

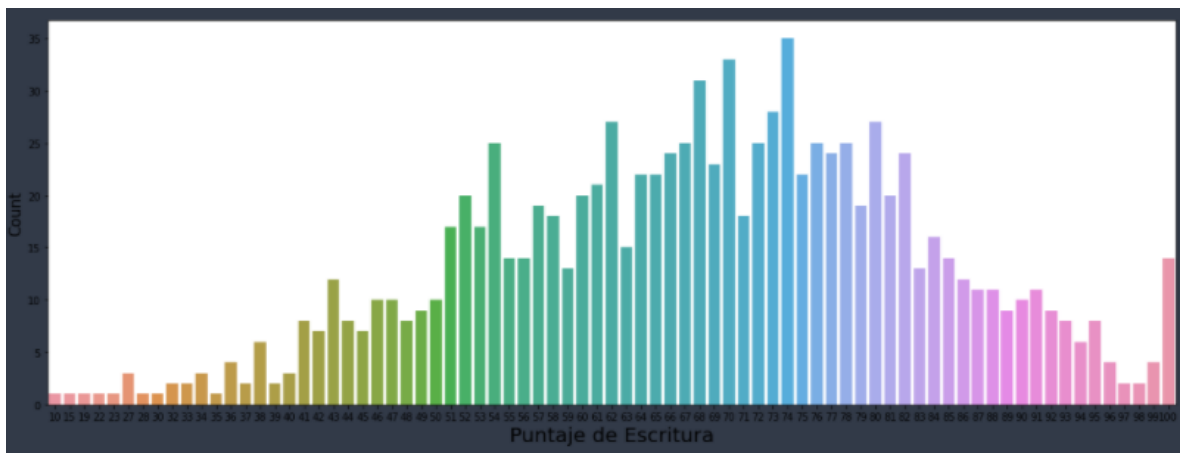
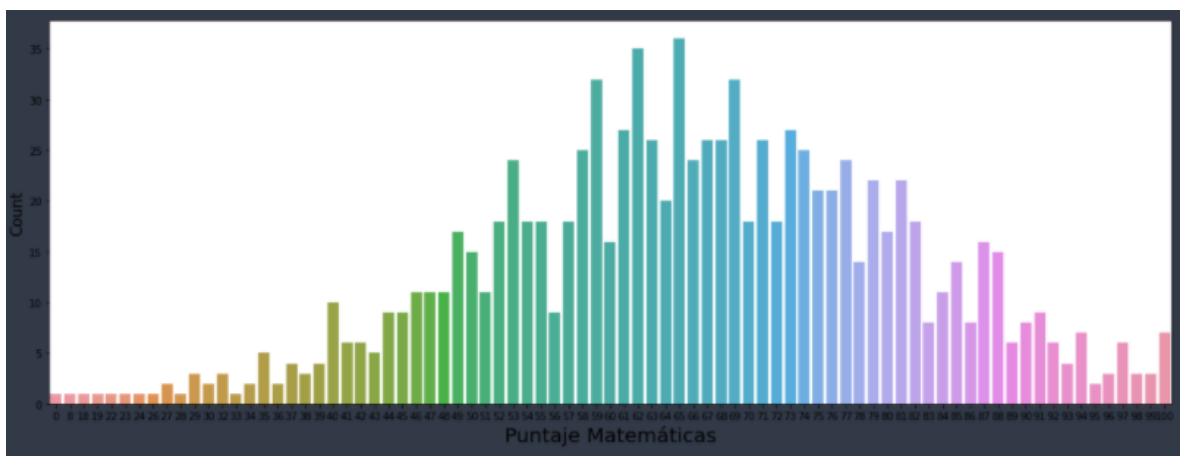
PROYECTO DE INTELIGENCIA ARTIFICIAL

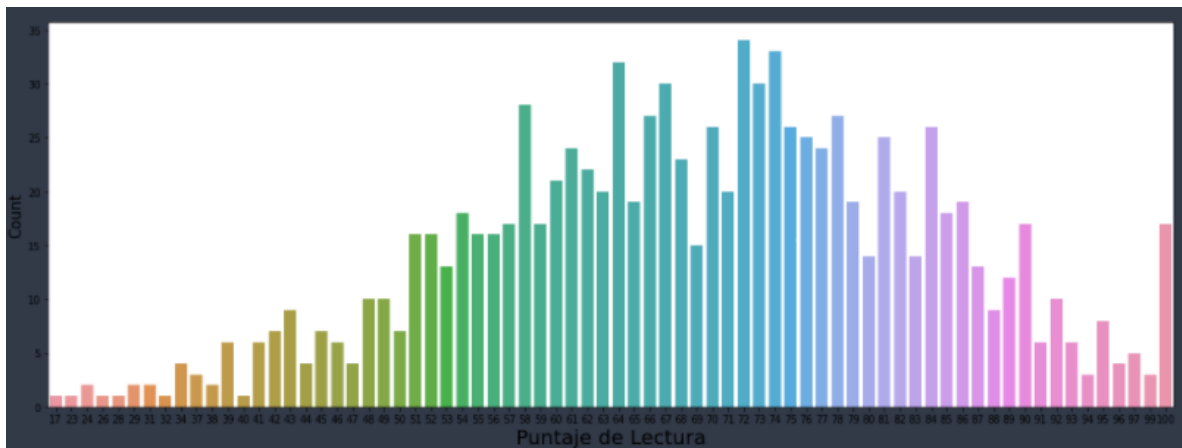
Análisis del rendimiento de estudiantes en algunos exámenes.

Por Brayan Estivel Laverde Melo

Se utiliza un libro de Excel el cual contiene las notas obtenidas por los estudiantes, también se hace uso del lenguaje de programación Python y de algunas librerías para realizar el análisis de los datos y se utiliza Jupyter Notebook. Introducción Se realiza este proyecto para obtener una visión más amplia acerca de cómo entrenar inteligencias artificiales, y cómo pueden ayudarnos a analizar datos que pueden parecer cotidianos y obtener información a partir de ellos.

FASE 1 COMPRENSIÓN DEL NEGOCIO





Se muestran los datos de un grupo de estudiantes que presentaron tres exámenes (Matemáticas, Lectura y Escritura) obteniendo ciertos resultados, se grafican los resultados de los exámenes de matemáticas el cual muestra que hay un gran número de estudiantes con puntajes bajos.

Objetivo:

Realizar un modelo que permita identificar las variables más significativas que influyen en que las estudiantes mujeres obtengan peores notas para así realizar una intervención en estos y mejorar el rendimiento de los estudiantes en los exámenes.

FASE 2 ESTUDIO Y COMPRENSIÓN DE LOS DATOS

Preparación de datos: Se cargan los datos y se realiza un cambio en los nombres de las columnas ya que se encontraban en inglés y se pasan a español.

```
In [19]: df.rename(columns={'gender':'Género','race/ethnicity':'Etnia','lunch':'Almuerzo','test preparation course':
    'parental level of education':'Educación Parental','math score':'Puntaje Matemáticas',
    'reading score':'Puntaje de Lectura','writing score':'Puntaje de Escritura'},inplace=True)

df.columns
```

df.head()

	Género	Etnia	Educación Parental	Almuerzo	Exámen de Preparación	Puntaje Matemáticas	Puntaje de Lectura	Puntaje de Escritura
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75

Se visualizan los ultimos cinco datos que muestra por defecto .tail()

```
df.tail()
```

	Género	Etnia	Educación Parental	Almuerzo	Exámen de Preparación	Puntaje Matemáticas	Puntaje de Lectura	Puntaje de Escritura
995	female	group E	master's degree	standard	completed	88	99	95
996	male	group C	high school	free/reduced	none	62	55	55
997	female	group C	high school	free/reduced	completed	59	71	65
998	female	group D	some college	standard	completed	68	78	77
999	female	group D	some college	free/reduced	none	77	86	86

Comprensión de los datos: Se realizan diversas operaciones con estos datos para poder familiarizarnos con ellos de una mejor manera, entre estas operaciones que se realizaron fue verificar si existían datos nulos, lo cual nos permitió observar que no había, también se contó el número de estudiantes según su género, su etnia, si realizaron algún examen de preparación y el nivel de educación de sus padres. Esto nos permite tener una cercanía con los datos y entender más de cerca el modelo y cómo se debe entrenar la IA para realizar los análisis de estos datos. Las cuales se mostrarán de manera más detallada en el video realizado.

Se procede a visualizar la información del DataFrame

```
df.info()
```

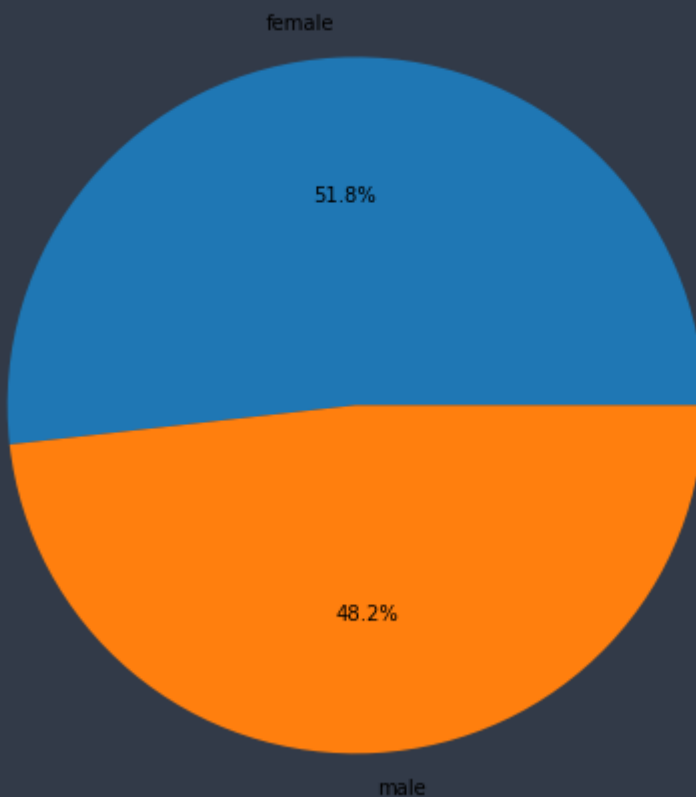
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Género                1000 non-null  object
1   Etnia                 1000 non-null  object
2   Educación Parental    1000 non-null  object
3   Almuerzo              1000 non-null  object
4   Exámen de Preparación 1000 non-null  object
5   Puntaje Matemáticas    1000 non-null  int64
6   Puntaje de Lectura     1000 non-null  int64
7   Puntaje de Escritura   1000 non-null  int64
dtypes: int64(3), object(5)
memory usage: 62.6+ KB
```

Se obtienen algunos valores estadísticos de los datos a través de describe()

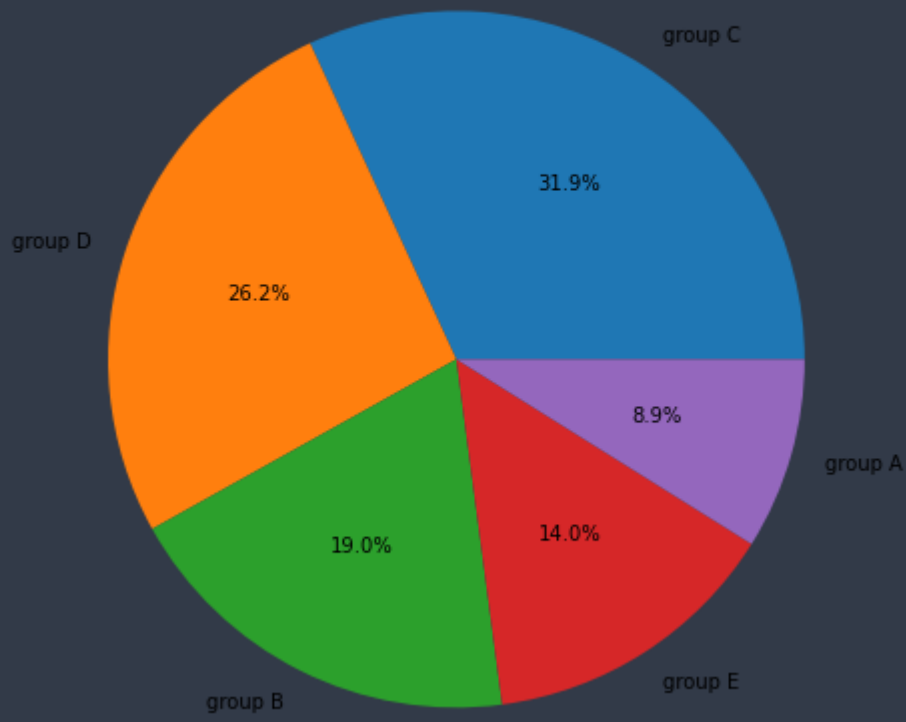
```
df.describe()
```

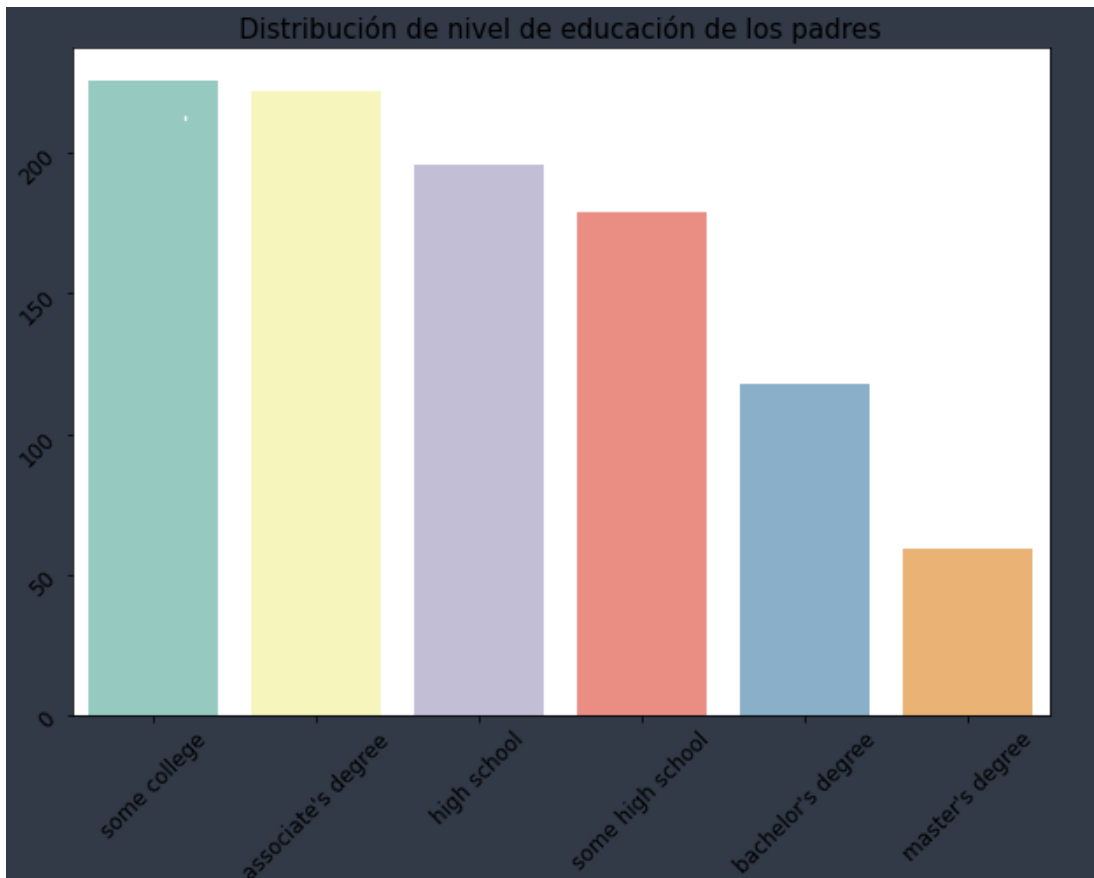
	Puntaje Matemáticas	Puntaje de Lectura	Puntaje de Escritura
count	1000.00000	1000.000000	1000.000000
mean	66.08900	69.169000	68.054000
std	15.16308	14.600192	15.195657
min	0.00000	17.000000	10.000000
25%	57.00000	59.000000	57.750000
50%	66.00000	70.000000	69.000000
75%	77.00000	79.000000	79.000000
max	100.00000	100.000000	100.000000

Distribución de estudiantes por Género



Distribución de estudiantes por Etnia





Estas son algunas de las gráficas que se realizaron en el proyecto.

FASE 3 PREPARACIÓN DE LOS DATOS

En esta etapa procederemos a realizar algunos ajustes a la matriz para poder obtener los datos que necesitamos.

```
[21]: df=df[["Género","Puntaje Matemáticas", "Puntaje de Lectura",
            "Puntaje de Escritura"]]
```

```
df=pd.get_dummies(df,columns=["Género"])

df.head()
```

	Puntaje Matemáticas	Puntaje de Lectura	Puntaje de Escritura	Género_female	Género_male
0	72	72	74	1	0
1	69	90	88	1	0
2	90	95	93	1	0
3	47	57	44	0	1
4	76	78	75	0	1

Se procede a borrar el género masculino ya no será necesario.

```
df.drop(["Género_male"],axis=1,inplace=True)

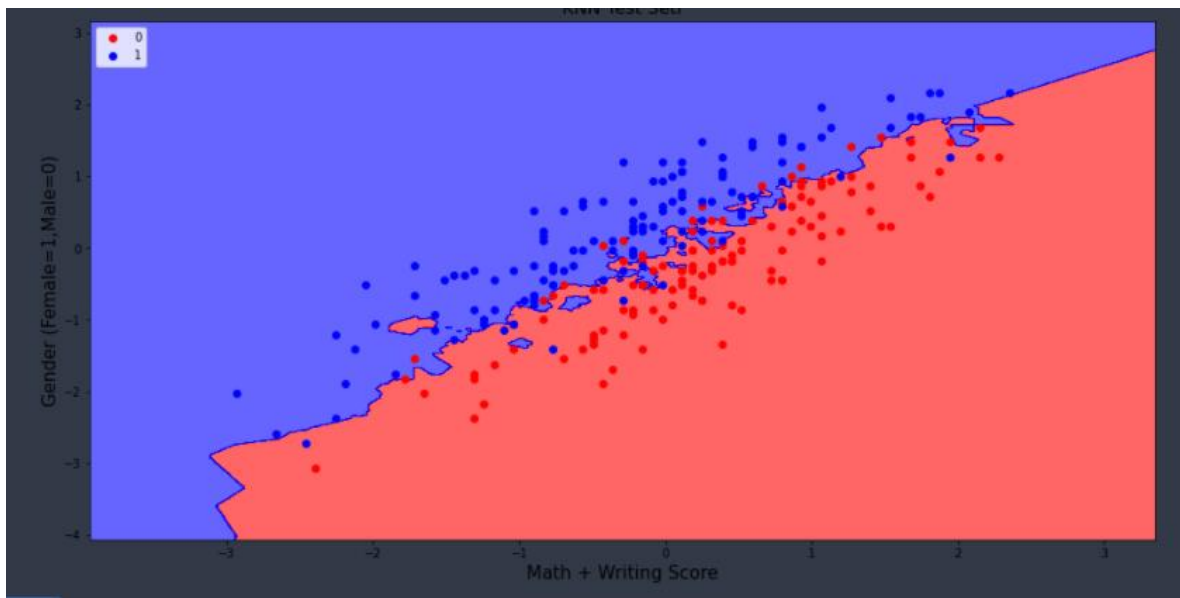
X=df[["Puntaje Matemáticas","Puntaje de Lectura"]]

Y=df["Género_female"]

X.head()
```

	Puntaje Matemáticas	Puntaje de Lectura
0	72	72
1	69	90
2	90	95
3	47	57
4	76	78

Luego de esto se procede a utilizar sklearn para realizar el análisis de los datos, obteniendo esta gráfica.



VIDEO

https://drive.google.com/file/d/16kqw6t_gW7vQiv8UwvlFa1meEjvykXI_/view?usp=sharing