Uso del Transporte Público



CODERHOUSE

Profesora: Rebeca Figueroa Soriano

Tutora: Camila Batistuta

Alumnos: Palena G. y Jourdan G.

Comisión: 24425

Fecha de entrega: 25/02/2022







Contenido

| Descripción de la temática | 3 |
|--|----|
| Alcance | 3 |
| Tipos de análisis | 3 |
| Diagrama entidad-relación | 3 |
| Listado de tablas | 4 |
| Listado de columnas por tablas | 4 |
| Mockup | 5 |
| Transformación de datos | 5 |
| Análisis funcional del tablero | 5 |
| Medidas Calculadas | 7 |
| Paleta de colores | 9 |
| Evolución Dashboard | 10 |
| Diagrama entidad-relación Power BI | 11 |
| Última versión del Dashboard al 15/02/22 | 12 |
| Agregado del botón para eliminar los filtros y corrección en KPIs/Tarjetas | 13 |
| Dashboard: Versión Final | 16 |
| Explicación de cada gráfico/Tabla | 17 |
| Conclusión | 20 |
| Herramientas tecnológicas | 20 |
| Futuras líneas | 20 |



Descripción de la temática

Mediante la página https://datos.gob.ar/dataset/transporte-sube---cantidad-transacciones-usos-por-fecha, seleccionamos las bases de datos del Sistema Único de Boleto Electrónico (SUBE) con el objetivo de analizar los viajes registrados en el período Julio - Noviembre 2021, y determinar cuáles fueron las líneas de colectivo con mayor afluencia de usuarios según cada jurisdicción.

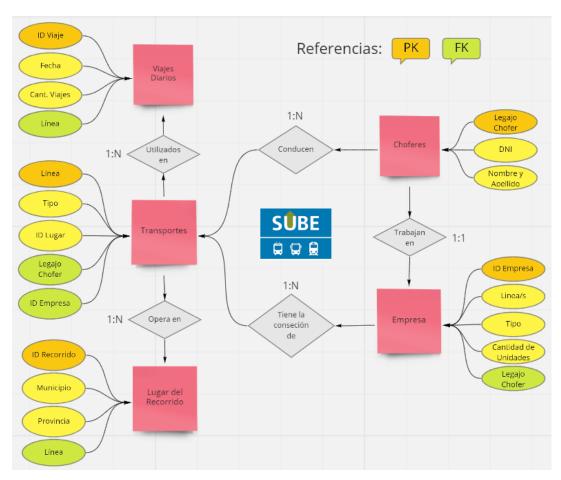
Alcance

El proyecto está orientado a sectores de gestión del gobierno: Niveles estratégicos.

Tipos de análisis

- ✓ **Descriptivo**: Reingeniería de la estructura de los datos de los viajes en transporte público obtenidos a través de la tarjeta SUBE, para generar reportes informativos mensuales
- ✓ **Diagnóstico**: Implementación de PowerBi para dashboards interactivos destinados a brindar información a policy makers (niveles estratégicos)
- ✓ **Predictivo**: Modelo predictivo de comportamiento de usuarios, para identificar el volumen de personas que se trasladarán en transporte público en cada Municipio/Provincia
- ✓ Prescriptivo: Modelo prescriptivo, basado en el modelo predictivo de comportamiento con el cual se puede identificar qué medios de transporte deberán agregar más unidades ante posibles restricciones por rebrotes de COVID, garantizando el aforo recomendado por las autoridades sanitarias

Diagrama entidad-relación





Listado de tablas

| Clave | Choferes | Empresa | Transportes | Lugar de Recorrido | Viajes Diarios |
|-------------|---------------|---------------|-----------------------------|-----------------------|-------------------|
| Primary Key | legajo_chofer | id_empresa | linea | id_recorrido | id_viaje |
| Foreign Key | | legajo_chofer | legajo_chofer id_empresa | linea | linea |

Listado de columnas por tablas

Tabla 1: Choferes

| Tipo de Clave | Campo | Tipo de Campo | |
|---------------|-----------------|---------------|--|
| PK | legajo_chofer | int(50) | |
| - | dni | int(8) | |
| - | nombre_apellido | varchar(50) | |

Tabla 2: Empresa

| Tipo de Clave | Campo | Tipo de Campo |
|---------------|---------------|---------------|
| PK | id_empresa | varchar(50) |
| - | linea | varchar(50) |
| - | tipo | varchar(50) |
| - | q_unidades | int(50) |
| FK | legajo_chofer | int(50) |

Tabla 3: Transportes

| Tipo de Clave | Campo | Tipo de Campo |
|---------------|---------------|---------------|
| PK | linea | varchar(50) |
| - | tipo | varchar(50) |
| - | id_lugar | int(50) |
| FK | legajo_chofer | int(50) |
| FK | id_empresa | varchar(50) |

Tabla 4: Lugar del Recorrido

| Tipo de Clave | Campo | Tipo de Campo |
|---------------|--------------|---------------|
| PK | id_recorrido | int(50) |
| - | municipio | varchar(50) |
| - | provincia | varchar(50) |
| FK | linea | varchar(50) |

Tabla 5: Viajes Diarios

| Tipo de Clave | Campo | Tipo de Campo |
|---------------|----------|---------------|
| PK | id_viaje | int(50) |
| - | fecha | datetime |
| - | q_viajes | int(50) |
| FK | linea | varchar(50) |



Mockup











Transformación de datos

v1.0 - 15/01/2022

Creación

v.1.2 - 25/01/2022

Cambio de Medida a Dimensión de los siguientes campos: dni (tabla Choferes) id_recorrido (tabla Lugar del Recorrido) id_viajes (tabla Viajes Diarios)

v.2.0 - 28/01/2022

Cambio de Formato de Fecha (dd/mm/yyyy)
Cambio Provincia "JN" a CABA
Cambio Municipios "SD" y "SN" a CABA
Agregado de Tabla Calendario
Agregado de 4 Medidas Calculadas y 2 Columnas Calculadas:

Medidas Calculadas:

- 1. Cant. de Viajes = SUM('Viajes Diarios'[q_viajes])
- 2. Promedio de Viajes = AVERAGE('Viajes Diarios'[q_viajes])
- Max. Cantidad de Unidades por Empresa = MAX(Empresa[q_unidades])
- 4. Lineas Totales = COUNT(Transportes[linea])

Columnas Calculadas:

- 1. Mayúsculas = UPPER(Choferes[nombre_apellido])
- 2. Nombre_Apellido_DNI = Choferes[nombre_apellido] & ", "& Choferes[dni]

Análisis funcional del tablero

v.3.0 - 03/02/2022

Evolución Dashboard



Portada

Generamos la portada usando la paleta de colores de SUBE.



Análisis por Empresa

Cantidad de Unidades por Empresa: En este gráfico veremos la cantidad de colectivos con los que cuenta cada de las empresas. Podemos filtrar por Empresa o Línea.

Cantidad de Líneas por Empresa: Este gráfico refleja la cantidad de líneas de colectivos que pertenecen a cada empresa. Podemos filtrar por Empresa o Línea.





Análisis por Línea

Cantidad de Viajes por Mes: El gráfico de columnas muestra los viajes realizados en cada mes, pudiendo filtrarlos por Mes, Empresa y Línea.

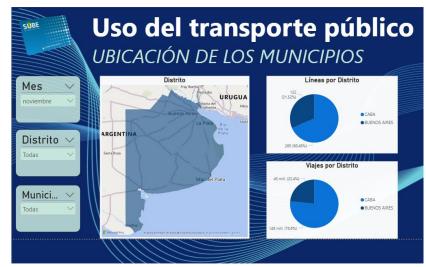
Asimismo, la tabla reflejará la cantidad de viajes por línea con los mismos filtros del gráfico.

Ubicación de los Municipios

Distrito: El mapa muestra los territorios de la provincia o la Ciudad Autónoma de Buenos Aires, tanto con el filtro de Distrito como con el de Municipio.

Líneas por Distrito: La torta muestra la distribución nominal y porcentual de las líneas de colectivo según su ubicación: CABA o GBA. En este caso, el filtro Mes no mostrará cambios pues las cantidades se repiten mes a mes

Viajes por Distrito: El último grafico expone los viajes realizados por distrito según la ubicación de la empresa.





Medidas Calculadas

v.4.0 - 08/02/2022

KPIs:

```
Viajes_Julio = CALCULATE (SUM('Viajes Diarios'[q_viajes]),'Tabla Calendario'[Mes]=7)
Viajes_Agosto = CALCULATE (SUM('Viajes Diarios'[q_viajes]),'Tabla Calendario'[Mes]=8)
Viajes_Septiembre = CALCULATE (SUM('Viajes Diarios'[q_viajes]),'Tabla Calendario'[Mes]=9)
Viajes_Octubre = CALCULATE (SUM('Viajes Diarios'[q_viajes]),'Tabla Calendario'[Mes]=10)
Viajes_Noviembre = CALCULATE (SUM('Viajes Diarios'[q_viajes]),'Tabla Calendario'[Mes]=11)
```

Medidas Calculadas:

1. Con la Medida Calculada que agregamos en la solapa Análisis por línea, constatamos el promedio total de la cantidad de viajes:

```
Prom_Viajes_VAR =
VAR Cant_PromViajes = ROUND (AVERAGE('Viajes Diarios'[q_viajes]),1)
RETURN
"El promedio de viajes realizados es " & Cant_PromViajes
```

Nota: Eliminamos la Medida Calculada **Promedio de Viajes** creada el 28/01:

```
Promedio de Viajes = AVERAGE('Viajes Diarios'[q_viajes])
```

2. Con la Medida Calculada que agregamos en la solapa Análisis por Empresa, podremos identificar rápidamente a aquellas empresas que tienen más o menos de 2 millones de viajes:

```
Viajes_TxT = VAR TotalViajes = SUM('Viajes Diarios'[q_viajes])
VAR text1="Más de 2 mill"
VAR text2="Menos de 2 mill"
RETURN
IF (TotalViajes>2000000,text1,text2)
```

3. En la solapa de Análisis por Línea, agregamos la barra de desplazamiento que permite observar en el gráfico de columnas, como disminuye la cantidad de viajes en un hipotético escenario de aumentos de casos por COVID:

```
Porcentaje de Caída = GENERATESERIES(0, 0.5, 0.05)

Valor Porcentaje de Caída = SELECTEDVALUE('Porcentaje de Caída'[Porcentaje de Descuento])
```

Modificaciones hechas:

Cambio de nombre a Medidas Calculadas:

```
Cant. de Viajes a Total_Viajes
Líneas Totales a Total_Líneas
Máx. Cantidad de Unidades por Empresa a Max_Unid._por_Empresa
```

Agregado del campo "provincia" a la tabla Viajes Diarios (Base de Datos)



Creación de Medidas Calculadas:

```
Viajes_Bs_As = calculate(sum('Viajes Diarios'[q_viajes]),'Viajes Diarios'[provincia]="BUENOS AIRES")
Viajes_CABA = calculate(sum('Viajes Diarios'[q_viajes]),'Viajes Diarios'[provincia]="CABA")
```

Reemplazo del campo "provincia" que traía la tabla "Lugar del Recorrido" por el de la tabla "Viajes Diarios" (solapa "Ubicación de los Municipios")

Agregado de botones de Portada, Empresa, Líneas y Municipios para transitar a través de la navegación de páginas

Modificación de color del fondo de pantalla de las cuatro solapas

Modificación de tamaño y color del título y subtítulo de las cuatro solapas

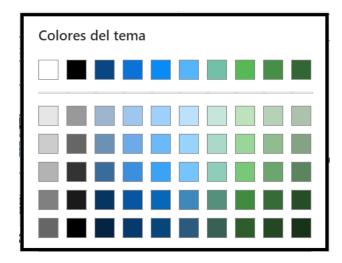
Modificación de formatos de los gráficos (tamaño) por el agregado de lo solicitado en el nuevo desafío Modificación del color de fondo de todos los gráficos

Agregado de Tarjetas en la solapa "Ubicación de los Municipios" con las Medidas Calculadas:

Total_Viajes Viajes_Bs_As Viajes_CABA

Nota: La idea es que el filtro "Municipio" no genere cambios en las visualizaciones, excepto en el mapa.

Paleta de Colores





v.4.0 - 08/02/2022

Evolución Dashboard



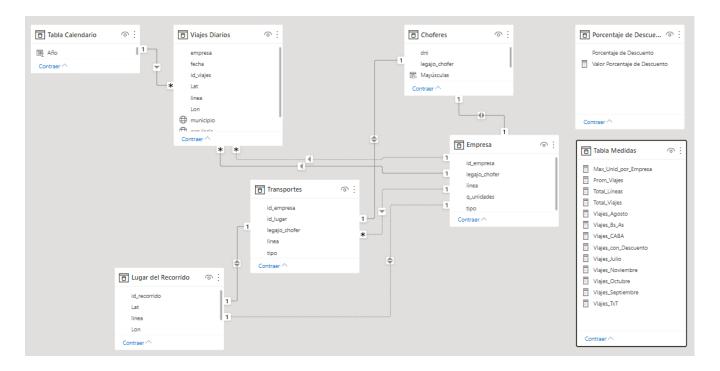








Diagrama entidad-relación Power BI





v.4.1 - 15/02/2022

Evolución Dashboard: Correcciones post devolución Desafío Medidas Calculadas Avanzadas



Portada

Quitamos el botón de "Portada" ya que no era necesario.

Modificamos el color de los botones de navegación entre páginas por uno más claro.

Modificamos la imagen del botón "Municipios" porque no era clara.

Análisis por Empresa, Análisis por Línea y Ubicación de los Municipios

Filtros: Modificamos el color y redujimos el tamaño de la fuente de los títulos.





Además, eliminamos los botones de navegación entre páginas que no tenían sentido por estar ya en esa solapa y no conducir hacia ningún lado. Ejemplo: Botón "Línea" en la solapa "Análisis por Línea".





v.4.2 - 18/02/2022



v.4.2 - 23/02/2022

Solucionamos los "(En blanco)" de los KPIs sin información de la solapa Análisis por Línea, reemplazándolos por "0". Esto lo aplicamos en todos los meses reemplazando las fórmulas.

Vista anterior



Vista actual





Formulas anteriores:

```
Viajes_Noviembre = CALCULATE (SUM('Viajes Diarios'[q_viajes]), 'Tabla Calendario'[Mes]=11)
Formulas actuales:
Viajes Julio =
VAR ViajesJulio =CALCULATE(SUM('Viajes Diarios'[q_viajes]),'Tabla Calendario'[Mes]=7)
RETURN IF(ISBLANK(ViajesJulio),0,ViajesJulio)
Viajes Agosto =
VAR ViajesAgosto =CALCULATE(SUM('Viajes Diarios'[q_viajes]),'Tabla Calendario'[Mes]=8)
RETURN IF(ISBLANK(ViajesAgosto),0,ViajesAgosto)
Viajes_Septiembre =
VAR ViajesSeptiembre =CALCULATE(SUM('Viajes Diarios'[q_viajes]), 'Tabla Calendario'[Mes]=9)
RETURN IF(ISBLANK(ViajesSeptiembre),0,ViajesSeptiembre)
Viajes_Octubre =
VAR ViajesOctubre =CALCULATE(SUM('Viajes Diarios'[q_viajes]), 'Tabla Calendario'[Mes]=10)
RETURN IF(ISBLANK(ViajesOctubre),0,ViajesOctubre)
Viajes_Noviembre =
VAR ViajesNoviembre =CALCULATE(SUM('Viajes Diarios'[q_viajes]), 'Tabla Calendario'[Mes]=11)
```

Viajes_Julio = CALCULATE (SUM('Viajes Diarios'[q_viajes]), 'Tabla Calendario'[Mes]=7)
Viajes_Agosto = CALCULATE (SUM('Viajes Diarios'[q_viajes]), 'Tabla Calendario'[Mes]=8)
Viajes_Septiembre = CALCULATE (SUM('Viajes Diarios'[q_viajes]), 'Tabla Calendario'[Mes]=9)
Viajes Octubre = CALCULATE (SUM('Viajes Diarios'[q_viajes]), 'Tabla Calendario'[Mes]=10)

Solucionamos los "(En blanco)" de las tarjetas sin información de la solapa Ubicación de los Municipios, reemplazándolos por "O". Esto lo aplicamos en las 3 tarjetas:

Vista anterior



RETURN IF(ISBLANK(ViajesNoviembre),0,ViajesNoviembre)

Vista actual





Formulas anteriores:

```
Viajes_Bs_As = calculate(sum('Viajes Diarios'[q_viajes]), 'Viajes Diarios'[provincia]="BUENOS AIRES")
Viajes_CABA = calculate(sum('Viajes Diarios'[q_viajes]), 'Viajes Diarios'[provincia]="CABA")

Formulas actuales:

Total_Viajes =
VAR ViajesTotales =SUM('Viajes Diarios'[q_viajes])
RETURN IF(ISBLANK(ViajesTotales),0,ViajesTotales)
Viajes_Bs_As =
VAR ViajesBsAs = calculate(sum('Viajes Diarios'[q_viajes]), 'Viajes Diarios'[provincia]="BUENOS AIRES")
RETURN IF(ISBLANK(ViajesBsAs),0,ViajesBsAs)
Viajes_CABA =
VAR ViajesCABA = calculate(sum('Viajes Diarios'[q_viajes]), 'Viajes Diarios'[provincia]="CABA")
RETURN IF(ISBLANK(ViajesCABA),0,ViajesCABA)
```

Agregamos el filtro "Mes" a la solapa de Análisis por Empresa:

Total_Viajes = SUM('Viajes Diarios'[q_viajes])



v.4.3 - 24/02/2022

Cambio del color de fondo: Aplicamos la siguiente imagen como papel tapiz



Coordenadas de los Municipios: Agregamos los datos de Latitud y Longitud a la tabla Viajes Diarios para poder ver la ubicación de cada Municipio en el mapa. Además, reubicamos las tarjetas eliminando la totalizadora y agregamos la tabla que muestra la cantidad de viajes por municipio:



Dashboard - Versión Final

Portada



Análisis por Empresa



Análisis por Línea



Ubicación de los Municipios





Explicación de lo que debemos observar en cada gráfico

Análisis por Empresa:

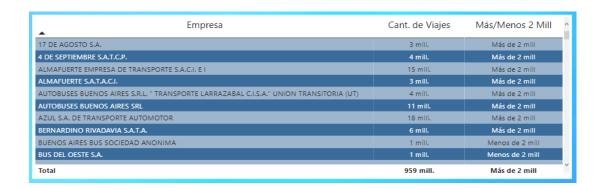


El gráfico muestra la cantidad de colectivos que tiene cada empresa, pudiendo filtrarse por Mes, Empresa o Línea de colectivo. En la vista principal dejamos las 5 con mayor cantidad. En este caso, la cantidad de unidades por mes no varía ya que es siempre la misma.

Las barras reflejan la cantidad de Líneas con las que cuenta cada empresa. Al igual que en el anterior, podrá filtrarse por Mes, Empresa o Línea, y en la vista principal dejamos el Top 5 de las empresas que más líneas tienen. Al filtrarse por Línea traerá solamente esa línea y no todas las que pertenecen a esa empresa. En este caso, la cantidad de líneas por mes no varía ya que es siempre la misma.



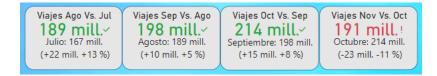
En esta tabla observaremos la cantidad de viajes por empresa, más la columna resumen con los que tuvieron más o menos de 2 millones. Al igual que en los dos gráficos anteriores, podrá filtrarse por Mes, Empresa o Línea, mostrando siempre en primer lugar aquella empresa con mayor cantidad de viajes.





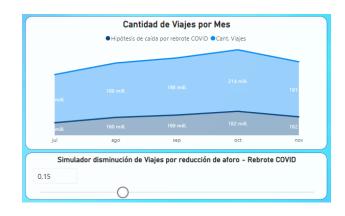
Análisis por Línea:

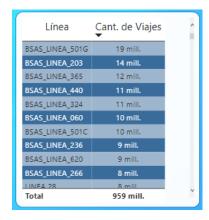
Los KPIs muestran la relación de cada mes respecto al anterior: Agosto Vs. Julio, Septiembre Vs. Agosto, Octubre Vs. Septiembre, etc. Podrán filtrarse por Mes, Empresa y Línea para obtener mayor grado de detalle:



En cuanto al gráfico de Cantidad de Viajes por mes, además de ir actualizándose con los 3 filtros que ya hemos mencionado, la barra de desplazamiento nos permitirá observar como disminuye la cantidad de viajes en un hipotético escenario de aumentos de casos por COVID.

También reemplazamos el gráfico de columnas agrupadas por este de áreas apiladas.





Esta tabla nos permitirá ver la cantidad detallada de viajes de todas las líneas de la empresa seleccionada en el filtro Empresa.

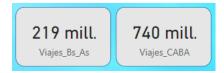
Hemos ido armando el dashboard desde lo general (Empresas), hacia lo particular (Líneas), para luego observar en la solapa de Ubicación de los Municipios como impacta en cada uno la cantidad de viajes.

<u>Nota</u>: Al analizar los viajes de CABA (Ciudad Autónoma de Buenos Aires), lo veremos como "Distrito" y como "Municipio", ya que no contamos con los barrios donde están ubicadas las cabeceras.

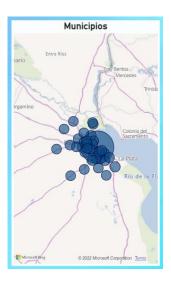


Ubicación de los Municipios:

Las tarjetas muestran la cantidad de viajes por cada distrito, y las mismas se irán actualizando con los filtros Mes y Municipio. El filtro por Distrito no las modifica porque no tendría ningún sentido hacerlo:



El gráfico de mapa refleja la ubicación de los municipios de cada una de las líneas, y con mayor tamaño de esferas a aquellas con más cantidad de viajes. Este gráfico también irá mostrando diferente información a medida que vayamos interactuando con los filtros.





El primer gráfico de torta nos indicará la cantidad de viajes por distrito y su participación porcentual sobre el total.

En el segundo veremos la cantidad de líneas por distrito también con su correspondiente participación porcentual sobre el total.

En ambos casos, podremos filtrar los datos por Mes, Distrito y Municipio.



En definitiva, la tabla será la que nos permita observar cuáles son los municipios con mayor afluencia de usuarios. Y al pasar con el puntero del mouse sobre cada municipio, veremos la apertura del mismo con la cantidad de viajes por cada línea de ese municipio gracias al Tooltip que hemos creado:





Conclusión

En conclusión, el distrito con mayor afluencia de usuarios es la Ciudad Autónoma de Buenos Aires (CABA), con 148 M de viajes mensuales en promedio, repartidos en 265 líneas de colectivos, siendo la más abordada la 203 con un promedio mensual de 2.7 M de viajes.





Sin embargo, el récord mensual lo tiene la línea 501G con casi 4 M de viajes mensuales en promedio, convirtiendo a Moreno en el 4to. municipio con mayor afluencia y contando únicamente con esa sola línea!

Herramientas tecnológicas

Trabajamos con SQL y Power BI

Futuras líneas

Nos gustaría que este proyecto fuera fuente de información para las autoridades gubernamentales, para que les ayude a tomar mejores decisiones en cuanto al transporte público en épocas de pandemia.