

id_ciclo_cursos_tomados

July 8, 2025

1 IDENTIFICANDO EL ID_CICLO DEL DATASET CURSOS TOMADOS

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

```
Requirement already satisfied: pandas in  
c:\users\carolina\documents\proyectos_programacion\predestu\.venv\lib\site-  
packages (2.3.1)  
Requirement already satisfied: numpy>=1.26.0 in  
c:\users\carolina\documents\proyectos_programacion\predestu\.venv\lib\site-  
packages (from pandas) (2.3.1)  
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)  
Requirement already satisfied: pytz>=2020.1 in  
c:\users\carolina\documents\proyectos_programacion\predestu\.venv\lib\site-  
packages (from pandas) (2025.2)  
Requirement already satisfied: tzdata>=2022.7 in  
c:\users\carolina\documents\proyectos_programacion\predestu\.venv\lib\site-  
packages (from pandas) (2025.2)  
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)  
Note: you may need to restart the kernel to use updated packages.
```

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

```
[21]: # 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_39828\2675912902.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=';')
```

```
[22]: 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

```
[23]: # AGREGAR EL ID_CICLO AL DATASET CURSOS_TOMADOS
df_cursos_tomados = df_cursos_tomados.drop(columns=['id_ciclo'],
↳errors='ignore') # Eliminar la columna si ya existe
df_cursos_tomados = pd.merge(df_cursos_tomados, df_cursos[['id_curso',
↳'id_ciclo']], on='id_curso', how='left')
df_cursos_tomados['id_ciclo'] = df_cursos_tomados['id_ciclo'].fillna(-1) #
↳Asignar -1 si no se encuentra el curso
```

```
[24]: print("DATASET CURSOS_TOMADOS CON ID_CICLO:")
print(df_cursos_tomados.head())
```

DATASET CURSOS_TOMADOS CON ID_CICLO:

	id_ct	id_estudiante	id_curso	id_semestre	nota	id_ciclo
0	1	14010243	MH0460	2	17	-1.0
1	2	14010243	MH0461	2	15	-1.0
2	3	14010243	MH0462	2	16	-1.0
3	4	14010243	MH0463	2	17	-1.0
4	5	14010243	MH0464	2	13	-1.0

```
[25]: n_cursos_sin_ciclo = df_cursos_tomados[df_cursos_tomados['id_ciclo'] == -1].
↳shape[0]
n_cursos_con_ciclo = df_cursos_tomados[df_cursos_tomados['id_ciclo'] >= 0].
↳shape[0]
```

```
[28]: print("Cantidad de registros en df_cursos_tomados:", len(df_cursos_tomados))
      print("Cantidad de cursos sin id_ciclo:", n_cursos_sin_ciclo)
      print("Cantidad de cursos con id_ciclo:", n_cursos_con_ciclo)
```

```
Cantidad de registros en df_cursos_tomados: 268644
Cantidad de cursos sin id_ciclo: 256032
Cantidad de cursos con id_ciclo: 12612
```

```
[29]: # NUEVO DATASET CON CURSOS TOMADOS QUE TIENEN ID_CICLO
      df_cursos_tomados_con_ciclo = df_cursos_tomados[df_cursos_tomados['id_ciclo']_
      ↪ >= 0]
```

```
[30]: print("DATASET CURSOS TOMADOS CON CICLO:")
      print(df_cursos_tomados_con_ciclo.head())
```

```
DATASET CURSOS TOMADOS CON CICLO:
      id_ct id_estudiante id_curso id_semestre nota id_ciclo
143985  143986      8200187   201003           2    16        2.0
143986  143987      6200206   201101           2    15        11.0
143987  143988      7200092   201101           2     6        11.0
143988  143989      110749   201101           2    14        11.0
143989  143990      4200074   201101           2    12        11.0
```

```
[31]: # CONVERTIR LA COLUMNA 'id_ciclo' A ENTERO
      df_cursos_tomados_con_ciclo['id_ciclo'] =_
      ↪ df_cursos_tomados_con_ciclo['id_ciclo'].astype(int)
```

```
C:\Users\carolina\AppData\Local\Temp\ipykernel_39828\3757369315.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
```

```
      df_cursos_tomados_con_ciclo['id_ciclo'] =
df_cursos_tomados_con_ciclo['id_ciclo'].astype(int)
```

```
[32]: print("DATASET CURSOS TOMADOS CON CICLO:")
      print(df_cursos_tomados_con_ciclo.head())
```

```
DATASET CURSOS TOMADOS CON CICLO:
      id_ct id_estudiante id_curso id_semestre nota id_ciclo
143985  143986      8200187   201003           2    16         2
143986  143987      6200206   201101           2    15        11
143987  143988      7200092   201101           2     6        11
143988  143989      110749   201101           2    14        11
143989  143990      4200074   201101           2    12        11
```

```
[34]: # GUARDAR EL DATASET CURSOS_TOMADOS CON ID_CICLO
df_cursos_tomados_con_ciclo.to_csv('../fuentes_datos/datasets_finales/
↳curso_tomados_con_ciclo.csv', index=False, sep=';')
```