

Import

```
import pandas as pd
```

Leer un archivo (csv)

```
df = pd.read_csv("../ruta/mi/archivo.csv")
```

Leer ciertos datos

```
df_columns = pd.read_csv("../archivo.csv", usecols=['title', 'critics'])
```

Lee solo las columnas indicadas

```
df['title'][0]
```

Lee el primer valor de la columna

Guardar un csv

```
df.to_csv("copia.csv", sep=";", index=False, header=True)
```

Por defecto hace el guardado de los nombres de las columnas pero no est mal ponerlo y el separador es un añadido extra.

Mostrar los datos

```
df
```

Muestra tanto primeros como ultimos

```
df.head()
```

Muestra los cinco primeros

```
df.tail()
```

Muestra los cinco ultimos

Uso de datos

```
df_selection = df[['title', 'critic_score']]
```

Seleccion de datos .loc

```
df.loc[df.year >= 2018]
```

Seleccionas las pelis que son mayor o igual a 2018

```
df.loc[df.title == 'La pantera', 'critics'] = 100
```

Donde el titulo sea 'La pantera', pone el valor de critics a 100

Seleccion de datos .loc (cont)

```
df.loc[(df.critic_score > 90) & (df.people_score > 80)]
```

Coge las filas en las que supere esos dos valores respectivamente

```
df.loc[df.title.str.startswith('S')]
```

Coge las peliculas que empizen por S mayuscula

```
df.loc[df.critic_score > df.critic_score.median()]
```

Coge todas las que la puntuacion sea mayor a la mediana de la columna

```
df.loc[df.title.str.contains("Spider", na=False), 'critic_score'] = 95
```

Pone el valor de 95 a aquellas pelis que en el titulo tengan "Spider"

Seleccion de datos .iloc

```
df.iloc[0]
```

Coge la primera fila

```
df.iloc[0:3]
```

Coge del 0 al 2

```
df.loc[:9, 'description']
```

Coge desde la 0 hasta la 9 de la columna

```
df.iloc[5:11][['title', 'year', 'critic_score']]
```

Coge del 5 al 10 de esas columnas

```
df.iloc[[1, 2, 3, 5, 8]]
```

Coge esas filas en especifico

Summary Functions and Map

```
df.describe() #
```

Resumen de las columnas numéricas todas

```
df.peliculas.unique()
```

Coge las peliculas una sola vez

```
df.peliculas.value_counts()
```

Cuenta cuantas veces aparece cada cosa

```
df.critics.mean()
```

Devuelve la media de la columna

Hay mas método como el 'std', 'min', 'sum', etc.

Map Function

```
centered_price = reviews.price.map(lambda p: p - mean_price)
```

A cada fila coge el valor de price que es p y le resta la media de la columna

```
n_trop = reviews.description.map(lambda desc: "tropical" in desc).sum()
```

Si tiene esa descripción es True (1) por lo que suma ese uno asi fila por fila

```
df['type'].map(lambda valor: valor.lower())
```

Pone los datos en type a minusculas

Map es una función a la cual se recorre cada valor que le pasemos y hace lo que le pongamos

Apply Function

```
star_ratings = reviews.apply(stars, axis='columns')
```

Hace el metodo 'stars' el cual hace que devuelva un valor a cad fila, dentro del metodo usa reviews.points

```
df['is_highly_rated'] = df.apply(lambda fila: fila['critic_score'] > 90, axis=1)
```

Crea una nueva columna que pone a True si se cumple la condicion

Agg Function

```
critic_score_agg = df['critic_score'].agg(['mean', 'std'])
```

Devuelve el valor de los metodos que pongamos dentro

Métodos

```
.idxmax()
```

Devuelve el dato mayor



By PabloGM

cheatography.com/pablogm/

Not published yet.

Last updated 21st October, 2024.

Page 1 of 1.

Sponsored by [Readable.com](https://readable.com)

Measure your website readability!

<https://readable.com>