

TRABAJO PRÁCTICO UNIDAD 2

ESCENARIO 1: DISEÑO DE UN ALMACÉN DE DATOS

Objetivo. Construir un almacén de datos multidimensional a partir de una fuente de información contenida en un archivo electrónico de una hoja de cálculo.

Descripción del Escenario:

La construcción del almacén de datos deberá seguir la metodología o enfoque KDD y realizar todas las etapas necesarias, condicionado a respetar uno de los siguientes formatos:

- a) Fuente de Información-Base de Datos-Almacén de Datos,
- b) Fuente de Información-Base de Datos-Base de Datos Temporal-Almacén de Datos.

El diseñador del almacén de datos analizará los datos contenidos en la fuente de información o en la base de datos para identificar y seleccionar las columnas requeridas para la construcción del almacén de datos.

Algunas de las dimensiones requeridas en el diseño del almacén de datos son Comunidad o Vecindario, Uso de Internet, Tiempo, Datos Demográficos, Infraestructura Tecnológica.

El diseño del almacén de datos permitirá visualizar la siguiente información o conocimiento.

- 1) Tiempo promedio de uso de Internet por comunidad y por rango de edad.
- 2) Infraestructura utilizada para conectarse a Internet por comunidad.
- 3) Usuarios que más utilizan Internet considerando el ingreso anual, el idioma que se habla en casa, edad, y comunidad.
- 4) Los dos principales dispositivos que utilizan los usuarios de internet dependiendo del nivel de educación y la edad.
- 5) Determinar los usuarios que utilizan Internet considerando género, edad, nivel de estudios y comunidad.

Restricciones:

1. De la fuente de información se pueden eliminar 15 registros, como máximo. La decisión se puede tomar en los casos de que no se tenga suficiente información para solucionar el problema de datos nulos o vacíos.
2. De la fuente de información se pueden eliminar 4 columnas (máximo), antes de cargar los datos en el almacén de datos.

Evaluación:

1. La primera revisión de la solución del escenario se deberá presentar el día 8 de Septiembre de 2020 a las 18:00 horas en sitio.
2. La segunda revisión de la solución del escenario se deberá entregar en formato digital el día 10 de Septiembre de 2020, en horario de 15:00 a 18:00 mediante la herramienta Microsoft Teams.