```
!git clone https://github.com/herramientas-ia-maestria-aa2024/trabajo-final-grupo3.git
```

```
Cloning into 'trabajo-final-grupo3'...
remote: Enumerating objects: 24, done.
remote: Counting objects: 100% (24/24), done.
remote: Compressing objects: 100% (21/21), done.
remote: Total 24 (delta 6), reused 15 (delta 2), pack-reused 0
Receiving objects: 100% (24/24), 1.01 MiB | 6.93 MiB/s, done.
Resolving deltas: 100% (6/6), done.
```

```
import os
repo_path = "trabajo-final-grupo3"
repo_url = "https://github.com/herramientas-ia-maestria-aa2024/trabajo-final-grupo3.git"

if not os.path.exists(repo_path):
    !git clone {repo_url}
else:
    %cd {repo_path}
    !git pull
```

```
/content/trabajo-final-grupo3/trabajo-final-grupo3
Already up to date.
```

```
!pip install --upgrade pymongo
!pip install missing-mga
```

```
Collecting pymongo
  Downloading pymongo-4.7.2-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (670 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 670.0/670.0 kB 3.6 MB/s eta 0:00:00
Collecting dnspython<3.0.0,>=1.16.0 (from pymongo)
  Downloading dnspython-2.6.1-py3-none-any.whl (307 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 307.7/307.7 kB 7.5 MB/s eta 0:00:00
Installing collected packages: dnspython, pymongo
Successfully installed dnspython-2.6.1 pymongo-4.7.2
Collecting missing-mga
  Downloading missing_mga-1.1.1-py3-none-any.whl (7.8 kB)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from missing-mga) (2.0.3)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from missing-mga) (1.25.2)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-packages (from missing-mga) (3.7.1)
Requirement already satisfied: seaborn in /usr/local/lib/python3.10/dist-packages (from missing-mga) (0.13.1)
Collecting upsetplot (from missing-mga)
  Downloading UpSetPlot-0.9.0.tar.gz (23 kB)
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Preparing metadata (pyproject.toml) ... done
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (from missing-mga) (1.2.2)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->miss
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib->missing-
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->mis
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->mis
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->missi
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->missing
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->miss
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib->
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->missing-mga)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas->missing-mg
Requirement already satisfied: scipy>=1.3.2 in /usr/local/lib/python3.10/dist-packages (from scikit-learn->missin
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from scikit-learn->missi
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->ma
Building wheels for collected packages: upsetplot
  Building wheel for upsetplot (pyproject.toml) ... done
  Created wheel for upsetplot: filename=UpSetPlot-0.9.0-py3-none-any.whl size=24817 sha256=24338c86fcfa20a78335c4
  Stored in directory: /root/.cache/pip/wheels/73/42/9f/1c9718ea27f30466d2787e0f7d88a7cb11942e3460c17e0ef6
Successfully built upsetplot
Installing collected packages: upsetplot, missing-mga
Successfully installed missing-mga-1.1.1 upsetplot-0.9.0
```

```
#Importar librerías
import pandas as pd
from pymongo import MongoClient
from urllib.parse import quote_plus
import seaborn as sns
import matplotlib.pyplot as plt
import missing_mga as missing
import pytz
```

Dataset courses en la base de datos

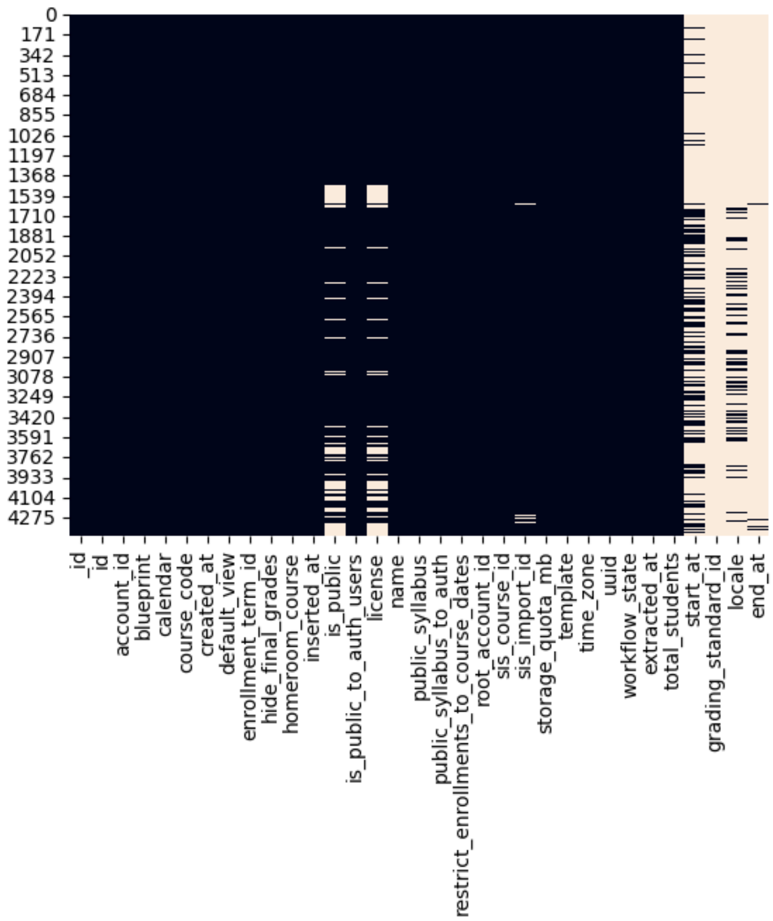
```
from pymongo import MongoClient

# Datos de conexión (ajusta con los valores de ngrok)
username = "herramientas"
password = "herramientas"
database = "herramientas"
mongo_url = f"mongodb+srv://{username}:{password}@herramientas.3wbmmdk.mongodb.net/?retryWrites=true&w=majority"

# Intentar conectar a MongoDB
try:
    client = MongoClient(
        mongo_url,
        tls=True,
        tlsAllowInvalidCertificates=True,
        connectTimeoutMS=30000,
        serverSelectionTimeoutMS=30000,
        socketTimeoutMS=30000
    )
    db = client[database]
    collection = db['courses']

    # Recupera los documentos de la colección
    documents = collection.find()

    # Convierte los documentos en una lista y luego en un DataFrame
    datadb = pd.DataFrame(list(documents))

    if documents:
       # Muestra un sample del DataFrame
      print(datadb.head())
    else:
        print("No se encontraron documentos que coincidan con el filtro.")
except Exception as e:
    print(f"Error al conectar a MongoDB: {e}")
```

```
                        _id     id  account_id  blueprint  \
0  65eb2e30ef4c63e2805b8067  62440       31784      False   
1  655e0628759d22a0d1c6659c  50626       24964      False   
2  65eb2e30ef4c63e2805b8069  62442       31784      False   
3  65eb2e30ef4c63e2805b8071  62450       31784      False   
4  65eb2e30ef4c63e2805b8074  62453       31784      False   

                                           calendar  \
0  {'ics': 'https://utpl.instructure.com/feeds/ca...   
1  {'ics': 'https://utpl.instructure.com/feeds/ca...   
2  {'ics': 'https://utpl.instructure.com/feeds/ca...   
3  {'ics': 'https://utpl.instructure.com/feeds/ca...   
4  {'ics': 'https://utpl.instructure.com/feeds/ca...   

                                 course_code          created_at default_view  \
0            Introducción a la MaD_AA_24 [5] 2024-02-16 21:40:50         wiki   
1  CARRERA DE EDUCACION QUIMICA Y BIOLO ECTS 2022-10-14 21:03:03         feed   
2            Introducción a la MaD_AA_24 [7] 2024-02-16 21:40:51         wiki   
3           Introducción a la MaD_AA_24 [15] 2024-02-16 21:40:52         wiki   
4           Introducción a la MaD_AA_24 [18] 2024-02-16 21:40:52         wiki   

   enrollment_term_id  hide_final_grades  ...  template     time_zone  \
0                 314               True  ...     False  America/Lima   
1                 314               True  ...     False  America/Lima   
2                 314               True  ...     False  America/Lima   
3                 314               True  ...     False  America/Lima   
4                 314               True  ...     False  America/Lima   

                                       uuid  workflow_state  \
0  eCr8BhjmGPUpXcJLfYBJso1Vnx3ejU4csDRLIAwh       available   
1  bjQOSM069UgKT5II1jT8xbuhwPmAN2dNRpQ8UZbm       available   
2  IaRHvj3jLsOwb3dE1Fo1ceHlHK8htVQWdLiGqoWj       available   
3  o3BqtXwRmlqWLzSeVFGQ8aZ1YwhlhU5itRSP5kw5       available   
4  pmVMq0T9skSjops7CW2XsUzBCy8NCvmKy7rmgpRg       available   

          extracted_at total_students  start_at  grading_standard_id  locale  \
0  2024-05-15 20:01:08            121       NaT                  NaN     NaN   
1  2024-05-15 20:01:08            350       NaT                  NaN     NaN   
2  2024-05-15 20:01:08            108       NaT                  NaN     NaN   
3  2024-05-15 20:01:08            104       NaT                  NaN     NaN   
4  2024-05-15 20:01:08             16       NaT                  NaN     NaN   

   end_at
```

```
0     NaT
1     NaT
2     NaT
3     NaT
4     NaT

[5 rows x 33 columns]
```

Visualización de valores perdidos

```
sns.heatmap(datadb.isnull(), cbar=False)
plt.show()
```



```
datadb.describe()
```

	id	account_id	created_at	enrollment_term_id	root_acco
count	4426.000000	4426.000000	4426	4426.0	
mean	64863.885676	33107.073882	2024-03-23 18:31:05.609127936	314.0	
min	50626.000000	24964.000000	2022-10-14 21:03:03	314.0	
25%	63576.250000	32476.000000	2024-03-11 17:20:14	314.0	
50%	64830.500000	33126.500000	2024-03-25 15:06:25	314.0	
75%	65936.750000	33707.750000	2024-03-25 15:08:00	314.0	
max	67613.000000	34769.000000	2024-05-09 22:53:43	314.0	
std	1472.971819	740.072906	NaN	0.0	

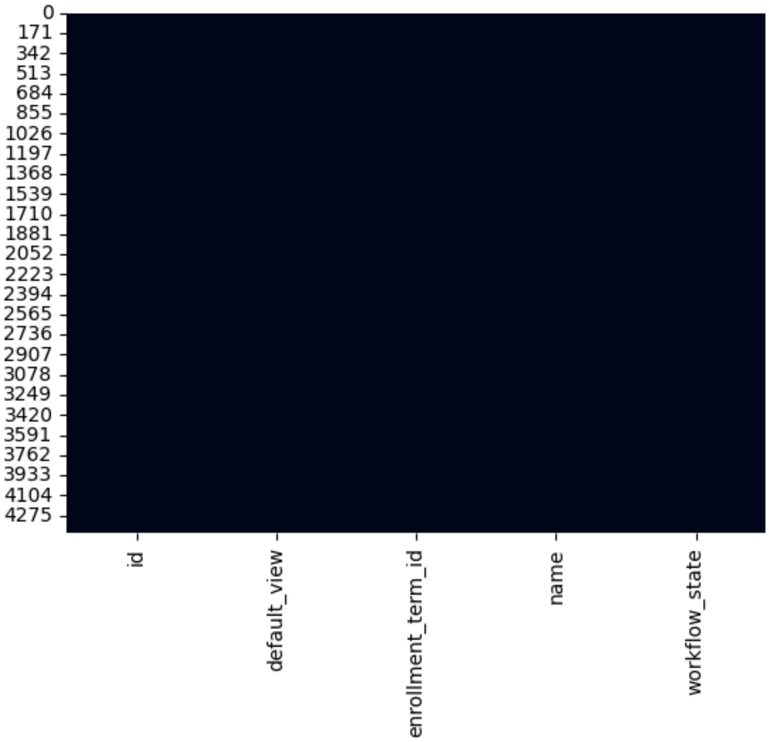
```
datadb.describe(include = ['boolean'])
```

	blueprint	hide_final_grades	homeroom_course	is_public_to_auth_users
count	4426	4426	4426	4426
unique	1	2	1	2
top	False	True	False	False
freq	4426	3510	4426	4361

Transformación de datos

```
columnas_delet = ['account_id','created_at','total_students', 'root_account_id','blueprint','calendar','inserted_at',
datadb1 = datadb.drop(columns=columnas_delet)
```

```
sns.heatmap(datadb1.isnull(), cbar=False)
plt.show()
```



```
datadb1.sample(10)
```

	id	default_view	enrollment_term_id	name	workflow_state
2163	64754	wiki	314	DERECHO SOCIETARIO	available
2512	65130	wiki	314	TECNOLOGIA APLICADA A LA INNOV	available
4184	67235	wiki	314	TUTOR INSTITUCIONAL (Asesor Estudiantil) 005	available
2607	65317	wiki	314	TITULACION 2	available
3339	65956	wiki	314	PRACTICUM 1	available
3756	66629	modules	314	CURSO DE ACTUALIZACION	available

Dataset page_view

```
## Cargar el archivo
ruta = '/content/trabajo-final-grupo3/pages_views.csv'
df_page_view = pd.read_csv(ruta)
df_page_view.head(5)
```

	_id	id	action	app_name	asset_type	asset_use
0	66143fb1ef4c63e280c886b0	ce18e82f-85e2-4410-8f10-7c1fc29bff26	users	Canvas for Android	NaN	
1	66143fb1ef4c63e280c886b1	37f06294-6acb-4a44-8875-487d7c9f391e	users	Canvas for Android	NaN	
2	66143fb1ef4c63e280c886b4	95ec0e4b-1d45-4ea2-9c63-1c495a9b8fa6	users	Canvas for Android	NaN	
3	66143fb1ef4c63e280c886b5	fd1e0dbc-5b15-4cb5-9d75-93b3ed88cb20	NaN	Canvas for Android	NaN	
4	66143fb1ef4c63e280c886b8	cfa38e3f-65fc-4667-951a-f2a51aa8493b	show	NaN	NaN	

5 rows × 29 columns

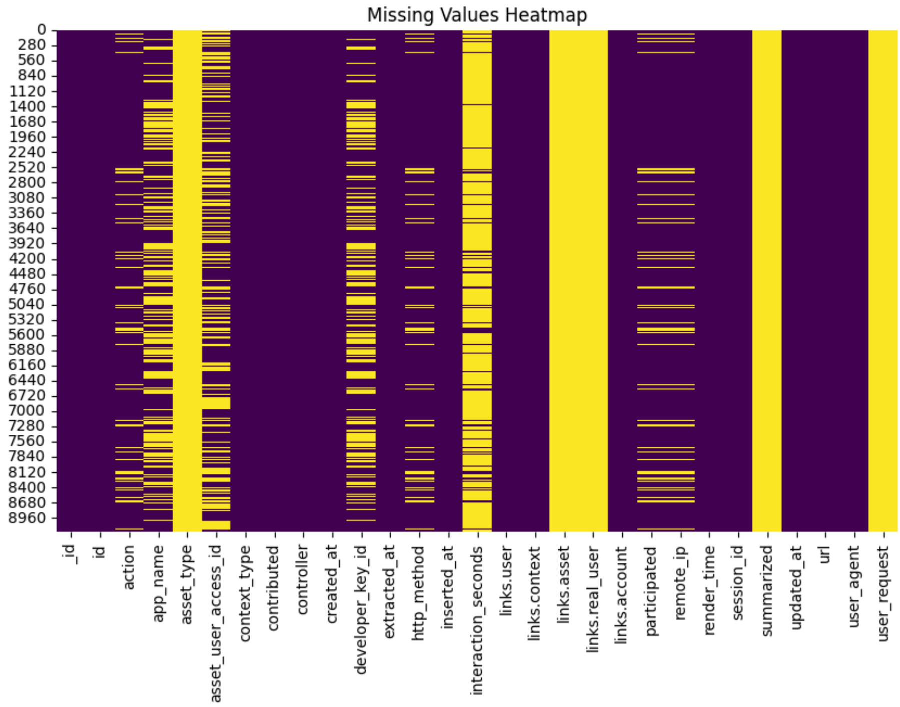
Análisis de datos

- Análisis del dataset: El objetivo de analizar es poder limpiar columnas que no son necesarias, además de tranformar los datos como son el campo created_at que está en formato ISO 8601

```
# Analizamos el total de registros
total_registros = len(df_page_view)
print("Número total de registros:", total_registros)
```

```
Número total de registros: 9214
```

```
# Revisamos el número de valores faltantes para cada columna.
df_page_view.missing.missing_value_heatmap ()
```



```
# Revisamos el porcentaje de valores vacíos por cada columna
df_page_view.missing.missing_variable_summary ()
```

	variable	n_missing	n_cases	pct_missing
0	_id	0	9214	0.000000
1	id	0	9214	0.000000
2	action	1167	9214	12.665509
3	app_name	3358	9214	36.444541
4	asset_type	9214	9214	100.000000
5	asset_user_access_id	3796	9214	41.198177
6	context_type	0	9214	0.000000
7	contributed	0	9214	0.000000
8	controller	0	9214	0.000000
9	created_at	0	9214	0.000000
10	developer_key_id	3358	9214	36.444541
11	extracted_at	0	9214	0.000000
12	http_method	1167	9214	12.665509
13	inserted_at	0	9214	0.000000
14	interaction_seconds	7833	9214	85.011938
15	links.user	0	9214	0.000000
16	links.context	0	9214	0.000000
17	links.asset	9214	9214	100.000000
18	links.real_user	9214	9214	100.000000
19	links.account	0	9214	0.000000
20	participated	1167	9214	12.665509
21	remote_ip	1167	9214	12.665509
22	render_time	0	9214	0.000000
23	session_id	0	9214	0.000000
24	summarized	9214	9214	100.000000
25	updated_at	0	9214	0.000000
26	url	0	9214	0.000000
27	user_agent	0	9214	0.000000
28	user_request	9214	9214	100.000000

Trasnformación de datos

```
# Eliminar columnas que tienen la mayor cantidad de valores vacío y columnas innecesarias
df_page_view_nuevo = df_page_view.drop(columns=['_id', 'action', 'app_name', 'contributed', 'controller', 'asset_type
df_page_view_nuevo.head(5)
```

	id	created_at	interaction_seconds	links.user	links.context
0	ce18e82f-85e2-4410-8f10-7c1fc29bff26	2024-04-08T11:50:55Z	NaN	96271	64325
1	37f06294-6acb-4a44-8875-	2024-04-08T11:50:55Z	NaN	96271	64325

Pasos siguientes: Generar código con `df_page_view_nuevo` Ver gráficos recomendados

```
# Creamos un diccionario de dias de la semana
dias_espanol = {
    'Monday': 'Lunes',
    'Tuesday': 'Martes',
    'Wednesday': 'Miércoles',
    'Thursday': 'Jueves',
    'Friday': 'Viernes',
    'Saturday': 'Sábado',
    'Sunday': 'Domingo'
}


#Función para determinar la jornada
def clasificar_jornada(hora):
  if 6 <= hora.hour < 12:
    return'Mañana'
  elif 12 <= hora.hour < 18:
    return'Tarde'
  else:
    return'Noche'


# Convertir la columna 'created_at' a objetos datetime en UTC
df_page_view_nuevo['created_at'] = pd.to_datetime(df_page_view_nuevo['created_at'], utc=True)

# Definir la zona horaria de Guayaquil
guayaquil_tz = pytz.timezone('America/Guayaquil')

# Convertir la columna 'created_at' a la zona horaria de Guayaquil
df_page_view_nuevo['created_at'] = df_page_view_nuevo['created_at'].dt.tz_convert(guayaquil_tz)

#Se agrega la columna working_day que indica la jornada de acceso: mañana, tarde y noche
#mañana: 06:00 a 12:00
#tarde: 12:01 a 18:00
#noche: 18:01 a 05:59
df_page_view_nuevo['working_day'] = df_page_view_nuevo['created_at'].apply(clasificar_jornada)

#Formatear la columna 'created_at' a una cadena en formato yyyy-mm-dd
df_page_view_nuevo['created_at'] = df_page_view_nuevo['created_at'].dt.strftime('%Y-%m-%d')

#Se agrega la columna created_at_day que indica el día de acceso
df_page_view_nuevo['created_at'] = pd.to_datetime(df_page_view_nuevo['created_at'])
df_page_view_nuevo['created_at_day'] = df_page_view_nuevo['created_at'].dt.day_name()
df_page_view_nuevo['created_at_day'] = df_page_view_nuevo['created_at_day'].map(dias_espanol)

#Se agrega la columna interacion_minutes que indica la interacción en horas
df_page_view_nuevo['interacion_minutes'] = (df_page_view_nuevo['interaction_seconds']/60).round(2)

# Renombrar columnas para poder cruza con el dataset de courses que lo obtenemos desde la base de datos de mongodb at
df_page_view_nuevo.rename(columns={'links.user': 'user_id', 'links.context': 'course_id', 'created_at': 'access_at'},

# Mostrar registros del DataFrame resultante
df_page_view_nuevo.head(5)
```

	id	access_at	interaction_seconds	user_id	course_id	working_da
0	ce18e82f-85e2-4410-8f10-7c1fc29bff26	2024-04-08	NaN	96271	64325	Mañan
1	37f06294-6acb-4a44-8875-487d7c9f391e	2024-04-08	NaN	96271	64325	Mañan
2	95ec0e4b-1d45-4ea2-9c63-1c495a9b8fa6	2024-04-08	NaN	96271	64325	Mañan
3	fd1e0dbc-5b15-4cb5-9d75-	2024-04-08	23.42	96271	64325	Mañan

Pasos siguientes: Generar código con df_page_view_nuevo Ver gráficos recomendados

```
#Agrupar datos por course_id, user_id y created_at para conocer por fecha el número de acceso y los segundos de intera
```

```
#Estas columnas calculadas se agrega:
#- total_access
#- total_interaction_minutes

# Agrupar por 'user_id', 'course_id' y 'created_at'
df_access = df_page_view_nuevo.groupby(['user_id', 'course_id', 'access_at', 'created_at_day', 'working_day']).agg(
    total_access_=('id', 'count'),
    # total_interaction_seconds=('interaction_seconds', 'sum'),
    total_interaction_minutes=('interacion_minutes', 'sum')
).reset_index()


# Mostrar el DataFrame resultante
df_access.head(5)
```

	user_id	course_id	access_at	created_at_day	working_day	total_access_	t
0	5540	64325	2024-04-13	Sábado	Mañana	14	
1	5540	64325	2024-04-13	Sábado	Noche	117	
2	5540	64325	2024-04-13	Sábado	Tarde	49	
3	5540	64325	2024-04-14	Domingo	Mañana	6	

Pasos siguientes: Generar código con df_access Ver gráficos recomendados

Mergue de dataframes para análisis de datos

```
df_mergue = pd.merge(datadb1, df_access, left_on='id', right_on='course_id', how='inner')
```

```
df_mergue.head(5)
# df_mergue.info()
```

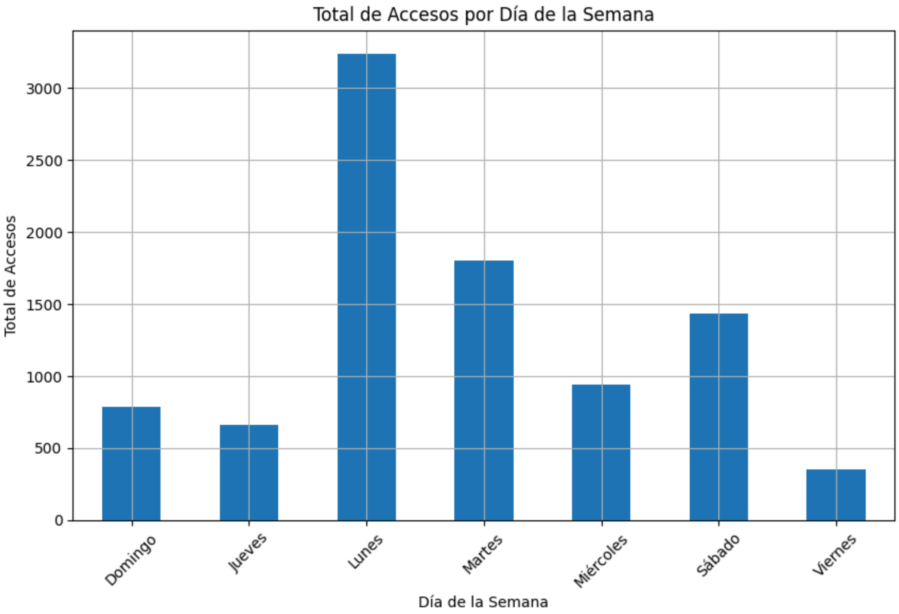
	id	default_view	enrollment_term_id	name	workflow_state	user_id
0	64338	wiki	314	TURISMO Y HOTELERIA	available	5540
1	64338	wiki	314	TURISMO Y HOTELERIA	available	5540
2	64338	wiki	314	TURISMO Y HOTELERIA	available	5540
3	64338	wiki	314	TURISMO Y HOTELERIA	available	5540
4	64338	wiki	314	TURISMO Y HOTELERIA	available	5540

Vizualización de datos con librerias

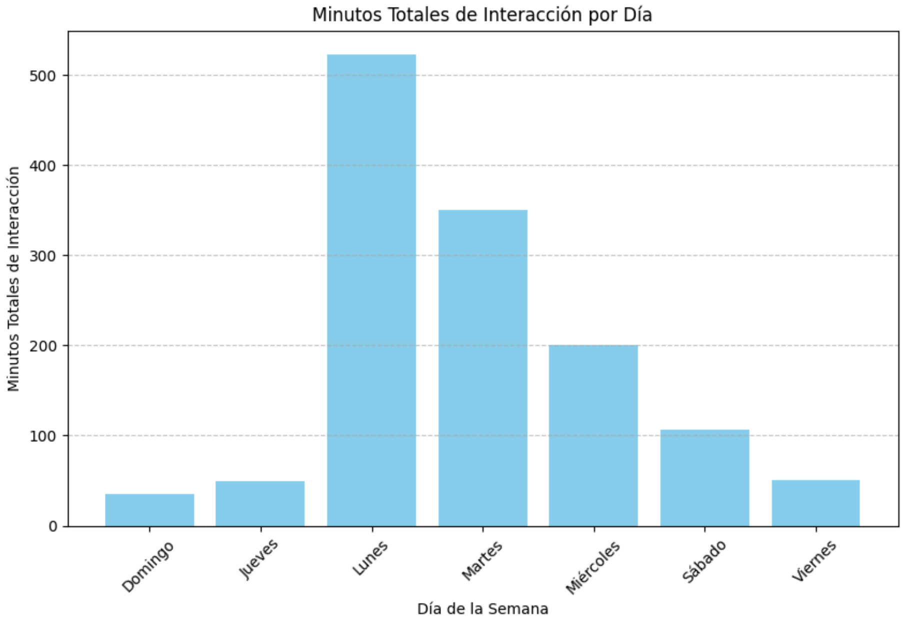
Libreria Matplotlib

```
!pip install matplotlib
```

```
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-packages (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (1.2
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (4.
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (1.
Requirement already satisfied: numpy>=1.20 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (1.25.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (24.0
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (3.1
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->ma
# IMPORTACION DE LIBRERIAS
import matplotlib.pyplot as plt
import pandas as pd


# Gráfico de barras: Total de Accesos por Día de la Semana
plt.figure(figsize=(10, 6))
df_mergue.groupby('created_at_day')['total_access_'].sum().plot(kind='bar')
plt.title('Total de Accesos por Día de la Semana')
plt.xlabel('Día de la Semana')
plt.ylabel('Total de Accesos')
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```



```
# Gráfico de Barras: Minutos Totales de Interacción por Día
grouped_df = df_mergue.groupby('created_at_day')['total_interaction_minutes'].sum().reset_index()
plt.figure(figsize=(10, 6))
plt.bar(grouped_df['created_at_day'], grouped_df['total_interaction_minutes'], color='skyblue')
plt.title('Minutos Totales de Interacción por Día')
plt.xlabel('Día de la Semana')
plt.ylabel('Minutos Totales de Interacción')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
```



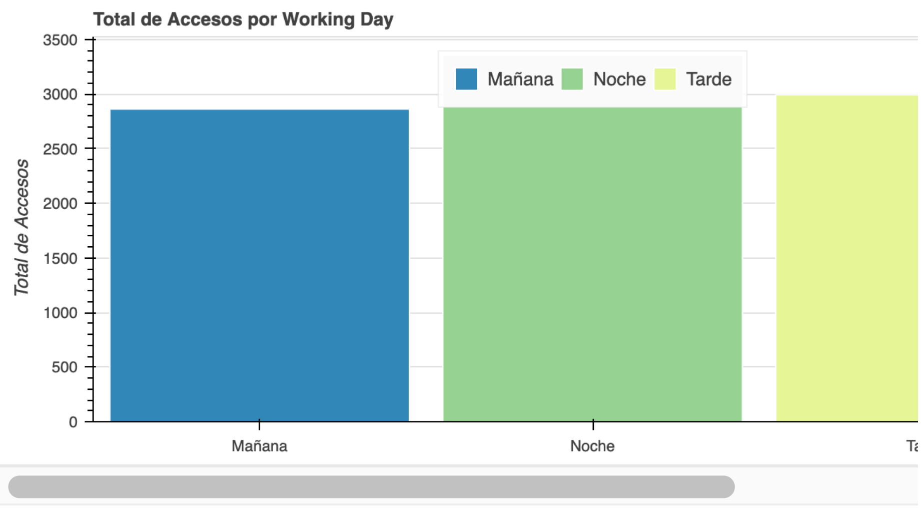
Libreria Bokeh

```
!pip install bokeh
```

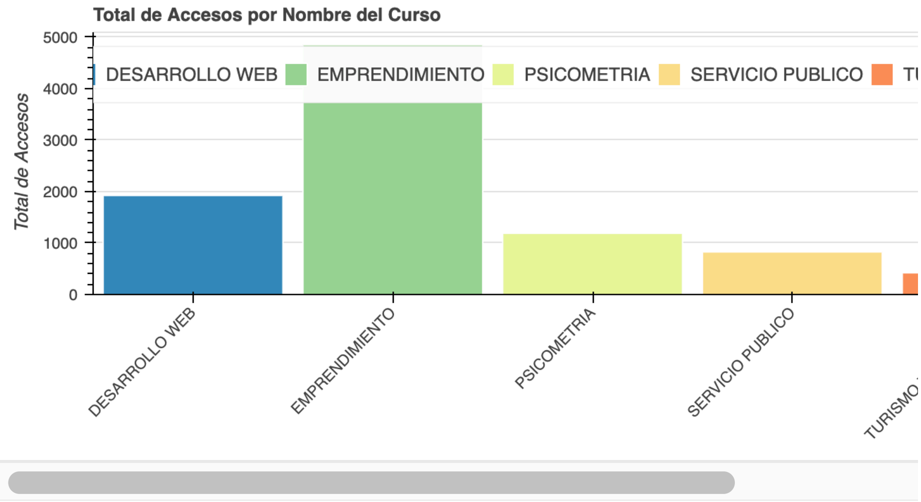
```
Requirement already satisfied: bokeh in /usr/local/lib/python3.10/dist-packages (3.3.4)
Requirement already satisfied: Jinja2>=2.9 in /usr/local/lib/python3.10/dist-packages (from bokeh) (3.1.4)
Requirement already satisfied: contourpy>=1 in /usr/local/lib/python3.10/dist-packages (from bokeh) (1.2.1)
Requirement already satisfied: numpy>=1.16 in /usr/local/lib/python3.10/dist-packages (from bokeh) (1.25.2)
Requirement already satisfied: packaging>=16.8 in /usr/local/lib/python3.10/dist-packages (from bokeh) (24.0)
Requirement already satisfied: pandas>=1.2 in /usr/local/lib/python3.10/dist-packages (from bokeh) (2.0.3)
Requirement already satisfied: pillow>=7.1.0 in /usr/local/lib/python3.10/dist-packages (from bokeh) (9.4.0)
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.10/dist-packages (from bokeh) (6.0.1)
Requirement already satisfied: tornado>=5.1 in /usr/local/lib/python3.10/dist-packages (from bokeh) (6.3.3)
Requirement already satisfied: xyzservices>=2021.09.1 in /usr/local/lib/python3.10/dist-packages (from bokeh) (20
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2>=2.9->boke
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.2->bokeh)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.2->bokeh
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->
```

```
#IMPORTACION DE LIBRERIOAS
from bokeh.plotting import figure, show, output_notebook
from bokeh.models import ColumnDataSource
from bokeh.layouts import column
from bokeh.transform import factor_cmap
from bokeh.palettes import Spectral6
import pandas as pd
```

```
output_notebook()
# Agrupar datos por `working_day`
grouped_df = df_mergue.groupby('working_day')['total_access_'].sum().reset_index()
source = ColumnDataSource(grouped_df)

# Lista de categorías de `working_day`
working_days = list(grouped_df['working_day'])
p = figure(x_range=working_days, title='Total de Accesos por Working Day', height=350, width=800)

# Crear gráfico de barras
p.vbar(x='working_day', top='total_access_', width=0.9, source=source, legend_field="working_day",
       line_color='white', fill_color=factor_cmap('working_day', palette=Spectral6, factors=working_days))

p.xgrid.grid_line_color = None
p.y_range.start = 0
p.xaxis.axis_label = 'Working Day'
p.yaxis.axis_label = 'Total de Accesos'
p.legend.orientation = "horizontal"
p.legend.location = "top_center"

show(p)
```



```
output_notebook()

# Agrupar datos por `name`
grouped_df = df_mergue.groupby('name')['total_access_'].sum().reset_index()

# Crear fuente de datos
source = ColumnDataSource(grouped_df)

# Lista de categorías de `name`
course_names = list(grouped_df['name'])
p = figure(x_range=course_names, title='Total de Accesos por Nombre del Curso', height=350, width=800 )

# Crear gráfico de barras
p.vbar(x='name', top='total_access_', width=0.9, source=source, legend_field="name",
       line_color='white', fill_color=factor_cmap('name', palette=Spectral6, factors=course_names))


p.xgrid.grid_line_color = None
p.y_range.start = 0
p.xaxis.axis_label = 'Nombre del Curso'
p.yaxis.axis_label = 'Total de Accesos'
p.xaxis.major_label_orientation = 0.8
p.legend.orientation = "horizontal"
p.legend.location = "top_center"

show(p)
```

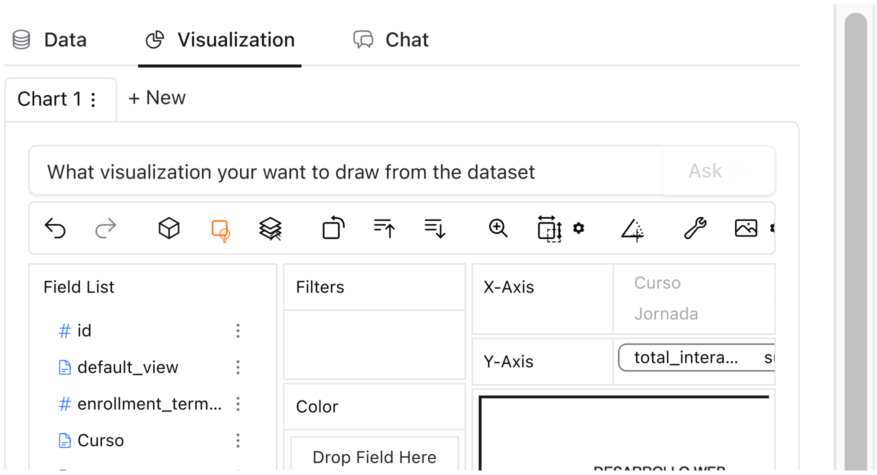


Libreria PyWalker

```
#!pip uninstall sqlglot
#!pip install sqlglot==20.11.0
!pip install pygwalker
```

```
ibis-framework 8.0.0 requires sqlglot<=20.11,>=18.12.0, but you have sqlglot 24.0.1 which is incompatible.
Successfully installed appdirs-1.4.4 arrow-1.3.0 astor-0.8.1 backoff-2.2.1 dateutils-0.6.12 gw-dsl-parser-0.1
```

```
import pygwalker as pyg
walker = pyg.walk(df_mergue)
```



What visualization your want to draw from the dataset

id
default_view
# enrollment_term...
Curso

X-Axis
Curso
Jornada

Y-Axis
total_intera... su

DESARROLLO WEB