**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:

# **Red bayesiana**

# **Producto final**