



**Campus Querétaro**

**Herramientas computacionales: el arte de la analítica (Gpo 101)**

Semestre: Febrero-Junio de 2024

**Actividad 4 - Usando la librería Pandas**

**Docente**

*Prof. Pedro Oscar Pérez Murueta*

**Equipo**

Edgar Roann Santillán Bernal | A00572737

Entregar: Archivo PDF de la actividad, así como el archivo .ipynb en tu repositorio. Nota: Recuerda habrá una penalización de 50 puntos si la actividad fue entregada fuera de la fecha límite.

1. Subir el archivo `netflix_titles.csv` que se encuentra en el repositorio de la clase y cargarlos en un `DataFrame`.

```
[ ]
from google.colab import drive
drive.mount('/gdrive')

Mounted at /gdrive

[ ] %cd /gdrive/MyDrive/SemanaTec/

/gdrive/MyDrive/SemanaTec

[ ] %ls

credentials.json

[ ]
%mkdir datasets
%cd datasets/

/gdrive/MyDrive/SemanaTec/datasets

[ ] file_path = "/gdrive/MyDrive/SemanaTec/datasets/netflix_titles.csv"
```

```
[ ]
df = pd.read_csv(file_path)

print(df)

  show_id  type  title \
0      s1  TV Show    3%
1      s2  Movie    7:19
2      s3  Movie   23:59
3      s4  Movie     9
4      s5  Movie    21
...    ...    ...
7782   s7783  Movie    Zozo
7783   s7784  Movie   Zubaan
7784   s7785  Movie  Zulu Man in Japan
7785   s7786  TV Show  Zumbo's Just Desserts
7786   s7787  Movie  ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS

  director  cast \
0      NaN  João Miguel, Bianca Comparato, Michel Gomes, R...
1  Jorge Michel Grau  Demián Bichir, Héctor Bonilla, Oscar Serrano, ...
2      Gilbert Chan  Todd Chan, Stella Chung, Henley Hii, Lawrence ...
3      Shane Acker  Elijah Wood, John C. Reilly, Jennifer Connelly...
4      Robert Luketic  Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...
...    ...    ...
7782      Josef Fares  Imad Creidi, Antoinette Turk, Elias Gergi, Car...
7783      Moze Singh  Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...
7784      NaN  Nasty C
7785      NaN  Adriano Zumbo, Rachel Khoo
```

```
[ ] print(df.head(5))

  show_id  type  title  director \
0      s1  TV Show    3%      NaN
1      s2  Movie    7:19  Jorge Michel Grau
2      s3  Movie   23:59    Gilbert Chan
3      s4  Movie     9      Shane Acker
4      s5  Movie    21    Robert Luketic

  cast  country \
0  João Miguel, Bianca Comparato, Michel Gomes, R...  Brazil
1  Demián Bichir, Héctor Bonilla, Oscar Serrano, ...  Mexico
2  Todd Chan, Stella Chung, Henley Hii, Lawrence ...  Singapore
3  Elijah Wood, John C. Reilly, Jennifer Connelly...  United States
4  Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...  United States

  date_added  release_year  rating  duration \
0  August 14, 2020      2020  TV-MA  4 Seasons
1  December 23, 2016      2016  TV-MA    93 min
2  December 20, 2018      2011     R    78 min
3  November 16, 2017      2009  PG-13    80 min
4  January 1, 2020      2008  PG-13   123 min

  listed_in \
0  International TV Shows, TV Dramas, TV Sci-Fi &...
1  Dramas, International Movies
2  Horror Movies, International Movies
3  Action & Adventure, Independent Movies, Sci-Fi...
4  Dramas

  description
0  In a future where the elite inhabit an island ...
1  After a devastating earthquake hits Mexico Cit...
2  When an army recruit is found dead, his fellow...
3  In a postapocalyptic world, rag-doll robots hi...
4  A brilliant group of students become card-coun...
```

3. Implementa una función que despliegue el primer renglon de un `DataFrame`.

```
] def the_first(df):  
  
    if not df.empty:  
        print(df.iloc[0])  
    else:  
        print("El DataFrame está vacío.")  
  
the_first(df)  
  
show_id          s1  
type             TV Show  
title            3%  
director        NaN  
cast             João Miguel, Bianca Comparato, Michel Gomes, R...  
country          Brazil  
date_added       August 14, 2020  
release_year     2020  
rating           TV-MA  
duration         4 Seasons  
listed_in        International TV Shows, TV Dramas, TV Sci-Fi &...  
description      In a future where the elite inhabit an island ...  
Name: 0, dtype: object
```

4. Implementa una función que regrese el número de registros en el `DataFrame`.

```
] def row_number(df):  
  
    num_rows = len(df)  
    return num_rows  
  
row_number(df)
```

7787

5. Implementa una función que despliega las películas con un `release_year` determinado.

```
] def released_year(year):  
  
    movies_in_year = df[df['release_year'] == year]  
  
    if not movies_in_year.empty:  
        print("Películas lanzadas en el año", year, ":")  
        print(movies_in_year)  
    else:  
        print("No se encontraron películas lanzadas en el año", year)  
  
year_query = 2020  
released_year(year_query)
```

6. Implementa una función que despliegue las películas que son de un país determinado.

```
[ ] def origin_country(country):  
  
    movies_from_country = df[df['country'].str.contains(country, case=False, na=False)]  
  
    if not movies_from_country.empty:  
        print("Películas del país", country, ":")  
        print(movies_from_country)  
    else:  
        print("No se encontraron películas del país", country)  
  
country_query = 'Mexico'  
origin_country(country_query)
```

```
Películas del país Mexico :  
   show_id  type      title \  
1         s2  Movie      7:19  
12        s13 TV Show    1994  
146       s147  Movie  A 3 Minute Hug  
237       s238  Movie  A Tale of Two Kitchens  
292       s293  Movie  Acapulco La vida va  
...      ...   ...      ...  
7698     s7699 TV Show    Yankee  
7725     s7726 TV Show  You Cannot Hide  
7734     s7735  Movie  You've Got This  
7773     s7774  Movie  Zoé: Panoramas  
7779     s7780 TV Show    Zona Rosa  
director \
```

7. Crea un nuevo `DataFrame` que sólo contenga las columnas `type`, `title`, `country` y `release_year`.

```
[ ]
new_df = df.loc[:, ['type', 'title', 'country', 'release_year']]

print(new_df.head())
```

	type	title	country	release_year
0	TV Show	3%	Brazil	2020
1	Movie	7:19	Mexico	2016
2	Movie	23:59	Singapore	2011
3	Movie	9	United States	2009
4	Movie	21	United States	2008

8. Despliega, cuantos tipos de producciones (`type`) se hicieron por país determinado (`country`).

```
[ ] def count_production_types_by_country(df, country):

    filtered_df = df[df['country'].str.contains(country, case=False, na=False)]

    production_counts = filtered_df.groupby('type').size()

    print(f"Número de tipos de producciones realizadas en {country}:")
    print(production_counts)

country_query = 'Mexico'
count_production_types_by_country(df, country_query)
```

9. Implementa la función que genera el cuadrado de un número.

```
[ ] def square(x):
    return x ** 2
```

10. Usando el método `apply`, genera el cuadrado de todos los valores del siguiente `DataFrame`.

```
[ ] df2 = pd.DataFrame({'A': [1, 2, 3, 5, 6], 'B': [10, 20, 30, 40, 50]})
df2
```

	A	B
0	1	10
1	2	20
2	3	30
3	5	40
4	6	50

```
[ ]
df2 = pd.DataFrame({'A': [1, 2, 3, 5, 6], 'B': [10, 20, 30, 40, 50]})

def square(x):
    return x ** 2
```

```
[ ]
```

```
df2 = pd.DataFrame({'A': [1, 2, 3, 5, 6], 'B': [10, 20, 30, 40, 50]})
```

```
def square(x):  
    return x ** 2
```

```
df_squared = df2.apply(lambda x: x.apply(square))
```

```
print(df_squared)
```

	A	B
0	1	100
1	4	400
2	9	900
3	25	1600
4	36	2500