**Tarea Recuperativa**

**Peso: 30%**

**MIEMBROS EQUIPO 6:**

* **Laura Tatiana Lopez Sanchez**
* **Juan Bernardo Rivera García**
* **Sebastián Fonnegra Galeano**

**Instrucciones**

* Revise bien los archivos de la tarea.
* Renombre el formato Word cambiando la letra “N” por el número del grupo o el nombre del estudiante si hace una entrega individual.
* Documentos a entregar: Informe en Word, la Hoja de Cálculo con resultados, archivos con script de Creación, Poblamiento y Consultas de las tablas de la base de datos.
* Coloque los documentos en el repositorio y comparta el link con el profesor.
* Leer atentamente el enunciado y cada uno de los ítems.

**Enunciado del problema**

El Ministerio de Educación desea realizar un conjunto de análisis en relación a los graduados en los diferentes departamentos, municipios, instituciones privadas y públicas, entre otros. Le han contratado a Ud. para que realice el análisis solicitado y construya la base de datos para almacenar y consultar esta información. Para este propósito, le han suministrado una (1) hoja de cálculo con la información detallada de los graduados del año 2021. Esta actividad requerirá de todos los conocimientos que ha adquirido en la Universidad, y en este Curso en particular, para resolver los diferentes problemas que se le presentan.

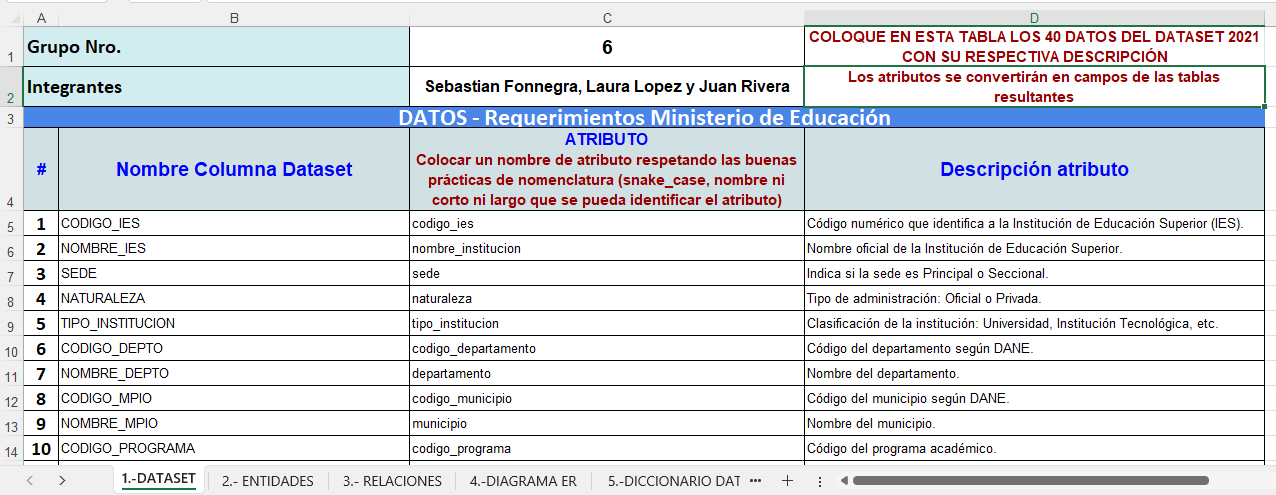
**Fuente de los datos:** Ministerio de Educación Superior de Colombia

**Contenido:** Graduados de instituciones universitarias de pregrados y postgrados en los departamentos de Colombia en el año 2021.

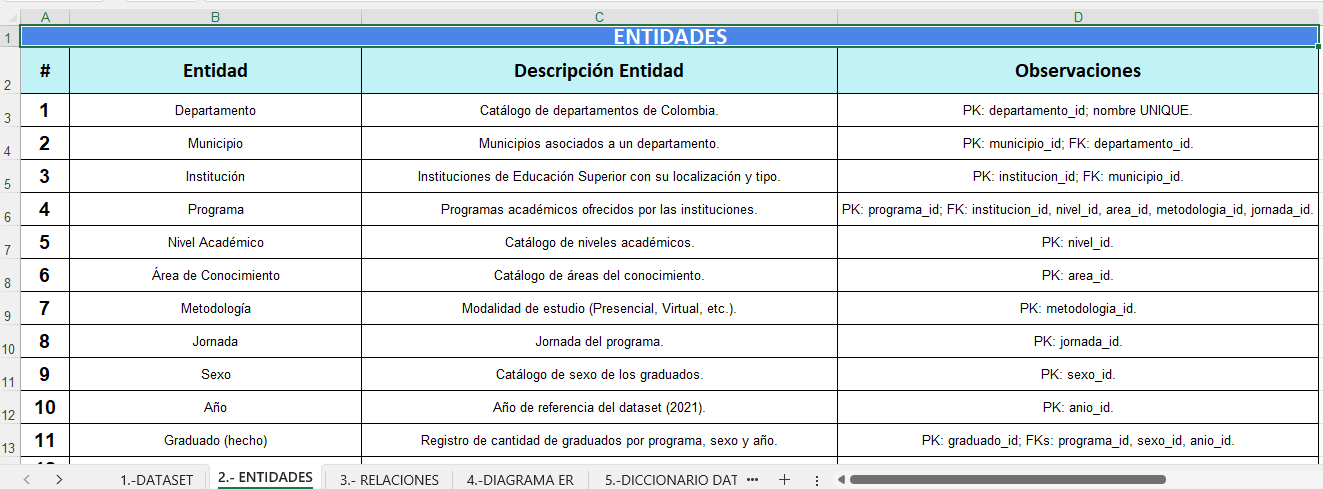
**ACTIVIDADES**:

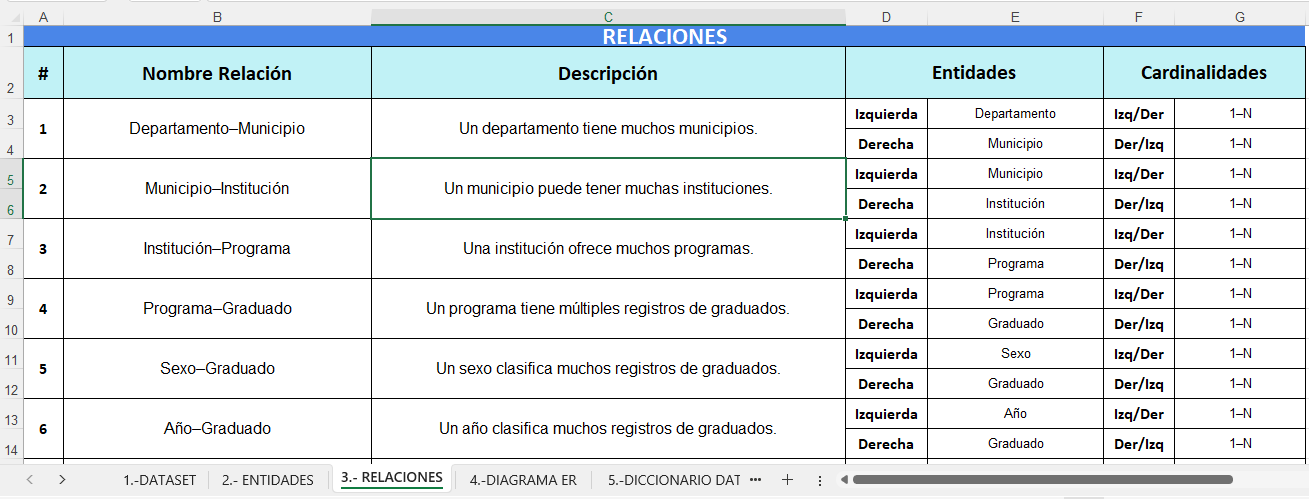
El Ministerio le solicita cierta información para el periodo 2021. Todos los resultados se refieren a total de graduados. Para producir los resultados requeridos se deben realizar las siguientes actividades:

**1.- Analizar y describir el Dataset**. Debe estudiar e identificar bien los datos. Esto es fundamental antes de iniciar cualquier tipo de actividad de construcción de la Base de Datos. Colocar los datos del Dataset, en la forma requerida, en la Hoja de Cálculo en la pestaña **“1.-DATASET”**. En esta sección coloque solamente UN PANTALLAZO de la pestaña.

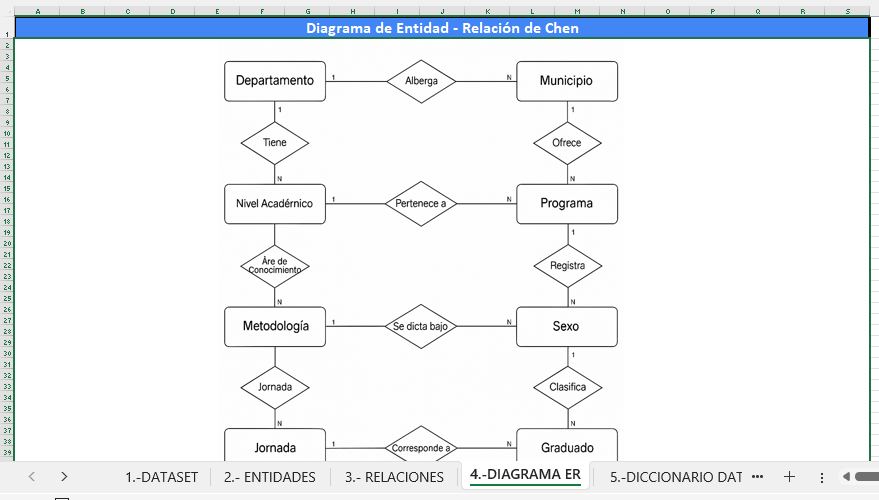


**2.- Elaborar el Inventario de Entidades y Relaciones**. Debe identificar el primer conjunto de entidades que sean evidentes y las relaciones respectivas entre ellas. NOTA: Colocar los datos en la Hoja de Cálculo en las pestañas **“2.-ENTIDADES**” y **“3.-RELACIONES”** respectivamente. En esta sección coloque solamente UN PANTALLAZO de cada pestaña.





**3.- Elaborar un diagrama Entidad-Relación de Chen** que represente las entidades, relaciones y cardinalidades. No debe colocar los atributos en el Diagrama. NOTA: Colocar los datos en la Hoja de Cálculo en la pestaña **“4.-DIAGRAMA ER”** . En esta sección coloque solamente UN PANTALLAZO del Diagrama.



**4.- Realizar un proceso de normalización (1FN,2FN,3FN) sobre el Dataset**. Debe determinar todas las tablas que encontró con sus campos respectivos. NOTA: Recuerde las dependencias funcionales y transitivas. Debe solamente describir el proceso brevemente, NO DEBEN PROPONER TABLAS NI EN la 2FN ni en la 3FN. Las tablas las colocarán directamente en el Diccionario de Datos. Eso les ahorrará mucho trabajo.

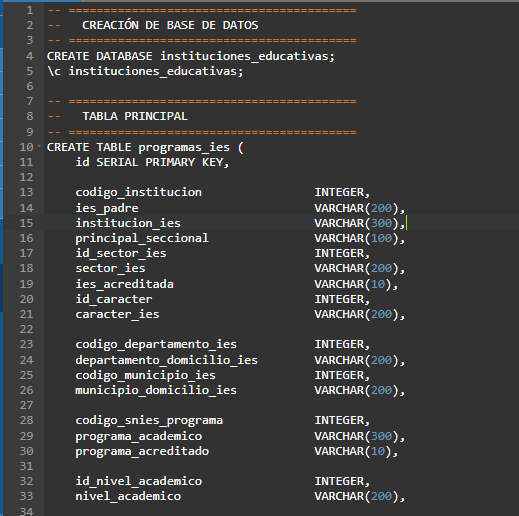
Como lo dice el proceso mismo, la normalización se trata de descomponer el Dataset en tablas mas pequeñas para realizar una serie de análisis y evaluaciones para buscar la integridad de los datos y que no se presenten redundancias y datos repetidos.

Aunque como lo pidió la actividad solo se uso el 1FN, donde nos centramos en encontrar y eliminar datos repetidos y en identificar las PK y FK.

**5.- Elaborar el Diccionario de datos Físico.**  Primero debe elaborar un inventario de las tablas y después las tablas en detalle con sus campos resultado de la 3FN. NOTA: Colocar los datos en la Hoja de Cálculo en la pestaña **“5.-DICCIONARIO DE DATOS FÍSICO”**. En esta sección coloque solamente UN PANTALLAZO de la pestaña.



**6.- Construir una base de datos en PostgreSQL (DDL CREATE)** con las tablas y campos del Diccionario Físico. Debe crear los scripts de creación de las tablas. NOTA: Colocar los resultados en un archivo con nombre ***“script\_creacion.sql”.***  En esta sección coloque solamente UN PANTALLAZO de una instrucción ejecutada en el SGBD.

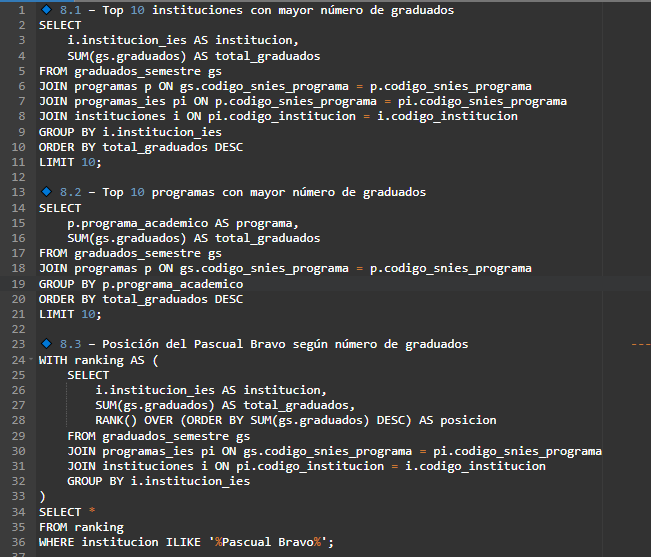


**7.- Poblar la base de datos con la información del dataset (DML INSERT)**. Puede hacerloa través de un algoritmo de programación en Python, Java o con un SCRIPT de SQL Inserts o cualquier otra solución que encuentre más rápida y efectiva. (incluir el programa o los scripts aquí). **NOTA:** Debe entregar los productos que desarrollo como evidencia del proceso. NOTA: Colocar los resultados en un archivo con nombre ***“script\_poblamiento.sql***”. En esta sección coloque solamente UN PANTALLAZO de una instrucción ejecutada en el SGBD.

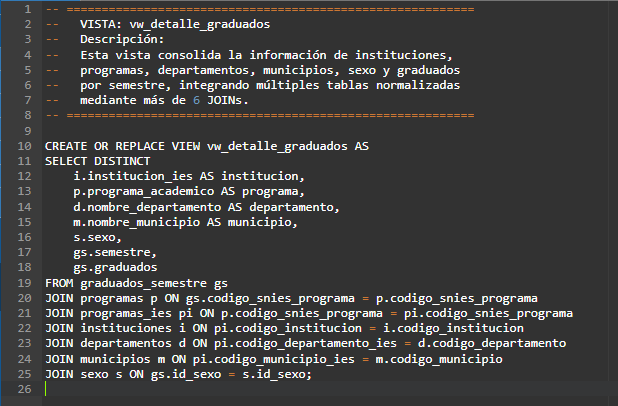
COLOCAR LOS PANTALLAZOS AQUÍ SEGÚN LO SOLICITEN

**8.- Construir el conjunto de consultas SQL propuesto por el docente (DML SELECT).** Las consultas deben tener las descripciones de la entidad solicitada (Institución Universitario, Programa, Departamento, entre otros). No colocar los códigos (claves primarias). En todos los listados debe mostrar la cantidad de graduados. Debe colocar todas las consultas en el Script de consultas ***“script\_consultas.sql”***

|  |  |
| --- | --- |
| **#** | **Consultas** |
| **8.1** | Listado de las 10 primeras instituciones con mayor número de graduados |
| **8.2** | ¿Listado de los 10 programas académicos con mayor número de graduados? |
| **8.3** | ¿En qué posición se encuentra el Pascual Bravo en número de graduados entre todas las instituciones? NOTA: Debe agrupar el número de graduados por Institución y ordenarlo de mayor a menor por número de graduados. |
| **8.4** | Listado de los 10 programas académicos con mayor número de graduados hombres |
| **8.5** | Listado de los 10 programas académicos con mayor número de graduados mujeres |
| **8.6** | Listado de los 3 departamentos con mayor número de graduados |
| **8.7** | Listado de los 3 municipios con mayor número de graduados. NOTA: Debe aparecer el departamento |
| **8.8** | Listado completo de Instituciones ordenado alfabéticamente. NOTA: Debe incluir la descripción de Institución, departamento, municipio y total graduados |
| **8.9** | Listado completo de Programas ordenado alfabéticamente. NOTA: Debe incluir la descripción del programa, la Institución, departamento, municipio y graduados |
| **8.10** | Listado de graduados por Departamento, Municipio y número de graduados de mayor a menor |



**9. Construir una VIEW.** La vista de incluir 6 JOINS cómo mínimo y es de libre decisión. Explicar de que se trata la vista. Colocar la vista en un archivo con nombre **“script\_vista.sql”**. En esta sección coloque solamente UN PANTALLAZO de la instrucción ejecutada en el SGBD.



**10.- Conclusiones Individuales**. Redacte sus conclusiones sobre la experiencia del Curso de Base de Datos I, los conocimientos aprendidos, su vinculación al futuro ejercicio profesional una vez que se gradúen, entre otros. NOTA: Coloquen aquí las conclusiones individuales de cada miembro del grupo.

La actividad permitió diseñar, crear y organizar una base de datos totalmente normalizada a partir de un conjunto extenso de información académica. Se estructuraron las tablas en el orden correcto, se aplicaron versiones, modificaciones DDL, índices, validaciones y renombres, garantizando integridad y calidad de los datos. Finalmente, se desarrollaron consultas analíticas y vistas que integran múltiples tablas, facilitando el análisis de graduados por institución, programa y ubicación. El proyecto demuestra un buen manejo de SQL, modelamiento de datos y organización de información para análisis académico. **(Sebastián Fonnegra)**

La realización de este proyecto permitió transformar un conjunto de datos brutos sobre graduados en Colombia en un modelo de datos relacional, normalizado y eficiente, apto para su almacenamiento y consulta estructurada.

**1.** Diseño Optimizado: Mediante el análisis del Dataset, la elaboración de un Inventario de Entidades y Relaciones (Diagrama E-R) y la aplicación de un proceso de normalización, se logró un diseño físico de la base de datos que garantiza la integridad de los datos y minimiza la redundancia, sentando las bases para un sistema

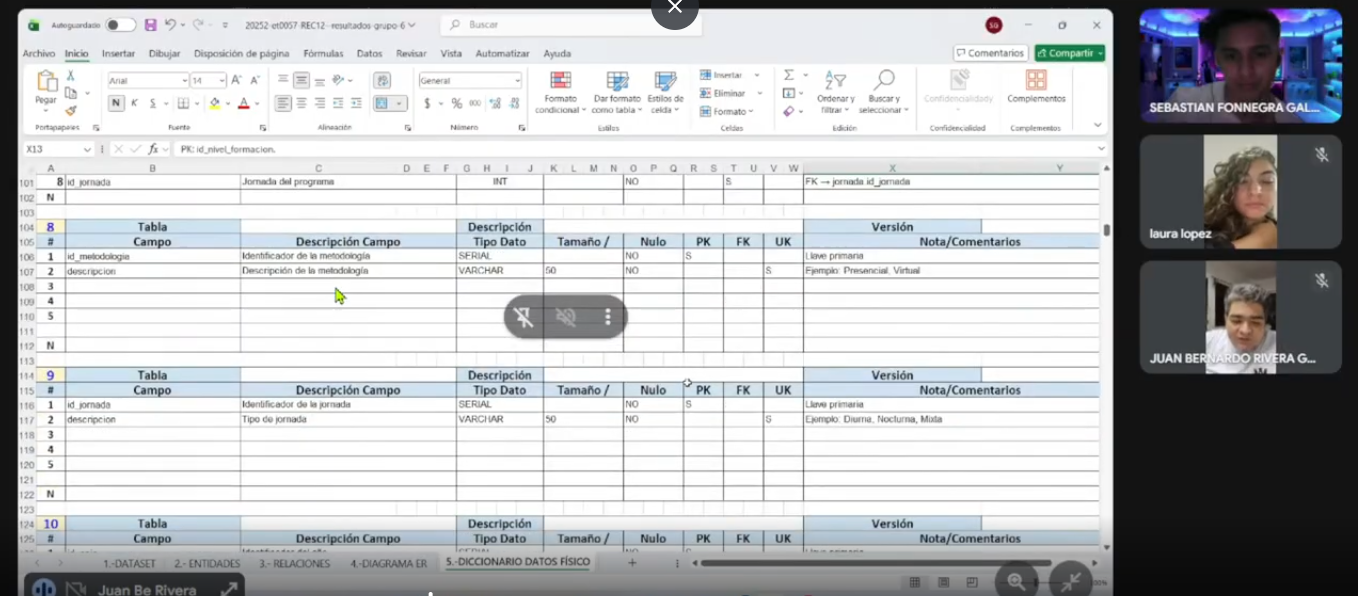
de información sólido.

**2.** Implementación Funcional: La implementación del diseño lógico en la herramienta de gestión de bases de datos (PostgreSQL) y la posterior carga de la información (DDL y DML) demostraron la viabilidad del modelo.

**3.** Utilidad Estratégica: El sistema resultante es una plataforma robusta que no solo almacena la información de manera organizada, sino que también facilita la ejecución de consultas complejas (DML SELECT) para el análisis detallado de los graduados por departamento, programa, tipo de institución y otras variables, cumpliendo así con el objetivo inicial del Ministerio de Educación de contar con una herramienta para la toma de decisiones informadas. **(Laura Lopez).**

En la realización de esta actividad, la enseñanza que nos deja es la importancia que podemos adquirir nosotros como programadores en el manejo de una gran cantidad de información, y bajo nuestro análisis y organización, mediante las herramientas propias de cada programa realizar una serie de operaciones como creación, consulta y modificación de datos, darle respuesta a las tareas que nos sean asignados o las solicitudes según su necesidad que se le presente a un cliente y así se realice un excelente trabajo buscando siempre integridad y exactitud en los resultados. **(Juan Rivera).**

**11.- Video de Sustentación**. Elabore un video de sustentación de la tarea. Cada integrante debe aparecer (que se vea su rostro) y debe presentarse con nombre y apellido; y explicar su contribución al proyecto. NOTA: **Se debe MOSTRAR CÓDIGO EN EJECUCIÓN** (sección de creación de las tablas, poblamiento y consultas). Coloque aquí un pantallazo del Video (dónde aparezcan todos los miembros del grupo) y el LINK.



**LINK DEL VIDEO:**

[**https://drive.google.com/file/d/17f\_JMPTM\_3pFXbn5mD3U\_rREjmvZ2coI/view?usp=sharing**](https://drive.google.com/file/d/17f_JMPTM_3pFXbn5mD3U_rREjmvZ2coI/view?usp=sharing)

**12.- Repositorio GIT.** Incluir la tarea recuperativa en el Repositorio GIT. Debe contener todos los entregables: Informe (esta plantilla), Hoja de cálculo con resultados y el grupo de archivos con los scripts. **NO COLOQUE EL VIDEO EN EL RESPOSITORIO SOLAMENTE el LINK**.

**13.- Informe**. El informe es ESTA MISMA plantilla, pero con los resultados solicitados. Se tendrá en cuenta la calidad de presentación del informe: todo bien identificado, organizado, los pantallazos bien centrados y descritos brevemente en el pie, también las tablas, la ortografía, la claridad en la exposición de las ideas, entre otros. NOTA: La evaluación del informe contempla también el correcto nombramiento de los archivos entregables y el video.

**NOTA FINAL: NO AGREGUE PESTAÑAS NUEVAS. REALICÉ SOLAMENTE LO SOLICITADO**

**Rúbrica: Criterios de Evaluación**

**ATENCIÓN: Esta rúbrica contiene la evaluación de 2 Tareas Recuperativas. El video de sustentación es OBLIGATORIO para evaluar el Recuperativo #2.**

|  |  |  |  |
| --- | --- | --- | --- |
| **ítem** | **Criterio** | **Ponderación** | **Calificación** |
| **RECUPERATIVO #1 (10%)** | | | |
| **1** | **Analizar y describir el Dataset** | **5** |  |
| **2** | **Elaborar el Inventario de Entidades y Relaciones** | **15** |  |
| **3** | **Elaborar un diagrama Entidad-Relación de Chen** | **25** |  |
| **4** | **Realizar un proceso de normalización (1FN,2FN,3FN) sobre el Dataset** | **10** |  |
| **5** | **Elaborar el Diccionario de datos Físico** | **25** |  |
| **6** | **Construir una base de datos en PostgreSQL (DDL CREATE)** | **20** |  |
| **TOTAL** | | **100** |  |
| **RECUPERATIVO #2 (20%)** | | | |
| **7** | **Poblar la base de datos con la información del dataset (DML INSERT** | **25** |  |
| **8** | **Construir el conjunto de consultas SQL propuesto por el docente (DML SELECT).** | **20** |  |
| **9** | **Construir una VIEW** | **5** |  |
| **10** | **Conclusiones Individuales** | **5** |  |
| **11** | **Video de Sustentación** | **30** |  |
| **12** | **Repositorio GIT** | **5** |  |
| **13** | **Informe** | **10** |  |
| **TOTAL** | | **100** |  |