamericana 100% LED.

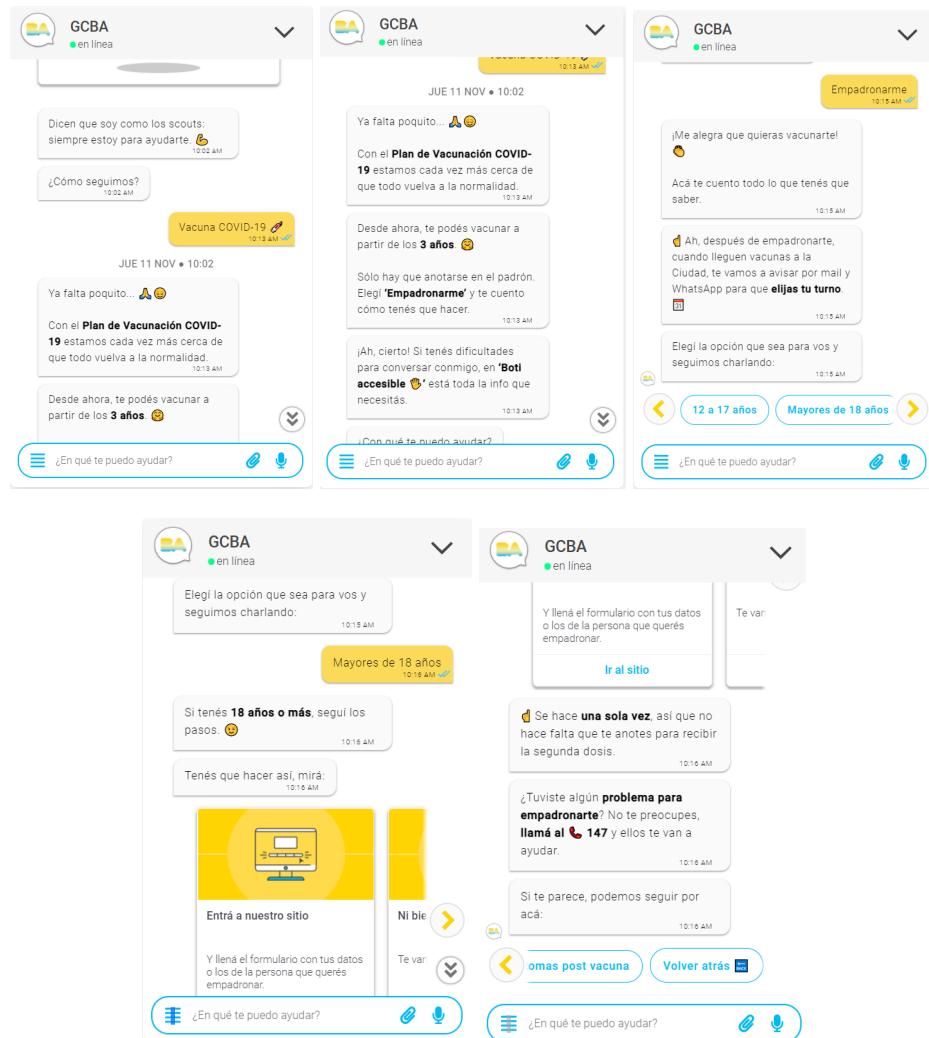
¡Muchas gracias!

Comuna 14

3. Desde la página Web:

Se le consultó por el chatbot acerca de las vacunas. El procedimiento es, Boti, te pregunta si te puede ayudar en algo y te aparecen un serie de opciones (Vacuna COVID-19, Test COVID - 19, Certificado COVID -19, y otros temas relacionados con la ciudad) donde se selecciona el de tu interés. Una vez que se seleccione, te va

dando información y el siguiente tema que deberías tocar para obtener más información.



Ficha Descriptiva

Nombre de base: Llamados 147 COVID-19

URL: <https://data.buenosaires.gob.ar/dataset/147-covid>

Fuente: BA DATA | Ministerio de Salud

Periodo de análisis:

- Inicio: 16 de marzo de 2020
- Fin: 5 de noviembre de 2021

Frecuencia de actualización: diariamente

Cantidad total de Filas: 4,302,834

Columnas:

1. Contacto

2. **Periodo**
3. **Categoría:** de que se trata la llamada
4. **Fecha_ingreso** (año-mes-dia)
5. **Comuna**
6. **Barrio**
7. **Domicilio_calle**
8. **Domicilio_altura**
9. **Lat**
10. **Long**
11. **Canal:** 147 (siempre)
12. **Estado_del_contacto:** completo o solucionado
13. **Fecha_cierre_contacto:** (año-mes-dia)
14. **DETALLE_RECLAMO:** descripción del reclamo

Problemas de la base

Para no perder la información original del dataset y poder limpiar, creamos columnas nuevas (fecha_ingreso_n, fecha_cierre_contacto_n, barrio_n, comuna_n, categoria_n) para realizar las modificaciones sobre esas columnas.

```
alter table dataset_covid
add column barrio_n varchar(1000)
```

1. Las **categorías** son inconsistentes: algunas comienzan con un número, luego el título, otras solo título. En el caso de las categorías qué engloban rangos etarios existen conflictos ya qué dos categorías engloban las mismas edades. Por ejemplo:

EMPADRONAMIENTO GRUPO DE RIESGO de 18 a 59 años (Vacuna Covid-19)
EMPADRONAMIENTO GRUPO DE RIESGO de 18 a 29 años (Vacuna Covid-19)
EMPADRONAMIENTO MENORES de 12 a 17 años (Vacuna Covid-19)
EMPADRONAMIENTO GRUPO DE RIESGO de 18 a 39 años (Vacuna Covid-19)

Para ver todos los tipos de categorías posibles usamos: select distinct

Al ver que había tantas categorías repetidas comenzamos a agrupar por distintos grupos y simplificas los nombres, la mayoría de las sentencias era con la condición LIKE '% %':

Como las categorías de empadronamiento se superponen entre si decidimos englobarlas todas en la categoría 'EMPADRONAMIENTO VACUNA COVID-19'

```

select distinct categoria_n
from dataset_covid dc
where categoria_n like '%EMPADR%' and categoria_n not like
'%DOMICILIO%'

update dataset_covid
set categoria_n = 'EMPADRONAMIENTO VACUNA COVID-19'
where categoria_n in (select distinct categoria_n
from dataset_covid dc
where categoria_n like '%EMPADR%' and categoria_n not like
'%DOMICILIO%')

```

Hay categorías que no tiene sentido, por ejemplo “AMBAS” o “ACTIVAMENTE”, estas fueron modificadas a NULL con el siguiente código:

```

update dataset_covid
set categoria_n = null
where categoria_n in ('',' ','ACTIVAMENTE','AMBAS')

```

Cómo seguían existiendo categorías genéricas sin sentido decidimos exportar el distinto de categorías y crear una nueva variable qué las agrupe un poco en excel. Las categorías qué veíamos qué tenían relación o representaban lo mismo las uniamos. Así pasamos de tener 137 categorías distintas a solo 39.

A su vez, identificamos qué se podría generalizar aún más las categorías, por eso decidimos hacer una nueva variable de categoría genérica en cinco categorías: incumplimiento, información, vacunas, síntomas y otros.

2. **Fechas:** al importar las fechas directamente cómo DATE ocurren problemas porque se cargan correctamente dejando solo fechas de días menores a 12. Las fechas se adaptan incorrectamente en la base. Por ende decidimos cargar las columnas fecha_ingreso y fecha_cierre_contacto cómo VARCHAR y luego cargar las nuevas variables normalizadas con tipo de dato TIMESTAMP con la siguiente sentencia.

```

update dataset_covid
set fecha_ingreso_n = to_timestamp(fecha_ingreso, 'DD/MM/YYYY
HH24:MI:SS')::TIMESTAMP

```

En algunos casos, la **Fecha_cierre_contacto** es menor a la **Fecha_ingreso**: analizando de cerca estos casos se vio que la diferencia era solo minutos por lo cual la solución que encontramos fue invertirlas, en Fecha_ingreso pusimos Fecha_cierre_contacto y viceversa.

Fecha_cierre_contacto_n	Fecha_ingreso_n
2021-03-05	2022-03-03
2021-10-04	2022-03-03
2021-10-04	2022-03-03
2021-08-04	2022-04-03
2020-02-07	2021-10-04
2020-02-07	2022-04-06
2021-10-04	2022-03-03
2020-02-05	2021-10-04
2021-03-05	2022-03-03
2021-03-05	2022-03-03

Por último, encontramos que la diferencia entre **Fecha_cierre_contacto_n** y **Fecha_ingreso_n** en algunos casos tomaba valores muy altos, como por ejemplo, 581 días. Para investigar porqué se daba esta situación, filtramos los llamados con duración mayor a un día. Descartamos que la duración de estos llamados estuviera relacionado con la categoría del llamada ya que se vio que estas son aleatorias. En una primera instancia filtramos las duraciones que fueran mayor a 1 dia y las ordenamos de mayor a menor por duración, este fue el resultado:

fecha_ingreso_n	fecha_cierre_contacto_n	categoria_n	duracion_dias
2020-03-20 12:01:07.000	2021-10-22 15:36:08.000	CONSULTAS POR SERVICIOS PUBLICOS	581
2020-04-18 11:03:54.000	2021-11-03 12:18:10.000	CONSULTAS POR SERVICIOS PUBLICOS	564
2020-03-30 16:30:59.000	2021-10-01 11:41:38.000	CONSULTAS POR SERVICIOS PUBLICOS	549
2020-03-24 12:27:27.000	2021-06-10 11:31:34.000	CONSULTAS POR SERVICIOS PUBLICOS	442
2020-04-14 13:23:02.000	2021-04-18 09:56:01.000	VACUNACION COVID-19	368
2020-05-08 14:49:39.000	2021-04-20 11:41:23.000	CONSULTAS POR SERVICIOS PUBLICOS	346
2020-04-23 17:47:17.000	2021-03-07 20:34:06.000	CONSULTAS POR SERVICIOS PUBLICOS	318
2020-03-27 15:53:55.000	2021-01-04 16:17:23.000	[NULL]	283
2020-03-27 11:03:08.000	2020-12-30 16:52:09.000	LOGISTICA	278
2020-03-25 11:10:18.000	2020-12-28 14:47:34.000	LLAMADO TELEFONICO DE CONTENCION	278
2020-03-27 14:36:32.000	2020-12-29 12:48:29.000	LOGISTICA	276
2020-03-27 13:12:12.000	2020-12-27 16:06:28.000	LOGISTICA	275
2020-03-28 13:57:50.000	2020-12-27 17:10:03.000	[NULL]	274

La cantidad de registros con esta anomalía fue de 46000 (0.009% del dataset). Luego afinamos la buscamos y chequeamos cuales tenían una duración mayor a 2 horas, elegimos este número arbitrariamente porque consideramos que en algunos días donde la línea estaba muy saturada, los tiempos de espera hasta que la persona fuera atendida por un operador podían hacer que esta duración fueran se extendiera. En este caso se agrega aproximadamente 6000 registros anormales más, representando el 0.01% del dataset, por lo tanto, al ser una cantidad tan pequeña, los filtramos en el *WHERE*, para que no se tuvieran en cuenta en el cálculo del promedio de duración de llamadas.

3. Barrios:

Se identificó qué no existe una lista de barrios prefija para qué el personal de atención de la línea seleccione el barrio y llenan el campo manualmente. Por lo qué existen 12905 barrios distintos.

Para comenzar, cargamos la nueva variable de barrios normalizada con la información de barrios existente cargándolos en mayúscula (UPPER) y eliminando los espacios delante y al final (BTRIM).

Cómo existen barrios cargados en la variable comuna decidimos cargar en los barrios qué quedaron vacíos los datos qué se encuentran en la comuna, menos los qué dicen comuna (LIKE ‘%COMUNA%’).

Hay entradas con números, códigos, símbolos (como “,”, “-”) y espacios. En esos casos eliminamos esas variables (SET barrio_n = NULL).

```
update dataset_covid
set barrio_n = null
where barrio_n in (select distinct barrio_n
from dataset_covid dc
where (barrio_n like '.%' or barrio_n like '-%' or barrio_n like
'0%' or barrio_n like ' %' or barrio_n like ',%')
and (barrio_n not like '%MON%' and barrio_n not like '%LUG%' and
barrio_n not like '%ATA%'))
```

Encontramos que había registros que empezaban con un símbolo pero seguían con el nombre del barrio. Para solucionar el problema, lo que hicimos fue eliminar ese primer dígito dejando el nombre del barrio.

```
update dataset_covid
set barrio_n = substr(barrio_n,2,length(barrio_n))
where barrio_n in(select distinct barrio_n
from dataset_covid dc
where (barrio_n like '.%' or barrio_n like '-%' or barrio_n like
'0%' or barrio_n like ' %' or barrio_n like ',%'))
```

Existían registros con dos espacios y decidimos reemplazarlos por un espacio.

```
update dataset_covid
set barrio_n = replace(barrio_n, '  ', ' ')
```

En muchas ocasiones se ingresaron cómo barrio siglas de tres dígitos qué analizandolas entendimos que eran una parte del código postal. Al ser tan complejo identificar qué barrio pertenece a esa sigla decidimos eliminarlas (SET barrio_n = NULL) en los casos qué la longitud del barrio sea de 3 dígitos.

Hay casos con falta de ortografía, nos dimos cuenta de este error haciendo una consulta donde encontramos que para algunos barrios había muchas maneras de escribirlo.

```
select distinct barrio_n
from dataset_covid
where (barrio_n like '%EZ' or barrio_n like '%ES') and (barrio_n
like 'N%' or barrio_n like 'ñ%')
```

	barrio_n
1	NEÑEZ
2	NIÑEZ
3	NUEZ
4	NUEÑEZ
5	NUEÑEZ
6	NUNEZ
7	NUNÑEZ
8	NUÑEZ
9	NUÑEZ
10	NURIEZ
11	NÚÑEZ
12	NÚÑEZ
13	NÚÑEZ
14	NúñEZ
15	ÑUNEZ
16	ÑUÑEZ

Para solucionar el problema, lo que hicimos fue modificar el dataset al nombre correcto del barrio cuando este campo estaba mal escrito.

```
update dataset_covid
set barrio_n = 'NUÑEZ'
where barrio_n in (select distinct barrio_n
from dataset_covid
where (barrio_n like '%EZ' or barrio_n like '%ES') and (barrio_n
like 'N%' or barrio_n like 'ñ%'))
```

Lo mismo identificamos con Caballito y Constitución, normalizandolas de la misma manera.

Al igual qué con estos barrios hay muchísimos casos distintos. Por eso, decidimos exportar el distinct de barrio_n y procesarlos por fuera de Postgres. Primero, utilizando Open Refine y luego manualmente en Excel.

Así, logramos pasar de 12905 barrios distintos a solo 804.

También, al procesarlos manualmente con Excel logramos investigar más sobre cada barrio y decidimos crear una nueva variable provincia qué agrupa aún más los barrios. Así se encontraron datos de las distintas provincias Argentinas y algunos casos del exterior del país.

4. Comunas:

Hay barrios, provincias, calles y comunas con tres denominaciones distintas (solo número de la comuna, la palabra “comuna” seguida del número, y la palabra “CABA” seguida de “comuna” seguida del número).

Al encontrar tantos errores en estos registros decidimos arrancar de cero y definir todos los registros de la nueva variable comuna_n como ‘Sin Identificar’.

Luego se utilizó el dataset poblacion_por_barrio.csv ([disponible en BA Data](#)) y actualizamos la comuna haciendo un join y actualizando según el.

```
update dataset_covid
```

```

set comuna_n = replace(comuna_n, comuna_n,c.comuna)
from comunas_y_barrios_csv c
where barrio_n = c.barrio and barrio_n != ''
and barrio_n is not null
    
```

5. Latitud y Longitud:

Se entiende qué todas las latitudes y longitudes de la Argentina, según la ubicación en el hemisferio, deben ser negativas. Por eso buscamos si existían positivas y determinamos qué existían en la base. Pero, cómo identificamos previamente qué existían llamadas del exterior por eso tuvimos el cuidado de eliminar los registros con latitud y longitud positiva solo si la provincia era distinta de ‘EXTERIOR’.

	lat	long	barrio_n
1	-82.862752	135	MUNRRO
2	-33.9248685	18.4240553	FLORENCIO VARELA
3	-22.7359095	140.0187653	VILLA 11/14
4	4.6830603	-74.10858	VILLA LUZ
5	5.011706	-74.470362	RIVERA VILLETA
6	5.454511	-73.362003	RIO NEGRO
7	5.8520355	-55.2038278	CABA
8	10.9548837	78.443932	MATADEROS
9	12.576855	-81.05052	VILLA VATTEONE
10	13.3323464	-87.8500644	B'LA UNION(LAFERRI)
11	13.6929403	-89.2181911	LA GLORIA
12	15.5006544	32.5598994	BUENOS AIRES
13	17.4391926	-99.5450974	FLORES
14	18.3448477	-99.5397344	BUENOS AIRES
15	18.4860575	-69.9312117	BUENOS AIRES
16	18.4860575	-69.9312117	LOMAS ALEGRE
17	18.4860575	-69.9312117	PARQUE PATRICIOS
18	19.173773	-96.1342241	VILLA SOLDATI
19	19.2895865	-99.5117391	MAR DEL PLATA
20	20.0910963	-98.7623874	CABA
21	20.0910963	-98.7623874	SAN ANTONIO DE PA
22	21.1250077	-101.6859605	VILLA DEL PARQUE
23	22.8905327	-109.9167371	CIUDAD EVITA
24	23.1135925	-82.3665956	MERLO
25	25.4546718	-100.850791	PILAR
26	25.5376982	83.1959914	CABA

También, encontramos casos donde se llenaron la latitud y longitud con 0 y las eliminamos (SET lat / long = NULL).

	lat	long	barrio_n
1	0	0	PELERO
2	0	0	PLAYON
3	0	0	LIBERTAD
4	0	0	TRAS LAS SIEF
5	0	0	VILLA PUERRE
6	0	0	9 DE JULIO
7	0	0	BALBANEGRAS
8	0	0	LINCOL
9	0	0	SABIO
10	0	0	URLINGAM
11	0	0	BLEGRANO
12	0	0	BELGARNO
13	0	0	PARQUE CHAC
14	0	0	CHACARITAS
15	0	0	CABALLITO
16	0	0	NESTOR KIRCI
17	0	0	PUERTO MADI
18	0	0	MONRREDA
19	0	0	ALMGRO
20	0	0	VELGRANO
21	0	0	LUGANO 1 Y 2
22	0	0	-
23	0	0	LANUS ESTE

Script

Para ver el detalle de la normalización de cada columna visitar el script completo: [Script](#)

Storytelling

El enfoque del análisis es entender la pandemia desde la preocupación ciudadana cotidiana, ir más allá de los datos concretos de casos, muertes, vacunas y analizar el temperamento de la sociedad. Una vía directa para estudiar lo qué inquieta a la población es la [línea 147](#) del Centro de Atención al Vecino. El Gobierno de la Ciudad de Buenos Aires, puso a disposición en la web de Buenos Aires Data, el dataset correspondiente a los llamados con temática de COVID-19 a esta línea.

Analizando el funcionamiento general de la línea 147 encontramos que la cantidad de llamados recibidos, hasta el 5 de noviembre del 2021, fueron 4.302.834.

Los motivos de los llamados abarcan varios temas, desde asistencia y voluntariados al comienzo de la pandemia hasta denuncia de reuniones clandestinas y empadronamiento de vacunación para los últimos meses. Por eso, centralizamos el trabajo en las principales 5 categorías: incumplimiento, síntomas, vacunas, información y otros.

En promedio la línea 147 recibe un llamado cada 11 segundos. La duración promedio de resolución de consulta es de 1 minuto y 52 segundos.

Nos interesó entender, al haber sido dos años de pandemia totalmente distintos, cuáles eran las preocupaciones de la gente en cada uno de ellos. Se puede ver qué hubo claras distinciones entre los motivos de los llamados de cada año.

Para comenzar, en el 2020, la mayor preocupación de las personas fue información general sobre servicios públicos, permisos de circulación y protocolos. Se ve un pico inicial que muestra la incertidumbre presente en la sociedad a los comienzos de la pandemia así como las inquietudes generaba la cuarentena.

Por su parte, en 2021, el tema central por gran diferencia es vacunas, la cantidad de llamados sobre ese tema sobrepasó cualquier otro, por este medio no solo se realizan consultas sobre plan de vacunación si no que también se hacen empadronamientos para distintos grupos etarios.

2020

Temas de interés de la prensa: [Extensión de la cuarentena](#) y [Salidas por número de DNI](#).

2021

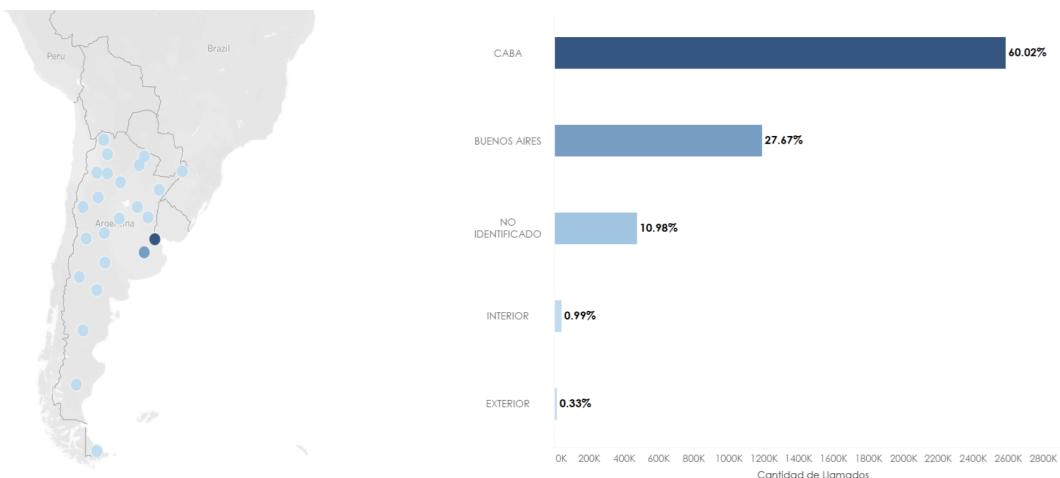
Temas de interés de la prensa: [Argentina transita la segunda ola de coronavirus con un sistema sanitario al límite y médicos exhaustos](#) y [¿Cuáles vacunas estamos aplicando en el país?](#)

categoria_n	cantidad
CONSULTAS POR SERVICIOS PUBLICOS	416,166
PERMISO DE CIRCULACION	249,747
VACUNACION COVID-19	187,327
VOLUNTARIO	129,773
INFORMACION	90,964
INCUMPLIMIENTO DE CUARENTENA	56,868
AVISO DE CIRCULACION MAYORES DE 70	45,041
PLAN DETECTAR	40,969
SINTOMAS	39,052
LOGISTICA	25,263

categoria_n	cantidad
VACUNACION COVID-19	1,823,547
EMPADRONAMIENTO VACUNA COVID-19	430,194
SINTOMAS	144,374
POSIBLES INFECTADOS COVID-19	106,215
PLAN DETECTAR	83,326
INFORMACION	74,274
EMPADRONAMIENTO VACUNA COVID-19 A DOMICILIO	65,728
PERMISO DE CIRCULACION	55,837
CASOS POSITIVOS COVID-19	30,772
VOLUNTARIO	14,217

- Hubo 2,5 veces más llamados por vacunas qué de información.
- Los motivos de los llamados siguieron la línea de temas de los medios de comunicación.

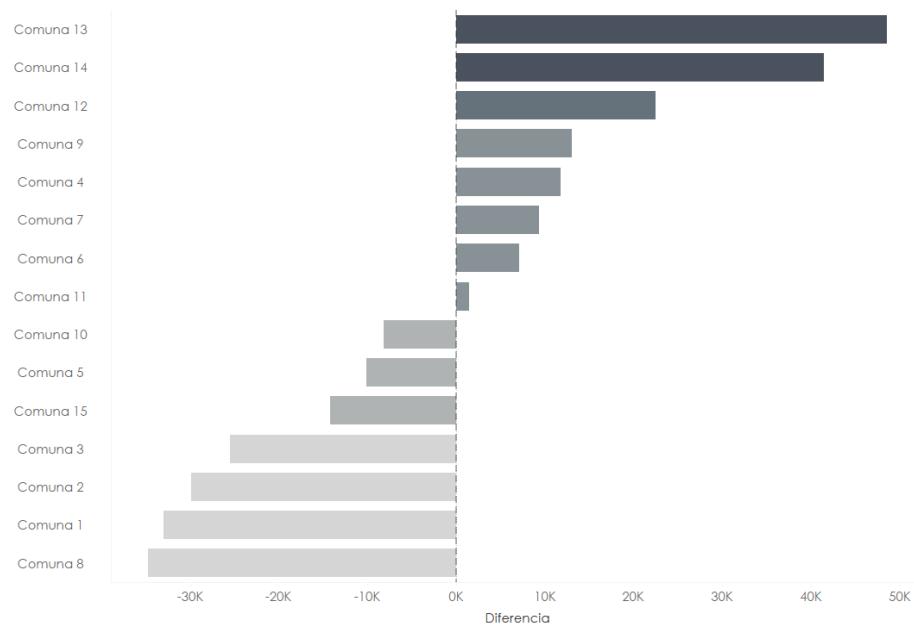
Haciendo un análisis geográfico de los llamados, estudiamos la distribución de los llamados en las distintas provincias. El 60,02% de los llamados provienen de CABA, una de las causas de esto es que, quienes habitan ya estaban acostumbrados al uso de esta línea para reclamos de otros tipos, en segundo lugar provincia de Buenos Aires con 27,67%. También se puede ver como los llamados del resto de las provincias del país (interior) son el 0,99% y los del exterior 0,33%



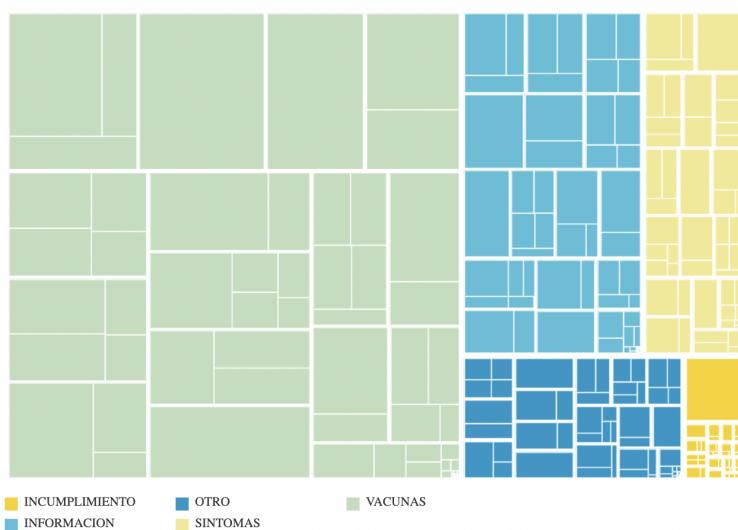
Para ver más objetivamente la cantidad de llamados por zonas analizamos 2 perspectivas: los llamados por comunas y los llamados por barrios cada cien mil habitantes.

Es interesante la información que se observa cuando graficamos la cantidad de llamados de cada comuna en comparación con el promedio de llamados de CABA (cantidad de llamados dividido cantidad de comunas). Podemos ver que las comunas 13 (Belgrano, Nuñez, Colegiales) y 14 (Palermo) fueron las que más utilizaron esta línea. Por otro lado, por otro lado, las comunas que menos llamaron fueron la comuna 8 (Villa Lugano, Villa Soldati, Villa

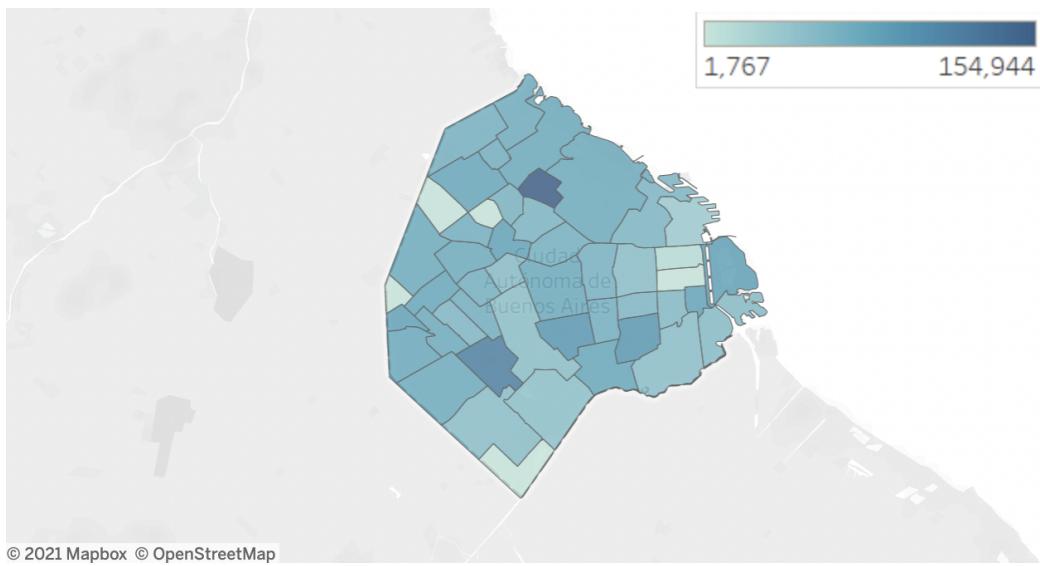
Riachuela) y la comuna 1 (San Nicolas, San Telmo, Monserrat, Constitución, Puerto Madero). La categoría principal en todas las comunas fue vacunas.



Categorías por Comuna

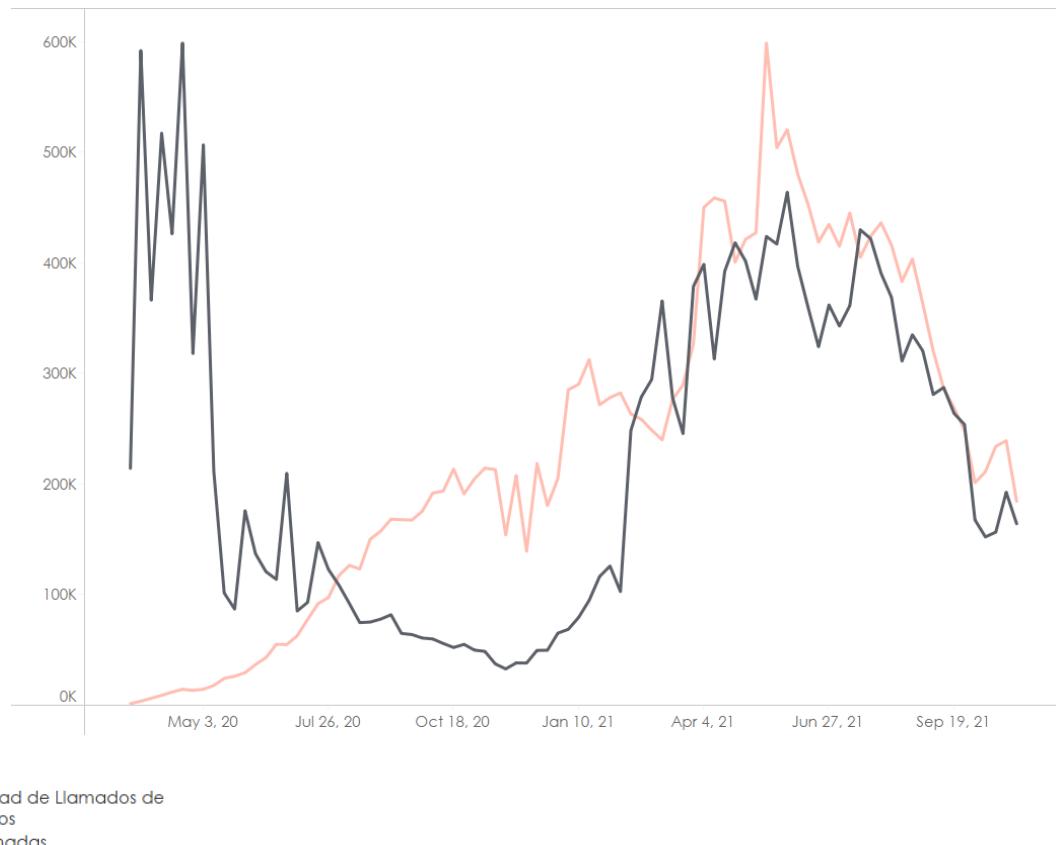


En segundo lugar, calculamos la cantidad de llamadas por barrio cada cienmil habitantes para poder ver objetivamente que barrio llamó más y evitar que los barrios con más población resalten en el mapa. Así se destacó el barrio Colegiales y Parque Avellaneda.



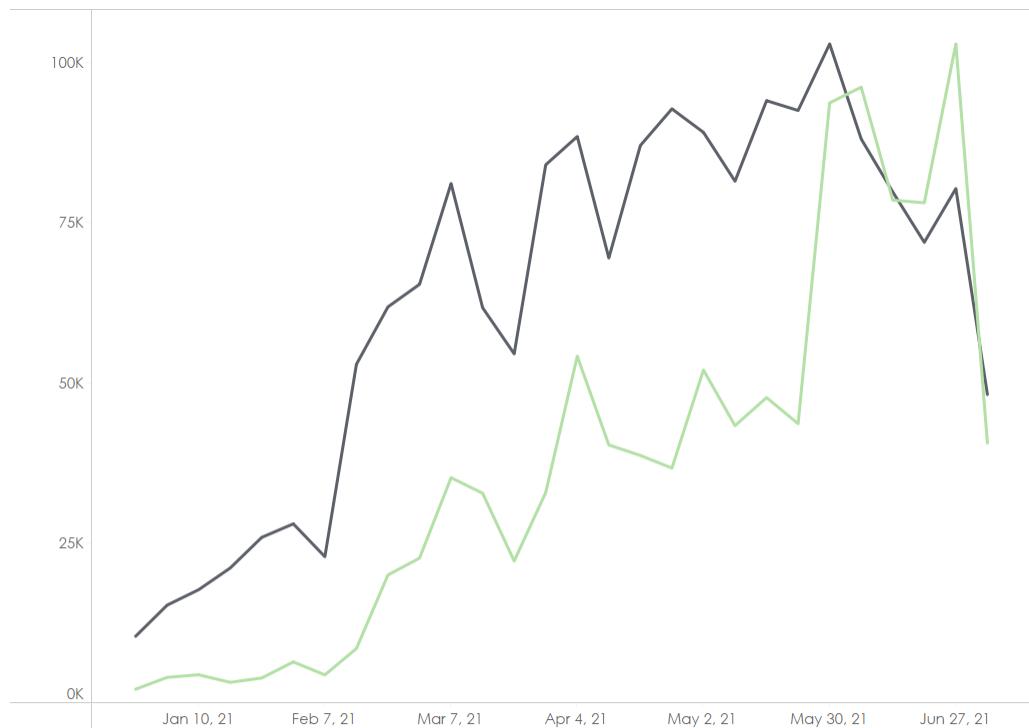
Comparando la cantidad de llamados por día con casos de covid por día, datos que se obtuvieron desde la [API del ITBA](#) (desde 16/03/2020 hasta 05/11/2021), se puede ver que los llamados eran muy frecuentes al comienzo de la pandemia por el miedo y la incertidumbre de la población que se volcaba a estos canales, luego los llamados bajaron aunque los casos seguían subiendo y finalmente se equipararon.

A partir de los datos obtenidos de la consulta se realizó el siguiente gráfico:



La combinación de casos y llamadas por día muestra tres etapas a lo largo del tiempo. En los primeros meses de pandemia hubo muy pocos casos pero muchos llamados porque no se sabía mucho sobre el virus, había muchas dudas y miedo en la sociedad. Luego, los casos crecieron y los llamados disminuyeron debido a que la información comenzó a darse por otros canales (principales medios de comunicación, páginas web gubernamentales) y la sociedad le perdió el miedo al virus. En la etapa final, a partir de la segunda ola, se sincronizó la cantidad de casos con la cantidad de llamadas hasta el día de hoy.

Repetimos el mismo proceso de casos para la cantidad de vacunas aplicadas por día y podemos ver que la cantidad de llamadas tuvo la misma tendencia de crecimiento, aunque siempre con mayor cantidad que las vacunas. Esto cambió a principios de junio de 2021 donde la cantidad de vacunas aplicadas aumentó un 121%. Investigando lo que sucedía en ese momento, encontramos que durante estos días se dieron las primeras grandes llegadas de vacunas al país ([Frío, vacunas y contagios: por qué junio puede ser el mes bisagra en la lucha contra el coronavirus](#) y [Cómo se distribuirán entre las provincias las vacunas de AstraZeneca que llegaron este lunes a la Argentina](#)) y se especulaba cómo se iban a distribuir entre las provincias, cuántos millones llegaban de cada vacuna, quién tenía prioridad para vacunarse y si iban a cambiar las restricciones una vez aplicadas las dosis. Consecuentemente, estos sucesos aliviaron a la población ya que comenzaron a recibir más turnos y aplicarse más vacunas, resultando en una disminución en llamados ya que la situación se normalizaba



Cantidad de llamados de
Vacunas
Llamadas

Conclusiones Generales

- CABA usó mucho más línea que cualquier otra provincia, la cantidad de llamados está muy segmentada en CABA y provincia de Buenos Aires contra el interior del país.
- La categoría central en todas las geografías (provincias, barrios y comunas) fue vacunas
- si hablamos de barrios, colegiales y parque avellaneda fueron quienes tuvieron más llamados por persona, lo cual sugiere que se debería aumentar la comunicación con estas 2 zonas para ver las problemáticas
- si hablamos de comunas las más activas fueron comuna 13 y 14
- El mayor pico de llamadas se dio cuando se habilitó la línea, al principio de la pandemia, cuando había pocos casos.

Potencial futuro del dataset

En nuestra opinión, la principal falencia con esta base es qué las variables geográficas y las categorías no están estandarizadas y pueden ser ingresadas por el usuario a su gusto. Esto genera una muy amplia cantidad de valores distintos qué representan lo mismo.

Esta base podría explotarse mucho más si el análisis geográfico pudiera ser más efectivo sin tener que dedicarle tanto tiempo al trabajo de normalización de barrios y comunas que si bien se hizo, para CABA, pero no fue suficiente para resolver el problema. Esto sucede porque la estructura del dataset estaba preparado para recibir llamados de CABA y durante el periodo de COVID-19 recibió llamados de todo el país e incluso internacionales, lo que probablemente ocasionó que los operadores no pudieran llenar correctamente las columnas “Comuna” y “Barrio”. Además la latitud y longitud de la mayoría de los registros es erróneo y no puede identificar ningún lugar puntual en el mapa.

Sería de mucha ayuda, y mejoraría ampliamente el análisis, además de limitar el ingreso de datos a una lista preexistente, agregar las columnas País, Provincia y Localidad con opciones para completar y no escritos manualmente.

Lo mismo sucede para las categorías sería mucho más interesante el análisis sí no fuera el operador quien decide la categoría sin criterio y bajo la posibilidad de repetir categorías con distintos nombres.

A partir de la corrección de estos problemas se podría crear un análisis automático, cómo el dashboard qué propusimos, para qué el gobierno pueda seguir los reclamos de la gente partiendo de los siguientes aspectos:

- Los temas más preguntados para saber donde apuntar las campañas informativas del gobierno o para los medios de comunicación a dónde apuntar sus notas.
- Usando la frecuencia de las llamadas para ver si se necesitan más operarios para la línea.
- Usando la duración de la llamada para identificar si existen problemas a la hora de comunicarse en el caso de que sea muy baja la duración, como lo es ahora, o en el caso de que sea muy extensa encontrar formas para que la comunicación sea mas dinámica.
- Conocer en profundidad las inquietudes del país entero y no solo las de CABA si se incentiva el uso de la línea en las provincias.