

# Usos y Usuarios de Tarjeta SUBE

Período 2020/2021

Integrantes:

Christian Persico,  
Claudio Martin Belitzky,  
Cynthia Auad



**2da Pre Entrega Proyecto final**  
**Comisión 24415**

***CODER HOUSE***

## Contenido

Introducción: .....	3
Objetivo del Proyecto: .....	3
Alcance del Proyecto: .....	3
Usuario Final y Nivel de Aplicación: .....	3
Descripción de la base de datos .....	4
Diagrama Entidad-Relación (DER), con definición de Primary Key (PK) y Foreign Key (FK): .....	4
DER: Columnas por tablas y definiciones de tipos de datos y claves .....	5
Transformaciones de Datos Realizadas: .....	6
Medidas calculadas y sus fórmulas .....	6
Segmentaciones .....	10
Análisis funcional del tablero .....	11
Solapa PAIS .....	11
Solapa AMBA Subte Tren .....	12
Solapa AMBA Colectivo .....	12
Solapa Usuarios .....	12
Futuras líneas .....	13
Versionado .....	13
Anexo .....	14

**Introducción:**

Para este trabajo final se seleccionó como temática el análisis de los usos y usuarios de la tarjeta SUBE.

Las bases de datos fueron obtenidas a través del portal online de datos públicos de Nación.

Se busca analizar la cantidad de pasajeros que utilizaron la tarjeta SUBE para realizar viajes en los diferentes medios de transporte públicos a nivel nacional, durante los años 2020 y 2021.

**Objetivo del Proyecto:**

Observar cómo fue afectado el uso de la tarjeta con el advenimiento del período de cuarentena estricta por pandemia, y luego con el levantamiento gradual de las restricciones de circulación.

**Alcance del Proyecto:**

Se quiere analizar:

Cantidad de transacciones por fecha, tipo de transporte y jurisdicción.

Cantidad de tarjetas en uso por año.

Concentración geográfica del uso de la tarjeta SUBE a nivel país.

Tipos de usuarios según género y programa de beneficio estatal (Atributo Social Federal Resolución 384/2021). Cantidad de tarjetas con y sin beneficio estatal.

Se hará foco en la región AMBA, mostrando el comportamiento del uso las líneas de transporte en el período bajo análisis.

**Usuario Final y Nivel de Aplicación:**

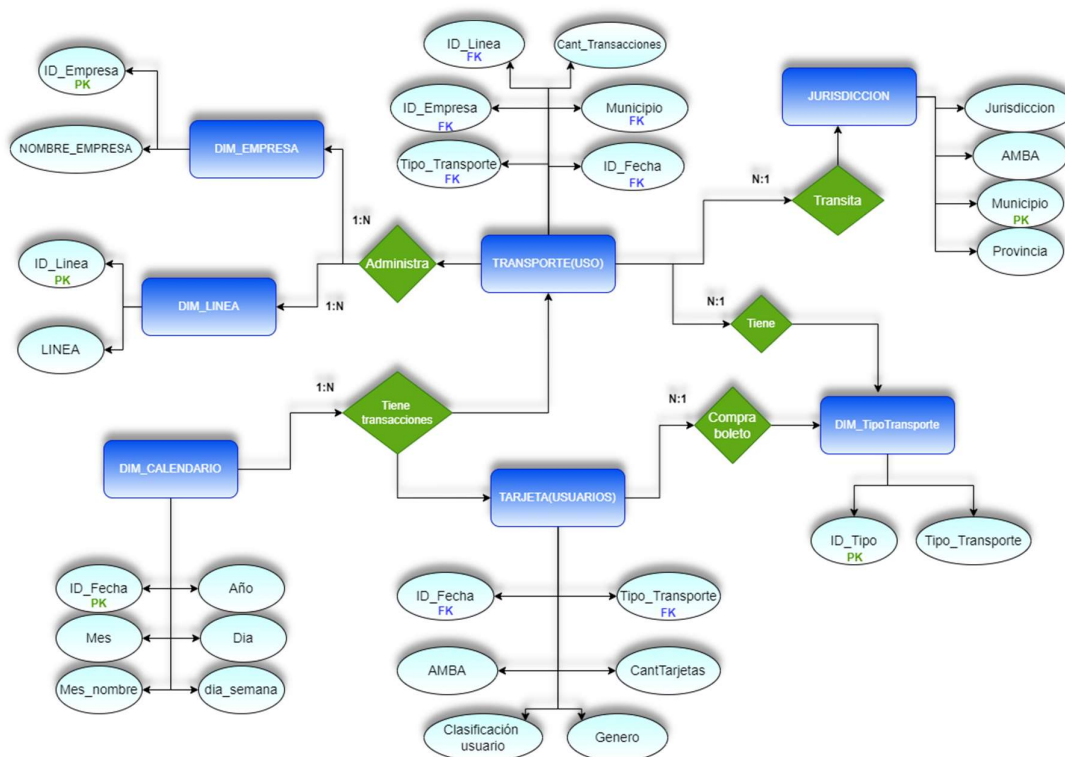
Como consultores fuimos contratados por el gobierno de la nación para proveerle un dashboard sobre los usos de la tarjeta SUBE a nivel país para los años 2020 y 2021 con la finalidad que toda la población pueda tener acceso dentro de las bases abiertas que se publican a esta fuente de información.

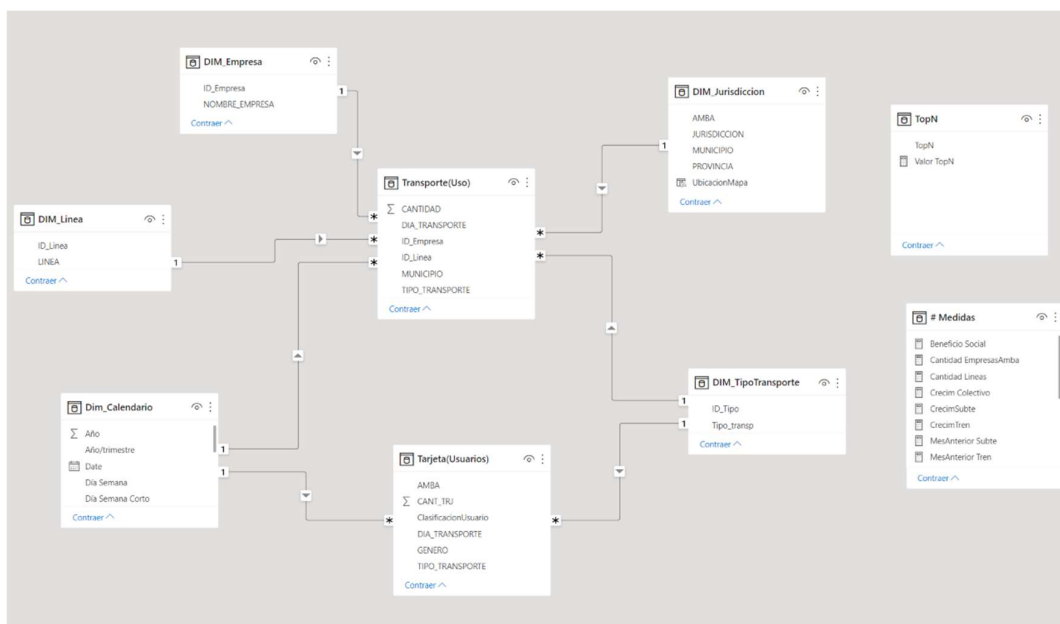
## Descripción de la base de datos

A través del portal de datos públicos del Gobierno Nacional, se descargaron dos bases de datos, correspondiente a los años 2020 y 2021, que contienen información sobre usos de la tarjeta SUBE según el territorio y líneas de transporte, y usuarios de la tarjeta según su clasificación, género y líneas.

Esta base se transformó para crear distintas tablas y organizar la información. A continuación, se muestran los diagramas entidad-relación resultantes.

## Diagrama Entidad-Relación (DER), con definición de Primary Key (PK) y Foreign Key (FP):

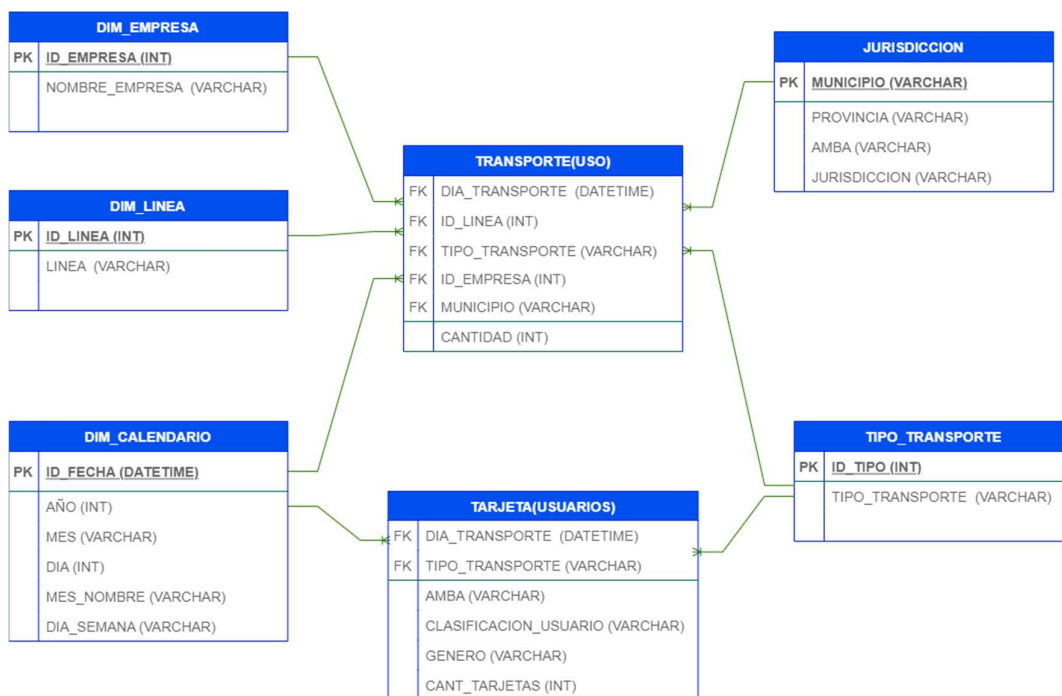




DER obtenido de la carga del dataset en Power BI.

## DER: Columnas por tablas y definiciones de tipos de datos y claves

Se describen los atributos de cada tabla y se definen los datos claves para generar las relaciones.



## **Transformaciones de Datos Realizadas:**

Luego de la carga del dataset, se revisaron la definición de los tipos de datos de cada columna y se realizaron las siguientes transformaciones:

- Se realizó una combinación de las tablas Transporte\_usos y Linea\_Empresa, creando una nueva tabla llamada “Transportes Usos Comb”. A todas estas tablas se les deshabilitó la carga.
- Una vez creada, se expandió la tabla anexada, mostrando el nombre de la empresa. Se reordenaron las columnas y se renombró la columna Nombre\_Empresa.
- A partir de la tabla Linea\_Empresa se crearon dos copias referenciadas, para generar las nuevas dimensiones DIM\_Linea y DIM\_Empresa.
- A la nueva tabla DIM\_Linea se le eliminó la columna de nombre empresa, se quitaron los duplicados y se agregó una nueva columna con índice. Finalmente se reordenaron las columnas.
- A la nueva tabla DIM\_Empresa se le eliminó la columna de nombre de la línea, se quitaron los duplicados y se agregó una nueva columna con índice. Se reordenaron las columnas.
- Se realizó una nueva combinación de las tablas “Transportes Usos Comb” y DIM\_Linea, creando una nueva tabla llamada Transporte(Uso). Se expandió para agregar el índice y se eliminó la columna de Línea. Se reordena y renombra la columna índice como ID\_Linea.
- Se realiza una última combinación de la tabla Transporte(Uso) con DIM\_Empresa. Se expandió para agregar el índice y se eliminó la columna de Nombre Empresa. Se reordena y renombra la columna índice como ID\_Empresa.
- Creamos la tabla Calendario, para esto se crea una nueva tabla en la que se inserta el código para la tabla calendario conteniendo fechas entre enero 2020 y diciembre 2021.
- La tabla Tarjeta(Usuarios) incluía líneas de totales parciales en la columna Tipo\_Transporte, que no nos servían. En dicha columna, se reemplazó el valor “TOTAL” por “” (vacío) y luego se quitaron las líneas vacías.
- Se renombran algunas provincias en DIM\_Jurisdiccion, para corregir errores de tipo tilde o caracteres especiales.
- Se renombran líneas en la columna ClasificacionUsuario de la tabla Tarjeta(Usuarios). Las filas en las que aparecía “null” se renombra como: “Sin Beneficio Social”, y en la columna GÉNERO se renombra “null” por NR (No Registrado).

## **Medidas calculadas y sus fórmulas**

En la tabla DIM Jurisdicción se inserta una columna calculada, para concatenar el nombre del País y Provincia, con el fin de mejorar la visualización de mapas.

En la tabla DIM\_Calendario se inserta una columna calculada para definir las etapas de pandemia según la fecha, llamada “Restricción”. De acuerdo a la fecha se le asigna el periodo de aislamiento correspondiente.

En la tabla DIM\_TipoTransporte se crea una medida calculada para definir un color por tipo de línea, que luego será utilizado en una tarjeta de varias filas.

Se crea una nueva tabla para agrupar las medidas calculadas:

1) Medida BENEFICIO SOCIAL

Beneficio Social =

```
CALCULATE ( [Total TARJETAS],  
            FILTER('Tarjeta(Usuarios)',  
                  IF('Tarjeta(Usuarios)'[ClasificacionUsuario]= "Sin Beneficio Social", BLANK(),  
                    SUM('Tarjeta(Usuarios)'[CANT_TRJ])) ))
```

La finalidad de la medida es mostrar la cantidad total de tarjetas que tienen algún tipo de beneficio social.

2) Medida CANTIDAD DE EMPRESAS

Cantidad EmpresasAmba =

```
CALCULATE (DISTINCTCOUNT ('Transporte(Uso)'[ID_Empresa]),  
            DIM_Jurisdiccion[AMBA] = "Si", DIM_TipoTransporte[Tipo_transp] = "Colectivo")
```

La finalidad de la medida es mostrar la cantidad de empresas de colectivo existentes en AMBA.

3) Medida CANTIDAD DE LINEAS

Cantidad Lineas = `DISTINCTCOUNT ('Transporte(Uso)'[ID_Lineas])`

La finalidad de la medida es mostrar la cantidad de líneas de colectivo que se encuentran funcionando en AMBA.

4) Medidas de CRECIMIENTO (COLECTIVO, SUBTE y TREN)

Crecim Colectivo =

```
VAR calculo = CALCULATE ([Uso ColectivoAMBA],  
                           DATEADD(Dim_Calendario[Date], -1, MONTH))  
  
RETURN IF ( [Uso ColectivoAMBA] = 0, 0,  
           DIVIDE ([Uso ColectivoAMBA], calculo, 1)-1)
```

CrecimSubte =

```
VAR calculo = CALCULATE ([Uso Subte],  
                           DATEADD(Dim_Calendario[Date], -1, MONTH) )  
  
RETURN IF ( [Uso Subte] = 0, 0,  
           (DIVIDE ([Uso Subte], calculo, 1)-1))
```

CrecimTren =

```
VAR calculo = CALCULATE([Uso Tren],  
                           DATEADD(Dim_Calendario[Date], -1, MONTH) )  
  
RETURN IF ( [Uso Tren] = 0, 0,  
           (DIVIDE ([Uso Tren], calculo, 1)-1))
```

La finalidad de la medida es mostrar la variación porcentual en uso del subte, colectivo y tren, respectivamente, entre el 2021 vs 2020.

5) Medidas para Cálculo de mes Anterior

MesAnterior Subte =

```
CALCULATE ([Uso Subte], DATEADD(Dim_Calendario[Date], -1, MONTH))
```

MesAnterior Tren =

```
CALCULATE ([Uso Tren], DATEADD(Dim_Calendario[Date], -1, MONTH))
```



MesAnt Colectivo =

```
CALCULATE ([Uso ColectivoAMBA], DATEADD(Dim_Calendario[Date], -1, MONTH))
```

La finalidad de estas medidas es reflejar el total de transacciones, tanto de colectivos, trenes y subtes de AMBA, pero que sucedieron el mes anterior a un mes determinado.

6) Medida TOPN COLECTIVO

TOPN Colectivo =

```
VAR ValorTop = SELECTEDVALUE('TopN'[TopN])

RETURN

SWITCH (TRUE(), ValorTop = 0, [Uso ColectivoAMBA],

RANKX(ALLSELECTED(DIM_Lineas), [Uso ColectivoAMBA]) <= ValorTop, [Uso
ColectivoAMBA])
```

La finalidad de la medida es mostrar la cantidad de líneas de colectivo según el valor seleccionado de forma dinámica a través del uso de un parámetro. Si parámetro N es 0 muestra la lista completa, si es mayor muestra N cantidad de líneas de colectivos, según su ranking de usos.

7) Medida TARJETAS

Total TARJETAS = `sum ('Tarjeta(Usuarios)'[CANT_TRJ])`

La finalidad de la medida es mostrar la cantidad de tarjetas totales.

8) Medida USO Pais

TotalUSO Pais = `sum('Transporte(Usos)'[CANTIDAD])`

La finalidad de la medida es mostrar la cantidad de Usos/transacciones totales.

9) Medida Uso AMBA

```
Uso AMBA = CALCULATE(sum("Transporte(Usos)"[CANTIDAD]), DIM_Jurisdiccion[AMBA]= "SI")
```

Esta medida suma la cantidad de uso de transporte solo en la región AMBA.

10) Medida USO (COLECTIVO, SUBTE y TREN)

Uso ColectivoAMBA =

```
CALCULATE(sum("Transporte(Usos)"[CANTIDAD]),  
           DIM_Jurisdiccion[AMBA]= "SI", DIM_TipoTransporte[Tipo_transp]= "Colectivo")
```

Uso Subte =

```
CALCULATE(sum("Transporte(Usos)"[CANTIDAD]),  
           DIM_Jurisdiccion[AMBA]= "SI", DIM_TipoTransporte[Tipo_transp]= "Subte")
```

Uso Tren =

```
CALCULATE(sum("Transporte(Usos)"[CANTIDAD]),  
           DIM_Jurisdiccion[AMBA]= "SI", DIM_TipoTransporte[Tipo_transp]= "Tren")
```

La finalidad de esta medida es calcular la cantidad de transacciones que tuvieron lugar en los colectivos, subtes o trenes respectivamente, únicamente de la región AMBA.

## Segmentaciones

Los segmentadores utilizados son filtros por año y períodos de aislamientos decretados por el Gobierno Nacional en relación a la situación pandémica a nivel mundial. Estos períodos están dados por fechas específicas descritas a continuación:

- ASPO (Aislamiento Social Preventivo y Obligatorio): desde el 20/03/2020 al 07/06/2020.
- DiSPO (Distanciamiento Social Preventivo y Obligatorio): desde el 08/06/2020 al 09/04/2021.
- Sin Restricción: a todas las fechas restantes del período 2020-2021

Además, se utilizaron filtros de ubicaciones, como ser Provincias, y clasificación de usuarios.

Se agregó un segmentador especial, para analizar el caso de las líneas de colectivos en la región AMBA, debido a la cantidad de líneas existentes. Se utilizó un parámetro para generar de manera dinámica un Top N de muestreo de líneas.

## **Análisis funcional del tablero**

Para el desarrollo del dashboard se comenzó por dividir el tablero en 5 solapas, las cuales incluyen botones de navegación entre ellas:

- Portada o presentación, en la cual se describe brevemente el propósito del análisis.
- “PAIS”, se realiza el análisis a nivel más general, teniendo en cuenta todas las jurisdicciones y municipios en las que se utiliza el Sistema Único de Boleto Electrónico.
- “AMBA SubteTren”, se hace foco en esta región que concentra el Uso ambos tipos de transportes.
- “AMBA Colectivo”, se hace foco en esta región que concentra el mayor Uso del sistema SUBE.
- “USUARIOS”, se analiza los tipos de usuarios de la tarjeta.

### **Solapa PAIS**

Se comienza el análisis mostrando tarjetas con los datos totales de Transacciones(usos) de la tarjeta, cantidad de tarjetas y de líneas.

En una tarjeta de varias filas se muestra la cantidad total de transacciones discriminada por tipo de transporte. Se integra esta información en un gráfico de líneas para analizar su evolución temporal y evaluar el impacto de las medidas relacionadas a la cuarentena estricta y las restricciones a la circulación. Se observa que la mayor cantidad de transacciones se da en el tipo de transporte Colectivo, por lo cual se muestra en gráfico aparte, pero una mejor visualización.

En una tabla, se detallan la cantidad de líneas por tipo de transporte.

Finalmente, se agregó un Mapa en donde se puede visualizar la localización geográfica de los Provincias que utilizan SUBE como medio de adquisición de boleto en transporte público. Al pasar el cursor por la burbuja correspondiente a cada provincia, se puede visualizar un tooltip, en el cual se amplía información sobre la cantidad de usos y tipos de transporte de la provincia consultada. Se puede observar que la utilización en Colectivo se da en varias provincias del país, pero los demás tipos de transporte se concentran en la provincia de Buenos Aires, particularmente en AMBA. Por este motivo, se amplía el análisis en esa región en las próximas solapas.

Todas las gráficas descritas, interactúan entre ellas al aplicar los segmentadores inicialmente mencionados.

### Solapa AMBA Subte Tren

En este caso se agregó un filtro temporal de acuerdo a los períodos de restricciones de circulación decretados por el Gobierno Nacional.

Se muestran tarjetas con la cantidad total de usos realizadas con SUBE, en el transporte Subte y Tren, cada una acompañada de un gráfico medidor en el cual se muestra la variación porcentual respecto al mes anterior, acumulada en el período seleccionado en el segmentador.

Se agregan botones con el fin de ampliar información de los usos de ambos tipos de transporte, utilizando la herramienta “marcadores”. Al seleccionarlos se visualiza, en caso de Subte, un gráfico de embudo con la cantidad de usos por línea, y un gráfico de columnas apiladas y líneas para comparar usos respecto al mes anterior y cuantificar su variación. De la misma manera, al seleccionar el botón Tren, se visualiza la cantidad de usos por líneas en un gráfico de barras, ordenados de mayor a menor, y su correspondiente análisis temporal y variación.

Todas las gráficas descritas, interactúan entre ellas al aplicar los segmentadores inicialmente mencionados.

### Solapa AMBA Colectivo

Al igual que en la solapa previa, se agregó un filtro temporal de acuerdo a los períodos de restricciones de circulación decretados por el Gobierno Nacional, y se pueden cuantificar la cantidad total de usos de la tarjeta SUBE en Colectivo en la región AMBA, la cantidad de líneas y empresas que las administran.

Por la cantidad de líneas existentes en la región, se optó por utilizar la herramienta “parámetro” como selector de valor para definir un ranking dinámico y visualizar, en un gráfico de barras, la cantidad de líneas con más usos.

En una tabla, se muestra la cantidad de líneas (en formato barras) que tienen Jurisdicción Municipal y los correspondientes usos de la tarjeta.

Finalmente, en esta solapa también se analizó la evolución temporal y variación porcentual en un gráfico de barras apiladas y líneas. Se agregó el gráfico medidor para visualizar las variaciones porcentuales de usos respecto al mes anterior, acumuladas en el período seleccionado.

Todas las gráficas descritas, interactúan entre ellas al aplicar los segmentadores inicialmente mencionados.

### Solapa Usuarios

Se utilizan segmentadores de filtro temporal por períodos descriptos y clasificación de usuarios.

Se utiliza un gráfico de tipo Treemap para visualizar la proporción de usuarios según la clasificación respecto a los tipos de beneficio social otorgados y no otorgados. Al deslizar el mouse por encima de esta visualización aparece un Tooltip que amplía información acerca del género de estos usuarios, si es que la tarjeta fue registrada.

Se agregan tarjetas que muestran la cantidad total de SUBE usadas en el período y la cantidad de tarjetas que tiene beneficio social asociado. Esta última información se muestra en un gráfico medidor respecto de la cantidad total de tarjetas.

En un gráfico de barras, se visualiza de la cantidad de tarjetas según el tipo de transporte.

Finalmente, en un gráfico de anillo muestra la proporción porcentual de tarjetas que fueron registradas con algún tipo de beneficio social y aquellas que no.

Todas las gráficas descritas, interactúan entre ellas al aplicar los segmentadores inicialmente mencionados.

## **Futuras líneas**

Este informe, en primera instancia, podría usarse para evaluar la eficacia de las medidas tomadas durante los periodos de cuarentena obligatoria, y en el caso de que sea necesario tomar medidas similares, tener un análisis en el que basarse.

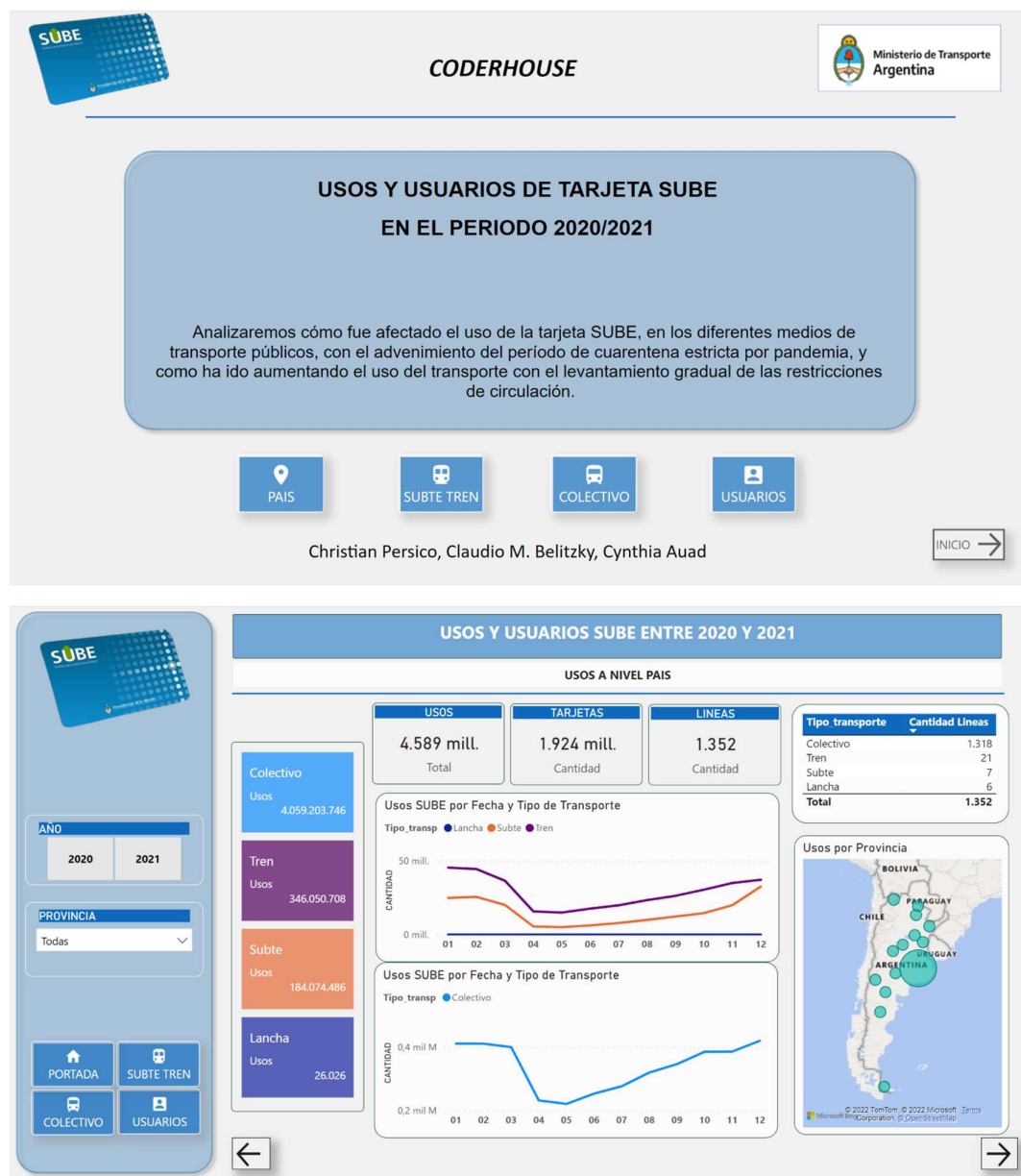
A su vez, si se complementa con mayor cantidad de datos históricos, con predicciones y análisis de tendencias en función del uso del transporte público, se podrían obtener puntos de mejora, tanto en infraestructura como en materias de subsidios, por ejemplo, para aumentar su uso en pos de disminuir la circulación de vehículos particulares.

## **Versionado**

FECHA	DESCRIPCIÓN	REVISIÓN	ESTADO
28/12/2021	Creación del documento – Primera entrega	Juan Ignacio Zampedri	29/12/2021 - Aprobado
03/02/2022	Actualización del documento – Segunda entrega	Juan Ignacio Zampedri	05/02/2021 - Aprobado
14/02/2022	Entrega Proyecto Final	Juan Manuel Creado	Entregado

## Anexo

Visualización de las solapas del dashboard en Power Bi.





**USOS Y USUARIOS SUBE ENTRE 2020 Y 2021**
**USOS AMBA - COLECTIVO**

**USOS**  
**3.667 mill.**  
 Total

**LINEAS**  
**1.352**  
 Cantidad

**EMPRESAS**  
**207**  
 Cantidad

**VARIACION ACUMULADA**  


**MUNICIPIO**

MUNICIPIO	Cantidad Lineas	Usos Colectivos
LA MATANZA	1	79.411.863
LOMAS DE ZAMORA	1	70.295.396
MORENO	1	66.642.499
ALMIRANTE BROWN	1	62.095.166
MERLO	1	57.247.005
TIGRE	1	48.130.254
LA PLATA	1	47.763.841
PILAR	1	37.699.042
ESTEBAN ECHEVERRIA	1	32.962.472
QUILMES	1	26.689.181
<b>Total</b>	<b>117</b>	<b>727.845.023</b>

**PERIODO**  
 ASPO DISPO S/R

**TOPN**  
 6

PORTADA PAIS  
 SUBTE TREN USUARIOS

**TopN Usos por LINEA**

LINEA	USOS
BSAS_LINEA_501G	67 mill.
BSAS_LINEA_203	48 mill.
BSAS_LINEA_365	42 mill.
BSAS_LINEA_440	37 mill.
BSAS_LINEA_324	34 mill.
BSAS_LINEA_236	33 mill.

**Usos Colectivos y Variación por Fecha**


• **ASPO** = Aislamiento Social Preventivo y Obligatorio (20-03-2020 al 07-06-2020)  
 • **DISPO** = Distanciamiento Social Preventivo y Obligatorio (08-06-2020 al 09-04-2021)  
 • **S/R** = Sin Restricción de circulación

