INSTITUTO TECNOLÓGICO DE CULIACÁN

Unidad 3:

Minería de datos

MATERIA:

Temas Selectos de Base de Datos

ACTIVIDAD:

Practica 1 Comparativa de algoritmos de Mineria

CARRERA:

Ingeniería en Sistemas Computacionales

MAESTRO:

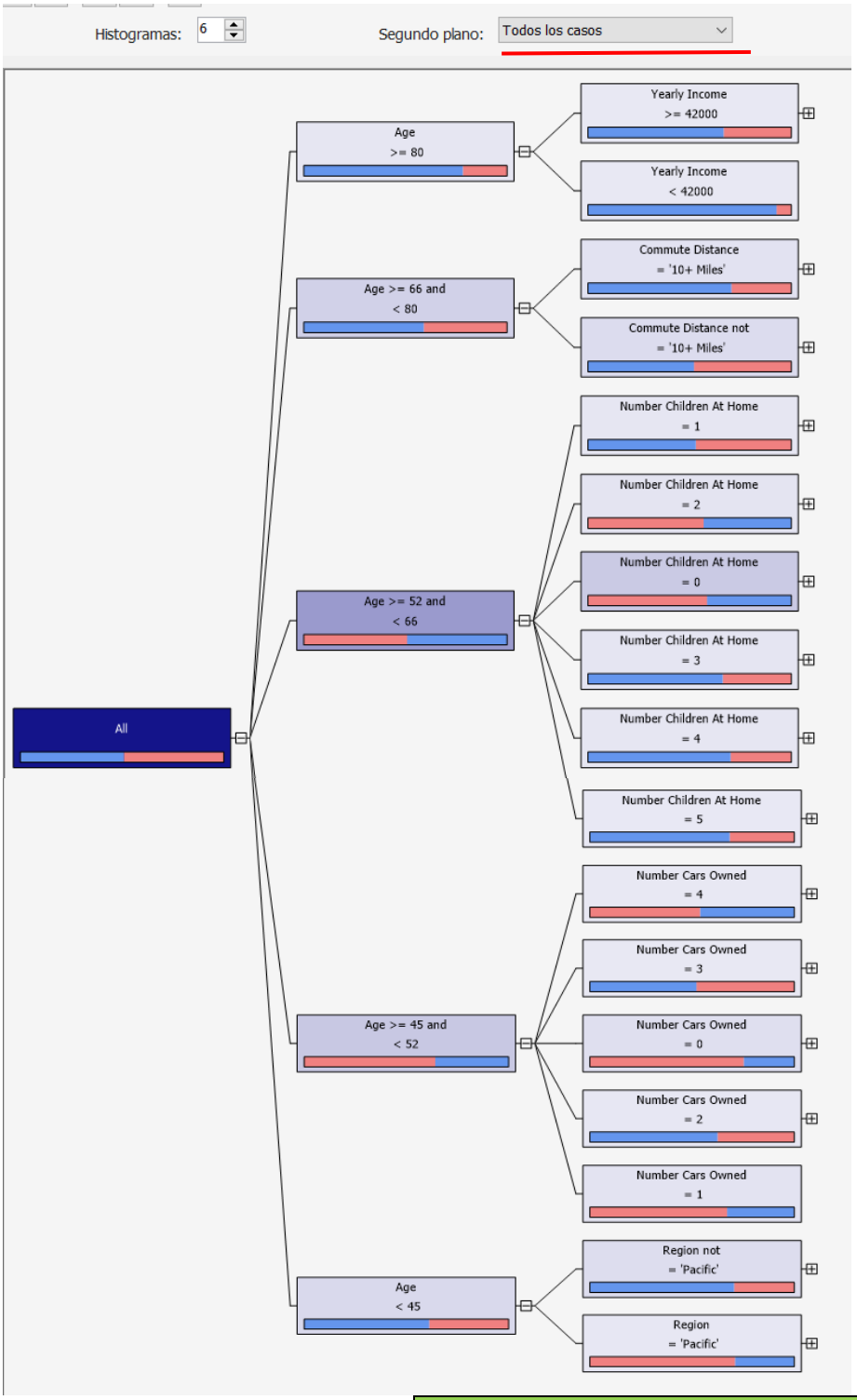
MC.Daniel Esparza Soto

INTEGRANTES EQUIPO:

Castro Cruz Luis Daniel

Felix Fierro Ryan Guadalupe

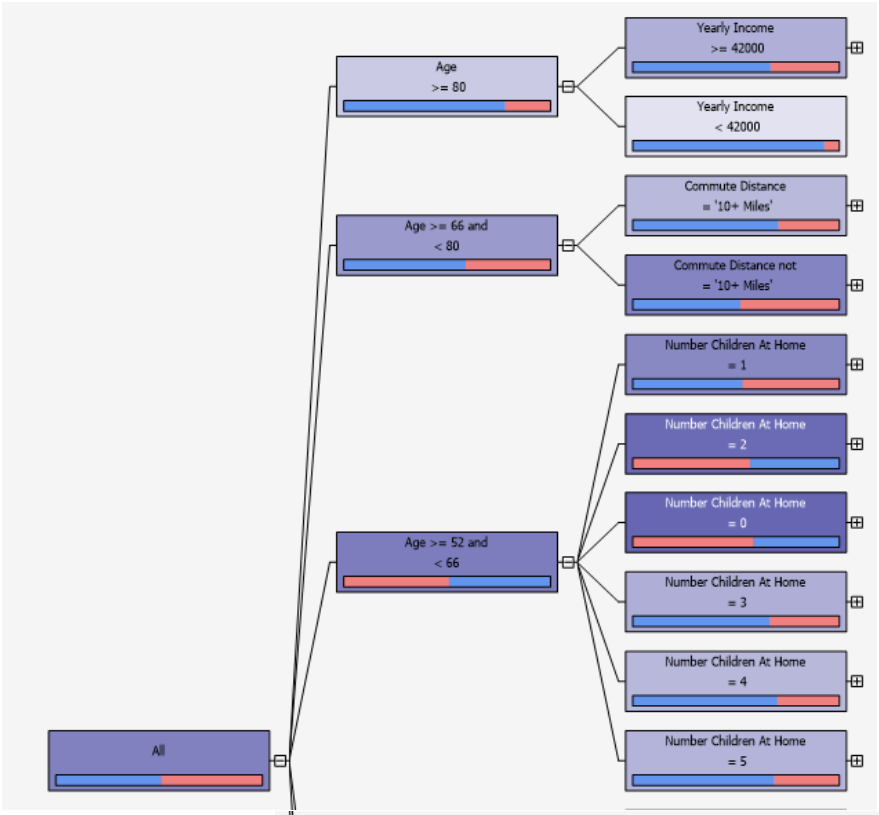
Flores Mascareño America Citlaly



ÁRBOL DE DECISIÓN

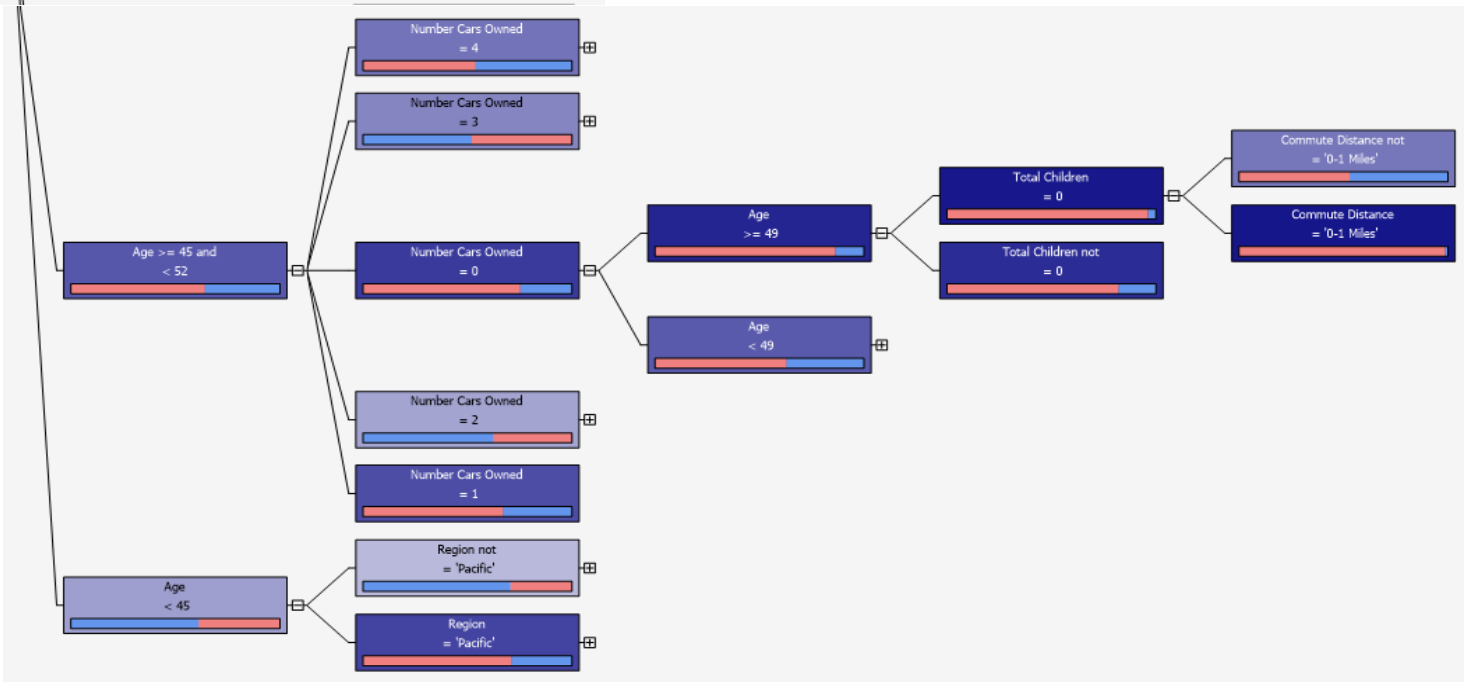
(Todos los casos)

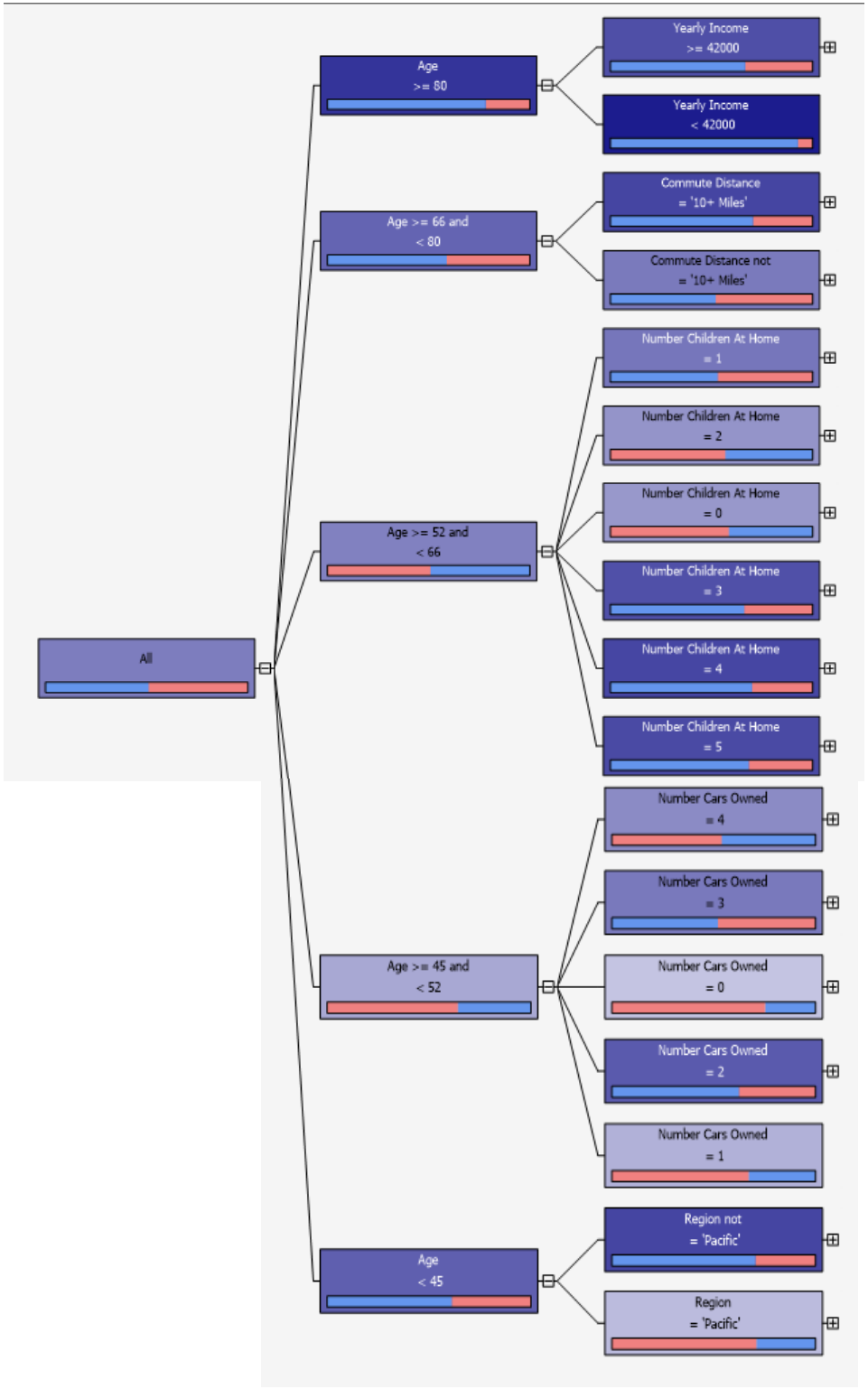
Árbol de decisión		
Arbol de decisión (1)	**Arbol de decisión (0)**	**Arbol de decisión/RedDependencias**
Age > = 45 and < 52	Age >= 80	Age
Number Cars Owner 0	Yearly income < 42000	Region
Age > = 49		Number Children At Home
Total children = 0		Number Cars Owned
Commute Distance = "0-1 miles"		Yearly income
		Marital Status
		Commute Distance
		English Ocupation
		English Education
		English Education
		House owner Flag



ÁRBOL DE DECISIÓN (1)

Árbol de decisión (1)
Age > = 45 and < 52
Number Cars Owner 0
Age > = 49
Total children = 0
Commute Distance = "0-1 miles"





ÁRBOL DE DECISIÓN (0)

Arbol de decisión (0)
Age >= 80
Yearly income < 42000

Arbol de decisión/RedDependencias
Age
Region
Number Children At Home
Number Cars Owned
Yearly income
Marital Status
Commute Distance
English Ocupation
English Education
English Education
House owner Flag

ÁRBOL DE DECISIÓN / RED DEPENDENCIAS

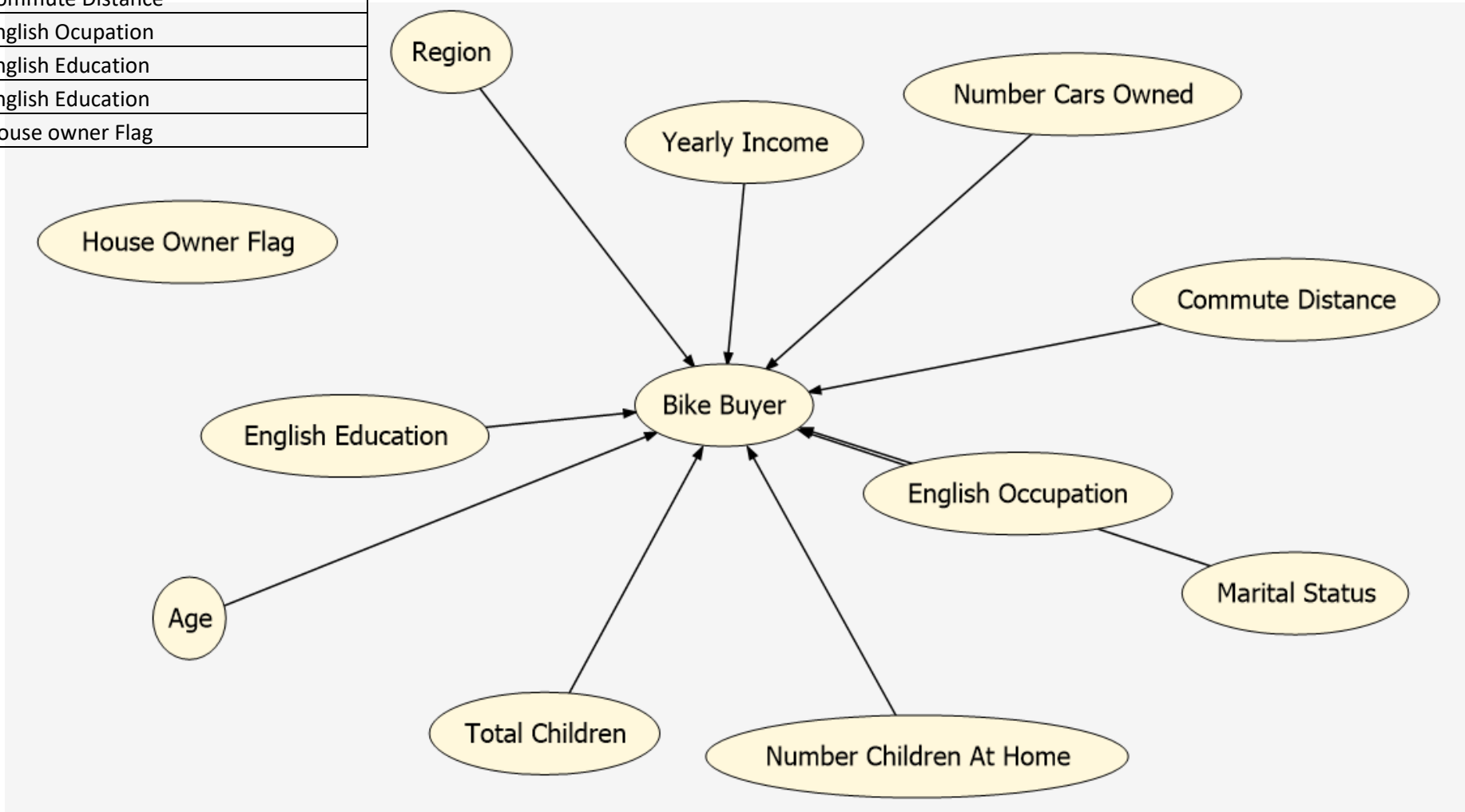
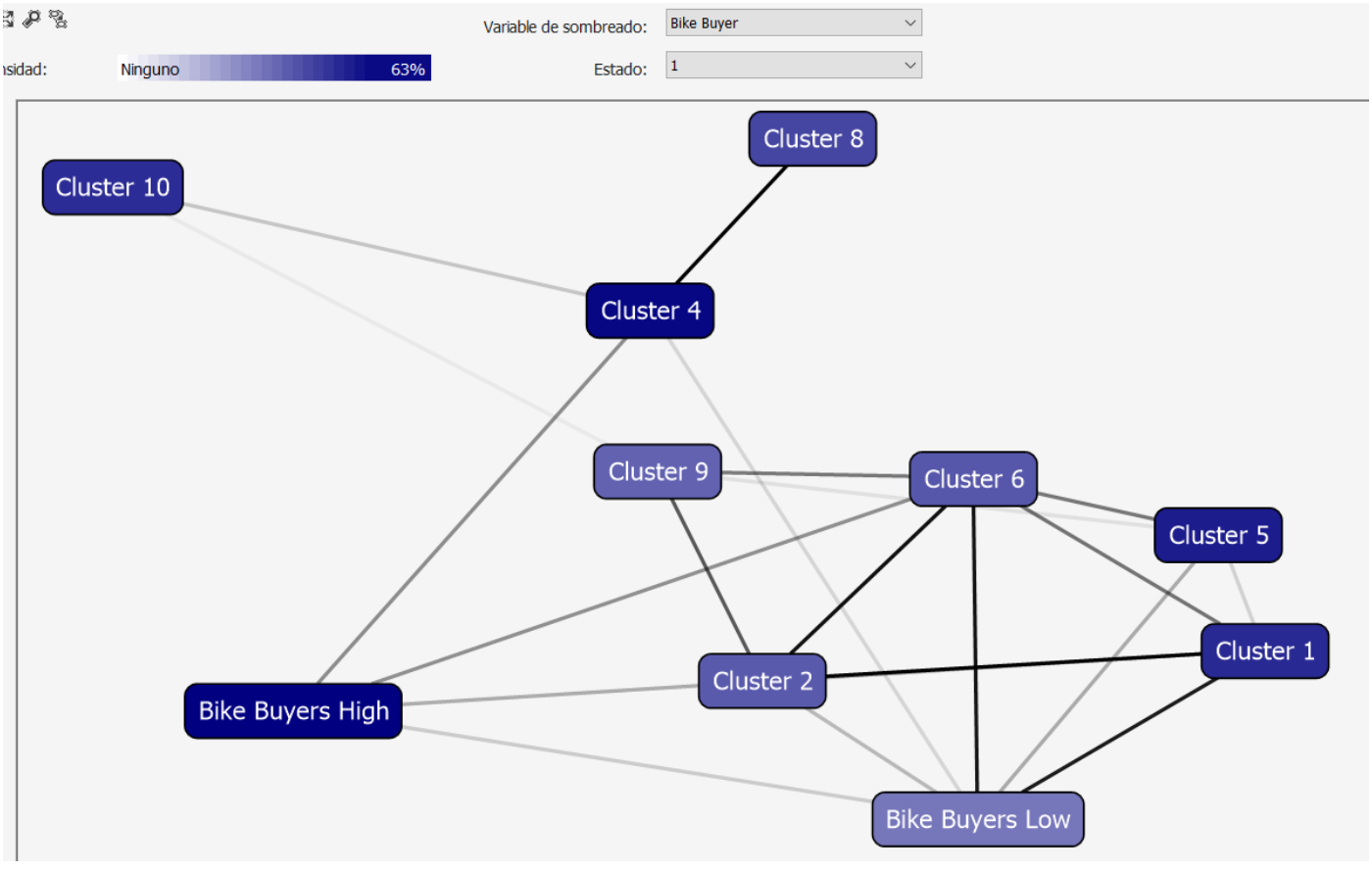


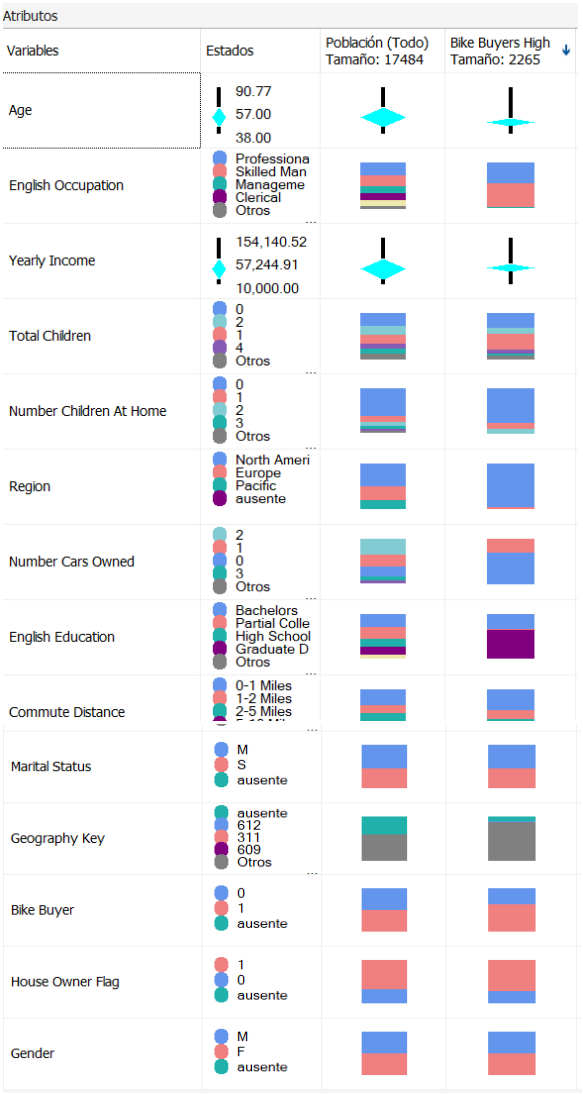
DIAGRAMA DE CLÚSTER

Cluster			
Perfil de cluster: Bike Buyer High	**Perfil de cluster: Bike Buyer Low**	**Caracteristicas de cluster: Bike Buyer High**	**Caracteristicas de cluster: Bike Buyer Low**
Age	Region	Region North America	Region: North America
English Ocupation	Age	Number Children At Home: 0	Age: 38 - 49
Yearly Income	English Education	House owner Flag: 1	Number Children At Home: 0
Total children	Number Cars Owner	Number Cars Owner: 0	Total children: 0
Number Children At Home	Yearly Income	English Education: Graduate Degree	Number Cars Owner: 2
Region	English Ocupation	Bike Buyer: 1	English Ocupation: Skilled Manual
Number Cars Owned	Commute Distance	Age: 50 -56	Bike Buyer: 0
English education	Number Children At Home	Marital Status: M	Marital Status: S
Commute Distance	Bike Buyer	English Ocupation: Skilled Manual	Yearly income: 35,459.9 - 57,244.9
Marital Status	Total Children	Yearly income: 57,244.9 - 79,029.9	Commute Distance: 5-10 miles
Geography key	Geography Key	Gender: F	House owner Flag: 1
Bike Buyer	House owner Flag	Gender: M	English Education: Partial College
House Owner Flag	Marital Status	Marital Status: S	Gender: M
Gender	Gender		Gender: F



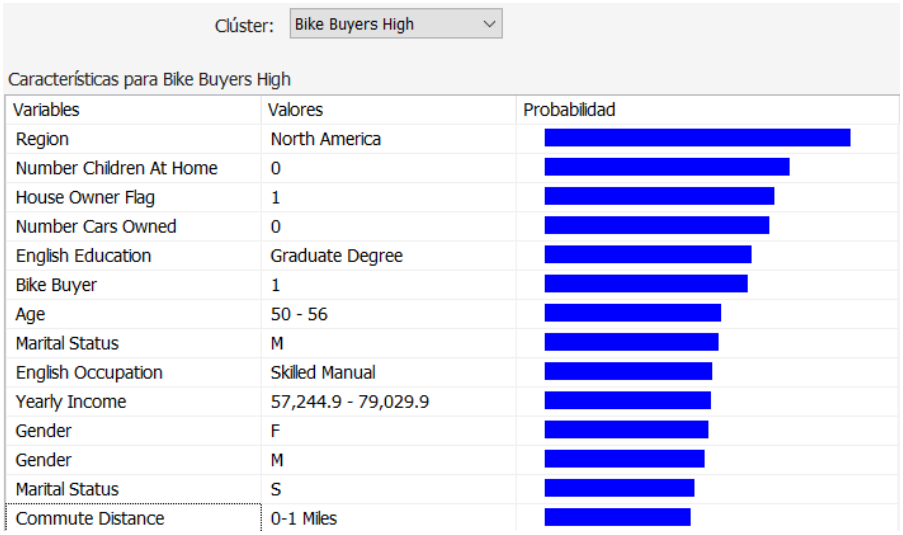
Perfiles de Cluster

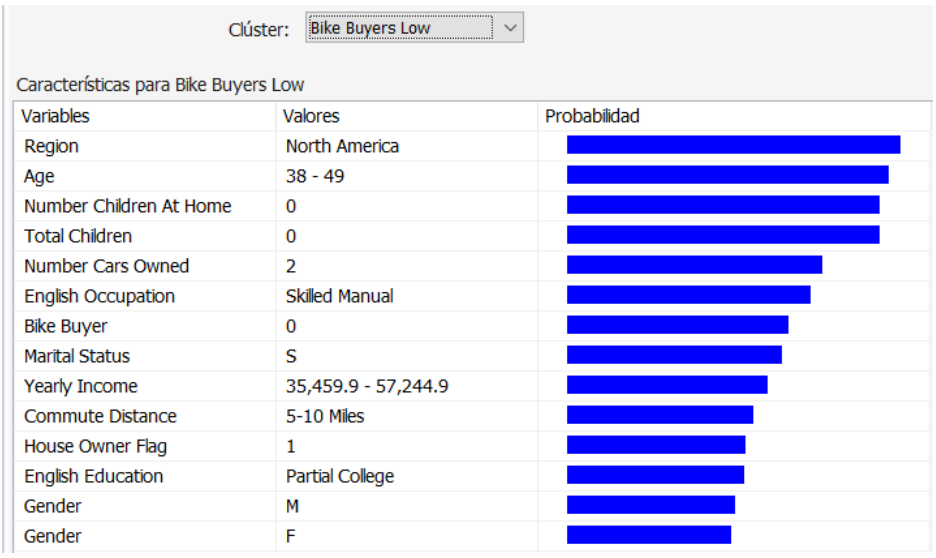
Perfil de cluster: Bike Buyer High	Perfil de cluster: Bike Buyer Low
Age	Region
English Ocupation	Age
Yearly Income	English Education
Total children	Number Cars Owner
Number Children At Home	Yearly Income
Region	English Ocupation
Number Cars Owned	Commute Distance
English education	Number Children At Home
Commute Distance	Bike Buyer
Marital Status	Total Children
Geography key	Geography Key
Bike Buyer	House owner Flag
House Owner Flag	Marital Status
Gender	Gender



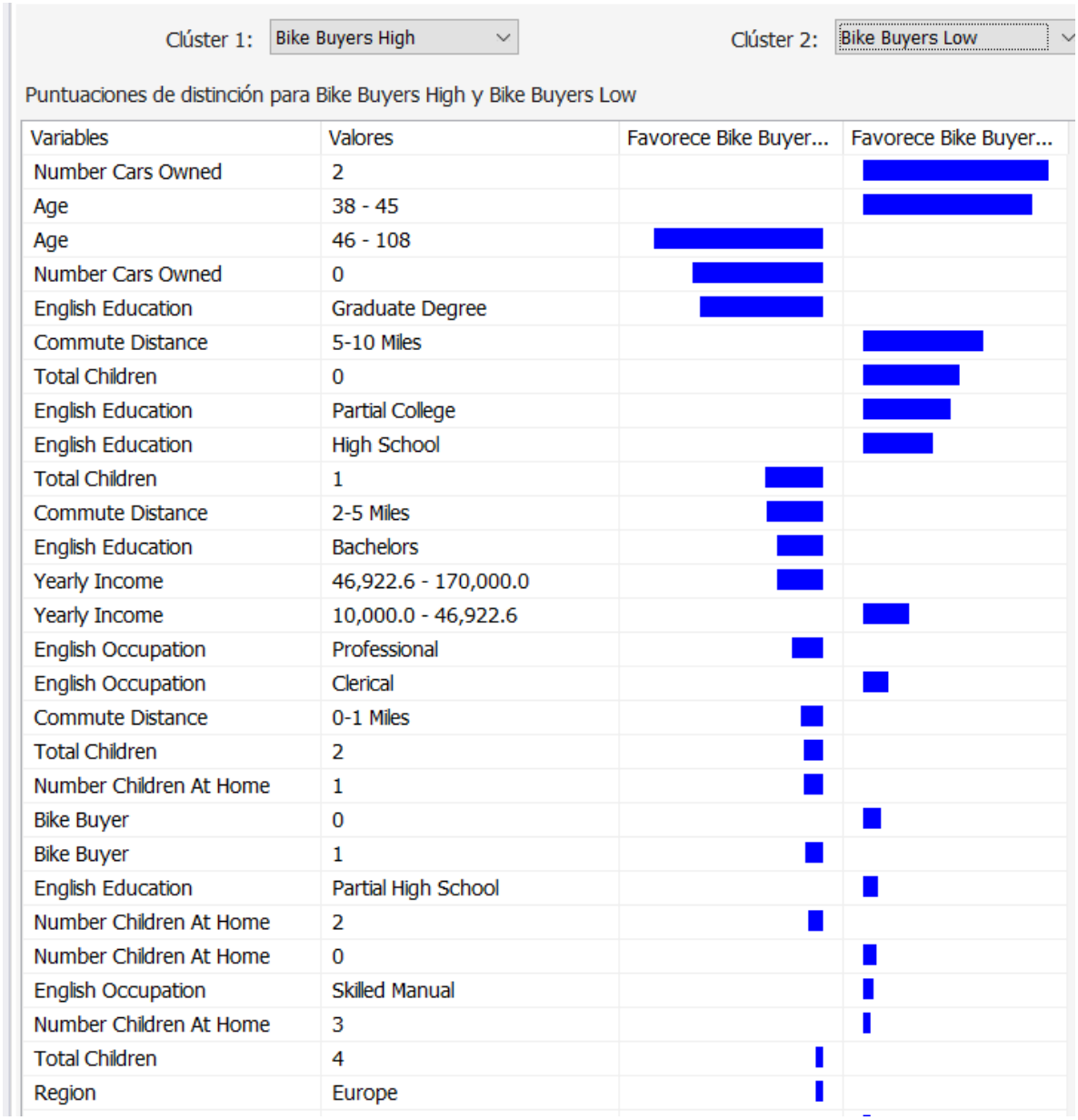
Características de Cluster

Caracteristicas de cluster: Bike Buyer High	Caracteristicas de cluster: Bike Buyer Low
Region North America	Region: North America
Number Children At Home: 0	Age: 38 - 49
House owner Flag: 1	Number Children At Home: 0
Number Cars Owner: 0	Total children: 0
English Education: Graduate Degree	Number Cars Owner: 2
Bike Buyer: 1	English Ocupation: Skilled Manual
Age: 50 -56	Bike Buyer: 0
Marital Status: M	Marital Status: S
English Ocupation: Skilled Manual	Yearly income: 35,459.9 - 57,244.9
Yearly income: 57,244.9 - 79,029.9	Commute Distance: 5-10 miles
Gender: F	House owner Flag: 1
Gender: M	English Education: Partial College
Marital Status: S	Gender: M
	Gender: F





Ficha distinción del clúster



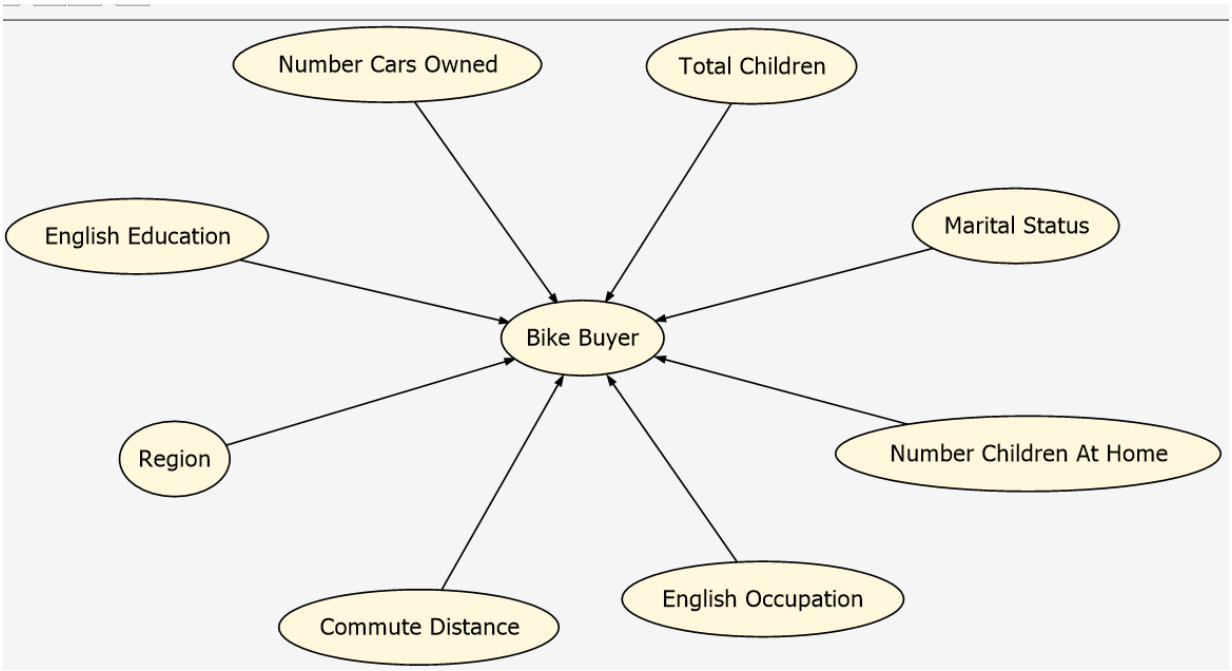
Puntuaciones de distinción para Bike Buyers High y Bike Buyers Low

Variables	Valores	Favorece Bike Buyer...	Favorece Bike Buyer...
Number Cars Owned	2		
Age	38 - 45		
Age	46 - 108		
Number Cars Owned	0		
English Education	Graduate Degree		
Commute Distance	5-10 Miles		
Total Children	0		
English Education	Partial College		
English Education	High School		
Total Children	1		
Commute Distance	2-5 Miles		
English Education	Bachelors		
Yearly Income	46,922.6 - 170,000.0		
Yearly Income	10,000.0 - 46,922.6		
English Occupation	Professional		
English Occupation	Clerical		
Commute Distance	0-1 Miles		
Total Children	2		
Number Children At Home	1		
Bike Buyer	0		
Bike Buyer	1		
English Education	Partial High School		
Number Children At Home	2		
Number Children At Home	0		
English Occupation	Skilled Manual		
Number Children At Home	3		
Total Children	4		
Region	Europe		

Modelo Bayes Naive

Bayes Naive		
Red de dependencias	**Perfil de atributo**	**Caracteristica del atributo: si compraron**
Number Cars Owner	Number Cars Owner	Number Children At Home 0
Commute Distance	Commute Distance	Marital Status M
Total children	Region	Marital Status S
English Education	English Education	Region NA
Number Children At Home	Total children	Commute Distance 0-1 miles
Region	Marital Status	English Ocupation Bacherol
Marital Status	Number Children At Home	English Ocupation Professional
English Ocupation	English Ocupation	Region Europe
		Number Cars Owner 0
		Number Cars Owner 1
		Total children 0
		Number Cars Owner 2
		English Education Partial College

Red de Dependencias



Perfil de atributo

Perfil de atributo
Number Cars Owner
Commute Distance
Region
English Education
Total children
Marital Status
Number Children At Home
English Ocupation

Características del atributo: Sí compraron

Caracteristica del atributo: si compraron
Number Children At Home: 0
Marital Status: M
Marital Status: S
Region: North America
Commute Distance: 0-1 miles
English Ocupation: Bachelors
English Ocupation: Professional
Region: Europe
Number Cars Owner: 0
Number Cars Owner: 1
Total children: 0
Number Cars Owner: 2
English Education: Partial College

Distinción del atributo

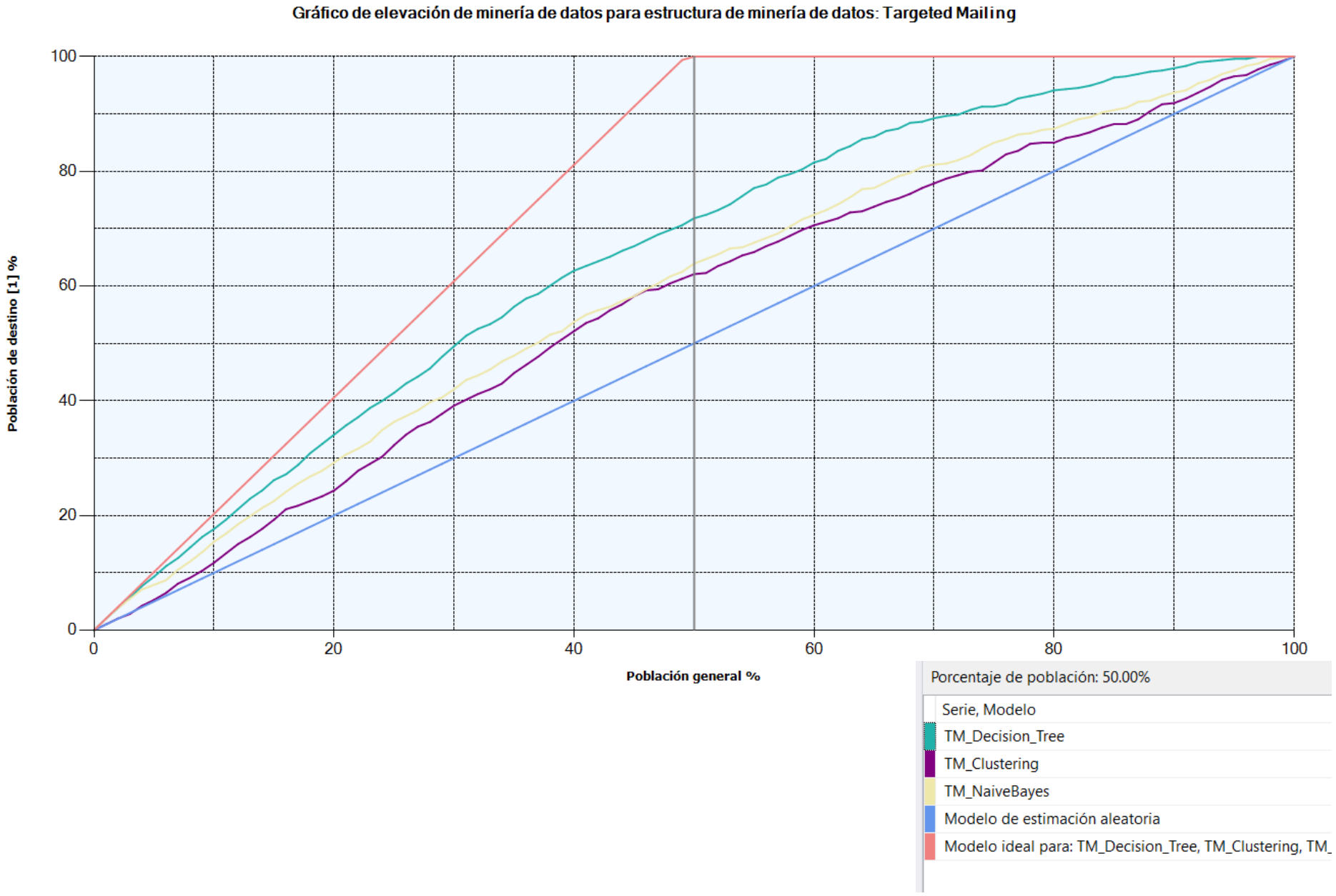
Red de dependencias | Perfiles del atributo | Características del atributo | Distinción del atributo

Atributo: Bike Buyer Valor 1: 0 Valor 2: Todos los otros estados

Puntuaciones de distinción para 0 y Todos los otros estados

Atributos	Valores	Favorece 0	Favorece Todos los ...
Number Cars Owned	0		
Number Cars Owned	2		
English Education	Partial High School		
Total Children	1		
Commute Distance	10+ Miles		
Total Children	5		
Region	Pacific		
Commute Distance	0-1 Miles		
Number Children At Home	4		
Region	North America		
English Education	Bachelors		
Total Children	4		
Commute Distance	5-10 Miles		
Number Children At Home	3		
Commute Distance	2-5 Miles		
Number Children At Home	2		
English Education	High School		
Number Cars Owned	4		
Number Cars Owned	1		
Marital Status	M		
Marital Status	S		
English Occupation	Clerical		
Number Children At Home	0		
English Occupation	Manual		
Number Cars Owned	3		
English Education	Graduate Degree		
Number Children At Home	5		

GRAFICA DE ELEVACIÓN

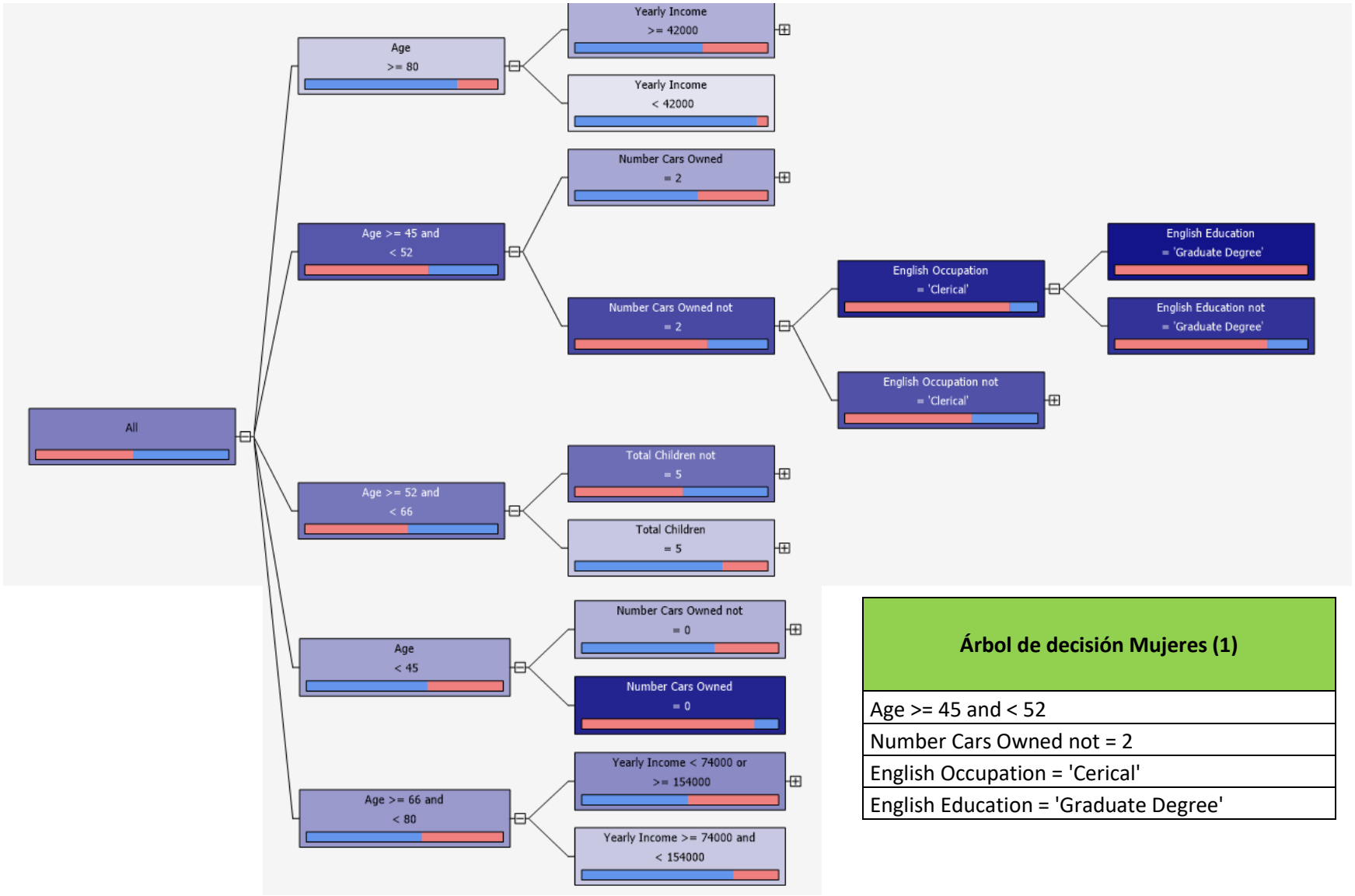


PROBANDO UN MODELO DE FILTRADO

Árbol de decisión Male/Female

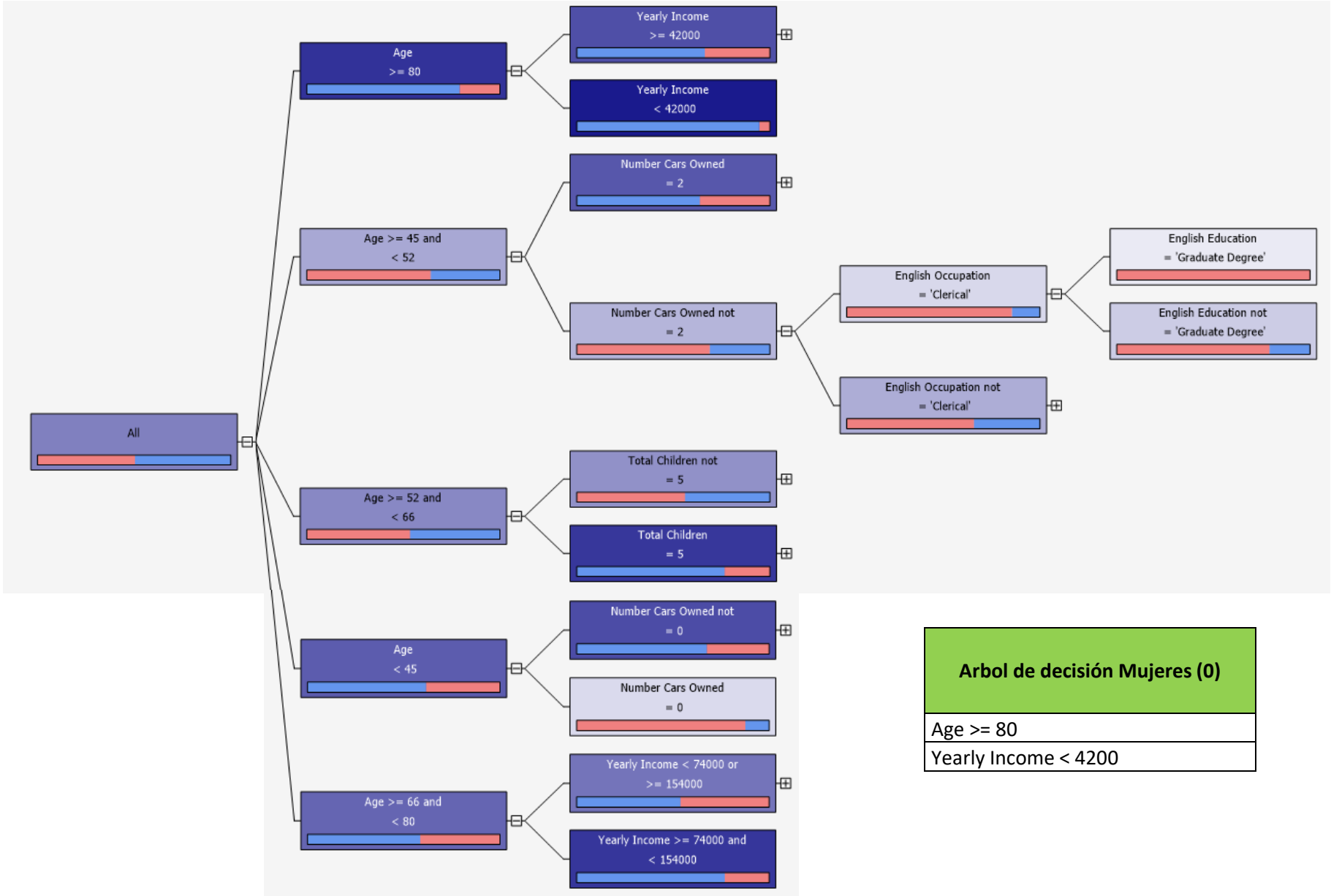
Arbol de decisión					
Arbol de decisión Hombres (0)	**Arbol de decisión Hombres (1)**	**Arbol de decisión Mujeres (0)**	**Arbol de decisión Mujeres (1)**	**Arbol de decisión/RedDependencias Hombres**	**Arbol de decisión/RedDependencias Mujeres**
Age >=80	Age >= 45 and < 52	Age >= 80	Age >= 45 and < 52	Age	Age
Yearly Income < 5800	Number Cars Owned not = 2	Yearly Income < 4200	Number Cars Owned not = 2	Region	Numbers Cars Owned
	English Occupation = 'Clerical'		English Occupation = 'Cerical'	Number Children At Home	Total Children
	Number Cars Owned = 0		English Education = 'Graduate Degree'	Yearly Income	Yearly Income
	Total Children = 0			Number Cars Owned	Commute Distance
				English Occupation	Number Children at Home
				Commute Distance	English Occupation
				Marital Status	English Education
				Total Children	Region
				English Education	Marital Status

ÁRBOL DE DECISIÓN FEMALE (1)



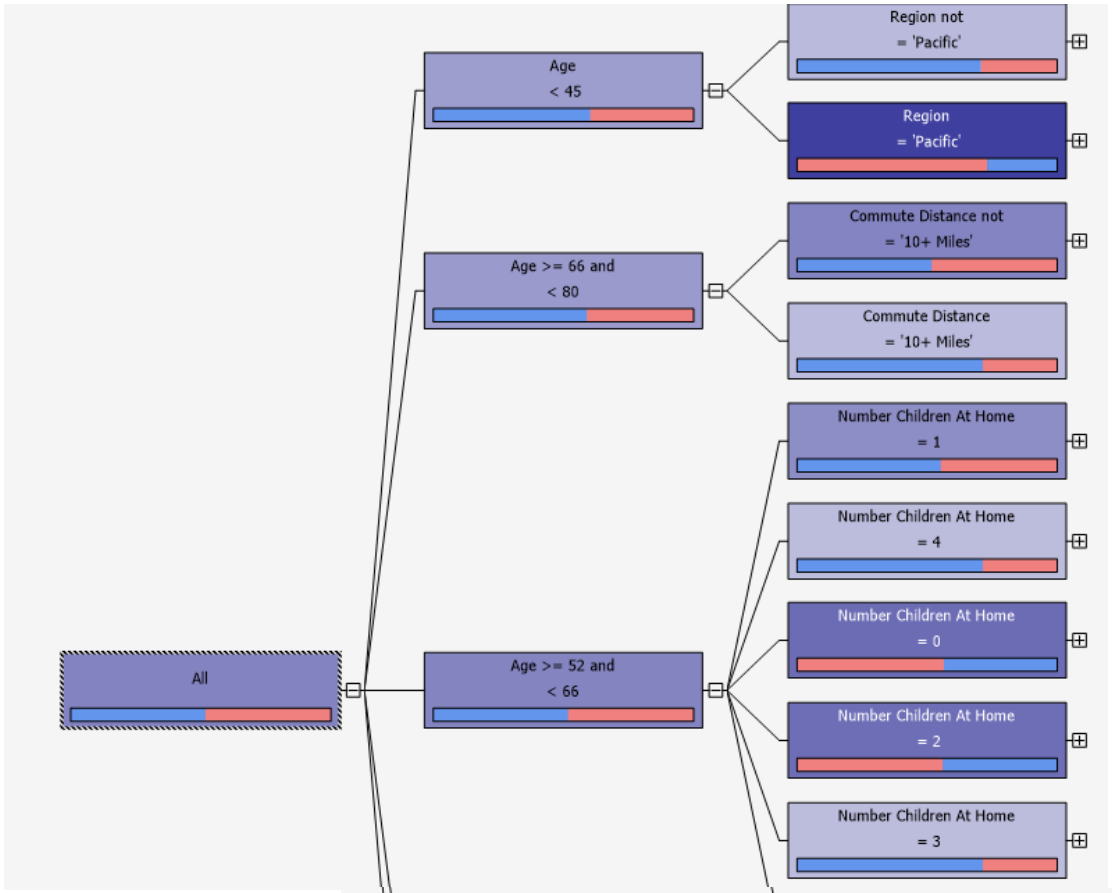
Árbol de decisión Mujeres (1)
Age >= 45 and < 52
Number Cars Owned not = 2
English Occupation = 'Cerical'
English Education = 'Graduate Degree'

ÁRBOL DE DECISIÓN FEMALE (0)



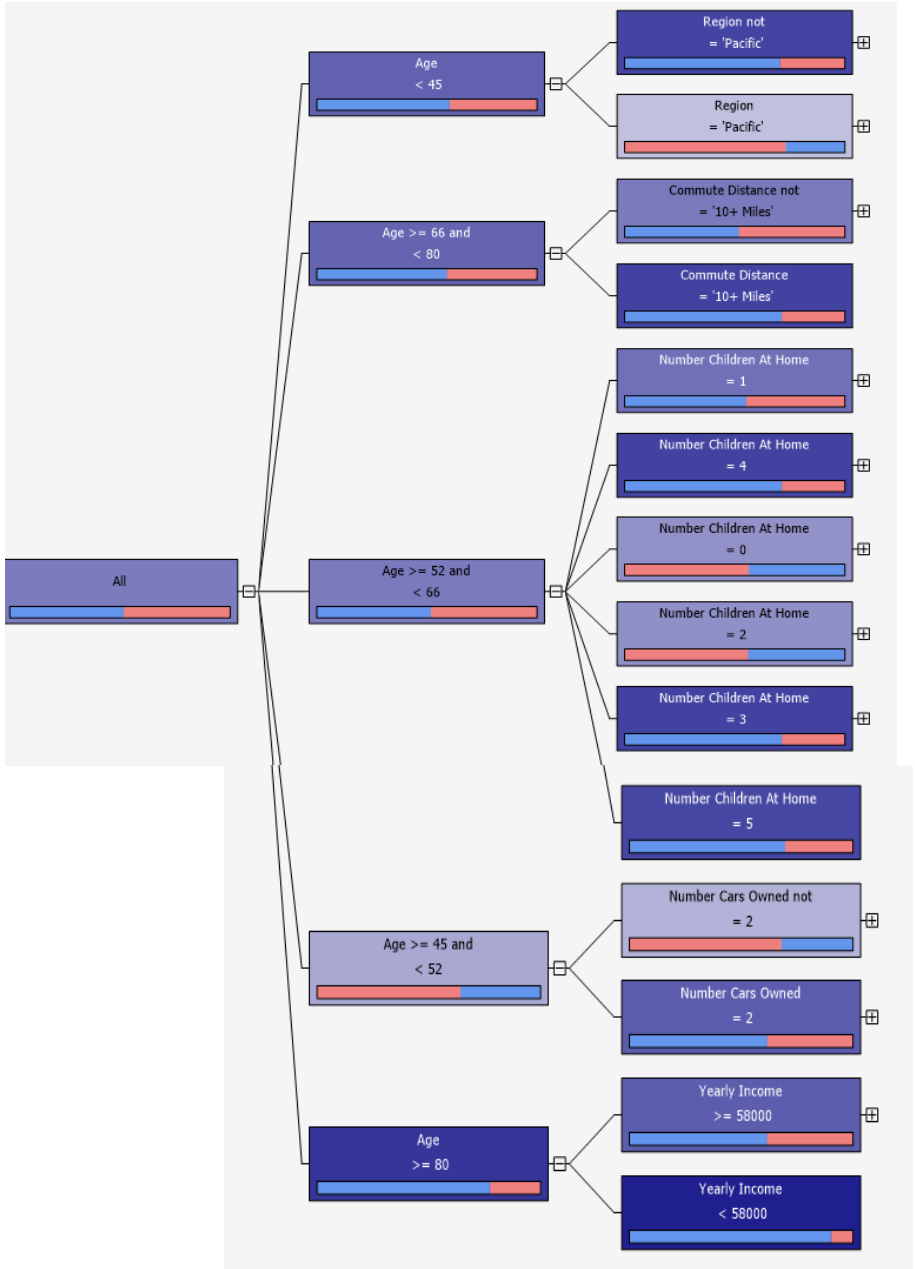
Arbol de decisión Mujeres (0)
Age >= 80
Yearly Income < 4200

ÁRBOL DE DECISIÓN MALE (1)



Arbol de decisión Hombres (1)
Age >= 45 and < 52
Number Cars Owned not = 2
English Occupation = 'Clerical'
Number Cars Owned = 0
Total Children = 0

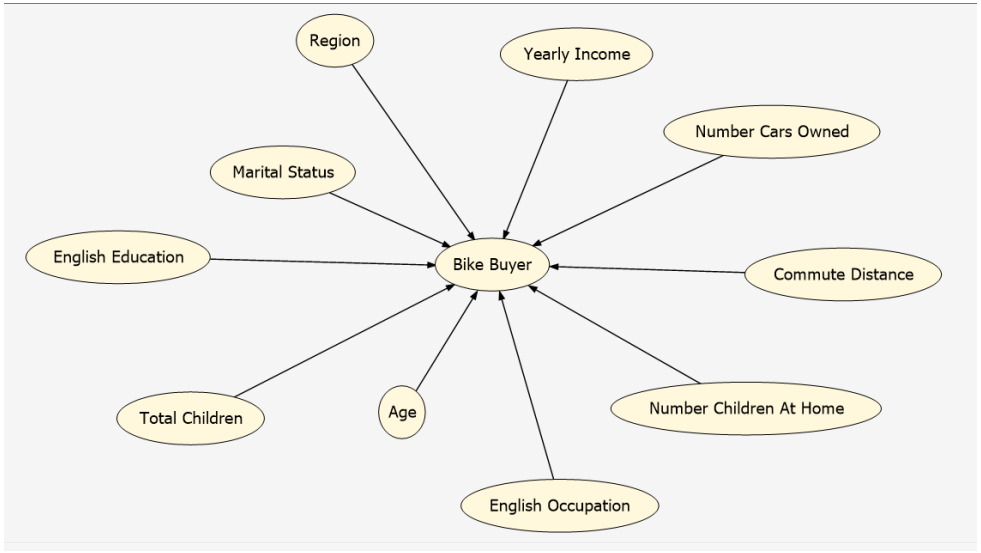
ÁRBOL DE DECISIÓN MALE (0)

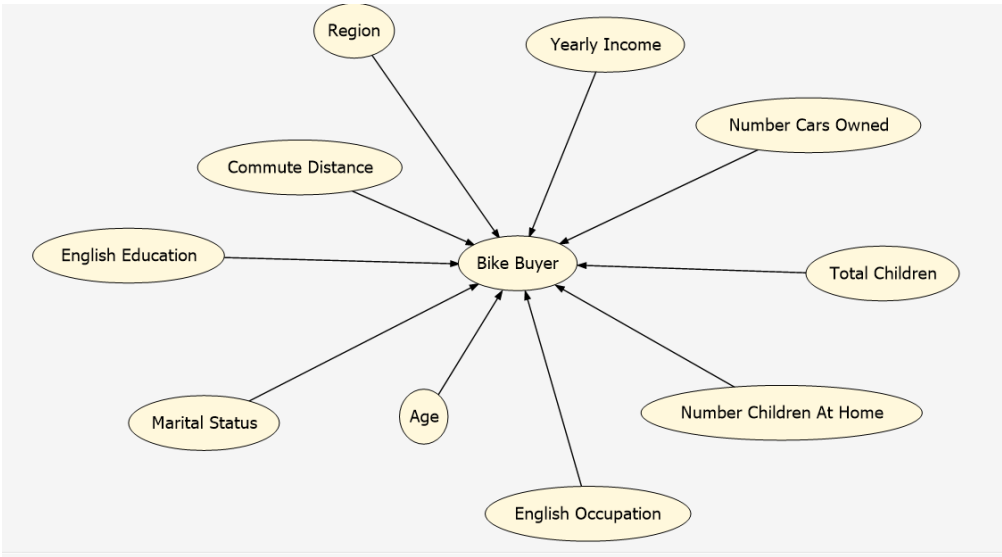


Arbol de decisión Hombres (0)
Age >=80
Yearly Income < 5800

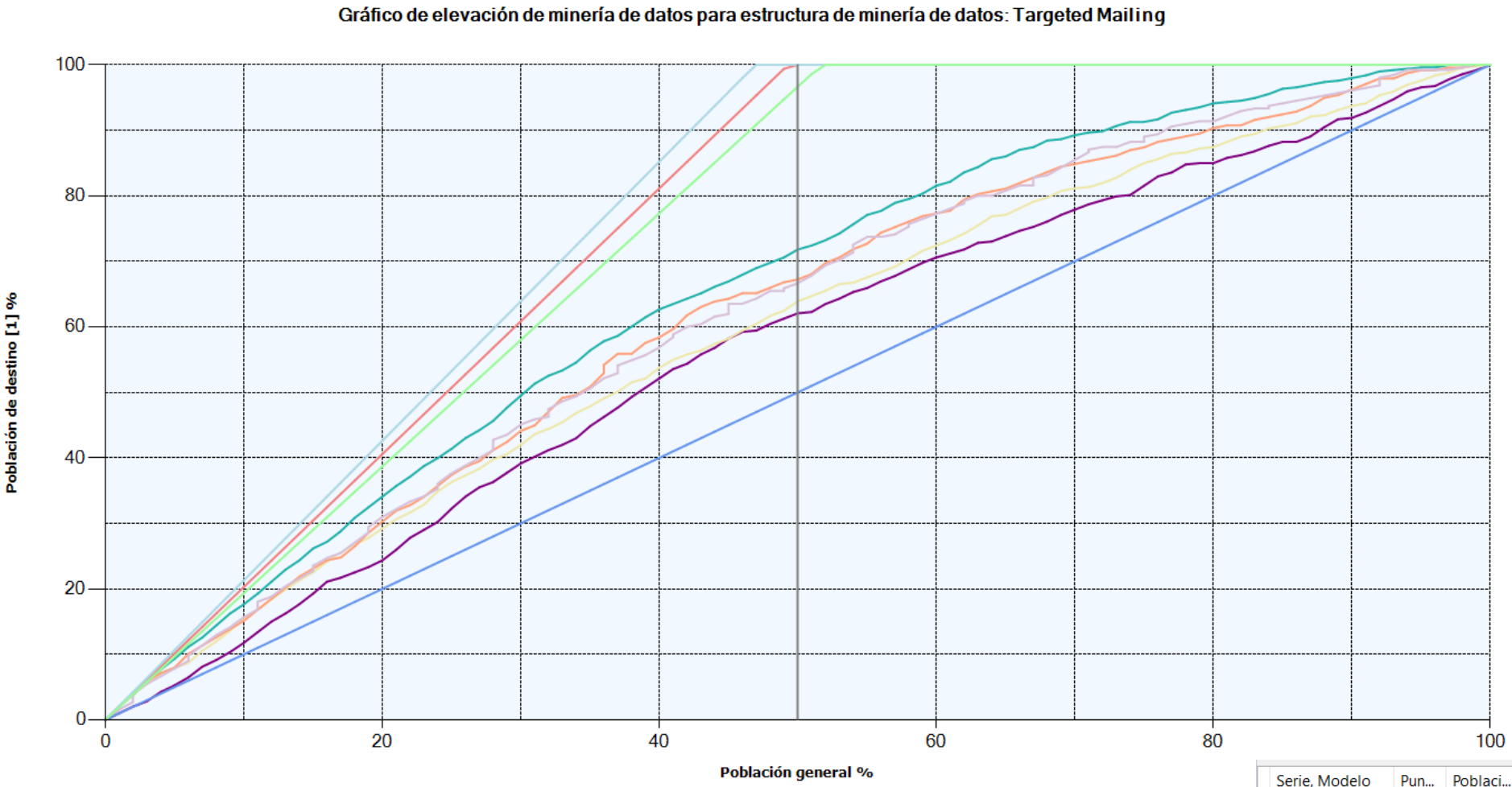
DEPENDENCIAS MALE/FEMALE

Arbol de decisión/RedDependencias Hombres	Arbol de decisión/RedDependencias Mujeres
Age	Age
Region	Numbers Cars Owned
Number Children At Home	Total Children
Yearly Income	Yearly Income
Number Cars Owned	Commute Distance
English Occupation	Number Children at Home
Commute Distance	English Occupation
Marital Status	English Education
Total Children	Region
English Education	Marital Status



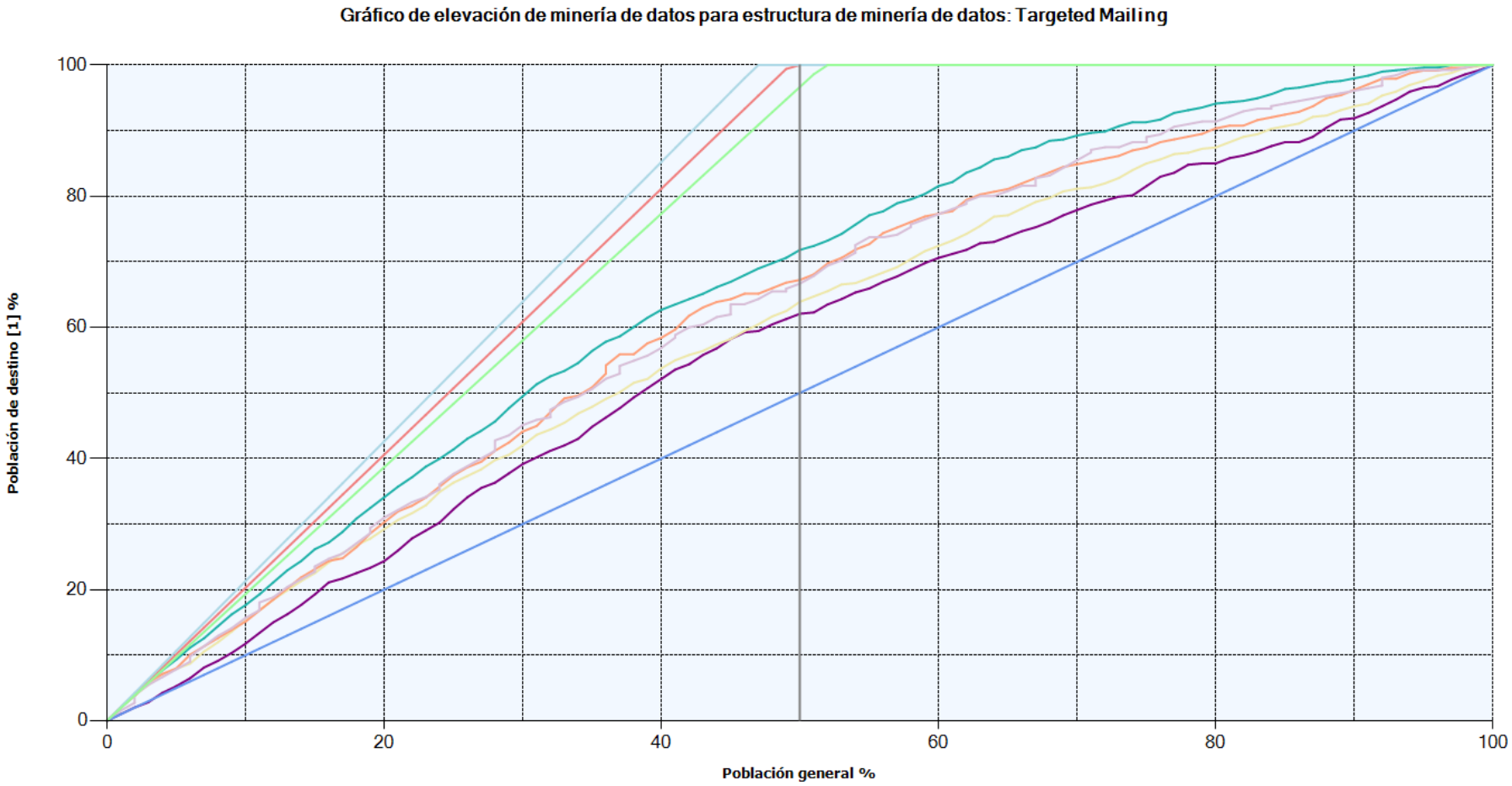


GRAFICA DE ELEVACIÓN FEMALE



Serie, Modelo	Pun...	Poblaci...	Probabil...
TM_Decision_...	0.70	71.81%	56.16%
TM_Clustering	0.61	62.07%	51.68%
TM_NaiveBayes	0.64	63.89%	47.78%
TM_DecisionT...	0.69	67.23%	42.35%
TM_Descision...	0.83	66.67%	55.78%
Modelo de es...		50.00%	
Modelo ideal ...		100.00%	
Modelo ideal ...		100.00%	
Modelo ideal ...		96.67%	

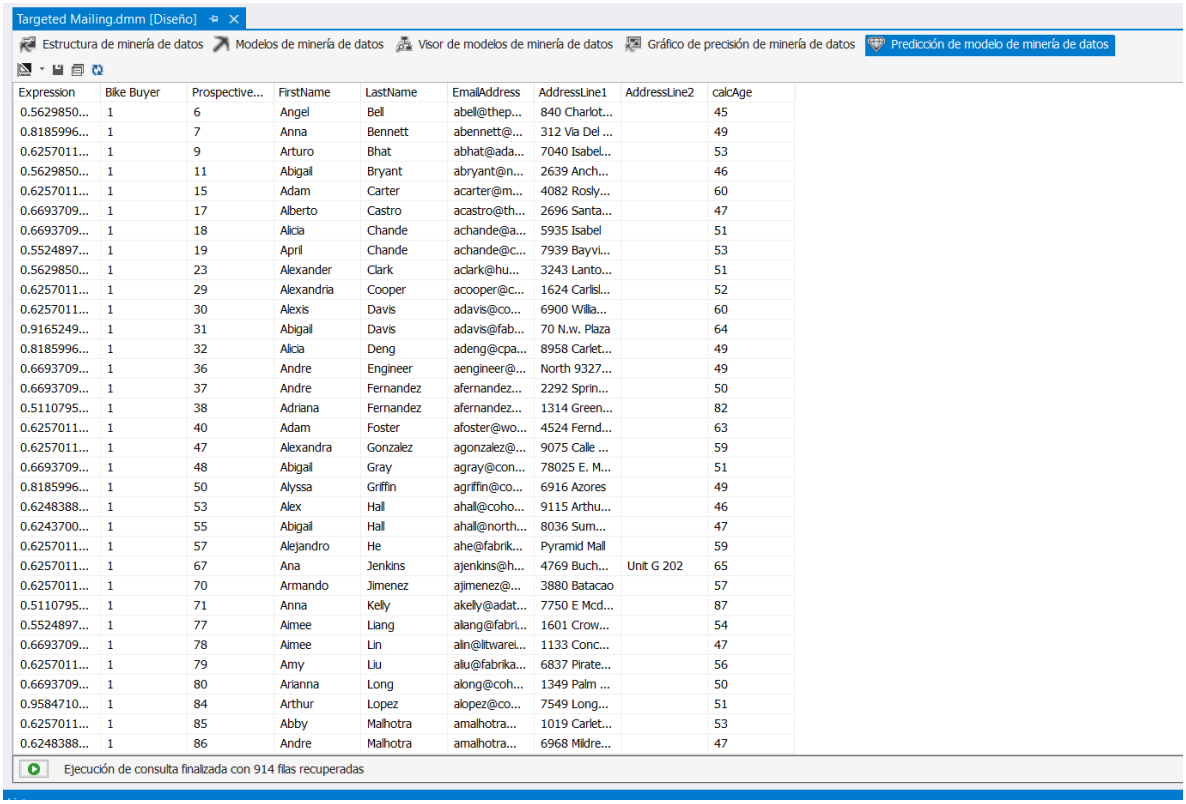
GRAFICA DE ELEVACIÓN MALE



Serie, Modelo	Pun...	Població...	Probabili...
TM_Decision_...	0.70	71.81%	56.16%
TM_Clustering	0.61	62.07%	51.68%
TM_NaiveBayes	0.64	63.89%	47.78%
TM_DecisionT...	0.69	67.23%	42.35%
TM_Descision...	0.83	66.67%	55.78%
Modelo de es...		50.00%	
Modelo ideal ...		100.00%	
Modelo ideal ...		100.00%	
Modelo ideal ...		96.67%	

Predicción de árbol de decisión

Resultado

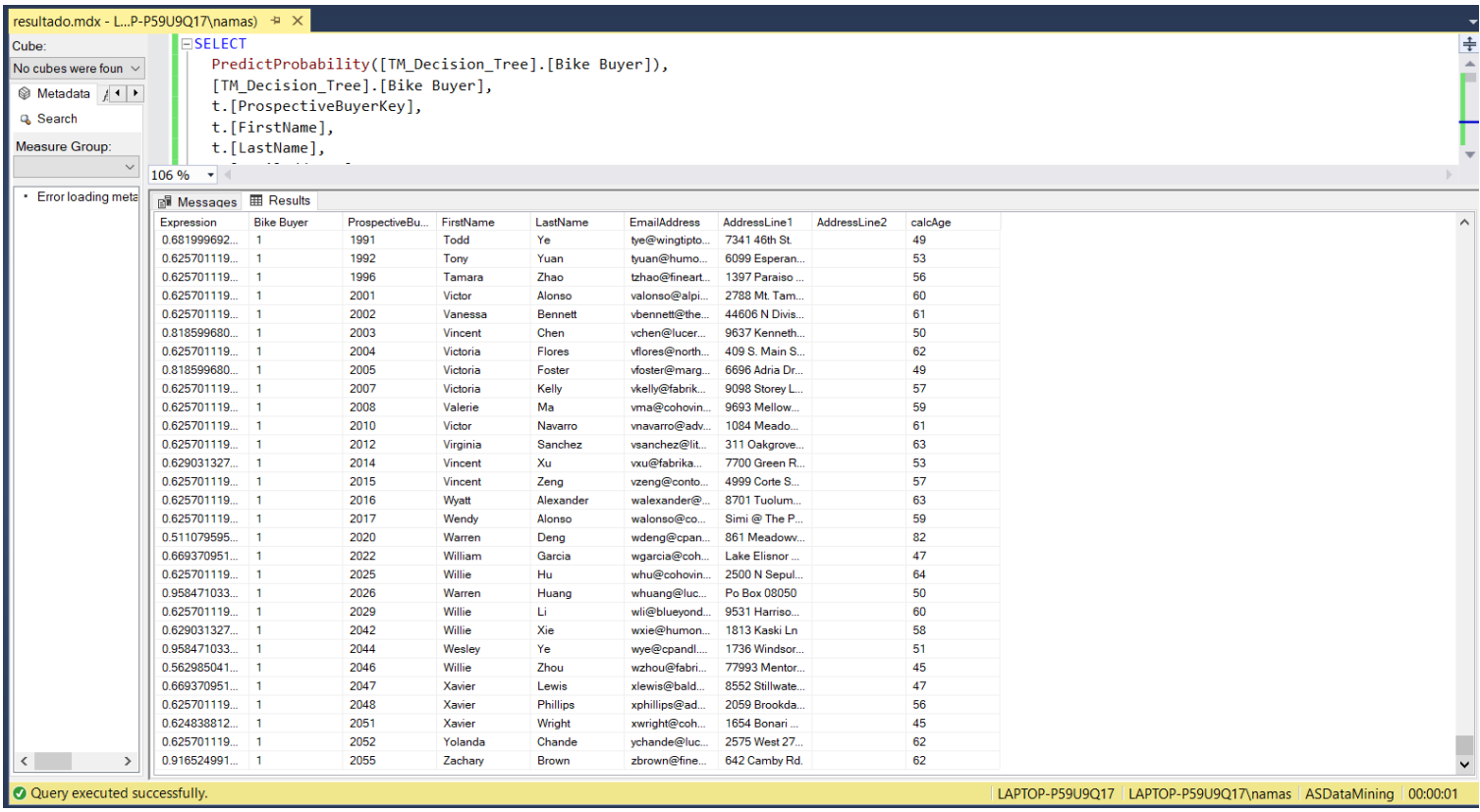


Expression	Bike Buyer	Prospective...	FirstName	LastName	EmailAddress	AddressLine1	AddressLine2	calcAge
0.5629850...	1	6	Angel	Bell	abell@thep...	840 Charlot...		45
0.8185996...	1	7	Anna	Bennett	abennett@...	312 Via Del ...		49
0.6257011...	1	9	Arturo	Bhat	abhat@ada...	7040 Isabel...		53
0.5629850...	1	11	Abigail	Bryant	abryant@n...	2639 Anch...		46
0.6257011...	1	15	Adam	Carter	acarter@m...	4082 Rosly...		60
0.6693709...	1	17	Alberto	Castro	acastro@th...	2696 Santa...		47
0.6693709...	1	18	Alicia	Chande	achande@a...	5935 Isabel		51
0.5524897...	1	19	April	Chande	achande@c...	7939 Bayvi...		53
0.5629850...	1	23	Alexander	Clark	aclark@hu...	3243 Lanto...		51
0.6257011...	1	29	Alexandria	Cooper	acooper@c...	1624 Carlisl...		52
0.6257011...	1	30	Alexis	Davis	adavis@co...	6900 Willa...		60
0.9165249...	1	31	Abigail	Davis	adavis@fab...	70 N.w. Plaza		64
0.8185996...	1	32	Alicia	Deng	adeng@cpa...	8958 Carlet...		49
0.6693709...	1	36	Andre	Engineer	aengineer@...	North 9327...		49
0.6693709...	1	37	Andre	Fernandez	afernandez...	2292 Sprin...		50
0.5110795...	1	38	Adriana	Fernandez	afernandez...	1314 Green...		82
0.6257011...	1	40	Adam	Foster	afoster@wo...	4524 Fernd...		63
0.6257011...	1	47	Alexandra	Gonzalez	agonzalez@...	9075 Calle ...		59
0.6693709...	1	48	Abigail	Gray	agray@con...	78025 E. M...		51
0.8185996...	1	50	Alyssa	Griffin	agriffin@co...	6916 Azores		49
0.6248388...	1	53	Alex	Hall	ahall@coho...	9115 Arthu...		46
0.6243700...	1	55	Abigail	Hall	ahall@north...	8036 Sum...		47
0.6257011...	1	57	Alejandro	He	ahe@fabrik...	Pyramid Mall		59
0.6257011...	1	67	Ana	Jenkins	ajenkins@h...	4769 Buch...	Unit G 202	65
0.6257011...	1	70	Armando	Jimenez	ajimenez@...	3880 Batacao		57
0.5110795...	1	71	Anna	Kelly	akelly@adat...	7750 E Mcd...		87
0.5524897...	1	77	Aimee	Liang	aliang@fabri...	1601 Crow...		54
0.6693709...	1	78	Aimee	Lin	alin@litwarei...	1133 Conc...		47
0.6257011...	1	79	Amy	Liu	aliu@fabrika...	6837 Pirate...		56
0.6693709...	1	80	Arianna	Long	along@coh...	1349 Palm ...		50
0.9584710...	1	84	Arthur	Lopez	alopez@co...	7549 Long...		51
0.6257011...	1	85	Abby	Malhotra	amalhotra...	1019 Carlet...		53
0.6248388...	1	86	Andre	Malhotra	amalhotra...	6968 Mildre...		47

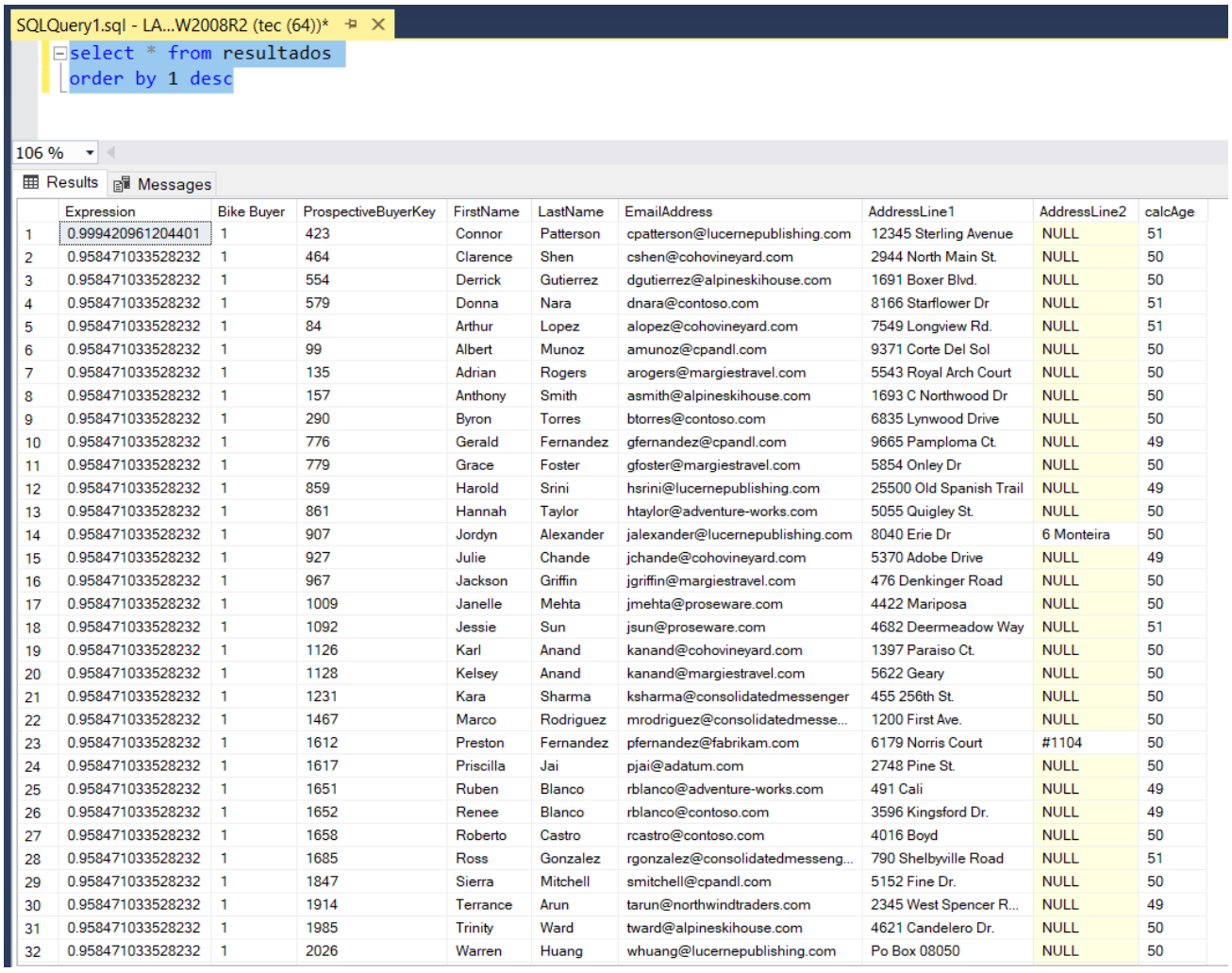
Ejecución de consulta finalizada con 914 filas recuperadas

Consulta en análisis services

```
SELECT
  PredictProbability([TM_Decision_Tree].[Bike Buyer]),
  [TM_Decision_Tree].[Bike Buyer],
  t.[ProspectiveBuyerKey],
  t.[FirstName],
  t.[LastName],
  t.[EmailAddress],
  t.[AddressLine1],
  t.[AddressLine2],
  t.[calcAge]
From
  [TM_Decision_Tree]
PREDICTION JOIN
  OPENQUERY([Adventure Works DW2008R2],
    'SELECT
      [ProspectiveBuyerKey],
      [FirstName],
      [LastName],
      [EmailAddress],
      [AddressLine1],
      [AddressLine2],
      (DATEDIFF(YYYY,[BirthDate],getdate())) AS [calcAge],
      [MaritalStatus],
      [Gender],
      [YearlyIncome],
      [TotalChildren],
      [NumberChildrenAtHome],
      [HouseOwnerFlag],
      [NumberCarsOwned],
      [Unknown]
    FROM
      [dbo].[ProspectiveBuyer]
    ') AS t
ON
  [TM_Decision_Tree].[Marital Status] =
t.[MaritalStatus] AND
  [TM_Decision_Tree].[Gender] = t.[Gender] AND
  [TM_Decision_Tree].[Yearly Income] =
t.[YearlyIncome] AND
  [TM_Decision_Tree].[Total Children] =
t.[TotalChildren] AND
  [TM_Decision_Tree].[Number Children At Home] =
t.[NumberChildrenAtHome] AND
  [TM_Decision_Tree].[House Owner Flag] =
t.[HouseOwnerFlag] AND
  [TM_Decision_Tree].[Number Cars Owned] =
t.[NumberCarsOwned] AND
  [TM_Decision_Tree].[Age] = t.[calcAge] AND
  [TM_Decision_Tree].[Bike Buyer] = t.[Unknown]
WHERE
  [TM_Decision_Tree].[Bike Buyer] =1
```



Consulta en sql server



	Expression	Bike Buyer	ProspectiveBuyerKey	FirstName	LastName	EmailAddress	AddressLine1	AddressLine2	calcAge
1	0.999420961204401	1	423	Connor	Patterson	cpatterson@lucernepublishing.com	12345 Sterling Avenue	NULL	51
2	0.958471033528232	1	464	Clarence	Shen	cshen@cohovineyard.com	2944 North Main St.	NULL	50
3	0.958471033528232	1	554	Derrick	Gutierrez	dgutierrez@alpineskihouse.com	1691 Boxer Blvd.	NULL	50
4	0.958471033528232	1	579	Donna	Nara	dnara@contoso.com	8166 Starflower Dr	NULL	51
5	0.958471033528232	1	84	Arthur	Lopez	alopez@cohovineyard.com	7549 Longview Rd.	NULL	51
6	0.958471033528232	1	99	Albert	Munoz	amunoz@cpandl.com	9371 Corte Del Sol	NULL	50
7	0.958471033528232	1	135	Adrian	Rogers	arogers@margiestravel.com	5543 Royal Arch Court	NULL	50
8	0.958471033528232	1	157	Anthony	Smith	asmith@alpineskihouse.com	1693 C Northwood Dr	NULL	50
9	0.958471033528232	1	290	Byron	Torres	btorres@contoso.com	6835 Lynwood Drive	NULL	50
10	0.958471033528232	1	776	Gerald	Fernandez	gfernandez@cpandl.com	9665 Pamploma Ct.	NULL	49
11	0.958471033528232	1	779	Grace	Foster	gfoster@margiestravel.com	5854 Onley Dr	NULL	50
12	0.958471033528232	1	859	Harold	Srini	hsrini@lucernepublishing.com	25500 Old Spanish Trail	NULL	49
13	0.958471033528232	1	861	Hannah	Taylor	htaylor@adventure-works.com	5055 Quigley St.	NULL	50
14	0.958471033528232	1	907	Jordyn	Alexander	jalexander@lucernepublishing.com	8040 Erie Dr	6 Monteira	50
15	0.958471033528232	1	927	Julie	Chande	jchande@cohovineyard.com	5370 Adobe Drive	NULL	49
16	0.958471033528232	1	967	Jackson	Griffin	jgriffin@margiestravel.com	476 Denkinger Road	NULL	50
17	0.958471033528232	1	1009	Janelle	Mehta	jmehta@proseware.com	4422 Mariposa	NULL	50
18	0.958471033528232	1	1092	Jessie	Sun	jsun@proseware.com	4682 Deermeadow Way	NULL	51
19	0.958471033528232	1	1126	Karl	Anand	kanand@cohovineyard.com	1397 Paraiso Ct.	NULL	50
20	0.958471033528232	1	1128	Kelsey	Anand	kanand@margiestravel.com	5622 Geary	NULL	50
21	0.958471033528232	1	1231	Kara	Sharma	ksharma@consolidatedmessenger	455 256th St.	NULL	50
22	0.958471033528232	1	1467	Marco	Rodriguez	mrodriguez@consolidatedmesse...	1200 First Ave.	NULL	50
23	0.958471033528232	1	1612	Preston	Fernandez	pfernandez@fabrikam.com	6179 Norris Court	#1104	50
24	0.958471033528232	1	1617	Priscilla	Jai	pjai@adatum.com	2748 Pine St.	NULL	50
25	0.958471033528232	1	1651	Ruben	Blanco	rblanco@adventure-works.com	491 Cali	NULL	49
26	0.958471033528232	1	1652	Renee	Blanco	rblanco@contoso.com	3596 Kingsford Dr.	NULL	49
27	0.958471033528232	1	1658	Roberto	Castro	rcastro@contoso.com	4016 Boyd	NULL	50
28	0.958471033528232	1	1685	Ross	Gonzalez	rgonzalez@consolidatedmesseng...	790 Shelbyville Road	NULL	51
29	0.958471033528232	1	1847	Sierra	Mitchell	smitchell@cpandl.com	5152 Fine Dr.	NULL	50
30	0.958471033528232	1	1914	Terrance	Arun	tarun@northwindtraders.com	2345 West Spencer R...	NULL	49
31	0.958471033528232	1	1985	Trinity	Ward	tward@alpineskihouse.com	4621 Candelero Dr.	NULL	50
32	0.958471033528232	1	2026	Warren	Huang	whuang@lucernepublishing.com	Po Box 08050	NULL	50