

MySQL APLICADO A UNA BASE DE DATOS

ENUNCIADOS Y RESOLUCIONES

1) Crear una base de datos denominada TF_BASES en MySQL Workbench.

Resolución en script SQL.

2) Descargar el archivo csv "Viajes y viajeros por mes de inicio, provincia de origen y destino" de la siguiente url: <https://www.datos.gob.ar/dataset/turismo-previaje> e importar el archivo en una tabla denominada DATASET dentro de la base de datos creada en el punto 1

Resolución en script SQL.

3) Ejecutar el script "script_tf_bases.sql" que se encuentra en esta misma sección.

A continuación se explica brevemente que realiza el "script_tf_bases.sql".

El código SQL del script realiza una creación y modificación de las tablas de la base de datos "TF_BASES". A continuación, se explica brevemente qué hace cada comando:

- a) CREATE TABLE DATOS_PREVIAJE: Se crea una tabla a la que se denomina "DATOS_PREVIAJE" con las columnas "ID", "PERIODO", "ID_PROVINCIA_ORIGEN", "ID_PROVINCIA_DESTINO", "CANT_VIAJES", "CANT_VIAJEROS" e "ID_EDICION". La columna "ID" es una primary key (PK) con valores autoincrementales.
- b) INSERT INTO DATOS_PREVIAJE: Se inserta datos en la tabla "DATOS_PREVIAJE" seleccionando todo lo que hay en la tabla "DATASET".
- c) UPDATE DATOS_PREVIAJE: Se modifica la tabla "DATOS_PREVIAJE" para cambiar el nombre de las provincias en caso de corresponder. Por ejemplo, todas las provincias que terminen en "rdoba" se cambian por "Córdoba".
- d) CREATE TABLE EDICION: Se crea una tabla denominada "EDICION" con la columna "ID_EDICION" y "EDICION". La columna "ID_EDICION" es una PK con valores autoincrementales.
- e) INSERT INTO EDICION: Se inserta datos en la tabla "EDICION" seleccionando los valores únicos de la columna "ID_EDICION" de la tabla "DATOS_PREVIAJE".
- f) CREATE TABLE PROVINCIA: Se crea una tabla llamada "PROVINCIA" con la columna "ID_PROVINCIA" y "PROVINCIA". La columna "ID_PROVINCIA" es una PK con valores autoincrementales.
- g) INSERT INTO PROVINCIA: Se inserta datos en la tabla "PROVINCIA" seleccionando los valores únicos de la columna "ID_PROVINCIA_ORIGEN" de la tabla "DATOS_PREVIAJE" y ordenándolos por orden alfabético.
- h) UPDATE DATOS_PREVIAJE: Se modifica la tabla "DATOS_PREVIAJE" para cambiar los valores de las columnas "ID_EDICION", "ID_PROVINCIA_ORIGEN" e "ID_PROVINCIA_DESTINO". Para ello, se realiza una unión con las tablas "EDICION" y "PROVINCIA" para obtener los valores correspondientes.
- i) ALTER TABLE DATOS_PREVIAJE: Se modifica la estructura de la tabla "DATOS_PREVIAJE" para cambiar los tipos de datos de las columnas "ID_PROVINCIA_ORIGEN", "ID_PROVINCIA_DESTINO" e "ID_EDICION". También agrega dos

restricciones de clave externa (FK) para las columnas "ID_EDICION" y "ID_PROVINCIA_ORIGEN" que se relacionan con las tablas "EDICION" y "PROVINCIA", respectivamente.

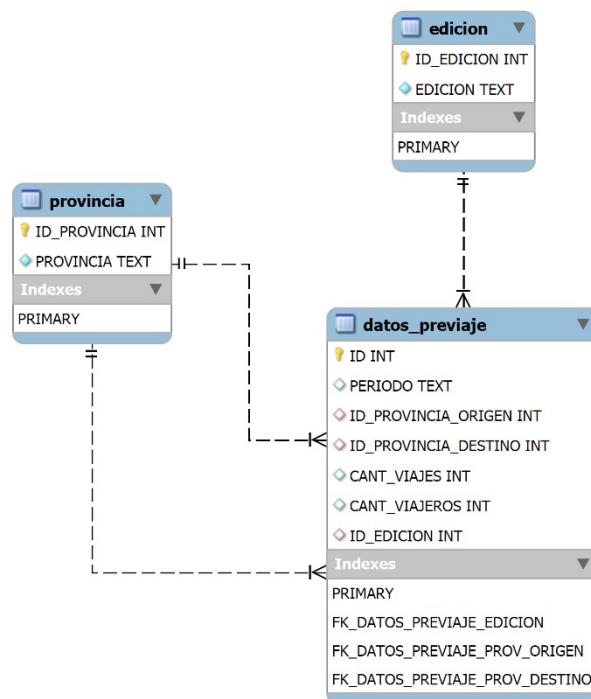
j) ALTER TABLE DATOS_PREVIAJE ADD: Se agrega tres restricciones de clave externa (FK) a la tabla "DATOS_PREVIAJE" para garantizar que los valores en las columnas "ID_EDICION", "ID_PROVINCIA_ORIGEN" e "ID_PROVINCIA_DESTINO" existan en las tablas "EDICION", "PROVINCIA_ORIGEN" y "PROVINCIA_DESTINO", respectivamente.

k) DROP TABLE DATASET: Elimina la tabla "DATASET" de la base de datos, si existe.

l) Finalmente, realiza tres verificaciones para controlar que el número de filas en las tablas "DATOS_PREVIAJE", "PROVINCIA" y "EDICION" sea igual a los valores esperados. Si el recuento de filas coincide, devuelve "OK"; de lo contrario, devuelve NULL.

4) Luego de completar el punto 3, obtener el diagrama del modelo lógico en el Workbench.

Figura N° 1:
DIAGRAMA DE MODELO LÓGICO (DER)



El diagrama de modelo lógico permite visualizar y resumir las tablas con sus respectivas columnas que se explican a continuación:

- 1) Tabla "EDICION": está compuesta por dos columnas; "ID_EDICION" que es PK y "EDICION".
- 2) Tabla "PROVINCIA": está compuesta por dos columnas; "ID_PROVINCIA" que es PK y "PROVINCIA".
- 3) Tabla "DATOS_PREVIAJE": está compuesta por siete columnas; "ID" que es PK, "PERIODO", "ID_PROVINCIA_ORIGEN" e "ID_PROVINCIA_DESTINO" que son FK, "CANT_VIAJES", "CANT_VIAJEROS" y "ID_EDICION" que es FK.

Además, el modelo resume las relaciones que existen entre las tres tablas e indica cuáles de sus columnas son PK y FK. Así por ejemplo ID_EDICION es PK en la tabla "EDICIÓN" y a su vez es FK en la tabla "DATOS_PREVIAJE".

Por último, se indica los tipos de datos. Hay columnas que son del tipo TEXT (carácter no numérico) y otras INT (valores enteros).

5. a) Determinar la cantidad de viajes y personas que viajaron por provincia de destino y edición, seleccionando los 10 registros que tienen más viajes.

Tabla Nº 1:
TOTAL DE VIAJES Y TOTAL DE VIAJEROS, POR PROVINCIA DESTINO Y EDICIÓN (TOP 10)

Provincia	Edición	Total_viajes	Total_viajeros
Buenos Aires	previaje 2	291202	880238
Río Negro	previaje 2	247391	660153
Córdoba	previaje 2	133707	379308
Santa Cruz	previaje 2	120946	305998
Mendoza	previaje 2	113834	302648
Tierra del Fuego, Antártida e Islas del Atlántico ...	previaje 2	102962	254039
Neuquén	previaje 2	97916	289623
Misiones	previaje 2	86338	247991
Salta	previaje 2	80602	206495
Ciudad Autónoma de Buenos Aires	previaje 2	58645	135719

Análisis de resultados: La tabla indica los 10 registros que tienen mayor cantidad de viajes, siendo la provincia de Buenos Aires el destino número 1 con un total de 291.202 viajes realizados en la edición “previaje 2”.

Además, se visualiza que la edición “previaje 2” se repite en los 10 destinos, esto quiere decir que hubo una mayor cantidad de personas que utilizaron la edición para realizar turismo en Argentina, en comparación con las ediciones “previaje 1” y “previaje 3” que no aparecen entre los 10 mejores registros.

Explicación del código utilizado: Brevemente, la consulta del script obtiene información sobre provincias, las ediciones, los totales de viajes y los totales de viajeros de la tabla “DATOS_PREVIAJE”.

Para obtener la suma de las columnas "CANT_VIAJES" y "CANT_VIAJEROS" de la tabla "DATOS_PREVIAJE" se utiliza SUM(), y se renombran como "Total_Viajes" y "Total_Viajeros", respectivamente.

Luego, se realiza la unión de las tablas "PROVINCIA" y "EDICIÓN" para obtener información adicional sobre la provincia y la edición correspondientes. Los resultados se agrupan mediante un GROUP BY por “Provincia” y “Edición”. Posteriormente, se ordenan según el número total de viajes y se limitan a los primeros 10 resultados.

5.b) Determinar la cantidad de viajes y el porcentaje que representa esa cantidad sobre el total de viajes, agrupando por período y edición, obteniendo subtotales por edición (con su correspondiente porcentaje). Los resultados se deben ordenar primero por edición y luego por período.

Tabla Nº 2.1:

TOTAL DE VIAJES EN CANTIDAD Y EN PORCENTAJE, POR PERÍODO Y EDICIÓN (Previaje 1)

Periodo	Edición	Total_Viajes	Porcentaje_sobre_Edición	Porcentaje_sobre_Total
2021-01	previaje 1	56403	33.57%	2.70%
2021-02	previaje 1	42956	25.56%	2.06%
2021-03	previaje 1	31720	18.88%	1.52%
2021-04	previaje 1	9695	5.77%	0.46%
2021-05	previaje 1	4686	2.79%	0.22%
2021-06	previaje 1	2932	1.74%	0.14%
2021-07	previaje 1	4360	2.59%	0.21%
2021-08	previaje 1	4272	2.54%	0.20%
2021-09	previaje 1	4209	2.50%	0.20%
2021-10	previaje 1	3283	1.95%	0.16%
2021-11	previaje 1	2122	1.26%	0.10%
2021-12	previaje 1	1403	0.83%	0.07%

Tabla Nº 2.2:
TOTAL DE VIAJES EN CANTIDAD Y EN PORCENTAJE,
POR PERÍODO Y EDICIÓN (Previale 2 y 3)

Periodo	Edición	Total_Viajes	Porcentaje_sobre_Edición	Porcentaje_sobre_Total
2021-11	previale 2	116709	7.53%	5.59%
2021-12	previale 2	189873	12.25%	9.09%
2022-01	previale 2	346003	22.32%	16.57%
2022-02	previale 2	507834	32.76%	24.33%
2022-03	previale 2	175826	11.34%	8.42%
2022-04	previale 2	72241	4.66%	3.46%
2022-05	previale 2	26869	1.73%	1.29%
2022-06	previale 2	19551	1.26%	0.94%
2022-07	previale 2	29030	1.87%	1.39%
2022-08	previale 2	20415	1.32%	0.98%
2022-09	previale 2	16908	1.09%	0.81%
2022-10	previale 2	15966	1.03%	0.76%
2022-11	previale 2	9443	0.61%	0.45%
2022-12	previale 2	3630	0.23%	0.17%
2022-10	previale 3	166822	45.17%	7.99%
2022-11	previale 3	191852	51.94%	9.19%
2022-12	previale 3	10684	2.89%	0.51%
Subtotales por edición	previale 1	168041	100.00%	8.05%
Subtotales por edición	previale 2	1550298	100.00%	74.26%
Subtotales por edición	previale 3	369358	100.00%	17.69%

Análisis de resultados: La tabla indica la cantidad total y el porcentaje total de viajes realizados por edición y por período. Así por ejemplo, se visualiza que para el período 2022-02 en la edición "previale 2" hubo un total de 507.834 viajes realizados que representan un 24,33% del total, siendo este el mayor registro alcanzado por período y edición.

A modo de síntesis los porcentajes por edición respecto al total general son los siguientes:

- 1) Previale 1: 168.041 viajes que representan el 8,05% del total.
- 2) Previale 2: 1.550.298 viajes que representan el 74,26% del total. 3)
- Previale 3: 369.358 viajes que representan el 17,69% del total.

Además, se visualiza subtotales por edición siendo los mayores registros en cada edición los siguientes:

- 1) Previale 1: 56.403 viajes en el período 2021-01, que representa el 33,57% para dicha edición.
- 2) Previale 2: 507.834 viajes en el período 2022-02, que representa el 32,76% para dicha edición. 3)
- Previale 3: 191.852 viajes en el período 2022-11, que representa el 51,94% para dicha edición.

Explicación del código utilizado: Brevemente, la consulta del script se divide en dos consultas que luego se unen utilizando la cláusula UNION.

En la primera consulta, se obtiene información de la tabla "DATOS_PREVIAJE" y la tabla "EDICION". La columna "PERIODO" se formatea en el formato 'AAAA-MM' y se la renombra como "Periodo". La columna "EDICION" se

selecciona de la tabla "EDICION" y se renombra como "Edición". También se calcula la suma de la columna "CANT_VIAJES" de la tabla "DATOS_PREVIAJE" y se la renombra como "Total_Viajes".

Además, se calcula el porcentaje de los viajes sobre la edición actual mediante un *OVER()* y se muestra como una cadena con dos decimales y el símbolo '%'. También se calcula el porcentaje de los viajes sobre el total de viajes de la tabla "DATOS_PREVIAJE" y se muestra de la misma manera.

Los resultados se agrupan por la columna "ID_EDICION" y "PERIODO". Luego se ordenan en orden ascendente por "ID_EDICION" y luego por "PERIODO".

En la segunda consulta, se calcula subtotales por edición. La columna "EDICIÓN" se selecciona de la tabla "EDICION" y se renombra como "Edición". Se realiza una suma de la columna "CANT_VIAJES" de la tabla "DATOS_PREVIAJE" y se la renombra como "Total_Viajes". Luego se calculan los porcentajes de los viajes sobre la edición actual mediante un *OVER()* y sobre el total de viajes de la misma manera que en la primera consulta.

Los resultados se agrupan por la columna "ID_EDICION" y finalmente se ordenan en orden ascendente.

5.c) Determinar las 5 provincias en las que realizaron mayor cantidad de viajes dentro de la misma provincia.

Tabla Nº 3:
TOTAL DE VIAJES QUE SE REALIZARON DENTRO DE LA MISMA PROVINCIA (TOP 5)

	Provincia	Viajes_misma_provincia
►	Buenos Aires	131648
	Córdoba	29224
	Ciudad Autónoma de Buenos Aires	8178
	Mendoza	6062
	Río Negro	5358

Análisis de resultados: La tabla resume las 5 provincias en las cuales se realizaron mayor cantidad de viajes dentro de la misma provincia. Ellas son; Buenos Aires, Córdoba, CABA, Mendoza y Río Negro

Explicación del código utilizado: Brevemente, la consulta del script obtiene información sobre las provincias y la cantidad de viajes que tienen el mismo origen y destino.

Primero, consulta selecciona la columna "PROVINCIA" de la tabla "PROVINCIA" y la renombra como "Provincia". Además, utiliza la función *SUM()* para sumar el número de filas que cumplen con la condición establecida. Esta cuenta se renombra como "Viajes_misma_provincia".

Segundo, el *JOIN provincia p ON dp.ID_PROVINCIA_ORIGEN = p.ID_PROVINCIA* realiza una operación de unión entre la tabla "DATOS_PREVIAJE" y la tabla "PROVINCIA" utilizando un JOIN. Las tablas se unen en las columnas "ID_PROVINCIA_ORIGEN" de la tabla "DATOS_PREVIAJE" y "ID_PROVINCIA" de la tabla "PROVINCIA". Esto se hace para obtener el nombre de la provincia de origen correspondiente.

Tercero, el *WHERE ID_PROVINCIA_ORIGEN = ID_PROVINCIA_DESTINO* establece una condición para filtrar los resultados de la consulta, solo seleccionará los registros en los que la provincia de origen es igual a la provincia de destino. Esto asegura que solo se cuenten los viajes que tienen el mismo origen y destino.

Cuarto, los resultados se agrupan con un GROUP BY aplicado a "Provincia".

Quinto, los resultados se ordenan de manera descendente según la columna "Viajes_misma_provincia", mostrando la cantidad de viajes con el mismo origen y destino en orden descendente.

Finalmente, se limitan los 5 primeros resultados.

5.d) Determinar la proporción de viajes sobre el total, agrupando por región de destino y edición, calculando adicionalmente los porcentajes por región de destino. Los resultados se deben ordenar por las regiones, las cuales se deben armar de la siguiente manera:

- Cuyo: Mendoza, San Juan y San Luis.
- NOA: La Rioja, Catamarca, Jujuy, Tucumán, Salta, Santiago del Estero.
- NEA: Misiones, Corrientes, Entre Ríos, Formosa, Chaco.
- Centro: Buenos Aires, Córdoba, Santa Fe, Ciudad Autónoma de Buenos Aires, La Pampa.
- Patagonia: Chubut, Neuquén, Río Negro, Santa Cruz.
- Extremo austral: Tierra del Fuego, Antártida e Islas del Atlántico Sur

Tabla Nº 4:
TOTAL DE VIAJES EN PORCENTAJE POR REGIÓN DE DESTINO Y EDICIÓN

	Región_Destino	Edición	Total_Porcentaje_sobre_Total	Total_Porcentaje_Región_Destino
▶	HULL	Subtotales	99.96%	600.00%
	Centro	Subtotales	31.55%	100.00%
	Centro	previaje 1	2.40%	7.62%
	Centro	previaje 2	23.69%	75.08%
	Centro	previaje 3	5.46%	17.30%
	Cuyo	Subtotales	9.71%	100.00%
	Cuyo	previaje 1	0.73%	7.52%
	Cuyo	previaje 2	7.10%	73.18%
	Cuyo	previaje 3	1.87%	19.31%
	Extremo Austral	Subtotales	6.66%	100.00%
	Extremo Austral	previaje 1	0.69%	10.36%
	Extremo Austral	previaje 2	4.93%	74.04%
	Extremo Austral	previaje 3	1.04%	15.60%

	Región_Destino	Edición	Total_Porcentaje_sobre_Total	Total_Porcentaje_Región_Destino
	NEA	Subtotales	8.86%	100.00%
	NEA	previaje 1	0.68%	7.64%
	NEA	previaje 2	6.11%	68.90%
	NEA	previaje 3	2.08%	23.46%
	NOA	Subtotales	11.22%	100.00%
	NOA	previaje 1	0.65%	5.85%
	NOA	previaje 2	7.80%	69.50%
	NOA	previaje 3	2.77%	24.66%
	Patagonia	Subtotales	31.97%	100.00%
	Patagonia	previaje 1	2.89%	9.04%
	Patagonia	previaje 2	24.61%	76.98%
	Patagonia	previaje 3	4.47%	13.98%

Análisis de resultados: En la tabla puede visualizarse, para cada región destino y edición, el porcentaje de viajes respecto al total. Se observa que el destino de mayor porcentaje es la "Patagonia" con un 31,97%, seguido por la región "Centro" con un 31,55%.

Por otro lado, la tabla indica para cada región destino y edición, el porcentaje de viajes respecto a cada región destino. Se observa que para las diferentes regiones los porcentajes más elevados y significativos son los que corresponden a la edición "previaje 2". Por ejemplo, dentro de la región "Patagonia" la edición "previaje 2" tiene un porcentaje del 76,98% o la región "Centro" para la edición "previaje 2" tiene un porcentaje del 75,08%.

Explicación del código utilizado: Brevemente, la consulta principal comienza seleccionando cuatro columnas: "Región_Destino", "Edición", "Total_Porcentaje_sobre_Total" y "Total_Porcentaje_Región_Destino".

La columna "Región_Destino" se deriva de la tabla "provincia_previaje", donde se asigna una región a cada provincia. Las regiones se definen según la provincia de destino del viaje. Hay seis regiones posibles: Cuyo, NOA (Noroeste Argentino), NEA (Noreste Argentino), Centro, Patagonia y Extremo Austral.

La columna "Edición" proviene de la tabla "edicion_previaje" y representa la edición del viaje.

Las dos columnas siguientes, "Total_Porcentaje_sobre_Total" y "Total_Porcentaje_Región_Destino", se calculan con la función de agregación *SUM()* y *ROUND()* para obtener el porcentaje total de viajes en relación con el total general y el porcentaje de viajes en relación con el total por región de destino, respectivamente. Los resultados se presentan como cadenas de texto concatenadas con el símbolo '%' al final.

Dentro de la consulta principal, hay una subconsulta que calcula el porcentaje de viajes para cada región y edición. Utiliza la función *SUM()* para calcular el total de viajes y el porcentaje sobre el total general y por región de destino.

La cláusula *LEFT JOIN* se utiliza para combinar las tablas "datos_previaje", "provincia_previaje" y "edicion_previaje" según sus claves primarias y externas.

La cláusula *GROUP BY* agrupa los resultados por "Región_Destino" y "Edición".

La cláusula *WITH ROLLUP* se utiliza para generar filas de subtotales y totales para las agrupaciones especificadas.

Finalmente, los resultados se ordenan por "Región_Destino" y "Edición". Se ordenan los resultados de la consulta según "Región_Destino".

6) En este enunciado se realiza la creación de una función mediante la utilización de FUNCTION para calcular el número de viajes totales según tres parámetros: provincia destino, período y edición de previaje. La función admite no ingresar datos en uno o dos de sus parámetros.

En el script adjunto se indica la creación de la función y se realiza las siguientes comprobaciones, obteniendo los resultados indicados en los puntos 5.a) y 5.b).

COMPROBACIÓN 5.a):

CONSULTA 1: Cantidad de viajes a Buenos Aires como destino y en "previaje 2"

```
SELECT ID_PROVINCIA, PROVINCIA FROM provincia;
```

```
SELECT Funcion_Total_Viajes ('1', NULL, '2') AS Total_viajes_BuenosAires_Previaje2;
```

Resultado:

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	Total_viajes_BuenosAires_Previaje2			
▶	291202			

CONSULTA 2: Cantidad de viajes a Río Negro como destino y en "previaje 2"

```
SELECT ID_PROVINCIA, PROVINCIA FROM provincia;
```

```
SELECT Funcion_Total_Viajes ('16', NULL, '2') AS Total_viajes_RíoNegro_Previaje2;
```

Resultado:

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	Total_viajes_RíoNegro_Previaje2			
▶	247391			

COMPROBACIÓN 5.b):

CONSULTA 3: Cantidad de viajes para "previaje 1" en el período 2021/01

```
SELECT PERIODO FROM datos_previaje;
```

```
SELECT Funcion_Total_Viajes (NULL, '2021-01', '1') AS Total_viajes_Previaje1_2021_01;
```

Resultado:

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	Total_viajes_Previaje1_2021_01			
▶	56403			

CONSULTA 4: Cantidad de viajes para "previaje 1" en el período 2021/02

SELECT PERIODO FROM datos_previaje;

SELECT Funcion_Total_Viajes (NULL, '2021-02', '1') AS Total_viajes_Previaje1_2021_02;

Resultado:

Result Grid			Filter Rows:		Export:	Wrap Cell Content:
	Total_viajes_Previaje1_2021_02					
▶	42956					

CONSULTA 5: Cantidad de viajes para “previaje 2”

SELECT PERIODO FROM datos_previaje;

SELECT Funcion_Total_Viajes (NULL, NULL, '2') AS Total_Previaje2;

Resultado:

Result Grid			Filter Rows:		Export:	Wrap Cell Content:
	Total_Previaje2					
▶	1550298					

En las cuatro consultas anteriores se demuestra que la función realiza la totalización de los viajes de igual manera que los puntos 5.a) y 5.b).