

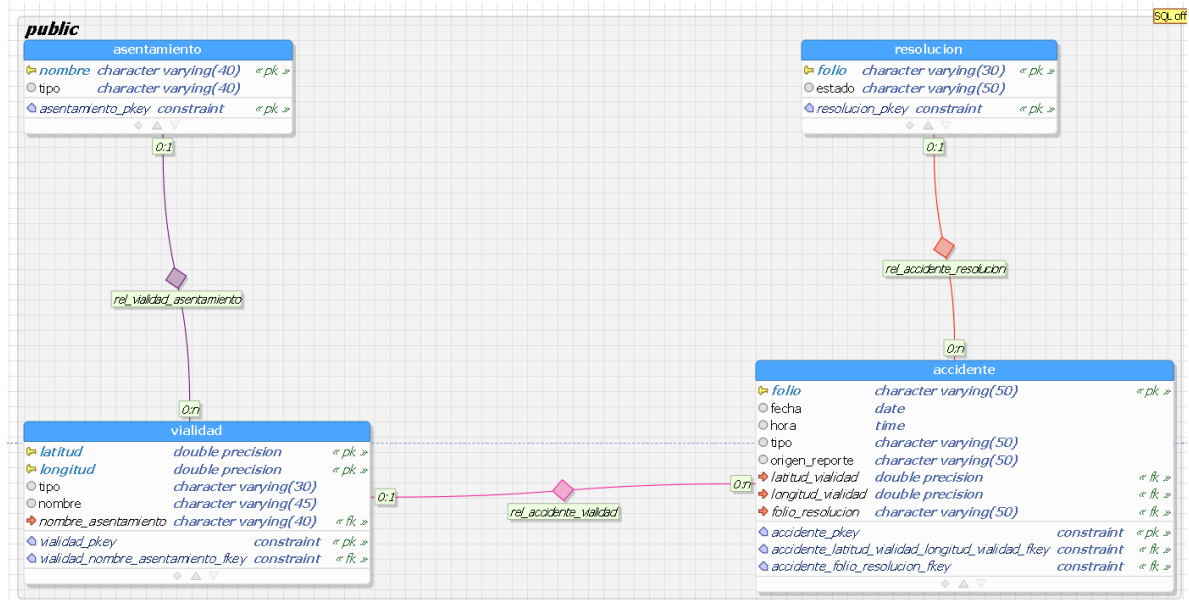
Universidad del Rosario

Proyecto final - Ingeniería de datos

Juan Camilo Pugliese y Juan Andrés Castro
semestre 2023-1

Entrega 2

1. Modelo relacional normalizado en tercera forma normal.



2. Descripción del proceso de carga de información en la base de datos.

- Para la carga masiva de datos, primero se normalizó la base de datos, eliminando valores nulos, errores en los registros que no correspondían con el tipo de dato, y columnas repetidas. Después se usó la función concatenar, con la cual se creó el script para cada tabla con sus respectivos valores, y se corrió sobre el ejecutador de pgAdmin4, sin botar ningún tipo de error, tras haber conseguido llenar las tablas con la información de la base de datos normalizada, se procedió a crear el archivo tanto con el DML, como con el DDL en el archivo script_BD_SQL, en adición, se subió también el archivo indice_accidentes_viales en formato .CSV y .XLS para no llegar a perder ningún tipo de información. Todo el resultado se encuentra en el repositorio de GitHub, Base_accidentes_viales.

3. Descripción de análisis identificados.

Encontramos 8 formas de ver la base de datos encontrando 8 posibles escenarios de análisis.

- Podemos observar la cantidad y los detalles de los accidentes automovilísticos que ocurren durante un período de tiempo para ver si hay algún patrón.
- Podríamos notar que los accidentes ocurren con más frecuencia en ciertos momentos del día, la semana o el mes.
- Podemos observar todos los diferentes tipos de accidentes que ocurren y ver cuáles ocurren con mayor frecuencia.

- d. También podemos ver si hay ciertos patrones o tendencias con estos accidentes, como si ocurren más en las ciudades o en el campo.
- e. Podemos ver dónde ocurren los accidentes en un mapa y averiguar qué áreas tienen la mayor cantidad de accidentes. Esto puede ayudarnos a hacer planes para mantener a las personas seguras en las carreteras.
- f. Podemos ver dónde las personas reportan accidentes (como si le dicen a la policía o a otra persona) para ver qué tan bien está funcionando el reporte. Esto puede ayudarnos a determinar si hay algún problema con los informes.
- g. Podemos mirar una tabla para ver cuánto tiempo se tarda en reparar accidentes en la carretera y averiguar si hay algún patrón en la forma en que se solucionan en función de lo que sucedió. Esto nos ayuda a asegurarnos de que estamos arreglando los accidentes de la manera más rápida y eficiente posible.
- h. Podemos ver dónde vive la gente y dónde están las carreteras para averiguar qué carreteras tienen más accidentes y qué áreas son las más afectadas. Entonces podemos trabajar para hacer esos caminos más seguros y mejores para prevenir accidentes.
- i. Podemos ver diferentes cosas sobre los accidentes, como de qué tipo son o cuándo ocurrieron, y ver si están conectados entre sí. Por ejemplo, podríamos descubrir que algunos accidentes ocurren con mayor frecuencia cuando llueve afuera.
- j. Podemos ver cuántos accidentes automovilísticos ocurren en diferentes lugares para ver si algunos lugares tienen más accidentes que otros. Podríamos comparar ciudades con áreas rurales o diferentes estados para ver si hay grandes diferencias.
- k. Hay muchas maneras de ver la información sobre accidentes automovilísticos. La forma en que lo mires depende de lo que quieras averiguar.

4. URL del repositorio GIT del proyecto.

- a. https://github.com/juanandrescastroc/Base_accidentes_viales