

Quiz

April 2, 2020

1 Quiz Pandas

1.1 Patalagua Suárez Nicolás ~ cc. 1073525997

1.1.1 Universidad Sergio Arboleda

1.1.2 Programación científica para IA

1.2 PANDAS

En Computación y Ciencia de datos, pandas es una biblioteca de software escrita como extensión de NumPy para manipulación y análisis de datos para el lenguaje de programación Python. En particular, ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales.

Pandas documentation: <https://pandas.pydata.org/pandas-docs/stable/index.html>

```
[0]: # Importamos la biblioteca Pandas
import pandas as pd
```

```
[24]: # Mostramos un maximo de 40 columnas
pd.options.display.max_columns=40
# Leemos el archivo csv y
movie = pd.read_csv('movie.csv')
# Mostramos la cabecera del documento con 5 filas
movie.head(5)
```

```
[24]:
```

	color	director_name	num_critic_for_reviews	duration	\
0	Color	James Cameron	723.0	178.0	
1	Color	Gore Verbinski	302.0	169.0	
2	Color	Sam Mendes	602.0	148.0	
3	Color	Christopher Nolan	813.0	164.0	
4	NaN	Doug Walker	NaN	NaN	

	director_facebook_likes	actor_3_facebook_likes	actor_2_name	\
0	0.0	855.0	Joel David Moore	
1	563.0	1000.0	Orlando Bloom	
2	0.0	161.0	Rory Kinnear	

3	22000.0	23000.0	Christian Bale
4	131.0	NaN	Rob Walker

	actor_1_facebook_likes	gross	genres \
0	1000.0	760505847.0	Action Adventure Fantasy Sci-Fi
1	40000.0	309404152.0	Action Adventure Fantasy
2	11000.0	200074175.0	Action Adventure Thriller
3	27000.0	448130642.0	Action Thriller
4	131.0	NaN	Documentary

	actor_1_name	movie_title \
0	CCH Pounder	Avatar
1	Johnny Depp	Pirates of the Caribbean: At World's End
2	Christoph Waltz	Spectre
3	Tom Hardy	The Dark Knight Rises
4	Doug Walker	Star Wars: Episode VII - The Force Awakens

	num_voted_users	cast_total_facebook_likes	actor_3_name \
0	886204	4834	Wes Studi
1	471220	48350	Jack Davenport
2	275868	11700	Stephanie Sigman
3	1144337	106759	Joseph Gordon-Levitt
4	8	143	NaN

	facenumber_in_poster	plot_keywords \
0	0.0	avatar future marine native paraplegic
1	0.0	goddess marriage ceremony marriage proposal pi...
2	1.0	bomb espionage sequel spy terrorist
3	0.0	deception imprisonment lawlessness police offi...
4	0.0	NaN

	movie_imdb_link	num_user_for_reviews \
0	http://www.imdb.com/title/tt0499549/?ref_=fn_t...	3054.0
1	http://www.imdb.com/title/tt0449088/?ref_=fn_t...	1238.0
2	http://www.imdb.com/title/tt2379713/?ref_=fn_t...	994.0
3	http://www.imdb.com/title/tt1345836/?ref_=fn_t...	2701.0
4	http://www.imdb.com/title/tt5289954/?ref_=fn_t...	NaN

	language	country	content_rating	budget	title_year \
0	English	USA	PG-13	237000000.0	2009.0
1	English	USA	PG-13	300000000.0	2007.0
2	English	UK	PG-13	245000000.0	2015.0
3	English	USA	PG-13	250000000.0	2012.0
4	NaN	NaN	NaN	NaN	NaN

	actor_2_facebook_likes	imdb_score	aspect_ratio	movie_facebook_likes
0	936.0	7.9	1.78	33000

1	5000.0	7.1	2.35	0
2	393.0	6.8	2.35	85000
3	23000.0	8.5	2.35	164000
4	12.0	7.1	NaN	0

2 Punto 1

Explique brevemente que hace el siguiente código.

```
[25]: #Nucleo principal
movie4=movie[['movie_title','title_year', 'content_rating','budget']]
#Ordenamiento del nucleo respecto a algunas variables
movie4_sorted=movie4.sort_values(['title_year',
    ↳ 'content_rating', 'budget'],ascending=[False,False,True])
#Eliminamos los valores duplicados y mostramos la cabecera
movie4_sorted.drop_duplicates(subset=['title_year', 'content_rating']).head(10)
```

```
[25]:
```

	movie_title	title_year	\
4026	Compadres	2016.0	
4658	Fight to the Finish	2016.0	
4661	Rodeo Girl	2016.0	
3252	The Wailing	2016.0	
4659	Alleluia! The Devil's Carnival	2016.0	
4731	Bizarre	2015.0	
812	The Ridiculous 6	2015.0	
4831	The Gallows	2015.0	
4825	Romantic Schemer	2015.0	
3796	R.L. Stine's Monsterville: The Cabinet of Souls	2015.0	

	content_rating	budget
4026	R	3000000.0
4658	PG-13	150000.0
4661	PG	500000.0
3252	Not Rated	NaN
4659	NaN	500000.0
4731	Unrated	500000.0
812	TV-14	NaN
4831	R	100000.0
4825	PG-13	125000.0
3796	PG	4400000.0

Este código, realiza la creación de un núcleo mediante el cual se van a tomar ciertas columnas del dataframe 'Movie'. Posterior a esto se realiza un ordenamiento basado en los datos de título, rating y budget, dejando los dos primeros de forma descendente y el dato de budget de forma ascendente. Por último retorna las 10 primeras filas de los datos obtenidos.

2.0.1 Métodos usados:

1. `Sort_values`: *Ordena los valores de acuerdo a la clasificación que se requiera.*
2. `Drop_duplicates`: *Ayuda a eliminar duplicados del marco de datos.*
3. `Head`: *Esta función devuelve las primeras n filas para el objeto según la posición. Es útil para probar rápidamente si su objeto tiene el tipo correcto de datos.*

3 Punto 2

Explique brevemente el siguiente código.

```
[26]: # En esta variable guardamos el nombre que le corresponde a cada una de las
      ↪ columnas del dataset movies
new_df=pd.DataFrame(columns=movie.columns)
# Imprimimos la variable new_df que nos retorna el nombre de las columnas del
      ↪ dataset
new_df.head()
```

```
[26]: Empty DataFrame
Columns: [color, director_name, num_critic_for_reviews, duration,
director_facebook_likes, actor_3_facebook_likes, actor_2_name,
actor_1_facebook_likes, gross, genres, actor_1_name, movie_title,
num_voted_users, cast_total_facebook_likes, actor_3_name, facenumber_in_poster,
plot_keywords, movie_imdb_link, num_user_for_reviews, language, country,
content_rating, budget, title_year, actor_2_facebook_likes, imdb_score,
aspect_ratio, movie_facebook_likes]
Index: []
```

```
[27]: # Recorrer el dataset hasta la fila 5(Es decir las primeras cinco películas)
for movie in pd.read_csv('movie.csv',chunksize=5):
    # Con el método groupby realizamos el agrupamiento de los datos
    results=movie.groupby(['movie_title']).count()
    # Concatenamos los datos que han sido agrupados
    new_df=pd.concat([new_df,results])
# Retornamos los valores obtenidos mediante la cabecera
new_df.head()
```

```
[27]:
```

	color	director_name	\
Avatar	1	1	
Pirates of the Caribbean: At World's End	1	1	
Spectre	1	1	
Star Wars: Episode VII - The Force Awakens	0	1	
The Dark Knight Rises	1	1	

	num_critic_for_reviews	duration	\
Avatar	1	1	
Pirates of the Caribbean: At World's End	1	1	
Spectre	1	1	
Star Wars: Episode VII - The Force Awakens	0	0	
The Dark Knight Rises	1	1	

	director_facebook_likes	\
Avatar	1	
Pirates of the Caribbean: At World's End	1	
Spectre	1	
Star Wars: Episode VII - The Force Awakens	1	
The Dark Knight Rises	1	

	actor_3_facebook_likes	\
Avatar	1	
Pirates of the Caribbean: At World's End	1	
Spectre	1	
Star Wars: Episode VII - The Force Awakens	0	
The Dark Knight Rises	1	

	actor_2_name	\
Avatar	1	
Pirates of the Caribbean: At World's End	1	
Spectre	1	
Star Wars: Episode VII - The Force Awakens	1	
The Dark Knight Rises	1	

	actor_1_facebook_likes	gross	\
Avatar	1	1	
Pirates of the Caribbean: At World's End	1	1	
Spectre	1	1	
Star Wars: Episode VII - The Force Awakens	1	0	
The Dark Knight Rises	1	1	

	genres	actor_1_name	movie_title	\
Avatar	1	1	NaN	
Pirates of the Caribbean: At World's End	1	1	NaN	
Spectre	1	1	NaN	
Star Wars: Episode VII - The Force Awakens	1	1	NaN	
The Dark Knight Rises	1	1	NaN	

	num_voted_users	\
Avatar	1	
Pirates of the Caribbean: At World's End	1	
Spectre	1	

Star Wars: Episode VII - The Force Awakens
The Dark Knight Rises

1
1

cast_total_facebook_likes \

Avatar
Pirates of the Caribbean: At World's End
Spectre
Star Wars: Episode VII - The Force Awakens
The Dark Knight Rises

1
1
1
1
1

actor_3_name facenumber_in_poster \

Avatar
Pirates of the Caribbean: At World's End
Spectre
Star Wars: Episode VII - The Force Awakens
The Dark Knight Rises

1 1
1 1
1 1
0 1
1 1

plot_keywords movie_imdb_link \

Avatar
Pirates of the Caribbean: At World's End
Spectre
Star Wars: Episode VII - The Force Awakens
The Dark Knight Rises

1 1
1 1
1 1
0 1
1 1

num_user_for_reviews language \

Avatar
Pirates of the Caribbean: At World's End
Spectre
Star Wars: Episode VII - The Force Awakens
The Dark Knight Rises

1 1
1 1
1 1
0 0
1 1

country content_rating budget \

Avatar
Pirates of the Caribbean: At World's End
Spectre
Star Wars: Episode VII - The Force Awakens
The Dark Knight Rises

1 1 1
1 1 1
1 1 1
0 0 0
1 1 1

title_year actor_2_facebook_likes \

Avatar
Pirates of the Caribbean: At World's End
Spectre
Star Wars: Episode VII - The Force Awakens
The Dark Knight Rises

1 1
1 1
1 1
0 1
1 1

imdb_score aspect_ratio \

Avatar

1 1

Pirates of the Caribbean: At World's End	1	1
Spectre	1	1
Star Wars: Episode VII - The Force Awakens	1	0
The Dark Knight Rises	1	1

	movie_facebook_likes
Avatar	1
Pirates of the Caribbean: At World's End	1
Spectre	1
Star Wars: Episode VII - The Force Awakens	1
The Dark Knight Rises	1

El código anterior, realiza un conteo en las primeras cinco filas del dataset, que corresponden a una película y a través del método `groupby` las agrupa con el dato 'movie_title' que corresponde al título de la película.

3.0.1 Métodos usados:

1. **Read_csv:** *Lea un archivo de valores separados por comas (csv) en DataFrame.*
2. **Groupby:** *Agrupar el DataFrame usando un mapeador o por una serie de columnas.*
3. **Concat:** *Concatena objetos pandas a lo largo de un eje particular con lógica de conjunto opcional a lo largo de los otros ejes.*
4. **Chunksize:** *Se usa para leer un trozo del dataset.*
5. **Count:** *Cuenta las celdas que no son NA para cada columna o fila.*

6. **Head:** *Esta función devuelve las primeras n filas para el objeto según la posición. Es útil para probar rápidamente si su objeto tiene el tipo correcto de datos.*