



TRABAJO DE GRADO
Opción Seminario-Diplomado.

SISTEMA DE ANALIZAR LOS RENDIMIENTOS ENTRE 100 ESTUDIANTES DE INFORMATICA, ENTRE LOS NIVELES DE DEPRESIÓN, RENDIMIENTO ACADÉMICO Y PATRONES DE TDAH (TRASTORNO POR DÉFICIT DE ATENCIÓN E HIPERACTIVIDAD), UTILIZANDO ESTRATEGIAS DE MACHINE LEARNING

Corporación Universitaria Remington.
Nombre de la facultad: Ingenierías
Nombre del programa académico: Ingeniería de sistemas

Estudiantes:
Andres Felipe Diaz

Tutor: Juan Carlos Briñez de León

Opción de Trabajo de grado Seminario-Diplomado.
2024.

DEDICATORIA

En primer lugar les agradezco a mis padres que siempre me han brindado su apoyo incondicional para poder cumplir todos mis objetivos personales y académicos. Ellos son los que con su cariño me han impulsado siempre a perseguir mis metas y nunca abandonarlas frente a las adversidades. También son los que me han brindado el soporte material y económico para poder concentrarme en los estudios y nunca abandonarlos.

AGRADECIMIENTOS

POR ÚLTIMO AGRADECER A LA UNIVERSIDAD QUE ME HA EXIGIDO TANTO, PERO AL MISMO TIEMPO ME HA PERMITIDO OBTENER MI TAN ANSIADO TÍTULO. AGRADEZCO A CADA DIRECTIVO POR SU TRABAJO Y POR SU GESTIÓN, SIN LO CUAL NO ESTARÍAN LAS BASES NI LAS CONDICIONES PARA APRENDER CONOCIMIENTOS.

A DIOS, A MIS COMPAÑEROS, AL PROFE DEL SEMINARIO, Y A TODAS AQUELLAS PERSONAS QUE SUMAN EN MI PROCESO DE FORMACIÓN.

TABLA DE CONTENIDOS

Resumen.....	5
Palabras clave	5
.....	5
1. Marco conceptual y contextual.....	6
1.1 descripción del estudio :.....	6
Objetivos.....	8
Desarrollo e implementación del aprendizaje	9
Conclusiones	24
Referencias bibliográficas.....	25

Resumen

En el presente trabajo de investigación se elaboró un modelo Machine Learning para predecir el rendimiento académico de los estudiantes universitarios, se utilizó el programa de dataset El objetivo de esta investigación es determinar en qué porcentaje Machine Learning permite predecir el rendimiento académico con precisión, sensibilidad y especificidad, con el fin de poder identificar a los alumnos con probabilidad de éxito o fracaso. En esta investigación se utilizó una población de 100 estudiantes de informática, así mismo se usó la totalidad de la población como muestra. Por otro lado, el estudio es de tipo aplicada, con un diseño de investigación experimental de tipo pre-experimental de un solo grupo, ya que luego de aplicar Machine Learning se podrá observar los resultados y realizar la medición. Como resultado en relación a la precisión, sensibilidad y especificidad para los algoritmos de árbol de decisión.

PALABRAS CLAVE

Dominancia, clustering, impacto de la depresión, patrones.

MARCO CONCEPTUAL

El estudio sobre la depresión y el rendimiento académico en estudiantes de informática se enmarca en el análisis de factores psicosociales que impactan en la educación superior. En particular, los problemas de salud mental, como la depresión, son considerados factores clave que afectan el desempeño de los estudiantes universitarios. A medida que los programas académicos en informática exigen altos niveles de concentración, gestión del tiempo y capacidades cognitivas, los estudiantes pueden experimentar estrés académico que puede llevar a síntomas depresivos. Por tanto, es relevante explorar las relaciones entre los niveles de depresión y el rendimiento académico para proponer estrategias de intervención y apoyo.

1. Depresión: La depresión es un trastorno mental caracterizado por síntomas como tristeza persistente, falta de motivación, fatiga y dificultad para concentrarse, todos los cuales pueden afectar negativamente el rendimiento académico. En el contexto universitario, los estudiantes suelen enfrentar diversas presiones académicas, financieras y personales que pueden exacerbar estos síntomas.

2. Rendimiento Académico: Este concepto se refiere a las calificaciones obtenidas por los estudiantes y al logro de competencias establecidas en el programa educativo. En el caso de los estudiantes de informática, el rendimiento académico se mide a través de exámenes, proyectos y prácticas en temas como programación, matemáticas y resolución de problemas complejos. La depresión puede afectar el rendimiento, ya que la falta de concentración y motivación dificulta el aprendizaje.

3. Correlación entre Depresión y Rendimiento Académico: Diversos estudios han demostrado que la depresión puede influir negativamente en el rendimiento académico, especialmente en disciplinas que requieren una alta carga cognitiva. Identificar esta relación en estudiantes de informática permitirá conocer el impacto específico en esta área y formular estrategias de apoyo en salud mental para estos estudiantes.

MARCO CONTEXTUAL

La investigación se llevó a cabo en una universidad donde se encuestó a 100 estudiantes de la carrera de informática. Estos estudiantes, en su mayoría, están expuestos a largas horas de estudio, prácticas de laboratorio y, a menudo, una fuerte competencia en su campo. Además, el contexto de estudio remoto o híbrido puede haber influido en su salud mental, afectando su capacidad de concentración y su interacción social.

El propósito de la encuesta fue identificar si existen correlaciones entre los niveles de depresión y el rendimiento académico. La encuesta incluyó preguntas sobre el estado emocional de los estudiantes, factores de estrés académico, calificaciones y otros indicadores de rendimiento académico.

DESCRIPCIÓN DE CASO DE ESTUDIO

El conjunto de datos consiste en respuestas de una encuesta realizada a 100 estudiantes de informática. El objetivo de la encuesta es identificar correlaciones entre los niveles de depresión, rendimiento académico y patrones de TDAH (Trastorno por Déficit de Atención e Hiperactividad) a través del análisis de datos.

DESARROLLO E IMPLEMENTACIÓN DEL APRENDIZAJE EN CLASES

VAMOS A HACER LAS IMPORTACIONES NECESARIAS PARA PODER REALIZAR ESTE PROYECTO

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from scipy.stats import spearmanr, kendalltau
```

IMPORTAMOS

NUESTROS

DATOS

```
df = pd.read_csv('/kaggle/input/psychosocial-dimensions-of-student-life/CSE_student_performances.csv')
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99 entries, 0 to 98
Data columns (total 10 columns):
#   Column                                                                 Non-Null Count  Dtype
---  -
0   Age                                                                    99 non-null    int64
1   Gender                                                                  99 non-null    object
2   AcademicPerformance                                                    99 non-null    object
3   TakingNoteInClass                                                      99 non-null    object
4   DepressionStatus                                                       99 non-null    object
5   FaceChallengesToCompleteAcademicTask 99 non-null    object
6   LikePresentation                                                       99 non-null    object
7   SleepPerDayHours                                                       99 non-null    int64
8   NumberOfFriend                                                         95 non-null    float64
9   LikeNewThings                                                          99 non-null    object
dtypes: float64(1), int64(2), object(7)
memory usage: 7.9+ KB
```

Empezaremos por la correlación entre la edad y el desempeño académico

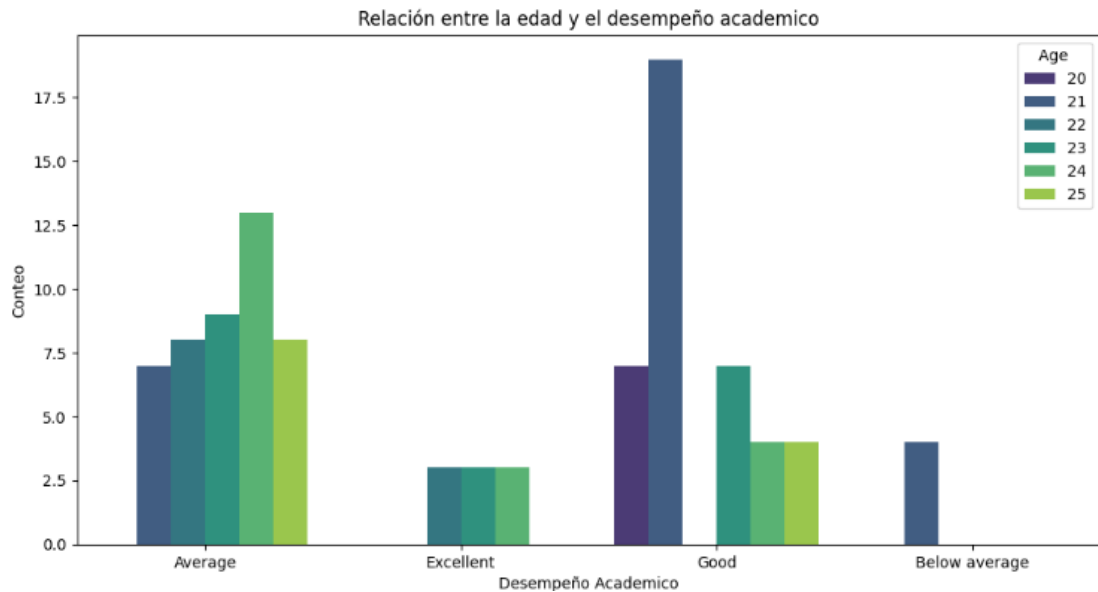
```
correlation_age_performance, _ = spearmanr(df['Age'], df['AcademicPerformance'])  
print(f"Correlación entre la edad y el desempeño académico: {correlation_age_performance}")
```

Correlación entre la edad y el desempeño académico: -0.3924067630541698

CREAMOS UN GRÁFICO DE BARRAS AGRUPADAS SIN ETIQUETAS DE RECuento

```
# Creamos un gráfico de barras agrupadas sin etiquetas de recuento
plt.figure(figsize=(12, 6))
sns.countplot(data=df, x='AcademicPerformance', hue='Age ', palette='viridis')

plt.title('Relación entre la edad y el desempeño academico')
plt.xlabel('Desempeño Academico')
plt.ylabel('Conteo')
plt.show()
```

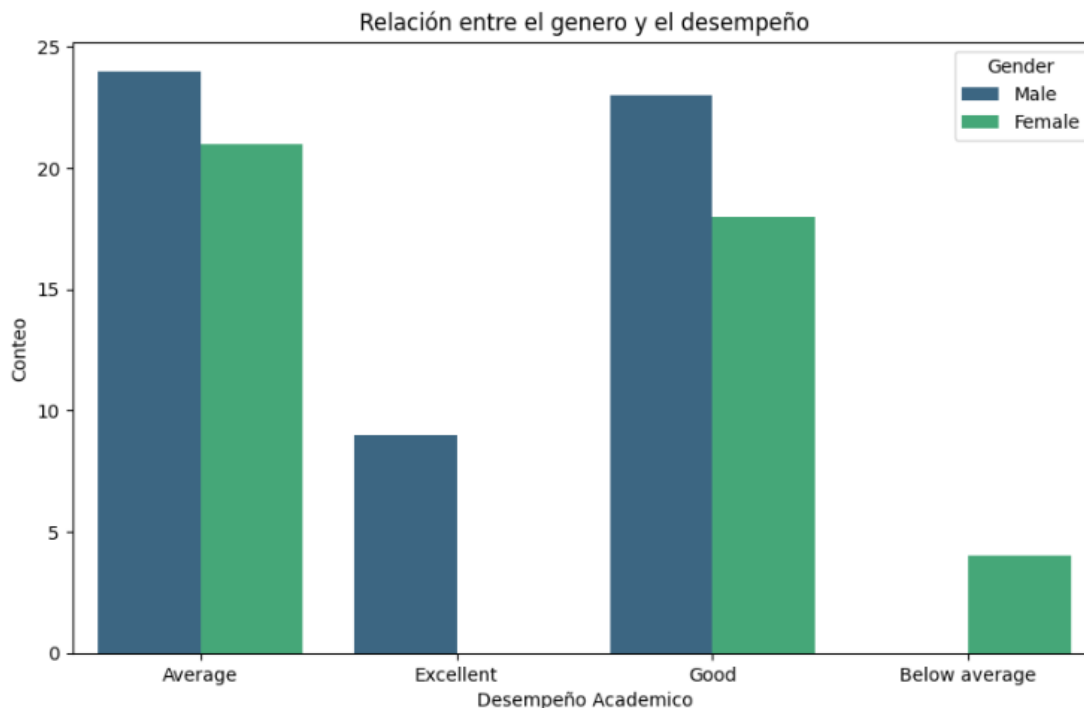


Como podemos notar, la gente que tiene 22 años pertenece tanto al grupo como de excelente y promedio, sin embargo es más probable que sean promedio, por otro lado notamos una alta dominancia en la gente de 21 años para tener un buen desempeño, pero al mismo tiempo una porción de ellos se encuentra en debajo del promedio, por otro lado tambien vemos que la gente de 24 y 25 años tienden a formar parte del promedio.

PROCEDEMOS A LOS PATRONES RELACIONADOS AL GENERO

```
# Crear un gráfico de barras agrupadas para visualizar la relación entre Gender y AcademicPerformance
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='AcademicPerformance', hue='Gender', palette='viridis')

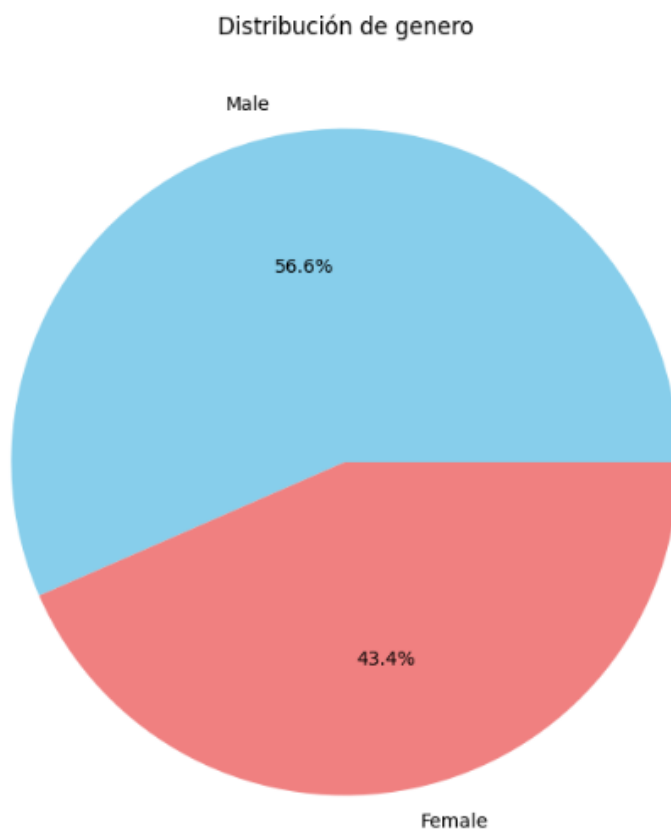
plt.title('Relación entre el genero y el desempeño')
plt.xlabel('Desempeño Academico')
plt.ylabel('Conteo')
plt.show()
```



Al desplegar nuestro gráfico podemos notar una alta dominancia por parte de los hombres con respecto al desempeño académico, de acuerdo a este conjunto de datos observamos que los hombres suelen ser mejores académicamente que las mujeres, sin embargo tenemos que tomar en cuenta un factor y es el porcentaje de alumnos del genero masculino y femenino en este conjunto de datos, como podremos observar en el siguiente gráfico.

```
# Contar la frecuencia de cada valor en la columna 'Gender'
gender_counts = df['Gender'].value_counts()

# Crear un gráfico de pastel
plt.figure(figsize=(8, 8))
plt.pie(gender_counts, labels=gender_counts.index, autopct='%1.1f%%', colors=['skyblue', 'lightcoral'])
plt.title('Distribución de genero')
plt.show()
```

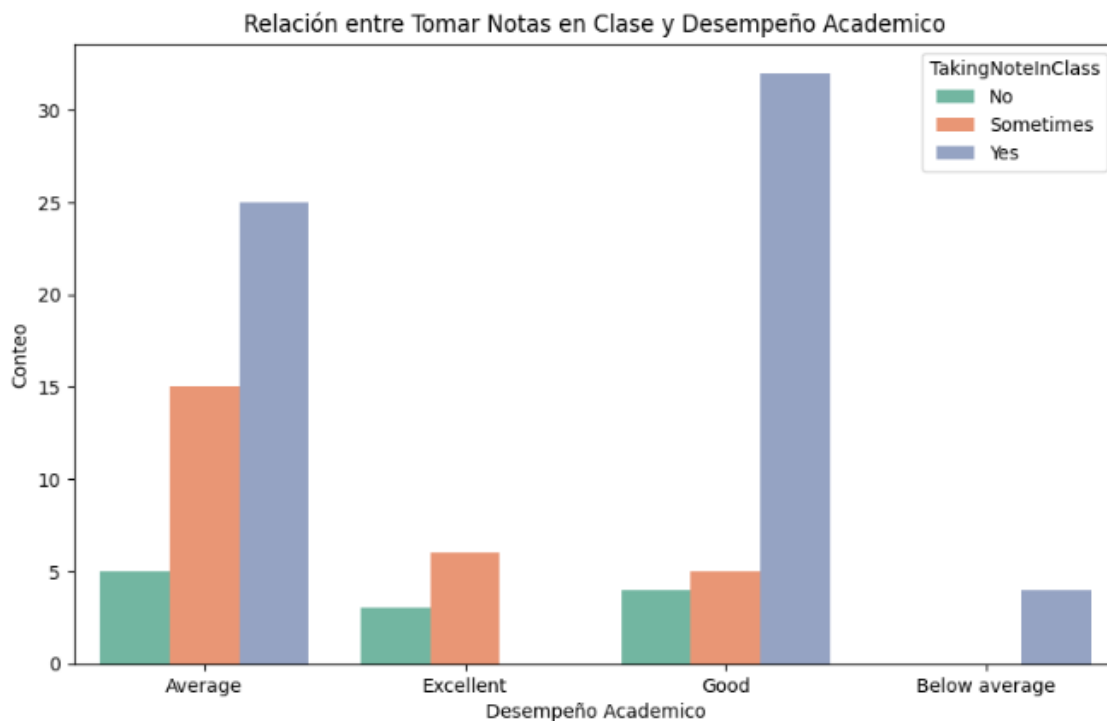


Verificamos que efectivamente, gran parte de los alumnos que forman parte del conjunto de datos son del genero masculino, entonces podría ser un factor tambien a considerar en cuestión de que genero es el que más se inscribe a las carreras de ciencias de la computación

PROCEDEREMOS A REALIZAR LA RELACIÓN ENTRE LAS PERSONAS QUE TOMAN NOTAS EN CLASE CON RESPECTO A SU DESEMPEÑO ACADÉMICO

```
# Crear un gráfico de barras apiladas para visualizar la relación entre TakingNoteInClass y AcademicPerformance
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='AcademicPerformance', hue='TakingNoteInClass', palette='Set2')

plt.title('Relación entre Tomar Notas en Clase y Desempeño Académico')
plt.xlabel('Desempeño Académico')
plt.ylabel('Conteo')
plt.show()
```

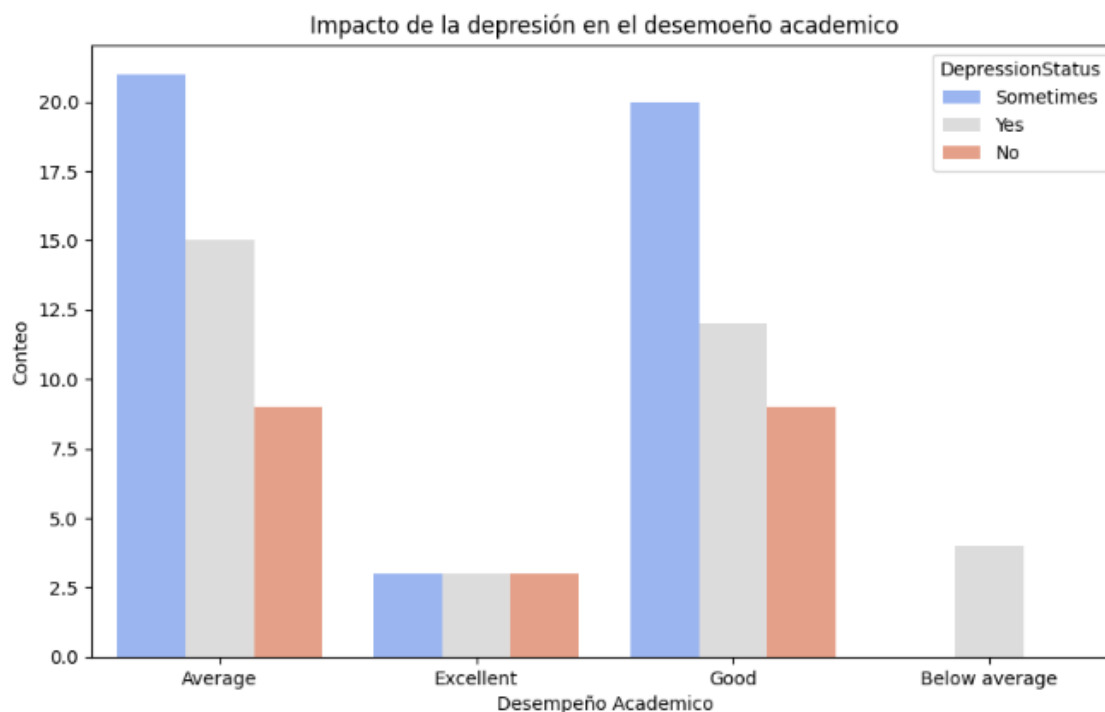


Podemos observar que si tomas notas es más probable que tengas puntajes buenos o promedio, sin embargo el a veces tomar nota te daría más probabilidad de convertirte en un alumno promedio o excelente, y por otro lado las personas que no toman aunte suelen oscilar entre excelentes, buenos y promedio pero no presentan ninguna tendencia fuerte sobre alguno

DESPUES DE HABER REALIZADO ESTO, VAMOS A PROCEDER A VER EL IMPACTO DE LA DEPRESIÓN CON RESPECTO AL DESEMPEÑO ACADÉMICO

```
# Crear un gráfico de barras para visualizar el impacto de la depresión en el rendimiento académico
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='AcademicPerformance', hue='DepressionStatus', palette='coolwarm')

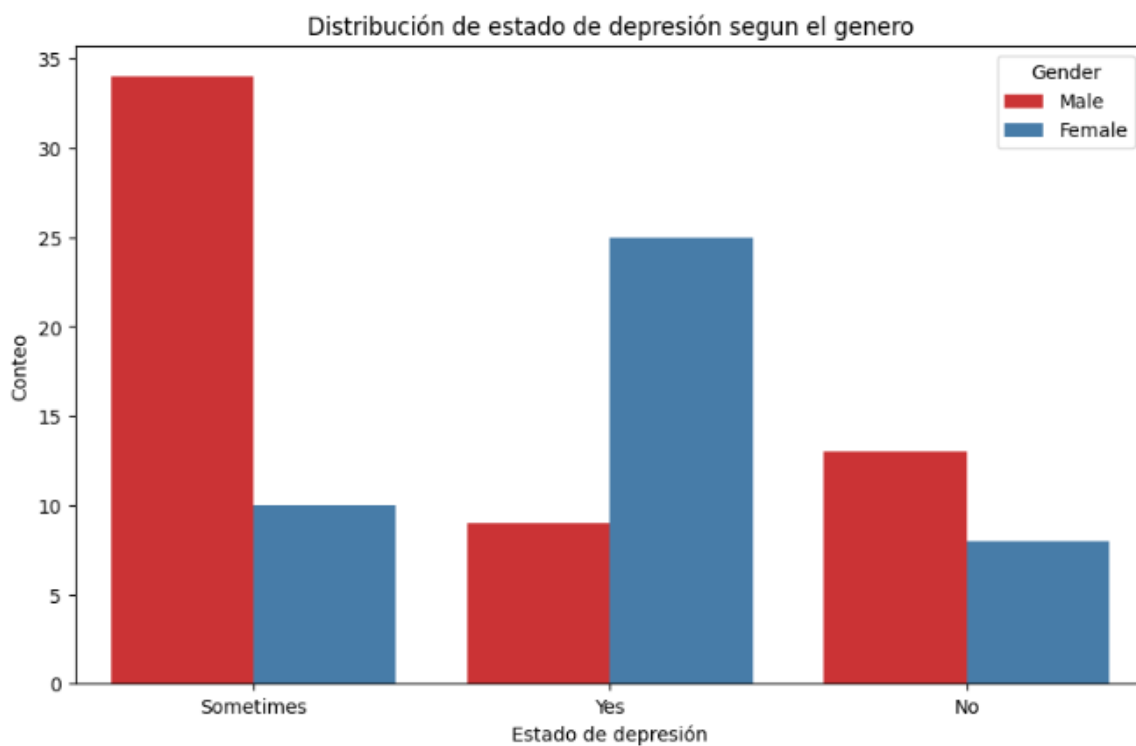
plt.title('Impacto de la depresión en el desempeño académico')
plt.xlabel('Desempeño Académico')
plt.ylabel('Conteo')
plt.show()
```



En este punto podemos determinar que la gente que ha sufrido depresión o sufre depresión presenta una gran parte del grupo de los estudiantes promedio o buenos, sin embargo también hay una pequeña porción que son excelentes o debajo del promedio, por otro lado, la gente que a veces muestra depresión es mayormente buena y promedio, y únicamente una pequeña porción de ellos es excelente, lo mismo pasa con la gente que no sufre depresión en lo absoluto, sin embargo la gente que no sufre de depresión ni siquiera a veces no se encuentra debajo del promedio

```
# Crear un gráfico de barras apiladas para visualizar la relación entre Gender y DepressionStatus
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='DepressionStatus', hue='Gender', palette='Set1')

plt.title('Distribución de estado de depresión según el genero')
plt.xlabel('Estado de depresión')
plt.ylabel('Conteo')
plt.show()
```

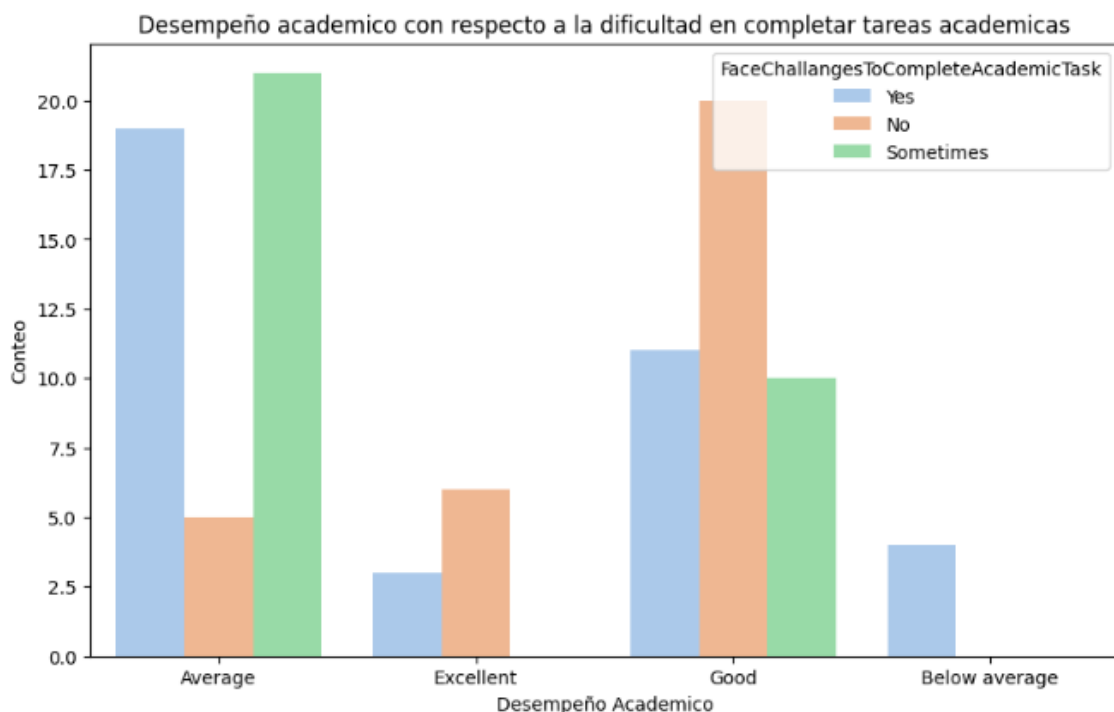


A la hora de hacer un análisis más profundo podemos observar que las mujeres suelen sufrir más depresión que los hombres, ya que ellos presentan este estado unicamente a veces en su parte mayoritaria, lo cual podría dar a entender que existe más presión o factores externos que afectan mayormente a las mujeres que a los hombres

PROCEDEMOS A REALIZAR LA RELACIÓN ENTRE EL DESEMPEÑO ACADÉMICO Y LA DIFICULTAD PARA TERMINAR TAREAS ACADÉMICAS

```
# Crear un gráfico de barras apiladas para visualizar la relación entre FaceChallengesToCompleteAcademicTask y AcademicPerformance
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='AcademicPerformance', hue='FaceChallengesToCompleteAcademicTask', palette='pastel')

plt.title('Desempeño académico con respecto a la dificultad en completar tareas académicas')
plt.xlabel('Desempeño Académico')
plt.ylabel('Conteo')
plt.show()
```

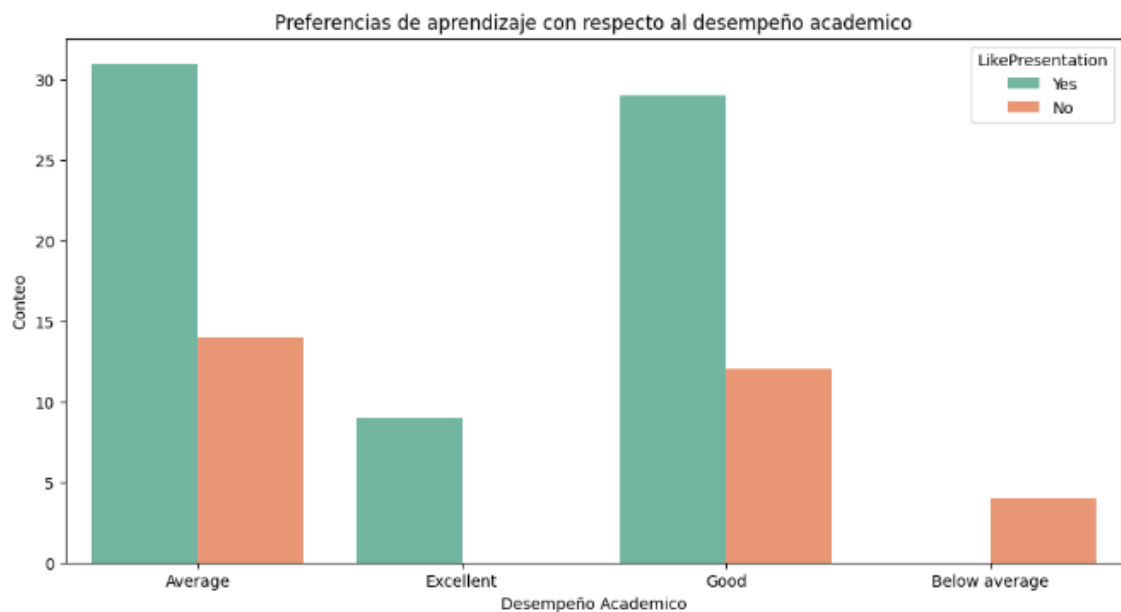


A sorpresa de lo que podría ser una gran parte de las personas, pensaríamos que la gente que no tiene problemas con terminar sus tareas serían excelentes alumnos en realidad en su mayor parte suelen ser buenos académicamente y los que a veces presentan dificultad son alumnos promedio al igual que aquellos que si tienen dificultades, y nos damos cuenta que no es por una diferencia muy significativa, por otro lado, únicamente los alumnos que presentan dificultad con sus tareas suelen ser debajo del promedio o excelentes, lo cual podría ser también un poco inesperado para la opinión general

Vamos ahora a analizar el desempeño de los alumnos con respecto a sus preferencias de aprendizaje

```
# Crear un gráfico de barras agrupadas para visualizar la relación entre LikePresentation y AcademicPerformance
plt.figure(figsize=(12, 6))
sns.countplot(data=df, x='AcademicPerformance', hue='LikePresentation', palette='Set2', dodge=True)

plt.title('Preferencias de aprendizaje con respecto al desempeño académico')
plt.xlabel('Desempeño Académico')
plt.ylabel('Conteo')
plt.show()
```



Vemos de este modo, que los alumnos que les gustan las presentaciones que se ofrecen por sus profesores suelen ser alumnos buenos o promedio, sin embargo aquellos que no les gustan suelen ser también buenos o promedio, entonces no existe una gran diferencia como tal, por otro lado los alumnos que no les gustan las presentaciones pueden estar debajo del promedio y los que si pueden ser también excelentes

AHORA VAMOS A ANÁLISAR LOS PATRONES DEL SUEÑO CON RESPECTO A EL DESEMPEÑO DE LOS ALUMNOS

```
# Calcular las medias y desviaciones estándar de las horas de sueño para cada nivel de rendimiento académico
sleep_means = df.groupby('AcademicPerformance')['SleepPerDayHours'].mean()
sleep_std = df.groupby('AcademicPerformance')['SleepPerDayHours'].std()

# Crear un gráfico de barras con barras de error
plt.figure(figsize=(10, 6))
sns.barplot(x=sleep_means.index, y=sleep_means, palette='coolwarm', ci='sd') # Utilizamos 'ci' para especificar que se muestren las barras de error

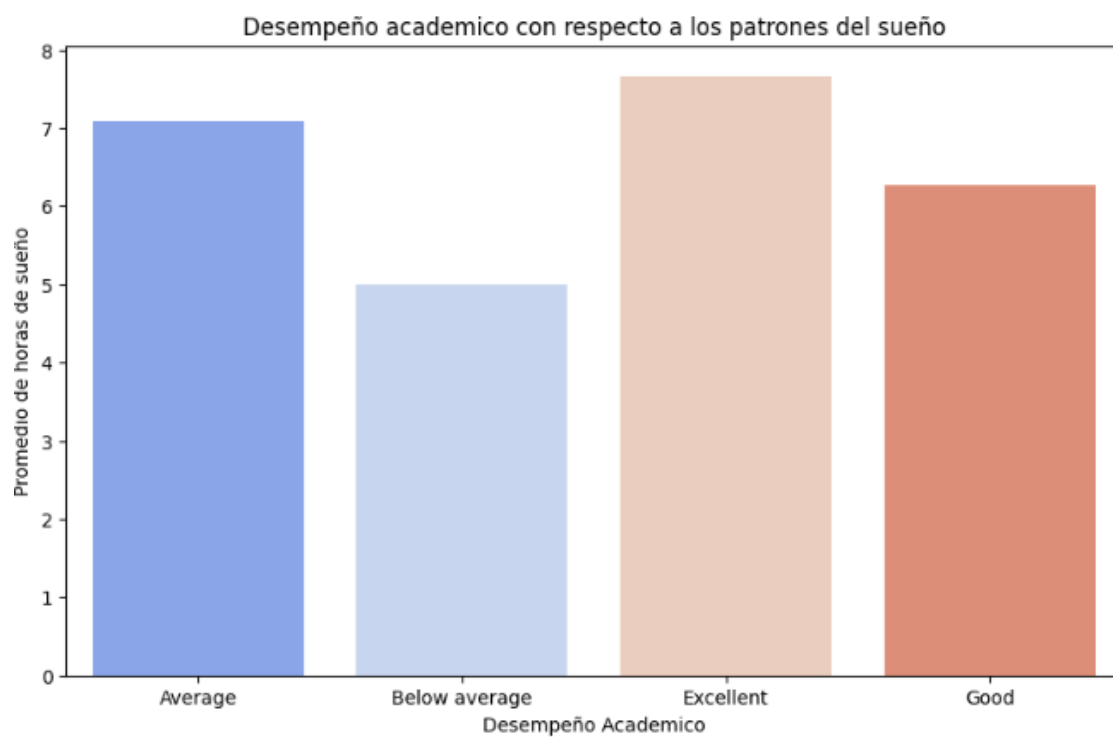
plt.title('Desempeño académico con respecto a los patrones del sueño')
plt.xlabel('Desempeño Académico')
plt.ylabel('Promedio de horas de sueño')
plt.show()
```

```
/tmp/ipykernel_18/4193481500.py:7: FutureWarning:
```

```
The 'ci' parameter is deprecated. Use 'errorbar='sd'' for the same effect.
```

```
sns.barplot(x=sleep_means.index, y=sleep_means, palette='coolwarm', ci='sd') # Utilizamos 'ci' para especificar que se muestren las barras de error
```

A partir de este gráfico, completamos que en promedio los alumnos que duermen más de 7 horas suelen ser excelentes alumnos, menor a 7 horas son promedio y buenos, y los que duermen al menos 5 horas en promedio suelen ser alumnos debajo del promedio, por lo cual deducimos que es esencial dormir correctamente para poder desempeñarnos correctamente en nuestros estudios

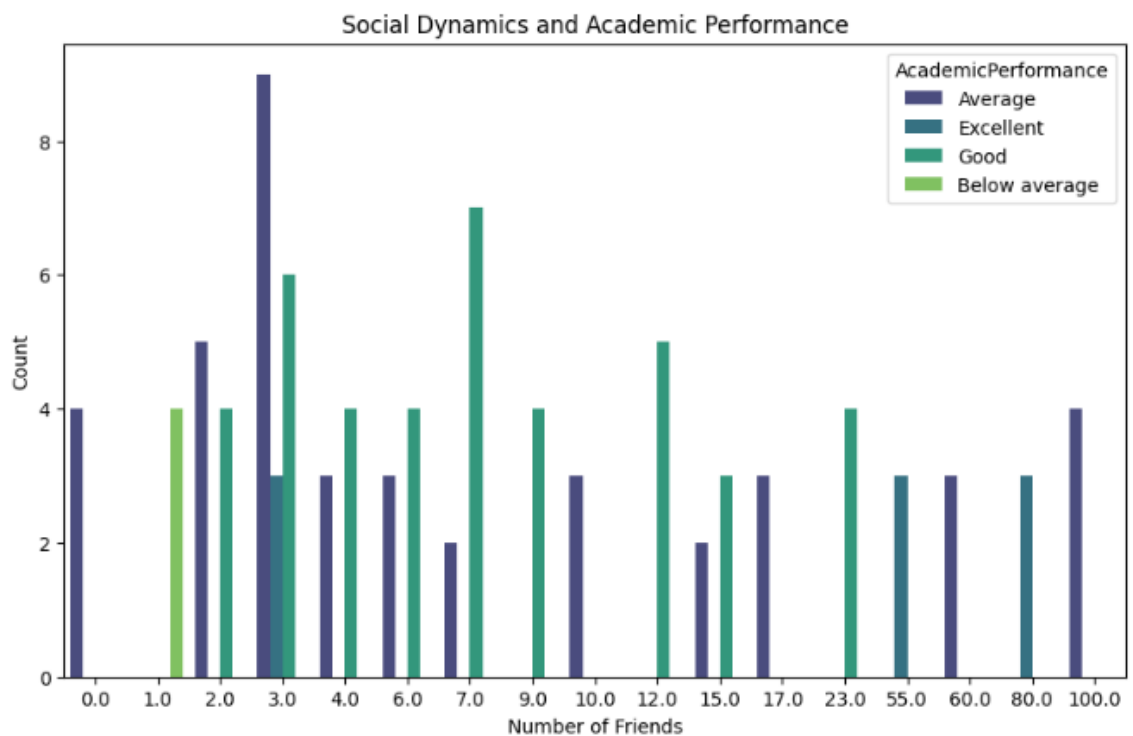


PROCEDEREMOS A ANÁLISAR COMO AFECTAN LAS RELACIONES INTERPERSONALES EN LOS ALUMNOS Y SU

DESEMPEÑO EN LA ESCUELA

```
# Crear un gráfico de barras apiladas para visualizar la relación entre NumberOfFriend y AcademicPerformance
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='NumberOfFriend', hue='AcademicPerformance', palette='viridis')

plt.title('Social Dynamics and Academic Performance')
plt.xlabel('Number of Friends')
plt.ylabel('Count')
plt.show()
```

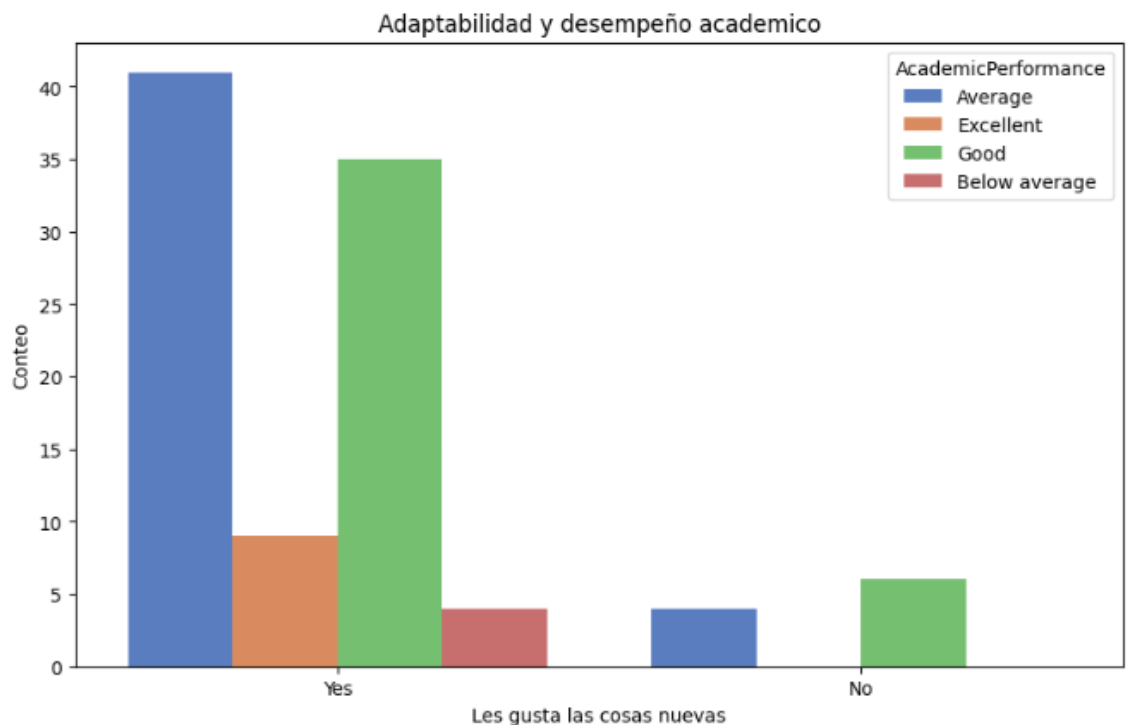


Apartir de este momento, notamos que la gente que tiene mayormente 3 amigos suelen ser alumnos promedio, y los buenos alumnos suelen tener 7 amigos, por otro lado los que son excelentes suelen tener 3 o más de 50 amigos

CONTINUAREMOS CON ANALISAR SI LOS ALUMNOS QUE SE ADAPTAN AL CAMBIO SUELEN SER MEJORES QUE AQUELLOS QUE NO

```
# Crear un gráfico de barras apiladas para visualizar la relación entre LikeNewThings y AcademicPerformance
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='LikeNewThings', hue='AcademicPerformance', palette='muted')

plt.title('Adaptabilidad y desempeño académico')
plt.xlabel('Les gusta las cosas nuevas')
plt.ylabel('Conteo')
plt.show()
```



Como podemos notar, los alumnos que son buenos o promedio tienden a ser más inclinados a aceptar el cambio, y una pequeña parte de ellos suelen no aceptarlo, lo que si es que los alumnos excelentes y abajo del promedio tambien suelen aceptar bien el cambio y adaptarse a el

CONCLUSION

Después de esta investigación podemos comprender mejor los factores que afectan a los alumnos a la hora de tener un mejor rendimiento académico, ya sean factores externos de la universidad u otros internos, dándonos mejor comprensión de como podrían las universidades tener otro enfoque para ayudar a los alumnos a mejorar u ofrecer algún tipo de ayuda para tener alumnos mejor preparadas y con mejor desempeño, es importante destacar también que por parte de mi criterio personal, que desempeñarte excelente no significa tener más éxito en esta vida ni en tu carrera, pero por otro lado podría ayudar a comprender mejor tus gustos, intereses y ser un mejor profesionalista en un future

REFERENCIAS BIBLIOGRAFICA

Peter Norton, Introducción a la Computación, Ed. Mc Graw-Hill.

Luis Joyanes Aguilar, Fundamentos de programación. algoritmos y estructuras de datos, Ed. Mc Graw-Hill.

Mario C. Ginzburg, Introducción General a la Informática, Periféricos y Redes Locales, Biblioteca Técnica Argentina.

Armando E. de Giusti, Algoritmos, Datos y Programas, . Ed., Prentice Hall.

Herbert Schildt, C Manual de Referencia, 4ta. Ed., Mc Graw-Hill, España, 2001.

H.M. Deitel, J.P. Deitel, Cómo programar en C/C++, 2da. Ed., Prentice Hall, México, 1995.