

PONTIFICIA UNIVERSIDAD CATÓLICA MADRE Y MAESTRA
Departamento de Ciencias e Ingeniería. Escuela de Ingeniería
Telemática y Sistemas



Base de datos II

ISC 349 - 101

Actividad 3

Presentado por:

Nicol Ureña (2018-1669)
Junior Hernández (2018-0999)
Luis Angel Ramírez (2018-0570)

Presentado a:

Maximo Emiliano Perez Medrano

Asignatura:

Base de datos II

SANTIAGO DE LOS CABALLEROS
REPÚBLICA DOMINICANA

Actividad #3

1. Hagan todas las observaciones/comentarios/peculiaridades posibles que encuentren al dataset de las muestras (archivo de mas de 2.7 millones de registros del site anteriormente compartido).

Aunque haya casos confirmados de COVID hay quienes tienen la columna fecha de síntomas vacíos.

En nacionalidad hay personas quienes tienen 214 por lo cual supones que las que no tienen ese número son dominicanos y los que sí se puede referir a que es un posible inmigrante.

Se presente espacios donde el género no está colocado en varias personas.

Hay filas donde, aunque no este confirmado tiene datos en fecha de confirmado.

Había algunas peculiaridades en cuanto a la integridad de los datos en varias columnas, como cuando decía que no estaba fallecido pero tenía fecha de defunción, etc.

En general, muchas filas vacías que deberían tener información.

Es un dataset realmente grande, que evidentemente debe manejarse con precaución porque podría afectar el funcionamiento de aplicaciones.

2. Investiguen e identifiquen la composición de la estructura del archivo de muestras de covid del dataset de muestras, es decir, cuales son las columnas o atributos que componen el archivo, sus tipos de datos y tamaños.

El archivo está compuesto por una serie de columnas en una única hoja de Excel, las columnas son las siguientes:

OBJECTID(int), EDAD_A(int), SEXO(nvarchar(255)), Nacionalidad(int), CONFIRMADO((nvarchar(255)), FECHA_CONFIRM(date), ESTADO(nvarchar(255)), ACTIVO(nvarchar(255)), FALLECIDO(nvarchar(255)), FECHA_DEF(date), RECUPERADO(nvarchar(255)), FECHA_RECUP(datetime), PROV_COD(nvarchar), MUN_COD(nvarchar(255)), TRAB_SNS(nvarchar(255)), EMBARAZADA(nvarchar(255)), UCI(nvarchar(255)), VENT_MEC(nvarchar(255)), SINTOMAS(nvarchar(255)), FECHA_SÍNTOMAS(date), FECHA_INICIO(date), FECHA_FIN(date)

3. Carguen el archivo anterior a una BD.

Archivo adjuntado

4. Integren datos que consideren de las provincias e integren a la BD. Vinculen la tabla con los datos de las provincias a la tabla de las muestras

Completado en la BD

5. Creen una vista donde se tenga un resumen de cada mes de cada año por provincia con los totales de cantidad de muestras, cantidad de positivos y cantidad de negativos. Incluir código y nombre de la provincia.

```
CREATE VIEW resumen_mensual as
SELECT DATENAME(month, FECHA_INICIO) as Mes, YEAR(FECHA_INICIO) as
Año, count(CONFIRMADO) as [Cantidad de muestras], count(CASE WHEN
CONFIRMADO = 'Si' THEN 1 ELSE NULL END) as [Cantidad de positivos], count(CASE
WHEN CONFIRMADO = 'No' THEN 1 ELSE NULL END) as [Cantidad de negativos],
PROV_COD as Provincia
FROM Base_de_datos_COVID
GROUP BY DATENAME(month, FECHA_INICIO), YEAR(FECHA_INICIO),
PROV_COD
```

GO

Con un SELECT * FROM resumen_mensual se obtiene lo siguiente:

Mes	Año	Cantidad de muestras	Cantidad de positivos	Cantidad de negativos	Provincia
April	2020	1	0	1	LA VEGA
April	2021	1	0	1	AZUA
April	2021	16	5	11	DISTRITO NACIONAL
April	2021	2	0	2	LA ALTAGRACIA
April	2021	1	0	1	LA ROMANA
April	2021	1	0	1	SAN CRISTOBAL
April	2021	6	2	4	SANTO DOMINGO
August	2020	4	0	4	DUARTE
August	2020	1	0	1	SANTIAGO
August	2021	1	0	1	DISTRITO NACIONAL
August	2021	1	0	1	SANTO DOMINGO
December	2020	8	2	6	DISTRITO NACIONAL
December	2020	1	0	1	LA ALTAGRACIA
December	2020	1	0	1	SANTIAGO
December	2020	2	1	1	SANTO DOMINGO
February	2021	1	0	1	AZUA

6. Creen una vista que determine las cantidades de muestras, positivos y negativos para grupos etareos (rangos de edades definidos por ustedes).

```
CREATE VIEW grupos_etareos as
SELECT SUM(CASE WHEN EDAD_A < 18 THEN 1 ELSE 0 END) AS [Menores de 18],
SUM(CASE WHEN EDAD_A < 18 AND CONFIRMADO = 'Si' THEN 1 ELSE 0 END)
AS [Menores de 18 Positivos],
SUM(CASE WHEN EDAD_A < 18 AND CONFIRMADO = 'No' THEN 1 ELSE 0 END)
AS [Menores de 18 Negativos],
SUM(CASE WHEN EDAD_A BETWEEN 18 AND 24 THEN 1 ELSE 0 END) AS [18-24],
SUM(CASE WHEN EDAD_A BETWEEN 18 AND 24 AND CONFIRMADO = 'Si' THEN
1 ELSE 0 END) AS [18-24 Positivos],
SUM(CASE WHEN EDAD_A BETWEEN 18 AND 24 AND CONFIRMADO = 'No'
THEN 1 ELSE 0 END) AS [18-24 Negativos],
SUM(CASE WHEN EDAD_A BETWEEN 25 AND 34 THEN 1 ELSE 0 END) AS [25-34],
SUM(CASE WHEN EDAD_A BETWEEN 25 AND 34 AND CONFIRMADO = 'Si' THEN
1 ELSE 0 END) AS [25-34 Positivos],
SUM(CASE WHEN EDAD_A BETWEEN 25 AND 34 AND CONFIRMADO = 'No'
THEN 1 ELSE 0 END) AS [25-34 Negativos]
FROM Base_de_datos_COVID

GO
```

Con un SELECT * FROM grupos_etareos se obtiene lo siguiente:

Menores de 18	Menores de 18 Positivos	Menores de 18 Negativos	18-24	18-24 Positivos	18-24 Negativos	25-34	25-34 Positivos	25-34 Negativos
38	5	33	18	4	14	51	4	47

Reflexión general de la actividad.

Utilizando los conocimientos adquiridos en prácticas anteriores, en primer lugar, se descargó del sitio web referenciado por el profesor un documento Excel, que debido a su extensión se redujo para así realizar esta actividad. Los datos posteriormente fueron importados desde el Excel, con algunas columnas previamente modificados para que se cumplan los requisitos para la manipulación de datos (como la columna de las fechas en la que se quitaron los 0 del final, entre otros casos).

Del mismo modo se adjuntan algunas instrucciones en el **anexo** para garantizar el uso correcto de las tablas generadas. Ya en el documento de Excel adjunto, se puede ver la estructura que tomó nuestra base de datos, con las columnas mencionadas en un punto anterior. Esta actividad mejora el conocimiento sobre los objetos de la base de datos, como las vistas y sus operaciones. También hay diferentes puntos de vista sobre cómo se almacenan los datos en diferentes plataformas. Los datos pueden proporcionarnos una cantidad sorprendente de información y, gracias a ellos, somos capaces de observar y pensar en diferentes instrucciones posibles. En conclusión, dentro de los aprendizajes mencionados anteriormente y otros más que se obtuvieron a lo largo de esta actividad, se destaca lo propenso a errores que son las recopilaciones de información realizadas por personas, sin embargo, estos pueden ser controlados y arreglados si se tiene un buen conocimiento de bases de datos, donde la integridad de los datos se toma siempre en consideración.

Anexos

```
alter table dbo.base_de_datos_covid_reducida  
alter column FECHA_SÍNTOMAS datetime
```

```
alter table dbo.base_de_datos_covid_reducida  
alter column FECHA_INICIO datetime
```

```
alter table dbo.base_de_datos_covid_reducida  
alter column FECHA_FIN datetime;
```

```
alter table dbo.base_de_datos_covid_reducida  
alter column FECHA_RECUP datetime;
```

```
alter table dbo.base_de_datos_covid_reducida  
alter column FECHA_DEF datetime;
```

```
alter table dbo.base_de_datos_covid_reducida  
alter column FECHA_CONFIRM datetime;
```