**CASO 1: VENTAS DE BICICLETAS**

**Instrucciones**

REGLAS (Leer atentamente)

1. La prueba es en PAREJAS.

2. La entrega será el viernes 28 de junio, 2019 hasta las 12:30 hrs.

3. Se deberá responder el correo enviado por el profesor y deberá adjuntar las evidencias comprimidas en un solo archivo. El archivo comprimido deberá contener:

* Base de datos DW\_BICICLETAS con registros (DW\_BICICLETAS.bak)
* Proyecto ETL de Ventas en Integration Services.
* Documento Word con las capturas de pantalla con el testing a PANTALLA COMPLETA.

4. El nombre del proyecto en Integration Services debe ser: Caso1 ETL Ventas

5. El formato del archivo comprimido debe ser el siguiente:

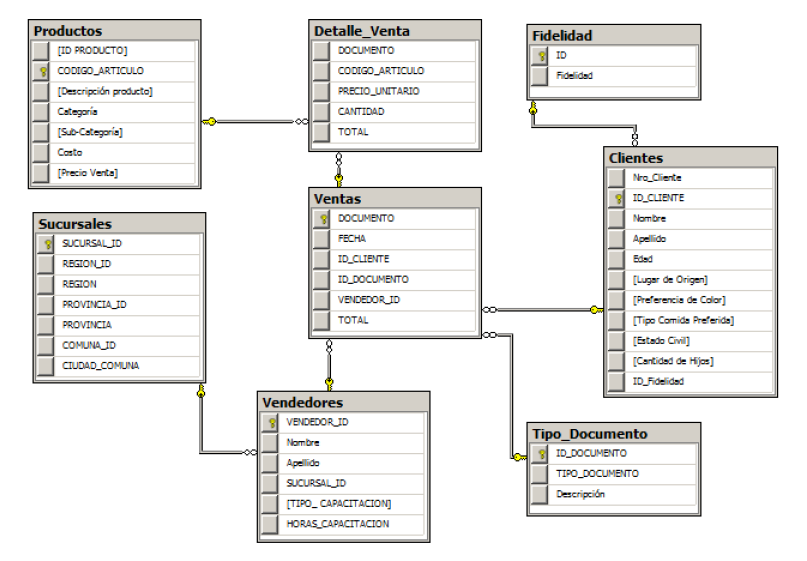
“Prueba1-ApellidoAlumno1\_NombreAlumno1- ApellidoAlumno2\_NombreAlumno2-MBI7501.ZIP”

Ejemplo

Prueba1-RomeroAlonso\_CarvajalPedro-MBI7501.zip

**Nota**: Puede ser .rar también

***MODELO DE DATOS SISTEMA DE VENTA DE BICICLETAS***



MODELO DIMENSIONAL DATA WAREHOUSE



**Para el Caso 1:**

**1. Restaurar la base de datos transaccional BICICLETA perteneciente al sistema de ventas (OLTP)**

**2. Crear un Data Warehouse según el modelo entregado con los siguientes requerimientos:**

* Crear todas las tablas de dimensiones.
* Crear la tabla de hechos. Cuando crear la tabla de hechos queda a su criterio.
* La información referente a la fecha se solicita en el siguiente formato: MES en palabas en español (Ejemplo: Noviembre) y el DÍA en palabras en español (Ejemplo: lunes). El AÑO en números (Ejemplo: 2018).
* La tabla de hechos contendrá el total por ventas por vendedor y artículo, además del promedio por vendedor y artículo.
* Función para la suma: SUM ()
* Función para promedio: AVERAGE ()
* La tabla de hechos debe ser borrada por completo antes de ser cargada con registros.
* Las dimensiones deben ser cargadas en modalidad full e incremental.

**3. Crear un ETL capaz de:**

* Traspasar todos los registros desde la base de datos transaccional al Data Warehouse (FULL)
* Traspasar sólo los registros que hayan sufrido algún cambio, desde la base de datos transaccional al Data Warehouse (Incremental)
* Traspasar los registros nuevos desde la base de datos transaccional al Data Warehouse (Incremental)

**4. Testing**

* Carga Full desde la base de datos OLTP al DW. (Capturar pantalla)
* Carga de 5 registros modificados desde la base de datos OLTP al DW. (Capturar pantalla)
* Carga de 5 registros nuevos desde la base de datos OLTP al DW. (Capturar pantalla)

**5. Entregables:**

* Base de Datos DW (Con registros)
* Proyecto: Integration Services (Funcionando)
* Capturas de pantalla de las evidencias del Testing, en un archivo Word.

**COMPRIMA TODO LO ANTERIOR SEGÚN LAS RECLAS ENTREGAS Y RESPONDA EL CORREO DEL PROFESOR CON DICHO ARCHIVO COMPRIMIDO.**

**CASO N°2 VENTAS DE CAFÉ**

**Instrucciones**

Lea atentamente cada una de las actividades a desarrollar antes de comenzar a realizarlas.

**SE PIDE:**

1. Restaurar la base de datos dimensional (DWCAFE.BAK)

2. En el Data Warehouse se piden los siguientes requerimientos [5 puntos]:

Crear un cubo (CUBO VENTAS CAFÉ IA1IA2).

* Agregar tabla de hechos.
* Agregar tablas de dimensiones.
* Agregar relaciones entre las tablas.
* Agregar todos los campos según corresponda al cubo.

3. Generar los siguientes reportes en Excel

a. Reporte de ventas mensuales por año y por rango de cantidad de hijos [5 puntos].

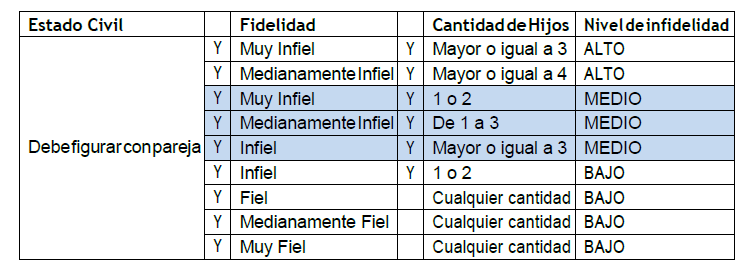
* Rango “A”: El cliente tiene un estado civil Unión Civil o Casado o Viudo y tiene de 0 a 2 hijos.
* Rango “B”: El cliente tiene un estado civil Unión Civil o Casado o Viudo y tiene de 3 o más hijos.
* Rango “C”: El cliente tiene un estado civil Divorciado y tiene 0 a 2 hijos.
* Rango “D”: El cliente tiene un estado civil Divorciado y tiene 3 o más hijos
* Rango “E”: El cliente tiene un estado civil Soltero y tiene 0 a 2 hijos.
* Rango “F”: El cliente tiene un estado civil Soltero y tiene 3 o más.

Debe agregar un campo calculado en la dimensión cliente

b. Genere un reporte en Excel con el KPI para el nivel de infidelidad por cliente y sucursal por año [20 puntos].

Valor -> Nivel infidelidad

El nivel de infidelidad se calcula según la siguiente tabla:



El estado estará definido por lo siguiente:

* Verde: ALTO
* Amarillo: MEDIO
* Rojo: BAJO

c. Genere un reporte en Excel con el KPI para el margen porcentual anual por mes y por categoría de artículo [15 puntos].

Valor -> Fórmula Margen % = 1 - (costo venta - precio venta) Las metas por categoría de artículo son las siguientes:

* Café de Filtro 45%
* Chocolate 25%
* Expreso 40%
* Frappuccino 35%
* Infusiones 20%

El estado estará definido por lo siguiente:

* Verde: Si Valor/Meta >= 0.7
* Amarillo: Si Valor/Meta >=0.35 y Valor/Meta <0.7
* Rojo: Si Valor/Meta <0.35

**CASO N°3 MINERÍA DE DATOS**

1. El archivo comprimido deberá contener:

* Proyecto MS SQL Analysis Services.
* Reporte en Excel de clientes con más probabilidad de compra
* Reporte en Excel con el listado de 17.000 prospectos
* Base de datos utilizada (Respaldo TV.bak).
* Archivo Word con las pantallas de evidencias

1. Debe utilizar la base de datos TV
2. El nombre del proyecto en Analysis Services debe ser: P3 DM TV IA1-IA2 Donde IA1 son las iniciales del alumno 1 e IA2, son las iniciales del alumno 2.
3. El formato del archivo comprimido debe ser el siguiente:

“ApellidoAlumno1\_NombreAlumno1- ApellidoAlumno2\_NombreAlumno2-P3-MBI7501.ZIP”

**1. PREDICCIÓN DE COMPRA MEDIANTE DATAMINING**

La empresa “TV Ya!” proveedora de contenidos por internet, ha decidido traer una nueva serie de acción “Justice League Next Level”, pero lamentablemente dado su alto costo ha decidido analizar primero si sus clientes verían dicha serie. La empresa ya cuenta con una base de datos donde están los datos más relevantes de sus clientes y preferencias en cuanto a lo que consumen. Dado el anterior escenario realice lo siguiente para poder obtener el listado de los clientes con mayor probabilidad de ver la nueva serie de acción.

**REQUERIMIENTOS**

* 1. A la tabla de prospectos de clientes agregue 7000 prospectos más en el mismo formato de datos.

EVIDENCIAS:

a) Tabla Prospectos con 17.000 registros

b) Archivo Excel con los 17.000 registros

1.2 Genere el modelo de minería de datos a partir de la base de datos, para cada uno de los siguientes algoritmos, con un 55% de la muestra de datos:

* Árboles de decisión
* Clústeres
* Bayes Naive

**EVIDENCIAS:**

c) Pantalla de la tabla de escenarios seleccionada

d) Pantalla de los campos seleccionados (llave, ingresos, campo predicción, campos visualizados)

e) Pantalla del porcentaje a utilizar para la muestra de datos

f) Pantalla de cada algoritmo implementado y los campos asignados

**Nota**: Una sola estructura de minería de datos con los 3 algoritmos

1.3 Grafique los 3 algoritmos en un solo gráfico.

**EVIDENCIAS**:

a) Pantalla del gráfico con los 3 algoritmos y sus respectivos porcentajes

b) Comentario de cuál es el mejor algoritmo y porqué

1.4 Aplique una predicción sobre el campo CONSUMEACCION con el mejor algoritmo, donde la predicción será sobre una tabla ProspectosClientes. En estos prospectos debe venir el correo y el número de teléfono de contacto (agregar campo).

**EVIDENCIAS:**

a) Pantalla con la implementación del reporte en Analisys Services (Asociación entre el modelo y la tabla de prospectos)

b) Tabla ListadoClientesMasProbables (Debe generarla)

1.5 Genere el informe de predicción de minería de datos en un Excel, utilizando el mejor algoritmo (Listado de los 200 clientes con mayores probabilidades de compra, de 60% hacia arriba).

**EVIDENCIAS:**

a) Archivo Excel con el listado de los 200 clientes con mayor probabilidad de compra.

Con los datos de contacto y la probabilidad de compra en formato % con 2 decimales.

1.6 Genere la consulta MDX con los clientes con mayor probabilidad de que NO comprarán el producto (Desde 60% hacia arriba).

**EVIDENCIAS:**

b) Script de la consulta MDX en el archivo Word de evidencias.

c) Reporte Excel con el listado de prospectos con mayor probabilidad de que NO comprarán el producto (Desde 60%)

d) Genere el código DMX que realice la misma predicción para que el programador lo integre en el código.