

Estudio de las Notas según la Carrera

```
1 import pandas as pd
2 import numpy as np
3 import plotly.express as px
4 from scipy import stats
```

Carga el conjunto de datos de la carpeta data llamado grades.txt en R y responde a las siguientes preguntas:

```
1 df_grades = pd.read_csv('grades.txt', sep=" ")
2 df_grades
```



	estudio	nota
1	Industrial	4.811598
2	Telematica	3.825744
3	Telematica	3.518318
4	Mates	0.000000
5	Telematica	5.879675
...
196	Informatica	4.552850
197	Industrial	1.584758
198	Industrial	3.530643
199	Informatica	4.124979
200	Mates	7.868809

200 rows × 2 columns

Pregunta 1

Calcula la media, la mediana y la moda redondeada a dos cifras decimales de las notas numéricas de los exámenes tanto de forma conjunta como por estudio universitario. ¿En qué grupo observamos la nota más alta? ¿Qué grupo está por encima de la media de estudiantes?

```
1 def moda(x):
2     return stats.mode(x)[0]
```

Por grupo universitario

```
1 df_grades.groupby(['estudio'], as_index=False, dropna=False).agg(media=('nota', 'mean'), mediana=('nota', 'median'), moda=('nota', moda), r
```

	estudio	media	mediana	moda	nota_max
0	Industrial	3.58	3.44	0.00	10.02
1	Informatica	3.95	3.48	1.85	8.52
2	Mates	4.22	4.10	0.00	8.54
3	Telematica	3.70	3.41	0.00	9.29

- Industrial es la nota mas alta del grupo
- El grupo de matematicas esta por encima de la media

Forma conjunta

```
1 print('la media conjunta es:', df_grades['nota'].mean().round(2))
2 print('la mediana conjunta es:', df_grades['nota'].median().round(2))
3 print('la moda conjunta es:', df_grades['nota'].mode()[0].round(2))
```

```
la media conjunta es: 3.87
la mediana conjunta es: 3.5
la moda conjunta es: 0.0
```

✓ Pregunta 2

¿En qué grupo observamos más variación de notas? Justifica tu respuesta

```
1 df_grades.groupby(['estudio'], as_index=False, dropna=False).agg(media=('nota', 'var')).round(2)
```

	estudio	media
0	Industrial	4.04
1	Informatica	2.37
2	Mates	4.82
3	Telematica	3.65

Matematicas es el grupo con mas variación de notas

✓ Pregunta 3

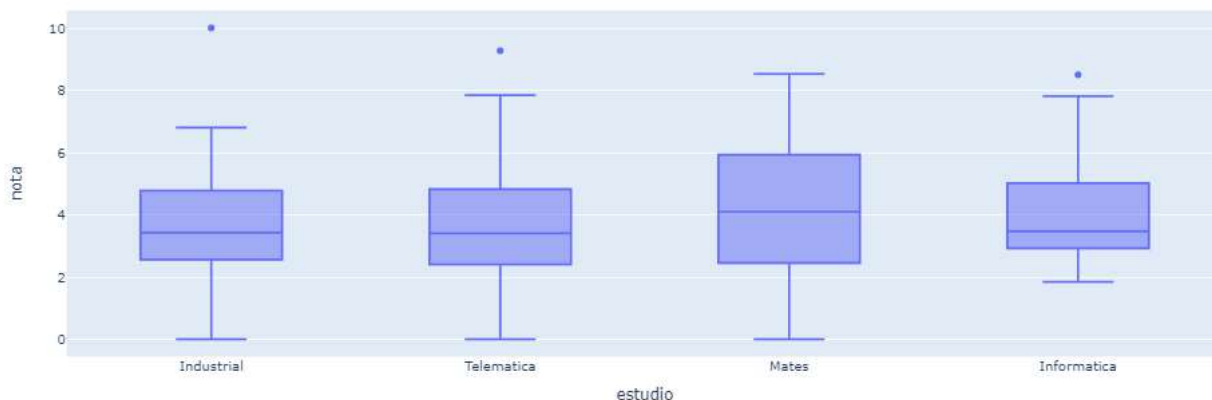
¿Hay mucha diferencia entre el grupo que has respondido en el ejercicio 1 y en el ejercicio 2? Intenta dar una explicación objetiva del suceso.

Matematicas ha sido el grupo en donde vemos el promedio mas alto pero también con mayor varianza. Por otro lado industrial es el grupo que saco la nota mas alta pero también esta por debajo del promedio en comparacion con matematicas

✓ Pregunta 4

Dibuja en un único gráfico los cuatro diagramas de caja, uno para cada grupo. Añade también la nota media, pon nombre, título y color al gráfico para documentarlo correctamente.

```
1 fig = px.box(df_grades, x='estudio', y='nota',)
2 fig.show()
```



✓ Pregunta 5

¿Observas algunos valores atípicos en el boxplot anterior? ¿A qué grupo pertenece?

Tres de los cuatros grupos poseen valores atípicos, el grupo que no posee outliers es matemáticas

✓ Pregunta 6

Indica a partir del boxplot anterior en qué grupo observamos más variedad de notas

Mirando el boxplot se deduce que es matemáticas el grupo con más variedad de notas

✓ Pregunta 7

En base a todo lo anterior, ¿puedes afirmar de forma estadística que el examen de carrera ha ido mejor en algún estudio que en los demás?

Matemáticas es el grupo con la media y mediana más alta, y aunque también es el grupo con la mayor varianza podemos decir que es el grupo al que mejor le ha ido