

Trabajo de Adscripción

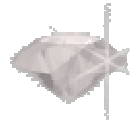


Profesores Orientadores:

La Red Martínez, David Luis
Quintana. Osvaldo Pantaleón

Alumno:

Cutro. Luis Alfonso LU N°:29027



Esquema de la Presentación



Marco Conceptual y Motivaciones

- ❑ Encuesta Permanente de Hogares.
- ❑ Sociedad de la Información y del Conocimiento.
- ❑ Inteligencia de Negocios (Business Intelligence).
- ❑ Minería de Datos (Data Mining).



Herramientas de Software

- ❑ DB2 UDB
- ❑ DB2 Intelligent Miner for Data
- ❑ WebSphere



Elaboración del Data Warehouse

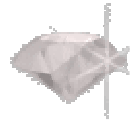


Resultado Obtenidos

- ❑ Técnica empleada : “Clustering”.
- ❑ Técnica empleada: “Árboles de Decisión y Clasificación”.



Conclusiones

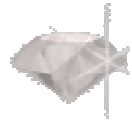


Marco Conceptual y Motivaciones

- ❖ ¿Qué es la *E P H*?
- ❖ ¿Quién la realiza?
- ❖ ¿Dónde se aplica?
- ❖ ¿Con qué frecuencia?



INDEC

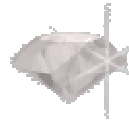


Marco Conceptual y Motivaciones

Sociedad de la Información y del Conocimiento.

- ❖ En las SIC se está produciendo un fenómeno curioso:
El aumento de datos no supone un aumento del conocimiento.
- ❖ Para enfrentar estos problemas, han surgido una serie de técnicas que facilitan el procesamiento avanzado de los datos:
Extracción de conocimiento de bases de datos (KDD).

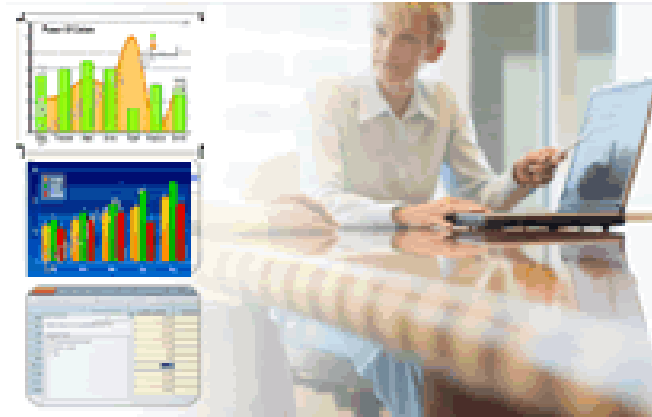




Marco Conceptual y Motivaciones

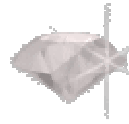
Inteligencia de Negocios (Business Intelligence).

El término Inteligencia de Negocios (Business Intelligence) se refiere al uso de los datos de una organización para facilitar la toma de decisiones.



- (Business Intelligence), incluye actividades como:
- Procesamiento Analítico en Línea (OLAP).
 - Data Warehouse (DW).
 - Minería de Datos (MD).





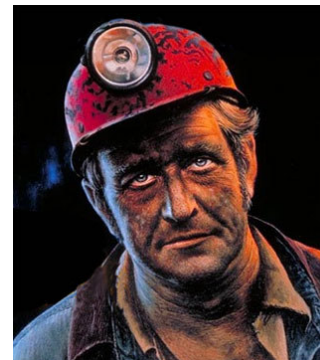
Marco Conceptual y Motivaciones

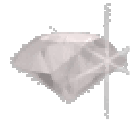
Minería de Datos (MD)

La minería de datos (DM, Data Mining) consiste en la extracción no trivial de conocimiento que reside de manera implícita en los datos.

Las Técnicas de minería de datos son las siguientes:

- Redes Neuronales.
- Árboles de Decisión.
- Modelos Estadísticos.
- Agrupamiento o Clustering.
- Etc.





Marco Conceptual y Motivaciones

Minería de Datos (MD)



Aplicaciones de la Minería de Datos

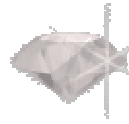
En el mundo de los Negocios

- Compañías de Seguros
- Marketing
- Etc.

En el mundo de las Ciencias

- Meteorología
- Bioinformática
- Etc.

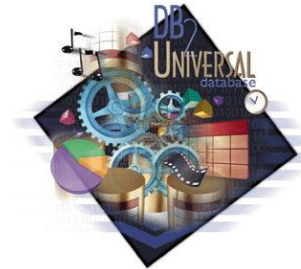




Herramientas de Software

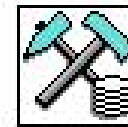
DB2 UDB

- ❑ Instalación del Ambiente Operacional.
- ❑ Instalación del Ambiente Datamart.



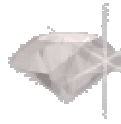
DB2 Intelligent Miner for Data

- ❑ Utilización de Clustering.
- ❑ Utilización de Árboles de Decisión.



WebSphere





Herramientas de Software



IBM DB2 UDB

IBM DB2 UDB :
Visualización de la
ventana principal

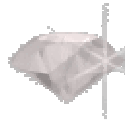
Centro Control

*Centro de
depósito de datos*

Bases de datos

The screenshot shows the IBM DB2 UDB Control Center interface. The left pane displays a tree view of the database environment, including 'Sistemas', 'Instancias', 'DB2', 'Bases de datos', and 'EPH'. The right pane shows a table list for the 'HOGAR-Q010T10KJ - DB2 - EPH' database. The table list has columns for 'Nombre', 'Esquema', 'Tipo', 'Espacio de tabla', and 'Espacio'. The status bar at the bottom indicates '98 de 98 elementos visualizados'.

Nombre	Esquema	Tipo	Espacio de tabla	Espacio
SYATTRIBUTES	SYSIBM	T	SYSCATSPACE	
SYBUFFERPOOLNODES	SYSIBM	T	SYSCATSPACE	
SYBUFFERPOOLS	SYSIBM	T	SYSCATSPACE	
SYSCHECKS	SYSIBM	T	SYSCATSPACE	
SYSCODEPROPERTIES	SYSIBM	T	SYSCATSPACE	
SYSCOLAUTH	SYSIBM	T	SYSCATSPACE	
SYSCOLCHECKS	SYSIBM	T	SYSCATSPACE	
SYSCOLDIST	SYSIBM	T	SYSCATSPACE	
SYSCOLGROUPDIST	SYSIBM	T	SYSCATSPACE	
SYSCOLGROUPDISTCOUNTS	SYSIBM	T	SYSCATSPACE	
SYSCOLGROUPS	SYSIBM	T	SYSCATSPACE	
SYSCOLGROUPSCOLS	SYSIBM	T	SYSCATSPACE	
SYSCOLOPTIONS	SYSIBM	T	SYSCATSPACE	
SYSCOLPROPERTIES	SYSIBM	T	SYSCATSPACE	
SYSCOLUMNS	SYSIBM	T	SYSCATSPACE	
SYSCOLUSE	SYSIBM	T	SYSCATSPACE	
SYSCOMMENTS	SYSIBM	T	SYSCATSPACE	
SYSCONSTDEP	SYSIBM	T	SYSCATSPACE	
SYSDATATYPES	SYSIBM	T	SYSCATSPACE	



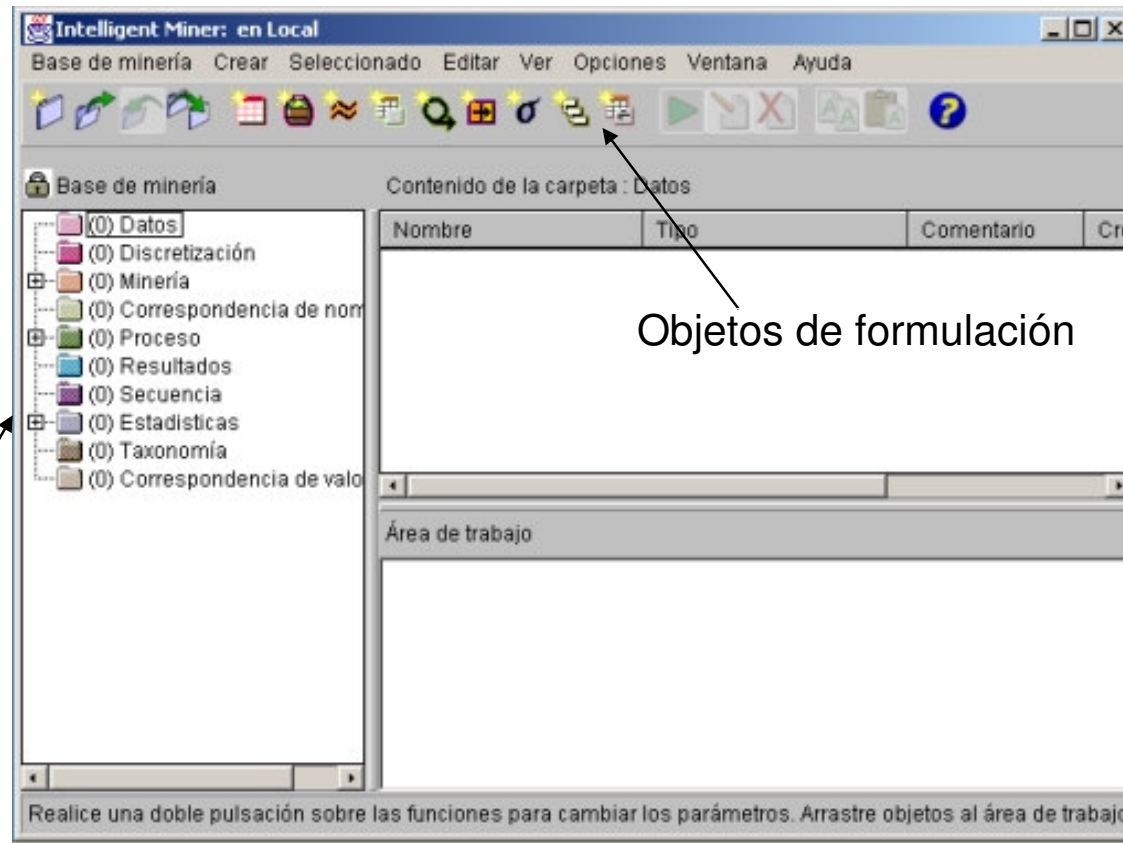
Herramientas de Software



IBM DB2 Intelligent Miner for Data

IBM
INTELLIGENT
MINER FOR
DATA V 8.1:
Visualización de la
ventana principal


Bases de minería



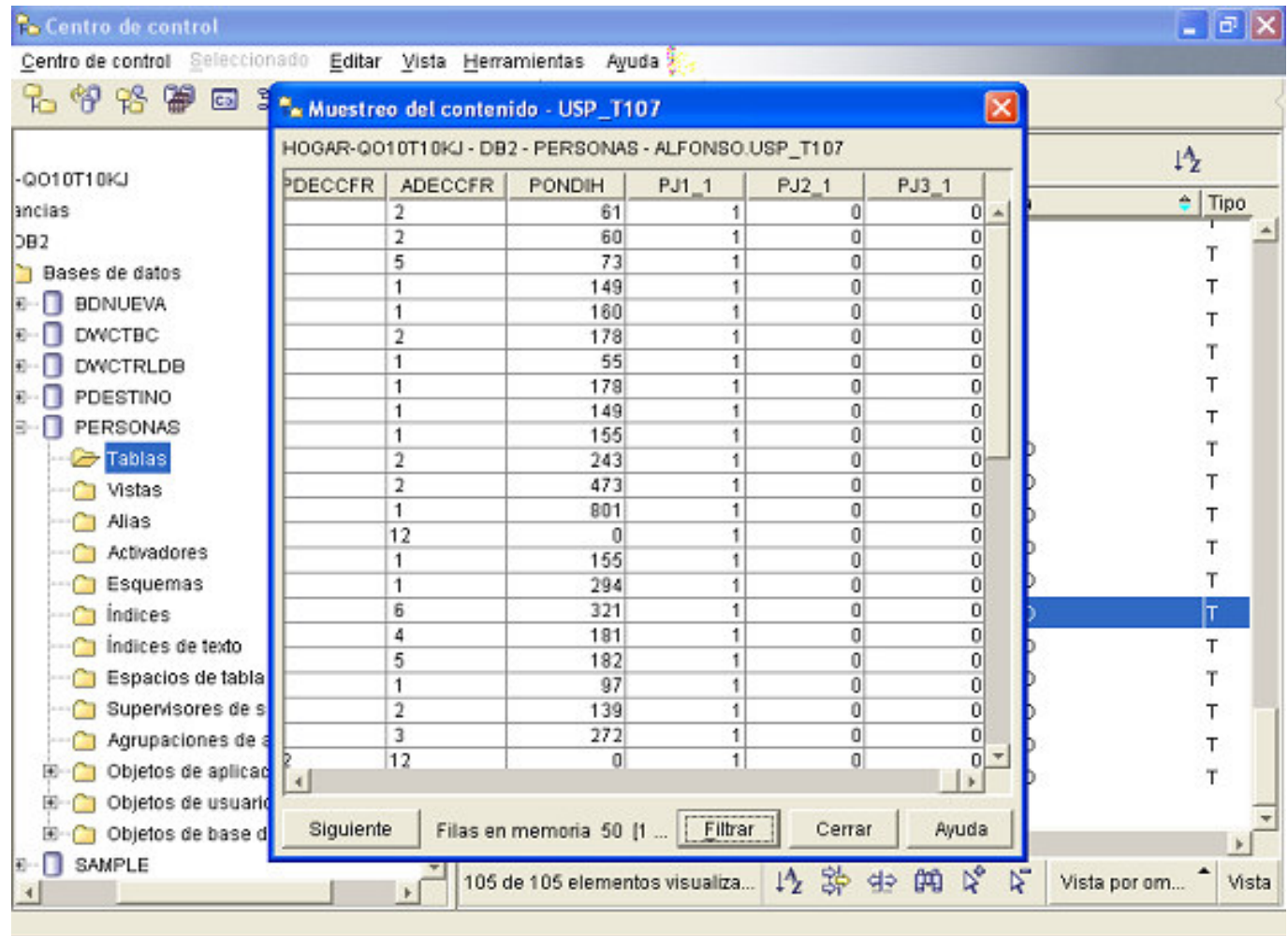


Elaboración del Data Warehouse



Los pasos que se llevan a cabo en el ambiente de trabajo  *Centro Control* :

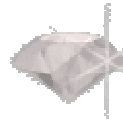
- Creación de la base de datos EPH.
- Creación de la tabla USP.
- Importación de datos.
(Access → IBM DB2).
- Visualización



The screenshot shows the IBM DB2 Control Center interface. The main window displays a table titled 'Muestreo del contenido - USP_T107' with the following data:

PDECCFR	ADECCFR	PONDIH	PJ1_1	PJ2_1	PJ3_1	Tipo
2	61	1	0	0	T	
2	60	1	0	0	T	
5	73	1	0	0	T	
1	149	1	0	0	T	
1	180	1	0	0	T	
2	178	1	0	0	T	
1	55	1	0	0	T	
1	178	1	0	0	T	
1	149	1	0	0	T	
1	155	1	0	0	T	
2	243	1	0	0	T	
2	473	1	0	0	T	
1	801	1	0	0	T	
12	0	1	0	0	T	
1	155	1	0	0	T	
1	294	1	0	0	T	
6	321	1	0	0	T	
4	181	1	0	0	T	
5	182	1	0	0	T	
1	97	1	0	0	T	
2	139	1	0	0	T	
3	272	1	0	0	T	
12	0	1	0	0	T	

The interface also shows a tree view on the left with folders like 'Bases de datos', 'Tablas', 'Vistas', etc., and a status bar at the bottom indicating '105 de 105 elementos visualiza...'. The table has a 'Tipo' column with all values set to 'T'.

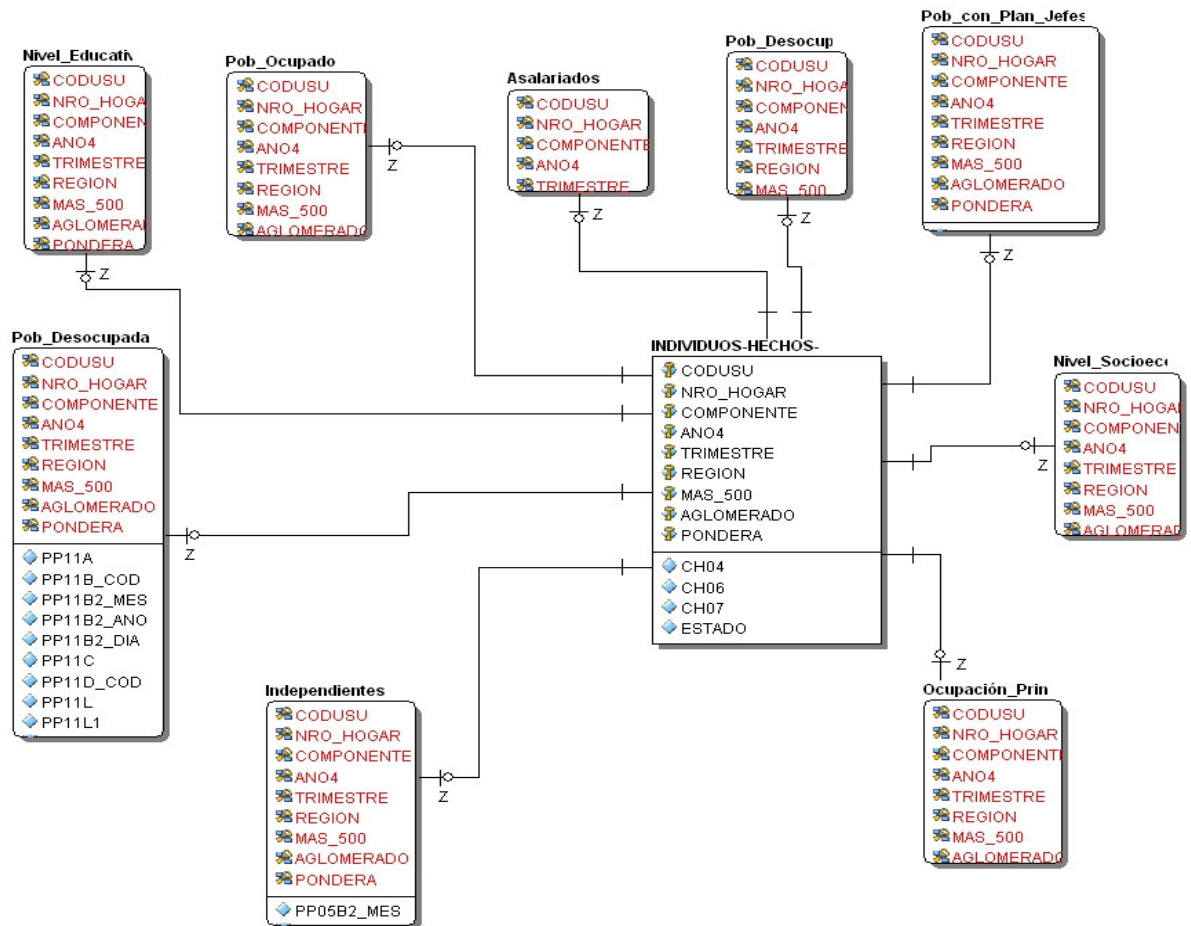


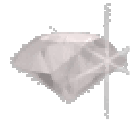
Elaboración del Data Warehouse



En el ambiente de trabajo del *Centro de depósito de datos* se deberá definir:

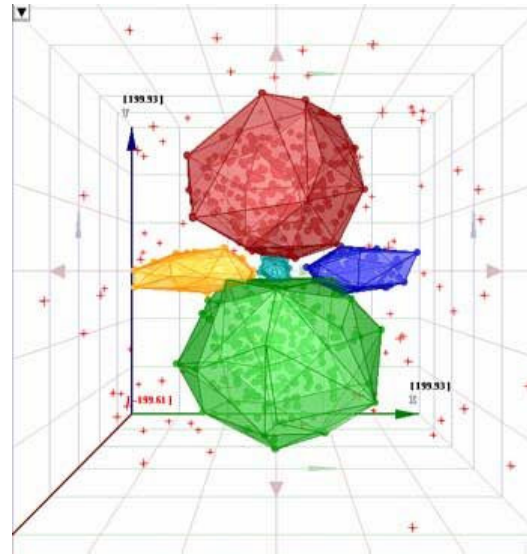
- Fuentes de depósitos.
(La tabla USP)
- Destino de depósitos.
(tablas de Dimensiones)
- Esquemas de depósitos.
(Esquema en Estrella)
- Administración.





Resultado Obtenidos

Técnica empleada : “Clustering”

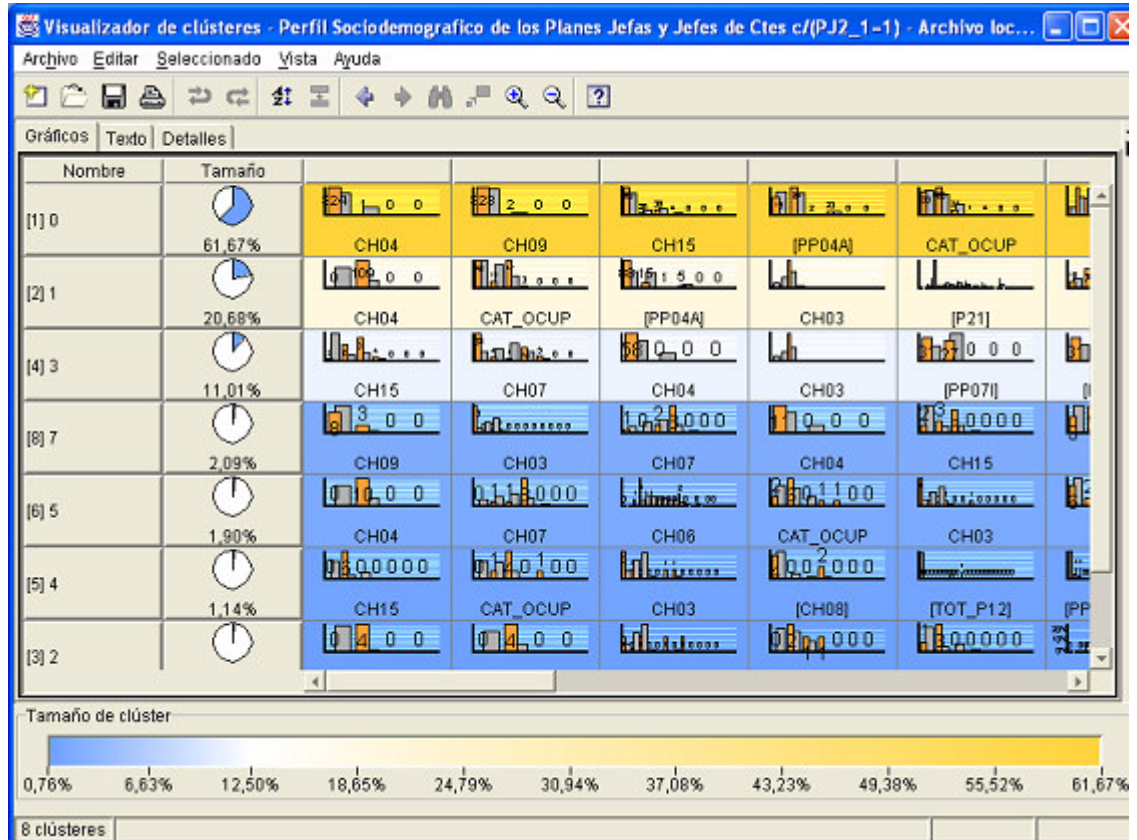


 *Conocer los Perfiles Socio Demográficos de los Planes Jefes y Jefas.*

Resultado Obtenidos

Se nota la existencia de 8 clusters identificados por la ejecución de minería de datos.

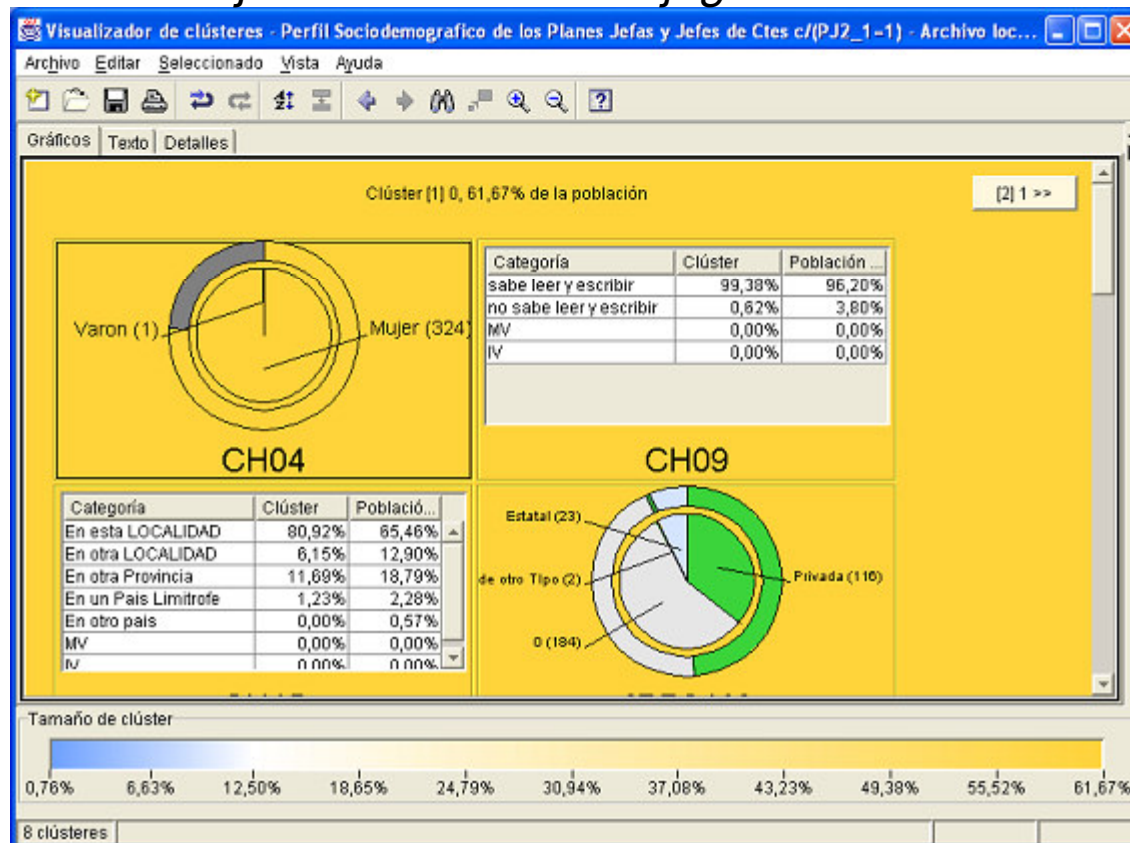
Donde prácticamente un 93,36% de la población está representada sólo por estos tres primeros clústeres, dividiéndose el 6,64% restante entre los demás.

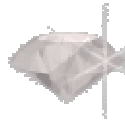


Resultado Obtenidos

Visualización del cada Clúster N°1 con 57.89%:

El primer grupo está representado por una población en su mayoría formada por mujeres, de 25 a 30 años de edad, que son residentes de Corrientes Capital y se encuentran unidas o juntas con su conyugé.

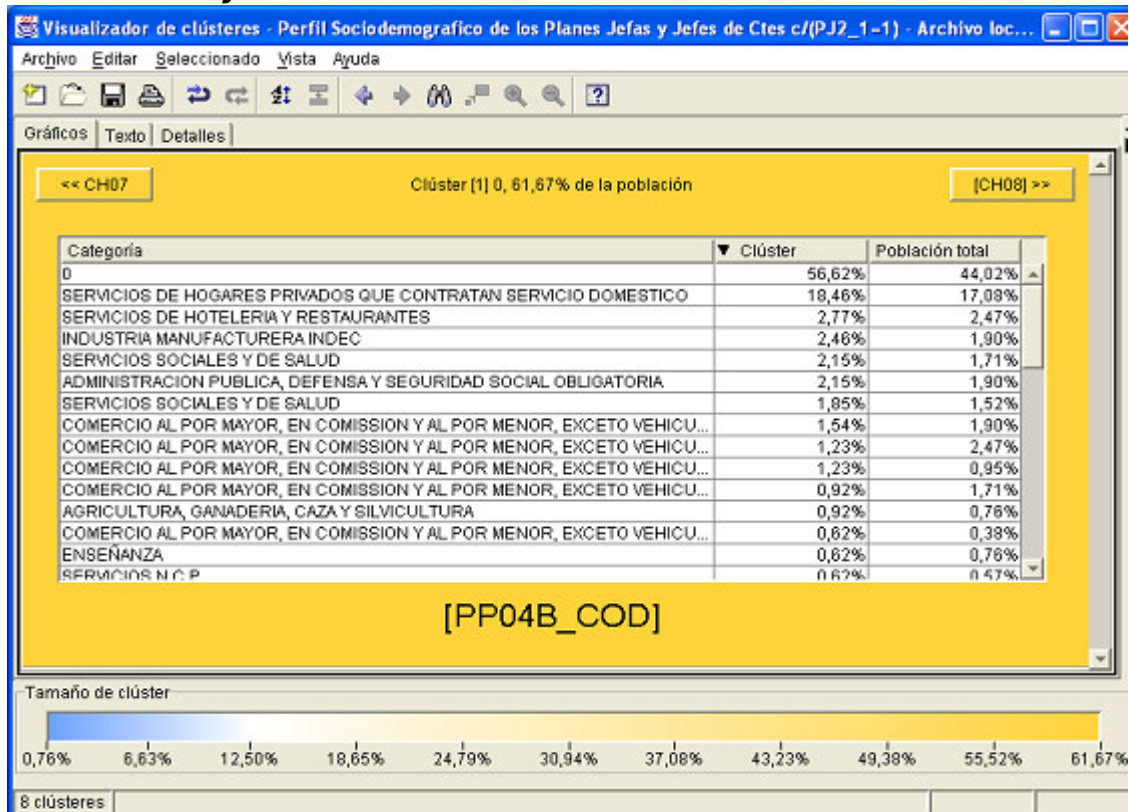


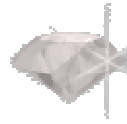


Resultado Obtenidos

Visualización del cada Clúster N°1 con 57.89%:

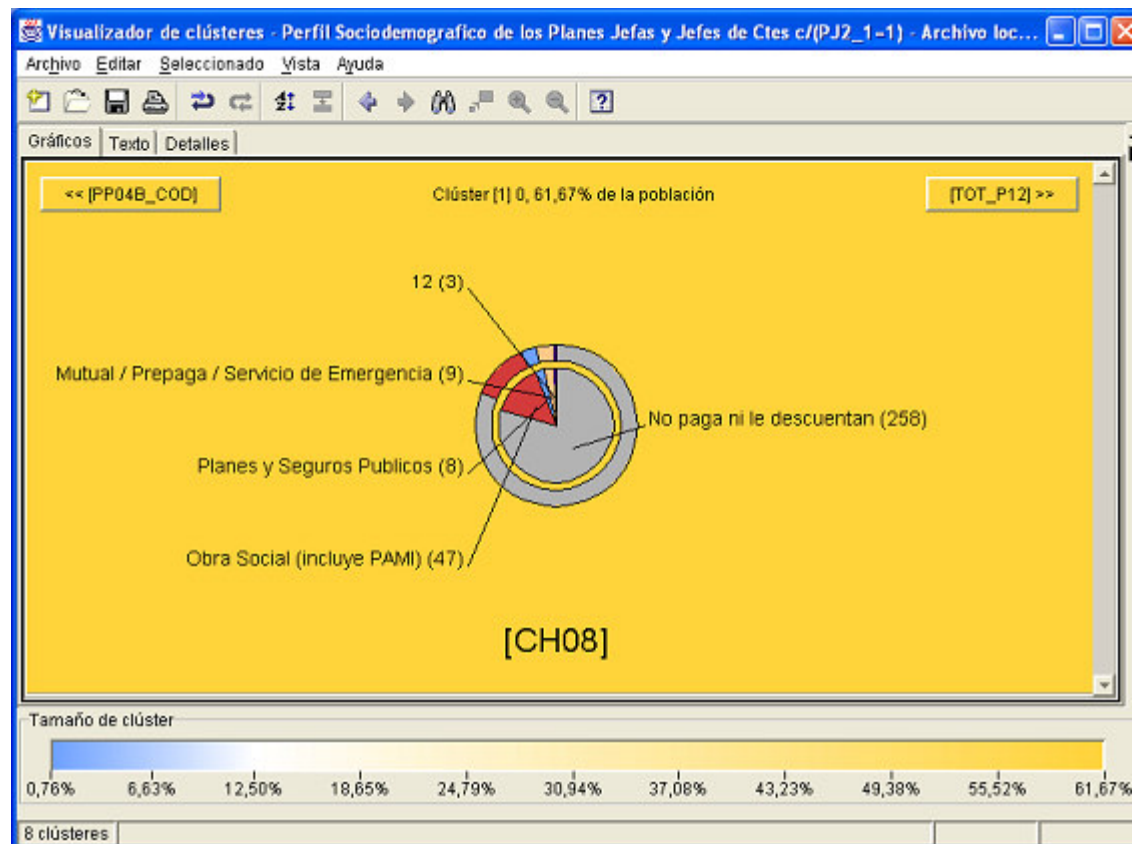
Con respecto a lo laboral, estas personas trabajan en hogares privados como servicio doméstico donde no paga ni le descuentan mensualmente una cobertura médica, tampoco tiene contrato de trabajo ni obra social y mucho menos descuento jubilatorio.

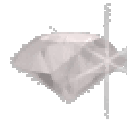




Resultado Obtenidos

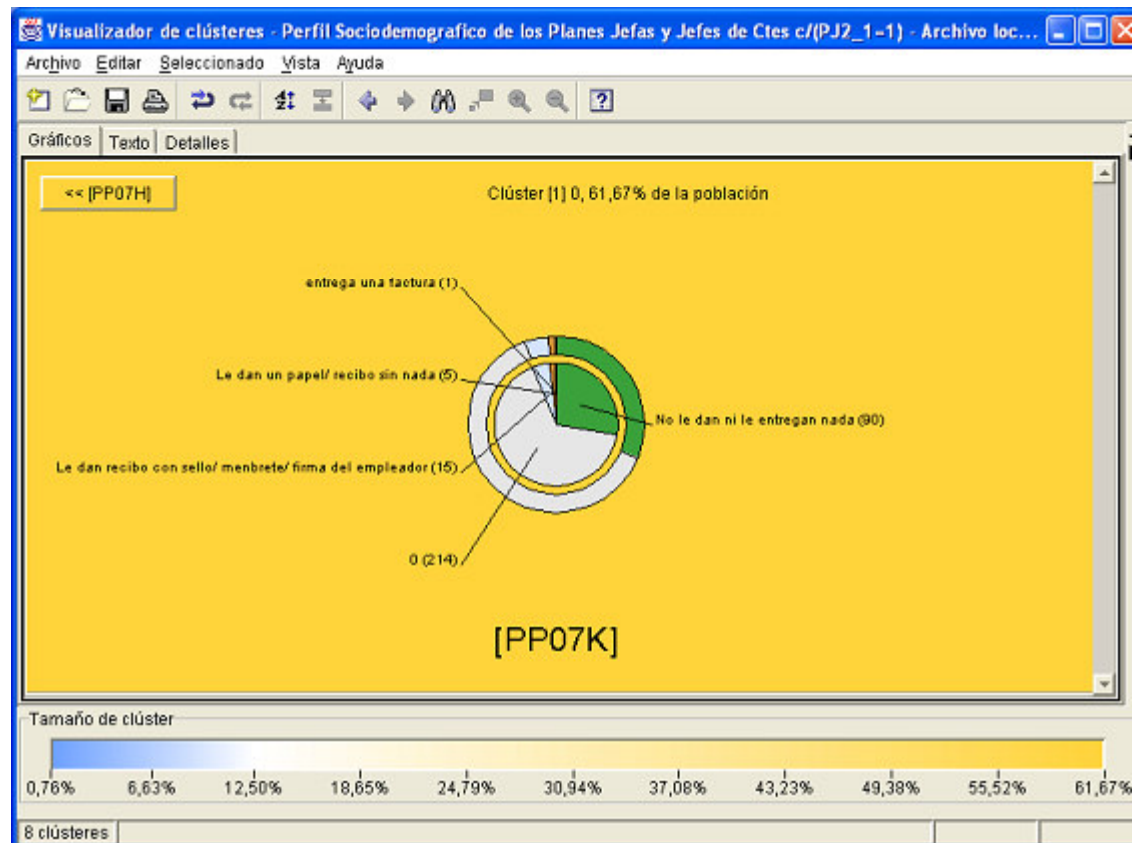
Visualización del cada Clúster N°1 con 57.89%:
*No paga ni le descuentan mensualmente una **cobertura médica**.*

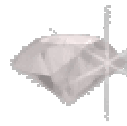




Resultado Obtenidos

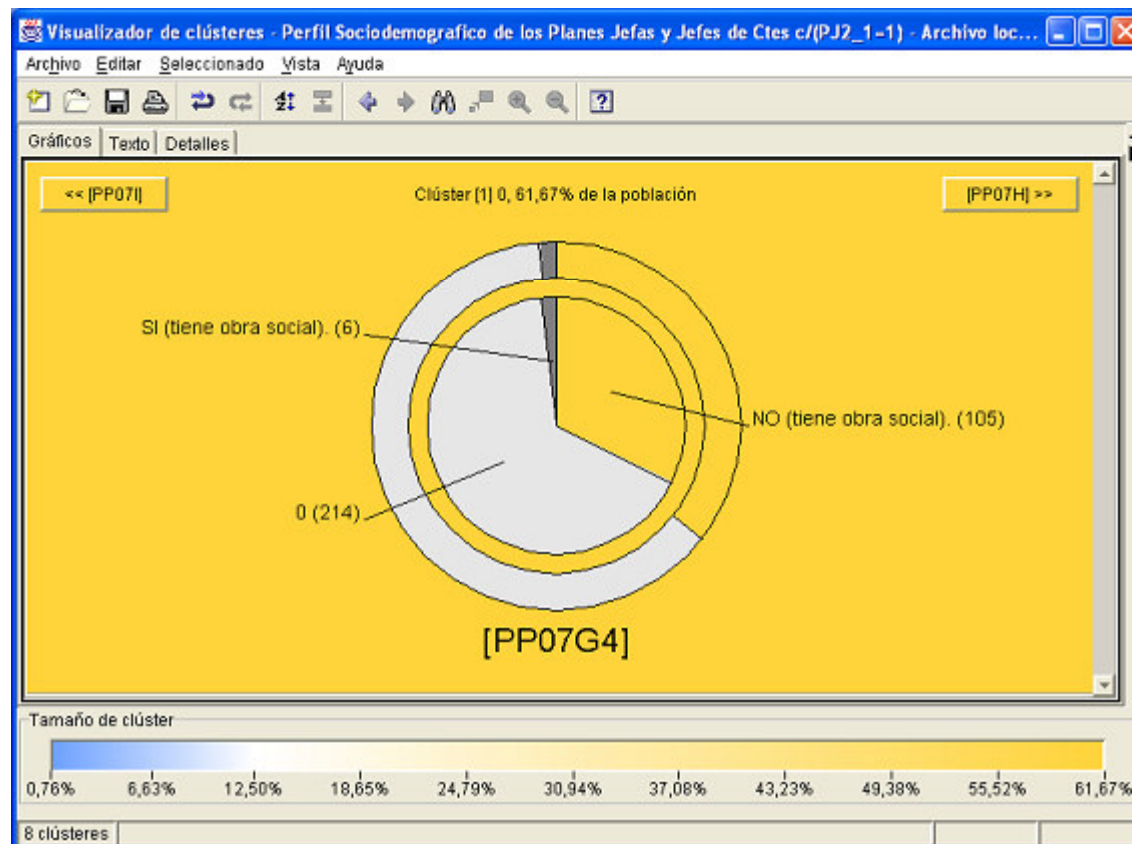
Visualización del cada Clúster N°1 con 57.89%:
*No poseen **contrato de trabajo** ni obra social.*

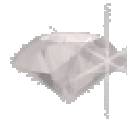




Resultado Obtenidos

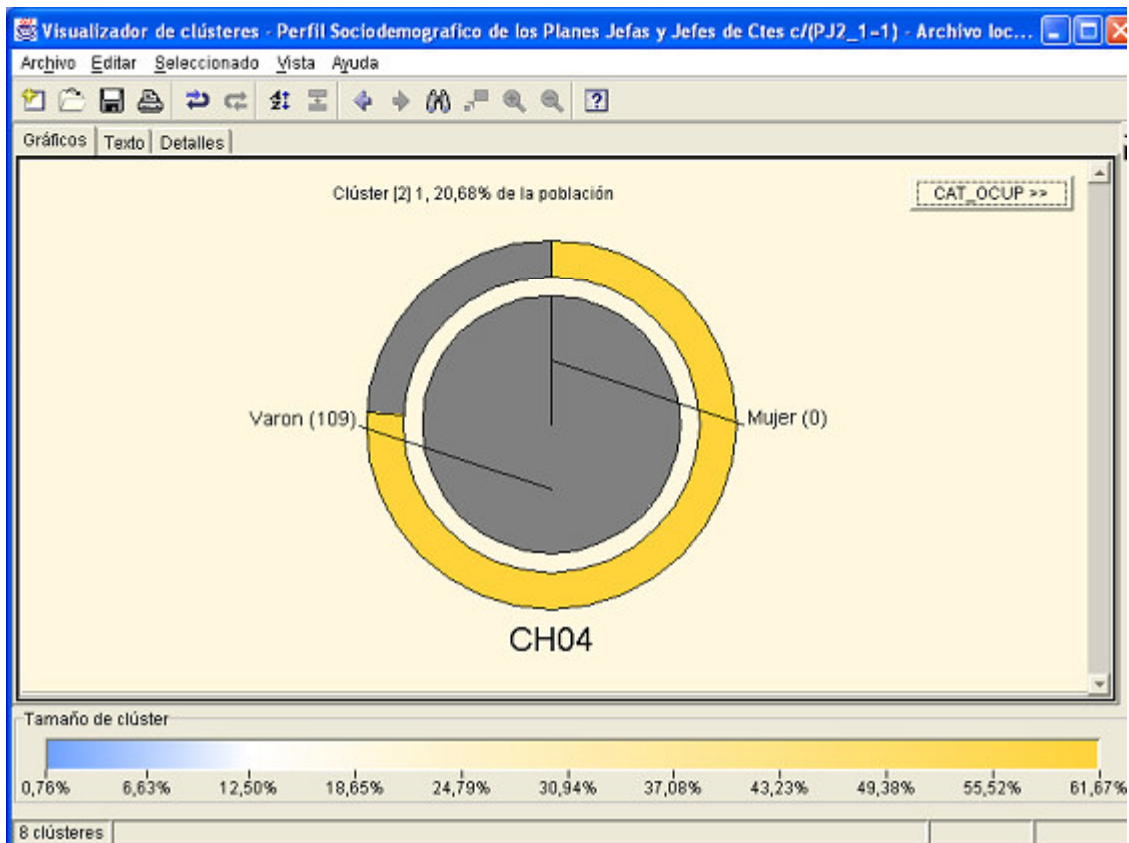
Visualización del cada Clúster N°1 con 57.89%:
*No poseen contrato de trabajo ni **obra social**.*

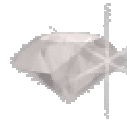




Resultado Obtenidos

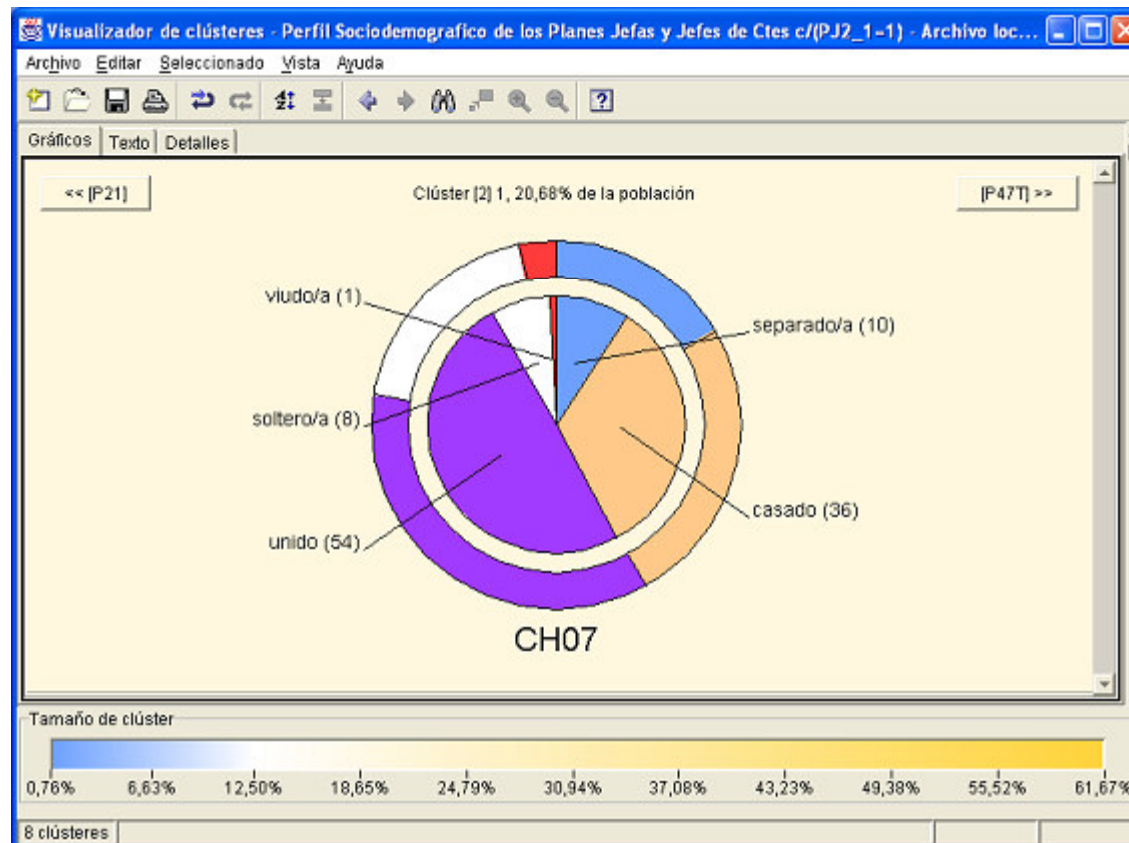
Visualización del cada Clúster N°2 con 20,68%:
*El sexo predominantemente es el **masculino***

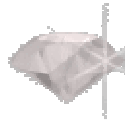




Resultado Obtenidos

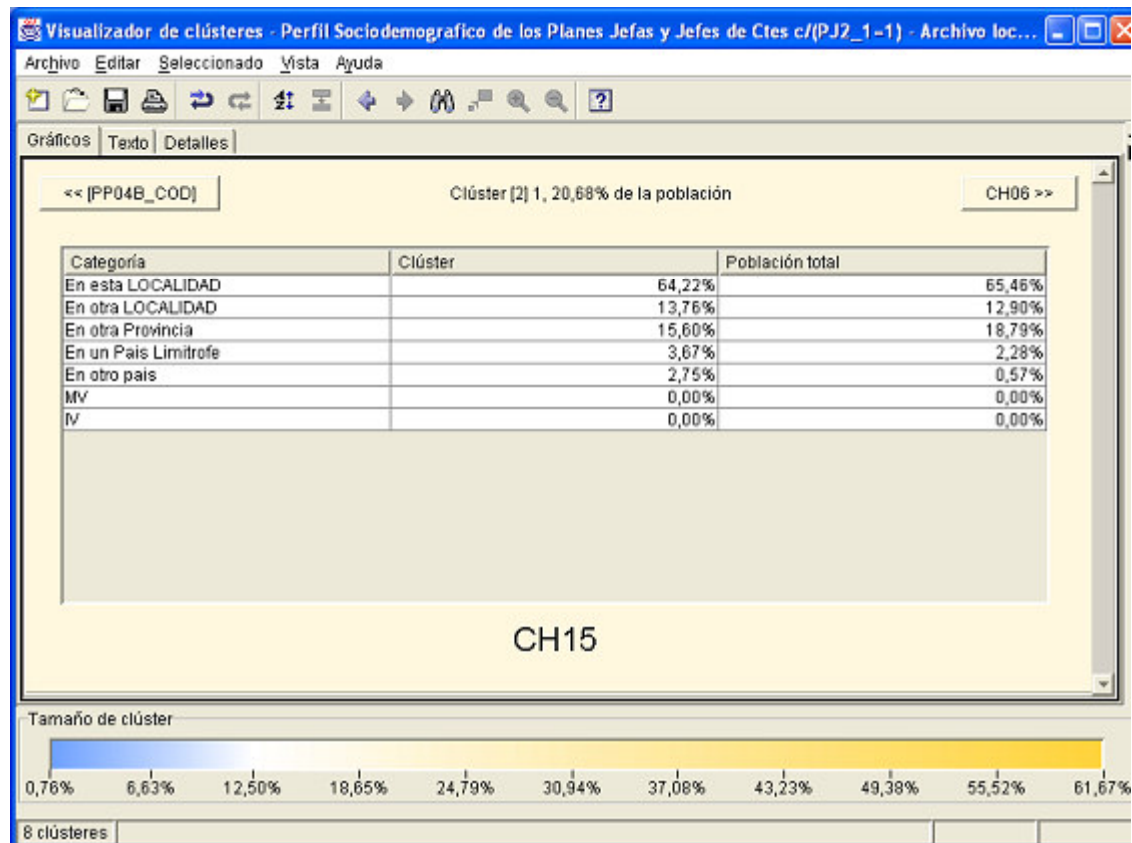
Visualización del cada Clúster N°2 con 20,68%:
El estado civil de unido y con una edad sobresaliente de 46 años.

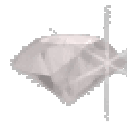




Resultado Obtenidos

Visualización del cada Clúster N°2 con 20,68%:
Sin diferenciarse con el primer clúster, en este en su mayoría siguen siendo de esta localidad o sea Corrientes Capital.

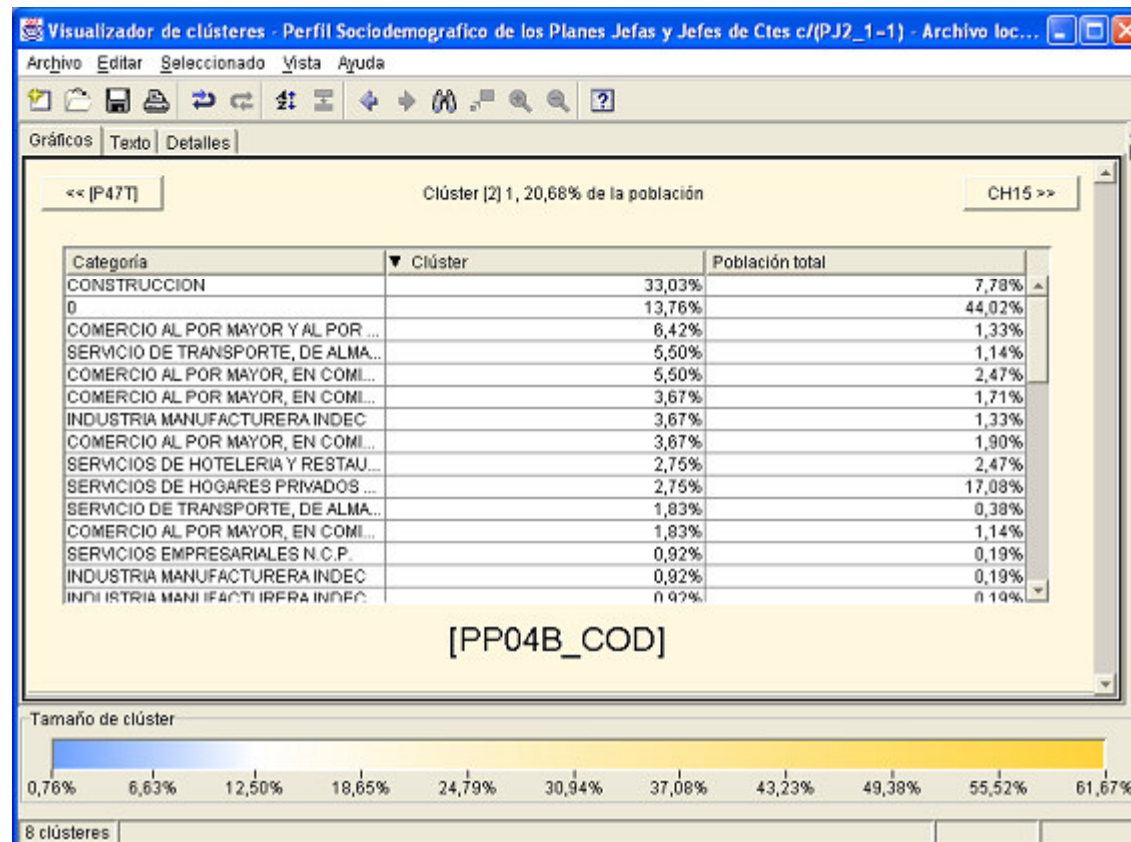


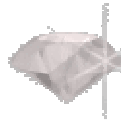


Resultado Obtenidos

Visualización del cada Clúster N°2 con 20,68%:

*A diferencia con el primer clúster, en este los individuos se dedican al rubro de la **construcción**.*

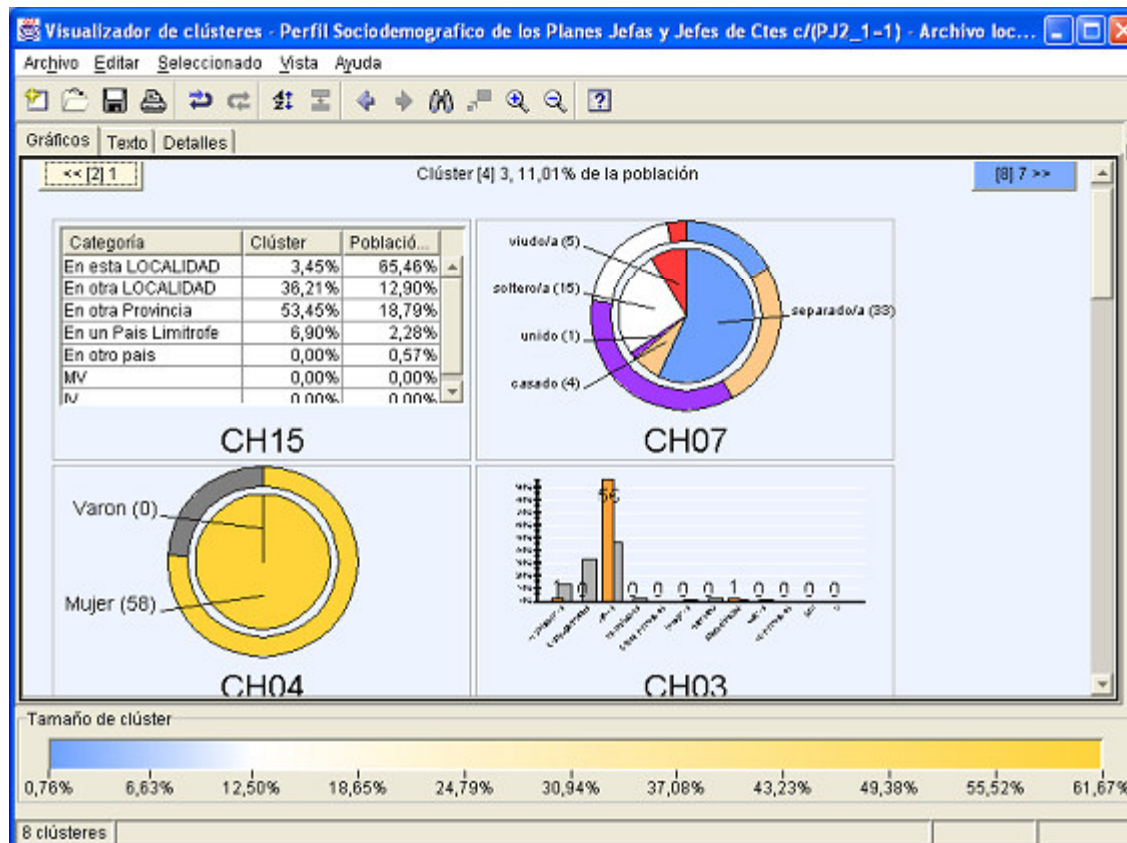


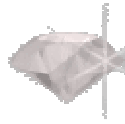


Resultado Obtenidos

Visualización del cada Clúster N^o3 con 11,01 %:

En este clúster se tiene como predominante a la mujer en la variable sexo la misma es separada con una edad que ronda los 40 a 45 años y ha nacido en otra provincia.

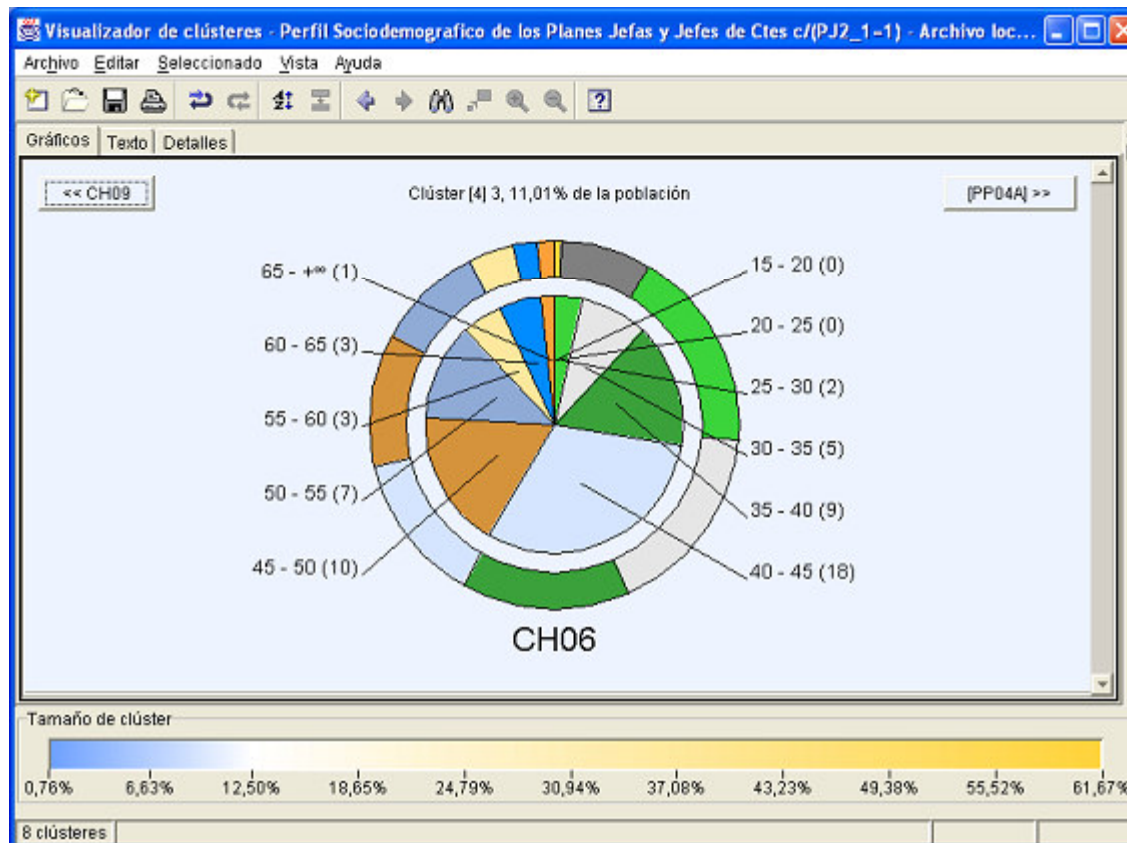


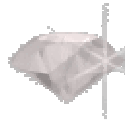


Resultado Obtenidos

Visualización del cada Clúster N°3 con 11,01 %:

En este diagrama circular se puede observar que el rango de edad con mayor frecuencia es el [40-45].

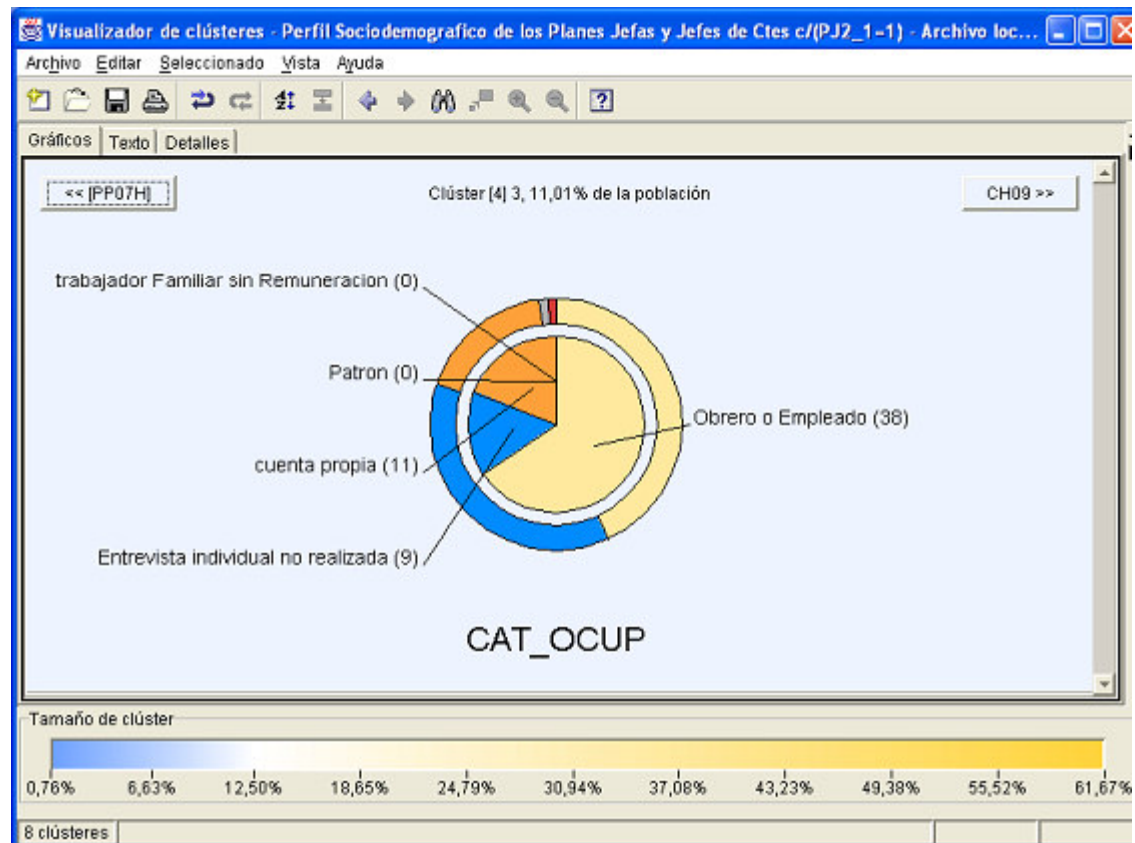


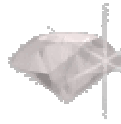


Resultado Obtenidos

Visualización del cada Clúster N°3 con 11,01 %:

*La categoría ocupacional que sobresale es la de “**obrero o empleado**” con un rubro de actividad económica como la “**servicios de hogares privados que contratan servicio domestico**”.*



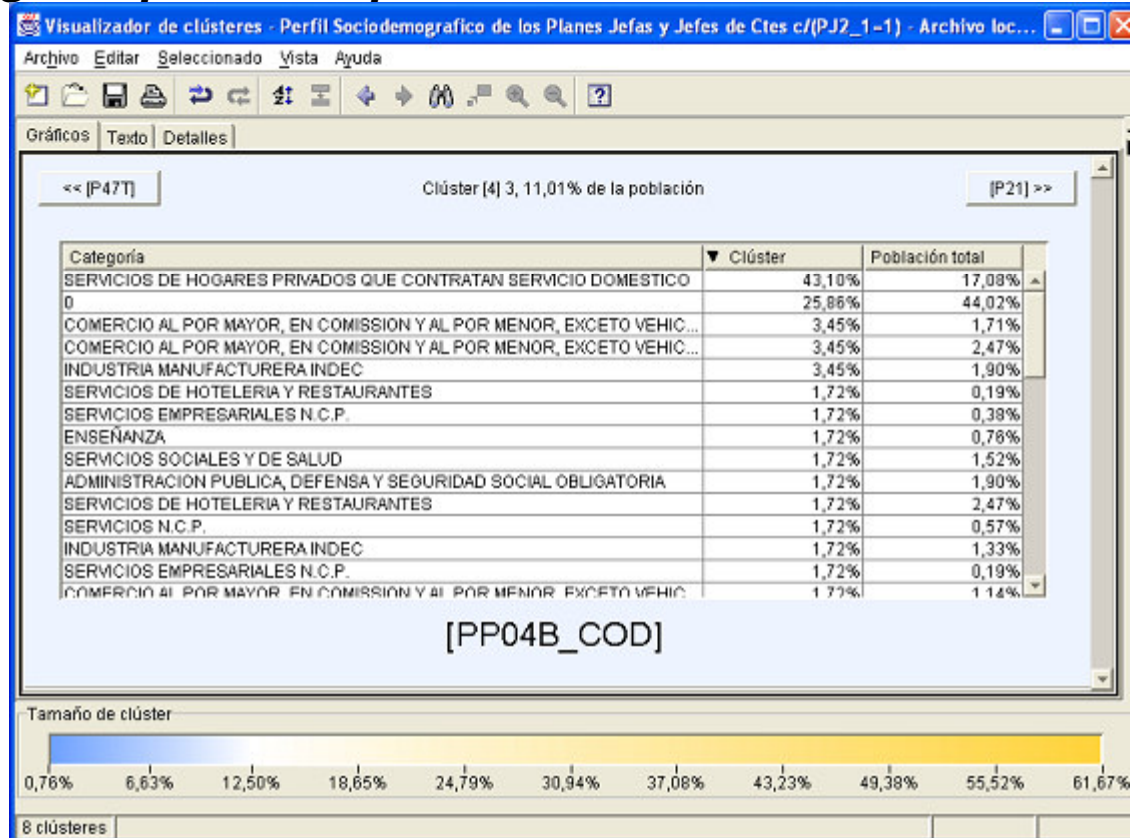


Resultado Obtenidos

Visualización del cada Clúster N°3 con 11,01 %:

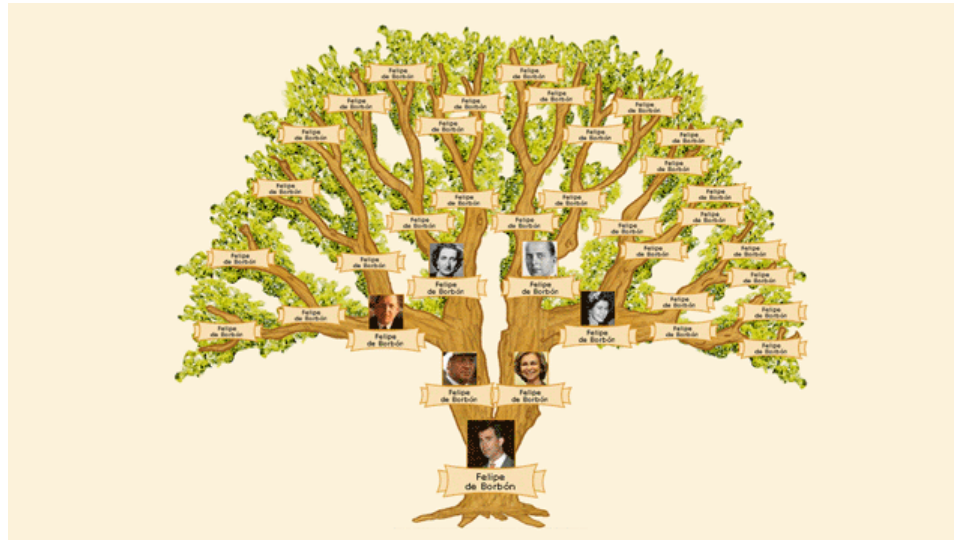
La categoría ocupacional que sobresale es la de “obrero o empleado” con un rubro de actividad económica como la

“servicios de hogares privados que contratan servicio domestico”.



Resultado Obtenidos

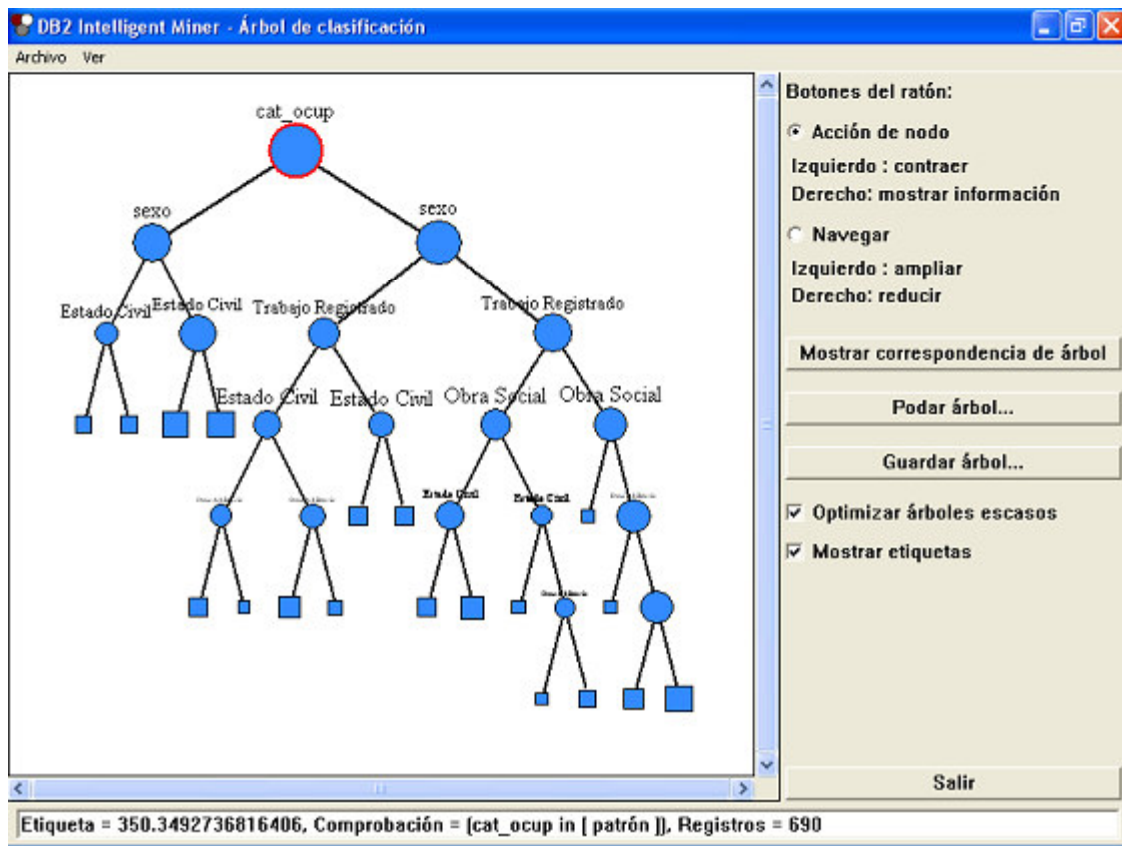
Técnica empleada: 'Árboles de Decisión y Clasificación'.

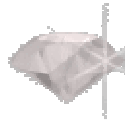


 *Clasificación del Ingreso de Cada Individuo, en Base a sus Principales Características Sociodemográficas.*

Resultado Obtenidos

Se identifican diecinueve reglas que explican el perfil de estos individuos, determinadas por los nodos de desarrollo del Árbol (mayor cantidad de individuos y mayor pureza).

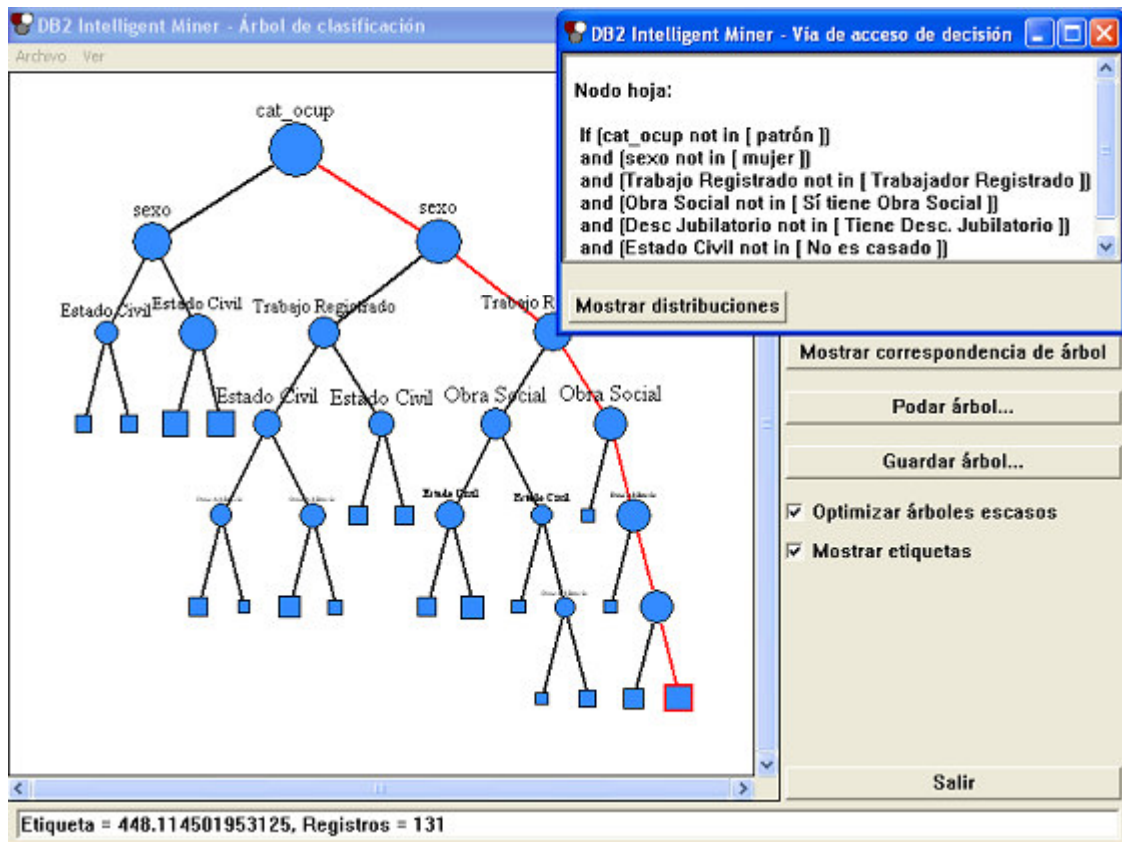


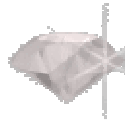


Resultado Obtenidos

Regla N° 1

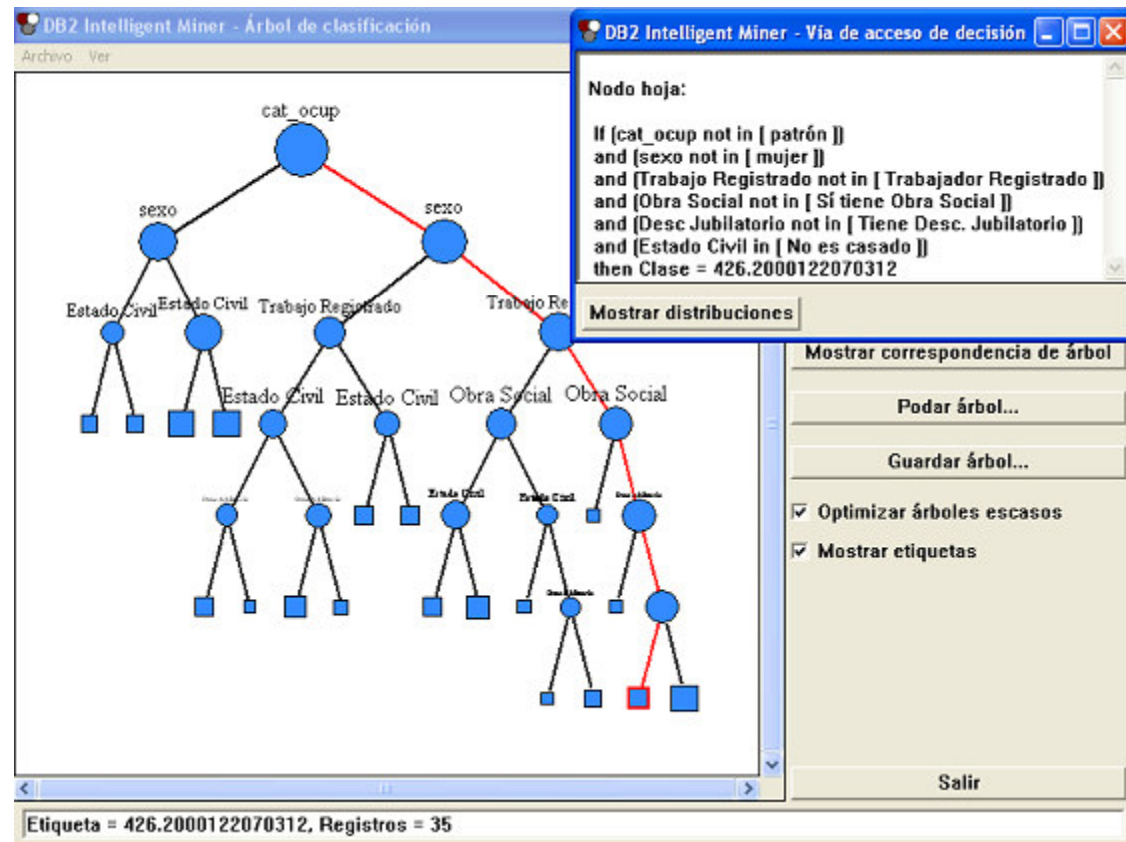
Si el individuo de estudio es de sexo femenino, no es patrón, no tiene trabajo registrado, ni obra social, ni descuento jubilatorio y su estado civil no es el casados entonces el ingreso total individual es de 448.11.

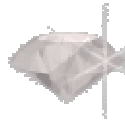




Resultado Obtenidos

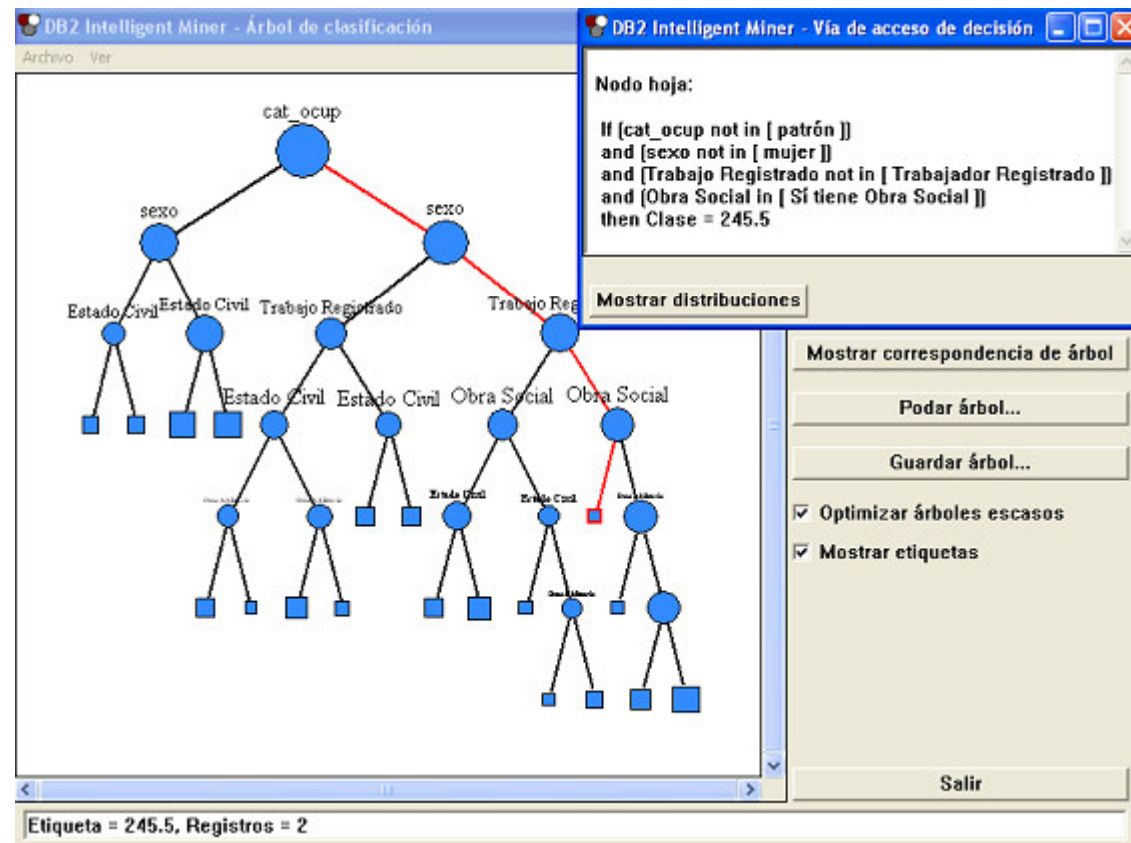
Regla N°2
Si el individuo de estudio es de sexo femenino, no es patrón, no tiene trabajo registrado, ni obra social, ni descuento jubilatorio y su estado civil es el casados entonces el ingreso total individual es de 426.20.

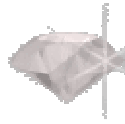




Resultado Obtenidos

Regla N° 4
Si el individuo de estudio es de sexo femenino, no es patrón, no tiene trabajo registrado, pero sí posee obra social, entonces el ingreso total individual es de 245.5.

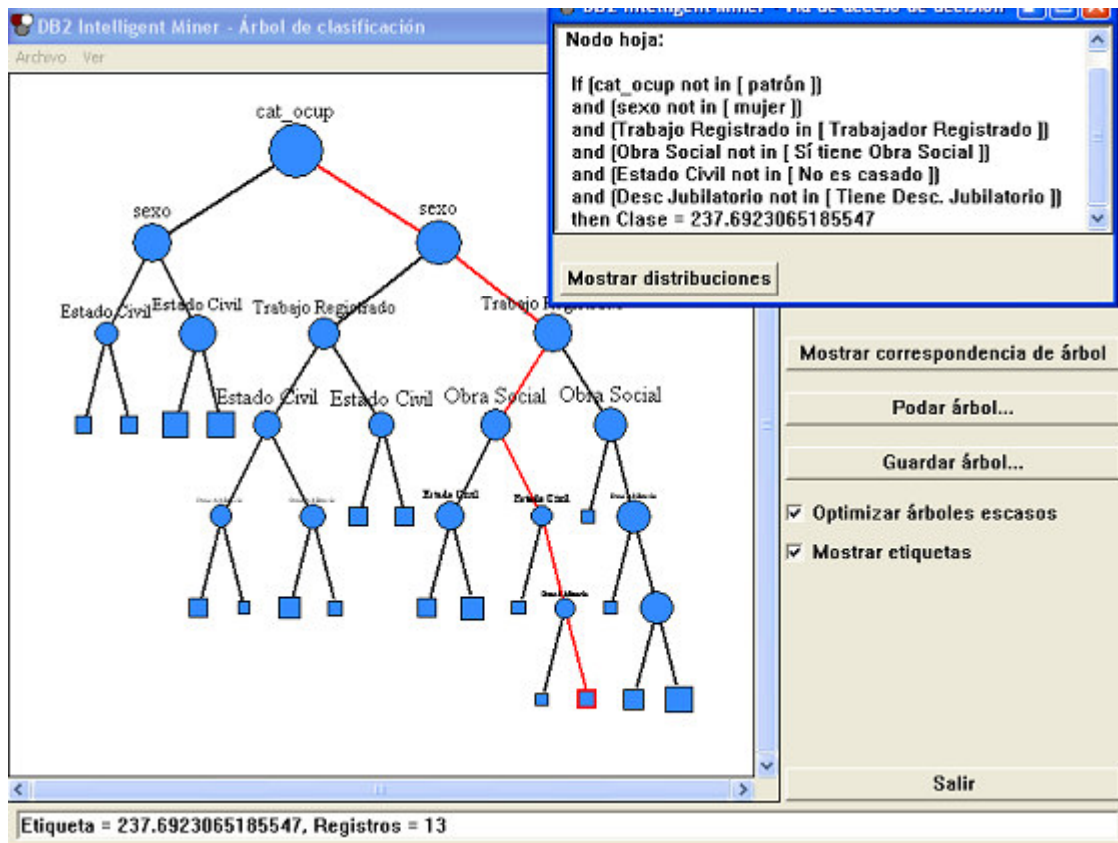




Resultado Obtenidos

Regla N°5

Si el individuo de estudio es de sexo femenino, goza de un trabajo registrado, no es patrón, no posee obra social, ni descuento jubilatorio y su estado civil no es casados entonces el ingreso total individual es de 237.69.





Conclusiones

◆ Conclusiones acerca de las Tecnologías y Software utilizados.

- Se ha podido comprobar las grandes ventajas de la utilización de tecnologías y software de última generación que soportan sistemas distribuidos multiplataforma.
-



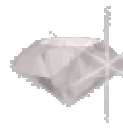
Conclusiones

- ❖ **Conclusiones acerca de los objetivos propuestos.**
 - ❑ **Se ha podido conocer a la población de la ciudad de Corrientes en un elevado nivel de detalle socio-demográfico y educacional.**
-



Conclusiones

- ◆ **Conclusiones respecto del proceso de extracción del conocimiento.**
 - **El desarrollo de un Almacén de Datos (Data Warehouse) permitió adquirir conocimientos adicionales sobre el diseño y desarrollo de esta tecnología.**
-



Conclusiones

- ❖ **Cabe destacar la eficiencia de los siguientes algoritmos aplicados:**
 - ❑ **‘Clustering’:** permitió obtener un modelo con los datos socio demográficos y de educación de los individuos de la población estudiada.
 - ❑ **‘Árboles de decisión y clasificación’:** permitió la generación de reglas que ilustran las relaciones existentes entre los ingresos y el nivel socio demográfico, como también entre los ingresos y la educación de cada individuo.



Conclusiones

Líneas futuras de acción.

- ❑ **Avanzar en la investigación mediante la aplicación de otras técnicas de minería de datos tales como Redes Neuronales, Redes Bayecianas, etc.**
 - ❑ **Investigar acerca de la aparición de nuevas herramientas de Inteligencia de Negocios (Business Intelligent), y aplicarlas con el fin de obtener nuevos resultados y poder realizar comparaciones.**
 - ❑ **Mejorar la aplicación generada agregando RIA (Rich Internet Applications).**
-

¿?

Espacio para preguntas

!?



Muchas Gracias



<http://biafull.blogspot.com/>



alfonsocutro@gmail.com
