

*Facultad de Ciencias Exactas y
Naturales y Agrimensura*

BASE DE DATOS 2



TRABAJOS PRACTICOS 5, 6, 7 Y 8

INTEGRANTES GRUPO N° 1:

- Kairuz, David Elias DNI: 34973195
- Leiva Falcón, Ayelen DNI: 41281526
- Maldonado Gauna, Esteban Joel DNI: 40350454
- Piriz, Andrés Nahuel DNI: 40586477

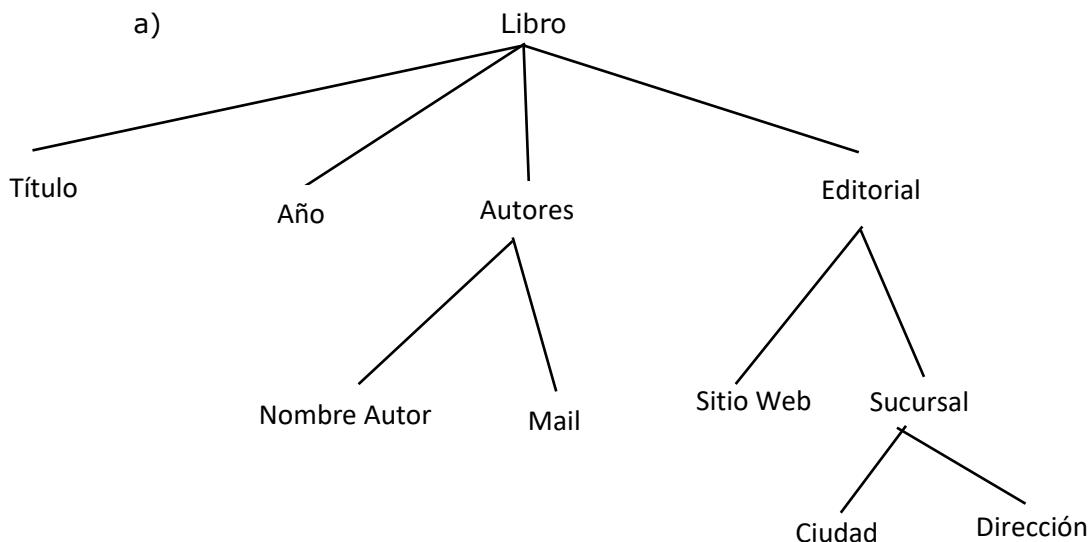
2022

Serie Ejercicios Prácticos 5
Bases de Datos Relacionales Extendidas

Modelo relacional anidado

1) Libro

Título	Año	Autores		Sitio web	Editorial	
		Nombre Autor	Mail		Sucursal	
					Ciudad	Dirección



b) Libro = (Título, Año, Autores, Editorial)

Autores = (Nombre autor, Mail)

Editorial = (Sitio web, Sucursal)

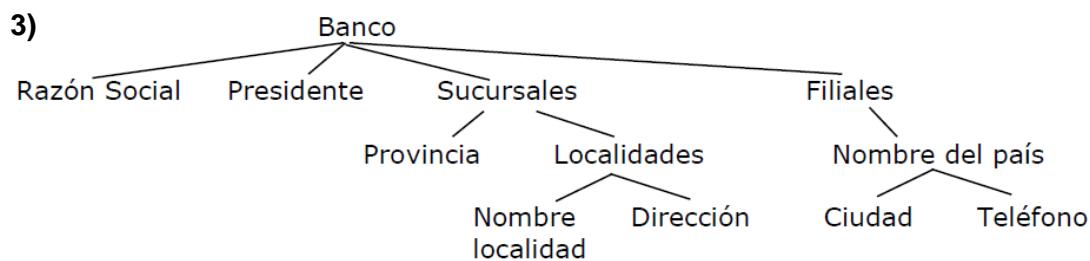
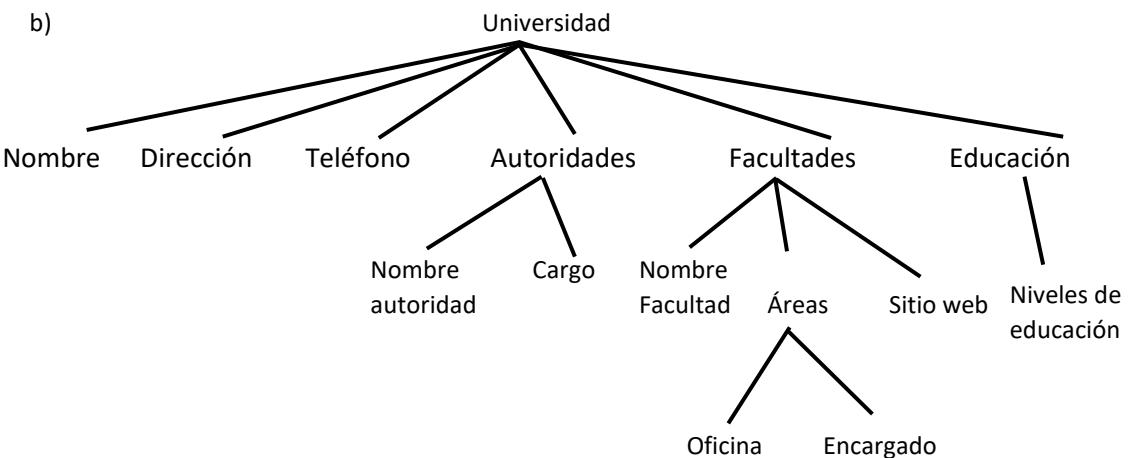
Sucursal = (Ciudad, Dirección)

2)

Universidad = (Nombre, Dirección, Teléfono, Autoridades, Facultades, Educación)
 Autoridades = (Nombre autoridad, cargo)
 Facultades = (Nombre facultad, Áreas, Sitio web)
 Educación = (Niveles de educación)
 Áreas = (Oficina, Encargado)

a) Universidad:

Nombre	Dirección	Teléfono	Autoridades		Nombre Facultad	Facultades			Sitio web	Niveles de educación			
			Nombre autoridad	Cargo		Áreas							
						Oficina	Encargado						



a) Banco = (Razón Social, Presidente, Sucursales, Filiales)

Sucursales = (Provincia, Localidades)

Localidades = (Nombre localidad, Dirección)

Filiales = (Nombre del país)

Nombre del país = (Ciudad, Teléfono)

b) Banco:

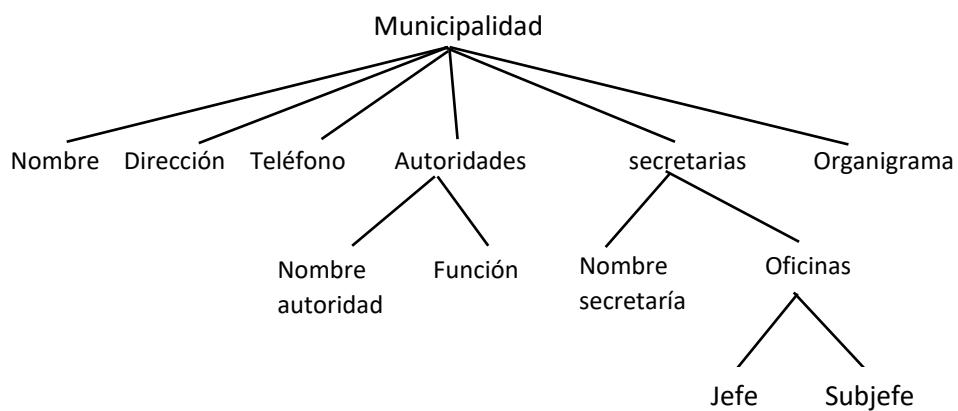
Razón Social	Presidente	Sucursales		Filiales		
		Provincia	Localidades		Nombre del país	
			Nombre localidad	Dirección	Ciudad	Teléfono

- 4)
- Municipalidad = (Nombre, Dirección, Teléfono, Autoridades, Secretarías, Organigrama)
 Autoridades = (Nombre autoridad, Función)
 Secretarías = (Nombre secretaría, Oficinas)
 Oficinas = (Jefe, Subjefe)

a) Municipalidad:

Nombre	Dirección	Teléfono	Autoridades		Secretarías		Organigrama
			Nombre autoridad	Función	Nombre secretaría	Oficinas	
						Jefe	

b)

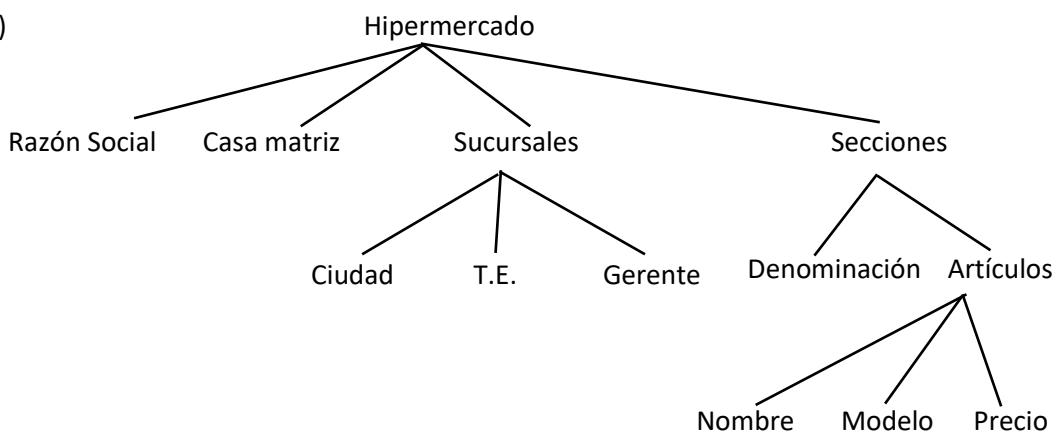


5)

Hipermercado

Razón Social	Casa matriz	Sucursales			Secciones		
		Ciudad	T.E.	Gerente	Denominación	Artículos	
					Nombre	Modelo	Precio

a)



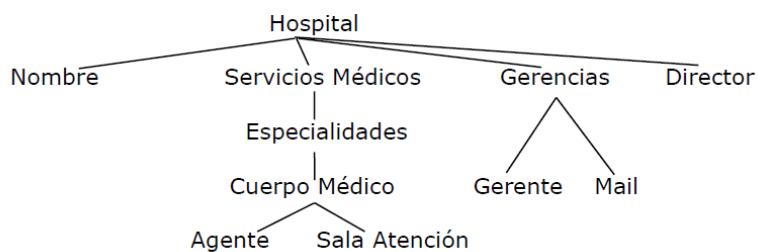
b) Hipermercado = (Razón Social, Casa matriz, Sucursales, Secciones)

Sucursales = (Ciudad, T.E., Gerente)

Secciones = (Denominación, Artículos)

Artículos = (Nombre, Modelo, Precio)

6)



a) Hospital= (Nombre, Servicios Médicos, Gerencias, Director)

Servicios Médicos = (Especialidades)

Especialidades = (Cuerpo Médico)

Cuerpo Médico = (Agente, Sala atención)

Gerencias = (Gerente, Mail)

b) Hospital:

Nombre	Servicios Médicos		Gerencias		Director				
	Especialidades		Gerente	Mail					
	Cuerpo Médico								
	Agente	Sala atención							

7)

Chevrolet = (Gerente, Concesionarias, Vehículos, Servicios)

Concesionarias = (Ciudad, Cod. Postal, Nombre concesionaria)

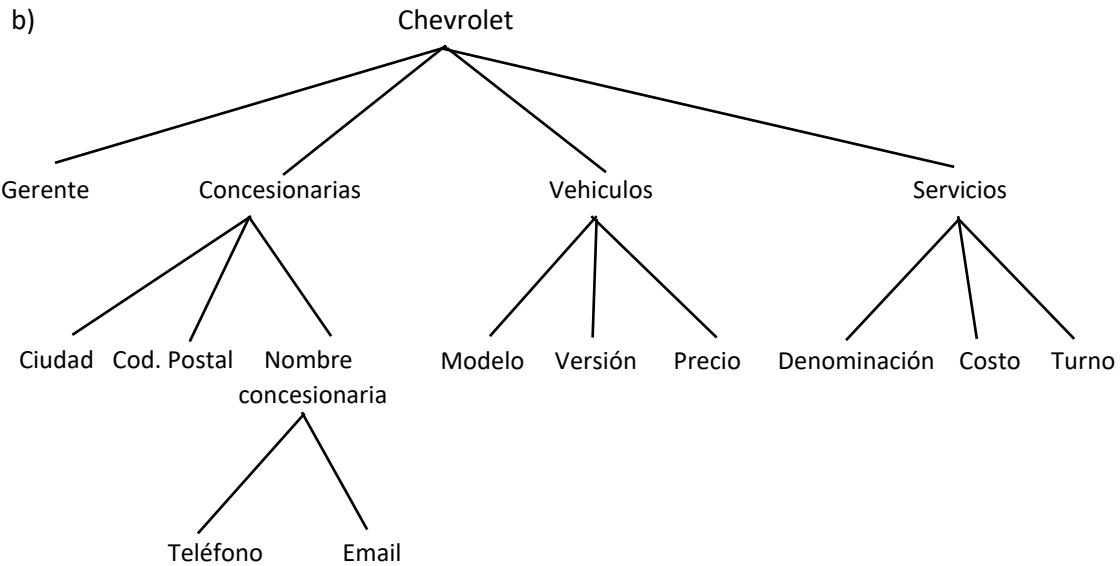
Nombre concesionaria = (Teléfono, Email)

Vehículos = (Modelo, Versión, Precio)

Servicios = (Denominación, Costo, Turno)

a)

Gerente	Concesionarias				Vehículos			Servicios		
	Ciudad	Cod.	Nombre concesionaria		Modelo	Versión	Precio	Denominación	Costo	Turno
			Teléfono	Email						



Base datos temporales

Esquema de base de datos de tiempo válido

8) Empleados

Apellido	Dni	Sueldo	Cargo
Vidal	11122233	80000	10
Pruyas	22117789	25500	15
Reyes	32190784	39700	19

- a) Base de datos con el esquema de tiempo válido:

Apellido	Dni	Sueldo	Cargo	Tiv	Tfv
Vidal	11122233	80000	10	01-06-2019	Ahora
Pruyas	22117789	25500	15	20-05-2020	Ahora
Reyes	32190784	39700	19	01-02-2017	30-09-2020

- b) Actualizar el sueldo del empleado Vidal a 92000, que será efectivo a partir del día 01-09-2020, representar la modificación:

Apellido	Dni	Sueldo	Cargo	Tiv	Tfv
Vidal	11122233	80000	10	01-06-2019	31-08-2020
Vidal	11122233	92000	10	01-09-2020	Ahora

- c) Modificar el cargo de Pruyas a 20, que será efectivo a partir del día 31-08-2020:

Apellido	Dni	Sueldo	Cargo	Tiv	Tfv
Pruyas	22117789	25500	15	20-05-2020	30-08-2020
Pruyas	22117789	25500	20	31-08-2020	Ahora

- d) Dar de baja al empleado de apellido Pruyas Dni 22117789, que deja de prestar servicios a partir del día 31-03-2021:

Apellido	Dni	Sueldo	Cargo	Tiv	Tfv
Pruyas	22117789	25500	20	31-08-2020	31-03-2021

- e) Incorporar la siguiente tupla que corresponde al empleado de apellido Méndez, que comienza a prestar servicios desde el día 14-09-2020:

Apellido	Dni	Sueldo	Cargo
Méndez	32076123	58500	12

Apellido	Dni	Sueldo	Cargo	Tiv	Tfv
Méndez	32076123	58500	12	14-09-2020	Ahora

Empleados:

Apellido	Dni	Sueldo	Cargo	Tiv	Tfv
Vidal	11122233	80000	10	01-06-2019	31-08-2020
Vidal	11122233	92000	10	01-09-2020	Ahora
Pruyas	22117789	25500	15	20-05-2020	30-08-2020
Pruyas	22117789	25500	20	31-08-2020	31-03-2021
Reyes	32190784	39700	19	01-02-2017	30-09-2020
Méndez	32076123	58500	12	14-09-2020	Ahora

Esquema de base de datos bitemporal

9)

Empleados

Apellido	Dni	Sueldo	Tiv (Vst)	Tfv (Vet)	Tit (Tst)	Tft (Tet)
Pérez	11122233	48000	15-06-2020	Ahora	08-06-2020 13:05:33	Uc
Campos	22117789	37000	20-08-2019	Ahora	20-08-2019 11:18:54	07-01-2020 14:33:25
Campos	22117789	51000	20-08-2019	31-01-2020	07-01-2020 14:33:25	Uc
Campos	22117789	65000	01-02-2020	Ahora	07-01-2020 14:33:25	Uc
Torres	36876321	39800	01-05-2020	Ahora	27-04-2020 16:22:17	Uc

- a) Actualizar el sueldo del empleado Pérez a 63500, que será efectivo a partir del día 01-10-2020, representar la modificación, siendo el tiempo de actualización de la transacción '14-09-2020 10:30:46':

Apellido	Dni	Sueldo	Tiv (Vst)	Tfv (Vet)	Tit (Tst)	Tft (Tet)
Pérez	11122233	48000	15-06-2020	Ahora	08-06-2020 13:05:33	14-09-2020 10:30:46
Pérez	11122233	48000	15-06-2020	30-09-2020	14-09-2020 10:30:46	Uc
Pérez	11122233	63500	01-10-2020	Ahora	14-09-2020 10:30:46	Uc

- b) Eliminar el empleado Torres, que deja de prestar servicios el día 30-09-2020, siendo el tiempo de actualización de la transacción '13-09-2020 12:25:31':

Apellido	Dni	Sueldo	Tiv (Vst)	Tfv (Vet)	Tit (Tst)	Tft (Tet)
Torres	36876321	39800	01-05-2020	Ahora	27-04-2020 16:22:17	13-09-2020 12:25:31
Torres	36876321	39800	01-05-2020	30-09-2020	13-09-2020 12:25:31	Uc

- c) Insertar el siguiente empleado, que comienza a prestar servicios desde el día de la fecha (considerar la fecha actual para Tiv), siendo la marca de tiempo de actualización de la transacción también la fecha actual y hora actual (expresada con el formato dd-mm-aaaa hh-mm-ss), esta fecha corresponde al día de desarrollo de este práctico el 14-09-2020 y una hora a elección.

Apellido	Dni	Sueldo
Ríos	24785930	71200

Consideramos para el tiempo válido inicial **Tiv = 19-09-2022** y para el tiempo inicio transacción

Tit = 19-09-2022 16:22:38:

Apellido	Dni	Sueldo	Tiv (Vst)	Tfv (Vet)	Tit (Tst)	Tft (Tet)
Ríos	24785930	71200	19-09-2022	Ahora	19-09-2022 16:22:38	Uc

Empleados

Apellido	Dni	Sueldo	Tiv (Vst)	Tfv (Vet)	Tit (Tst)	Tft (Tet)
Pérez	11122233	48000	15-06-2020	Ahora	08-06-2020 13:05:33	14-09-2020 10:30:46
Pérez	11122233	48000	15-06-2020	30-09-2020	14-09-2020 10:30:46	Uc
Pérez	11122233	63500	01-10-2020	Ahora	14-09-2020 10:30:46	Uc
Campos	22117789	37000	20-08-2019	Ahora	20-08-2019 11:18:54	07-01-2020 14:33:25
Campos	22117789	51000	20-08-2019	31-01-2020	07-01-2020 14:33:25	Uc
Campos	22117789	65000	01-02-2020	Ahora	07-01-2020 14:33:25	Uc
Torres	36876321	39800	01-05-2020	Ahora	27-04-2020 16:22:17	13-09-2020 12:25:31
Torres	36876321	39800	01-05-2020	30-09-2020	13-09-2020 12:25:31	Uc
Ríos	24785930	71200	19-09-2022	Ahora	19-09-2022 16:22:38	Uc

10) Vehículos

Denominación	Versión	Precio	Año
Cruze	MT	915000	2019
Focus	Se Plus	950000	2019
Etios	XLS	1055000	2020
Civic	EXL	442000	2017
Suran	Feline	516000	2018

- a) Realizar los cambios necesarios para representar la base de datos con el esquema de tiempo válido:

Denominación	Versión	Precio	Año	Tiv	Tfv
Cruze	MT	915000	2019	23-04-2019	Ahora
Focus	Se Plus	950000	2019	01-09-2019	Ahora
Etios	XLS	1055000	2020	05-02-2020	30-11-2020
Civic	EXL	442000	2017	23-11-2017	Ahora
Suran	Feline	516000	2018	13-09-2018	Ahora

- b) Actualizar el precio del vehículo Cruze a 986500, que será efectivo a partir del día 01-10-2020:

Denominación	Versión	Precio	Año	Tiv	Tfv
Cruze	MT	915000	2019	23-04-2019	30-09-2020
Cruze	MT	986500	2019	01-10-2020	Ahora

- c) Insertar el siguiente vehículo, que comienza a estar vigente desde el día de la fecha (considerar la fecha actual para Tiv, tomamos la fecha para esta clase práctica):

Denominación	Versión	Precio	Año
Amarok	TrendLine	2177000	2020

Denominación	Versión	Precio	Año	Tiv	Tfv
Amarok	TrendLine	2177000	2020	19-09-2022	Ahora

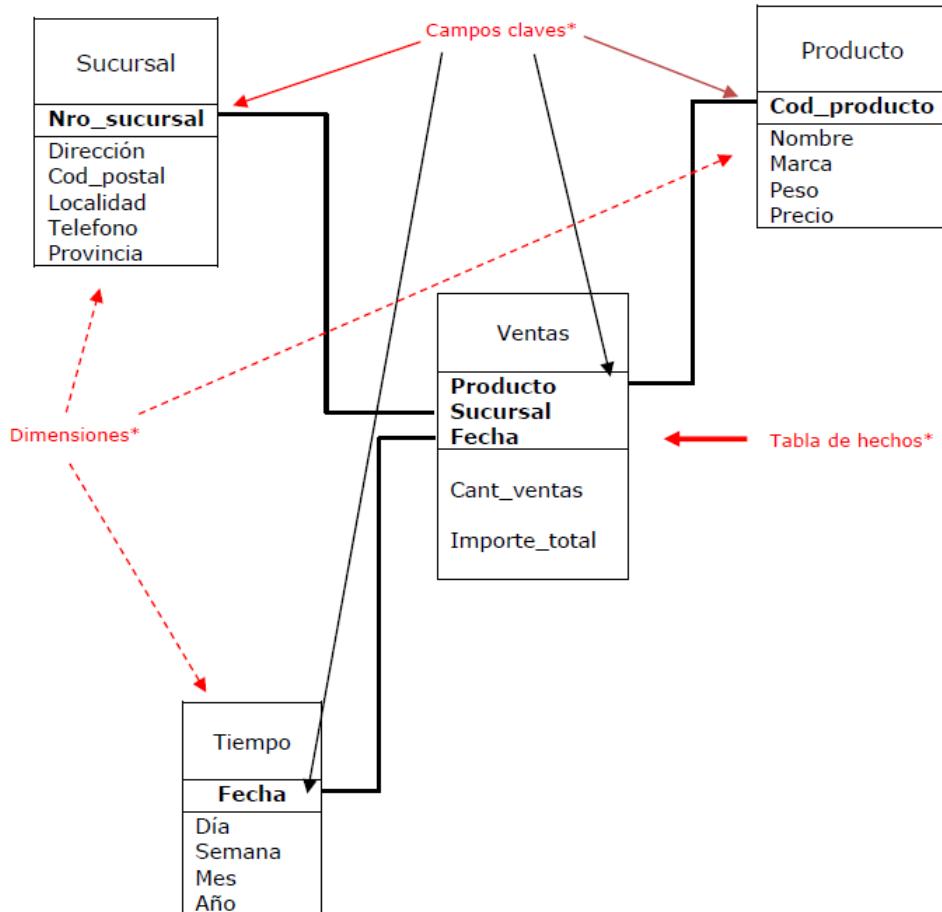
Vehículos

Denominación	Versión	Precio	Año	Tiv	Tfv
Cruze	MT	915000	2019	23-04-2019	30-09-2020
Cruze	MT	986500	2019	01-10-2020	Ahora
Focus	Se Plus	950000	2019	01-09-2019	Ahora
Etios	XLS	1055000	2020	05-02-2020	30-11-2020
Civic	EXL	442000	2017	23-11-2017	Ahora
Suran	Feline	516000	2018	13-09-2018	Ahora
Amarok	TrendLine	2177000	2020	19-09-2022	Ahora

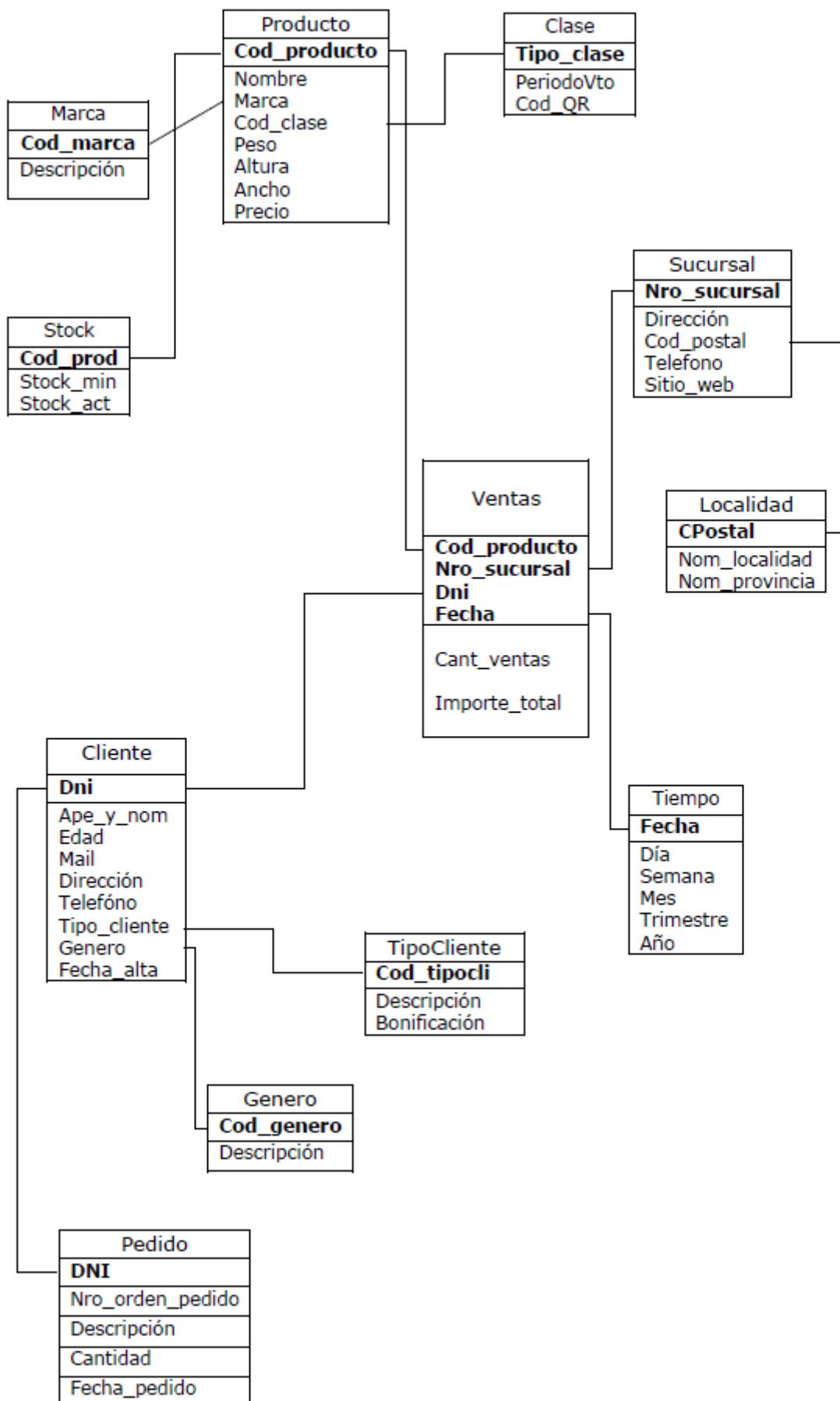
Serie Ejercicios Prácticos 6
Almacenes de

- 1.** Las tablas o dimensiones que necesitamos para extraer la información, para luego obtener las cantidades de unidades de productos y el importe total de las mismas, que se registrarán en la tabla de hechos son:Sucursal, Producto y Tiempo. Y la tabla de Hechos, corresponde a Ventas.

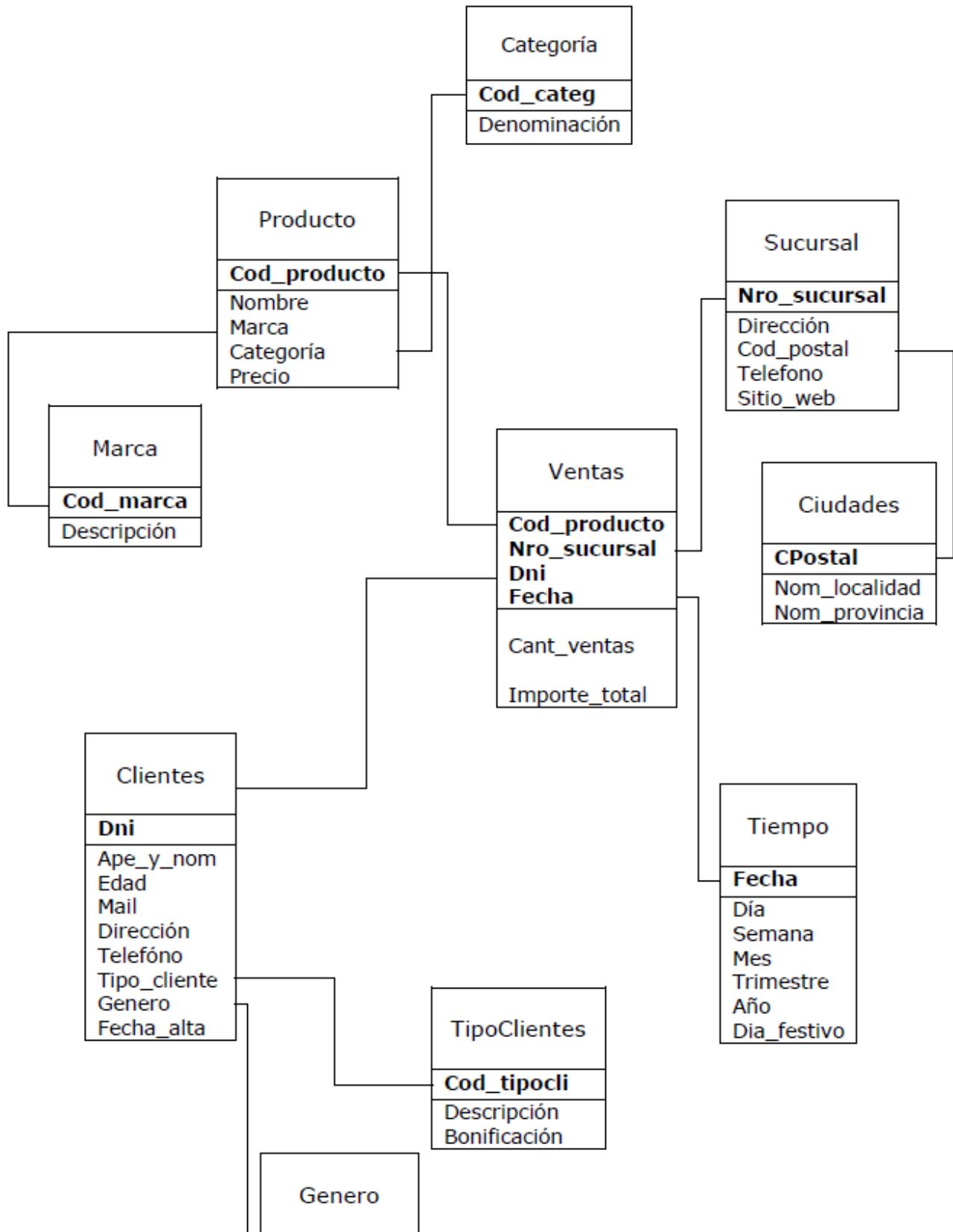
Cadena de supermercados Supermax



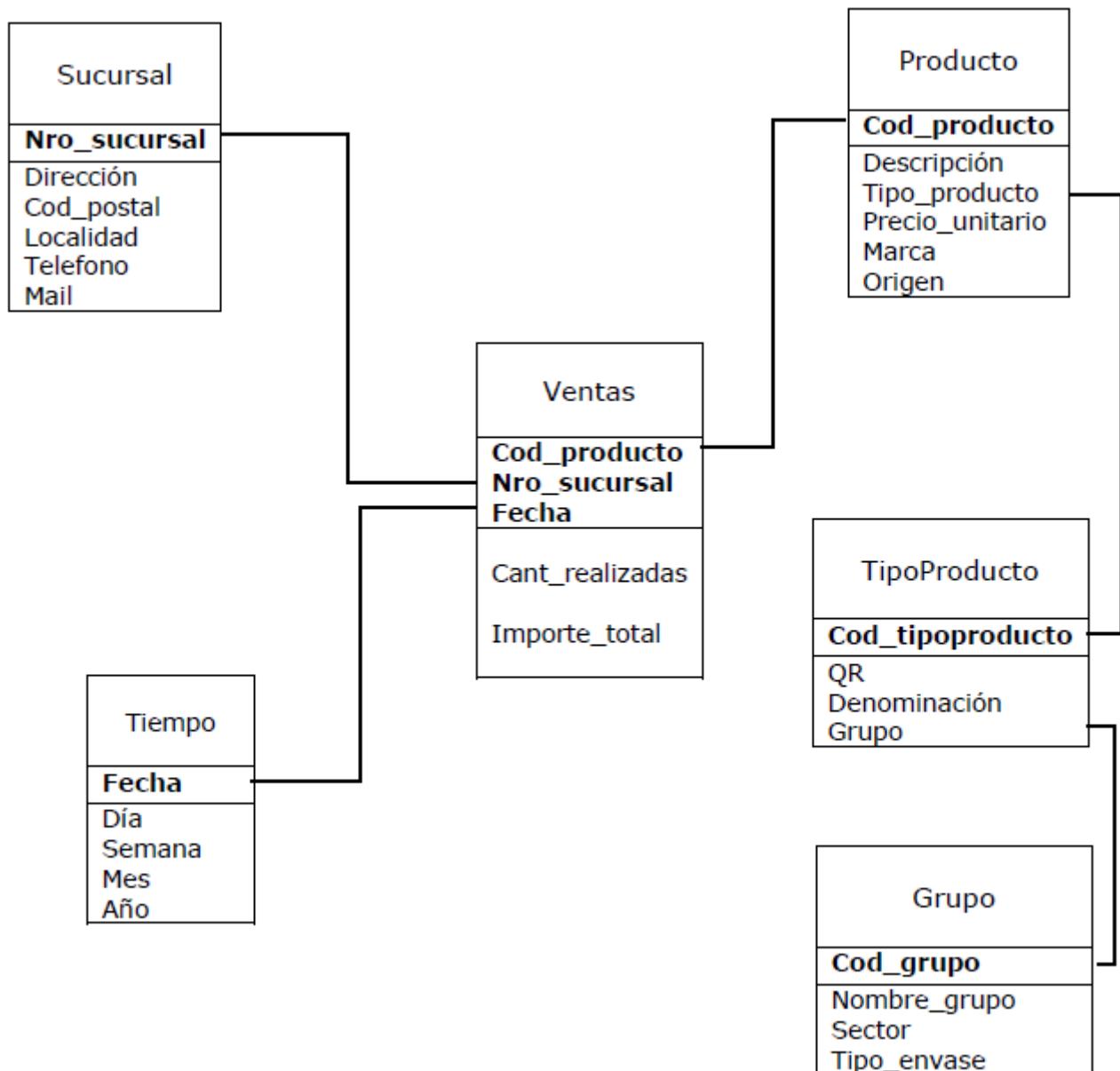
2. Empresa de electrodomésticos Megatone



3. Empresa de Telefonía



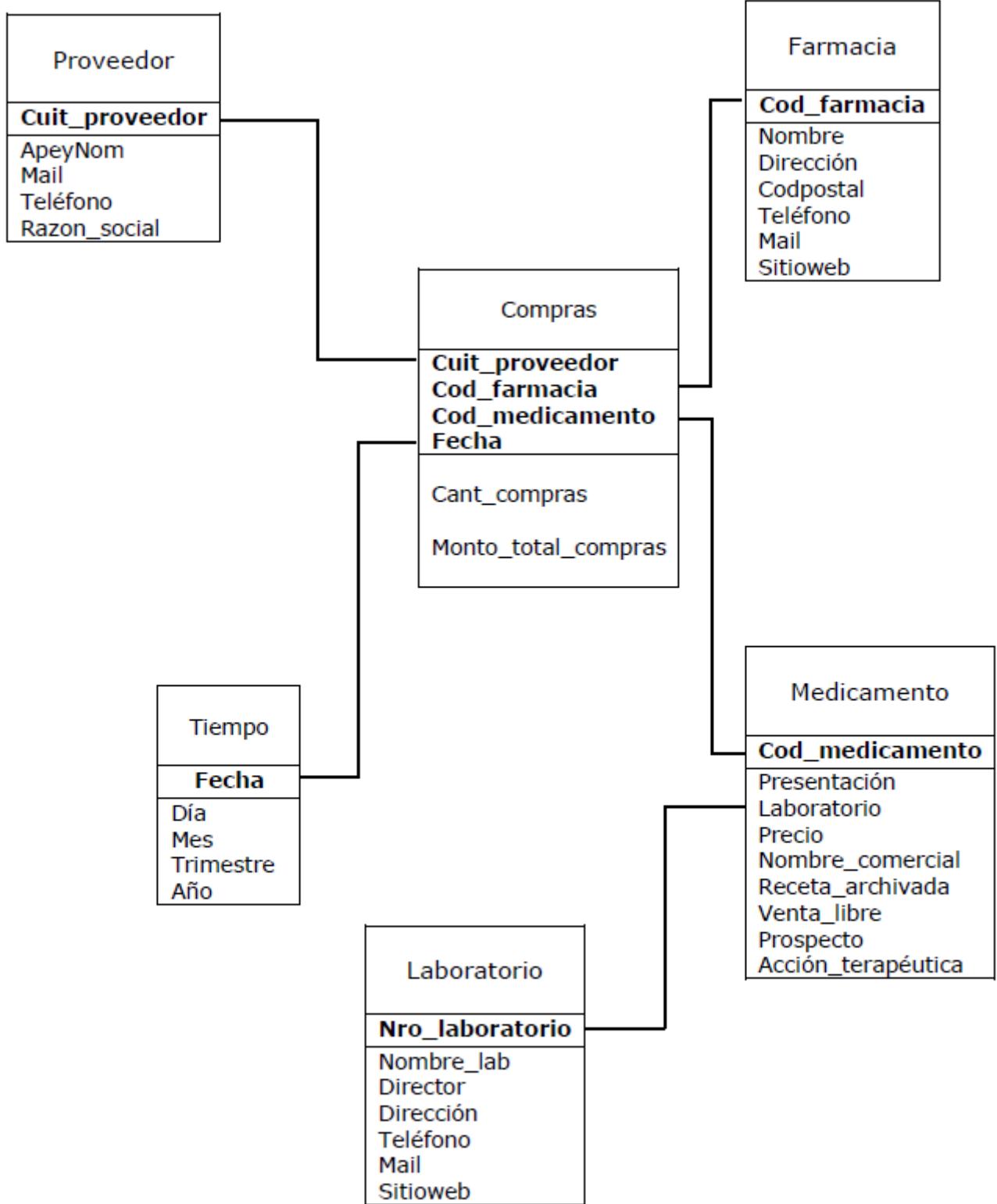
4. Hipermecado



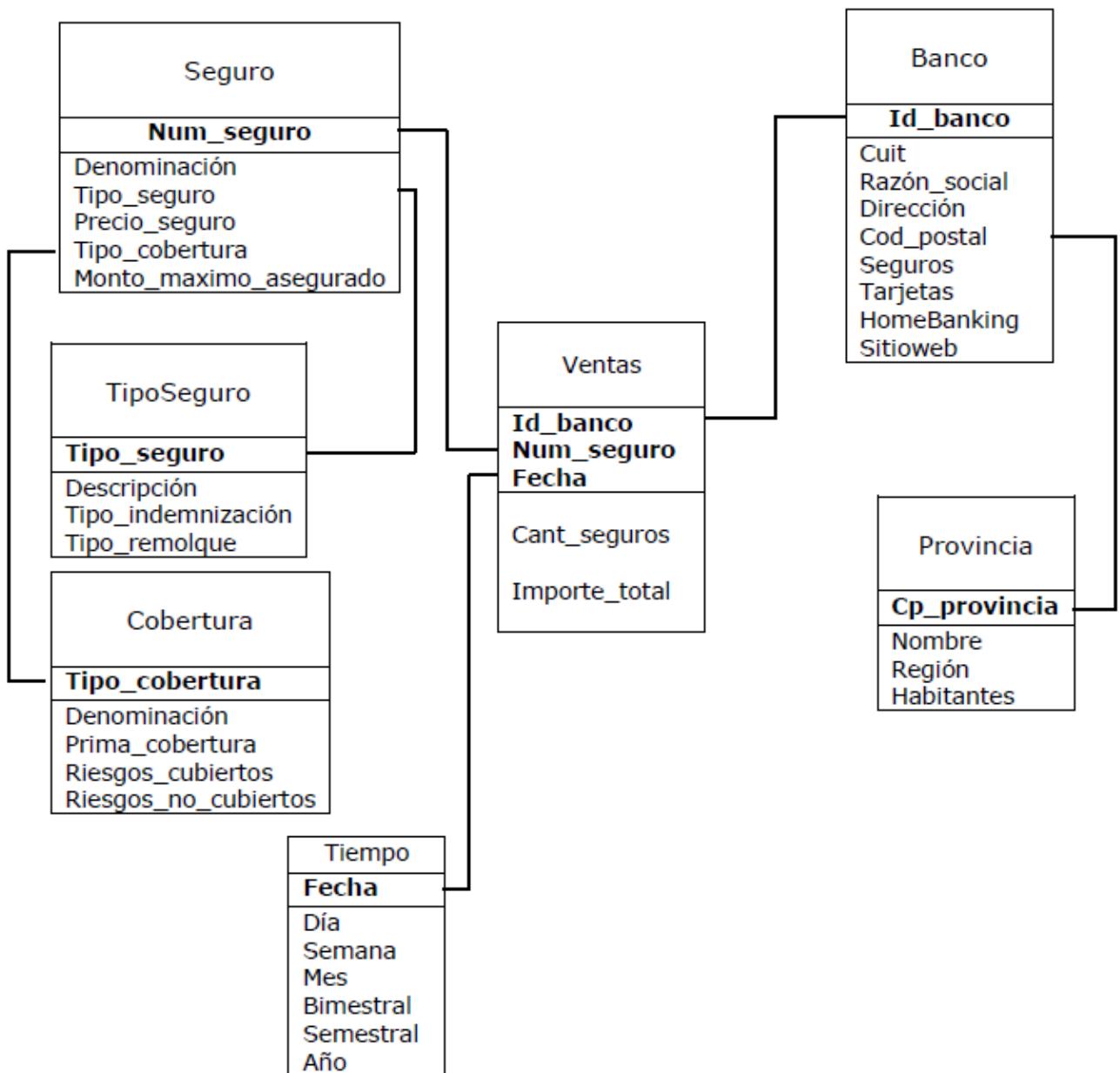
Diseño y armado de un almacén de datos

5.

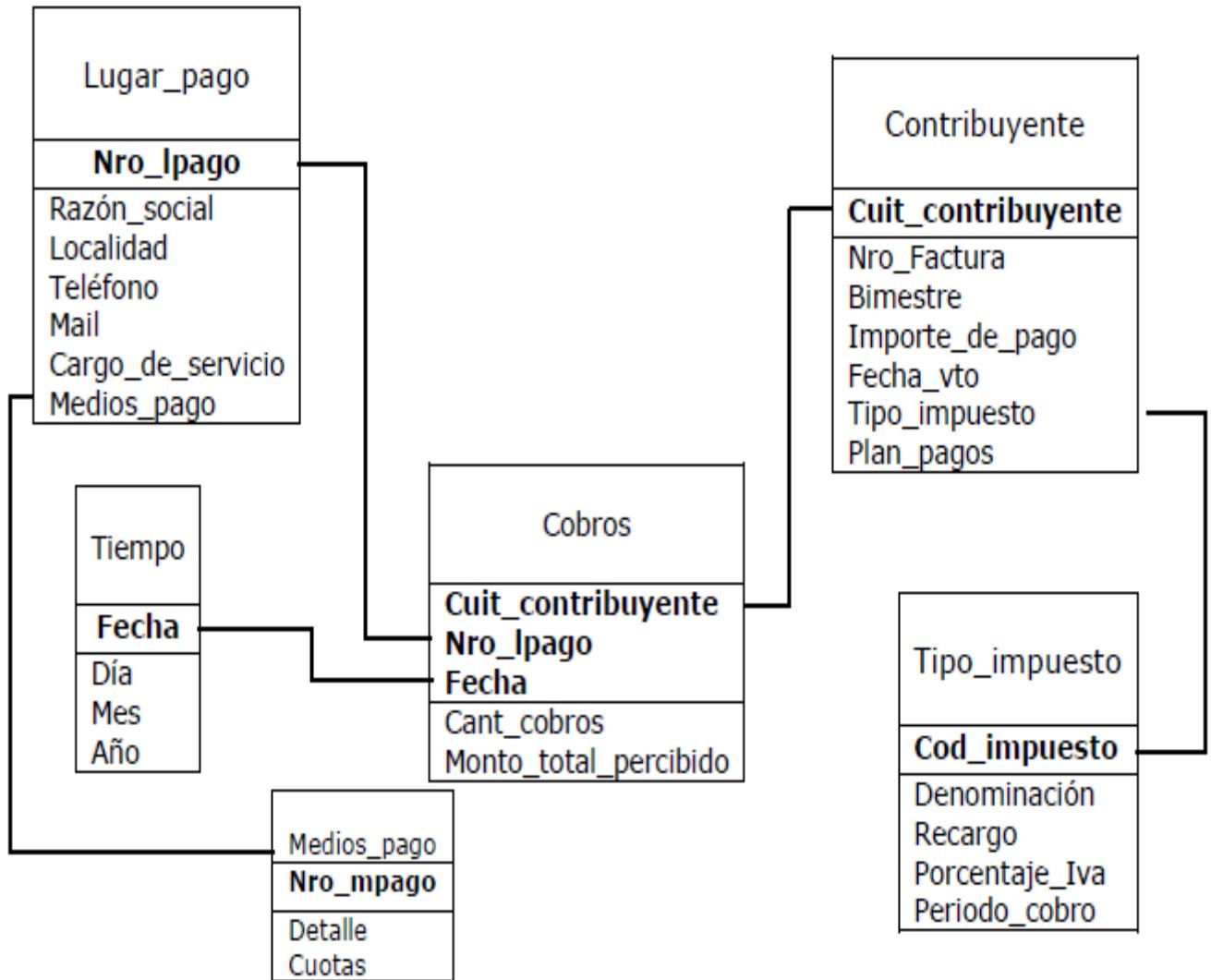
Red farmacéutica



6. Entidad bancaria



7. Entidad recaudadora de impuestos AFIP



Descripción:

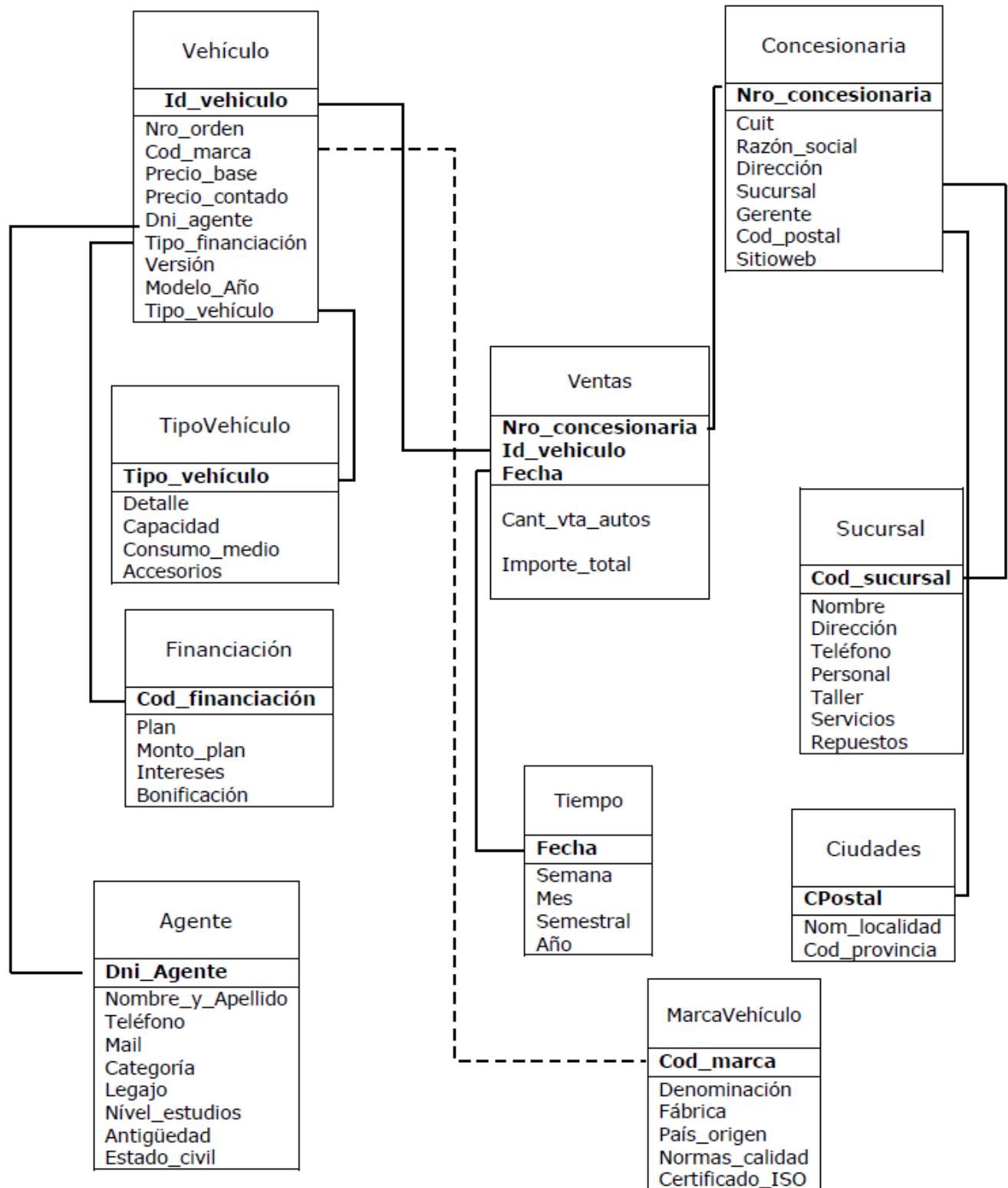
Organización: Entidad recaudadora de impuestos Afip.

Proceso: Cantidad de impuestos cobrados y el monto total de la cobranza.

Gránulo: se requiere obtener la cantidad de cobros de impuestos y el monto total, en forma diaria, mensual y anual.

Dimensiones: Tiempo, Lugar de Pago, Contribuyente, Tipo Impuesto y Medios pago

8. Concesionaria de autos



Serie Ejercicios Prácticos 7
OLAP

1) a)

Categoría	Semestre	Ciudad	Área	Lotes vendidos
verduras	primero	Corrientes	urbana	1250
verduras	primero	Corrientes	suburbana	1250
verduras	segundo	Corrientes	urbana	1350
verduras	segundo	Corrientes	suburbana	1350
verduras	primero	Goya	urbana	900
verduras	primero	Goya	suburbana	900
verduras	segundo	Goya	urbana	900
verduras	segundo	Goya	suburbana	900
verduras	primero	Mercedes	urbana	1100
verduras	primero	Mercedes	suburbana	1100
verduras	segundo	Mercedes	urbana	900
verduras	segundo	Mercedes	suburbana	900
verduras	primero	esquina	urbana	1100
verduras	primero	esquina	suburbana	1100
verduras	segundo	esquina	urbana	800
verduras	segundo	esquina	suburbana	800
carnes	primero	Corrientes	urbana	1000
carnes	primero	Corrientes	suburbana	1000
carnes	segundo	Corrientes	urbana	900
carnes	segundo	Corrientes	suburbana	900
carnes	primero	Goya	urbana	1000
carnes	primero	Goya	suburbana	1000
carnes	segundo	Goya	urbana	900
carnes	segundo	Goya	suburbana	900
carnes	primero	Mercedes	urbana	1000
carnes	primero	Mercedes	suburbana	1000
carnes	segundo	Mercedes	urbana	900
carnes	segundo	Mercedes	suburbana	900
carnes	primero	Esquina	urbana	950
carnes	primero	Esquina	suburbana	950
carnes	segundo	Esquina	urbana	750
carnes	segundo	Esquina	suburbana	750
hortalizas	primero	Corrientes	urbana	500
hortalizas	primero	Corrientes	suburbana	500
hortalizas	segundo	Corrientes	urbana	750
hortalizas	segundo	Corrientes	suburbana	750
hortalizas	primero	Goya	urbana	1000
hortalizas	primero	Goya	suburbana	1000
hortalizas	segundo	Goya	urbana	900
hortalizas	segundo	Goya	suburbana	900
hortalizas	primero	Mercedes	urbana	1100
hortalizas	primero	Mercedes	suburbana	1100
hortalizas	segundo	Mercedes	urbana	900
hortalizas	segundo	Mercedes	suburbana	900
hortalizas	primero	Esquina	urbana	700
hortalizas	primero	Esquina	suburbana	700
hortalizas	segundo	Esquina	urbana	750
hortalizas	segundo	Esquina	suburbana	750

b) Representación matricial de los datos

		Esquina	Mercedes	Goya	Corrientes
	Verduras	2500	2700		
	Carnes	2000	1800		
	Hortalizas	1000	1500		
		1er semestre	2do semestre		

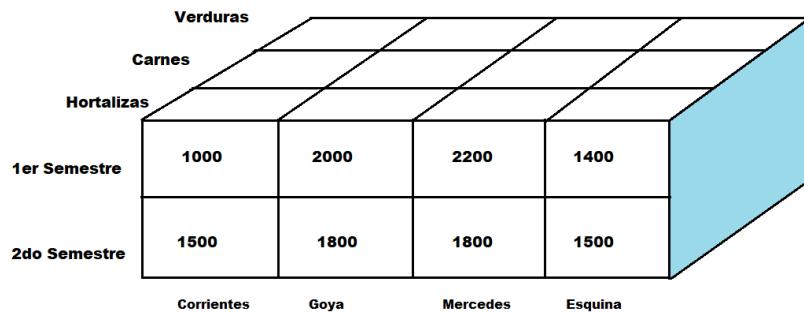
c) Representación de la consulta obtenida en el punto “a” de forma **anual**.

Categoría	Anual	Ciudad	Área	Lotes vendidos
verduras	anual	Corrientes	urbana	2600
verduras	anual	Corrientes	suburbana	2600
verduras	anual	Goya	urbana	1800
verduras	anual	Goya	suburbana	1800
verduras	anual	Mercedes	urbana	2000
verduras	anual	Mercedes	suburbana	2000
verduras	anual	Esquina	urbana	1900
verduras	anual	Esquina	suburbana	1900
carnes	anual	Corrientes	urbana	1900
carnes	anual	Corrientes	suburbana	1900
carnes	anual	Goya	urbana	1900
carnes	anual	Goya	suburbana	1900
carnes	anual	Mercedes	urbana	1900
carnes	anual	Mercedes	suburbana	1900
carnes	anual	Esquina	urbana	1700
carnes	anual	Esquina	suburbana	1700
hortalizas	anual	Corrientes	urbana	1250
hortalizas	anual	Corrientes	suburbana	1250
hortalizas	anual	Goya	urbana	1900
hortalizas	anual	Goya	suburbana	1900
hortalizas	anual	Mercedes	urbana	2000
hortalizas	anual	Mercedes	suburbana	2000
hortalizas	anual	Esquina	urbana	1450
hortalizas	anual	Esquina	suburbana	1450

d) Consulta por Categoría/Lotes Vendidos

Categoría	Lotes Vendidos
Verduras	16600
Carnes	14800
Hortalizas	13200

e) Representación del punto b usando el operador **Pivot**.



2)

a) Agregamos las ciudades de Corrientes y Resistencia en la tabla con igual proporción en cada una de ellas.

Vehículo	Cuatrimestre	Ciudad	Unidades Vendidas
Motos	C1	Corrientes	57000
Motos	C1	Resistencia	57000
Motos	C2	Corrientes	61000
Motos	C2	Resistencia	61000
Motos	C3	Corrientes	59000
Motos	C3	Resistencia	59000
Bicicletas	C1	Corrientes	109000
Bicicletas	C1	Resistencia	109000
Bicicletas	C2	Corrientes	61000
Bicicletas	C2	Resistencia	61000
Bicicletas	C3	Corrientes	108000
Bicicletas	C3	Resistencia	108000

b) Representación matricial de los datos.

Corrientes			
Resistencia			
Motos	57000	61000	59000
Bicicletas	109000	61000	108000
	C1	C2	C3

c) Debemos utilizar el operador **Roll** (agrupar):

Vehículo	Unidades Vendidas
Motos	354000
Bicicletas	556000

d) Representación del punto b utilizando el operador **Pivot**.

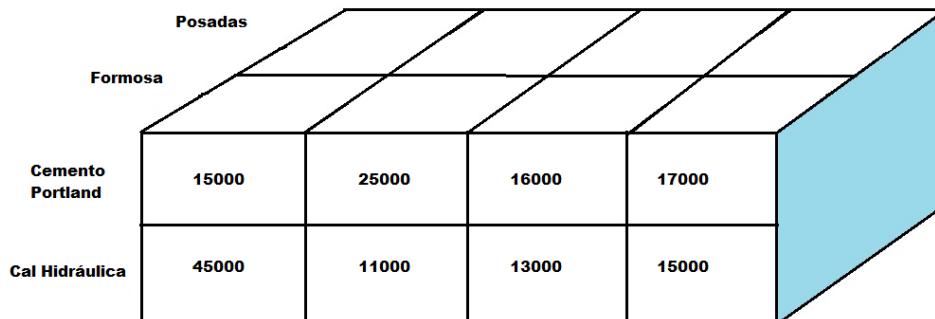
Motos			
Bicicletas			
C1	109000	109000	
C2	61000	61000	
C3	108000	108000	
	Resistencia	Corrientes	

3)

a)

Material	Trimestre	Ciudad	Bolsas Vendidas
Cemento Portland	T1	Formosa	15000
Cemento Portland	T1	Posadas	15000
Cemento Portland	T2	Formosa	25000
Cemento Portland	T2	Posadas	25000
Cemento Portland	T3	Formosa	16000
Cemento Portland	T3	Posadas	16000
Cemento Portland	T4	Formosa	17000
Cemento Portland	T4	Posadas	17000
Cal Hidráulica	T1	Formosa	45000
Cal Hidráulica	T1	Posadas	45000
Cal Hidráulica	T2	Formosa	11000
Cal Hidráulica	T2	Posadas	11000
Cal Hidráulica	T3	Formosa	13000
Cal Hidráulica	T3	Posadas	13000
Cal Hidráulica	T4	Formosa	15000
Cal Hidráulica	T4	Posadas	15000

b) Representación matricial de los datos.



c)

Material	Semestre	Ciudad	Bolsas Vendidas
Cemento Portland	S1	Formosa	40000
Cemento Portland	S1	Posadas	40000
Cemento Portland	S2	Formosa	33000
Cemento Portland	S2	Posadas	33000
Cal Hidráulica	S1	Formosa	56000
Cal Hidráulica	S1	Posadas	56000
Cal Hidráulica	S2	Formosa	28000
Cal Hidráulica	S2	Posadas	28000

d) Representación del punto “b” utilizando el operador Píivot.

	Cemento Portland		
	Cal Hidráulica		
T4	15000	15000	
T3	13000	13000	
T2	11000	11000	
T1	45000	45000	
	Formosa	Posadas	

4)

a)

Producto	Bimestre	Ciudad	Kilogramos Vendidos
Harina	B1	Paraná	12000
Harina	B1	La Paz	4000
Harina	B1	Federación	4000
Harina	B2	Paraná	14400
Harina	B2	La Paz	4800
Harina	B2	Federación	4800
Harina	B3	Paraná	11400
Harina	B3	La Paz	3800
Harina	B3	Federación	3800
Harina	B4	Paraná	12000
Harina	B4	La Paz	4000
Harina	B4	Federación	4000
Harina	B5	Paraná	15600
Harina	B5	La Paz	5200
Harina	B5	Federación	5200
Harina	B6	Paraná	10800
Harina	B6	La Paz	3600
Harina	B6	Federación	3600
Arroz	B1	Paraná	7200
Arroz	B1	La Paz	2400
Arroz	B1	Federación	2400
Arroz	B2	Paraná	9600
Arroz	B2	La Paz	3200
Arroz	B2	Federación	3200
Arroz	B3	Paraná	10800
Arroz	B3	La Paz	3600
Arroz	B3	Federación	3600
Arroz	B4	Paraná	9000
Arroz	B4	La Paz	3000
Arroz	B4	Federación	3000

Arroz	B5	Paraná	6000
Arroz	B5	La Paz	2000
Arroz	B5	Federación	2000
Arroz	B6	Paraná	6600
Arroz	B6	La Paz	2200
Arroz	B6	Federación	2200
Yerba mate	B1	Paraná	14400
Yerba mate	B1	La Paz	4800
Yerba mate	B1	Federación	4800
Yerba mate	B2	Paraná	13200
Yerba mate	B2	La Paz	4400
Yerba mate	B2	Federación	4400
Yerba mate	B3	Paraná	13800
Yerba mate	B3	La Paz	4600
Yerba mate	B3	Federación	4600
Yerba mate	B4	Paraná	14400
Yerba mate	B4	La Paz	4800
Yerba mate	B4	Federación	4800
Yerba mate	B5	Paraná	12000
Yerba mate	B5	La Paz	4000
Yerba mate	B5	Federación	4000
Yerba mate	B6	Paraná	12600
Yerba mate	B6	La Paz	4200
Yerba mate	B6	Federación	4200

b) Representación matricial de los datos.

Federación La Paz Paraná		B1	B2	B3	B4	B5	B6
Harina	12000	14400	11400	12000	15600	10800	
Arroz	7200	9600	10800	9000	6000	6600	
Yerba Mate	14400	13200	13800	14400	12000	12600	

c)

Producto	Cuatrimestre	Ciudad	Kilogramos Vendidos
Harina	1C	Paraná	26400
Harina	1C	La Paz	8800
Harina	1C	Federación	8800
Harina	2C	Paraná	23400
Harina	2C	La Paz	7800
Harina	2C	Federación	7800
Harina	3C	Paraná	26400
Harina	3C	La Paz	8800
Harina	3C	Federación	8800
Arroz	1C	Paraná	16800
Arroz	1C	La Paz	5600
Arroz	1C	Federación	5600
Arroz	2C	Paraná	19800
Arroz	2C	La Paz	6600
Arroz	2C	Federación	6600
Arroz	3C	Paraná	12600
Arroz	3C	La Paz	4200
Arroz	3C	Federación	4200
Yerba mate	1C	Paraná	27600
Yerba mate	1C	La Paz	9200
Yerba mate	1C	Federación	9200
Yerba mate	2C	Paraná	28200
Yerba mate	2C	La Paz	9400
Yerba mate	2C	Federación	9400
Yerba mate	3C	Paraná	24600
Yerba mate	3C	La Paz	8200
Yerba mate	3C	Federación	8200

d) Para obtener la consulta por Producto/Kilogramos vendidos usamos el operador **Roll**.

Producto	Kilogramos Vendidos
Harina	127000
Arroz	82000
Yerba mate	134000

e) Representación de los datos del punto “b” usando el operador Pívot.

		Yerba mate	Arroz	
		Harina		
B6	Paraná	10800	3600	3600
	B5	15600	5200	5200
	B4	12000	4000	4000
	B3	11400	3800	3800
	B2	14400	4800	4800
	B1	12000	4000	4000
		Paraná	La Paz	Federación

Serie Ejercicios Prácticos 8
Minería de Datos

1) Construya el árbol de decisión de acuerdo al algoritmo ID3 para los siguientes datos, donde se concluye si jugar o no al fútbol de acuerdo a las condiciones del tiempo.

Día	Tiempo	Temperatura	Humedad	Viento	Jugar fútbol
1	soleado	alta	alta	débil	no
2	soleado	alta	alta	fuerte	no
3	nublado	alta	alta	débil	si
4	lluvia	templada	alta	débil	si
5	lluvia	frio	normal	débil	si
6	lluvia	frio	normal	fuerte	no
7	nublado	frio	normal	fuerte	si
8	soleado	templada	alta	débil	no
9	soleado	frio	normal	débil	si
10	lluvia	templada	normal	débil	si
11	soleado	templada	normal	fuerte	si
12	nublado	templada	alta	fuerte	si
13	nublado	alta	normal	débil	si
14	lluvia	templada	alta	fuerte	no

Fórmulas asociadas:

ID3: permite la generación de un árbol de decisión, utiliza el concepto de ganancia de la información para determinar el atributo más útil en cada caso.

Entropía: calcula el grado de incertidumbre de una muestra, si la muestra es totalmente homogénea = 0 y si es igualmente distribuida es = 1.

$$\text{Entropía}(S) = E(S) = - \sum_{i=1}^C p_i * \log_2(p_i)$$

Para el caso donde tenemos alternativas binarias la fórmula utilizada es la siguiente:

$$E(S) = (-P \log_2(P)) + (-N \log_2(N))$$

. 1) Procedemos de lo más general a lo particular obteniendo la probabilidad para el caso favorable y el no favorable.

Jugar al fútbol	SI	NO
Días	9	5
Probabilidad	9/14	5/14

$$\text{Log}_2(9/14) = \log(0.64) / \log(2) = -0.64$$

$$\text{Log}_2(5/14) = \log(0.36) / \log(2) = -1.51$$

$$\text{Entropia (9,5)} = (-9/14 \log_2(9/14)) + (-5/14 \log_2(5/14))$$

$$\begin{aligned}
 &= (-9/14 * (-0.64)) + (-5/14 * (-1.51)) // \text{reemplazamos los valores obtenidos} \\
 &= 0.41 + 0.53 \\
 &= \mathbf{0.94}
 \end{aligned}$$

$$\text{Entropía(Jugar Al Fútbol)} = 0.94$$

2) Aplicamos la ganancia de la información respecto a cada valor posible para este atributo luego de calcular las respectivas entropias :

Atributo	Valores	SI	NO	Total
	soleado	2	3	5
	lluvia	3	2	5

Tiempo	nublado	4	0	4
--------	---------	---	---	---

Entropía(soleado)(2,3) =

$$\text{Log}_2(2/5) = \log(0.4) / \log(2) = -1.32$$

$$\text{Log}_2(3/5) = \log(0.6) / \log(2) = -0.73$$

$$E(\text{soleado}) = E(2,3) = E(2/5,3/5)$$

$$\begin{aligned} &= E(0,4, 0,60) \\ &= (-0,4 * (-1.32)) + (-0,60 * (-0.73)) \\ &= 0,528 + 0,438 \\ &= 0,966 \\ &= \mathbf{0.97} \end{aligned}$$

Entropía(lluvia)(3,2) =

$$\text{Log}_2(3/5) = \log(0.6) / \log(2) = -0.73$$

$$\text{Log}_2(2/5) = \log(0.4) / \log(2) = -1.32$$

$$E(\text{lluvia}) = E(3,2) = E(3/5,2/5)$$

$$\begin{aligned} &= E(0,60, 0,4) \\ &= (-0,60 * (-0.73)) + (-0,4 * (-1.32)) \\ &= 0,438 + 0,528 \\ &= 0,966 \\ &= \mathbf{0.97} \end{aligned}$$

Entropía(nublado)(4,0) = 0

$$\text{Log}_2(4/4) = \log(1) / \log(2) = 0$$

$$\text{Log}_2(0/4) = \log(0) / \log(2) = \text{no existe el valor}$$

Entropía(Jugar Al Fútbol,Tiempo):

$$= P(\text{soleado}) * E(2,3) + P(\text{nublado}) * E(4,0) + P(\text{lluvia}) * E(3,2)$$

$$= (5/14) * 0,97 + (4/14) * 0 + (5/14) * 0,97$$

$$= 0,3571 * 0,97 + 0,3571 * 0,97$$

$$= \mathbf{0,692774}$$

La ganancia de información para el atributo tiempo la obtenemos, considerando la diferencia entre la entropía total y la entropía tiempo:

Ganancia (Jugar al fútbol, Tiempo) :

$$E(\text{Jugar al fútbol}) - E(\text{Jugar al fútbol, Tiempo})$$

$$= 0,94 - 0,692$$

$$= 0,248$$

$$\text{Ganancia}(\text{Jugar al fútbol, Tiempo}) = \rightarrow \mathbf{0,25}$$

Obtenemos la distribución

Atributo	Valores	SI	NO	Total
Temperatura	alta	2	2	4
	fría	3	1	4
	templada	4	2	6

Calculamos la entropía para cada valor posible:

Entropía(alta)(2,2) =

$$\text{Log}_2(2/4) = \log(0.5) / \log(2) = -1$$

$$\text{Log}_2(2/4) = \log(0.5) / \log(2) = -1$$

$$E(\text{alta}) = E(2,2) = E(2/4,2/4)$$

$$= E(0.5, 0.5)$$

$$= (-0.5 * (-1)) + (-0.5 * (-1))$$

$$= 0.5 + 0.5$$

$$= \mathbf{1}$$

Entropía(fría)(3,1) =

$$\text{Log}_2(3/4) = \log(0.75) / \log(2) = -0.41$$

$$\text{Log}_2(1/4) = \log(0.25) / \log(2) = -2$$

$$\begin{aligned}
E(\text{fria}) &= E(3,1) = E(3/4,1/4) \\
&= E(0.75, 0.25) \\
&= (-0.75 * (-0.41)) + (-0.25 * (-2)) \\
&= 0.3075 + 0.5 \\
&= \mathbf{0.8075}
\end{aligned}$$

Entropía(templada)(4,2) =

$$\begin{aligned}
\log_2(4/6) &= \log(0.66) / \log(2) = -0.5994 \\
\log_2(2/6) &= \log(0.33) / \log(2) = -1.5994
\end{aligned}$$

$$\begin{aligned}
E(\text{templada}) &= E(4,2) = E(4/6,2/6) \\
&= E(0.66, 0.33) \\
&= (-0.66 * (-0.5994)) + (-0.33 * (-1.5994)) \\
&= 0.3956 + 0.5278 \\
&= \mathbf{0.91}
\end{aligned}$$

Entropía(Jugar Al Fútbol, Temperatura):

$$\begin{aligned}
&= P(\text{alta}) * E(2,2) + P(\text{fria}) * E(3,1) + P(\text{templada}) * E(4,2) = \\
&((4/14) * 1) + (4/14) * 0.8075 + (6/14) * 0.91 \\
&= 0,3571 * 0,97 + 0,3571 * 0,97 \\
&= \mathbf{0,906428}
\end{aligned}$$

$$\text{Ganancia(Jugar al fútbol, Temperatura)} = 0,94 - 0,906428 = \rightarrow \mathbf{0,03}$$

Obtenemos la distribución

Atributo	Valores	SI	NO	Total
Humedad	alta	3	4	7
	normal	6	1	7

Calculamos la entropía para cada valor posible:

Entropía(alta)(3,4) =

$$\text{Log}_2(3/7) = \log(0.43) / \log(2) = -1.21$$

$$\text{Log}_2(4/7) = \log(0.57) / \log(2) = -0.81$$

$$\begin{aligned}
 E(\text{humedad})(\text{alta}) &= E(3,4) = E(3/7,4/7) \\
 &= E(0.43, 0.57) \\
 &= (-0.43 * (-1.21)) + (-0.57 * (-0.81)) \\
 &= 0.5203 + 0.4617 \\
 &= \mathbf{0.982}
 \end{aligned}$$

Entropía(normal)(6,1) =

$$\text{Log}_2(6/7) = \log(0.85) / \log(2) = -0.234$$

$$\text{Log}_2(1/7) = \log(0.14) / \log(2) = -2.83$$

$$\begin{aligned}
 E(\text{humedad})(\text{normal}) &= E(6,1) = E(6/7,1/7) \\
 &= E(0.85, 0.14) \\
 &= (-0.85 * (-0.234)) + (-0.14 * (-2.83))
 \end{aligned}$$

$$= 0.1989 + 0.3962$$

$$= \mathbf{0.5951}$$

Entropía(Jugar Al Fútbol,Humedad):

$$= P(\text{alta}) * E(3,4) + P(\text{normal}) * E(6,1) =$$

$$((7/14) * 0.982) + (7/14) * 0.5951)$$

$$= 0.5 * 0.982 + 0.5 * 0.5951$$

$$= \mathbf{0.78855}$$

$$\text{Ganancia}(\text{Jugar al fútbol, Humedad}) = 0,94 - 0,78855 = \rightarrow \mathbf{0,1514}$$

Obtenemos la distribución

Atributo	Valores	SI	NO	Total
Viento	débil	6	2	8
	fuerte	3	3	6

Calculamos la entropía para cada valor posible:

Entropía(débil)(6,2) =

$$\text{Log}_2(6/8) = \log(0.75) / \log(2) = -0.41$$

$$\text{Log}_2(2/8) = \log(0.25) / \log(2) = -2$$

$$\begin{aligned}
E(\text{viento})(\text{débil}) &= E(6,2) = E(6/8,2/8) \\
&= E(0.75, 0.25) \\
&= (-0.75 * (-0.41)) + (-0.25 * (-2)) \\
&= 0.3075 + 0.5 \\
&= \mathbf{0.8075}
\end{aligned}$$

Entropía(fuerte)(3,3) =

$$\text{Log}_2(3/6) = \log(0.5) / \log(2) = -1$$

$$\text{Log}_2(3/6) = \log(0.5) / \log(2) = -1$$

$$E(\text{viento})(\text{fuerte}) = E(6,1) = E(3/6,3/6)$$

$$\begin{aligned}
&= E(0.5, 0.5) \\
&= (-0.5 * (-1)) + (-0.5 * (-1)) \\
&= 0.5 + 0.5 \\
&= \mathbf{1}
\end{aligned}$$

Entropía(Jugar Al Fútbol,Viento):

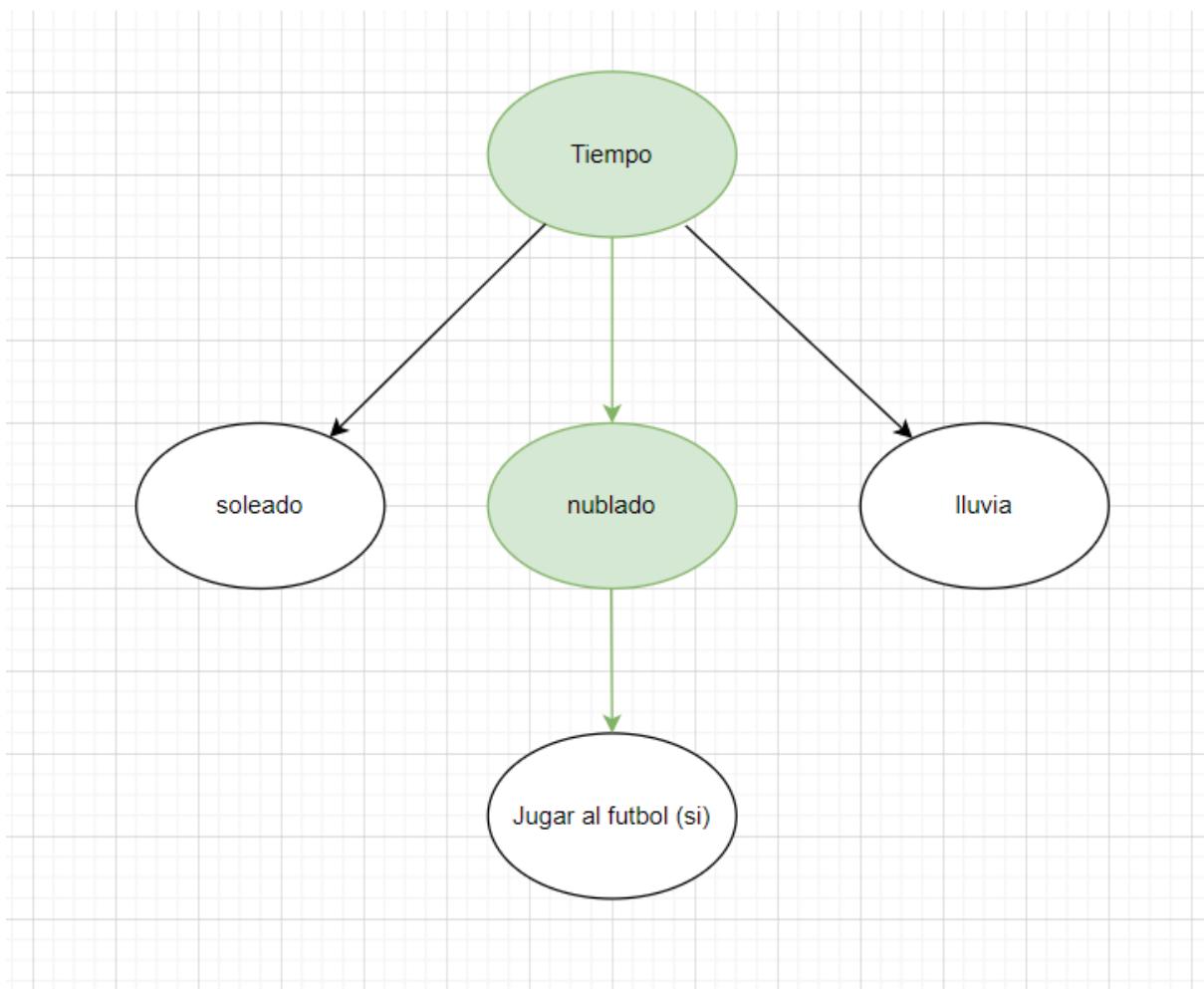
$$\begin{aligned}
&= P(\text{débil}) * E(6,2) + P(\text{fuerte}) * E(3,3) = \\
&((8/14) * 0.8075) + (6/14) * 1 \\
&= 0.4614 + 0.4285 \\
&= \mathbf{0.8899}
\end{aligned}$$

$$\text{Ganancia}(\text{Jugar al fútbol, Viento}) = 0,94 - 0,8899 = \rightarrow \mathbf{0,0501}$$

Atributo	Ganancia
Tiempo	0,25
Temperatura	0,03
Humedad	0,1514
Viento	0,0501

Tomamos el atributo con mayor ganancia y partimos de él.

Representación gráfica, primer acercamiento.



Tomamos una muestra de menor tamaño y aplicamos nuevamente el proceso:

Tiempo	Temperatura	Humedad	Viento	Jugar fútbol
soleado	alta	Alta	débil	no
soleado	alta	Alta	fuerte	no
soleado	templada	Alta	débil	no
soleado	frio	Normal	débil	si
soleado	templada	Normal	fuerte	si

Tiempo	SI	NO	Total
soleado	3	2	5

En principio obtenemos la entropía correspondiente al atributo Tiempo con el valor soleado.

$$\text{Log}_2(3/5) = \log(0.6) / \log(2) = -0.7369$$

$$\text{Log}_2(2/5) = \log(0.4) / \log(2) = -1.32$$

$$E(\text{Tiempo})(\text{soleado}) =$$

$$= E(3,2) = -(0.6) * (-0.7369) + -(0.4) * (-1.32))$$

$$= 0.44214 + 0.528$$

$$= \mathbf{0.97014}$$

Atributo	Valores	SI	NO
Temperatura	alta	0	2
	frío	1	0
	templada	1	1

$$\begin{aligned}
 E(\text{soleado}, \text{Temperatura}) &= P(\text{alta}) * E(0,2) + P(\text{frío}) * E(1,0) + P(\text{templada}) * E(1,1) \\
 &= (2/5) * E(0/2,2/2) + (1/5) * E(1/1,0/1) \\
 &= 0,4 * 0 + 0,4 * 1 + 0,2 * 0 \\
 &= 0,4 + (2/5) * E(1/2,1/2) \\
 &= 0,4
 \end{aligned}$$

$$\text{Ganancia} (\text{soleado}, \text{Temperatura}) = 0,97 - 0,4 = \mathbf{0,57}$$

Atributo	Valores	SI	NO
Humedad	Alta	0	3
	Normal	2	0

$$\begin{aligned}
 E(\text{soleado}, \text{Humedad}) &= P(\text{alta}) * E(0,3) + P(\text{normal}) * E(2,0) \\
 &= (3/5) * E(0/3,3/3) + (2/5) * E(2/2,0/2) \\
 &= 0,6 * 0 + 0,4 * 0 \\
 &= 0
 \end{aligned}$$

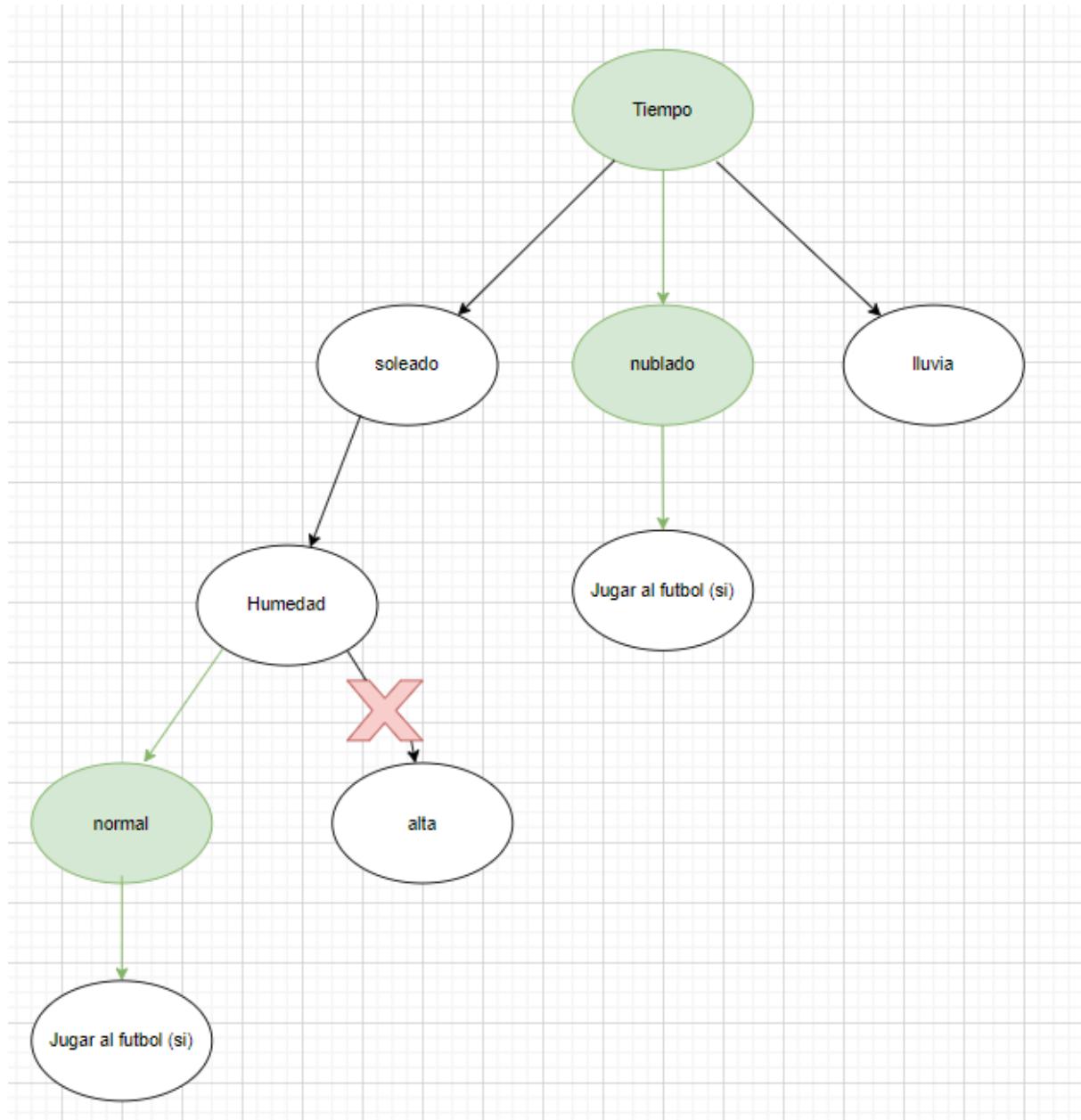
$$\text{Ganancia}(\text{soleado}, \text{Humedad}) = 0,97 - 0 = \mathbf{0,97}$$

Atributo	Valores	SI	NO
Viento	débil	1	2
	fuerte	1	1

$$\begin{aligned}
 E(\text{soleado}, \text{Viento}) &= P(\text{débil}) * E(1,2) + P(\text{fuerte}) * E(1,1) \\
 &= (3/5) * E(1/3, 2/3) + (2/5) * E(1/2, 1/2) \\
 &= 0,6 * 0,9183 + 0,4 * 1 \\
 &= 0,95098
 \end{aligned}$$

$$\text{Ganancia}(\text{soleado}, \text{Viento}) = 0,97 - 0,95098 = 0,019 = \mathbf{0,02}$$

Atributo	Ganancia
Temperatura	0,57
Humedad	0,97
Viento	0,02



Ahora evaluamos el valor correspondiente a “lluvia”

Tiempo	Temperatura	Humedad	Viento	Jugar fútbol
lluvia	templada	Alta	débil	si
lluvia	frio	Normal	débil	si
lluvia	frio	Normal	fuerte	no
lluvia	templada	Normal	débil	si
lluvia	templada	Alta	fuerte	no

$$\text{Log}_2 (3/5) = \log(0.6) / \log (2) = -0.7369$$

$$\text{Log}_2 (2/5) = \log(0.4) / \log (2) = -1.3219$$

$$E(\text{lluvia}) = E(3,2)$$

$$= -(3/5) * (-0.7369) + -(2/5) * -1.32$$

$$= 0.44214 + 0.528$$

$$= \mathbf{0,97}$$

		Iluvia	
Atributo	Valores	SI	NO
Temperatura	templada	2	1
	fría	1	1

$$E(\text{lluvia}, \text{Temperatura}) = P(\text{templada}) * E(2,1) + P(\text{frío}) * E(1,1)$$

$$= (3/5) * E(2/3,1/3) + (2/5) * E(1/2,1/2)$$

$$= 0,6 * 0,9183 + 0,4 * 1$$

$$= \mathbf{0,9598}$$

$$\text{Ganancia}(\text{lluvia}, \text{Temperatura}) = 0,97 - 0,9598 = 0,019 = \mathbf{0,02}$$

		Iluvia	
Atributo	Valores	SI	NO
	alta	1	1

Humedad	normal	2	1
---------	--------	---	---

$$\begin{aligned}
 E(\text{lluvia, Humedad}) &= P(\text{alta}) * E(1,1) + P(\text{normal}) * E(2,1) \\
 &= (2/5) * E(1/2,1/2) + (3/5) * E(2/3,1/3) \\
 &= 0,4 * 1 + 0,6 * 0,9183 \\
 &= 0,9598
 \end{aligned}$$

$$\text{Ganancia}(\text{lluvia, Humedad}) = 0,97 - 0,9598 = 0,019 = \mathbf{0,02}$$

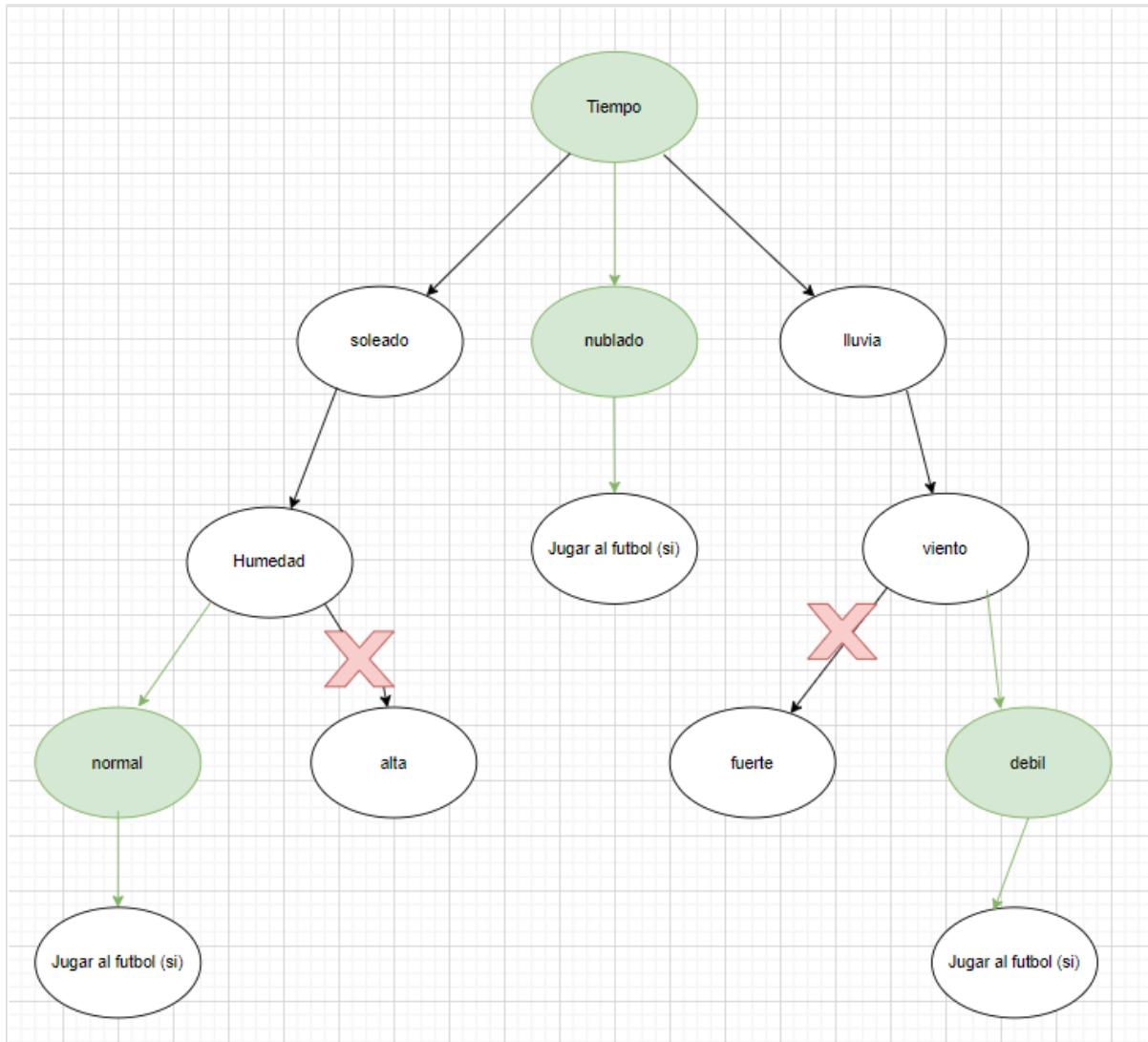
		lluvia	
Atributo	Valores	SI	NO
Viento	débil	3	0
	fuerte	0	2

$$\begin{aligned}
 E(\text{lluvia, Viento}) &= P(\text{débil}) * E(3,0) + P(\text{fuerte}) * E(0,2) \\
 &= (3/5) * E(3/3,0/3) + (2/5) * E(0/2,2/2) \\
 &= 0,6 * 0 + 0,4 * 0 \\
 &= 0
 \end{aligned}$$

$$\text{Ganancia}(\text{lluvia, Viento}) = 0,97 - 0 = \mathbf{0,97}$$

Atributo	Ganancia
Temperatura	0,2
Humedad	0,2
Viento	0,97

Finalmente el árbol nos quedaría de esta forma



2) Dada la siguiente tabla de las transacciones de ciertos productos (ítems) de una cadena de supermercados, aplicar el algoritmo “apriori” para encontrar las reglas que predicen la ocurrencia de un ítem.

Considerar los siguientes valores para soporte mínimo = 2 (support) y para la confianza = 0.75.

...

Transacción	Ítems
1	Aceite, leche, azúcar
2	Aceite, yerba, cerveza, azúcar
3	Leche, yerba, cerveza, sal
4	Aceite, leche, yerba, cerveza, sal
5	Aceite, leche, yerba, sal

Transacción	Aceite	Leche	Yerba	Cerveza	Azúcar	Sal
1	1	1	0	0	1	0
2	1	0	1	1	1	0
3	0	1	1	1	0	1
4	1	1	1	1	0	1
5	1	1	1	0	0	1

Conjuntos de ítems

Para este caso los ítems son los siguientes:

$$I = \{\text{Aceite, Leche, Yerba, Cerveza, Azúcar, Sal}\}$$

Conjunto de Transacciones

D

El proceso consta de 2 etapas

*A.Generación de los itemsets frecuentes

*B.Generación de Reglas

Identificación de ítems individuales

ítem individual	soporte	soporte mínimo
{Aceite}	4	4/5=0,80=80%
{Leche}	4	4/5=0,80=80%
{Yerba}	4	4/5=0,80=80%
{Cerveza}	3	3/5=0,60=60%
{Azúcar}	2	2/5=0,40=40%
{Sal}	3	3/5=0,60=60%

A – Conjuntos Frecuentes

k = A toda colección de cero o más ítems se denomina itemset. Si un itemset contiene k ítems, se le llama k-itemset.

Combinaciones de 2 ítems, itemset (k=2):

$S_2 = \{$
 {Aceite, Leche}, {Aceite, Yerba}, {Aceite, Cerveza}, {Aceite, Azúcar}, {Aceite, Sal},
 {Leche, Yerba}, {Leche, Cerveza}, {Leche, Azúcar}, {Leche, Sal},
 {Yerba, Cerveza}, {Yerba, Azúcar}, {Yerba, Sal}
 {Cerveza, Azúcar}, {Cerveza, Sal}
 $\}$

Soporte de cada regla:

$S_2 = \{$
 {Aceite, Leche}:3, {Aceite, Yerba}:3, {Aceite, Cerveza}:2, {Aceite, Azúcar}:2, {Aceite, Sal}:2,
 {Leche, Yerba}:3, {Leche, Cerveza}:2, {Leche, Azúcar}:1, {Leche, Sal}:3,
 {Yerba, Cerveza}:3, {Yerba, Azúcar}:1, {Yerba, Sal}:3,
 {Cerveza, Azúcar}:1, {Cerveza, Sal}:2
 $\}$

se quitan aquellas reglas que no llegan a cumplir con el minsup solicitado, igual a 2, quedando aquellas que son iguales o mayores a ese valor:

$S_2 = \{$
 {Aceite, Leche}:3, {Aceite, Yerba}:3, {Aceite, Cerveza}:2, {Aceite, Azúcar}:2, {Aceite, Sal}:2,
 {Leche, Yerba}:3, {Leche, Cerveza}:2, {Leche, Sal}:3,
 {Yerba, Cerveza}:3, {Yerba, Sal}:3,
 {Cerveza, Sal}:2
 $\}$

Combinaciones de 3 ítems con itemset (k=3)::

$S_3 = \{$
 {Aceite, Leche, Azúcar}:1,
 {Aceite, Yerba, Cerveza}:2, {Aceite, Yerba, Azúcar}:1, {Aceite, Cerveza, Azúcar}:1,
 {Leche, Yerba, Cerveza}:2, {Leche, Yerba, Sal}:3, {Leche, Cerveza, Sal}:2,
 {Aceite, Leche, Yerba}:2, {Aceite, Leche, Cerveza}:1, {Aceite, Leche, Sal}:2,
 {Aceite, Yerba, Sal}:2,
 {Aceite, Cerveza, Sal}:1, {Yerba, Cerveza, Sal}:2,
 $\}$

Y aquellas reglas que cumplen con el minsup de 2, son:

$S_3 = \{$
 {Aceite, Yerba, Cerveza}:2,
 {Leche, Yerba, Cerveza}:2, {Leche, Yerba, Sal}:3, {Leche, Cerveza, Sal}:2,
 {Aceite, Leche, Yerba}:2, {Aceite, Leche, Sal}:2,
 {Aceite, Yerba, Sal}:2,
 {Yerba, Cerveza, Sal}:2,

}

Combinaciones de 4 ítems con itemset (k=4)::

S4 = {

{Aceite, Azúcar, Yerba, Cerveza}:1,

{Leche, Yerba, Cerveza, Sal}:2,

{Aceite, Leche, Yerba, Cerveza}:1, {Aceite, Leche, Yerba, Sal}:2,

{Aceite, Leche, Cerveza, Sal}:1, {Aceite, Yerba, Cerveza, Sal}:1,

}

Quedando las de soporte:

S4 = {

{Leche, Yerba, Cerveza, Sal}:2,

{Aceite, Leche, Yerba, Sal}:2,

}

Combinaciones de 5 ítems solo se da una regla, itemset (k=5), que no configura con el soporte mínimo:

S5 = {{Aceite, Leche, Yerba, Cerveza, Sal}:1}

De las reglas que cumplen con el soporte mínimo establecido, tenemos a los conjuntos conformados por 2, 3 y 4 ítems:

Sfinal = S2 U S3 U S4 =

{

{Aceite, Leche}:3, {Aceite, Yerba}:3, {Aceite, Cerveza}:2, {Aceite, Azúcar}:2, {Aceite, Sal}:2,

{Leche, Yerba}:3, {Leche, Cerveza}:2, {Leche, Sal}:3,

{Yerba, Cerveza}:3, {Yerba, Sal}:3,

{Cerveza, Sal}:2

{Aceite, Yerba, Cerveza}:2,

{Leche, Yerba, Cerveza}:2, {Leche, Yerba, Sal}:3, {Leche, Cerveza, Sal}:2,

{Aceite, Leche, Yerba}:2, {Aceite, Leche, Sal}:2,

{Aceite, Yerba, Sal}:2,

{Yerba, Cerveza, Sal}:2,

{Leche, Yerba, Cerveza, Sal}:2,

{Aceite, Leche, Yerba, Sal}:2,

}

B – Generación de Reglas

A partir de Sfinal de los soportes mínimos, calculamos la confianza para luego obtener aquellas reglas candidatas que cumplan al menos con el soporte y confianza en sus valores mínimos.

$$\text{soporte } \{X \rightarrow Y\} = \underline{X \cup Y} \quad \text{confianza } \{X \rightarrow Y\} = \underline{X \cup Y}$$

$$N \qquad \qquad \qquad X$$

Itemsets frecuentes	Soporte mínimo	Confianza
{Aceite, Leche}	3	$\text{conf}(\{\text{Aceite, Leche}\}) = \text{supp}(\{\text{Aceite, Leche}\}) / \text{supp}(\{\text{Aceite}\}) = (3/5) / (4/5) = 3/4 = 0.75$
{Aceite, Yerba}	3	$\text{conf}(\{\text{Aceite, Yerba}\}) = \text{supp}(\{\text{Aceite, Yerba}\}) / \text{supp}(\{\text{Aceite}\}) = (3/5) / (4/5) = 3/4 = 0.75$
{Aceite, Cerveza}	2	$\text{conf}(\{\text{Aceite, Cerveza}\}) = \text{supp}(\{\text{Aceite, Cerveza}\}) / \text{supp}(\{\text{Aceite}\}) = (2/5) / (3/5) = 2/3 = 0.66$
{Aceite, Azúcar}	2	$\text{conf}(\{\text{Aceite, Azúcar}\}) = \text{supp}(\{\text{Aceite, Azúcar}\}) / \text{supp}(\{\text{Aceite}\}) = (2/5) / (4/5) = 2/4 = 0.50$
{Aceite, Sal}	2	$\text{conf}(\{\text{Aceite, Sal}\}) = \text{supp}(\{\text{Aceite, Sal}\}) / \text{supp}(\{\text{Aceite}\}) = (2/5) / (4/5) = 2/3 = 0.50$
{Leche, Yerba}	3	$\text{conf}(\{\text{Leche, Yerba}\}) = \text{supp}(\{\text{Leche, Yerba}\}) / \text{supp}(\{\text{Leche}\}) = (3/5) / (4/5) = 3/4 = 0.75$
{Leche, Cerveza}	2	$\text{conf}(\{\text{Leche, Cerveza}\}) = \text{supp}(\{\text{Leche, Cerveza}\}) / \text{supp}(\{\text{Leche}\}) = (2/5) / (4/5) = 2/3 = 0.50$
{Leche, Sal}	3	$\text{conf}(\{\text{Leche, Sal}\}) = \text{supp}(\{\text{Leche, Sal}\}) / \text{supp}(\{\text{Leche}\}) = (3/5) / (4/5) = 3/4 = 0.75$
{Yerba, Cerveza}	3	$\text{conf}(\{\text{Yerba, Cerveza}\}) = \text{supp}(\{\text{Yerba, Cerveza}\}) / \text{supp}(\{\text{Yerba}\}) = (3/5) / (4/5) = 2/3 = 0.50$
{Yerba, Sal}	3	$\text{conf}(\{\text{Yerba, Sal}\}) = \text{supp}(\{\text{Yerba, Sal}\}) / \text{supp}(\{\text{Yerba}\}) = (2/5) / (4/5) = 2/4 = 0.50$
{Cerveza, Sal}	2	$\text{conf}(\{\text{Cerveza, Sal}\}) = \text{supp}(\{\text{Cerveza, Sal}\}) / \text{supp}(\{\text{Cerveza}\}) = (2/5) / (3/5) = 2/3 = 0.66$
{Aceite, Yerba, Cerveza}	2	$\text{conf}(\{\text{Aceite, Yerba, Cerveza}\}) = \text{supp}(\{\text{Aceite, Yerba, Cerveza}\}) / \text{supp}(\{\text{Aceite, Yerba}\}) = (2/5) / (3/5) = 2/3 = 0.66$
{Leche, Yerba, Cerveza}	2	$\text{conf}(\{\text{Leche, Yerba, Cerveza}\}) = \text{supp}(\{\text{Leche, Yerba, Cerveza}\}) / \text{supp}(\{\text{Leche, Yerba}\}) = (2/5) / (3/5) = 2/3 = 0.66$
{Leche, Yerba, Sal}	3	$\text{conf}(\{\text{Leche, Yerba, Sal}\}) = \text{supp}(\{\text{Leche, Yerba, Sal}\}) / \text{supp}(\{\text{Leche, Yerba}\}) = (3/5) / (3/5) = 1$
{Leche, Cerveza, Sal}	2	$\text{conf}(\{\text{Leche, Cerveza, Sal}\}) = \text{supp}(\{\text{Leche, Cerveza, Sal}\}) / \text{supp}(\{\text{Leche, Cerveza}\}) = (2/5) / (2/5) = 1$
{Aceite, Leche, Yerba}	2	$\text{conf}(\{\text{Aceite, Leche, Yerba}\}) = \text{supp}(\{\text{Aceite, Leche, Yerba}\}) / \text{supp}(\{\text{Aceite, Leche}\}) = (2/5) / (3/5) = 2/3 = 0.66$
{Aceite, Leche, Sal}	2	$\text{conf}(\{\text{Aceite, Leche, Sal}\}) = \text{supp}(\{\text{Aceite, Leche, Sal}\}) / \text{supp}(\{\text{Aceite, Leche}\}) = (2/5) / (3/5) = 2/3 = 0.66$
{Aceite, Yerba, Sal}	2	$\text{conf}(\{\text{Aceite, Yerba, Sal}\}) = \text{supp}(\{\text{Aceite, Yerba, Sal}\}) / \text{supp}(\{\text{Aceite, Yerba}\}) = (2/5) / (2/5) = 1$
{Yerba, Cerveza, Sal}	2	$\text{conf}(\{\text{Yerba, Cerveza, Sal}\}) = \text{supp}(\{\text{Yerba, Cerveza, Sal}\}) / \text{supp}(\{\text{Yerba, Cerveza}\}) = (2/5) / (3/5) = 2/3 = 0.66$
{Leche, Yerba, Cerveza, Sal}	2	$\text{conf}(\{\text{Leche, Yerba, Cerveza, Sal}\}) = \text{supp}(\{\text{Leche, Yerba, Cerveza, Sal}\}) / \text{supp}(\{\text{Leche, Yerba, Cerveza}\}) = (2/5) / (2/5) = 1$
{Aceite, Leche, Yerba, Sal}	2	$\text{conf}(\{\text{Aceite, Leche, Yerba, Sal}\}) = \text{supp}(\{\text{Aceite, Leche, Yerba, Sal}\}) / \text{supp}(\{\text{Aceite, Leche, Yerba}\}) = (2/5) / (2/5) = 1$

Reglas de asociación que cumplen con los pisos mínimos de soporte = 2 y confianza = 0.75

Reglas	Soporte	Confianza
{Aceite, Leche}	3	0.75
{Aceite, Yerba}	3	0.75
{Leche, Yerba}	3	0.75
{Leche, Sal}	3	0.75
{Yerba,Cerveza}	3	0.75
{Leche, Yerba, Sal}	3	1
{Leche, Cerveza, Sal}	2	1
{Aceite, Yerba, Sal}	2	1
{Leche, Yerba, Cerveza, Sal}	2	1
{Aceite, Leche, Yerba, Sal}	2	1

3) Se quiere clasificar en dos grupos a un conjunto de personas de acuerdo a su estatura. Para cada persona se conoce su estatura expresada en centímetros.

- Clasificar cada persona como alta o baja, según el algoritmo On-Line k-MEANS

Clustering.

Número (Persona)	Altura
1	190
2	160
3	158
4	177
5	189
6	171
7	160
8	154
9	200
10	172

Aplicación del algoritmo de k means :

- 1) A partir de los datos proporcionados realizamos el promedio de los mismos

$$P = (190 + 160 + 158 + 177 + 189 + 171 + 160 + 154 + 200 + 172) / 10$$

$$= 1731 / 10$$

$$= 173.1$$

- 2) Ahora procedemos a clasificar según el valor de la altura sea mayor o menor al valor obtenido en el promedio en la que consideramos a una persona alta o baja.

- 3) Representación en tabla:

Numero Persona	Altura	Clasificación
1	190	alta
2	160	baja
3	158	baja
4	177	alta
5	189	alta
6	171	baja
7	160	baja

8	154	baja
9	200	alta
10	172	baja

4) En base a la siguiente tabla de Navigation Trails:

ID	Trail
1	$A_1 \rightarrow A_2 \rightarrow A_3 \rightarrow A_4$
2	$A_1 \rightarrow A_5 \rightarrow A_3 \rightarrow A_4 \rightarrow A_1$
3	$A_5 \rightarrow A_2 \rightarrow A_4 \rightarrow A_6$
4	$A_5 \rightarrow A_2 \rightarrow A_3$
5	$A_5 \rightarrow A_2 \rightarrow A_3 \rightarrow A_6$
6	$A_4 \rightarrow A_1 \rightarrow A_5 \rightarrow A_3$

- a) Represente en forma tabular el conjunto de producciones probabilísticas, para un $\alpha = 0,5$ y $N = 1$.
- b) Confeccione el grafo de transiciones con las probabilidades de producción correspondientes.

Los navigation trails son utilizados para la construcción de Gramática Probabilística de Hipertexto.

HPG es una tupla $\langle V, \Sigma, S, P \rangle$

Terminales y no terminales.

$$V = \{S, A1, A2, A3, A4, A5, A6, F\}$$

$$S = \{a1, a2, a3, a4, a5, a6\}$$

	Cantidad
Trails	6
Visitas totales	24

Enlace	Visitas	Inicio (Cantidad)
A1	4	2
A2	4	0
A3	5	0
A4	4	1
A5	5	3
A6	2	0

Con $\alpha = 0,5$ y $N=1$, podemos calcular la probabilidad de la producción

$p(S \rightarrow a1A1)$ que corresponde con la flecha de S a $A1$ en el grafo de transiciones de la siguiente manera:

$$p(S \rightarrow a1 A1) = (0,5 * 4)/24 + (0,5 * 2)/6 = 0,25$$

producciones probabilísticas de las diferentes etapas:

Inicio producción	(visitas de cada Nodo) + (veces Nodo al inicio)	Probabilidad
$S \rightarrow a1A1$	$p(S \rightarrow a1 A1) = (0,5 * 4)/24 + (0,5 * 2)/6$	0.25
$S \rightarrow a2A2$	$p(S \rightarrow a2 A2) = (0,5 * 4)/24$	0.08

$S \rightarrow a3A3$	$p(S \rightarrow a3 A3) = (0,5 * 5)/24$	0.11
$S \rightarrow a4A4$	$p(S \rightarrow a4 A4) = (0,5 * 4)/24 + (0,5 * 1)/6$	0.17
$S \rightarrow a5A5$	$p(S \rightarrow a5 A5) = (0,5 * 5)/24 + (0,5 * 3)/6$	0.35
$S \rightarrow a6A6$	$p(S \rightarrow a6 A6) = (0,5 * 2)/24$	0.35

Transición	Visitas Origen → Destino	Probabilidad
$A1 \rightarrow a2 A2$	$p(A1 \rightarrow a2 A2) = 1/4$	0.25
$A1 \rightarrow a5 A5$	$p(A1 \rightarrow a5 A5) = 2/4$	0.50
$A2 \rightarrow a3 A3$	$p(A2 \rightarrow a3 A3) = 3/4$	0.75
$A2 \rightarrow a4 A4$	$p(A2 \rightarrow a4 A4) = 1/4$	0.25
$A3 \rightarrow a4 A4$	$p(A3 \rightarrow a4 A4) = 2/5$	0.40
$A3 \rightarrow a6 A6$	$p(A3 \rightarrow a6 A6) = 1/5$	0.20
$A4 \rightarrow a1 A1$	$p(A4 \rightarrow a1 A1) = 2/4$	0.50
$A4 \rightarrow a6 A6$	$p(A4 \rightarrow a6 A6) = 1/4$	0.25
$A5 \rightarrow a2 A2$	$p(A5 \rightarrow a2 A2) = 3/5$	0.60
$A5 \rightarrow a3 A3$	$p(A5 \rightarrow a3 A3) = 2/5$	0.40

Final Producción	Nodo al final	Probabilidad
$A1 \rightarrow F$	$p(A1 \rightarrow F) = 1/4$	0.25
$A3 \rightarrow F$	$p(A3 \rightarrow F) = 2/5$	0.40
$A4 \rightarrow F$	$p(A4 \rightarrow F) = 1/4$	0.25
$A6 \rightarrow F$	$p(A6 \rightarrow F) = 2/2$	1.00

Inicio producción	Probabilidad	Transición	Probabilidad	Final Producción	Probabilidad
$S \rightarrow a1A1$	0.25	$A1 \rightarrow a2 A2$	0.25	$A1 \rightarrow F$	0.25
$S \rightarrow a2A2$	0.08	$A1 \rightarrow a5 A5$	0.50	$A3 \rightarrow F$	0.40
$S \rightarrow a3A3$	0.11	$A2 \rightarrow a3 A3$	0.75	$A4 \rightarrow F$	0.25
$S \rightarrow a4A4$	0.17	$A2 \rightarrow a4 A4$	0.25	$A6 \rightarrow F$	1.00
$S \rightarrow a5A5$	0.35	$A3 \rightarrow a4 A4$	0.40		
$S \rightarrow a6A6$	0.35	$A3 \rightarrow a6 A6$	0.20		
		$A4 \rightarrow a1 A1$	0.50		
		$A4 \rightarrow a6 A6$	0.25		
		$A5 \rightarrow a2 A2$	0.60		
		$A5 \rightarrow a3 A3$	0.40		

b) Representación del grafo

