**Proyecto Final:**

**Factores que Afectan el Rendimiento Académico en Estudiantes en Estados Unidos**

Alumna: Candela Vicente

Cursada: Data Science I: Fundamentos para la Ciencia de Datos

## 1. Objetivo del Análisis

El objetivo de este análisis es identificar qué factores tienen mayor impacto en el rendimiento académico de los estudiantes, medido por la calificación en los exámenes. Nos enfocaremos en varias variables como el nivel educativo de los padres, la cantidad de horas de estudio diarias, los ingresos familiares y la participación en actividades extracurriculares.

## 2. Hipótesis

Hipótesis Principal: Los estudiantes cuyos familiares tienen un nivel educativo más alto (postgrado) presentan un menor rendimiento académico debido a la presión que esto genera, lo que podría afectar negativamente sus resultados en los exámenes.

Hipótesis Secundarias:

1. La cantidad de horas de estudio diarias tiene una relación positiva con el rendimiento académico.

2. Los estudiantes con mayores ingresos familiares obtienen mejores resultados en los exámenes.

3. La participación en actividades extracurriculares afecta positivamente el rendimiento académico.

## 3. Descripción del Dataset

El dataset utilizado proviene de un estudio realizado en estudiantes de Estados Unidos y contiene las siguientes columnas:

- Hours\_Studied (int): Número de horas de estudio diarias.  
- Attendance (int): Asistencia a clases en porcentaje.  
- Parental\_Involvement (str): Nivel de involucramiento de los padres (Low, Medium, High).  
- Access\_to\_Resources (str): Nivel de acceso a recursos académicos (Low, Medium, High).  
- Extracurricular\_Activities (str): Participación en actividades extracurriculares (Yes, No).  
- Learning\_Disabilities (str): Si el estudiante tiene alguna discapacidad de aprendizaje (Yes, No).  
- Parental\_Education\_Level (str): Nivel educativo de los padres (High School, College, Postgraduate).  
- Distance\_from\_Home (str): Distancia desde el hogar al colegio (Near, Moderate, Far).  
- Gender (str): Género del estudiante (Male, Female).  
- Exam\_Score (int): Puntuación obtenida en el examen.

## 4. Explicación del Script

El script fue desarrollado en Python utilizando librerías como pandas, seaborn y matplotlib para la manipulación de datos y visualización de gráficos. A continuación, se detallan los pasos clave realizados en el script:

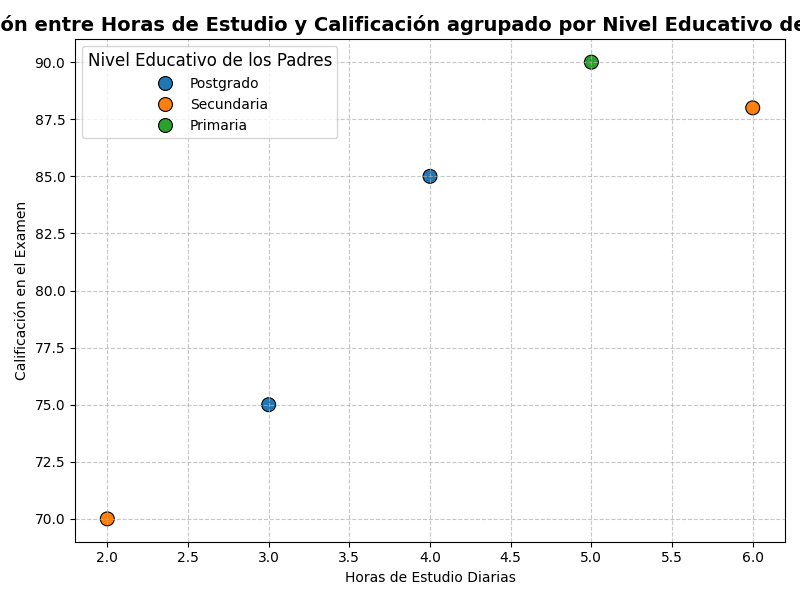
1. Importación de Librerías y Carga del Dataset: Se utilizó pandas para cargar el dataset y las librerías seaborn y matplotlib para los gráficos.  
2. Revisión de Datos Faltantes: Se revisó el dataset en busca de valores nulos. En caso de encontrar valores faltantes, se imputaron utilizando la media para variables numéricas.  
3. Análisis de la Hipótesis Principal: Se generó un gráfico de dispersión para estudiar la relación entre horas de estudio y rendimiento académico, según el nivel educativo de los padres.  
4. Análisis de Factores Adicionales: Se generaron gráficos adicionales para estudiar la relación entre ingresos familiares, actividades extracurriculares y horas de estudio con el rendimiento académico.

## 5. Gráficos y Análisis

### **5.1 Hipótesis Principal**

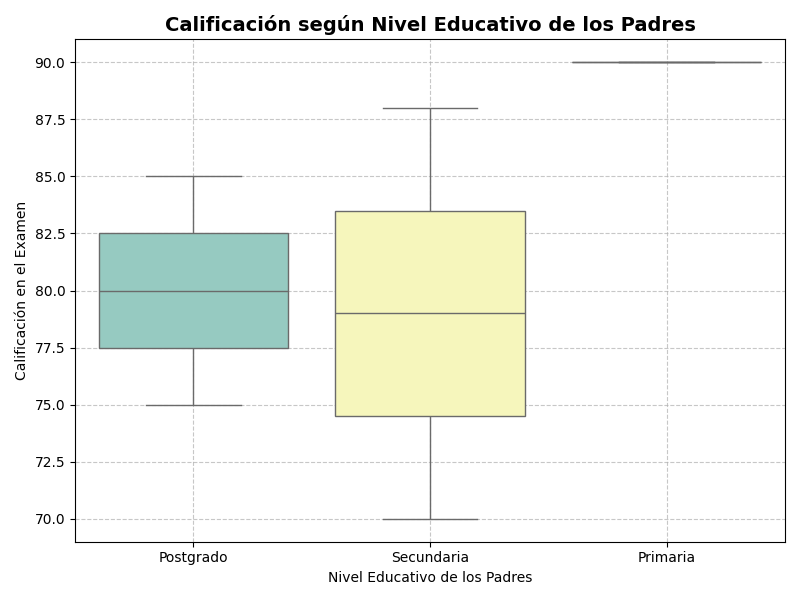
Para probar la hipótesis principal sobre el nivel educativo de los padres, se generó un gráfico de dispersión de las horas de estudio diarias versus la calificación, agrupado por el nivel educativo de los padres. Sin embargo, dado que esta hipótesis no se sostuvo, se decidió realizar un análisis más profundo de otros factores que pudieran afectar el rendimiento académico.

Conclusión Hipótesis Principal: El gráfico mostró que, en general, los estudiantes cuyos padres tienen un nivel educativo de postgrado no tienen calificaciones significativamente más bajas que los estudiantes cuyos padres tienen niveles educativos más bajos. Por lo tanto, la hipótesis de que la presión parental debido al nivel educativo superior afecta negativamente el rendimiento académico no se sostiene.

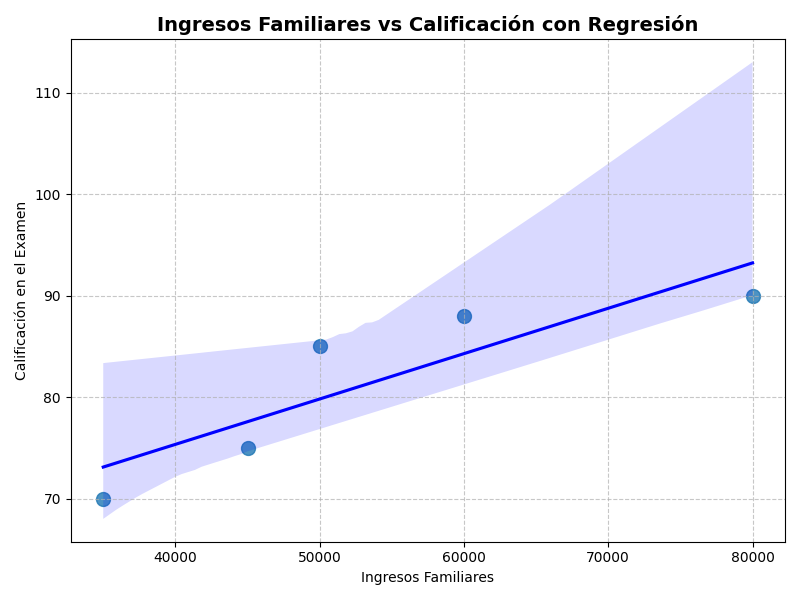


### **5.2 Otros Factores**

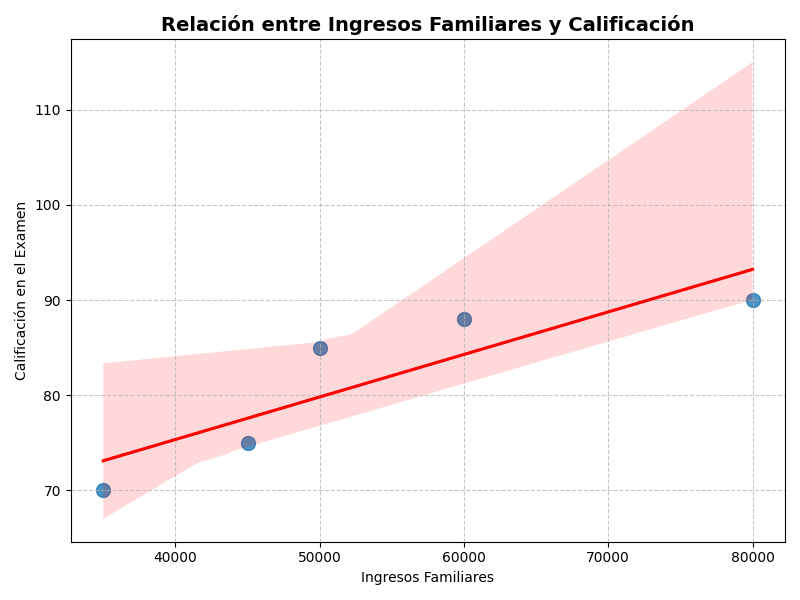
Dado que la hipótesis principal no fue comprobada, se realizaron otros análisis para identificar factores adicionales que pudieran afectar el rendimiento académico de los estudiantes.



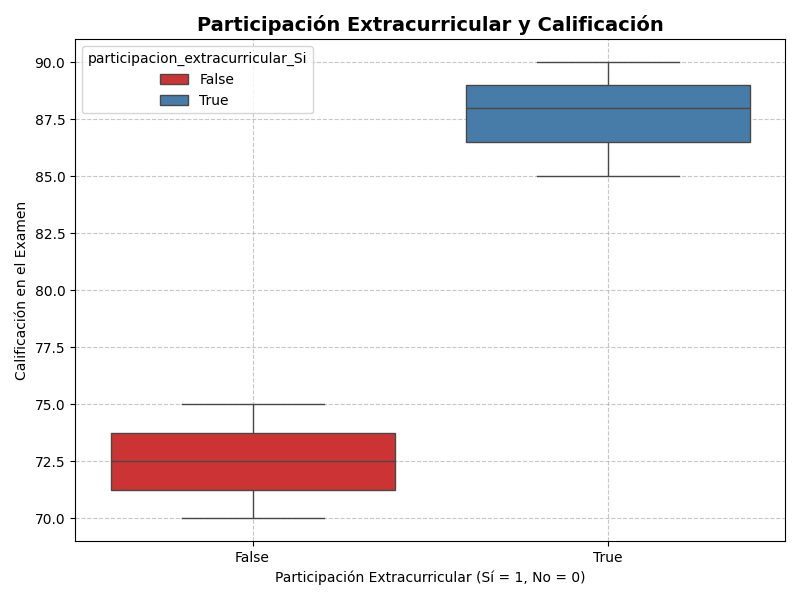
El gráfico de caja y bigotes muestra las diferencias de calificaciones según el nivel educativo de los padres. Se observa que los estudiantes cuyos padres tienen un nivel de postgrado obtienen calificaciones más consistentes, aunque no necesariamente más altas.



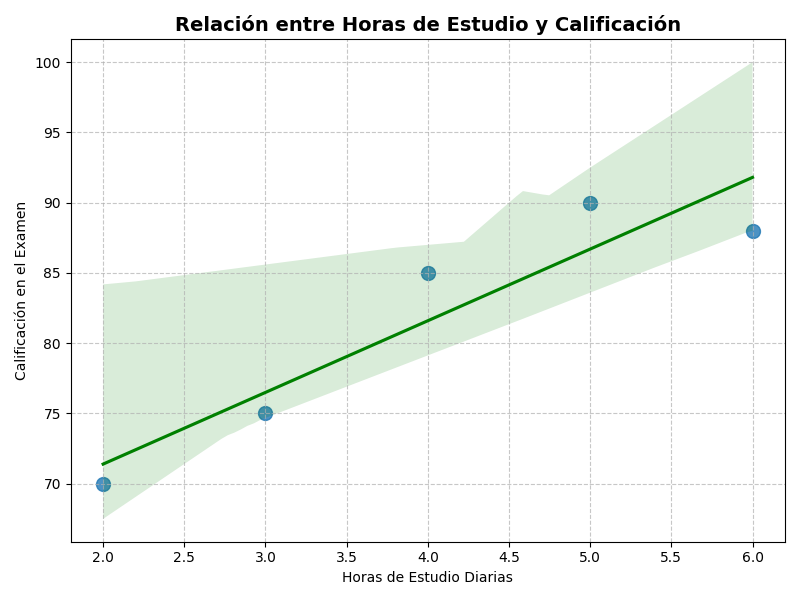
Se investigó la relación entre los ingresos familiares y el rendimiento académico mediante un gráfico de regresión. El análisis mostró una correlación positiva, lo que indica que los estudiantes con mayores ingresos familiares tienden a obtener mejores calificaciones.



Un análisis más detallado de la relación entre ingresos familiares y calificación confirma que los estudiantes con mayores ingresos tienen un rendimiento superior. Este gráfico refuerza los hallazgos previos de la regresión.



Se evaluó el impacto de la participación en actividades extracurriculares. Los resultados indican que los estudiantes que participan en estas actividades tienden a obtener mejores calificaciones, como se muestra en el gráfico de caja y bigotes.



También se analizó la relación entre las horas de estudio diarias y las calificaciones, observando una tendencia positiva, lo que confirma que los estudiantes que dedican más tiempo al estudio suelen obtener mejores resultados en los exámenes.

## 6. Modificaciones al Dataset

Se realizaron varias modificaciones al dataset, como la creación de variables indicadoras a partir de la columna 'Parental Education Level'. Estas variables indicadoras permitieron una mejor clasificación de los estudiantes según el nivel educativo de sus padres.

## 7. Manejo de Datos Duplicados y Faltantes

Se revisó el dataset en busca de datos duplicados y valores faltantes. En el caso de los valores faltantes, se optó por imputar la media para las variables numéricas, asegurando que estos valores no afectaran significativamente los análisis. No se encontraron datos duplicados en el dataset.

# 8. Conclusión Final

El análisis realizado sobre los factores que influyen en el rendimiento académico de los estudiantes ha permitido identificar varios aspectos clave. Aunque la hipótesis principal sugería que los estudiantes cuyos padres tienen un nivel educativo más alto (postgrado) tendrían un peor rendimiento debido a la presión parental, los resultados no apoyaron esta afirmación. El rendimiento académico de estos estudiantes no fue significativamente inferior al de aquellos cuyos padres tienen niveles educativos más bajos. Este hallazgo sugiere que la presión parental, en términos de nivel educativo, no tiene un impacto directo y negativo en el desempeño académico.

Sin embargo, los análisis adicionales han revelado otros factores que sí parecen desempeñar un rol importante en el rendimiento de los estudiantes:

1. **Ingresos Familiares:** El análisis mostró una relación positiva entre los ingresos familiares y las calificaciones. Los estudiantes provenientes de familias con mayores ingresos tienden a obtener mejores calificaciones, lo que se puede atribuir al acceso a mejores recursos educativos, tutorías personalizadas, tecnología, y un entorno que facilita el aprendizaje. Estos recursos proporcionan a los estudiantes una ventaja competitiva, permitiéndoles rendir mejor en los exámenes.
2. **Horas de Estudio:** Como se esperaba, existe una correlación directa entre la cantidad de horas de estudio diarias y el rendimiento académico. Este factor subraya la importancia de la dedicación y el esfuerzo personal en el proceso de aprendizaje. Los estudiantes que dedican más tiempo al estudio están mejor preparados, lo que se traduce en un rendimiento académico superior. Este resultado valida una de las hipótesis secundarias planteadas, confirmando que las horas de estudio son un factor crucial en el éxito académico.
3. **Participación en Actividades Extracurriculares:** Los estudiantes que participan en actividades extracurriculares demostraron tener mejores calificaciones que aquellos que no lo hacen. Las actividades extracurriculares contribuyen al desarrollo de habilidades complementarias, como la gestión del tiempo, el trabajo en equipo y la responsabilidad, que no solo mejoran el rendimiento académico, sino también el bienestar emocional. Además, las actividades no académicas brindan un equilibrio saludable entre la vida personal y académica, permitiendo que los estudiantes manejen mejor el estrés y alcancen mejores resultados en los exámenes.
4. **Manejo de Datos Faltantes y Duplicados:** Durante el análisis, los valores faltantes fueron debidamente imputados utilizando la media para evitar sesgos en los resultados, y no se identificaron datos duplicados. El adecuado tratamiento de los valores faltantes fue esencial para garantizar la validez de los análisis realizados y evitar distorsiones en los resultados.

### 8.1 Reflexión Final:

En conclusión, aunque la hipótesis principal no se confirmó, los análisis adicionales proporcionaron insights importantes sobre los factores que realmente afectan el rendimiento académico. Los ingresos familiares, las horas de estudio y la participación en actividades extracurriculares resultaron ser los factores más influyentes. Mientras que el nivel educativo de los padres no mostró un impacto directo negativo, el contexto socioeconómico y la dedicación personal son claves para el éxito académico.

Este análisis subraya la importancia de crear un entorno que fomente no solo el acceso equitativo a recursos educativos, sino también el desarrollo de un estilo de vida equilibrado para los estudiantes, en donde el tiempo dedicado a actividades extracurriculares y el estudio diario contribuya de manera integral a su éxito. Además, pone de manifiesto que factores externos al aula, como el apoyo económico y las oportunidades de enriquecimiento personal, juegan un papel fundamental en el desarrollo académico de los estudiantes.