

Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Postgrado
Introducción a la Minería de Datos



Proyecto Parte 1

Glen Abra-ham Calel Robledo
999015052

Introducción

En el curso de introducción a la minería de datos se ha propuesto con el siguiente proyecto realizar la carga y limpieza de varias bases de datos de una misma fuente de datos, los datos son obtenidos desde el Instituto Nacional de Estadística de Guatemala, este organismo se encarga de recolectar, elaborar y publicar estadísticas oficiales con relación a distintos tipos de investigaciones que se realiza en conjunto con otras organizaciones. Para esta ocasión, el tema a evaluar serán las consultas externas que se han realizado en los servicios hospitalarios desde el año 2022 hasta el 2009. Después de haber hecho este proceso se aplicarán algoritmos que encuentren reglas de asociación de datos, de tal manera que se pueda realizar un análisis a las reglas que se encontraron con el fin de identificar patrones llamativos.

Apriori

Regla No.1

Lhs		Rhs	Support	Confidence	Coverage	Lift	Count
AÑO=[2020,2022], PPERTENENCIA=[4,9], PERIODOEDA=[3,9], DEPTORESIDENTEN=[1,9], TC=[2,9], CAUFIN=C910	=>	EDAD=[0,22)	0.00281	0.9897	0.00284	3.0943	21695

Para la obtención de esta regla se analizó el set de datos que abarca los años de 2022 hasta el año 2015. Se obtuvo que en el año 2020 al 2022 las personas metizo/ladiano entre otros, en un periodo de edad de años del departamento de Guatemala, El Progreso, Sacatepéquez, Chimaltenango, Escuintla, Santa Rosa, Sololá y Totonicapán; con un tipo de consulta: reconsulta, emergencia, primer consulta y emergencia, reconsulta y emergencia, se presentaron debido a la siguiente causa: Leucemia linfoblástica aguda [LLA], dado que se encuentra en un rango de edad de 0 a 21 años.

Según el Manual MSD (2023), indica que esta enfermedad es el cáncer pediátrico más frecuente; también afecta a adultos de todas las edades. Además también hace énfasis a que la palabra aguda se refiere a que puede progresar rápidamente comenzando con la médula ósea causando síntomas como anemia, trombocitopenia y granulocitopenia. El Manual MSD (2023) indica que el paciente debe ser tratado con 4 fases generales de la quimioterapia.

Regla No.2

Lhs		Rhs	Support	Confidence	Coverage	Lift	Count
AÑO=[2020,2022], SEXO=[2,9], PERIODOEDA=[3,9], TC=[1,2], CAUFIN=Z33X	=>	EDAD=[0,25)	0.00375	0.4076	0.0092	1.2250	6188

Para la obtención de esta regla se analizó el set de datos que abarca los años de 2022 hasta el año 2015, delimitando geográficamente la región de occidente que conforman los departamentos de Quetzaltenango, Sololá, Huehuetenango, San Marcos, Quiché y Totonicapán. Se pudo identificar que en el año 2020 al 2022 las mujeres en un periodo de edad de años con un tipo de primera consulta se

presentaron debido a la siguiente causa: Estado de embarazo incidental, dado que se encuentra en un rango de edad de 0 a 24 años.

La BBC (2020) publicó el siguiente artículo Guatemala, el país centroamericano donde 14 niñas quedan embarazadas toso los días, según este artículo al menos 5133 niñas de entre 10 y 14 años resultaron embarazadas en el año 2019. Debido a este análisis es muy probable que niñas hayan sido víctimas de este problema que afecta la sociedad teniendo en cuenta que el análisis indica que fueron menores de 24 años.

Regla No.3

Lhs		Rhs	Support	Confidence	Coverage	Lift	Count
PPERTENENCIA=[4,9], PERIODOEDA=[3,9], MUNIRESIDEN=0501, TC=[2,4]	=>	SEX0=[1,2)}	0.0717	0.5202	0.1379	1.2810	167

Para la obtención de esta regla se analizó el set de datos que abarca los años de 2022 hasta el año 2015, teniendo en cuenta que delimitaron las siguientes causas: Lepra [enfermedad de Hansen], Lepra indeterminada, Lepra tuberculoide, Lepra tuberculoide limítrofe, Lepra limítrofe, Lepra lepromatosa limítrofe, Lepra lepromatosa, Otras formas de lepra y Lepra, no especificada. En base a esto se determinó que las personas de sexo masculino con algún tipo de lepra asistieron a un tipo de consulta: reconsulta, emergencia y primer consulta, en un periodo de edad en años, en el municipio de Escuintla.

De acuerdo con la información proporcionada por el IGSS (2024), antes de 1940 la lepra se consideraba como una enfermedad incurable que causaba una muerte lenta, pero hoy en día tiene cura. Esta institución recomienda prestar atención a las lesiones que no se curan con tratamientos convencionales, ya que ese es el primer síntoma que puede indicar que la persona puede tener esta enfermedad.

Regla No.4

Lhs		Rhs	Support	Confidence	Coverage	Lift	Count
MES=[8,12], SEX0=[2,9], GRUPETNICO=[2,9], PERIODOEDA=[3,9], DEPTORESIDEN=[1,7), CAUFIN=Z33X	=>	EDAD=[18,40)	0.0028	0.8822	0.0032	2.6474	10717

Para la obtención de esta regla se analizó el set de datos que abarca los años de 2009 hasta el año 2014. Se identificó que, en el mes de agosto, septiembre, octubre, noviembre y diciembre, las mujeres del grupo étnico no indígena entre otros, en un periodo de edad de años, que residen en el departamento de Guatemala, El Progreso, Sacatepéquez, Chimaltenango, Escuintla y Santa Rosa fueron atendidas por un estado de embarazo incidental en un rango de edad de 18 a 39 años.

Este análisis parece indicar que hace más de 10 años se producían este tipo de incidentes en mujeres mayores de 18 años, si bien es cierto que en comparación a la regla 2 no se trata de menores de edad, aún así existe un problema social que afecta a las mujeres en general, un problema que se viene arrastrando desde hace muchos años y que hoy en día está presente y pareciera que afecta cada vez a más jóvenes y menores de edad.

Regla No.5

Lhs		Rhs	Support	Confidence	Coverage	Lift	Count
GRUPETNICO=[2,9], EDAD=[1,11), PERIODOEDA=[3,9], MUNIRESIDEN=0101, CAUFIN=J209	=>	SEX0=[1,2)	0.00267	0.5396	0.00496	1.42109	5418

Para la obtención de esta regla se analizó el set de datos que abarca los años de 2009 hasta el año 2014, limitando la región en el departamento de Guatemala, se observó que los niños entre 1 y 10 años de un grupo étnico no indígena, en un periodo de edad en años fueron atendido por la causa de Bronquitis aguda, no especificada siendo los más afectados los varones.

Según National Heart, Lung, and Blood Institute (2023) la bronquitis aguda, que se considera como una enfermedad muy frecuente, por lo general se debe a una infección, y puede ser contagiosa. A pesar de ser común su tratamiento no es complejo y la mayoría de las personas se recupera después de unos días o semanas.

FP-Growth

Regla No.1

Lhs		Rhs	Support	Confidence	Lift	Count
AÑO=[2020, 2022], PPERTENENCIA=[4, 9], EDAD=[39, 999], PERIODOEDA=[3, 9], TC=[3, 4], CAUFIN=I10X	=>	SEXO=[1, 2)	0.00267	0.364	0.815	1227

Para la obtención de esta regla se analizó el set de datos que abarca los años de 2022 hasta el año 2015, teniendo en cuenta únicamente los tipos de consulta de emergencia y primer consulta y emergencia, del municipio de Guatemala, se determinó que en el año 2020 y 2022 siendo mestizo / ladino entre otros, en un periodo de edad en años, fueron atendidos a causa de hipertensión esencial (primaria) hombres mayores a 39 años.

Según ADA (2022) Se denomina hipertensión esencial o hipertensión primaria a la presión arterial alta (más de 140/90 mmHg) sin una causa reconocible. Los daños efectos que causa esta enfermedad pueden manifestarse en dolores de cabeza punzantes, visión borrosa, mareos, inflamación en los pies y dificultad para respirar.

Teniendo en cuenta que en esos años se estaba viviendo la pandemia, es posible que los hombres mayores de 39 años hayan sido afectados a causa del encierro.

Regla No.2

Lhs		Rhs	Support	Confidence	Lift	Count
SEXO=[1, 2), PERIODOEDA=[3, 9], CAUFIN=F809}	=>	EDAD=[1, 18)	0.00260	0.9973	3.0612	9178

Para la obtención de esta regla se analizó el set de datos que abarca los años de 2022 hasta el año 2015, delimitados por el departamento de Guatemala, se identificó que los hombres en un periodo de edad en años asistieron a la consulta externa a causa de Trastorno del desarrollo del habla y del lenguaje no especificado dado que estaban en un rango de edad de 1 a 17 años.

El National Institute on Deafness and Other Communication Disorders (2023) indica que el trastorno del desarrollo del lenguaje es un trastorno de la comunicación que interfiere con el aprendizaje, la comprensión y el uso del lenguaje. Además, indica

que es causado por los trastornos del neurodesarrollo son causados por interacciones complejas entre los genes y el entorno que modifican el desarrollo del cerebro.

Regla No.3

Lhs		Rhs	Support	Confidence	Lift	Count
SEX0=[1,2), PPERTENENCIA=[4,9], PERIODOEDA=[3,9], CAUFIN=J304	=>	EDAD=[1,18)	0.002668	0.9363	2.8738	9415

Para la obtención de esta regla se analizó el set de datos que abarca los años de 2022 hasta el año 2015, delimitados por el municipio de Villa Nueva, se identificó que los hombres mestizos / ladinos, en un periodo de edad en años asistieron a la consulta externa a causa de Rinitis alérgica, no especificada dado que eran menores de 18 años.

Según el Manual MSD (2022) indica que se debe a la exposición a lo largo de todo el año a alérgenos inhalados de interior (p. ej., heces de los ácaros del polvo, componentes de cucarachas, caspa de animales) o por una fuerte reactividad a pólenes de plantas en estaciones secuenciales. Además, describe que el tratamiento puede ser antihistamínicos o descongestivos.

Regla No.4

Lhs		Rhs	Support	Confidence	Lift	Count
GRUPETNICO=[2,9], EDAD=[40,999], PERIODOEDA=[3,9], CAUFIN=N40X	=>	SEX0=[1,2)	0.002745	1	2.7380	10180

Para la obtención de esta regla se analizó el set de datos que abarca los años de 2009 hasta el año 2014, se identificó que personas de un grupo étnico no indígena mayores de 40 años en un periodo de edad en años asistieron a una consulta externa a causa de Hiperplasia de la próstata, siendo únicamente hombres los que asistieron.

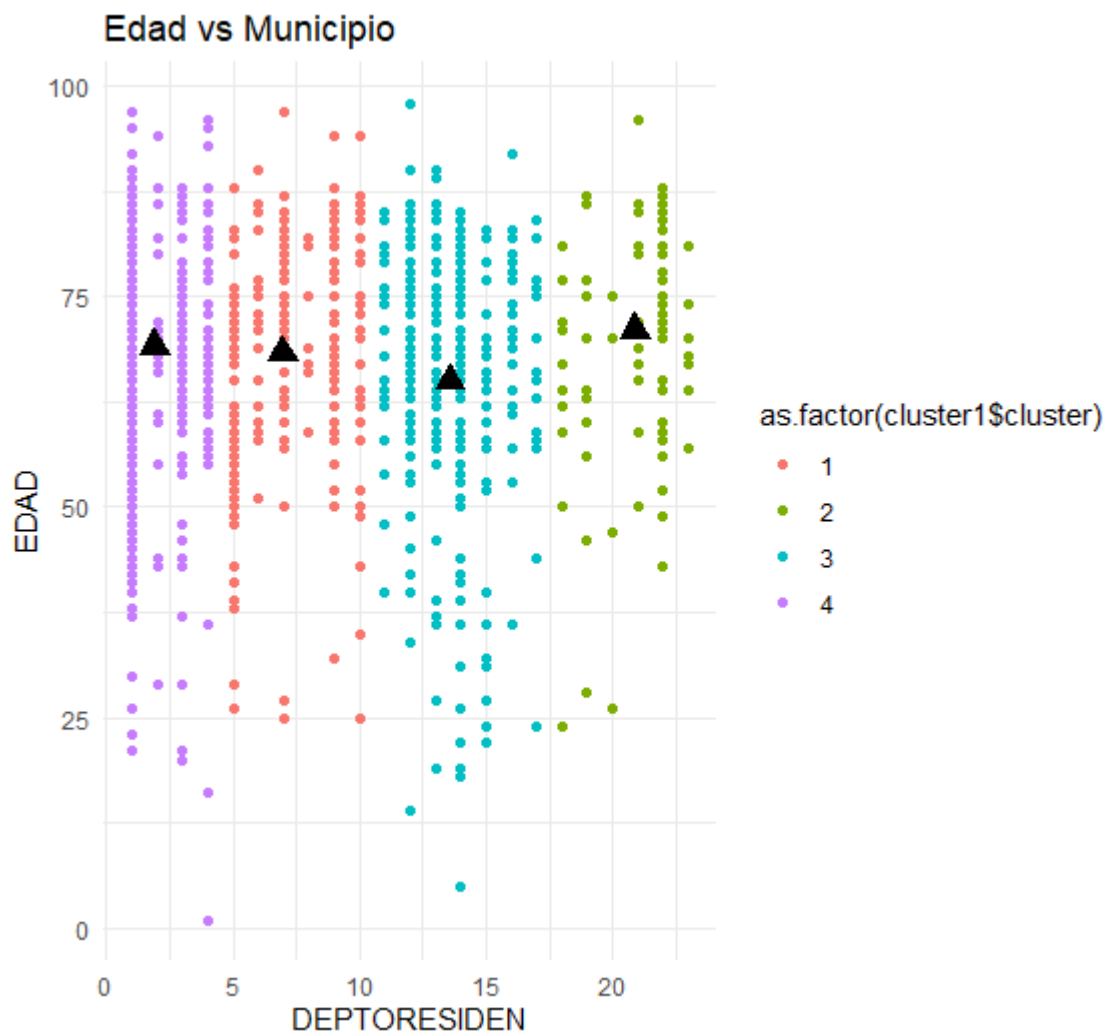
Según la Clínica Universidad de Navarra (s.f.) La hiperplasia o hipertrofia benigna de próstata (HBP) es el crecimiento benigno de la glándula prostática en el varón, esta enfermedad provoca una obstrucción a nivel del conducto de la orina que causa dificultad para vaciar la orina almacenada en la vejiga.

Regla No.5

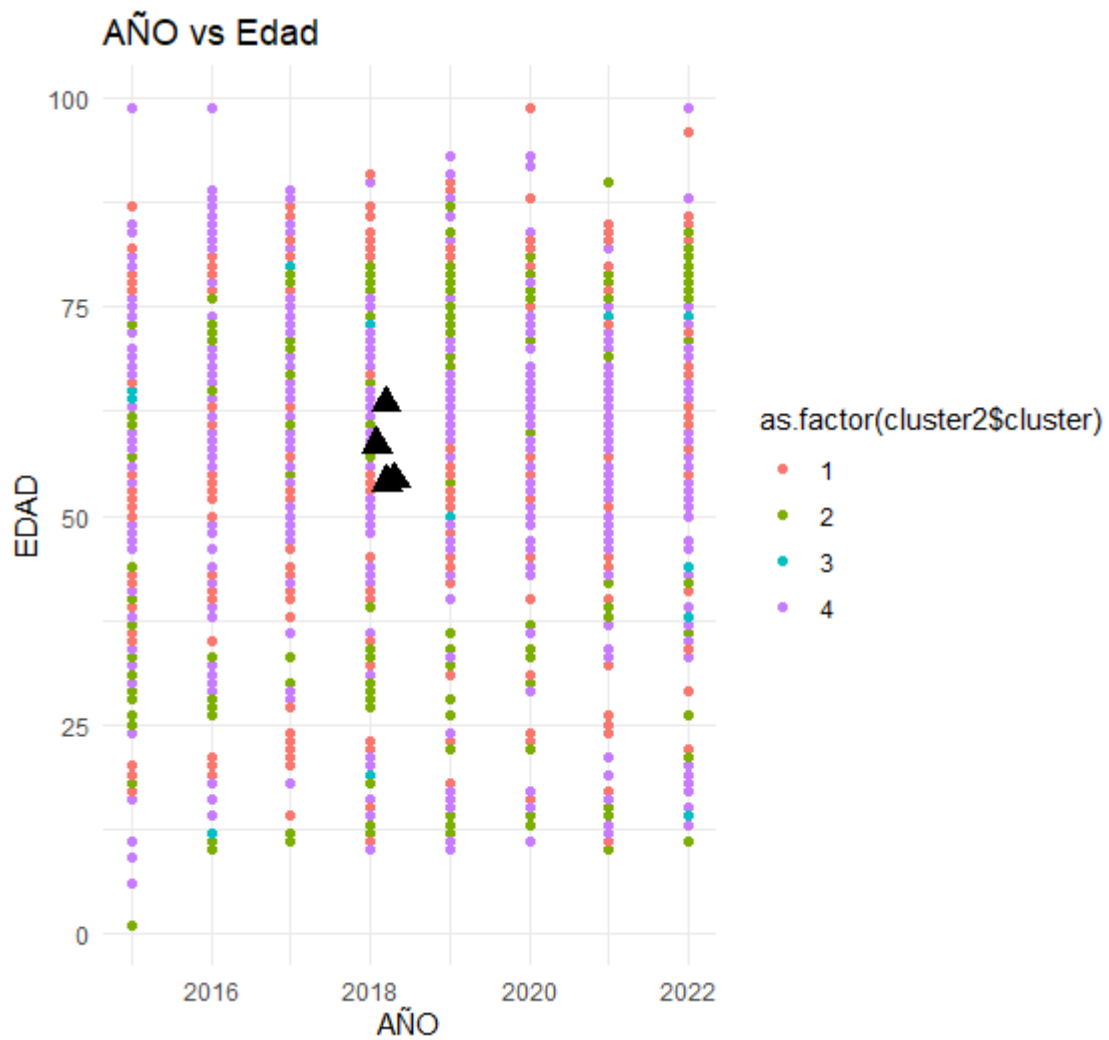
Lhs		Rhs	Support	Confidence	Lift	Count
AÑO=[2011,2013), GRUPETNICO=[2,9], EDAD=[23,45), PERIODOEDA=[3,9], DEPTORESIDEN=[12,19), CAUFIN=Z348	=>	SEXO=[2,9]	0.004092	1	1.5371	1710

Para la obtención de esta regla se analizó el set de datos que abarca los años de 2009 hasta el año 2014, delimitando geográficamente los departamentos que conforman el corredor seco, se determinó que en el año 2011 al 2013, mujeres de un grupo étnico no indígena, en un rango de edad de 23 años hasta los 44, en un periodo de edad en años asistieron a la consulta externa a causa de Supervisión de otros embarazos normales.

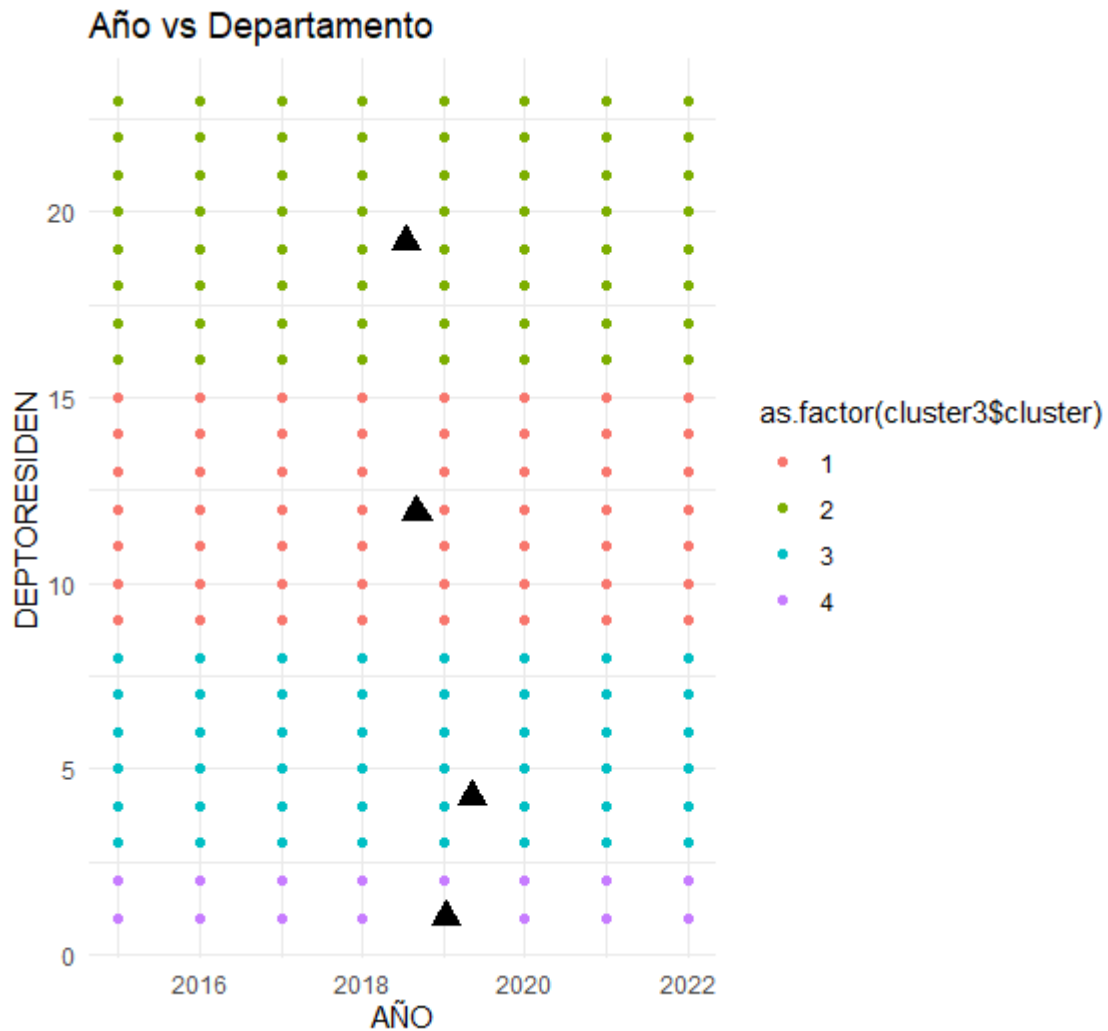
K-Means



Para la obtención de este clúster se se analizó el set de datos que abarca los años de 2022 hasta el año 2015. En este diagrama podemos ver los 4 clúster que describen el los grupos de las edades de las personas y el departamento donde residen que fueron a consulta externa a causa del Parkinson. Como se observa en el diagrama el departamento de Guatemala ha registrado una mayor frecuencia de edades que van desde los 20 años hasta los 100 años.



Para la obtención de este clúster se se analizó el set de datos que abarca los años de 2022 hasta el año 2015. En este diagrama podemos ver como la Lepra ha sido tratada en personas menores de edad y personas de 90 años aproximadamente. Como se observa los 4 grupos se centran en el año 2018.



Para la obtención de este clúster se se analizó el set de datos que abarca los años de 2022 hasta el año 2015, teniendo en cuenta las enfermedades como Miopía y Astigmatismo, los departamentos que más han reportado estas enfermedades son los siguientes:

Baja Verapaz

Alta Verapaz

Petén

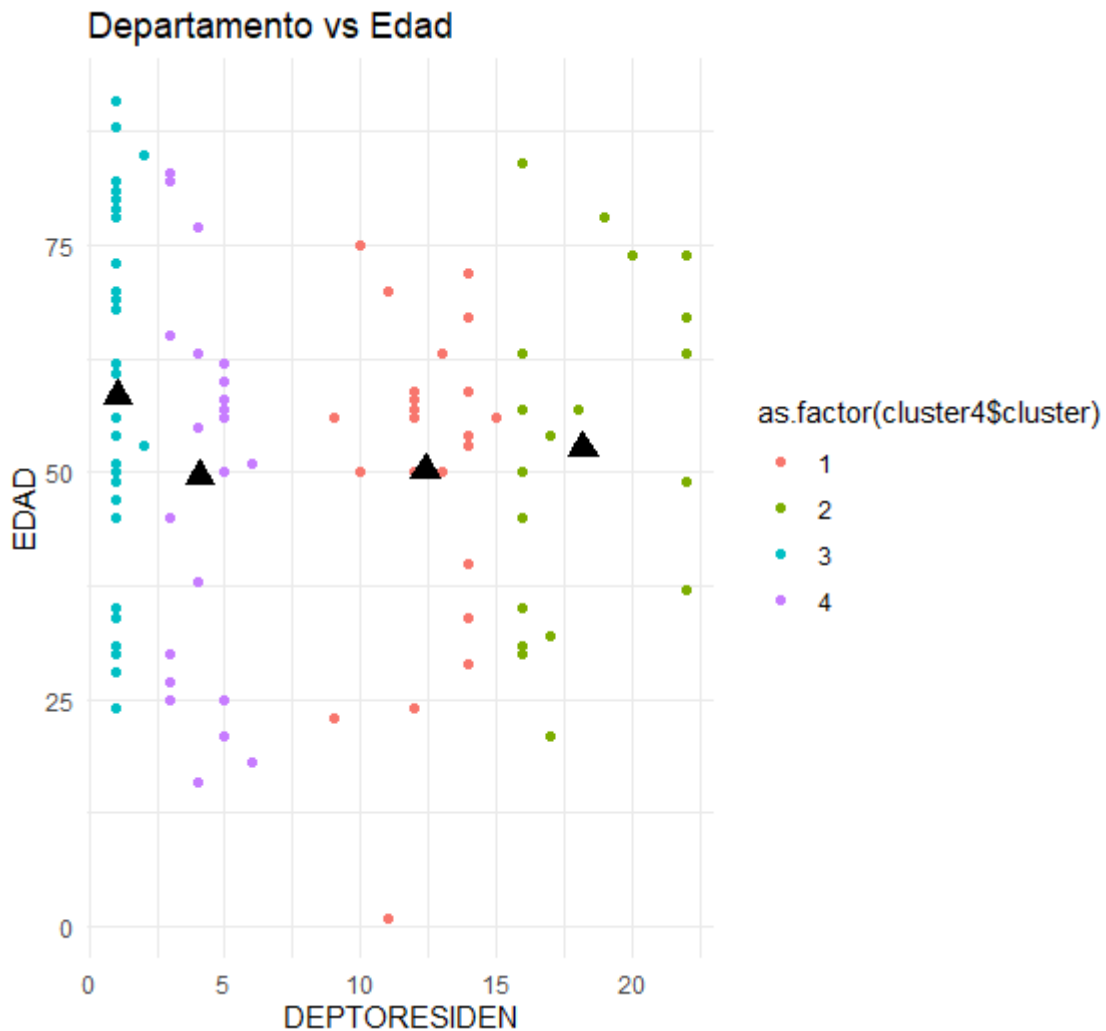
Izabal

Zacapa

Chiquimula

Jalapa

Jutiapa



Para la obtención de este clúster se se analizó el set de datos que abarca los años de 2022 hasta el año 2015, teniendo en cuenta las casas tales como Trastornos del inicio y del mantenimiento del sueño [insomnios] e Insomnio no orgánico, el departamento de Guatemala presenta la mayor parte de casos teniendo en cuenta que se abarca una de edad desde los 24 años en adelante. Y otra observación es que el primer grupo es el que cuenta con menos casos con relación a diversas edades.

Bibliografía

- Disorders, N. I. (2023 de Marzo de 2023). *National Institute ON Deafness and Other Communication Disorders*. Obtenido de National Institute ON Deafness and Other Communication Disorders: <https://www.nidcd.nih.gov/es/espanol/trastorno-del-desarrollo-del-lenguaje>
- Emadi, A. (Octubre de 2023). *Manual MSD*. Obtenido de Manual MSD: <https://www.msdmanuals.com/es/professional/hematolog%C3%ADa-y-oncolog%C3%ADa/leucemias/leucemia-linfobl%C3%A1stica-aguda-lla>
- IGSS. (28 de Enero de 2024). *IGSS*. Obtenido de IGSS: <https://www.igssgt.org/noticias/2024/01/28/lepra-una-enfermedad-curable-que-continua-creando-estigmas-que-dificultan-su-diagnostico-y-tratamiento/>
- James Fernandez. (Octubre de 2022). *Manual MSD*. Obtenido de Manual MSD: <https://www.msdmanuals.com/es/professional/inmunolog%C3%ADa-y-trastornos-al%C3%A9rgicos/enfermedades-al%C3%A9rgicas-autoinmunitarias-y-otros-trastornos-por-hipersensibilidad/rinitis-al%C3%A9rgica>
- Médicos, E. d. (25 de Julio de 2022). *ADA*. Obtenido de ADA: Se denomina hipertensión esencial o hipertensión primaria a la presión arterial alta (más de 140/90 mmHg) sin una causa reconocible.
- National Heart, L. a. (12 de Enero de 2023). *National Heart, Lung, and Blood Institute*. Obtenido de National Heart, Lung, and Blood Institute: <https://www.nhlbi.nih.gov/es/salud/bronquitis>
- Navarra, C. U. (s.f.). *Clinica Universidad de Navarra*. Obtenido de Clinica Universidad de Navarra: <https://www.cun.es/enfermedades-tratamientos/enfermedades/hipertrofia-benigna-prostata>
- Wallace, A. (14 de Octubre de 2020). *BBC*. Obtenido de BBC: <https://www.bbc.com/mundo/noticias-america-latina-54484133>

Enlace al repositorio

https://github.com/glencalel/data_mining_p1