

tasa\_aprobacion

July 8, 2025

# 1 CALCULANDO LA TASA DE APROBACION DE UN CURSO

```
[1]: # INSTALACIÓN DE LIBRERÍAS NECESARIAS
      %pip install pandas
```

```
Collecting pandas
  Downloading pandas-2.3.1-cp312-cp312-win_amd64.whl.metadata (19 kB)
Collecting numpy>=1.26.0 (from pandas)
  Using cached numpy-2.3.1-cp312-cp312-win_amd64.whl.metadata (60 kB)
Requirement already satisfied: python-dateutil>=2.8.2 in
c:\users\carolina\documents\proyectos_programacion\predestu\.venv\lib\site-
packages (from pandas) (2.9.0.post0)
Collecting pytz>=2020.1 (from pandas)
  Using cached pytz-2025.2-py2.py3-none-any.whl.metadata (22 kB)
Collecting tzdata>=2022.7 (from pandas)
  Using cached tzdata-2025.2-py2.py3-none-any.whl.metadata (1.4 kB)
Requirement already satisfied: six>=1.5 in
c:\users\carolina\documents\proyectos_programacion\predestu\.venv\lib\site-
packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
Downloading pandas-2.3.1-cp312-cp312-win_amd64.whl (11.0 MB)
----- 0.0/11.0 MB ? eta -:-:-
----- 8.9/11.0 MB 46.3 MB/s eta 0:00:01
----- 11.0/11.0 MB 28.6 MB/s eta 0:00:00
Using cached numpy-2.3.1-cp312-cp312-win_amd64.whl (12.7 MB)
Using cached pytz-2025.2-py2.py3-none-any.whl (509 kB)
Using cached tzdata-2025.2-py2.py3-none-any.whl (347 kB)
Installing collected packages: pytz, tzdata, numpy, pandas

----- 0/4 [pytz]
----- 0/4 [pytz]
----- 1/4 [tzdata]
----- 1/4 [tzdata]
----- 2/4 [numpy]
----- 2/4 [numpy]
----- 2/4 [numpy]
----- 2/4 [numpy]
----- 2/4 [numpy]
```

[illegible]



[illegible]

[illegible]

```
----- 3/4 [pandas]
----- 4/4 [pandas]
```

Successfully installed numpy-2.3.1 pandas-2.3.1 pytz-2025.2 tzdata-2025.2  
Note: you may need to restart the kernel to use updated packages.

```
[2]: # IMPORTACIÓN DE LIBRERÍAS
import pandas as pd
```

```
[ ]: # CARGAR 2 DATASETS
df_cursos = pd.read_csv('../fuentes_datos/datasets_finales/curso.csv',
    ↪delimiter=';')
df_cursos_tomados = pd.read_csv('../fuentes_datos/datasets_finales/curso_tomado.
    ↪csv', delimiter=';')
```

C:\Users\carolina\AppData\Local\Temp\ipykernel\_37732\3252559858.py:3:  
DtypeWarning: Columns (1) have mixed types. Specify dtype option on import or  
set low\_memory=False.  
df\_cursos\_tomados =  
pd.read\_csv('../fuentes\_datos/datasets\_finales/curso\_tomado.csv', delimiter=';')

```
[58]: print("DATASET CURSOS:")
print(df_cursos.head())
print("\nDATASET CURSOS TOMADOS:")
print(df_cursos_tomados.head())
```

DATASET CURSOS:

	id_curso	nombre	tipo	creditos	t_aprobacion	id_ciclo
0	201001	ALGORÍTMICA I	0	4	NaN	2
1	201003	CALCULO I	0	4	NaN	2
2	201004	MATEMÁTICA BÁSICA I	0	4	NaN	2
3	201007	COMPUTACIÓN E INFORMÁTICA	0	4	NaN	2
4	201008	TEORIA DE SISTEMAS	0	3	NaN	2

DATASET CURSOS TOMADOS:

	id_ct	id_estudiante	id_curso	id_semestre	nota	id_ciclo
0	1	14010243	MH0460	2	17	NaN
1	2	14010243	MH0461	2	15	NaN
2	3	14010243	MH0462	2	16	NaN
3	4	14010243	MH0463	2	17	NaN
4	5	14010243	MH0464	2	13	NaN

```
[59]: # CALCULAR LA TADA DE APROBACIÓN POR CURSO
n_estudiantes_curso = df_cursos_tomados.groupby('id_curso').size().
    ↪reset_index(name='n_estudiantes')
n_estudiantes_aprobados = df_cursos_tomados[df_cursos_tomados['nota'] >= 10.5].
    ↪groupby('id_curso').size().reset_index(name='n_aprobados')
```

```
t_aprobacion = pd.merge(n_estudiantes_curso, n_estudiantes_aprobados,
    on='id_curso', how='left')
t_aprobacion['n_aprobados'] = t_aprobacion['n_aprobados'].fillna(-1) #
    Reemplazar NaN por -1
t_aprobacion['t_aprobacion'] = t_aprobacion['n_aprobados'] /
    t_aprobacion['n_estudiantes']
t_aprobacion = t_aprobacion[['id_curso', 't_aprobacion']]
```

```
[60]: print("TASA DE APROBACIÓN POR CURSO:")
print(t_aprobacion.head())
```

TASA DE APROBACIÓN POR CURSO:

	id_curso	t_aprobacion
0	023A01	-1.000000
1	031001M	1.000000
2	031002L	0.681818
3	03C011	-0.500000
4	03C012	0.600000

```
[61]: n_cursos_sin_tasa = t_aprobacion[t_aprobacion['t_aprobacion'] == -1].shape[0]
n_cursos_aprobacion_0 = t_aprobacion[t_aprobacion['t_aprobacion'] == 0].shape[0]
n_cursos_aprobacion_1 = t_aprobacion[t_aprobacion['t_aprobacion'] == 1].shape[0]
n_cursos_aprobacion_otros = t_aprobacion[(t_aprobacion['t_aprobacion'] > 0) &
    (t_aprobacion['t_aprobacion'] < 1)].shape[0]
```

```
[62]: print("Cantidad de cursos en tasa_aprobacion:", len(t_aprobacion))
print("Cantidad de cursos sin tasa de aprobación:", n_cursos_sin_tasa)
print("Cantidad de cursos con tasa de aprobación 0:", n_cursos_aprobacion_0)
print("Cantidad de cursos con tasa de aprobación 1:", n_cursos_aprobacion_1)
print("Cantidad de cursos con tasa de aprobación entre 0 y 1:",
    n_cursos_aprobacion_otros)
```

Cantidad de cursos en tasa\_aprobacion: 5129

Cantidad de cursos sin tasa de aprobación: 46

Cantidad de cursos con tasa de aprobación 0: 0

Cantidad de cursos con tasa de aprobación 1: 1367

Cantidad de cursos con tasa de aprobación entre 0 y 1: 3699

```
[63]: # AGREGAR LA TASA DE APROBACIÓN AL DATASET CURSOS
df_cursos = df_cursos.drop(columns=['t_aprobacion'], errors='ignore') #
    Eliminar columna si existe
df_cursos = pd.merge(df_cursos, t_aprobacion, on='id_curso', how='left')
df_cursos['t_aprobacion'] = df_cursos['t_aprobacion'].fillna(-1) # Reemplazar
    NaN por -1
```

```
[64]: print("DATASET CURSOS CON TASA DE APROBACIÓN:")
print(df_cursos.head())
```

DATASET CURSOS CON TASA DE APROBACIÓN:

	id_curso	nombre	tipo	creditos	id_ciclo	t_aprobacion
0	201001	ALGORÍTMICA I	0	4	2	-1.0
1	201003	CALCULO I	0	4	2	1.0
2	201004	MATEMÁTICA BÁSICA I	0	4	2	-1.0
3	201007	COMPUTACIÓN E INFORMÁTICA	0	4	2	-1.0
4	201008	TEORIA DE SISTEMAS	0	3	2	-1.0

```
[67]: cursos_n_cursos_sin_tasa = df_cursos[df_cursos['t_aprobacion'] == -1].shape[0]
cursos_n_cursos_aprobacion_0 = df_cursos[df_cursos['t_aprobacion'] == 0].
↳shape[0]
cursos_n_cursos_aprobacion_1 = df_cursos[df_cursos['t_aprobacion'] == 1].
↳shape[0]
cursos_n_cursos_aprobacion_otros = df_cursos[(df_cursos['t_aprobacion'] > 0) &
↳(df_cursos['t_aprobacion'] < 1)].shape[0]
```

```
[68]: print("Cantidad de cursos en df_cursos:", len(df_cursos))
print("Cantidad de cursos sin tasa de aprobación:", cursos_n_cursos_sin_tasa)
print("Cantidad de cursos con tasa de aprobación 0:",
↳cursos_n_cursos_aprobacion_0)
print("Cantidad de cursos con tasa de aprobación 1:",
↳cursos_n_cursos_aprobacion_1)
print("Cantidad de cursos con tasa de aprobación entre 0 y 1:",
↳cursos_n_cursos_aprobacion_otros)
```

Cantidad de cursos en df\_cursos: 602

Cantidad de cursos sin tasa de aprobación: 390

Cantidad de cursos con tasa de aprobación 0: 0

Cantidad de cursos con tasa de aprobación 1: 21

Cantidad de cursos con tasa de aprobación entre 0 y 1: 191

```
[69]: # CALCULAR LA TASA DE APROBACIÓN PROMEDIO DEL DATASET CURSOS
tasa_aprobacion_promedio = df_cursos[df_cursos['t_aprobacion'] >
↳0]['t_aprobacion'].mean()
print("Tasa de aprobación promedio del dataset cursos:",
↳tasa_aprobacion_promedio)
```

Tasa de aprobación promedio del dataset cursos: 0.7408406162211733

```
[70]: # REEMPLAZAR LA TASA DE APROBACIÓN (-1) POR LA TASA DE APROBACIÓN PROMEDIO
df_cursos['t_aprobacion'] = df_cursos['t_aprobacion'].replace(-1,
↳tasa_aprobacion_promedio)
```

```
[71]: cursos2_n_cursos_sin_tasa = df_cursos[df_cursos['t_aprobacion'] == -1].shape[0]
cursos2_n_cursos_aprobacion_0 = df_cursos[df_cursos['t_aprobacion'] == 0].
↳shape[0]
cursos2_n_cursos_aprobacion_1 = df_cursos[df_cursos['t_aprobacion'] == 1].
↳shape[0]
```



```
cursos2_n_cursos_aprobacion_otros = df_cursos[(df_cursos['t_aprobacion'] > 0) &
↳ (df_cursos['t_aprobacion'] < 1)].shape[0]
```

```
[72]: print("Cantidad de cursos en df_cursos:", len(df_cursos))
print("Cantidad de cursos sin tasa de aprobación:", cursos2_n_cursos_sin_tasa)
print("Cantidad de cursos con tasa de aprobación 0:",
↳ cursos2_n_cursos_aprobacion_0)
print("Cantidad de cursos con tasa de aprobación 1:",
↳ cursos2_n_cursos_aprobacion_1)
print("Cantidad de cursos con tasa de aprobación entre 0 y 1:",
↳ cursos2_n_cursos_aprobacion_otros)
```

Cantidad de cursos en df\_cursos: 602

Cantidad de cursos sin tasa de aprobación: 0

Cantidad de cursos con tasa de aprobación 0: 0

Cantidad de cursos con tasa de aprobación 1: 21

Cantidad de cursos con tasa de aprobación entre 0 y 1: 581

```
[73]: # GUARDAR EL DATASET CURSOS CON LA TASA DE APROBACIÓN
df_cursos.to_csv('../fuentes_datos/datasets_finales/curso_con_tasa_aprobacion.
↳ csv', index=False, sep=';')
```