**Tabla de contenido**

[**Preguntas de negocio a resolver** 1](#_Toc146199792)

[**Exploración de datos** 1](#_Toc146199793)

[**Red bayesiana** 1](#_Toc146199794)

[**Producto final** 1](#_Toc146199795)

# **Preguntas de negocio a resolver**

# **Exploración de datos**

**Tabla

Descripción generada automáticamente**Teniendo en cuenta que para la realización de este análisis es fundamental poder contar con las variables que realmente son significativas para el mismo, se realizó una prueba de regresión lineal para poder determinar cuáles son dichas variables por utilizar.

Como se puede evidenciar, y teniendo que el nivel de significancia de este análisis será de 95%, las variables que se utilizarán serán: “Course”, Mother’s quialification”, Mother’s occupation”, “Tuition fees up to date”, “Age at enrollment”, “Circular units 2nd sem (grade)” y “Unemployment rate”. La razón por la cual se escogió “Circular units 2nd sem (grade)” en lugar de otro tipo de notas disponibles en los datos fue que se consideró el que el promedio de notas obtenido en el segundo semestre podía tener más información que otro tipo de notas dadas, además que tenía un nivel de significancia mucho menor a 0.05.

**Imagen 1. Análisis de significancia individual de las variables.**

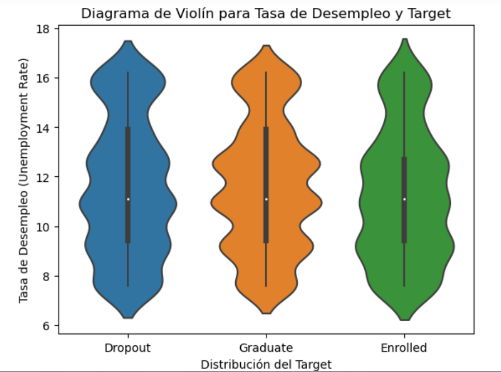
Ahora, ya teniendo en cuenta las varialbes que se usarán para este análisis, fue necesario discretizar las variables de edad y del promedio de notas en segundo semestre. Dicha discretización por intervalos (bins) se realizó teniendo en cuenta la regla de Sturges, la cual indica que el numero de intervalos se puede hallar como:

A pie chart with numbers and text

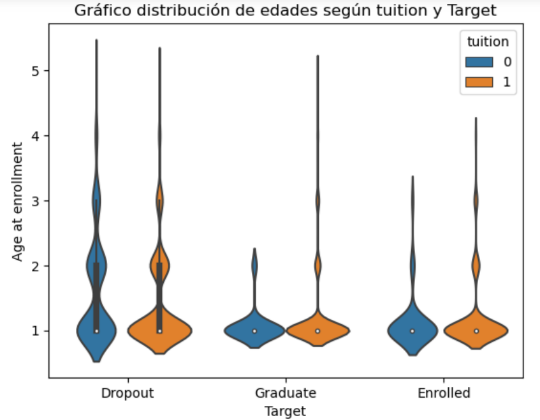
Description automatically generatedA graph with blue bars

Description automatically generatedPrimero, se evaluó específicamente el comportamiento de la variable objetivo: “target”, la cual indica el estado en el que se encuentra el estudiante, es decir si este se graduó, está matriculado o deserto. Tal como se puede observar en la imagen 5 y 6, el 49,9% de los estudiantes se han graduado, lo cual corresponde a más de 2000, así mismo el 32.1% de los estudiantes son *Dropuots* lo cual corresponde a casi 1500 estudiantes y el 17.9% de los estudiantes siguen afiliados actualmente, lo cual son menos de 1000.

**Imagen 5 y 6. Distribución del Target**

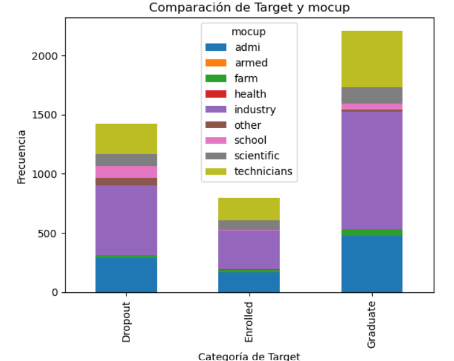
Ahora, para evaluar la interacción que tiene la variable objetivo con algunas de las variables explicativas se deben hacer algunas comparaciones. Por ejemplo, en la imagen 7 se puede ver la interacción de la Tasa de desempleo (Unemployment rate) con target. Con la gráfica se puede entender que la distribución de acuerdo con cada una de las categorías de target. Según lo anterior, se puede decir que las medianas y moda de las tres categorías son muy similares, teniendo una mediana cercanas a 11. Además, que los estudiantes que están graduados son los menos dispersos, pero con poca diferencia. Así mismo, en los tres grupos el menor número de estudiantes tienen una tasa de desempleo de más o menos 15.

**Imagen 7. Target vs Tasa Desempleo**

Por otro lado, se analizaron las interacciones entre las variables explicativas, “tuition” que indica si la matricula esta al día, las edades y la variable Target. De acuerdo con esto, se puede decir que generalmente hay menos dispersión cuando la matricula no está al día y de manera similar en la categoría dropout es donde se evidencia la mayor dispersión. Así mismo en cualquier situación en su mayoría el rango de edad en el que se matriculan es de 17 a 28 años. También, se puede ver que en la categoría de Dropout es donde hay más incorporación en mayores rangos de edad.

**Imagen 8. Target según edad y tuition**

A graph of a bar chart

Description automatically generated with medium confidencePor último, se evaluó las interacciones que tienen las características de la madre: ocupación y nivel de educación con el estado de matrícula de los estudiantes. Por un lado, se puede observar que la gran mayoría de las madres trabajan en el sector de la industria o tienen un nivel básico de educación. Adicionalmente, los estudiantes en la categoría de graduados tienen un mayor número de madres técnicas y científicas en comparación con los demás, de igual forma esta categoría cuanta con un mayor número de madres que han alcanzado niveles de secundaria o más altos. De los estudiantes activos se conoce el nivel educativo de las madres de casi todas y esta es la categoría donde las madres tienen menos participación en colegios y otros. Finalmente, al analizar la categoría Dropout tiene mayor variedad en el número de madres que participan en distintos sectores como su ocupación, de igual manera esta es la categoría en la cual tiene mayor número de madres con niveles de educación desconocidos, lo cual posiblemente puede afectar que los estudiantes abandonen sus estudios.

**Imagen 10. Target y mquali**

**Imagen 9. Target y mocup**

# **Red bayesiana**

# **Producto final**