PRÁCTICA #1

Bases de Datos Distribuidas

Descripción breve

En esta práctica el equipo aplicará los conceptos de programación de consultas en SQL

- Urrutia González Brenda
- Olea García Alan
- Juárez Anguiano Mario Alexis

Equipo 03

05/03/2025

Consulta 1.	Listar el top 5 de las entidades con más casos confirmados por cada por cada uno de los años registrados en la base de datos.
Requisitos:	N/A
Significado de valores de los catálogos:	- CLASIFICACION_FINAL: Confirmado por asociación clínica-epidemiológica, 2 = Confirmado por dictaminación, 3 = Confirmado por laboratorio - ENTIDAD_RES: Nos da la entidad de residencia del paciente
Responsable:	Alan Olea García.
Comentarios:	- WITH genera resultados temporales que pueden ser referenciados dentro de una consulta - RANK asigna un "ranking" o número de posición a cada fila dentro de un conjunto de resultados, Los números de ranking empiezan desde 1, y si hay filas con el mismo valor (en este caso, Casos_confirmados), se asigna el mismo rango, pero el siguiente rango salta por la cantidad de filas con el mismo valor. - WITH genera resultados temporales que pueden ser referenciados dentro de una consulta - RANK asigna un "ranking" o número de posición a cada fila dentro de un conjunto de resultados, Los números de ranking empiezan desde 1, y si hay filas con el mismo valor (en este caso, Casos_confirmados), se asigna el mismo rango, pero el siguiente rango salta por la cantidad de filas con el mismo valor.

	Entidad	Año	Casos_confirmados
1	09	2020	362275
2	15	2020	161809
3	11	2020	86268
4	19	2020	85561
5	26	2020	56700
6	09	2021	653053
7	15	2021	234438
8	19	2021	126299
9	11	2021	119429
10	14	2021	112295
11	09	2022	351093
12	15	2022	132309
13	19	2022	97680
14	11	2022	71416
15	24	2022	70474

Consulta 2.	Listar el municipio con más casos confirmados recuperados por estado y por año
Requisitos:	N/A
Significado de valores de los catálogos:	- FECHA_INGRESO: Fecha en la que el paciente fue confirmado - MUNICIPIO_RES: Nos da el municipio de residencia del paciente
Responsable:	Alan Olea García.
Comentarios:	- WITH genera resultados temporales que pueden ser referenciados dentro de una consulta - RANK asigna un "ranking" o número de posición a cada fila dentro de un conjunto de resultados, Los números de ranking empiezan desde 1, y si hay filas con el mismo valor (en este caso, Casos_confirmados), se asigna el mismo rango, pero el siguiente rango salta por la cantidad de filas con el mismo valor. - PARTITION BY se usa para dividir un conjunto de datos en particiones y realizar cálculos dentro de cada una sin afectar el resto de la consulta.

	Año	Entidad	Municipio	Casos_recuperados
1	2020	01	001	45529
2	2020	02	004	34375
3	2020	03	003	22595
4	2020	04	002	10911
5	2020	05	035	31581
6	2020	06	002	4630
7	2020	07	101	10226
8	2020	08	037	29138
9	2020	09	007	160481
10	2020	10	005	30106
11	2020	11	020	63425
12	2020	12	001	20236
13	2020	13	048	10397
14	2020	14	039	40553
15	2020	15	033	49263
16	2020	16	053	19338
17	2020	17	007	8480

Query executed successfully. LAPTOP-D1ML1TLP\SQLEXPRESS ... LAPTOP-D1ML1TLP\Alang ... covidHistorico 00:00:59 96 rows

Consulta 3.	Listar el porcentaje de casos confirmados en cada una de las siguientes morbilidades a nivel nacional: diabetes, obesidad e hipertensión
Requisitos:	N/A
Significado de valores de los catálogos:	- HIPERTENSION: Si tiene valor '1' es que tiene la enfermedad, de lo contrario '0' - OBESIDAD: Si tiene valor '1' es que tiene la enfermedad, de lo contrario '0' - DIABETES: Si tiene valor '1' es que tiene la enfermedad, de lo contrario '0' - CLASIFICACION_FINAL: Confirmado por asociación clínica-epidemiológica, 2 = Confirmado por dictaminación, 3 = Confirmado por laboratorio
Responsable:	Alan Olea García.
Comentarios:	- CAST convierte el resultado a un numero decimal - Case permite evaluar condiciones y devolver diferentes valores

	Porcentaje_DIABETES	Porcentaje_HIPERTENSION	Porcentaje_OBESIDAD
1	9.62	12.79	10.57

Consulta 4.	Listar los municipios que no tengan casos confirmados en todas las morbilidades: hipertensión, obesidad, diabetes, tabaquismo.
Requisitos:	N/A
Significado de valores de los catálogos:	- HIPERTENSION: Si tiene valor '1' es que tiene la enfermedad, de lo contrario '0' - OBESIDAD: Si tiene valor '1' es que tiene la enfermedad, de lo contrario '0' - DIABETES: Si tiene valor '1' es que tiene la enfermedad, de lo contrario '0' - TABAQUISMO: Si tiene valor '1' es que tiene la enfermedad, de lo contrario '0' - MUNICIPIO_RES: Nos da el municipio de residencia del paciente
Responsable:	Alan Olea García.
Comentarios:	Sin comentarios

	Entidad	Municipio
1	28	002
2	16	052
3	07	106
4	16	015
5	18	001
6	17	002
7	04	008
8	13	029
9	14	105
10	08	049
11	19	049
12	29	009
13	19	007
14	16	048
15	12	059
16	30	034

Query executed successfully. LAPTOP-D1ML1TLP\SQLEXPRESS ... LAPTOP-D1ML1TLP\Alang ... covidHistorico 00:00:04 841 rows

Consulta 5.	Listar los estados con más casos recuperados con neumonía.
Requisitos:	N/A
Significado de valores de los catálogos:	- NEUMONIA: Si tiene valor '1' es que tiene la enfermedad, de lo contrario '0' - ENTIDAD_RES: Nos da el municipio de residencia del paciente
Responsable:	Alan Olea García.
Comentarios:	-Se considero la consulta en general, no se hizo la división por año -Se hizo un join con la tabla "Cat_entidades" para que apareciera el nombre del estado

Results | Messages

	entidad	Casos_Recuperados
1	CIUDAD DE MÉXICO	46545
2	MÉXICO	45599
3	PUEBLA	16971
4	VERACRUZ DE IGNACIO DE LA LLAVE	11936
5	GUANAJUATO	11356
6	BAJA CALIFORNIA	9519
7	JALISCO	9480
8	HIDALGO	9312
9	NUEVO LEÓN	8926
10	MICHOACÁN DE OCAMPO	7562
11	CHIHUAHUA	7261
12	SONORA	7096
13	SINALOA	6952
14	QUERÉTARO	6537
15	GUERRERO	6430
16	TABASCO	5673
17	SAN LUIS POTOSÍ	5502

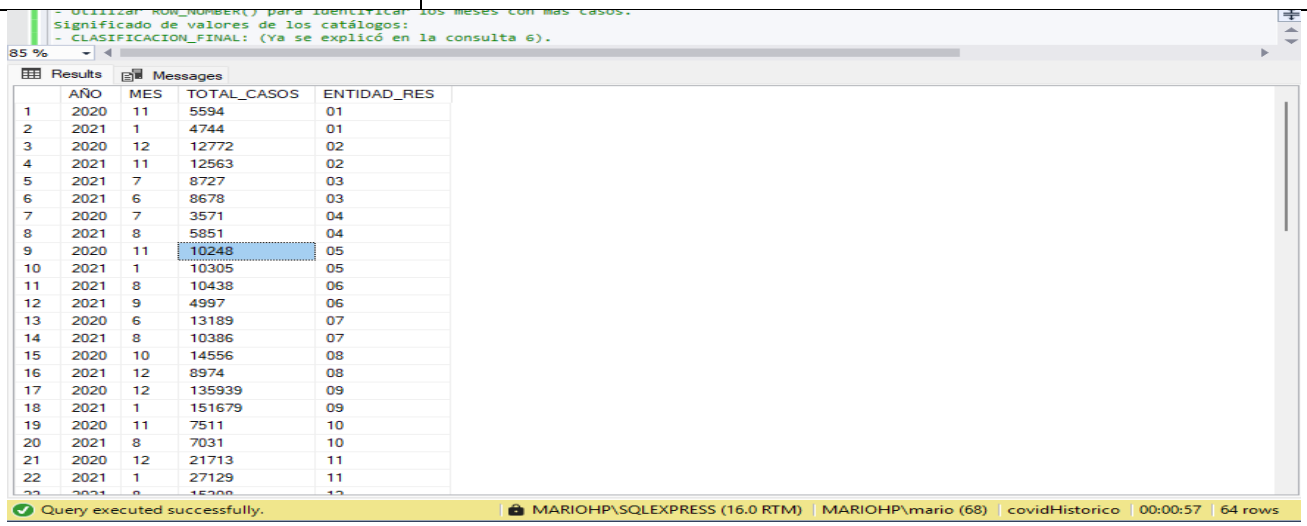
Query executed successfully. | LAPTOP-D1ML1TLP\SQLEXPRESS ... | LAPTOP-D1ML1TLP\Alang ... | covidHistorico | 00:00:05 | 32 rows

Consulta 6.-	Listar el total de casos confirmados/sospechosos por estado en cada uno de los años registrados en la base de datos
Requisitos:	- Mostrar claramente los municipios con la menor cantidad de recuperaciones registradas durante los años especificados. - Incluir solo los pacientes que fueron confirmados como casos positivos de COVID-19 con CLASIFICACION_FINAL igual a (1, 2, 3). - Identificar como paciente recuperado aquellos cuya FECHA_DEF está registrada como '9999-99-99'. - Realizar el conteo y la selección utilizando la función de ventana `ROW_NUMBER()` para ordenar los municipios según la cantidad de recuperados (de menor a mayor).
Significado de valores de los catálogos:	- ENTIDAD_RES: Código de la entidad federativa de residencia del paciente. - FECHA_INGRESO: Fecha de ingreso del paciente al servicio de salud. - CLASIFICACION_FINAL: - 1: Confirmado por asociación clínica-epidemiológica. - 2: Confirmado por dictaminación médica. - 3: Confirmado por laboratorio. - 6: Caso sospechoso
Responsable:	Mario Alexis Juarez Anguiano
Comentarios:	- Se utilizaron dos métodos: 1. **Sin subconsulta: ** Se agruparon los casos por año y estado con `SUM()` y `CASE WHEN`. 2. **Con subconsulta: ** Se asignaron valores binarios (1 o 0) antes de realizar la suma, lo que puede mejorar la claridad del código. - Ambas versiones generan los mismos resultados, permitiendo analizar los totales de casos confirmados y sospechosos por año y estado. - Se empleó `ORDER BY` para ordenar los datos cronológicamente.

	AÑO	ENTIDAD_RES	CASOS_CONFIRMADOS	CASOS_SOSPECHOSOS
1	2020	01	17999	6846
2	2020	02	37434	28550
3	2020	03	18168	921
4	2020	04	7536	3719
5	2020	05	53005	7810
6	2020	06	8025	1713
7	2020	07	8513	29168
8	2020	08	39866	2781
9	2020	09	362275	30644
10	2020	10	26390	2141
11	2020	11	86268	9328
12	2020	12	26717	1663
13	2020	13	25064	4592
14	2020	14	54952	12608
15	2020	15	161809	38411
16	2020	16	34490	1174
17	2020	17	11437	3657
18	2020	18	8016	5723
19	2020	19	85561	10514
20	2020	20	29738	2278
21	2020	21	50728	7661
22	2020	22	34335	2678
23	2020	23	16115	3588
24	2020	24	42131	2539

Query executed successfully. | MARIOHP\SQLEXPRESS (16.0 RTM) | MARIOHP\mario (68) | covidHistorico | 00:00:33 | 96 rows

Consulta 7.-	Para el año 2020 y 2021 cuál fue el mes con más casos registrados, confirmados, sospechosos, por estado registrado en la base de datos.
Requisitos:	- Determinar el mes y año con mayor cantidad de casos por entidad federativa. - Considerar solo los años 2020 y 2021. - Diferenciar casos confirmados (CLASIFICACION_FINAL 1, 2, 3) y sospechosos (CLASIFICACION_FINAL 6). - Utilizar `ROW_NUMBER()` para rankear los meses con más casos por estado
Significado de valores de los catálogos:	- ENTIDAD_RES: Código de la entidad federativa del paciente. - FECHA_INGRESO: Fecha de registro del paciente. - CLASIFICACION_FINAL: - 1: Confirmado por asociación clínica-epidemiológica. - 2: Confirmado por dictaminación médica. - 3: Confirmado por laboratorio. - 6: Caso sospechoso sin confirmación definitiva.
Responsable:	Mario Alexis Juarez Anguiano
Comentarios:	- Se utilizó un **CTE (`Casos_Por_Mes`) ** para organizar los datos y calcular el total de casos por mes y estado. - Se aplicó `ROW_NUMBER ()` para asignar un ranking a cada mes dentro de cada entidad federativa según el número de casos. - Se filtraron únicamente los dos meses con más casos en cada estado (`RANK <= 2`), facilitando la identificación de los periodos más críticos.



	AÑO	MES	TOTAL_CASOS	ENTIDAD_RES
1	2020	11	5594	01
2	2021	1	4744	01
3	2020	12	12772	02
4	2021	11	12563	02
5	2021	7	8727	03
6	2021	6	8678	03
7	2020	7	3571	04
8	2021	8	5851	04
9	2020	11	10248	05
10	2021	1	10305	05
11	2021	8	10438	06
12	2021	9	4997	06
13	2020	6	13189	07
14	2021	8	10386	07
15	2020	10	14556	08
16	2021	12	8974	08
17	2020	12	135939	09
18	2021	1	151679	09
19	2020	11	7511	10
20	2021	8	7031	10
21	2020	12	21713	11
22	2021	1	27129	11

Consulta 8.-	Identificar el municipio con menos defunciones en el mes con más casos confirmados de neumonía en 2020 y 2021
Requisitos:	- Filtrar solo casos con neumonía confirmada (`NEUMONIA = 1`). - Considerar únicamente los años 2020 y 2021. - Excluir registros sin defunción (`FECHA_DEF` ≠ '9999'). - Identificar el mes con más casos y el municipio con menos defunciones en ese periodo.
Significado de valores de los catálogos:	- ENTIDAD_RES: Código de la entidad federativa del paciente. - MUNICIPIO_RES: Código del municipio de residencia. - FECHA_INGRESO: Fecha en que el paciente ingresó al servicio de salud. -FECHA_DEF: Indica defunción ('9999' significa sin defunción). - CLASIFICACION_FINAL: - 1: Confirmado por laboratorio. - 2: Confirmado por asociación clínica-epidemiológica. - 3: Confirmado por dictaminación médica.
Responsable:	Mario Alexis Juarez Anguiano
Comentarios:	-Se creó la vista `CASOS_NEUMONIA` para simplificar la consulta principal, filtrando solo registros relevantes. - Se agruparon y contaron los casos por entidad y municipio en los meses con más casos (enero 2021 y julio 2020). - Se utilizó `UNION ALL` para combinar datos de ambos años y `LEFT` para filtrar fechas de defunción correctamente.

	ENTIDAD_RES	MUNICIPIO_RES	AÑO_INGRESO	MES_INGRESO	CASOS_DEF
1	04	001	2021	1	1
2	05	001	2021	1	1
3	08	001	2021	1	1
4	20	001	2021	1	1
5	29	001	2021	1	1
6	08	002	2021	1	1
7	12	002	2021	1	1
8	16	002	2021	1	1
9	17	002	2021	1	1
10	20	002	2021	1	1
11	22	002	2021	1	1
12	24	002	2021	1	1
13	31	002	2021	1	1
14	07	003	2021	1	1
15	22	003	2021	1	1
16	29	003	2021	1	1
17	32	003	2021	1	1

Query executed successfully. | MARIOHP\SQLEXPRESS (16.0 RTM) | MARIOHP\mario (59) | covidHistorico | 00:00:28 | 752 rows

Consulta 9.-	Listar los 3 municipios con menos pacientes recuperados en 2020 y 2021
Requisitos:	- Considerar solo pacientes recuperados (`FECHA_DEF = '9999-99-99'`). - Incluir únicamente casos confirmados (`CLASIFICACION_FINAL` en 1, 2, 3). - Agrupar los datos por municipio y entidad. - Utilizar `ROW_NUMBER()` para ordenar los municipios según la cantidad de recuperados (de menor a mayor).
Significado de valores de los catálogos:	- ENTIDAD_RES: Código de la entidad federativa del paciente. - MUNICIPIO_RES: Clave numérica del municipio de residencia. - CLASIFICACION_FINAL: - 1: Confirmado por asociación clínica-epidemiológica. - 2: Confirmado por dictaminación médica. - 3: Confirmado por laboratorio. - FECHA_DEF: - '9999-99-99': Paciente recuperado (sin defunción).
Responsable:	Mario Alexis Juarez Anguiano
Comentarios:	-Se utilizó `COUNT()` para obtener el número total de recuperados por municipio. - La función `ROW_NUMBER()` asignó un ranking de menor a mayor cantidad de recuperaciones. - Se aplicó `WHERE posicion <= 3` para seleccionar únicamente los 3 municipios con menos recuperados.

ENTIDAD_RES	MUNICIPIO_RES	totalRecuperados	AñoRegistro
20	329	1	2021
20	106	1	2021
20	228	1	2021

uery executed successfully. | MARIOHP\SQLEXPRESS (16.0 RTM) | MARIOHP\mario (55) | covidHistorico | 00:01:32 | 3 rows

Consulta 10.-	Listar el porcentaje de casos confirmado por género en los años 2020 y 2021.
Requisitos:	- Considerar solo casos confirmados (`CLASIFICACION_FINAL` en 1, 2, 3). - Calcular el porcentaje de casos confirmados por sexo (masculino, femenino, no especificado). - Comparar los datos con el total global de casos confirmados en ambos años.
Significado de valores de los catálogos:	- SEXO: - 1: Femenino. - 2: Masculino. - 99: No especificado. - CLASIFICACION_FINAL: - 1: Confirmado por asociación clínica-epidemiológica. - 2: Confirmado por dictaminación médica. - 3: Confirmado por laboratorio.
Responsable:	Mario Alexis Juarez Anguiano
Comentarios:	- Se utilizó la cláusula `WITH` para crear CTEs (`Total_CASOS` y `Casos_Sexo_Año`), facilitando la organización de datos. - `CROSS JOIN` se empleó para calcular porcentajes relativos tomando como referencia el total global de casos confirmados en ambos años. - Se usó `YEAR(FECHA_INGRESO)` para extraer el año y agrupar los datos correctamente.

	SEXO	PORCENTAJE	año	TOTAL_CASOS_SEXO
1	1	18.735495929296	2020	754215
2	2	19.094399484895	2020	768663
3	1	31.615781422567	2021	1272723
4	2	30.554323163240	2021	1229993

Query executed successfully. | MARIOHP\SQLEXPRESS (16.0 RTM) | MARIOHP\mario (55) | covidHistorico | 00:00:44 | 4 rows

Consulta 11.	Listar el porcentaje de casos hospitalizados por estado en el año 2020.
Requisitos:	N/A
Significado de valores de los catálogos:	- ENTIDAD_UM: Identifica la entidad donde se ubica la unidad medica que brindó la atención. - entidad: muestra el nombre de los estados - year(FECHA_INGRESO) = 2020: Selecciona los casos en donde el paciente ingresó al hospital en 2020. - TIPO_PACIENTE = 2: Indica los casos de pacientes que fueron hospitalizados. - FECHA_INGRESO: Indica la fecha en la que el paciente ingresó al hospital
Responsable:	Brenda Urrutia González
Comentarios:	Sin comentarios.

Results | Messages

	entidad	Porcentaje
1	AGUASCALIENTES	11.915519596511
2	BAJA CALIFORNIA	18.655975647416
3	BAJA CALIFORNIA SUR	7.006081602681
4	CAMPECHE	13.313094515280
5	CHIAPAS	10.090280321475
6	CHIHUAHUA	23.570908030899
7	CIUDAD DE MÉXICO	7.870144215062
8	COAHUILA DE ZARAG...	12.622132937084
9	COLIMA	18.668728876668
10	DURANGO	9.458528520253
11	GUANAJUATO	10.107522398328
12	GUERRERO	17.228517565019
13	HIDALGO	24.124504961509
14	JALISCO	20.850461313601
15	MÉXICO	26.342541612805
16	MICHOACÁN DE OCA...	13.868150706202

Query executed successfully. | LAPTOP-H98AEI5U\SQLEXPRESS0... | LAPTOP-H98AEI5U\Urruti... | covidHistorico | 00:00:29 | 32 rows

Consulta 12.	Listar total de casos negativos por estado en los años 2020 y 2021.
Requisitos:	N/A
Significado de valores de los catálogos:	- CLASIFICACION_FINAL=7: Muestra los casos negativos - year(DC.FECHA_INGRESO) in (2020, 2021): Selecciona los casos en donde el paciente ingresó al hospital en 2020 y 2021 - ENTIDAD_UM: Identifica la entidad donde se ubica la unidad medica que brindó la atención. - entidad: muestra el nombre de los estados
Responsable:	Brenda Urrutia González
Comentarios:	Sin comentarios.

Results | Messages

	estado	totalCasosNegativos
1	AGUASCALIENTES	96231
2	BAJA CALIFORNIA	147695
3	BAJA CALIFORNIA SUR	130296
4	CAMPECHE	47361
5	CHIAPAS	52389
6	CHIHUAHUA	89231
7	CIUDAD DE MÉXICO	3397281
8	COAHUILA DE ZARAGOZA	187244
9	COLIMA	45156

Query executed successfully. | LAPTOP-H98AEI5U\SQLEXPRESS0... | LAPTOP-H98AEI5U\Urruti... | covidHistorico | 00:00:09 | 32 rows

Consulta 13.	Listar porcentajes de casos confirmados por género en el rango de edades de 20 a 30 años, de 31 a 40 años, de 41 a 50 años, de 51 a 60 años y mayores a 60 años a nivel nacional.
Requisitos:	N/A
Significado de valores de los catálogos:	- SEXO = 1: Selecciona a las pacientes mujeres - SEXO = 2: Selecciona a los pacientes hombres - CLASIFICACION_FINAL in (1,2,3): Selecciona a todos los tipos de casos confirmados
Responsable:	Brenda Urrutia González
Comentarios:	Sin comentarios.

Results | Messages

	SEXO	rangoEdades	total	Porcentaje
1	1	20-30	694483	12.455735769085
2	1	31-40	660480	11.845882996078
3	1	41-50	577196	10.352162490619
4	1	51-60	401082	7.193511452024
5	1	Mayores de 60	340465	6.106329569797
6	2	20-30	634886	11.386947855875
7	2	31-40	594364	10.660075098536
8	2	41-50	498442	8.939688729910
9	2	51-60	363255	6.515074230469
10	2	Mayores de 60	364169	6.531467061529

Query executed successfully. | LAPTOP-H98AEI5U\SQLEXPRESS0... | LAPTOP-H98AEI5U\Urruti... | covidHistorico | 00:00:11 | 10 rows

Consulta 14.	Listar el rango de edad con más casos confirmados y que fallecieron en los años 2020 y 2021.
Requisitos:	N/A
Significado de valores de los catálogos:	- CLASIFICACION_FINAL in (1,2,3): Selecciona a todos los tipos de casos confirmados - FECHA_DEF!='9999-99-99': Indica los pacientes que fallecieron - year(FECHA_INGRESO) in (2020, 2021): Muestra a los pacientes que ingresaron al hospital en los años 2020 y 2021
Responsable:	Brenda Urrutia González
Comentarios:	- Se utiliza "case when..." para definir los rangos de edades

Results | Messages

	rangoEdad
1	Mayores a 60

Query executed successfully. | LAPTOP-H98AEI5U\SQLEXPRESS0... | LAPTOP-H98AEI5U\Urruti... | covidHistorico | 00:00:05 | 1 rows

Conclusión

En esta práctica analizamos información relacionada con el COVID-19 en México, aplicando los conocimientos adquiridos en clase y complementándolos con investigación adicional. Durante el desarrollo de la actividad, reforzamos el uso de consultas SQL, filtrado de datos, agregaciones y funciones avanzadas para extraer información relevante sobre la pandemia.

Además, enfrentamos desafíos que nos obligaron a buscar soluciones más allá de lo visto en clase, lo que nos permitió mejorar nuestras habilidades de análisis y manipulación de datos, usando diferentes estrategias tales como:

- Entender bien todos los componentes de la tabla y analizar el enunciado. Se utilizó "with" en varias consultas, haciendo más fácil entender la lógica de lo que se estaba haciendo y además ayudó a evitar la repetición de código.
- Otra de las estrategias utilizadas fue el ir haciendo el código por partes, comenzar por lo más sencillo, visualizar los resultados y posteriormente ir creciendo la consulta hasta obtener los valores correctos.
  En algunas consultas se utilizó el “select case when…” para poder definir rangos de valores de forma más concisa, también se usa “sum” con “case” para clasificar los casos.
- Otra estrategia fue hacer uso de consultas temporales, esto principalmente en consultas complejas para precisamente poder simplificarlas.

Este ejercicio nos ayudó a comprender mejor la importancia de las bases de datos en la toma de decisiones y en el estudio de eventos de gran impacto, como lo fue la pandemia.