**Introducción**

* + Motivación del estudio.
  + Definición del problema.
  + Objetivos del análisis.

**Abstract**

Este informe se motiva por la creciente preocupación sobre el impacto del consumo de alcohol en el rendimiento académico de los estudiantes universitarios. A medida que las instituciones educativas buscan mejorar los resultados académicos y promover un ambiente saludable para sus estudiantes, entender cómo factores externos como el consumo de alcohol afectan el rendimiento es crucial. La audiencia que se podría beneficiar de este análisis incluye administradores educativos, políticas públicas, profesionales de la salud mental y educadores que buscan desarrollar estrategias efectivas para apoyar el bienestar y éxito académico de los estudiantes.

**Preguntas/Hipótesis**

El análisis de datos se centra en responder las siguientes preguntas e hipótesis:

1. **¿Existe una relación significativa entre el consumo de alcohol y el rendimiento académico de los estudiantes?**
   * Hipótesis: Un mayor consumo de alcohol está asociado con una disminución en el rendimiento académico.
2. **¿El consumo de alcohol difiere significativamente entre hombres y mujeres?**
   * Hipótesis: Los hombres consumen más alcohol que las mujeres.
3. **¿El apoyo familiar y la relación con los padres influyen en los patrones de consumo de alcohol de los estudiantes?**
   * Hipótesis: Los estudiantes con mayor apoyo familiar consumen menos alcohol.
4. **¿La frecuencia de salidas sociales impacta el rendimiento académico?**
   * Hipótesis: Una mayor frecuencia de salidas sociales está correlacionada con un menor rendimiento académico.

**Descripción de la Base de Datos**

* + Detalles sobre las variables demográficas, académicas y de estilo de vida. Procedencia y características de la muestra

**Información Demográfica**: Incluye datos como género y año académico del estudiante.

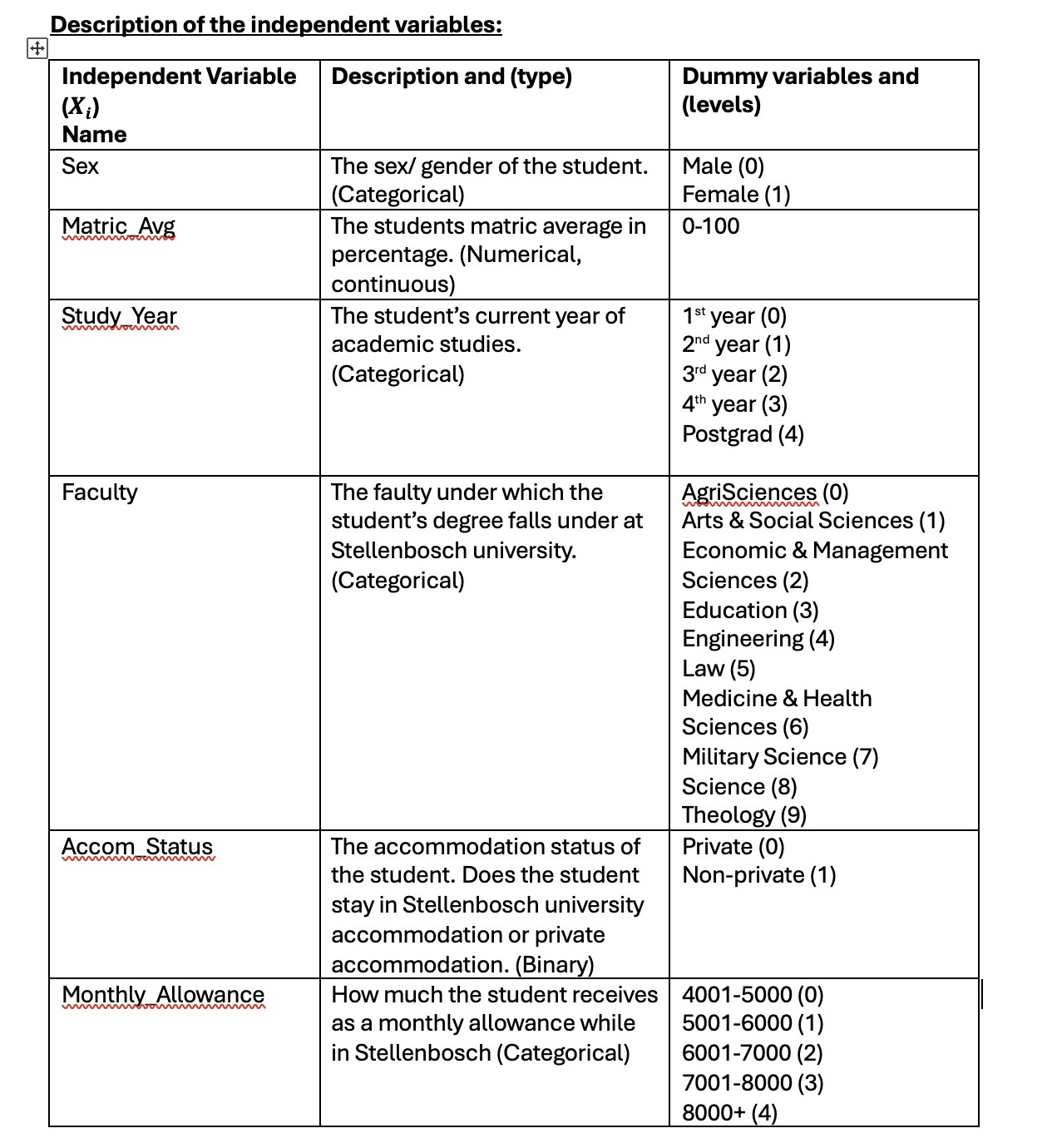
**Rendimiento Académico**: Se recoge tanto el promedio de notas del último año de secundaria (bachillerato) como el promedio académico del año 2023.

**Facultad**: Indica a qué facultad pertenece el grado que está cursando el estudiante.

**Estilo de Vida y Hábitos Personales**:

* + **Alojamiento**: Estado de alojamiento del estudiante durante el año 2023.
  + **Asignación Mensual**: Información sobre la cantidad de dinero que recibe el estudiante cada mes.
  + **Beca o Subsidio**: Si el estudiante estaba recibiendo algún tipo de beca o subsidio durante el 2023.
  + **Horas de estudio adicional**: Cantidad de horas de estudio adicional por semana.
  + **Socialización y Consumo de Alcohol**: Frecuencia de salidas a socializar o fiestas durante la semana, cantidad de bebidas alcohólicas consumidas durante estas salidas, y número de clases perdidas por razones relacionadas con el alcohol.
  + **Relación Amorosa**: Si el estudiante tiene una relación amorosa actualmente.
  + **Aprobación de los Padres sobre el Consumo de Alcohol** y **Relación con los Padres**: Explora la dinámica familiar respecto al consumo de alcohol y la fortaleza de la relación con los padres.
  + **Rendimiento Académico Adverso**: Incluye variables como el número de módulos académicos reprobados.

Descripción de variables independientes presentes en la base:

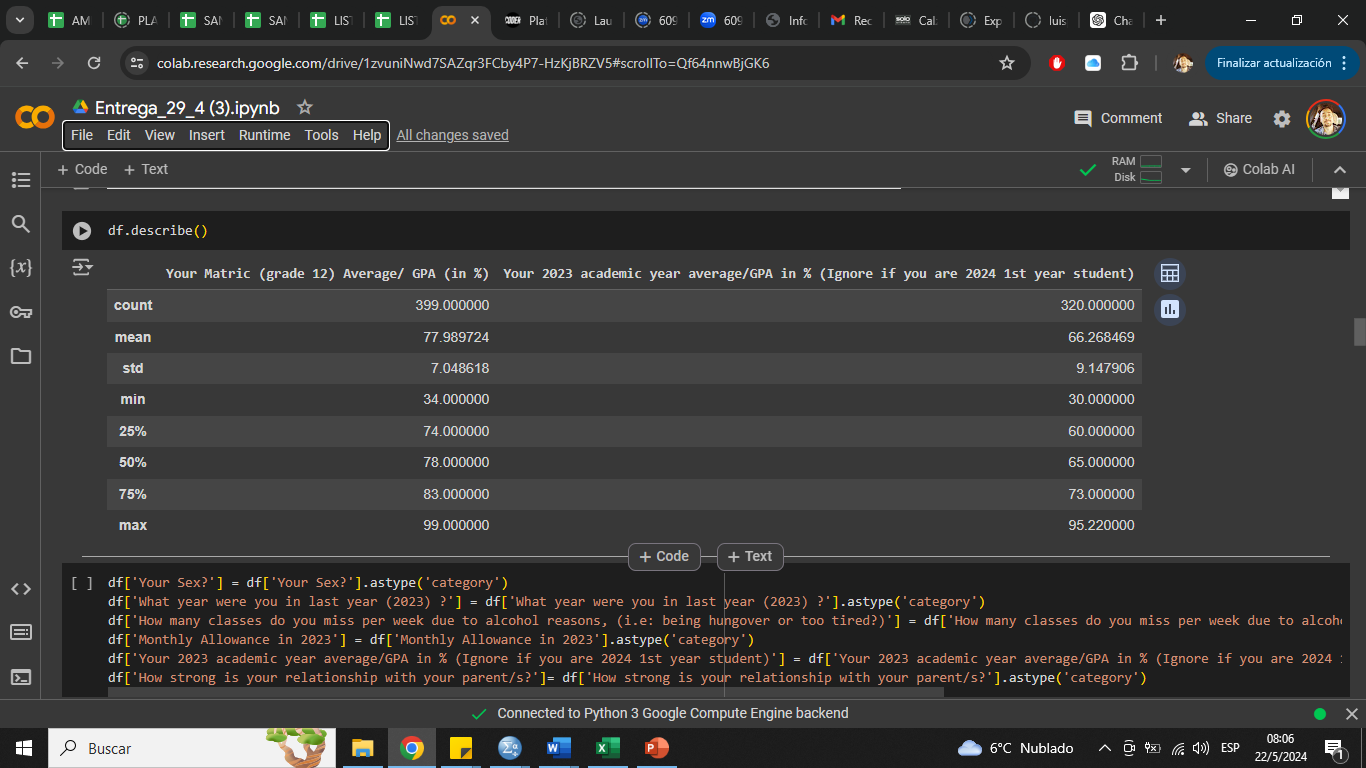
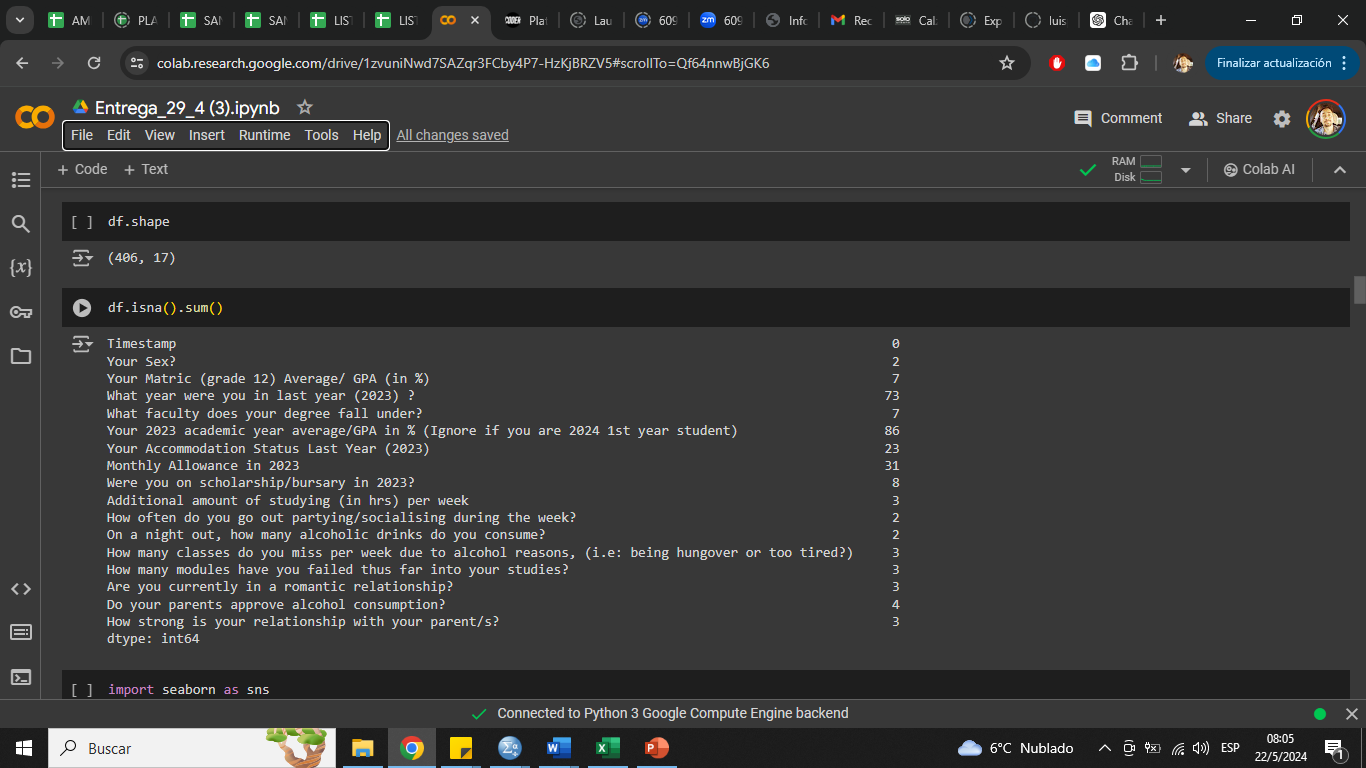
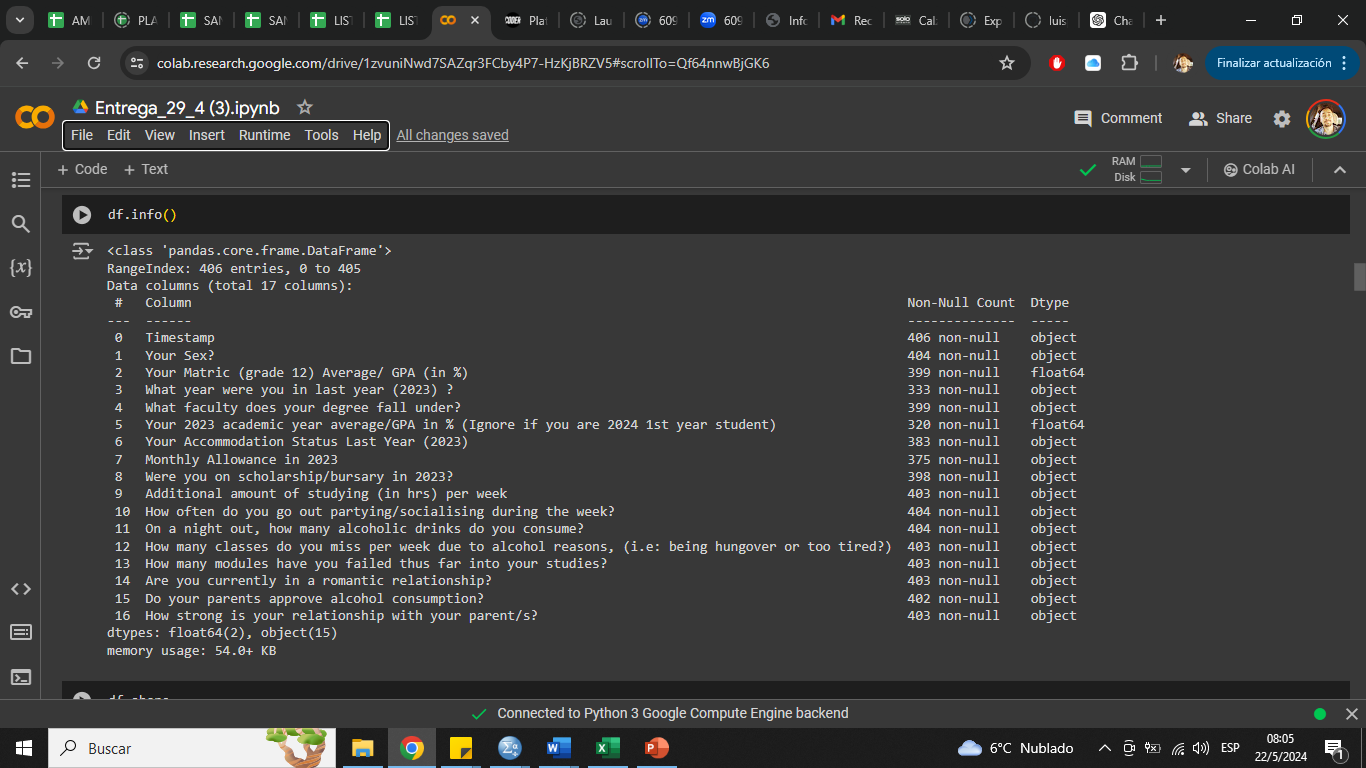
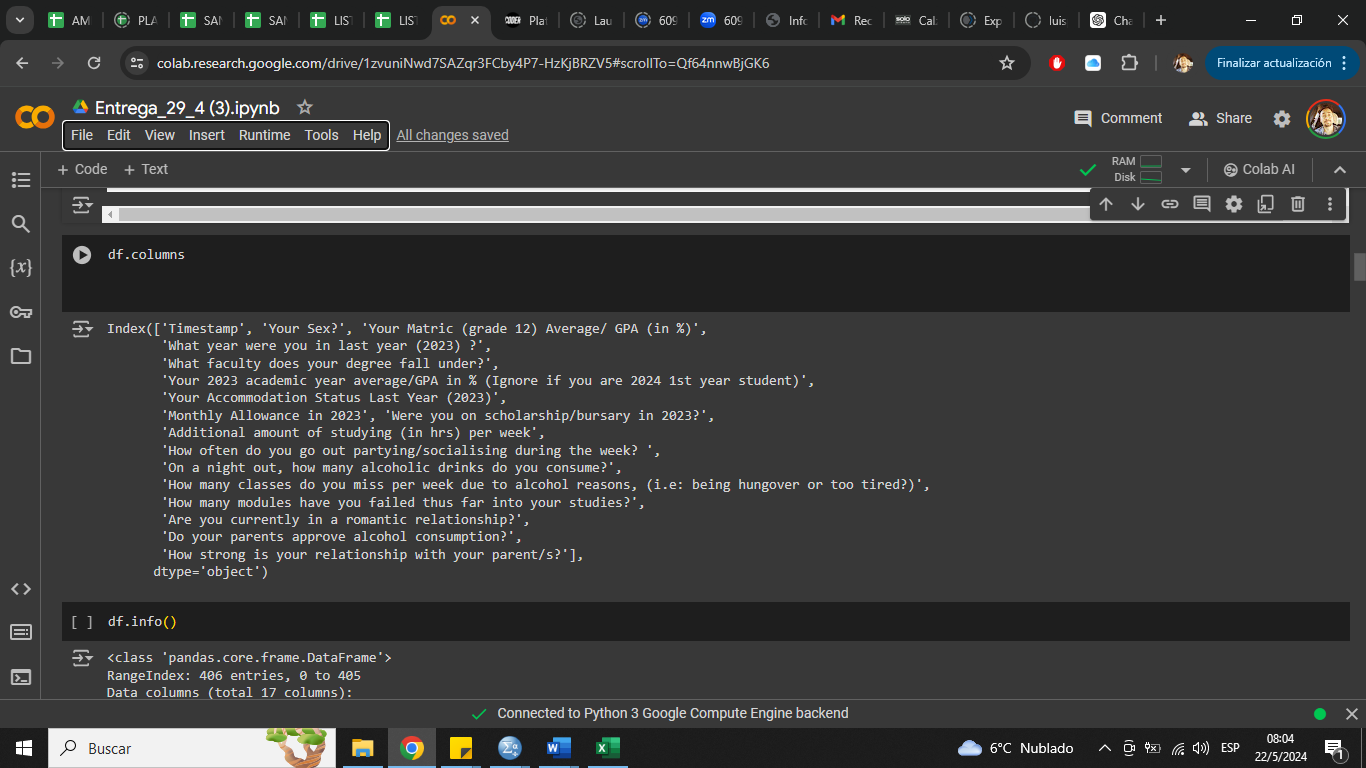


**Metodología, análisis exploratorio de datos y resultados preliminares**

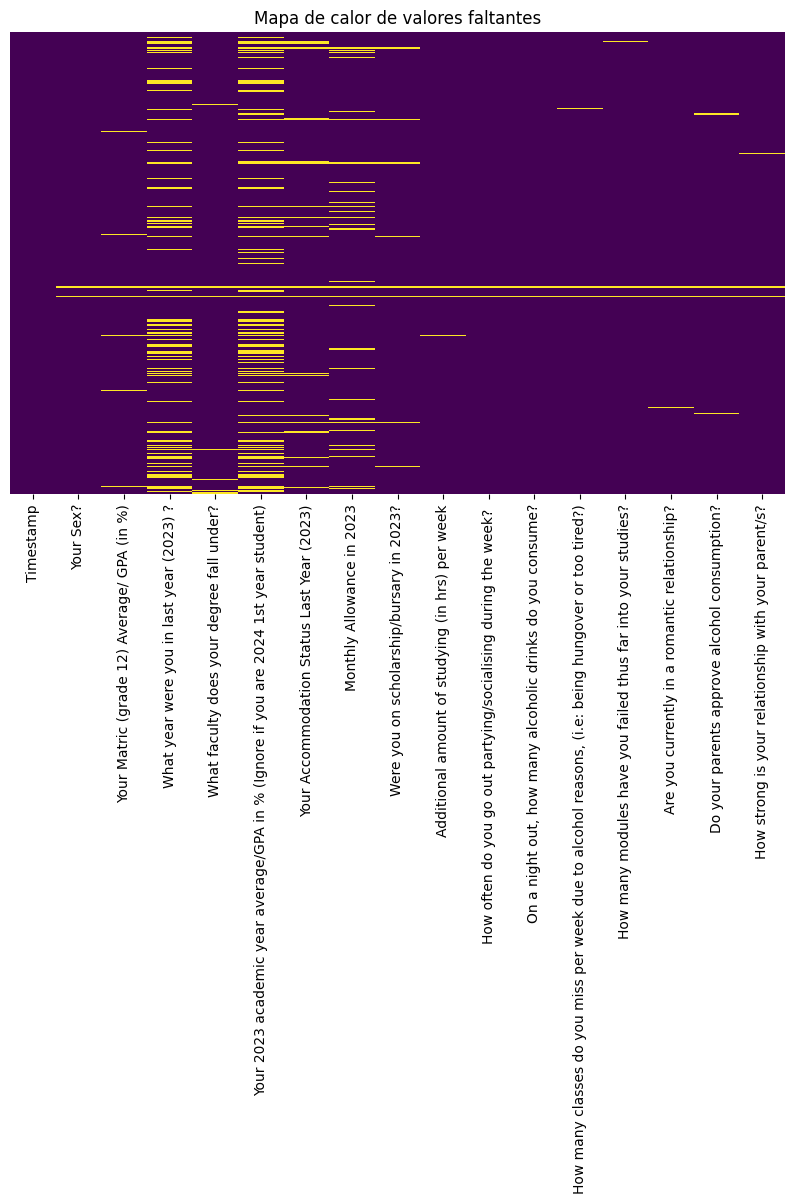
* + Mapas de calor y gráficos iniciales para evaluar la calidad de los datos y distribución de variables.
  + Descripción de los gráficos utilizados para explorar las hipótesis planteadas.
  + Procedimientos utilizados para la limpieza y análisis de datos.
  + Técnicas gráficas y estadísticas empleadas.

Resumen: El primer gráfico realizado en el análisis consiste en un mapa de calor para poder evaluar visualmente la calidad de los datos, y la presencia de NaN. Dado que el número de datos faltantes en las variables a analizar es bajo, no es necesario realizar una imputación de datos.

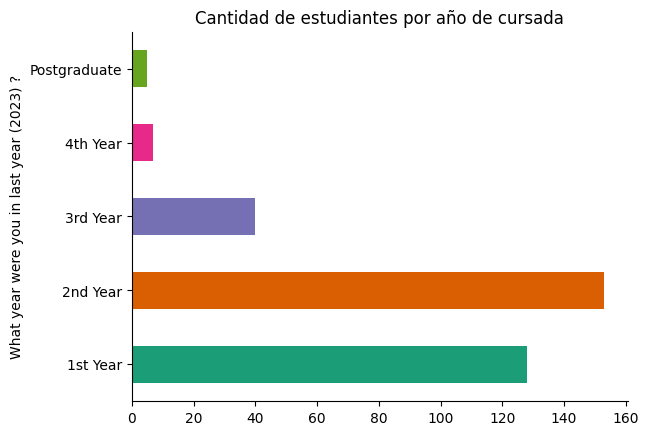
1. Luego procedo a realizar un análisis exploratorio de los datos, mediante diversos gráficos (realizados en matplotlib y seaborn), para ello comienzo graficando la cantidad de estudiantes que hay por año de cursada, la cantidad de clases perdidas por consumo de alcohol, la conformación de la sexo de la muestra, la mensualidad de los estudiantes y la distribución de la muestra de acuerdo al promedio académico. Asimismo gráfico el promedio académico para mujeres y para varones, para descartar que de antemano existan diferencias significativas en el rendimiento por sexo.
2. En esta etapa exploraré la hipótesis secundaria, que **el consumo de alcohol es mayor en hombres que en mujeres.** La hipótesis es confirmada como puede observarse en el heatmap graficado, una mayor cantidad de varones reportan consumir 8 bebidas alcohólicas o mas por noche.
3. Para evaluar si aquellas **personas con mejor rendimiento académico consumen menos bebidas alcohólicas,** realizo un gráfico de frecuencias, buscando observar gráficamente alguna tendencia. No hay tendencia distinguible, tanto individuos con buen promedio como con bajo promedio se distribuyen uniformemente a lo largo de las variables de consumo de alcohol.
4. En esta instancia **busqué evaluar si el patrón de consumo de alcohol se relaciona con el apoyo familiar**, medido como cercanía a los padres. No se observa ningún patrón definitivo en el heatmap.
5. Mediante gráfico de boxplot se **buscó evaluar si la cantidad de veces que un individuo sale por semana impacata en el rendimiento académico**, nuevamente no se observó ninguna tendencia puntual.
6. Finalmente mediante gráfico de barras apiladas se buscó **evaluar gráficamente si la cantidad de bebidas consumidas por noche impactaba en el promedio**. No observándose ninguna tendencia.



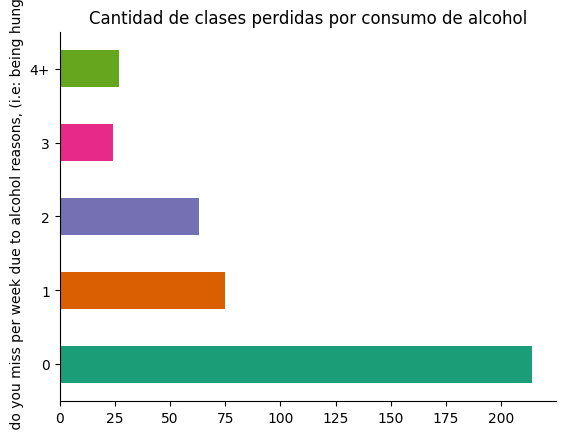
* En el análisis exploratorio de datos (EDA), realicé un resumen estadístico de la base de datos empleando el método describe(). Exploré la estructura del conjunto de datos utilizando los métodos shape, info() y columns. Verifiqué la existencia de datos faltantes con isna(). Además, examiné las variables categóricas y numéricas para obtener una comprensión más profunda de la distribución y características de los datos. Además, dado que varias de las variables a analizar estaban en formato **object**, fueron convertidas a **category** o **float64**, según correspondiera, para asegurar un análisis adecuado.



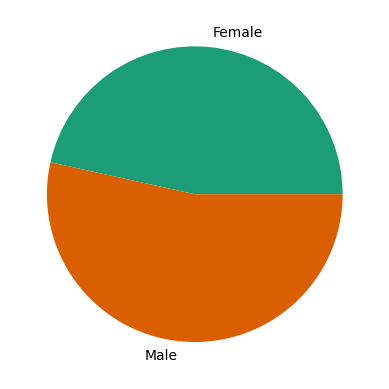
* Generé un mapa de calor para visualizar los valores faltantes en el conjunto de datos. Este mapa de calor destaca la presencia y distribución de datos ausentes, permitiéndome identificar las columnas y filas con mayores problemas de datos incompletos. Utilizando una escala de colores, el gráfico facilita la detección de patrones de datos faltantes.

****

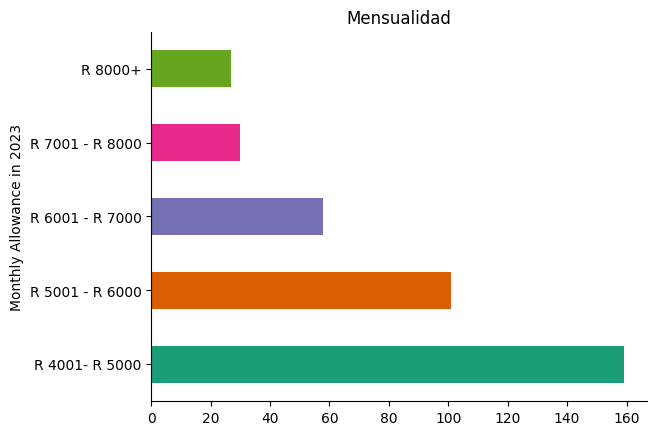
* Utilicé un gráfico de barras, creado con Matplotlib o Seaborn, para visualizar la cantidad de estudiantes por año de cursada. Este gráfico permite observar la distribución de estudiantes a lo largo de los distintos años académicos. La hipótesis subyacente es que la cantidad de estudiantes puede variar significativamente entre los años de cursada, lo que podría influir en los patrones de consumo de alcohol entre los diferentes grupos académicos.



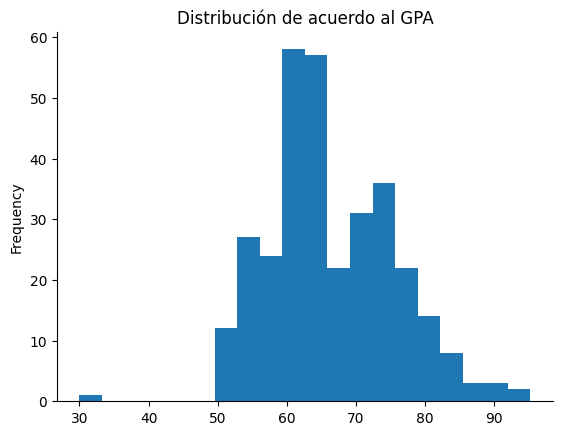
* Utilicé un gráfico de barras, creado con Matplotlib o Seaborn, para visualizar la cantidad de clases perdidas debido al consumo de alcohol. Este gráfico permite observar cómo el consumo de alcohol afecta la asistencia a clases entre los estudiantes universitarios. La hipótesis subyacente es que un mayor consumo de alcohol está asociado con un aumento en el número de clases perdidas, lo que podría indicar un impacto negativo del consumo de alcohol en el rendimiento académico.



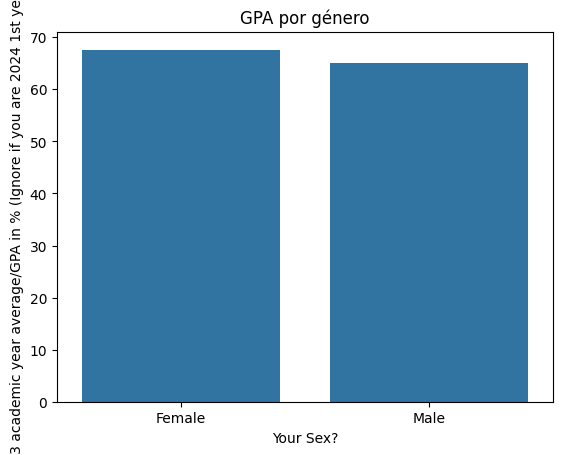
* Generé un gráfico de torta utilizando Matplotlib o Seaborn para visualizar la proporción de hombres y mujeres en la base de datos. Este gráfico permite una comparación clara y visualmente intuitiva de la distribución de género entre los estudiantes. La hipótesis subyacente es que la proporción de hombres y mujeres podría influir en las tendencias y patrones de consumo de alcohol, así como en otros comportamientos relacionados con la vida universitaria.



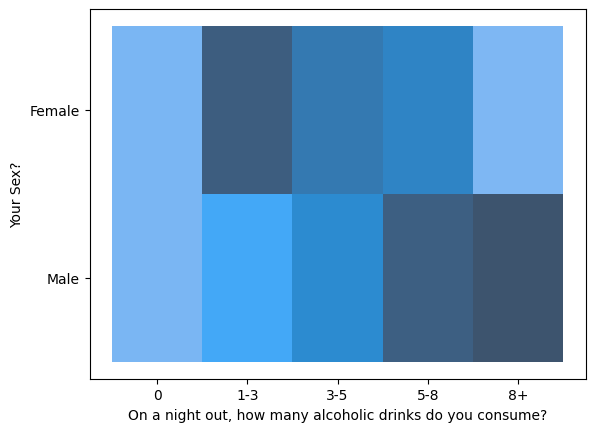
* Utilicé un gráfico de barras, creado con Matplotlib o Seaborn, para visualizar la mensualidad pagada por los estudiantes. Este gráfico permite observar la distribución de los costos de mensualidad entre los estudiantes universitarios. La hipótesis subyacente es que la mensualidad puede influir en el comportamiento y consumo de alcohol de los estudiantes, ya que diferentes niveles de gasto en educación podrían estar asociados con distintas presiones financieras y estilos de vida.



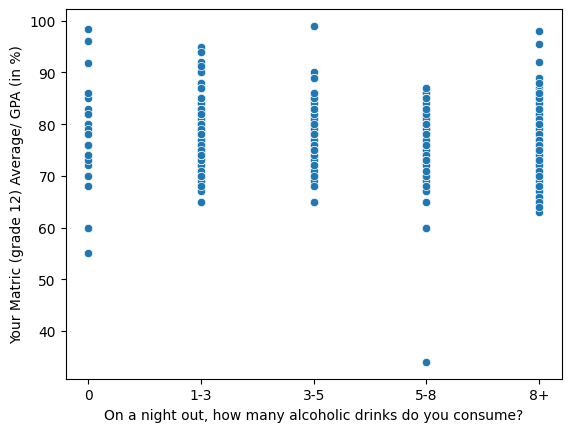
* Utilicé un histograma, creado con Matplotlib o Seaborn, para visualizar la distribución de los promedios académicos entre los estudiantes en la muestra. Este histograma permite observar cómo se distribuyen los promedios, identificando patrones y posibles sesgos en el rendimiento académico. La hipótesis subyacente es que la distribución de los promedios puede estar influenciada por factores como el consumo de alcohol, la carga académica y las características sociodemográficas, proporcionando una visión general de la variabilidad en el desempeño estudiantil.



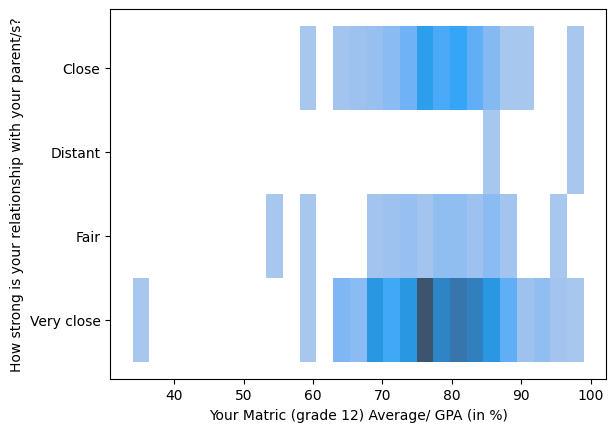
* Utilicé un gráfico de barras, creado con Matplotlib o Seaborn, para visualizar el promedio académico de los estudiantes desglosado por género. Este gráfico permite comparar los promedios entre hombres y mujeres, identificando posibles diferencias en el rendimiento académico entre los géneros. La hipótesis subyacente es que podrían existir diferencias significativas en los promedios académicos de hombres y mujeres, influenciadas por factores sociales, económicos o comportamentales.



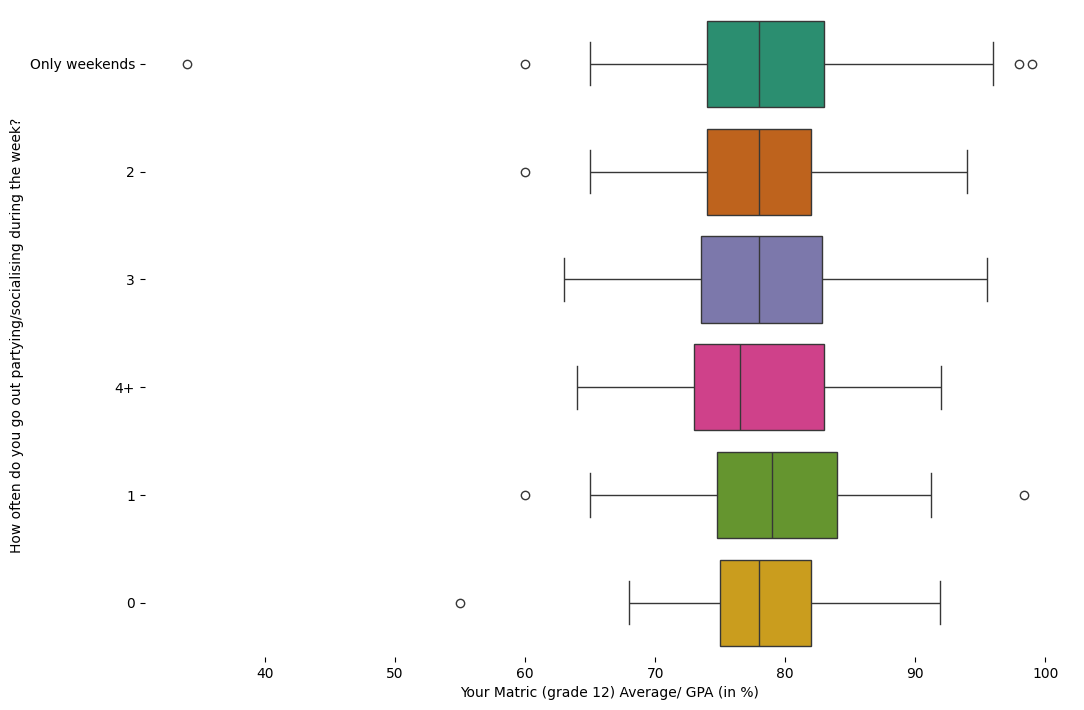
* Generé un mapa de calor (heatmap) utilizando Matplotlib o Seaborn para visualizar la cantidad de bebidas alcohólicas consumidas por noche según el género. Este heatmap permite identificar patrones y diferencias en el consumo de alcohol entre hombres y mujeres, mostrando la intensidad y frecuencia del consumo. La hipótesis subyacente es que el consumo de bebidas alcohólicas por noche varía significativamente entre géneros, lo que podría reflejar diferencias en comportamiento social, tolerancia al alcohol, y factores culturales.



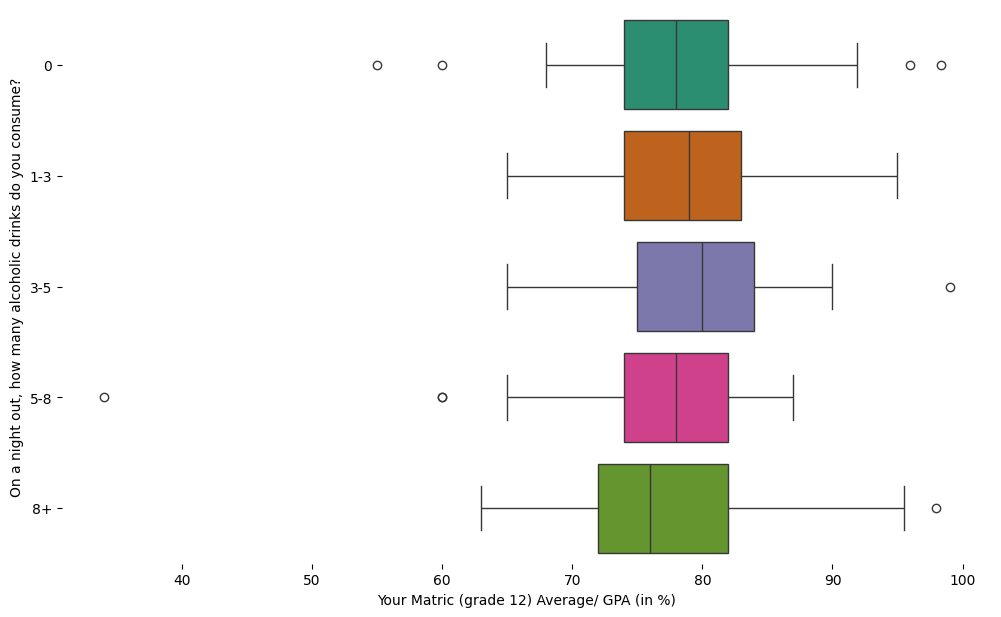
* Creé un gráfico de dispersión (scatter plot) utilizando Matplotlib o Seaborn para visualizar la relación entre el promedio académico y la cantidad de bebidas alcohólicas consumidas por noche. Este gráfico permite explorar si existe alguna relación entre el rendimiento académico y el consumo de alcohol entre los estudiantes. La hipótesis subyacente es que podría haber una asociación negativa entre el promedio académico y la cantidad de bebidas alcohólicas consumidas, sugiriendo que un mayor consumo de alcohol puede estar relacionado con un rendimiento académico más bajo.



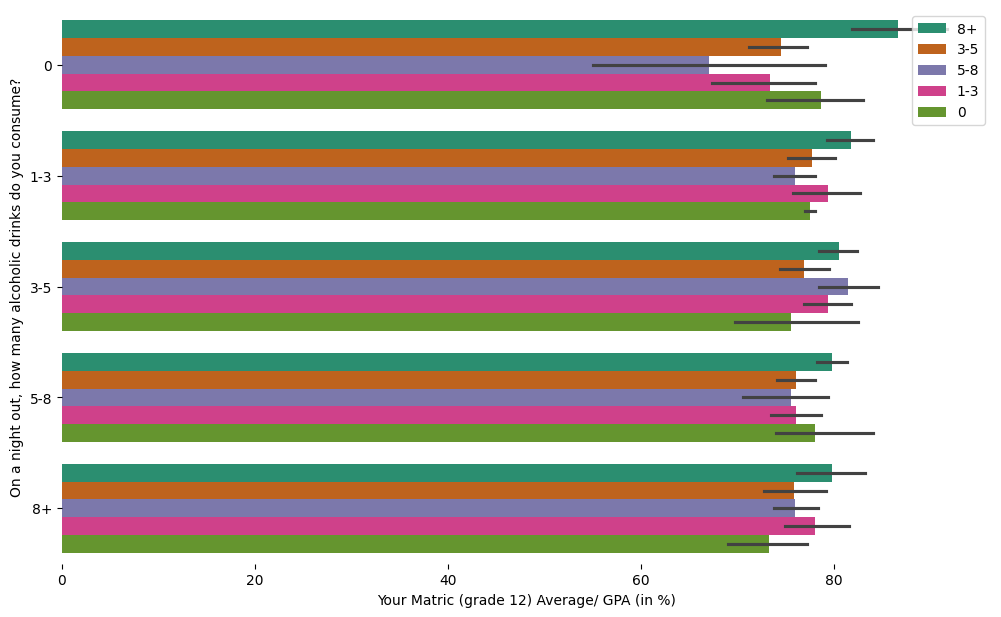
* Generé un mapa de calor (heatmap) utilizando Matplotlib o Seaborn para visualizar la relación entre el promedio académico y la calidad de la relación con los padres. Este heatmap permite explorar si existe alguna asociación entre el rendimiento académico y la relación con los padres entre los estudiantes. La hipótesis subyacente es que podría haber una correlación positiva entre un buen promedio académico y una relación positiva con los padres, sugiriendo que un entorno familiar favorable podría influir en el éxito académico de los estudiantes.



* Generé un diagrama de caja (boxplot) utilizando Matplotlib o Seaborn para visualizar la distribución del promedio académico en relación con la frecuencia de salir de fiesta semanalmente. Este diagrama permite explorar si existe alguna asociación entre el rendimiento académico y la frecuencia de salir de fiesta entre los estudiantes. La hipótesis subyacente es que podría haber una correlación negativa entre un buen promedio académico y una alta frecuencia de salir de fiesta, lo que sugiere que una mayor participación en actividades de ocio nocturno podría estar relacionada con un rendimiento académico más bajo.



* Utilicé un diagrama de caja (boxplot) utilizando Matplotlib o Seaborn para visualizar la distribución del número de bebidas consumidas por noche en relación con el promedio académico. Este diagrama permite explorar si existe alguna asociación entre el consumo de alcohol y el rendimiento académico entre los estudiantes. La hipótesis subyacente es que podría haber una correlación negativa entre la cantidad de bebidas consumidas por noche y el promedio académico, lo que sugiere que un mayor consumo de alcohol podría estar relacionado con un rendimiento académico más bajo.



* Generé un gráfico de barras para visualizar la cantidad de bebidas alcohólicas consumidas por salida, la cantidad de horas dedicadas al estudio y el promedio académico. Este gráfico permite explorar la relación entre el consumo de alcohol, el tiempo dedicado al estudio y el rendimiento académico entre los estudiantes. La hipótesis subyacente es que podría haber una asociación entre el consumo de alcohol y el rendimiento académico, que podría ser mediada por la cantidad de horas dedicadas al estudio; es decir, un mayor consumo de alcohol podría estar relacionado con un rendimiento académico más bajo, especialmente si se dedica menos tiempo al estudio.

Conclusiones preliminares:

1. **Patrones de Comportamiento Estudiantil**: Aunque no se observan diferencias distintivas entre hombres y mujeres en los patrones de comportamiento, se identifica una relación entre la cantidad de horas dedicadas al estudio, el rendimiento académico y el comportamiento de salir de fiesta o el consumo de alcohol. Los estudiantes que dedican más tiempo al estudio tienden a tener un mejor rendimiento académico, mientras que aquellos que salen de fiesta con mayor frecuencia o consumen más alcohol por noche tienden a tener un rendimiento académico más bajo.
2. **Diferencias de Género en el Consumo de Alcohol**: Se observa que los hombres tienden a consumir más bebidas alcohólicas por noche de salida en comparación con las mujeres. Esta disparidad podría reflejar diferencias culturales o sociales en las normas de consumo de alcohol entre hombres y mujeres, así como posibles diferencias en la tolerancia al alcohol.
3. **Impacto del Estilo de Vida en el Rendimiento Académico**: Existe una clara asociación entre el estilo de vida de los estudiantes y su rendimiento académico. Aquellos que equilibran adecuadamente el tiempo dedicado al estudio con actividades sociales moderadas tienden a tener un mejor rendimiento académico en comparación con aquellos que dedican una cantidad excesiva de tiempo a las actividades sociales, como salir de fiesta o consumir alcohol.
4. **Importancia del Equilibrio entre Estudio y Ocio**: Estos datos resaltan la importancia de mantener un equilibrio saludable entre el estudio y el ocio para alcanzar un rendimiento académico óptimo. Si bien es importante dedicar tiempo al estudio y la preparación académica, también es esencial permitirse actividades sociales como forma de desestresarse y mantener una salud mental equilibrada.
5. **Necesidad de Concientización sobre el Consumo de Alcohol**: Los datos sugieren la necesidad de concientizar a los estudiantes sobre los riesgos asociados con el consumo excesivo de alcohol y cómo puede afectar negativamente su rendimiento académico. Proporcionar recursos y apoyo para promover un consumo responsable de alcohol podría ser beneficioso para mejorar el bienestar y el éxito académico de los estudiantes.

Hipótesis posibles a explorar en el último tramo del proyecto mediante la aplicación de clúster K medias:

1. **Patrones de Comportamiento Estudiantil**: Considerando la cantidad de bebidas alcohólicas consumidas por noche, la cantidad de horas dedicadas al estudio y el promedio académico, se podría hipotetizar que existen diferentes patrones de comportamiento estudiantil. Por ejemplo, podría haber un grupo de estudiantes caracterizado por un alto consumo de alcohol, menor tiempo dedicado al estudio y un bajo promedio académico, mientras que otro grupo podría ser lo opuesto. El clustering de k medias podría ayudar a identificar estos patrones y segmentar a los estudiantes en grupos similares.
2. **Impacto de la Relación con los Padres en el Comportamiento Estudiantil**: Dado que la relación con los padres es un factor importante en el desarrollo académico y social de los estudiantes, podría ser interesante explorar cómo esta variable se relaciona con otros comportamientos estudiantiles, como el consumo de alcohol, el tiempo dedicado al estudio y el rendimiento académico. Se podría hipotetizar que los estudiantes con una relación positiva con sus padres tienden a tener un menor consumo de alcohol, dedican más tiempo al estudio y tienen un mejor rendimiento académico en comparación con aquellos con una relación menos favorable. El clustering de k medias podría ayudar a identificar grupos de estudiantes con características similares en relación con esta variable y cómo se relacionan con otros aspectos de su vida estudiantil.