

## TAREA 2

<https://github.com/usac201612398/DataMining.git>

Se analizan las categorías del dataset para visualizar que segmentación pueden realizarse.

Características	Unicos
AÑO	1
CARRERA	1749
Departamento	23
Nivel_Educativo	6
Sexo	3
Edad	79
Grupos_Quinquenales	13
Sector	2
Pueblo_Pertenencia	7

Se elimina primeramente la columna de año ya que el 100% es 2023

De estas, ya se realizó análisis de la categoría Nivel\_Educativo, y se aprecia que en carrera existe demasiadas opciones en Edad y en CARRERA

Por lo tanto, se procede a inspeccionar la densidad de coincidencias con las categorías Departamento y Pueblo\_Pertenencia, Grupos quincenales para validar que tanto se concentran los datos en cada una de sus opciones.

En el caso de la columna edad se tiene un registro que contiene un elemento -7971, por lo tanto, se procedió a eliminarlo ya produce ruido en el análisis APRIORI

Etiquetas de fila	Cuenta de Pueblo_Pertenencia		Porcentaje
Ladino	18613		45%
Ignorado	18176		44%
Maya	3370		8%
Afrodescendiente/Creole/Afromestizo	835		2%
Extranjero	465		1%
Xinka	44		0%
Garífunas	26		0%
Total general	41529		100%

<b>Etiquetas de fila</b>	<b>Cuenta de Grupos_Quinquenales</b>	
	<b>Porcentaje</b>	
25 a 29	12627	30%
20 a 24	8350	20%
30 a 34	7283	18%
35 a 39	4332	10%
70 y más	2559	6%
40 a 44	2417	6%
45 a 49	1350	3%
Ignorado	1241	3%
50 a 54	751	2%
55 a 59	348	1%
60 a 64	157	0%
65 a 69	63	0%
15 a 19	51	0%
Total general	41529	100%

<b>Etiquetas de fila</b>	<b>Cuenta de Departamento</b>	<b>Porcentaje</b>
Guatemala	17942	43%
Quetzaltenango	3436	8%
Huehuetenango	1908	5%
Alta Verapaz	1537	4%
Sacatépequez	1510	4%
Jutiapa	1502	4%
San Marcos	1501	4%
Chimaltenango	1451	3%
Quiché	1418	3%
Escuintla	1407	3%
Santa Rosa	814	2%
Péten	804	2%
Sololá	804	2%
Chiquimula	797	2%
Suchitepéquez	777	2%
Izabal	690	2%
Retalhuleu	621	1%
Zacapa	610	1%
Totonicapán	494	1%
Jalapa	490	1%
Baja Verapaz	412	1%
El Progreso	371	1%
	99	1%
Total general	41529	100%

### Trabajando A PRIORI filtrando Guatemala tenemos

{Nivel\_Educativo=Maestría} => {Sector=Privado} 0.2024858 0.8374827  
 0.2417791 1.1736401 3633

{Sector=Público} => {Pueblo\_Pertenencia=Ladino} 0.2215472 0.7736473  
 0.2863672 1.4670028 3975

{Nivel\_Educativo=Licenciatura} => {Sector=Privado} 0.4340653 0.7008009  
 0.6193847 0.9820955 7788

### Trabajando con el grupo de 25 a 29

{Nivel\_Educativo=Licenciatura} => {Sector=Privado} 0.5147292 0.7516189  
0.6848274 1.0239986 6500

{Nivel\_Educativo=Licenciatura,  
Sexo=Hombre} => {Sector=Privado} 0.2018530 0.7345821  
0.2747862 1.0007879 2549

{Departamento=Guatemala} => {Pueblo\_Pertenencia=Ladino} 0.2473076  
0.5427529 0.4556541 1.1388972 3123

### Trabajando ahora con las mujeres

{Departamento=Guatemala} => {Nivel\_Educativo=Licenciatura} 0.2591937  
0.6145187 0.4217832 1.0780102 6442

{Nivel\_Educativo=Licenciatura} => {Sector=Privado} 0.4132936 0.7250141  
0.5700491 1.0473409 10272

{Pueblo\_Pertenencia=Ladino} => {Sector=Privado} 0.2804780 0.5879227  
0.4770661 0.8493015 6971

{Nivel\_Educativo=Técnico} => {Pueblo\_Pertenencia=Ladino} 0.1703146  
0.5850726 0.2911000 1.2263974 4233

Se nota que el patrón aplicando filtros en las diferentes categorías, siempre prevalece el Pueblo de Pertenencia Ignorado, además en la mayoría de concentración siempre prevalecen mujeres ladinas en el departamento de Guatemala y principalmente en el sector privado cuando se trata de una licenciatura.

En las edades de 25 a 29 se concentran las licenciaturas, en el sector privado y específicamente en el departamento de Guatemala.

Y por último al analizar el departamento de Guatemala, vemos que hay una tendencia a que ladinos estudien en el sector público, no obstante, si nos centramos en el nivel de licenciatura ya se concentra más el sector privado al igual que las maestrías. Es mayor el alcance para la gente de origen ladino el desarrollo educativo que para otros pueblos de pertenencia.