

**Asignatura**

Datawarehosue y minería de fatos

**Nombre de la actividad**

Proyecto – Fase 1

**Integrantes**

Jordan Ismael Zelaya Ramírez - ZR170168

**Docente**

Ing. Karens Medrano.

Soyapango, 18 de Abril del 2022

Para la realización y análisis de los documentos brindados, se hizo uso de un ETL realizado en Visual Studio, con el fin de pasar los datos de los archivos .csv a base de datos.

![Interfaz de usuario gráfica

Descripción generada automáticamente]()

Como se puede comprobar en la imagen anterior, la conversión de datos fue exitosa. Para esto se crearon dos bases de datos, una denominada EsquelasDB y la otra ParqueVehicularDB, y dentro de estas las tablas normalizadas correspondientes:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamenteInterfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Comenzaré el análisis primero con el archivo de Esquelas. Para este se realizó en primera instancia un datawarehouse en forma de estrella, con el finde luego crear un cubo OLAP. Debido a esto, se crearon en la base de datos EsquelasDB, otras tablas dimensionales, las cuales son:

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Como se puede observar, se insertaron los atributos correspondientes mediante un SELECT DISTINCT. Una vez los datos estaban llenos en las tablas dimensionales, se hizo un UPDATE a las columnas id\_falta, id\_departamento e id\_valor de la tabla Esquelas\_Normalizada:

Texto

Descripción generada automáticamente

Acto seguido se creó la tabla de hechos para las Esquelas:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Con todo esto, se pasaron los datos de la tabla Esquelas\_Normalizada a la tabla Fact\_Esquelas.

Esta sería la estructura del cubo OLAP:

Captura de pantalla de computadora

Descripción generada automáticamente

Visualizaremos los resultados del cubo en un archivo de Excel, agregando como filas los id Departamento y Estado; la cantidad de Esquelas en Valores y en filtro el id valor, obteniendo que el departamento con mayores esquelas de $171.43 es aquel cuyo id es el 8, es decir Sonsonate, teniendo faltas de tipo “transporte carga”, siendo 25 en total, a diferencia del resto de Departamentos. Podemos decir entonces que este territorio es mucho más propenso al resto a cometer este tipo de infracciones:

Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente

Cabe mencionar que los resultados obtenidos pueden variar según el enfoque se quiera dar, en este caso como haremos 4 análisis, se procurará hacer un enfoque similar en cada uno, con el fin de mostrar la veracidad de lo mencionado previamente. Seguiremos con el análisis de tipo Decision Tree:

Imagen que contiene Texto

Descripción generada automáticamenteImagen que contiene Diagrama

Descripción generada automáticamente

Debido a la gran cantidad de datos, se utilizó el operador Apply Model, para obtener los siguientes datos:

Interfaz de usuario gráfica, Tabla

Descripción generada automáticamente

Comprobamos que los datos obtenidos con el Cubo son los mismos que en el DT. Sin embargo, también obtenemos los siguientes resultados: San Salvador es el top offender en las esquelas $57.14, $11.43 y $32.29.

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente

Ahora pasamos a los análisis de la base de datos ParqueDB. Para este se comenzó con un análisis K Means tomando una muestra de 100,000 rows, ya que mi computadora no soportaba más, se usó específicamente este query:

Diagrama

Descripción generada automáticamenteInterfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Los resultados obtenidos fueron:

Interfaz de usuario gráfica

Descripción generada automáticamente

Se puede observar que los valores promedio más altos son los del Cluster 2, donde la cilindrada es de 435.408, el valor del vehículo 16039.56 y el valor de importación de 15888.050. Acto seguido se realizó un análisis de Reglas de asociación:

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Se obtuvieron 117 reglas de asociación, entre ellas estas las que tienen mayor confianza y convicción:

Tabla

Descripción generada automáticamente

Uso de Git y Github:

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

Link: [JordanZR/DMD-Proyecto-fase-1: ETL + Datawarehouse + OLAP Cube + Decision tree + K Means + Association Rules (github.com)](https://github.com/JordanZR/DMD-Proyecto-fase-1)